UNIVERSITY OF CALIFORNIA SAN DIEGO

Molecular pathology of splicing and transcription factor mutations in myelodysplastic syndromes

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Biology

by

Yi-Jou Huang

Committee in charge:

Professor Dong-Er Zhang, Chair
Professor Rafael Bejar
Professor Xiang-Dong Fu
Professor David Traver
Professor James T Kadonaga

2019

The Dissertation of Yi-Jou Huang is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California San Diego

2019

DEDICATION

To my Family.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Dong-Er Zhang for her support as the chair of my committee, the graduate training environment she offered and the encouragement throughout the course of the PhD. I would also like to thank all the Zhang and Fu lab members for their insightful discussion, resource sharing and caring support during my time in the lab. It would not have been possible without all of them. I thank all my committee members Dr. Rafael Bejar, Dr. Xiang-Dong Fu, Dr. David Traver, Dr. James T Kadonaga. I am very grateful for their valuable input and the mentorship to shape me into a scientist. I also want to thank Dr. Scott Hansen for his supportive guidance and advices in the past few years. Last but not least, I would not be here without the love and support from my parents and sister. I deeply appreciate their understanding and care during my time in graduate school. I also want to thank Hanrui for his company and love and my UCSD biology classmates that are always here to support me in the past six years.

Chapter 1, in part includes reprint of the material as they appear in Molecular and Cellular Biology 2015 (Yukiko Komeno*, Yi-Jou Huang*, Jinsong Qiu, Leo Lin, YiJun Xu, Yu Zhou, Liang Chen, Dora D. Monterroza, Hairi Li, Russell C. DeKelver, Ming Yan, Xiang-Dong Fu, Dong-Er Zhang. “Srsf2 is essential for hematopoiesis and its myelodysplastic syndromes-related mutations dysregulate alternative pre-mRNA splicing.”) and Molecular Cell 2018 (Liang Chen*, Jia-Yu Chen*, Yi-Jou Huang*, Ying Gu, Jinsong Qiu, Hao Qian, Changwei Shao, Xuan Zhang, Jing Hu, Hairi Li, Shunmin He, Yu Zhou, Omar Abdel-Wahab, Dong-Er Zhang, Xiang-Dong Fu. “The augmented R-Loop is a unifying mechanism for myelodysplastic syndromes induced by high-risk splicing factor mutations.”) Yi-Jou Huang was one of the primary investigators and authors of these studies. I particularly like to thank co-author Dr. Yukiko Komeno, Dr. Liang Chen and Dr. Jia-Yu Chen for their guidance and collaborations.

Chapter 2, in full is being prepared for submission for publication by Yi-Jou Huang, Jia-Yu Chen, Ming Yan, Sayuri Miyauchi, Amanda Davis, Rafael Bejar, Liang Chen, Xiang-Dong Fu, and Dong-Er Zhang. "RUNX1 deficiency cooperates with SRSF2 mutation to further disrupt RNA splicing program and exacerbate myelodysplastic syndromes phenotypes." Yi-Jou Huang was the primary investigator and author of this material. I would like to thank Dr. Jia-Yu Chen and Dr. Xiang-Dong Fu for their support and bioinformatics expertise. I am very grateful for Dr. Ming Yan's generous assistance with animal experiments. For the eCLIP library preparation and analysis, I want to thank En-Ching Luo for her help. I would also like to acknowledge UCSD IGM genomics center for preparing the RNA-seq libraries and performing sequencing on the samples.

VITA

2012 Bachelor of Science, National Taiwan University

2019 Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Yukiko Komeno, Ming Yan, Shinobu Matsuura, Kentson Lam, Miao-Chia Lo, **Yi-Jou Huang**, Daniel G Tenen, James R Downing, Dong-Er Zhang. Runx1 exon 6 related alternative splicing isoforms differentially regulate hematopoiesis in mice. Blood. 2014 Jun.123 (24), 3760-3769.

Yukiko Komeno*, **Yi-Jou Huang***, Jingsong Qiu, Leo Lin, YiJun Xu, Yu Zhou, Liang Chen, Dora D Monterroza, Hairi Li, Russell C DeKelver, Ming Yan, Xing-Dong Fu, Dong-Er Zhang. Srsf2 is essential for hematopoiesis and its myelodysplastic syndromes-related mutations dysregulate alternative pre-mRNA splicing. Molecular and Cellular Biology. 2015 Sep. 35 (17), 3071-3082. (*indicates equal contribution)

Liang Chen*, Jia-Yu Chen*, **Yi-Jou Huang***, Ying Gu, Jingsong Qiu, Hao Qian, Changwei Shao, Xuan Zhang, Jing Hu, Hairi Li, Shumin He, Yu Zhou, Omar Abdel-Wahab, Dong-Er Zhang, Xing-Dong Fu. Excessive R-loop formation: A unifying mechanism for splicing factor mutation-Induced myelodysplastic syndromes. Molecular Cell. 2018 Feb. 69 (3), 412-425. e6. (*indicates equal contribution)

Chia-Jen Wu, Yi-Ting Tsai, I-Jung Lee, Ping-Yi Wu, Long-Sheng Lu, Wen-Shan Tsao, **Yi-Jou Huang**, Ching-Cheng Chang, Shuk-Man Ka, Mi-Hua Tao. Combination of radiation and interleukin 12 eradicates large orthotopic hepatocellular carcinoma through modulation of innate and adaptive immune responses. OncoImmunology. 2018 Sep. 7 (9), e1477459

FIELDS OF STUDY

Cancer immunotherapy
Professor Mi-Hua Tao, PhD

Hematology and myeloid malignancies
Professor Dong-Er Zhang, PhD

ABSTRACT OF THE DISSERTATION

Molecular pathology of splicing and transcription factor mutations in myelodysplastic syndromes

by

Yi-Jou Huang

Doctor of Philosophy in Biology

University of California San Diego, 2019

Professor Dong-Er Zhang, Chair

Myelodysplastic Syndromes (MDS) are hematopoietic disorders characterized by myeloid dysplasia, inefficient hematopoiesis, and the propensity to transform into acute myeloid leukemia (AML). Recent advances in sequencing technologies uncovered various types of mutations associated with the disease. In particular, mutations of the splicing factor SRSF2 were significantly associated with mutations of RUNX1, a transcription factor with an established role in hematological disorders. These findings suggest a critical role of these factors in hematopoiesis and disease pathogenesis through their cooperative effects in regulating gene expression and alternative splicing. First, we characterized the roles of SRSF2 mutations in blood cell development and the splicing alternations caused by these mutations. Second, we uncovered that, besides their canonical roles in splicing, SRSF2 mutations also trigger excessive R-loop formation, leading to replication stress and DNA damage response. Third, we discovered the convergent

effect of SRSF2 mutations and RUNX1 mutations in facilitating the development of MDS phenotypes, altering global splicing and gene expression programs, which elucidated the pathways critical for disease development. These findings laid a critical foundation to understand the contribution of individual mutations to MDS and the synergy effects among them and provided insights to develop novel therapeutic strategies in the future.

Chapter 1: The role of SRSF2 mutations in hematopoiesis and myelodysplastic syndromes development

1.1 Summary

The recent genomics studies in patients revealed the frequent presence of mutations in spliceosome genes in hematological disorders, yet their roles in normal hematopoiesis and disease pathogenesis required further investigation. Here we focused on one of the most prevalent mutations found in MDS patients, SRSF2. We characterized the consequences of the SRSF2 mutant expression in murine and human cellular models, the splicing alternations of different SRSF2 mutations and uncovered its novel roles in augmenting R-loop accumulation and DNA damage response. These findings unveil the canonical and non-canonical roles of SRSF2 mutations and expand our understanding of the molecular basis of the disease development.

1.2 Introduction

Myelodysplastic syndromes (MDS) are one of the most common myeloid malignancies, with around 10,000 new cases identified in the United States every year[1, 2]. Advanced age is the pre-dominant risk factor for developing MDS, patients with a median age at diagnosis of 71-76 years[3, 4]. MDS are a heterogenous group of myeloid neoplasms in a pre-leukemia state with ineffective hematopoiesis in pre-leukemia. MDS are characterized as clonal hematopoietic stem cell disorders with peripheral blood cytopenia, impaired myeloid differentiation, dysplasia and up to 30% of the patients further progress to secondary acute myeloid leukemia (AML)[5]. Many patients present symptoms such as fevers and recurrent infections because of neutropenia, fatigue due to anemia and bleeding related to thrombocytopenia. The clinical course is variable among the patients, some patients live for decades with minimal therapy required, while some patients rapidly

transform to AML. Clinically, MDS are categorized into several subtypes based on the karyotypes, lineages affected, degree of cytopenia, percentage of bone marrow blasts and presence of ring sideroblasts[6, 7]. The risk of MDS can be evaluated by several different prognostic systems, while International Prognostic Scoring System (IPSS) and the revised IPSS (IPSS-R) are more widely used lately[8-10]. Although the current systems do not include status of somatic mutations in individual genes, the genetic information is shown to predict patient outcomes independently of each of the scoring systems[11, 12].

Multiple classes of genetic aberrations have been reported as the cause of myelodysplastic syndromes (MDS) and played critical roles in disease development[13, 14], including mutations in transcription factors (RUNX1, ETV6, WT1, GATA2), DNA methylation (TET2, DNMT3A, IDH1, IDH2), chromatin modification (ASXL1, EZH2) and signal transduction (NRAS, KRAS, JAK2, CBL, PTPN11)[11, 15-17]. In 2011, several groups reported that mutations in spliceosome complex genes are highly associated with MDS, chronic myelomonocytic leukemia (CMML) and chronic lymphocytic leukemia (CLL)[18-25]. More recent studies also detected these mutations in AML using targeted sequencing panels with deeper coverage[26]. Mutations in spliceosome genes have been identified in over 60% of MDS/CMML patient, making it as the most prevalent class of genes mutated in MDS[27-29]. Due to their common presence and their roles in 3'splice site recognition, it suggests that defective spliceosome function is critical in MDS pathogenesis. This class of genes includes SRSF2, U2AF1, SF3B1and ZRSR2. Mutations in each of these are considered as early genetic events and occur as heterozygous missense mutations. In addition, mutations in spliceosome genes are mutually exclusive with others, possibly due to their overlapping cellular function and limit cellular tolerance of splicing perturbation[19, 27, 30].

Among them, SRSF2 (also known as SC35) mutations have been consistently associated with poor prognosis and adverse outcomes among MDS and AML patients[23, 27, 31, 32]. SRSF2 mutations occur in around 50% of CMML, 15% of MDS and approximately 20% sAML (secondary acute myeloid leukemia) patients[19, 27]. Most of the SRSF2 mutations occurred at proline 95. Among these mutations, the majority is changed to histidine (P95H), and less frequently changed to leucine (P95L), arginine (P95R) and in-frame deletion of 8 amino acids from P95 to R104 (Δ8aa)[18, 27, 31, 33, 34]. However, how these mutations affect disease development required further investigation.

SRSF2 is one of the founding members for the serine/arginine-rich (SR) protein family, and is generally a positive regulator of exon inclusion and is known to contributes to both constitutive and alternative splicing by binding to exonic splicing enhancer (ESE) sequences within pre-mRNA through its RNA recognition motif domain (RRM)[35-39]. Homozygous germline knockout mice of Srsf2 are embryonic lethal[40]. Analyses of conditional knockout mice of Srsf2 demonstrate various tissue-specific phenotypes including heart, T cells, pituitary gland and embryo fibroblasts[40-42]. To note, the downregulation of Srsf2 in mouse embryonic fibroblasts results in G2/M cell cycle arrest and genomic instability[41]. However, the systemic functional analysis of the function SRSF2 and its mutations in blood cell development remains lacking. Given the strong association of Srsf2 mutations with MDS patients, we further examined their biological functions in hematopoiesis using *in vivo* and *in vitro* models as well as profiled their splicing-dependent consequences by genomics tools in this chapter.

Besides the splicing outcomes caused by SRSF2 mutations, we also explored whether there are other splicing-independent roles of SRSF2 mutations may contribute to MDS disease etiology. One such mechanism might be related to increased genome instability in MDS[43], which

is consistent with increased DNA damage observed in Srsf1 and Srsf2-depleted cells[41, 44, 45]. In context of other splicing factor mutations context, DNA damage accumulation is also observed, including in murine cells overexpressing U2af1 S34F mutations[46] and in primary patient cells harboring SF3B1 mutations[47]. A recent MDS patient study with large cohort reveals a measurable increase in both mutation frequency and diversity during MDS progression to AML[48]. However, since leukemia in general has the lowest mutation rate among all human cancers[49], what would limit massive somatic mutations during the progression of hematopoietic malignancies? An possibility is the co-transcriptional induction of excessive R-loops accumulation[50]. Once formed, R-loops are thought to retard transcription and replication machinery[50-52], and the accumulated R-loops have been linked to genome instability[53, 54]. In this chapter, we further discovered the augmented R-loop formation and DNA damage response caused by SRSF2 mutations using *in vitro* human cell line and mouse models. Together, these findings provide the cellular and molecular explanation for the SRSF2 mutations during disease development.

1.3 Results

1.3.1 SRSF2 is essential for viability of murine primary bone marrow cells *in vitro*

To examine the functional dependency of Srsf2 in primary hematopoietic cells, we first utilized an *in vitro* model by infecting total bone marrow (BM) cells isolated from adult Srsf2+/+ and Srsf2f/f mice with MSCV-Cre-IRES-EYFP (Cre) or empty vector MSCV-IRES-EYFP (EV) retrovirus. Without Cre expression, Srsf2+/+ (+/+ EV) and Srsf2f/f (f/f EV) bone marrow cells did not demonstrated any significant difference in growth. However, in the presence of Cre expression, Srsf2f/f cells (f/f Cre) showed a clear growth disadvantage in liquid culture compared to Srsf2+/+ cells (+/+ Cre) (Figure 1-1A). Moreover, Cre-infected Srsf2f/f (f/f Cre) cells were more apoptotic

compared to other groups (Figure 1-1B). These results demonstrate that Srsf2 is essential for primary murine hematopoietic cell growth and survival *in vitro*.

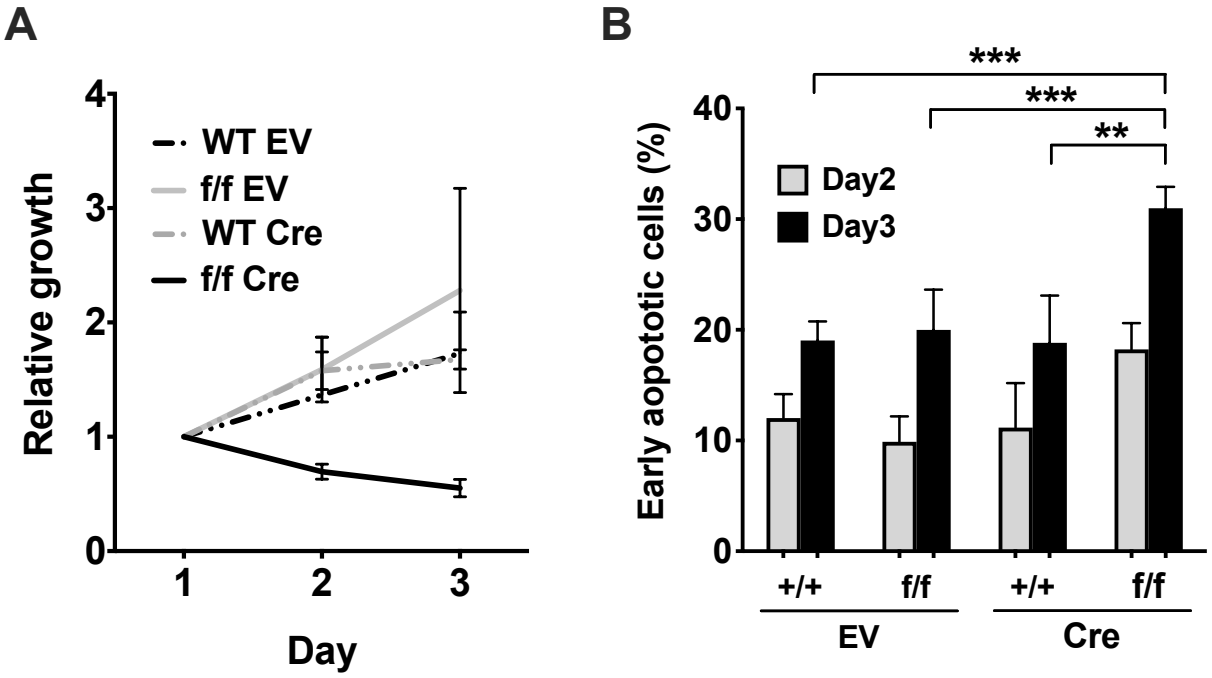


Figure 1-1. SRSF2 is essential for survival for murine adult bone marrow cells *in vitro*. (A) MSCV-IRES-EYFP empty vector (EV) or MSCV-Cre-IRES-EYFP (Cre) retrovirus-infected Srsf2 +/+ and f/f cells were seeded in duplicate, and cell number was counted with Trypan blue staining. The number of retrovirus-infected cells was determined based on EYFP expression. Cre-expressing Srsf2 conditional knockout cells (f/f Cre) demonstrated growth suppression. (B) Apoptosis of EYFP+ cells on day 3. EYFP+ cells were gated, and early apoptotic cells were defined as AnnexinV+ 7AAD- cells. Cre+Srsf2 f/f cells showed significantly enhanced apoptosis. Results of three independent experiments were used for statistical calculations. ** indicates $p < 0.01$; *** indicates $p < 0.001$.

1.3.2 SRSF2 WT and mutant overexpression causes a growth disadvantage in murine primary bone marrow cells

Single allele SRSF2 mutations have been commonly identified in MDS patients[19, 27, 31, 55]. However, our earlier work in lab showed that heterozygous deletion of Srsf2 in blood cells of Vav-iCre+Srsf2f/+ and polyIC-treated Mx1Cre+Srsf2f/+ mice did not demonstrate obvious phenotypes[56]. These results indicate that SRSF2 mutations in MDS are not simply loss-of-function mutations but likely are gain-of-function mutations. We examined the hypothesis by comparing the effect of WT SRSF2, MDS-associated P95H missense mutant and in-frame 8 amino-acid deletion mutant (Δ8aa; P95 to R102 deletion) in primary murine bone marrow cells. MSCV-IRES-puro (MIP) empty vector and MIP-WT, MIP-P95H and MIP-Δ8aa SRSF2 retroviruses were used to infect primary murine bone marrow cells. The exogenously expressed WT, P95H, and Δ8aa SRSF2 proteins are around 4-7 fold above the endogenous SRSF2 level (Figure 1-2A). Liquid culture of infected bone marrow cells showed growth inhibition by both WT and mutant SRSF2 under overexpression conditions (Figure 1-2B). Moreover, the cells expressing the mutants SRSF2 showed a significant increase in apoptosis relative to their WT counterpart (Figure 1-2C). Colony forming assay was used to evaluate the proliferation and colony formation ability of the infected cells. Consistent with the trend showed in liquid culture data, the colony forming assay showed significantly lower colony and cell numbers in WT-infected cells and were even lower in P95H and Δ8aa cells compared to the MIP infected cells (Figure 1-5D). The colony types (GEMM, BFU-E, CFU-G, CFU-GM and CFU-GM), surface markers (Gr-1 and CD11b), and cell morphology were not different among the groups. These results indicated that increased expression of SRSF2 affects cell survival but does not disrupt myeloid cell differentiation. In addition, overexpression

of the P95H and Δ8aa mutants showed stronger negative effects on colony formation ability than the WT SRSF2.

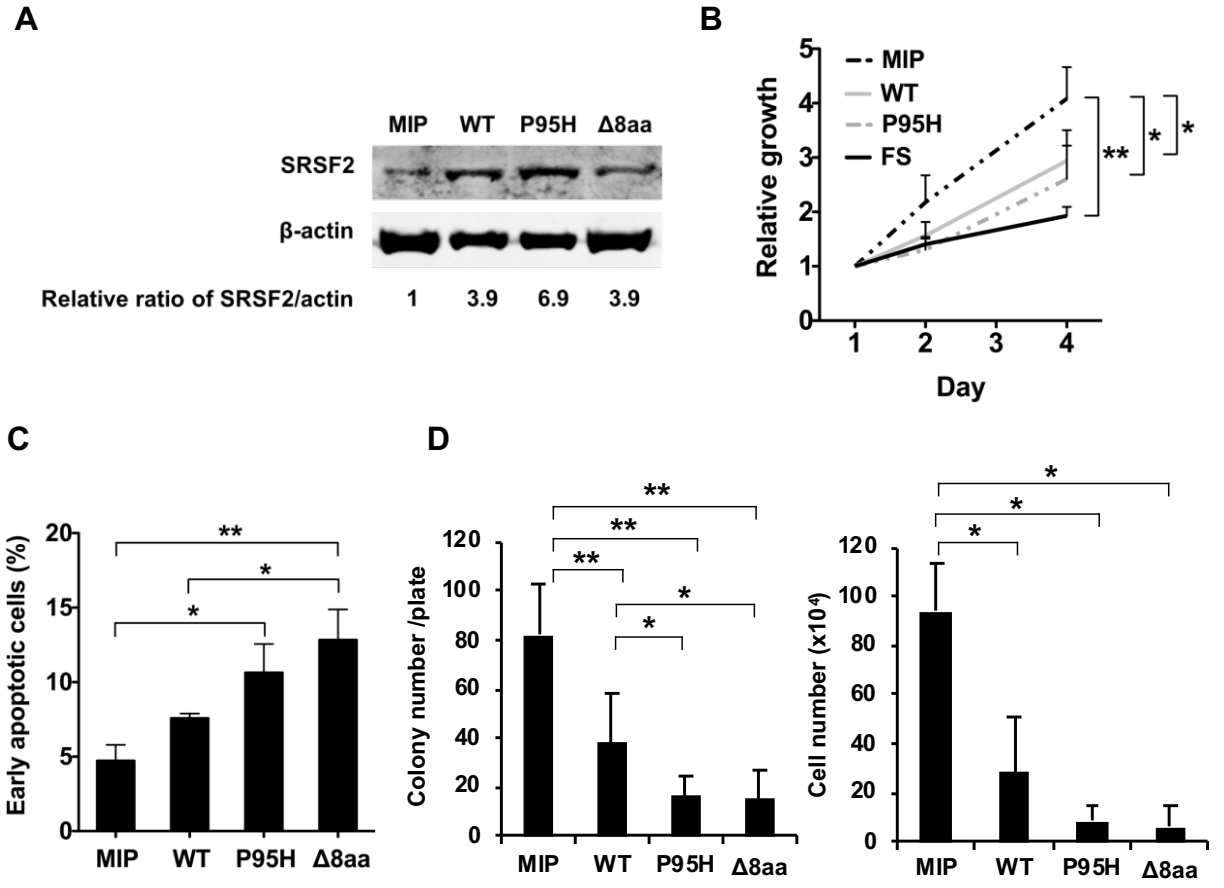


Figure 1-2. Overexpression of WT and mutant SRSF2 in primary murine bone marrow cells demonstrate growth suppression phenotype. (A) Protein expression of the overexpressed SRSF2 in mouse BM cells. β-actin serves as a loading control. Signal intensities of SRSF2 and loading control were used to calculate the relative SRSF2 expression level. (B) Growth curves of total BM cells in liquid culture. All of WT-, P95H- and Δ8aa-expressing retrovirus transduced cells showed growth disadvantage compared to MIP control cells. (C) Apoptosis assay of bone marrow cells on Day 3. P95H and Δ8aa cells showed enhanced apoptosis compared to MIP and WT cells. (D) Colony forming assay of bone marrow cells after one week of seeding in methylcellulose medium. Combined data of three independent experiments are shown. * indicates $p < 0.05$; ** indicates $p < 0.01$.

1.3.3 Low-level expression of SRSF2 WT and mutants caused growth arrest and increased apoptosis in human MDS-L cell lines

Our functional analysis of overexpressing SRSF2 is consistent with previous findings of overexpressed SR proteins interfering with developmental processes[57-59], and the data indicate that studying the mutant function of SRSF2 in cellular systems may enable us to recapitulate SRSF2 single allele mutations context in MDS patients. We next pursue a relatively low-level and inducible expression system instead of using retrovirus overexpression strategy. This expression system allowed lowering endogenous SRSF2 by shRNA and simultaneously mildly expressed the mutant SRSF2 protein in a disease-relevant human MDS cell line (MDS-L)[60, 61]. The cell line was derived from an MDS patient with chromosome 5q deletion but without splicing factor mutations (SRSF2 or U2AF1). The Tet-inducible lentivirus vectors were utilized to co-express an shRNA targeting the 3'-UTR of endogenous SRSF2 and an shRNA-resistant form of SRSF2 cDNA (WT/ P95H/Δ8aa) (Figure 1-3A). MDS-L cells were transduced with SRSF2 shRNA/cDNA lentivirus to establish pools of cells, and puromycin drug selection was performed to establish stable cell lines. After the stable cell lines were constructed, various dosages of Doxycycline (Dox) were added to culture medium in order to optimize an appropriate dosage of shRNA/cDNA induction.

Dox were added into culture medium for 48 hr to induce reduction of endogenous SRSF2 (shown by 3'-UTR expression) while inducing relatively lower level overexpression of total SRSF2 RNA (3 to 4 fold of the endogenous level) [shown by exon 2 coding sequence (CDS)]. The RNA and protein levels of SRSF2 were shown respectively (Figure 1-3B, C). After 48hr of induction, the mutant SRSF2 cells (P95H and Δ8aa) showed significant growth arrest and enhanced apoptosis at Day4 and Day5 post induction compared to WT cells (Figure 1-3D, E). Together, these data indicate that the system closely mimics certain MDS conditions in a disease-

relevant cell line and demonstrated SRSF2 mutants inhibit proliferation through induction of apoptosis.

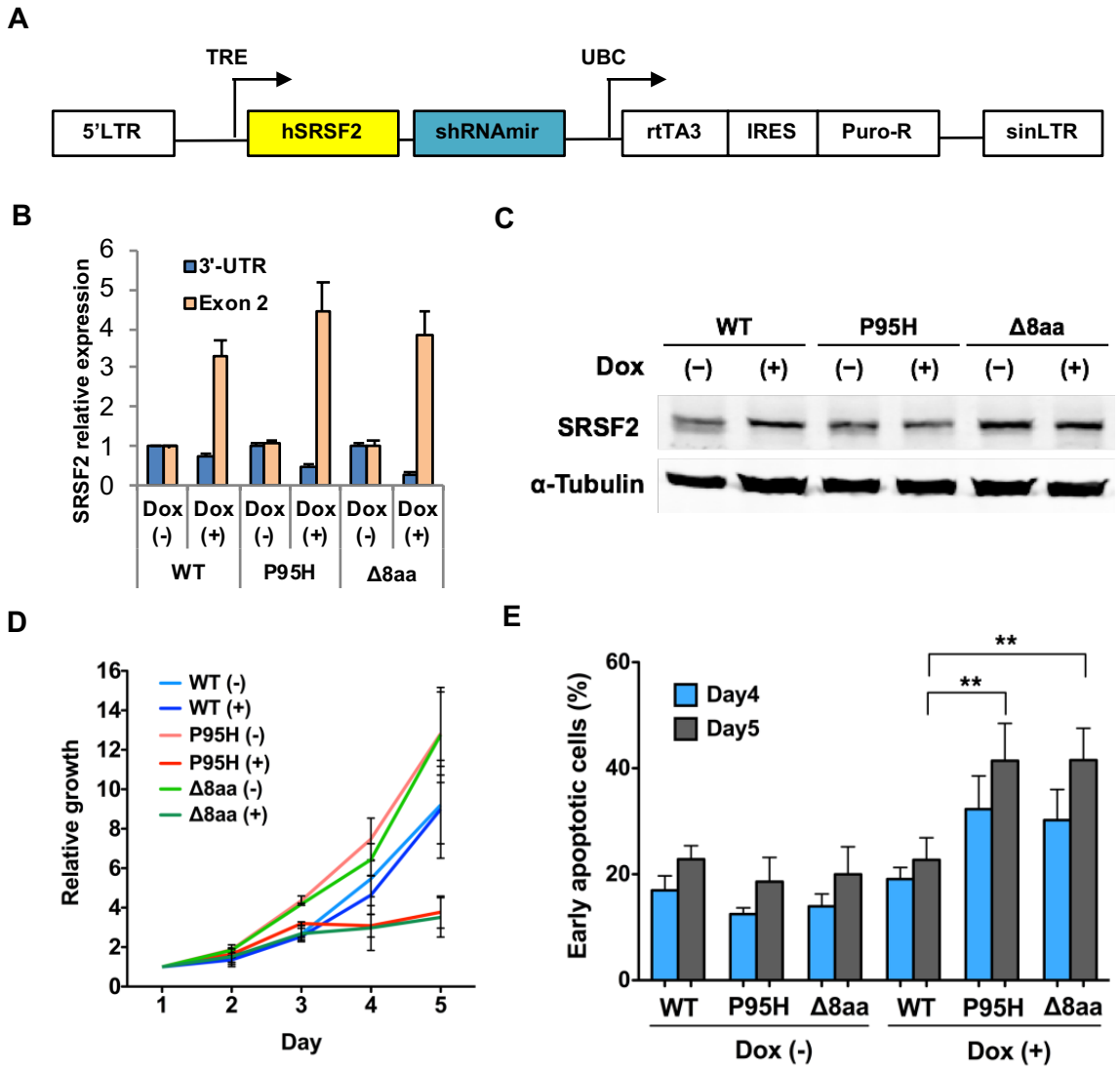


Figure 1-3. Inducible low-level expression of mutant SRSF2 causes apoptosis and growth arrest in human MDS-L cell lines. (A) Structure of pTRIPZ-SRSF2 constructs. Expression of microRNA-adapted shRNA against endogenous SRSF2 (shRNA-mir) and cDNA of shRNA-resistant WT or mutant human SRSF2 is driven by TRE. TRE, tetracycline inducible promoter. UBC, human ubiquitin C promoter. rtTA3, reverse tetracycline transactivator 3. IRES, internal ribosomal entry site. Puro-R, puromycin resistant gene. sinLTR, self-inactivating long terminal repeat. (B) Quantitative RT-PCR of SRSF2. The samples were measured in duplicate and normalized to GAPDH. Expression in untreated WT cells was set to 1. 3'-UTR, endogenous SRSF2 expression. Exon 2 CDS, exon 2 coding sequence representing total SRSF2 expression. (C) Protein expression of SRSF2 in Dox-treated MDS-L cells. α-Tubulin serves as a loading control. (D) The growth curves of the MDS-L cells. Cells were seeded in triplicate in four independent experiments and counted with Trypan blue. Dox-treated P95H and Δ8aa cells showed significant growth suppression compared to other groups. (E) Apoptosis assay. Cells were seeded in duplicate in four independent experiments. Early apoptotic cells are defined as AnnexinV+ 7AAD- cells. Dox-treated P95H and Δ8aa cells showed significantly enhanced apoptosis compared to other groups. ** indicates $p < 0.01$.

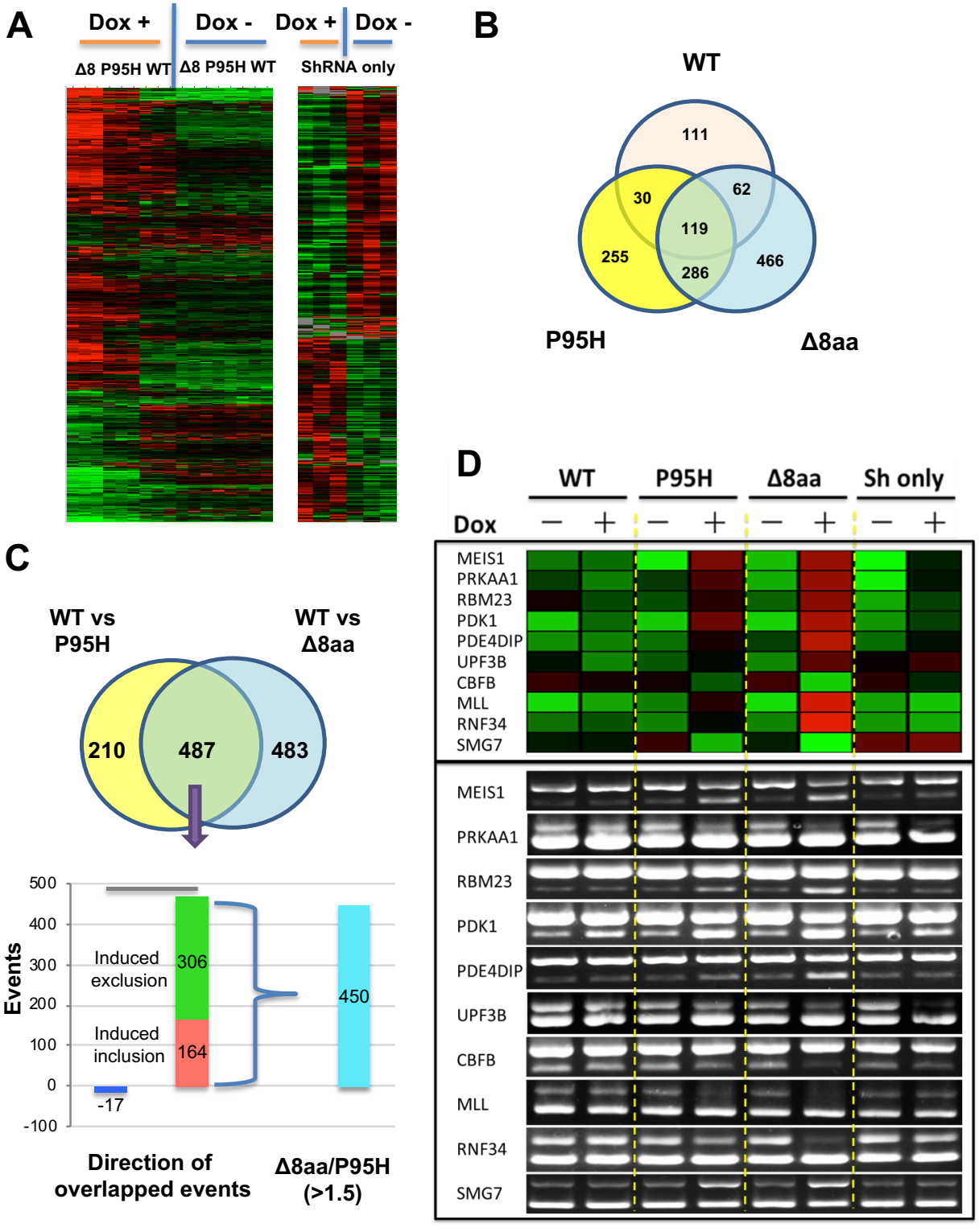
1.3.4 SRSF2 mutants altered a large RNA splicing program in MDS-L cell lines

Although it is anticipated that SRSF2 mutations may perturb splicing program because its critical roles in spliceosome complex, the functional differences in global splicing outcomes between expressing WT and SRSF2 mutants remained unclear. To further reveal the splicing responses caused by SRSF2 mutants which may contribute to MDS disease development, we examined the splicing perturbation by using RASL-seq (RNA-mediated oligonucleotide Annealing, Selection, and Ligation with Next-Gen sequencing) platform. This method was designed to quantitatively profile annotated 5530 alternatively spliced mRNA isoforms in the human genome[62]. MDS-L stable cell lines expressing Tet-inducible SRSF2 shRNA and cDNA of WT, P95H, or Δ8aa SRSF2 were cultured with or without Dox for 3 days and total RNA was isolated afterward from these cells for splicing profiling.

The unsupervised hierarchical clustering analysis show that splicing events of WT, P95H and Δ8aa (Δ8) cells in the absence of Doxycycline (Dox -) are similar and are separated from cells with activated shRNA/SRSF2 expression (Dox +). Δ8aa and P95H mutant expression samples are grouped together, with Δ8aa showing stronger signal (Figure 1-4A). Unsupervised hierarchical clustering of splicing events of ShSRSF2-only cells (without cDNA expression) cultured in the presence or absence of Dox are also shown. Using a stringent cutoff (a fold ratio change >1.5 and p-value <0.05), we detected both overlapped and separated changes in splicing upon expression of WT and mutant SRSF2 (Figure 1-4B), indicating that MDS-associated mutations have a shared function with and an independent effect from WT SRSF2 on splicing. Relative to WT cells, P95H and Δ8aa mutants induced a common set of 487 events, 470 of which were altered in the same directions, including 164 enhanced cassette exon inclusion and 306 enhanced cassette exon exclusion (Figure 1-4C). Importantly, P95H and Δ8aa mutants each induced 210 and 483 extra

events, respectively. Moreover, the change of the scale of exon inclusion or exclusion induced by Δ8aa mutant is over 1.5 times more in 450 out of 470 common events relative to the P95H mutant. Together, these data suggest that even though the two mutants showed certain overlapping effects on splicing, they were not functionally equivalent; the 8-amino acid deletion mutant may have a more profound impact on dysregulated splicing than the P95H mutation. Ten commonly changed splicing events in genes related to myeloid malignancies and hematopoiesis was selected and validated by RT-PCR (Figure 1-4D). Similar to shSRSF2 only, MEIS1, UPF38 and PRKAA1 had increased ratios of short to long isoforms in SRSF2 mutant cells, suggesting reduction of SRSF2 function of SRSF2 mutants in these splicing events. RBM23, PDK1, PDE4DIP, MLL and RNF34 had increased while CBFB and SMG7 had decreased short to long isoform ratios in SRSF2 mutant cells independent of shSRSF2, suggesting the gain of function effect of SRSF2 P95H and Δ8aa mutants in these splicing events.

Figure 1-4. Splicing alternation caused by WT and mutant SRSF2 profiled via RASL-seq analysis. (A) Unsupervised hierarchical clustering show that splicing events of WT, P95H and Δ8aa (Δ8) cells in the absence of Doxycycline (Dox -) and cells with activated shRNA/SRSF2 expression (Dox +) (left panel). Clustering of splicing events of ShSRSF2-only cells cultured in the presence or absence of Dox are also shown (right panel). Green: less short isoform, more long isoform; red: more short isoform, less long isoform. (B) Venn Diagram showing the overlap and difference of splicing events upon Dox activated expression of WT, P95H and Δ8aa SRSF2 compared to shRNA only expression. (C) Venn Diagram showing significant overlapped events (487) from P95H and Δ8aa groups relative to WT SRSF2 group. Among these shared events, over 96% are in the same direction. For these significant events changed in the same direction, Δ8aa induced changes in 450 out of 470 events are over 1.5 fold compared to that in P95H induced events. (D) Validation of 10 selected RASL-seq splicing events. Upper panel: heatmap demonstration of 10 events ratio (short isoform/long isoform). Green: less short isoform, more long isoform; red: more short isoform, less long isoform. Bottom panel: RT-PCR validation of 10 events.


A
Dox +
Dox -
Dox +
Dox -
Δ8 P95H WT
Δ8 P95H WT
ShRNA only
B
WT
111
30
62
119
255
286
466
P95H
Δ8aa
C
WT vs P95H
WT vs Δ8aa
210
487
483
500
400
300
200
100
0
-100
Events
Induced exclusion
306
Induced inclusion
164
-17
450
Direction of overlapped events
Δ8aa/P95H (>1.5)
D
WT
P95H
Δ8aa
Sh only
Dox
− + − + − + − +
MEIS1
PRKAA1
RBM23
PDK1
PDE4DIP
UPF3B
CBFB
MLL
RNF34
SMG7

1.3.5 The genes mis-spliced under SRSF2 mutant condition enriched pathways in MDS pathogenesis

To gain functional insights into such a dramatically altered splicing program, we reasoned that the altered splice events commonly affected by P95H and Δ8aa might be more related to MDS, while extra events induced by each mutant might contribute to enhance the disease phenotype. We thus focused on the commonly affected set of 470 genes by utilizing Ingenuity Pathway Analysis (IPA). Consistent with the functional consequences triggered by the mutations in the aforementioned cellular models, we observed cancer development and apoptosis pathways among the top 10-ranked canonical pathways (Figure 8A). These findings support the possibility that the MDS-associated mutations in SRSF2 promote the development of the disease phenotype with potential to induce a cascade of events that lead to both disease progression and more aggressive types of blood disorders.

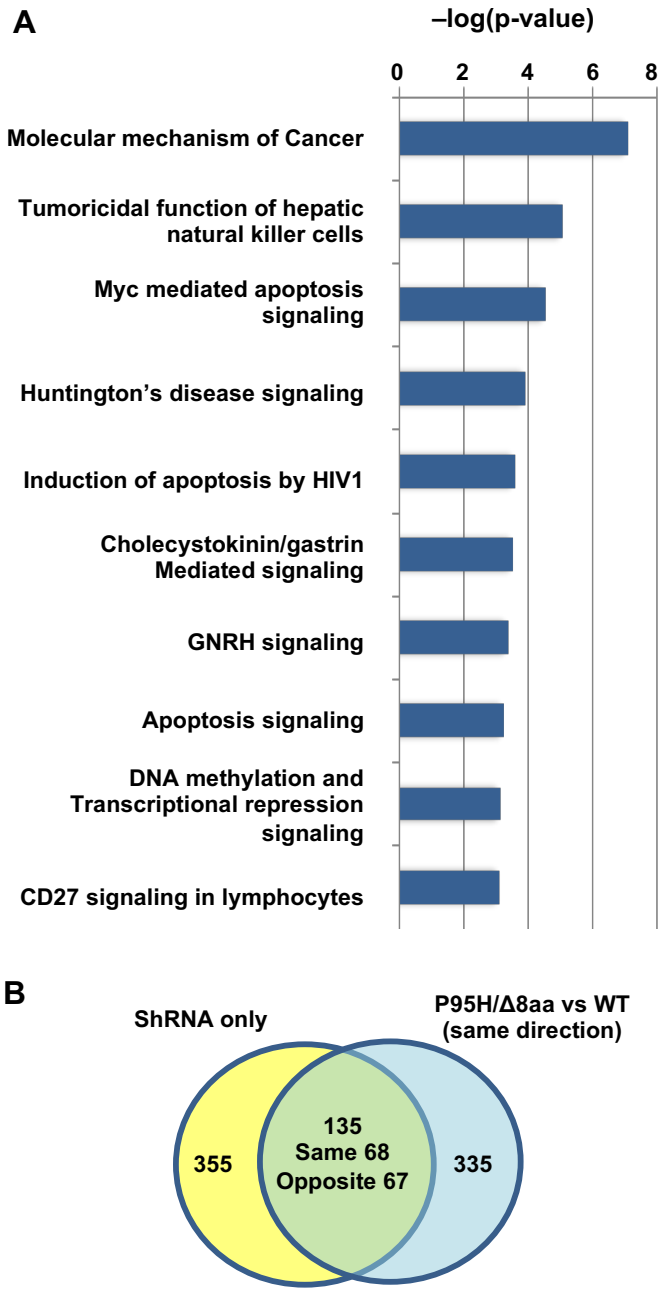


Figure 1-5. Pathway analysis of the target genes of WT and mutant SRSF2. (A) Top biological functions of the 470 common and same direction events of Δ8aa and P95H based on IPA analysis. The –log(p-value) indicates significant level of the pathway. (B) Venn diagram showing significant overlapped events of 470 common targets of P95H and Δ8aa over WT SRSF2 relative to cells expressing only shRNA. Events changed in the same or opposite directions are also listed separately.

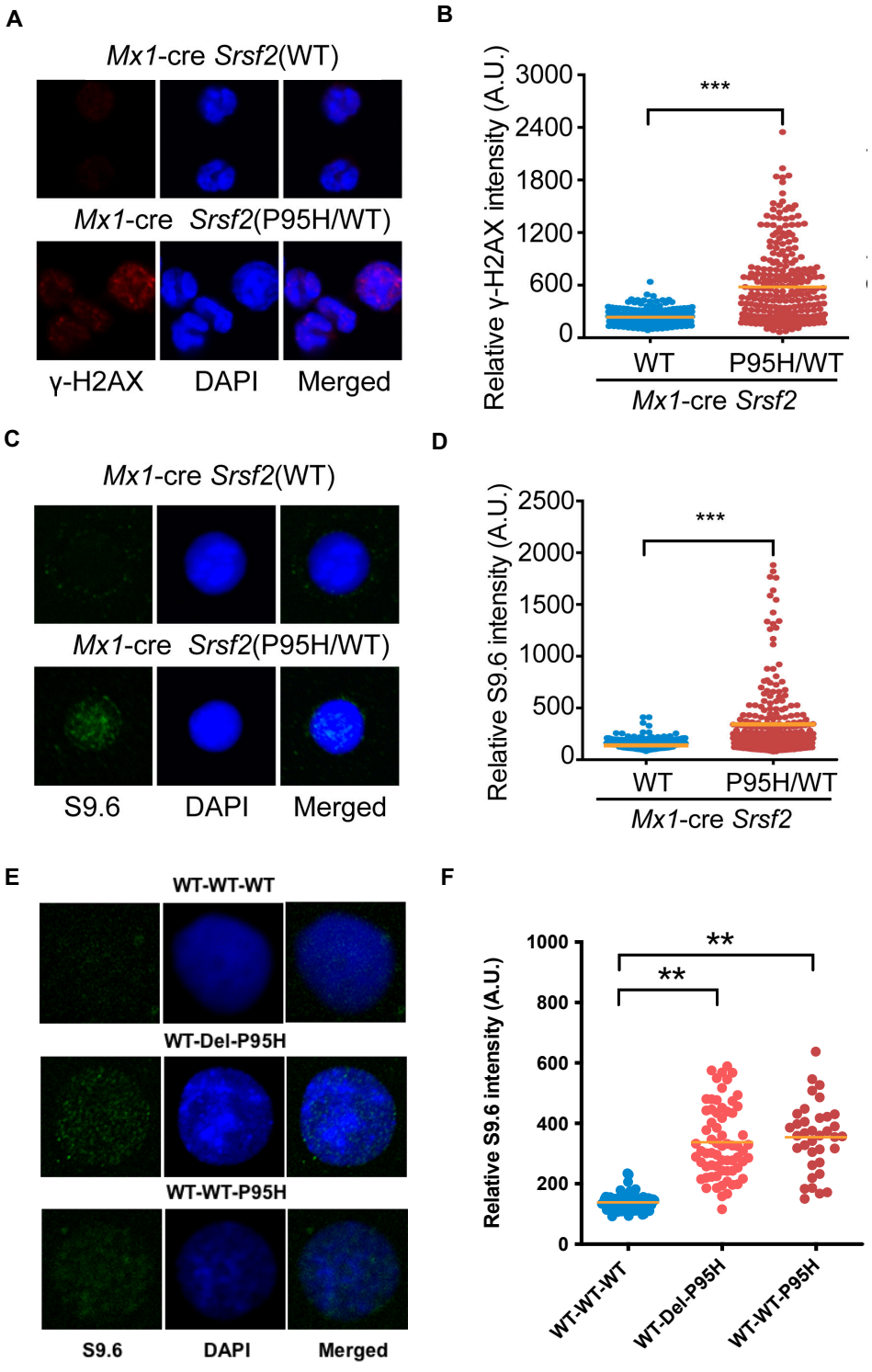
1.3.6 R-loop formation and DNA damage response caused by SRSF2 mutations are observed in murine and human hematopoietic systems

Because various MDS-associated spliceosome mutations caused cell growth inhibition and cell cycle arrest, while the aberrant splicing responses are non-uniformed between different spliceosome mutations and species. Thus, we further explored other general stress that splicing factor mutations could impose to cells. Based on previous studies have shown increased of DNA damage in response to Srsf2 depletion in mouse, SF3B1 mutations in primary patients samples and U2af1 S34F mutation expression in mouse[41, 46, 47], we first examined whether SRSF2 mutations expression also damage the genome in similar manners. Indeed, in 293T cells over-expressing SRSF2 P95H mutations showed elevated DNA damage signals detected by a pan DNA damage marker γ-H2AX[63]. To further validate the findings under non-overexpression condition, in MDS-disease relevant cell type and in heterozygous mutation manner, we next examined the hypothesis using mouse primary bone marrow c-Kit+ cells. We induced the expression of the knock-in mutant allele in the heterozygous Mx1-Cre Srsf2(P95H) knock-in mice by intraperitoneal injection of Poly(I:C)[64]. We next isolated total bone marrow cells, and enriched hematopoietic c-Kit+ progenitor cells for the γ-H2AX immunostaining and confirmed the increased DNA damage response (Fig 1-6 A-B).

Due to the previous findings that depletion of Srsf1 and Srsf2[44, 45, 63] results in increase of R-loops formation and genome instability, we further examined the induction of R-loops in the aforementioned bone marrow c-Kit+ progenitor cells derived from Mx1-Cre Srsf2(P95H) knock-in mice. We confirmed the elevated R-loops by immunostaining with S9.6 antibody (Fig 1-6 C-D) in primary murine cells. Besides murine system, we further seek validation of the R-loop accumulation in human context in order to closely mimic the scenario in patient disease

development. To address our question, we utilized CRISPR-edited K562 human cells harboring SRSF2 P95H mutations[65]. Though primary patient samples could also serve as an ideal source of cells for this examination, the direct effect on R-loop formation caused by the SRSF2 mutation could be better demonstrated under the isogenic mutation conditions. The WT control K562 cells (SRSF2 WT-WT-WT) and two SRSF2 mutant K562 cells bearing missense SRSF2 P95H mutation in one of the three alleles of SRSF2 (SRSF2 WT-Del-P95H; SRSF2 WT-WT-P95H) were used for examination. Both of the SRSF2 P95H mutant lines showed increased R-loop level when compared to SRSF2 WT counterpart (Fig 1-6 E-F).

Figure 1-6. R-loop formation and DNA damage response caused by SRSF2 mutations are observed in hematopoietic systems. (A and C) R-loop signals (A) and DNA damage response (C) detected by immunostaining with S9.6 and g-H2AX, respectively, in isolated bone marrow c-Kit+ murine hematopoietic progenitors from Srsf2(WT) or heterozygous Srsf2(P95H/WT) knockin mice 14 days after poly(I:C) injection. (B and D) The quantitative data of R-loop signals (B) and g-H2AX signals (D) were both based on 3 mice from each (WT or mutant) group. (E) R-loop signals detected by immunostaining with S9.6 in human K562 WT (WT-WT-WT) cells and two SRSF2 mutant (WT-Del-SRSF2; WT-WT-P95H) cells. (F) The quantitative data of R-loop signals in K562 cells. ** $p<0.01$, ***$p < 0.001$.


A
Mx1-cre Srsf2(WT)
Mx1-cre Srsf2(P95H/WT)
γ-H2AX
DAPI
Merged
B
Relative γ-H2AX intensity (A.U.)
3000
2400
1800
1200
600
0
***
WT
P95H/WT
Mx1-cre Srsf2
C
Mx1-cre Srsf2(WT)
Mx1-cre Srsf2(P95H/WT)
S9.6
DAPI
Merged
D
Relative S9.6 intensity (A.U.)
2500
2000
1500
1000
500
0
***
WT
P95H/WT
Mx1-cre Srsf2
E
WT-WT-WT
WT-Del-P95H
WT-WT-P95H
S9.6
DAPI
Merged
F
Relative S9.6 intensity (A.U.)
1000
800
600
400
200
0
**
**
WT-WT-WT
WT-Del-P95H
WT-WT-P95H

1.3.7 The cellular defects induced by SRSF2 mutations can be partially rescue in hematopoietic systems by RNASEH1 overexpression

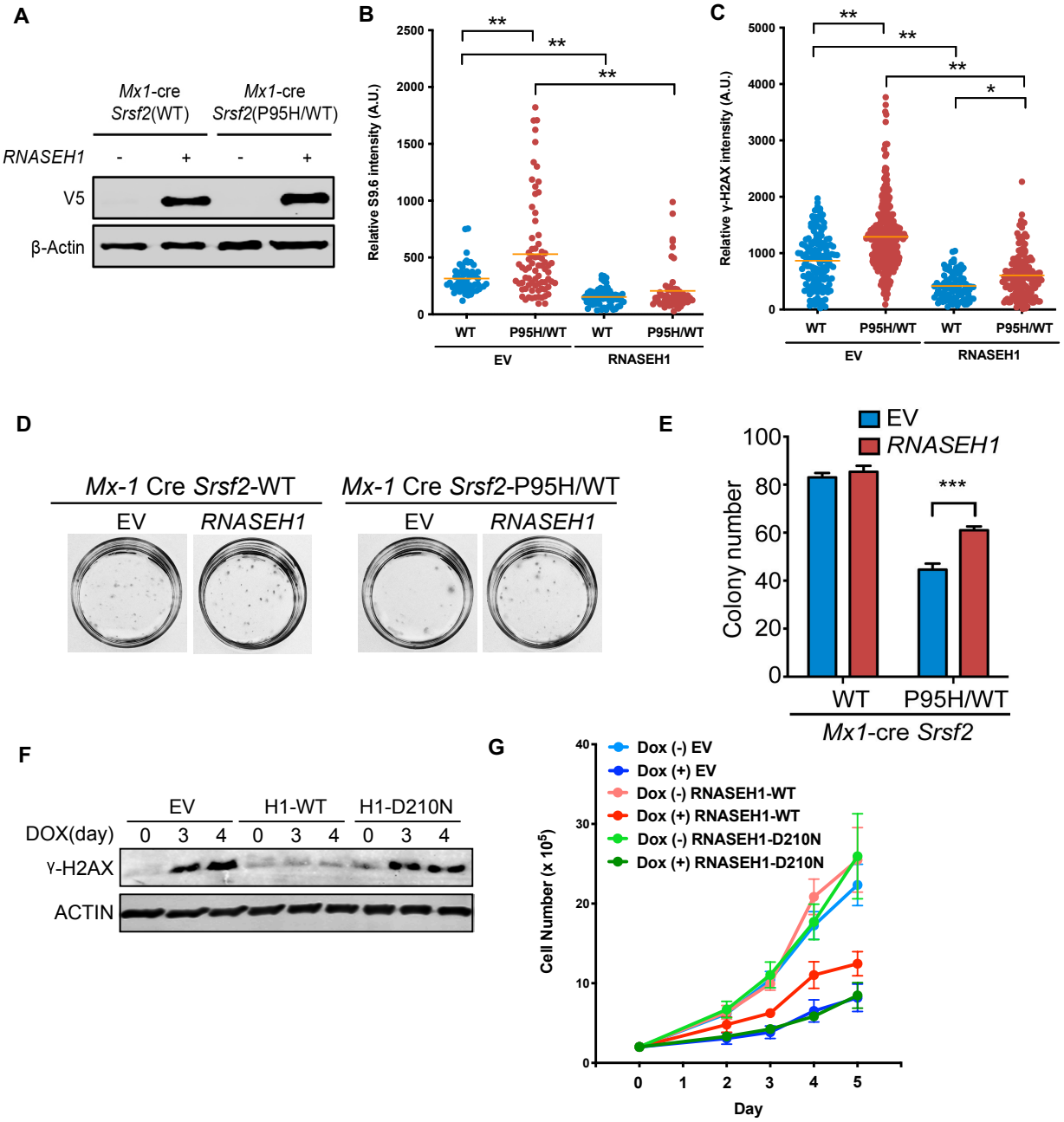
We next focused on addressing whether the observed various cellular defects caused by SRSF2 P95H mutations resulted from induced R-loops. Here we utilized a rescue strategy by suppressing R-loop level to see whether the cellular defects can be attenuated. RNaseH, a conserved endonuclease from bacteria to humans that specifically targets RNA-DNA hybrid, is known to recognize R-loops and performs R-loop resolution[66]. We therefore isolated bone marrow c-Kit+ progenitor cells of the Srsf2(WT) or Srsf2(P95H/WT) mice and infected the cells with MSCV-IRES-EGFP empty vector (EV) and MSCV-V5 tagged RNASEH1-IRES-EGFP (RNASEH1) retroviruses. The RNASEH1 expressing c-Kit+ cells were sorted out for analysis. The protein level of overexpression was validated by Western blotting (Fig 1-7 A). We detected R-loop suppression (Fig 1-7 B) and reduced γ-H2AX level (Fig 1-7 C) in Srsf2 P95H mutant cells overexpressing RNASEH1 when compared to empty vector expression.

These cells were also further examined using colony formation assay for evaluating the hematopoietic competence of early progenitor cells. Consistent with the previous report[64], mutant Srsf2(P95H) impaired the self-renewal potential of isolated early progenitor cells (Fig 1-7 D-E). Importantly, while RNASEH1 overexpression had no obvious effect on the colony forming ability of the wild-type progenitor cells, and the overexpression was able to partially rescue the defect of mutant Srsf2(P95H) cells. The degree of rescue is quite comparable to that observed with re-expressing the Ezh2 gene to compensate for its splicing effect, as reported earlier[64]. To note, although the R-loop level in Srsf2 P95H mutant cells after RNASEH1 expression are decreased to the degree comparable to Srsf2 WT cells (Fig 1-7 B), the level of DNA damage (Fig 1-7 C) was only partially rescued when compared to WT cells under the same condition. These results explain

the partial rescue effect observed in colony formation assay. Together, these data suggest that cell growth defects caused by mutant alleles in Srsf2 might result from a synergy between induced R-loops and other molecular alterations directly or indirectly linked to the altered splicing programs in the cells.

We also confirmed the rescue effect in human systems using MDS-L cell lines. The partial rescue of cell growth defects (Fig 1-7 F) and g-H2AX signals (Fig 1-7 G) in SRSF2 P95H mutant MDS-L cells by overexpressing WT RNASEH1 were observed, while the overexpression of catalytic dead mutant of RNASEH1 (D210N RNASEH1, for which RNA/DNA hybrid cleavage activity is inactivated[67]) was not able to rescue the cellular defects. These findings further re-enforce the concept that the compromised hematopoietic competence in MDS may results from combined molecular defects at multiple levels. Moreover, the data strongly suggest that augmented R-loop dynamics underlies, at least in part, the MDS phenotype, which is likely a common mechanism induced by various mutations in splicing factors and linked to the etiology of MDS.

Figure 1-7. Functional rescue of mutant SRSF2-induced cellular defects in hematopoietic systems by RNASEH1 overexpression. (A) Expression of RNASEH1 detected by Western blotting by using anti-V5 antibody in c-Kit+ progenitor cells isolated from Srsf2(WT) or Srsf2(P95H/WT) mice and transduced with retrovirus expressing the empty vector or RNASEH1. (B and C) R-loop signals (B) and DNA damage response (C) detected by immunostaining with S9.6 and g-H2AX, respectively, in isolated c-Kit+ murine hematopoietic progenitors from Srsf2(WT) or heterozygous Srsf2(P95H/WT) knockin mice after 48hr overexpression of EV or RNASEH1. (D-E) Colony formation assay, (D) Representative photographs showing colonies and quantification of colony numbers (E) of c-Kit+ progenitor cells isolated from Srsf2(WT) or Srsf2(P95H/WT) mice 14 days after overexpression of EV or RNASEH1. (F) Western blot analysis of g-H2AX levels in SRSF2(P95H) MDS-L cells expressing EV or RNASEH1. (G) Growth of MDS-L cells expressing empty vector or RNASEH1 (WT), RNASHE1 catalytic dead mutant (D210N) with or without induced expression of SRSF2(P95H) by Dox treatment for different days. Results of four independent experiments were shown. ** $p<0.01$, ***$p < 0.001$.


A
Mx1-cre Srsf2(WT)
Mx1-cre Srsf2(P95H/WT)
RNASEH1
V5
β-Actin
B
Relative S9.6 intensity (A.U.)
WT
P95H/WT
EV
RNASEH1
C
Relative γ-H2AX intensity (A.U.)
D
Mx-1 Cre Srsf2-WT
Mx-1 Cre Srsf2-P95H/WT
EV
RNASEH1
E
Colony number
Mx1-cre Srsf2
F
DOX(day)
γ-H2AX
ACTIN
H1-WT
H1-D210N
G
Dox (-) EV
Dox (+) EV
Dox (-) RNASEH1-WT
Dox (+) RNASEH1-WT
Dox (-) RNASEH1-D210N
Dox (+) RNASEH1-D210N
Cell Number (x 10^5)
Day

1.4 Discussion

The role of SRSF2 in hematopoiesis using Srsf2 conditional knockout mouse models was elucidated in our lab's earlier studies[56]. Our previous findings shown that SRSF2 plays critical role in embryonic development of hematopoietic system in blood cell-specific Cre transgenic mice[68] (Vav-iCre). The Vav-iCre Srsf2f/f embryos at E13.5 had smaller and paler fetal livers, which contained significantly fewer definitive blood cells and the cells demonstrated enhanced apoptosis and quiescence, which is compatible with the *in vitro* culture data (Figure 1-1). The peripheral blood of Vav-iCre Srsf2f/f embryos had significantly higher numbers of primitive red blood cells characterized by nucleated erythrocytes. Vav-iCre Srsf2f/f embryos died during embryonic development between E16.5 and E18.5 with severe anemia and edema phenotypes. However, unlike previous reports in several other cellular systems, we did not detect any significant delay of cell cycle in Srsf2-deficient blood cells. Interestingly, ablation of Srsf2 in pituitary cells and thymocytes results in decreased organ size but no increase of apoptosis[41, 42]. Developing thymocytes with Srsf2 deficiency showed a differentiation block from CD4/CD8 double-negative to double-positive stage[42]. Ablation of Srsf2 in the embryonic heart showed normal development but later suffered from dilated cardiomyopathy[40]. These different phenotypes of Srsf2 in various cell types may result from alteration of cell type and tissue-specific targets of Srsf2. For example, in thymocytes lacking Srsf2, the aberrant splicing of lymphoid cell-specific CD45 RNA has been related to their abnormal differentiation[41].

Besides the critical roles of Srsf2 in normal embryonic hematopoiesis, we next seek to examine the roles of MDS patient associated SRSF2 mutations to understand how they facilitate disease development. We first compared WT and mutant (P95H, and Δ8aa) SRSF2 by overexpressing them via MSCV retrovirus transduction. With this approach, exogenously

expressed SRSF2 protein is 4 to 7 fold above its endogenous level. Similar inhibitory effects of WT and both mutants on cell growth, colony formation, and *in vivo* blood cell repopulation were detected in these overexpression assays (Figure 1-2 B-D). These results support the hypothesis that WT and P95 mutant SRSF2 have certain similar functions, and such functions above normal levels could disrupt hematopoiesis. We further utilized the Tet-On system to induce co-expression of an SRSF2 shRNA and an shRNA-resistant WT or mutant SRSF2. Under this condition, we are able to mimic patient expression levels in the MDS-L cell line without high overexpression. Under this experimental setting, we detected cell growth inhibitory effects and increased apoptosis of mutant SRSF2, which is more likely due to the unique gain/loss functions of mutant SRSF2.

To compare the mis-splicing consequences caused by SRSF2 mutants, by using the RASL-seq platform, we observed that apoptosis and cancer-related pathways were among the top 10 ranked canonical pathways affected by SRSF2 mutations, which is consistent with pathology of MDS patient cells[11, 13, 14, 16]. MDS cells were noted to be highly apoptotic at early stages, likely reflecting a key mechanism to eliminate cells and eventually become malignant[43]. Thus, mutant forms of SRSF2 in MDS seem to trigger genomic instability, and promote accumulation of unwanted oncogenic mutations[69]. RASL-seq[62, 70] is a convenient way to screen the alternative splicing changes with multiple advantages, including (1) the ability to use limited amounts of total RNA for the analysis (typically from 0.1 to 1 μg of total RNA), (2) considerable tolerance of RNA quality to obtain robust data, (3) the sensitivity to obtain quantitative information on alternative splicing events in less abundant transcripts, and (4) the ability for highly parallel analyses of biological samples in a cost-effective manner. However, compared to RNA-seq, RASL-seq is limited to analysis of annotated alternative splicing events and cannot detect other aberrant RNA processing events. Despite these limitations, we successfully applied this technique to identify a

large cohort of alternative splicing events uniquely caused by mutant SRSF2, suggesting that the mutant splicing factor has the capacity to dramatically alter the splicing program in diseased cells.

It is striking that SRSF2 mutations in MDS mainly occur at proline 95 (P95), suggesting a critical role of P95 in SRSF2. SRSF2 has an N-terminal RNA recognition (RRM) domain and a C-terminal arginine and serine rich (RS) domain. The RRM domain is primarily involved in specific RNA recognition. The RS domain can be highly phosphorylated and is involved in both RNA binding and protein-protein interactions during spliceosome assembly[35]. P95 is generally believed to be located in the hinge region between the RRM domain and the RS domain. A recent structural analysis of the N-terminal 101 amino acids of SRSF2 bound to 6-oligonucleotide RNA targets showed that P95 directly contacts the C3 and G3 nucleotides in RNA containing the UCCAGU and UGGAGU motifs, which represent two consensus SRSF2 binding sequences[71]. It is possible that SRSF2 with mutations at P95 decrease the RNA binding specificity. This possibility is further supported by other report using Srsf2 P95H knock-in mice model and MDS patient samples, which show that unlike the WT SRSF2, the P95H mutant preferentially binds to CCAG motif[64]. Furthermore, the unique features of proline among the 20 protein-forming amino acids, including its cis-trans isomerization[72], CH/Pi hydrogen bond formation[73], and higher conformational rigidity in the secondary structure of proteins[72], may result in a different local structure to create a new surface for interaction with additional proteins and RNA sequences. These features are subject to future biochemical studies to provide deeper understand of SRSF2 mutations.

MDS is a highly heterogeneous disease that causes dysplasia of hematopoietic progenitor cells, implying that multiple mechanisms may underlie the disease progression. Recent identification of common mutations in several splicing factors, leading to research focusing on the aberrant splicing consequences induced by individual splicing factor mutations as potential causes

for the disease. To note, each splicing factor mutation appears to affect a largely distinct set of splicing targets, which raises the question of whether and how multiple independent splicing changes all cause a similar disease phenotype.

Besides the splicing perturbations caused by SRSF2 mutations, we next explore whether there are other common mechanisms that splicing mutations could contribute to MDS disease etiology. Interestingly, the increased genome instability in MDS[43, 48, 69, 74] and increased DNA damage in Srsf2-depleted cells[41] were reported in previous studies and raise the possibility of this angle. We discovered the increased R-loop and DNA damage signal associated with SRSF2 mutations, and the growth defects caused by SRSF2 mutations could be partially compensated under R-loop suppression conditions, suggesting a causative effect of the R-loop accumulation and the disease phenotypes. Fu lab also identified the activation of ATR pathway induced by SRSF2 and U2AF1 mutations but not ATM pathway[63], suggesting a more general mechanism for compromised hematopoietic competence, which is the activation of the ATR pathway that leads to cell cycle arrest, which agrees with the observation that the ATR pathway is selectively activated in high-risk MDS patients and that the ATR inhibitor, but not the ATM inhibitor, improves the differentiation of telomere dysfunctional common myeloid progenitor cells[74]. Moreover, the impaired transcription of two genes, CLSPN and PMS2, were found to be associated with SRSF2 mutations as well, and both of which have been implicated in DNA damage repair[63, 75, 76], suggesting that additional synergistic events may contribute to the induced DNA damage. Lately, other splicing mutations studies in U2AF1 S34F and SF3B1 K700E also demonstrate increased of R-loops and DNA damage[77, 78], which is aligned with our findings.

However, the preferred activation of the ATR pathway by splicing factor mutation is in contrast to many cases in solid tumors where more severe genome instability because of DSBs

leads to the activation of both the ATR and ATM pathways, which may accelerate the accumulation of additional oncogenic mutations[79, 80]. Instead, altered transcription and splicing programs may initially compromise cell fitness in general, generating constant pressure for affected cells to develop and accumulate compensatory changes that may ultimately become oncogenic. In this regard, augmented R-loop formation may represent a form of such selection pressure in diseased cells. To compensate for the activation of a cell cycle checkpoint, MDS cells may evolve various mechanisms to overcome such a checkpoint to become proliferative, which may not always be necessarily accomplished through the acquisition of new mutations in the genome. This points to a future effort in identifying genes whose altered expression may bypass R-loop-induced checkpoints.

In conclusion, we demonstrate that SRSF2 is essential for the survival of hematopoietic cells and that its loss results in apoptosis and growth suppression *in vitro*. The growth arrest and increased apoptosis were also clearly detected in cells expressing either P95H or Δ8aa mutants. Together, these results support that MDS-related SRSF2 mutations are not simply loss-of-function mutations. Furthermore, RASL-seq analysis demonstrates that MDS mutant forms of SRSF2 dysregulate RNA splicing. On top of the dysregulated splicing outcomes, we demonstrate the increase genome instability caused by SRSF2 mutations due to excessive R-loop. These alternations caused by SRSF2 mutations might further create synergy when two causal mutations co-exist in the same cell, giving rise to pressing stress to the genome and, thus, a more severe disease phenotype. Future studies are needed to further examine how mutations in SRSF2 may act in synergy with mutations in other key leukemia genes, such as TET2, IDH2, and RUNX1, to propel MDS progression though augmented splicing and transcription dysregulation to yield more aggressive blood cell disorders.

1.5 Materials and Methods

Mice

C57BL/6 (CD45.2) mice were obtained from Jackson Laboratory. Conditional Srsf2f/f mice[41, 42] and conditional Srsf2 P95H/WT mice were described previously[64]. Polyinosine-polycytosine (pIpC) (Sigma) was injected intraperitoneally (i.p.) to mice at 12μg/g every other day for three injections. Genotyping PCR was performed using the primers described previously. All the procedures were approved by the institutional animal care and use committee.

DNA Constructs

C-terminally HA-tagged human SRSF2 was subcloned in the EcoRI site of MSCV-IRES-GFP (MigR1), MSCV-IRES-puro (MIP), and pREV-TRE[41]. P95H and Δ8aa mutants were made by PCR mutagenesis. Primers used for mutagenesis are as follows.

P95H(F):5'-CAAATGGCGCGCTACGGCCGCCACCCGGACTCACACCACAGCCGC-3'

P95H(R): 5'-GCGGCTGTGGTGTGAGTCCGGGTGGCGGCCGTAGCGCGCCATTTG-3'

Δ8aa(F):5'-

GGGTGCAAATGGCGCGCTACGGCCGCCGGGGACCGCCACCCCGCAGGTACGGGG-3'

Δ8aa(R):5'-

CCCCGTACCTGCGGGGTGGCGGTCCCCGGCGGCCGTAGCGCGCCATTTGCACCCG-3'

pTRIPZ-SRSF2 constructs were made using shRNA against the 3'-UTR of human WT SRSF2 (CTCTCCCGATTGCTCCTGTGTA) and human SRSF2 cDNA sequences with or without mutations.

Human RNASEH1 CDS sequence tagged with the nuclear localization signal (NLS) at N terminus and V5 sequence at C terminus was cloned into the ppyCAG expression vector, which was kindly provided by Dr. Juan Calos Izpisua Belmonte. Site-directed mutagenesis was

performed to generate the D210N mutant form of RNASEH1, for which RNA/DNA hybrid cleavage activity is inactivated[67]. To overexpress RNASEH1 in MDS-L and mouse cells, C-terminally V5 tagged human full length RNASEH1 was cloned into the EcoRI and XhoI sites of the MSCV-IRES-GFP (MigRI) vector. Retrovirus production and infection procedure were performed as previously described[81].

Cell culture

293T cells and total BM cells were cultured as described before[81]. MDS-L cells[60, 61] were provided by Dr. Daniel Starczynowski, and cultured in RPMI supplemented with 10% fetal bovine serum, 10 mM glutamine, penicillin (100 IU)/streptomycin (100 μg/ml) and 10 ng/ml human IL-3 (hIL-3) (Peprotech). To induce the expression of shRNA and exogenous SRSF2 in MDS-L cell lines transduced with pTRIPZ vectors, 1 μg/ml Doxycycline (Sigma) was added every day to the cells for three days. To reach 50% SRSF2 expression in shRNA only cells, 2.5 μg/ml Doxycycline was used in the assay. Cell growth was evaluated in duplicate by the Trypan blue exclusion assay. As for colony forming unit (CFU) assays, specific number of cells (described in figure legends) were seeded into M3434 (STEMCELL Technologies). One week later, colony number was counted. CRISPR edited K562 cells were provided by Dr. Rafael Bejar[65] and cultured in RPMI with 10% fetal bovine serum, 10 mM glutamine and penicillin (100 IU)/streptomycin (100 μg/ml).

Flow cytometry

Primary bone marrow cells from mice were first treated with ACK buffer at room temperature for 5 minutes, then washed with PBS before antibody staining. For the apoptosis assay, Annexin V-APC (BioLegend) and 7AAD (BioLegend) were used as described by manufacturer's protocol. Data were collected on BD FACSCanto and analyzed by FACSDiva (BD) software.

Reverse transcription (RT) and quantitative polymerase chain reaction (qPCR)

Total RNA was extracted by using Trizol reagent (ThermoFisher Scientific, #15596026) according to manufacturer's instructions and treated with DNase I (Qiagen). RT reactions were carried out by using the qScript cDNA Synthesis Kit (Quanta Biosciences). qPCR was performed with the SYBR FAST qPCR Kit (KAPA Biosystems). For validation of the results of RASL-seq, OneStep RT-PCR kit (Qiagen) was used. Primers used for RT-PCR and qPCR are as follows.

mSrsf2(F): CGCGCTCCAGATCAACCTC.

mSrsf2(R): CTTGGACTCTCGCTTCGACAC.

mGAPDH(F): GGTGCTGAGTATGTCGTGGAGTCTA.

mGAPDH(R): AAAGTTGTCATGGATGACCTTGG.

hSRSF2 3'-UTR(F): GCACTAGGCGCAGTTGTGTA.

hSRSF2 3'-UTR(R): CAATCGGGAGAAAACAGGAA.

hSRSF2 Exon2 CDS(F): CTACAGCCGCTCGAAGTCTC.

hSRSF2 Exon2 CDS(R): TTGGATTCCCTCTTGGACAC.

hGAPDH(F): TCGCTCAGACACCATGGGGAAG.

hGAPDH(R): GCCTTGACGGTGCCATGGAATTTG.

Western blotting

Cell extracts from mouse bone marrow were prepared with Thermo Scientific Pierce IP Lysis Buffer (PI-87787) including the protease inhibitor cocktail (PI-88665) and phosphatase inhibitor cocktail (PI-88667). Western blotting was performed following standard procedures. Protein samples were denatured in 1× loading buffer [10% glycerol, 2% SDS, 10 mM DTT and 50 mM Tris-HCl (pH 6.8)]. Protein concentration was adjusted, and protein samples were loaded on SDS polyacrylamide gels after adding bromophenol blue (0.05%). Primary anti-SRSF2 antibody (ab28428; Abcam), anti-β-actin antibody (A1978; Sigma-Aldrich), V5 tag (sc-83849-R;

Santa Cruz) were used. Signals from fluorophore-conjugated secondary antibodies were detected with the Odyssey system (LI-COR).

Retrovirus infection

For the retrovirus infection for mouse bone marrow cells, 293T cells were transfected with 5 μg of retrovirus vectors and with 5 μg of Ecopac packaging vector using polyethylenimine (Polysciences Inc). The 293T medium was changed from DMEM to IMDM 10 hrs post transfection. Retroviral supernatants were harvested 48 hrs after transfection and filtered through a 0.45 μm filter. The supernatant was added to primary bone marrow cells, along with 4% IL3-CM, 4%SCF-CM, 1% HEPES and 0.1% polybrene (final concentration 4 μg/mL). The cells were spinoculated at 1200 × g for 3 hrs at 32°C. Infections were performed twice on consecutive days. For overexpression of WT and mutant SRSF2, BM cells were transduced with MIP vector or MIP-SRSF2 expression retrovirus, selected after two rounds of infection in 1 μg/mL puromycin for 3 days before assays.

For lentivirus production, 293T cells were transfected with pTRIPZ lentivirus vectors, pCMV-VSVG, and pCMV-dR8.2 using lipofectamine 2000 (Life Technologies) for 6 hrs and then the medium was changed. Forty-eight hours later, culture supernatant was harvested, and filtered through a 0.45 μm filter. For viral transduction, RPMI1640 medium with 10 ng/ml hIL-3 was diluted with viral supernatant at 1:1 ratio. The MDS-L cells were then cultured in the mixed medium with 8 μg/ml polybrene overnight. Next day, medium was changed, and the stable cell lines with incorporated viral DNA segments were selected in 2 μg/ml puromycin for 7 days.

RASL-seq

Doxycycline (Dox, Sigma) at indicated concentration was added to MDS-L cell culture every day for 3 days. Total RNA was extracted from cells and used for RASL-seq. Analysis of

splicing changes was described previously (25, 26). Gene function and pathway analyses were performed using QIAGEN's Ingenuity® Pathway Analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity). Original data is uploaded to GSE (GEO accession number: GSE61052).

RNASEH1 Overexpression in Mouse c-Kit+ Cells and *in vitro* Colony-forming Assays

c-Kit+ cells were enriched by anti-mouse CD117 Microbeads (Miltenyi Biotech) from 8-week-old Mx1-cre Srsf2(WT) or Mx1-cre Srsf2(P95H/WT) mice 2 weeks after Poly(I:C) injection. c-kit+ cells were infected with retrovirus expressing a MSCV-IRES-GFP empty vector or the vector containing the full length RNASEH1 cDNA. The viral infection procedure was performed as previously described[81]. GFP+ cells were sorted, 40,000 of which were seeded into methylcellulose medium (Methocult M3434; STEMCELL Technologies). Colonies were counted 14 days after seeding. Protein expression of RNASEH1 was confirmed by western blotting.

Immunofluorescence

Detection of R-loops by immunofluorescence using S9.6 antibody was performed according a previously described protocol[82]. Briefly, isolated mouse bone marrow c-Kit+ cells and K562 cells were suspended in the pre-warmed 75 mM KCl solution in a drop-wise manner and incubated for 12 min at 37°C. Several drops of freshly made, ice-cold fixative solution [methanol:acetic acid(3:1)] were added to cells, followed by centrifugation (800 rpm, 5 min, 4°C). The supernatant was aspirated down to 300 ml. Cells were resuspended in 5 mL of fixative solution in a drop-wise manner followed by incubation for 20 min on ice. Cells were spun down and then washed once with fixative solution. The supernatant was aspirated down to 20-30 ml. Fixed cells were resuspended by pipetting and spotted onto slides, which were then incubated for 30 min at 65°C C. For blocking, slides were incubated with blocking buffer (1X PBS, 0.2% Triton X-100,

5% BSA) for 1 hr at room temperature, followed by incubation with S9.6 antibody overnight at 4°C. On the next day, slides were washed three times with wash buffer (1X PBS, 0.1% Triton X-100) and then incubated with fluorescence- conjugated secondary antibody (Alexa Fluor 488 Donkey anti-Mouse IgG (H+L), 1:500 dilution) for 1hr at room temperature. Cells were then washed three times with 1X PBS and stained with DAPI for 30 min and mounted with Fluoromount-G (SouthernBiotech) for imaging under an Olympus FV1000 confocal microscope.

For g-H2AX staining, cells were fixed with 4% paraformaldehyde for 20 min and then incubated with 0.3% Triton X-100 in PBS for 30 min, both at room temperature. After washing with PBS, cells were then incubated with blocking buffer (10% BSA and 0.03% Trixon X-100 in PBS) for 1 hr at room temperature followed by incubation with primary antibody for g-H2AX (1:500 dilution) overnight at 4°C. After washing with wash buffer (0.03% Trixon X-100 in PBS), cells were incubated with the fluorescence-conjugated secondary antibody (Alexa Fluor 594 Goat anti-Mouse IgG (H+L), 1:500 dilution) diluted with wash buffer for 1 hr at room temperature. Cells were then washed three times and stained with DAPI for 30 min and mounted with Fluoromount-G for imaging under an Olympus FV1000 confocal microscope.

Statistics

Results were represented as mean ± standard deviation (SD) unless otherwise stated. Comparison between two groups was done by t-test. Whenever asterisks are used to indicate the statistical significances, *stands for $p < 0.05$; **for $p < 0.01$, and ***for $p < 0.001$.

1.6 Acknowledgements

Chapter 1, in part is a reprint of the material as it appears in Molecular Cellular Biology 2015, and Molecular Cell 2018 as following citation indicated. Yi-Jou Huang was one of the primary authors of these two papers.

Yukiko Komeno*, **Yi-Jou Huang***, Jingsong Qiu, Leo Lin, YiJun Xu, Yu Zhou, Liang Chen, Dora D Monterroza, Hairi Li, Russell C DeKelver, Ming Yan, Xing-Dong Fu, Dong-Er Zhang. Srsf2 is essential for hematopoiesis and its myelodysplastic syndromes-related mutations dysregulate alternative pre-mRNA splicing. Molecular and Cellular Biology. 2015 (*indicates equal contribution)

Liang Chen*, Jia-Yu Chen*, **Yi-Jou Huang***, Ying Gu, Jingsong Qiu, Hao Qian, Changwei Shao, Xuan Zhang, Jing Hu, Hairi Li, Shumin He, Yu Zhou, Omar Abdel-Wahab, Dong-Er Zhang, Xing-Dong Fu. Excessive R-loop formation: A unifying mechanism for splicing factor mutation-Induced myelodysplastic syndromes. Molecular Cell. 2018 (*indicates equal contribution)

I would like to thank Dr. Yukiko Komeno's guidance when I first joined the lab and started to work on SRSF2 studies. I also want to thank Dr. Liang Chen and Dr. Jia-Yu Chen's insightful discussion and feedback during our time working together. I want to thank Fu lab members for their intellectual discussion, techniques and equipment sharing throughout my time working on SRSF2 projects.

Chapter 2. RUNX1 deficiency cooperates with SRSF2 mutation to further disrupt RNA splicing program and exacerbate myelodysplastic syndromes phenotypes

2.1 Summary

Although SRSF2 is one of the most frequently mutated genes in MDS, its mutations rarely occur alone, indicating potential cooperation with other co-existed mutations to promote diseases. SRSF2 mutations are found to significantly associated with transcriptional factor RUNX1 mutations in MDS patients, and the co-existence is linked to poor prognosis of the disease. Transcriptional and RNA splicing are critical steps in controlling gene expression, but the understanding of know how the mutations in genes regulating these processes and facilitate pathogenesis remains elusive. Due to the complexity of MDS patient genetic background, isogenic models are required to further understand how these genetic alternations interact. Here we established isogenic double-mutations systems in both mouse and human, which includes SRSF2 P95H and RUNX1 deficiency. We found this co-existence exacerbates MDS phenotypes in mice and results in pan-cytopenia, skewing to myeloid at expense of B cell and competitive disadvantage. In global aspect, the splicing and gene expression programs are further perturbed and impact MDS-related pathways including RNA processing, DNA damage and cell cycle checkpoint. Together, these results identify the pathogenic crosstalk between two distinct classes of somatic mutations frequently occurred in MDS and provide functional consequences of the coexistence.

2.2 Introduction

MDS are genetically complex diseases that require multiple mutations during disease development. Acquisition of somatic mutations results in clonal diversity and different responses

to therapy. Multiple classes of mutations in splicing factors, transcription factors, epigenetic modifiers, and cell signaling proteins are the common molecular events mutated during disease evolution and those events rarely occur alone. Among these factors, somatic mutations in genes encoding spliceosome proteins SRSF2, SF3B1, U2AF1, and ZRSR2 have been found in ~60% of patients with MDS[11, 19, 21, 28, 83]. Mutations in these genes arise early in disease as initiating events[84, 85] and they are considered to have causative effect in MDS disease development because of their significant frequency, high allelic ratios and association with clinical outcomes.

The presence of SRSF2 mutations occurred in substantial percentage of MDS cases as discussed in Chapter 1, and the role of SRSF2 and its mutation in hematopoiesis and MDS development in mouse models and cellular models recently have been recently reported in multiple studies[56, 64, 86, 87]. To date, there are three independent groups reported Srsf2 P95H conditional knockin mouse models[64, 86, 87], these murine models recapitulate human MDS or MDS/MPN to varying extents. The mutations impact global RNA splicing outcomes by changing the RNA binding preference of SRSF2 in sequence-specific manner, leading to dysregulation of exon inclusion and mis-spliced transcripts critical to myeloid malignancies[64, 88]. Clinically, SRSF2 mutations are linked to significantly shorter overall survival, especially in the lower risk MDS patients when compared with patients without the mutation, and it also corelated to shorter time to AML progression[30, 31, 89, 90]. Moreover, SRSF2 mutations almost always present as heterozygous missense mutations and rarely occur alone. In a clinical MDS report, it shown that SRSF2 mutations occurred concurrently with at least one additional mutation in 85.3% of patients in the cohort (85.3%)[89], indicating the potential cooperative effect with other class of mutations. SRSF2 mutation is also shown to closely associated with epigenetic modifiers including TET2, ASXL1,

IDH2. The other class highly associated with SRSF2 mutations, is cell signaling/transcriptional regulators, including RUNX1 and NRAS[30, 31, 89, 91, 92].

Interestingly, the effect of co-mutation varied by partner gene, where patients with the SRSF2 and RUNX1 double mutations did significantly worse and the response rates and overall survival were lowest in patients with these two mutations in AML[93]. Moreover, a clinical report, which evaluated the chronology of the acquisition of mutations from MDS patients before (pre-progression) and after progression (post-progression) to a more advanced MDS subtype or to AML, showed SRSF2 mutations as earlier genetic events while RUNX1 mutations were found in the post-progression sample only in the majority of cases, indicating RUNX1 mutations are possibly late events that may cooperate with early events to promote disease progression[84].

Runt-related transcription factor 1 (RUNX1), also known as AML1, encodes DNA-binding subunit of the heterodimeric CBF complex. RUNX1 is critical for the development of definitive hematopoiesis[94, 95], and homozygous loss of Runx1 in mice results in embryonic lethality due to impaired hematopoiesis[96]. In adult mice, loss of Runx1 is not essential for adult hematopoietic compartment, but several hematopoietic abnormalities were found. Runx1 deficient mice demonstrated lineage specific effects on B and T cell maturation, decrease of platelet numbers, increase of neutrophils in peripheral blood and myeloid progenitor populations[97]. RUNX1 is one of the most frequently mutated genes in MDS, accounting for 10-15% of cases, and they are especially frequent in higher-risk MDS and closely associated with higher neutrophil counts and shorter overall survival[28, 98, 99]. RUNX1 mutations are very dispersed throughout the gene and can be defined into two functionally distinct classes: the N-terminal mutations, consisting the majority of RUNX1 mutations, which are located within runt homology domain (RHD) of the RUNX1 protein and directly disrupt DNA binding or interaction with CBFβ, indicating RUNX1

dysfunction is one of the major mechanisms; and the C-terminal mutations, which are located at activation domain (AD). The second class of mutations increases DNA binding but disrupts transcriptional activity in a dominant negative manner, suggesting additional functions can be gained in these mutations[100, 101].

Although RUNX1 is one of the most frequently co-mutated genes with SRSF2 in MDS patients, and their co-existence is linked to poor prognosis, it remains unclear how this combination of genetic alterations in different categories may have cooperative effects in gene regulation and disease etiology. Here we hypothesize that SRSF2 and RUNX1 mutations cooperate in MDS pathogenesis and assessed this biologically using isogenic mouse transgenic models and human myeloid cell models. These data provide a functional explanation for the genetic co-existence of splicing and transcription factor mutations and identify their convergent effect in global splicing and gene expression programs.

2.3 Results

2.3.1 Double mutations cause MDS-like phenotypes, including pan-cytopenia, dysplastic morphology and skewing to myeloid at expense of B cells

To investigate whether the coexistence of Srsf2 mutation and Runx1 deficiency could synergistically impair hematopoiesis *in vivo*, we established a new mouse model (Srsf2 P95H/+ Runx1 f/f Mx1-Cre) by crossing previously reported Mx1-Cre based conditional knock-in Srsf2-P95H mutation mice (P95H/+)[64] and Runx1 conditional knockout mice (Runx1 f/f)[97]. To determine whether the phenotypic effects of the mutations are in a hematopoietic cell-intrinsic manner, we performed non-competitive transplantation experiments by transplanting mouse bone marrow mononuclear cells collected from Mx1-Cre, Runx1 f/f Mx1-Cre, Srsf2 P95H/+ Mx1-Cre,

and Srsf2 P95H/+ Runx1 f/f Mx1-Cre mice into lethally irradiated recipients and inducing Cre expression by pIpC intraperitoneal injection four weeks after transplantation (Fig. 2-1).

Since MDS patients have ineffective hematopoiesis and results in limited production of mature cells in peripheral blood, we first examined the numbers of three mature lineages (red blood cells, platelets, white blood cells) in the single and double mutation mice. The presence of Srsf2 P95H either in either single mutation or double mutation context both resulted in macrocytic anemia in mice, demonstrating by decrease of hemoglobin and red blood cell number and increase of mean corpuscular volume (Fig. 2-2), while the Runx1 loss did not affect this parameter. The Runx1 loss caused decrease of platelet number (Fig. 2-3A,C), and the larger platelet size shown by mean platelet volume (Fig. 2-3B) indicating immature platelet formation. Interestingly, the thrombocytopenia phenotype is more severe in the Runx1 single mutation condition when compared to the double mutation, suggesting that existence of Srsf2 P95H partially attenuated this defect. In single mutation and double mutation context, leukopenia is always observed (Fig. 2-4A,C). Total number of white blood cells and lymphocytes decreased in single and double mutation mice, while number of granulocytes remained comparable between WT and all the mutation conditions (Fig 2-4 A-D). Overall, the double mutant mice recapitulated MDS-related phenotypes induced by either Srsf2 mutation or Runx1 deficiency. Moreover, the double mutation mice showed lower percentage of lymphocytes and higher percentage of granulocytes and monocytes (Fig 2-4 E-G), indicating skewing between lymphoid and myeloid lineages occurred in peripheral blood. Flow cytometric analysis of the peripheral blood cell lineages displayed more dramatic skewing to the myeloid lineage at expense of the B cell lineage when compared to single mutant mice, and it became more severe in double mutation mice when they are more aged (Fig. 2-5A-D). At 20-weeks post transplantation, we analyzed the spleen and bone marrow

compartments of the mice. In the spleen, we found the same enhanced skewing from B-cells to myeloid lineage as in the peripheral blood compartment (Fig. 2-5 E-F). However, in the bone marrow, single and double deficiency mice both demonstrated similar decreases in B cell percentage. Srsf2 P95H and double mutant mice also demonstrated dysplastic morphology, such as hyposegmented/ hypersegmented neutrophils and red blood cells with Howell-Jolly bodies in peripheral blood (Fig. 2-6). We also examined the hematopoietic stem and progenitor cell compartments of the single and double mutant mice. Under Runx1 deficienct context, the expansion of Lin- cells, Lineage- Sca-1+ c-Kit+ (LSK) cells, granulocyte-monocyte progenitor (GMP), megakaryocyte-erythrocyte progenitor (MEP) were observed in bone marrow and spleen (Fig. 2-7A-D, Fig. 2-8A-E). Though the double mutation mice also showed similar trend of expansion as Runx1 loss itself, there is not further increased of the expansion level, suggesting that presence of Srsf2 P95H does not have additive effect in stem cell expansion aspect.

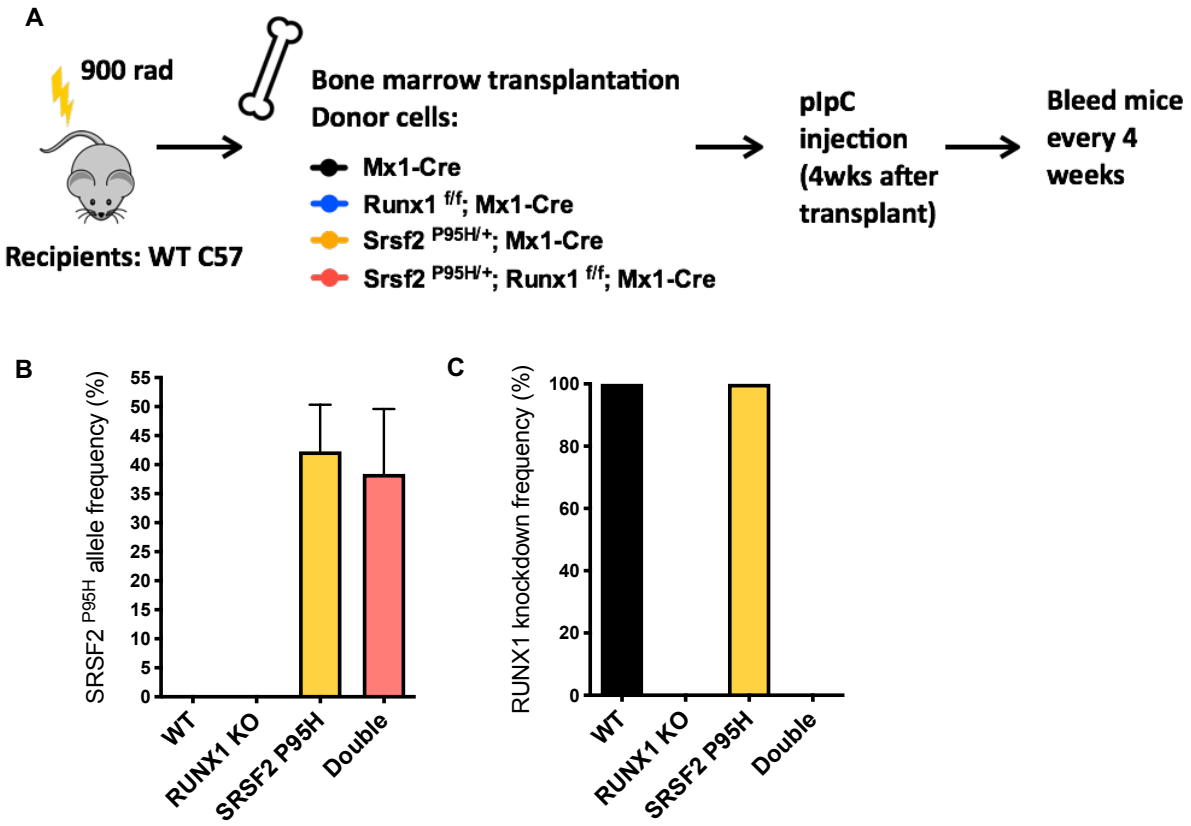


Figure 2-1. Generation of mice harboring conditional Runx1 f/f and Srsf2 P95H/+ by non-competitive bone marrow transplantation. (A) Schema for mouse bone marrow transplantation. (B) Srsf2 P95H allele frequency of mouse bone marrow LK cells quantified by RNA-seq after induction. (C) Runx1 knockdown efficiency of mouse bone marrow LK cells quantified by RNA-seq after induction.

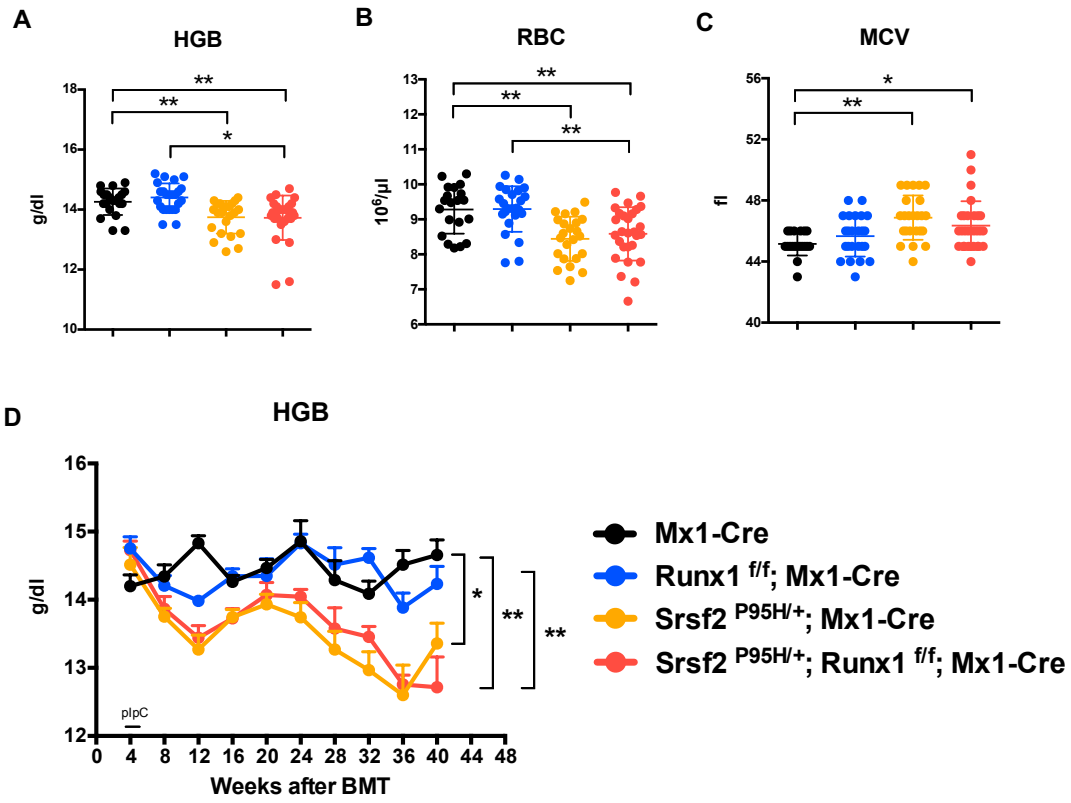


Figure 2-2. Anemia is developed in Srsf2 P95H and double mutation mice. (A-C) Total count of hemoglobin (HGB) (A), red blood cells (RBC) (B) and mean corpuscular volume (MCV) (C) in peripheral blood of recipient mice at week 16 post BMT. (D) Time course analysis of hemoglobin every 4 weeks. *p<0.05, **p<0.01.

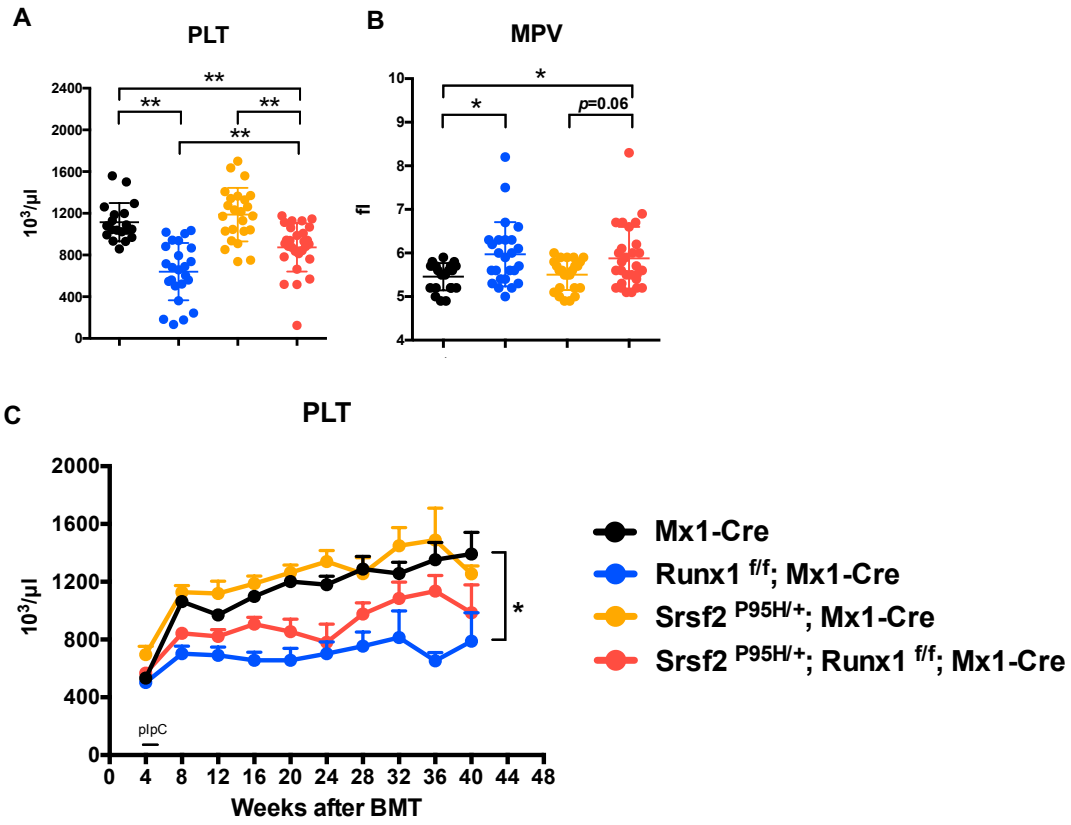


Figure 2-3. Thrombocytopenia is developed in Runx1 deficiency and double mutation mice. (A-B) Total count of platelets (PLT) (A), mean platelet volume (MPV) (B) in peripheral blood of recipient mice at week 16 post BMT. (C) Time course analysis of platelets every 4 weeks. *p<0.05, **p<0.01.

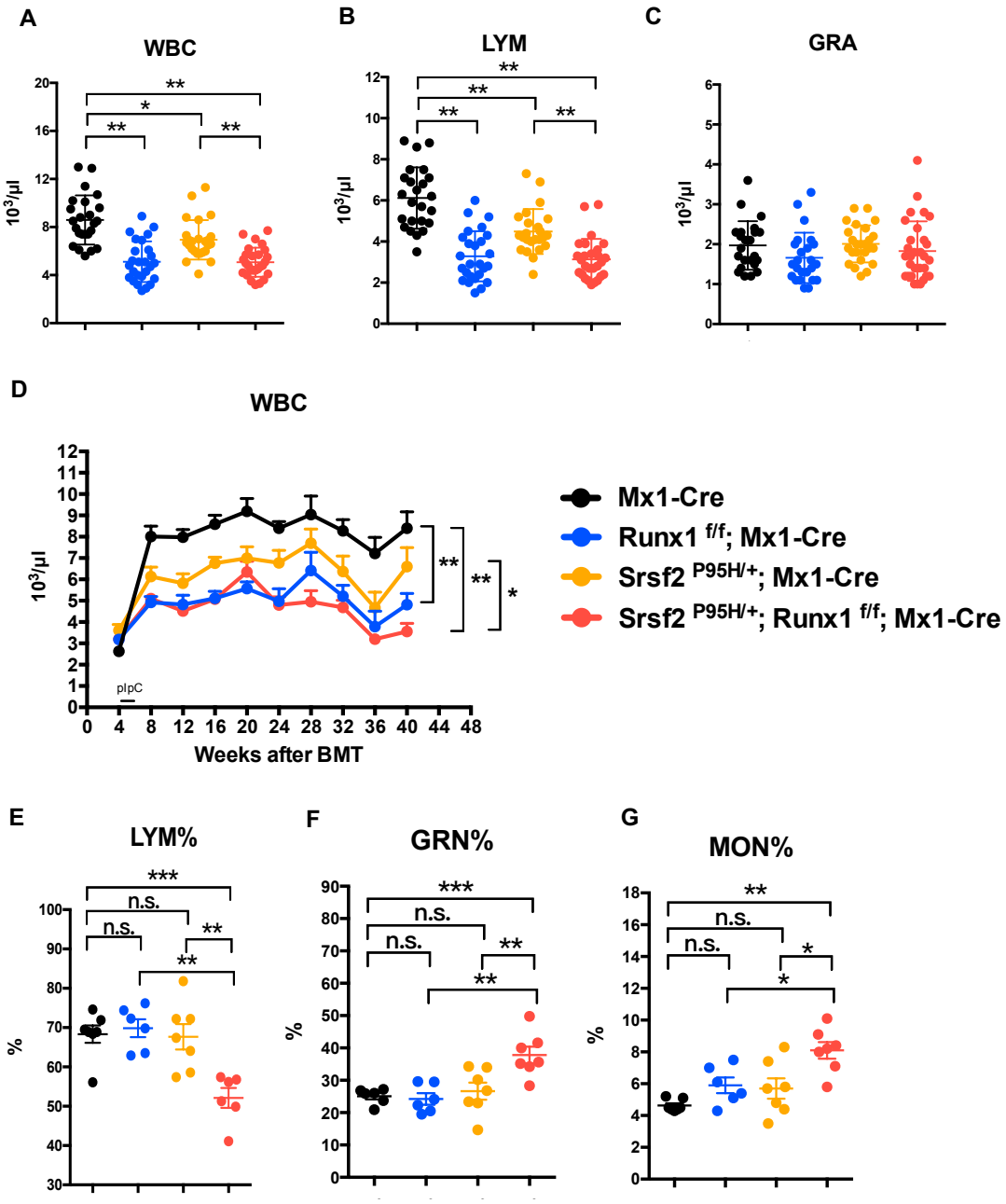


Figure 2-4. Leukopenia is developed in single mutation and double mutation mice. (A-C) Total count of white blood cells (WBC) (A), lymphocytes (LYM) (B) and granulocytes (GRA) (C) in peripheral blood of recipient mice at week 16 post BMT. (D) Time course analysis of white blood cells every 4 weeks. (E-G) Percentage of lymphocytes (E), granulocytes (F), monocytes (G) in peripheral blood of recipient mice 32 weeks post transplantation. *p<0.05, **p<0.01.

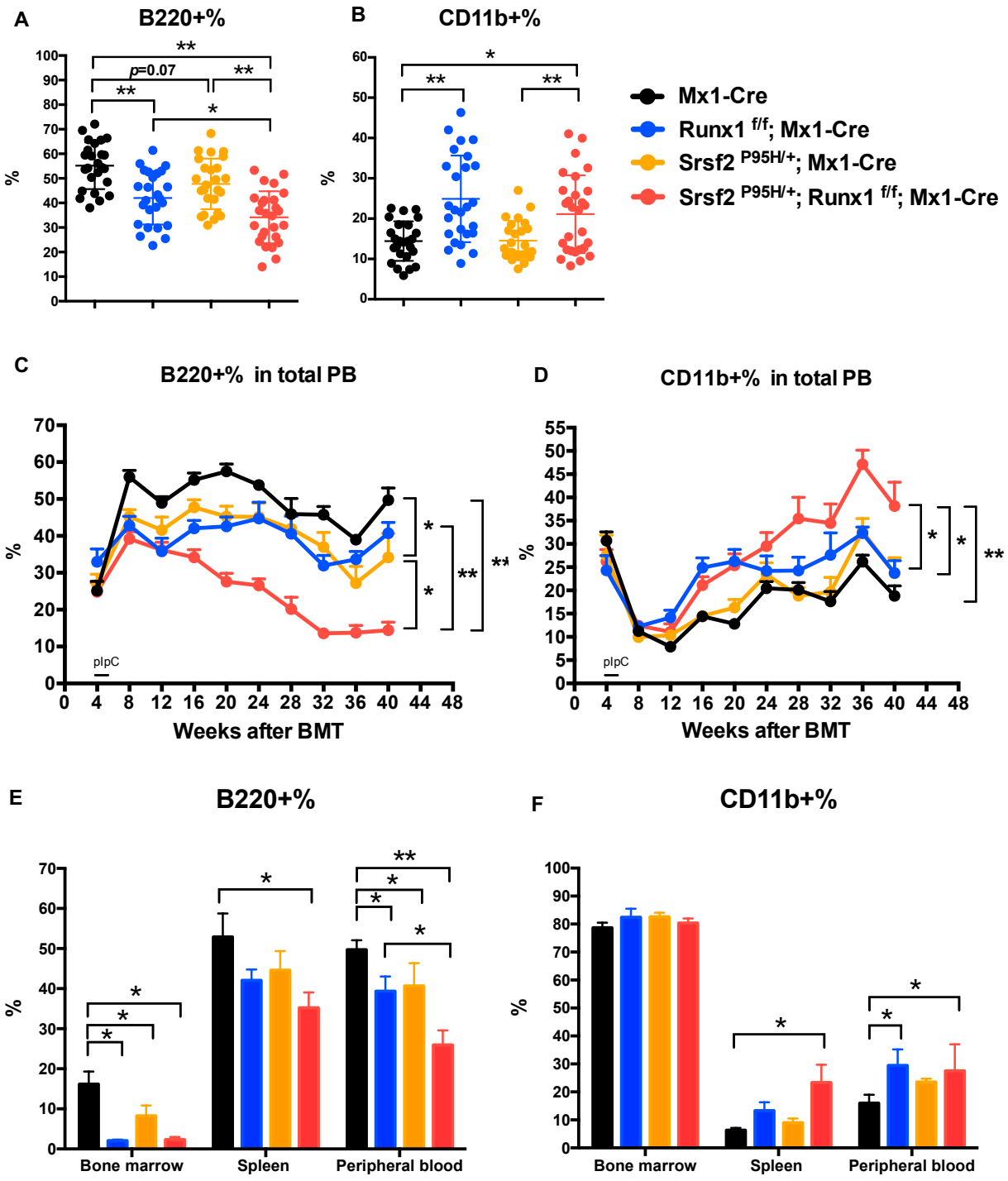


Figure 2-5. Skewing between peripheral B cells and myeloid cells occurred in double mutation mice. (A-B) Frequencies of B cells (B220+) (A) and myeloid cells (CD11b+) (B) in peripheral blood at week 16 post BMT. (C-D) Time course analysis of Frequencies of B cells (B220+) and myeloid cells (CD11b+) (B) every 4 weeks. (E-F) Frequencies of B cells (B220+) and (I) myeloid cells (CD11b+) in bone marrow, spleen and peripheral blood 20 weeks post transplantation. *p<0.05, **p<0.01.

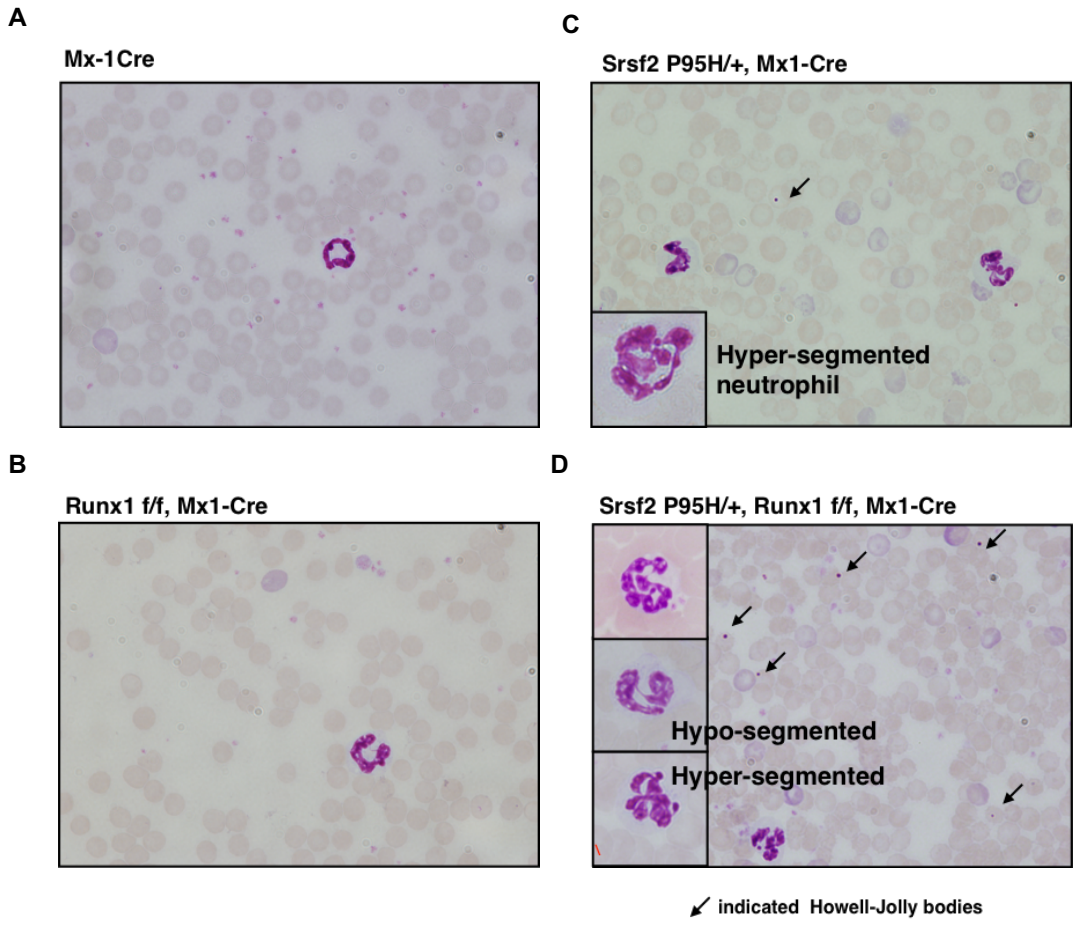


Figure 2-6. Peripheral blood of Srsf2 P95H and double mutation mice demonstrated dysplastic features. (A-D) Peripheral blood smear of WT (A), Runx1 deficiency (B), Srsf2 P95H (C) and double mutation (D) mice.

Figure 2-7: Double mutation caused mild hematopoietic stem cell and myeloid progenitor expansion in bone marrow. (A-D) Percentages of hematopoietic stem and progenitor populations from mouse bone marrow analyzed by flow cytometry after 20 weeks of bone marrow transplantation, including (A) Lineage- cells, (B) Lineage- Sca-1+ c-Kit+ (LSK) cells, (C) hematopoietic stem cells including restricted hematopoietic progenitor cell fractions 1 (HPC-1), short-term hematopoietic stem cells (ST-HSC), long-term hematopoietic stem cells (LT-HSC), and multipotent progenitor (MPP). (D) Percentage of granulocyte-monocyte progenitor (GMP), megakaryocyte-erythrocyte progenitor (MEP), and common myeloid progenitor (CMP). *$p<0.05$, **$p<0.01$.

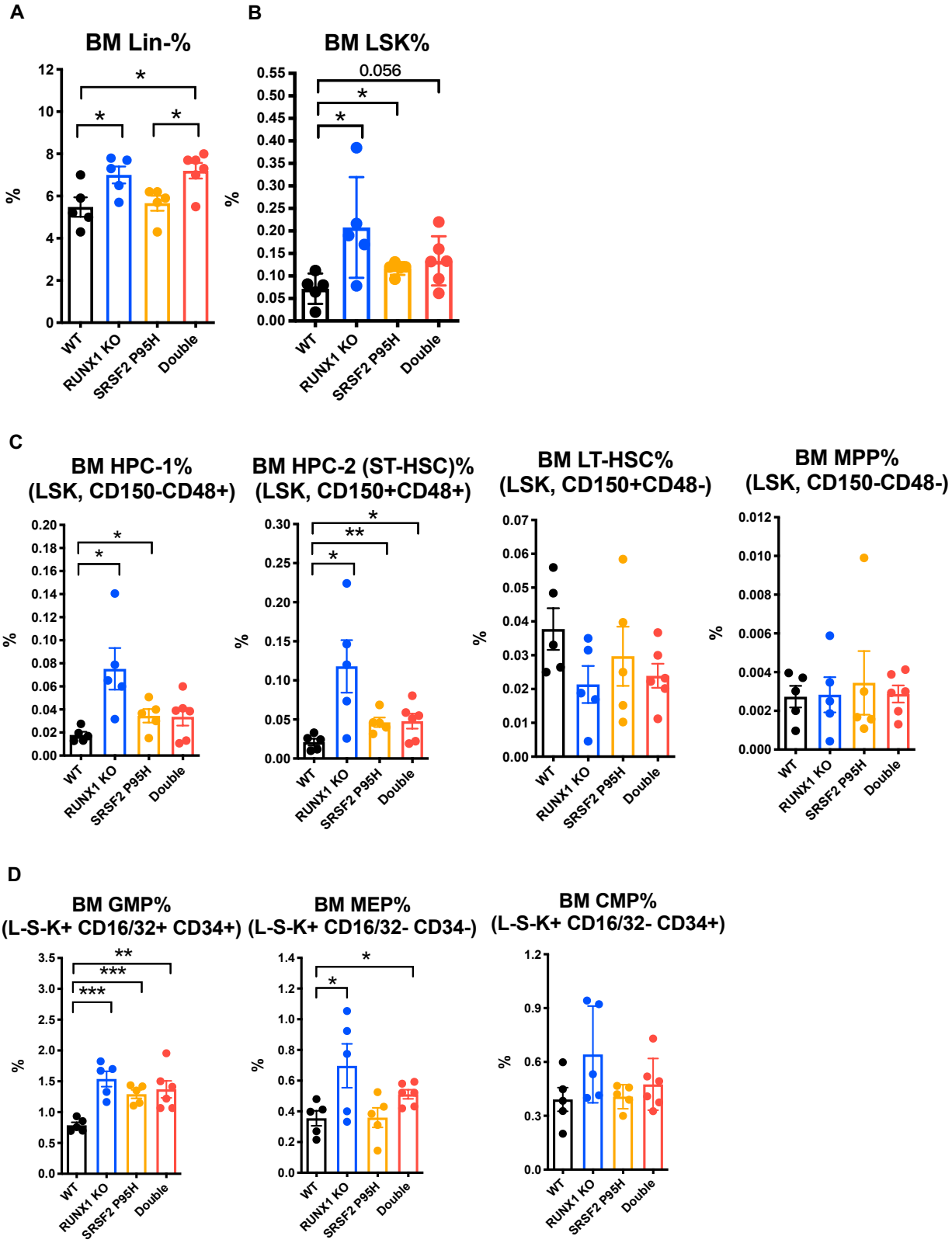
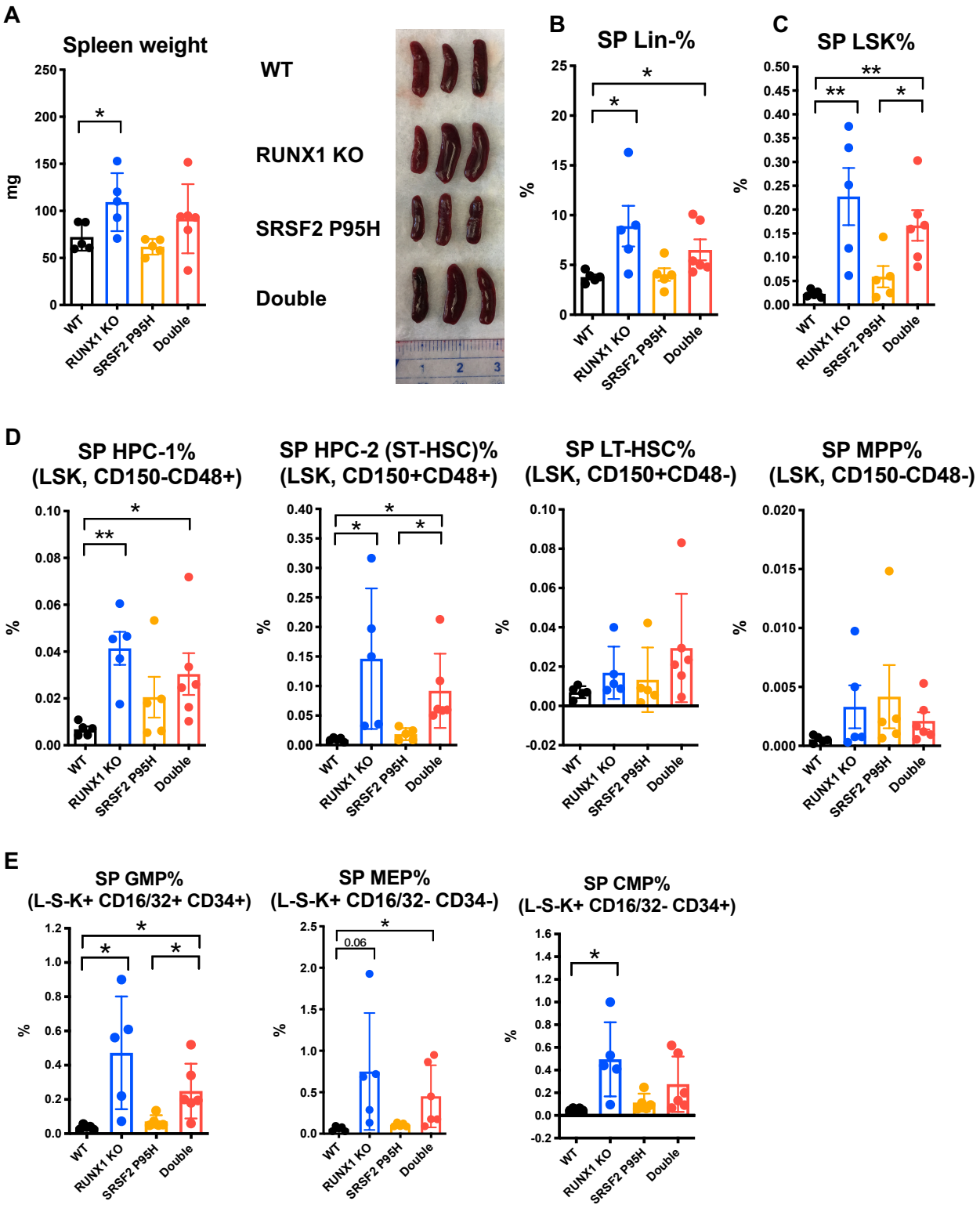

A
BM Lin-%
%
WT
RUNX1 KO
SRSF2 P95H
Double
B
BM LSK%
0.056
C
BM HPC-1%
(LSK, CD150-CD48+)
BM HPC-2 (ST-HSC)%
(LSK, CD150+CD48+)
BM LT-HSC%
(LSK, CD150+CD48-)
BM MPP%
(LSK, CD150-CD48-)
D
BM GMP%
(L-S-K+ CD16/32+ CD34+)
BM MEP%
(L-S-K+ CD16/32- CD34-)
BM CMP%
(L-S-K+ CD16/32- CD34+)

Figure 2-8: Double mutation caused mild hematopoietic stem cell and myeloid progenitor expansion in spleen. (A) Weight and photo of spleen in each genotype. (B-E) Percentages of hematopoietic stem and progenitor populations from mouse spleen analyzed by flow cytometry after 20 weeks of bone marrow transplantation, including (B) Lineage- cells, (C) Lineage- Sca-1+ c-Kit+ (LSK) cells, (D) hematopoietic stem cells including restricted hematopoietic progenitor cell fractions 1 (HPC-1), short-term hematopoietic stem cells (ST-HSC), long-term hematopoietic stem cells (LT-HSC), and multipotent progenitor (MPP). (E) Percentage of granulocyte-monocyte progenitor (GMP), megakaryocyte-erythrocyte progenitor (MEP), and common myeloid progenitor (CMP). *$p<0.05$, **$p<0.01$.


A
Spleen weight
mg
WT
RUNX1 KO
SRSF2 P95H
Double
B
SP Lin-%
%
C
SP LSK%
%
D
SP HPC-1%
(LSK, CD150-CD48+)
SP HPC-2 (ST-HSC)%
(LSK, CD150+CD48+)
SP LT-HSC%
(LSK, CD150+CD48-)
SP MPP%
(LSK, CD150-CD48-)
E
SP GMP%
(L-S-K+ CD16/32+ CD34+)
SP MEP%
(L-S-K+ CD16/32- CD34-)
0.06
SP CMP%
(L-S-K+ CD16/32- CD34+)

2.3.2 Runx1 loss exacerbates competitive disadvantage to Srsf2 P95H hematopoietic stem and progenitor cells in competitive bone marrow transplantation

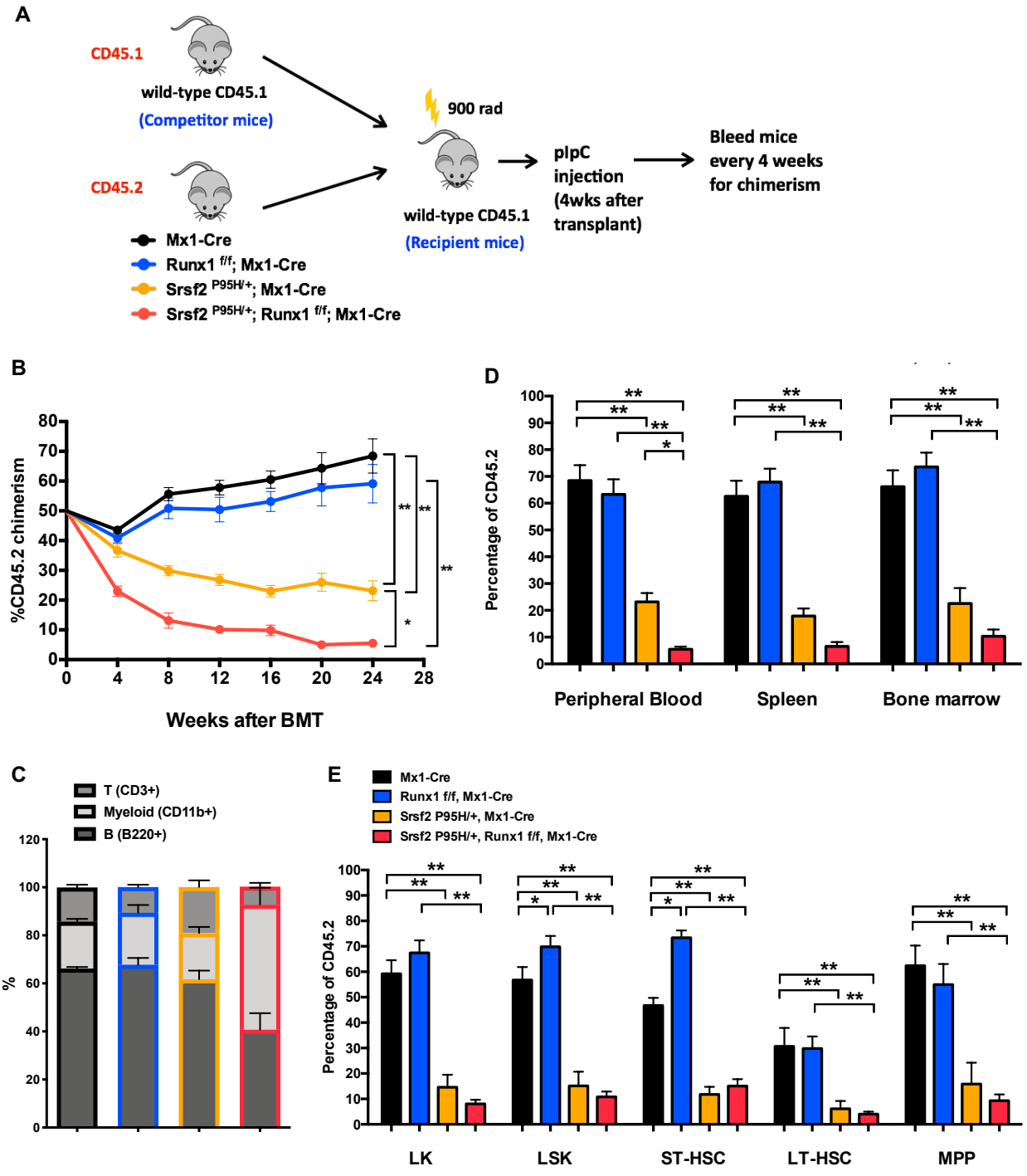
To evaluate hematopoietic stem cell and progenitor function *in vivo*, we performed competitive bone marrow transplantation. We mixed CD45.1 wild type bone marrow cells and CD45.2 (Mx1-Cre, Runx1 f/f Mx1-Cre, Srsf2 P95H/+ Mx1-Cre, Srsf2 P95H/+ Runx1 f/f Mx1-Cre) bone marrow cells in a 1:1 ratio, transplanted to lethally irradiated recipients, and injected pIpC 4 weeks after transplantation (Fig. 2-9A). Chimerism analysis is evaluated by percentage of CD45.1 and CD45.2 populations in peripheral blood.

The repopulation assay showed that Srsf2 P95H with Runx1 deficiency confer a competitive disadvantage (Fig. 2-9B). Within the CD45.2+ population of peripheral blood, the B220+ percentage of double mutation cells decreased and CD11b+ increased, which demonstrated similar trend as non-competitive bone marrow transplanted setting (Fig. 2-9C). After 24 weeks of transplantation, the chimerism of total cells in spleen and bone marrow were analyzed (Fig. 2-9D), similar to the trend in peripheral blood, the presence of Srsf2 P95H confer competitive disadvantage and the Runx1 loss further decreased the chimerism. Within the bone marrow population, the CD45.2 chimerism of stem and progenitor cell compartments (LK, LSK, ST-HSC, LT-HSC, MPP) were analyzed, Srsf2 P95H single mutation and double mutation demonstrated similar level of lower chimerism when compared to WT control cells or Runx1 single deficiency cells (Fig. 2-9E). These results demonstrating the loss of Runx1 further exacerbate the competitive disadvantage of Srsf2 P95H cells in *in vivo* competitive transplantation setting, which indicated the additive effect of double mutation leads to worsen the Srsf2 P95H MDS phenotypes.

Together, double mutation altered multi-lineage hematopoiesis in mice and showed MDS features including severe leukopenia in multiple lineages, anemia, and thrombocytopenia and mild

expansion of hematopoietic stem cells and myeloid progenitor cells. These results suggest that the deficiency of compound RUNX1 and SRSF2 P95H mutations exacerbates the disease phenotypes caused by single mutations alone, and the more severe phenotypes that may underlie the poor prognosis of patients with both mutations.

Figure 2-9. Runx1 loss and Srsf2 P95H cause competitive disadvantage to cells under competitive bone marrow transplantation condition. (A) Schematic of generation of competitive bone marrow transplantation model. (B)The percentage of CD45.2+ chimerism in peripheral blood. (C) Percentages of three lineages in CD45.2+ compartment in peripheral blood at week24 after transplantation. (D) CD45.2 chimerism of total cells in peripheral blood, spleen and bone marrow at week24 after transplantation. (E) CD45.2 chimerism of stem and progenitor cell compartments in bone marrow at week24 after transplantation. *$p<0.05$, **$p<0.01$.


A
CD45.1
wild-type CD45.1
(Competitor mice)
CD45.2
900 rad
wild-type CD45.1
(Recipient mice)
pIpC
injection
(4wks after
transplant)
Bleed mice
every 4 weeks
for chimerism
Mx1-Cre
Runx1 f/f; Mx1-Cre
Srsf2 P95H/+; Mx1-Cre
Srsf2 P95H/+; Runx1 f/f; Mx1-Cre
B
%CD45.2 chimerism
Weeks after BMT
D
Percentage of CD45.2
Peripheral Blood
Spleen
Bone marrow
C
T (CD3+)
Myeloid (CD11b+)
B (B220+)
%
E
Mx1-Cre
Runx1 f/f, Mx1-Cre
Srsf2 P95H/+, Mx1-Cre
Srsf2 P95H/+, Runx1 f/f, Mx1-Cre
Percentage of CD45.2
LK
LSK
ST-HSC
LT-HSC
MPP

2.3.3 The co-existence of SRSF2 P95H mutation and RUNX1 deficiency exacerbates global splicing defects

To understand the mechanistic basis for the magnified MDS phenotypes by SRSF2 and RUNX1 double mutations, we performed RNA-sequencing (RNA-seq) on mouse bone marrow hematopoietic progenitor Lineage- c-Kit+ (LK) cells, and on isogenic K562 cells, a human myeloid model cell line allowing expeditious elucidation and manipulation of critical molecular events. In K562 system, the SRSF2 P95H mutations was introduced by CRISPR knockin in one of the three alleles[65], while RUNX1 deficiency was introduced by shRNA knockdown (Fig. 2-10A). The RUNX1 knockdown efficiency was confirmed at protein level to 20-30% in all biological replicates (Fig. 2-10C). The mean allelic ratio of SRSF2 P95H was verified to be around 30% in both SRSF2 single mutant and double mutant cells (Fig. 2-10D). In mouse system, bone marrow LK progenitor cells were sorted from mice transplanted with bone marrow carrying single mutation or double mutations (Fig. 2-10B). With highly reproducible deep RNA-seq with 100M reads (Fig. 2-10E) we detected alternative splicing (AS) changes in all three genotypes. The altered splicing events could serve as molecular signature to well segregate different genotypes (Fig. 2-11A-B). In total, 553, 1,434 and 3,079 abnormal AS events were identified in human K562 SRSF2 P95H cells, RUNX1 deficiency cells and double mutant cells, respectively (Fig. 2-11C). The double mutation demonstrated the greatest number of splicing events alternations, which was also recapitulated in murine system (Fig. 2-11D), in other words, the loss of RUNX1 significantly cooperated with SRSF2 P95H to further induce global splicing defects at the level that were not observed in any single mutant in both human and murine context (Fig. 2-11C-D). We were next prompted to unveil the mechanism underlying the exacerbated splicing defects by RUNX1 in the presence of SRSF2 P95H mutant. Since RUNX1 was shown to be able to bind synthetic RNA

aptamers[102], it is interesting to know whether RUNX1 directly bound a set of RNA with specific sequence motifs for alternative splicing regulation in *vivo*. Within the types of abnormal AS events, skipped exon (SE) events were most frequently affected when SRSF2 P95H mutation presented (Fig. 2-12A-B), consistent with prior studies[64, 86, 87, 103, 104]. We therefore specifically investigated whether these SE events were potential targets of RUNX1 with specific motifs. As a control, the cassette exons promoted versus repressed in cells expressing SRSF2 P95H were enriched for CCNG and GGNG exonic splicing enhancer (ESE) motifs respectively as reported[64], independent of the condition of RUNX1 deficiency (Fig. 2-13). However, no evident consensus motifs could be recovered in single mutant cells with RUNX1 deficiency only, suggesting that RUNX1 either may not directly bind RNA *in vivo* or bind RNA with no sequence preference and it possibly indirectly affects splicing outcomes through its canonical role in gene expression regulation. We also compared the splicing inclusion level differences between single and double mutation contexts in both K562 and mouse LK cells (Fig. 2-14). Interestingly, the loss of RUNX1 attenuated the depth of global splicing changes caused by SRSF2 P95H in double mutant cells. To sum, these two genetics deficiency together caused more mis-splicing events and affected more genes, while the splicing depth is slightly attenuated when two mutations coexisted.

Figure 2-10. Double mutations induce changes in transcriptomes in human K562 cells and mouse bone marrow Lin- c-Kit+ progenitor cells. (A) Schema for K562 sample preparation for RNA-seq. (B) Experimental scheme of mouse LK cells sample preparation for RNA-seq. Samples are bone marrow Lineage- c-Kit+ progenitor cells sorted from mice transplanted with single mutation or double mutations bone marrow. (C) Knockdown efficiency of RUNX1 was confirmed in K562 cells. To quantitate depletion of RUNX1 of three biological replicates, the relative protein levels of RUNX1 were adjusted using β-Actin loading control and quantified relative to the protein level present in the WT controls (set as 1). (D) Expression of SRSF2 P95H alleles as percentage of mRNAs expressed from SRSF2 in K562 cells. (E) Unsupervised principal component analysis of total RNA reads number of genes expressed in three biological replicates of each genotype in K562 cells. **$p<0.01$.

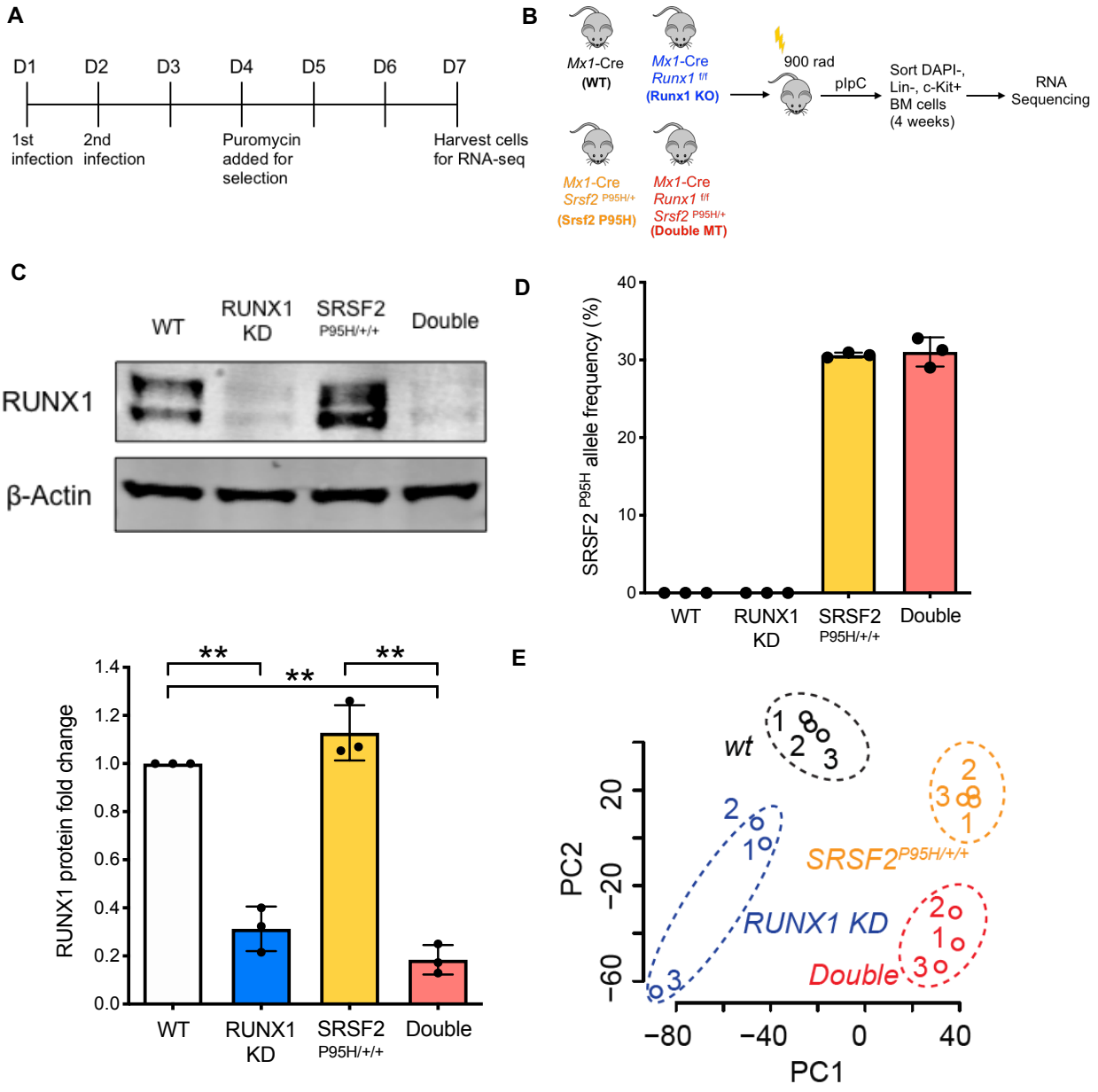
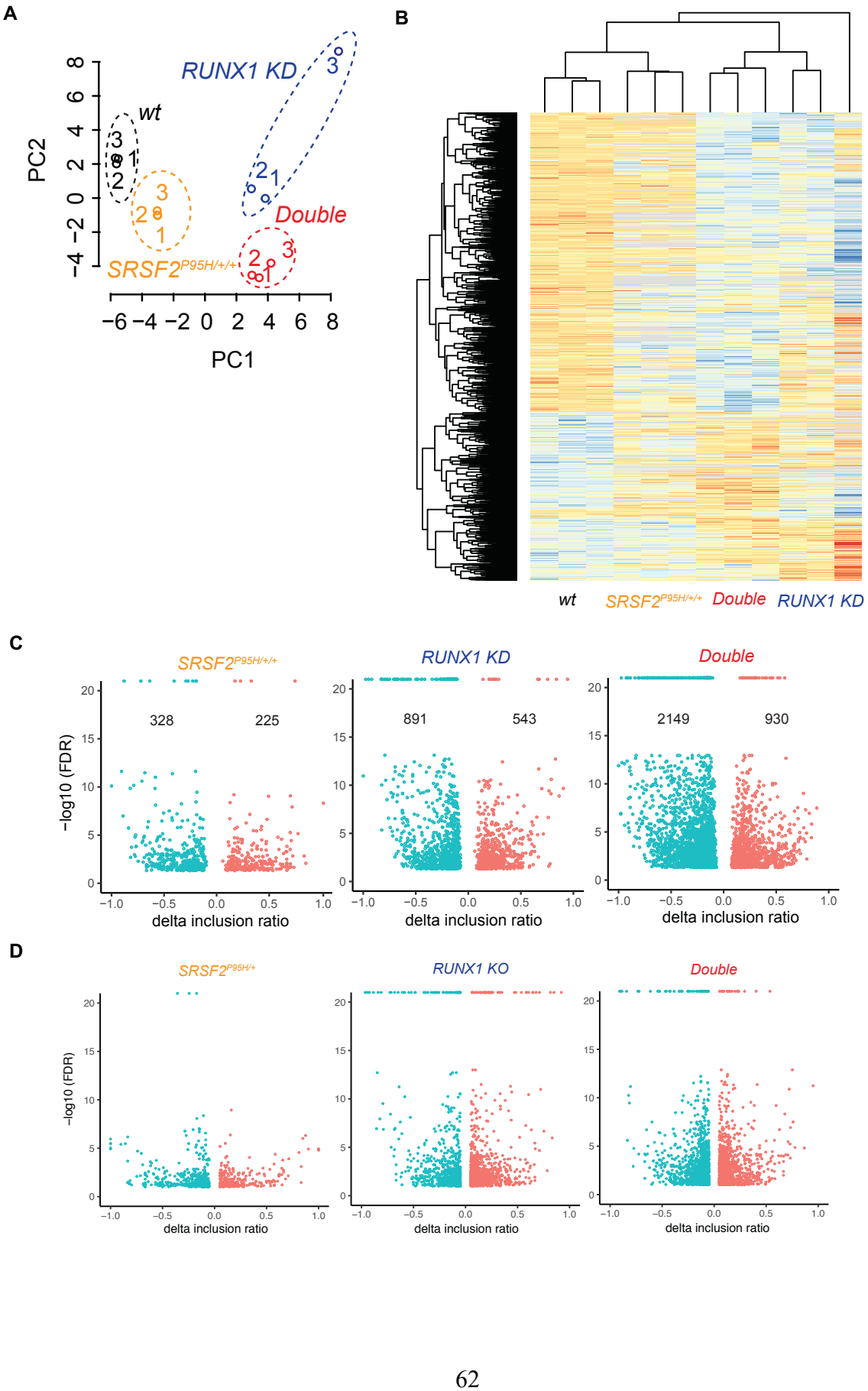

A
D1
D2
D3
D4
D5
D6
D7
1st infection
2nd infection
Puromycin added for selection
Harvest cells for RNA-seq
B
Mx1-Cre (WT)
Mx1-Cre Runx1 f/f (Runx1 KO)
Mx1-Cre Srsf2 P95H/+ (Srsf2 P95H)
Mx1-Cre Runx1 f/f Srsf2 P95H/+ (Double MT)
900 rad
pIpC
Sort DAPI-, Lin-, c-Kit+ BM cells (4 weeks)
RNA Sequencing
C
WT
RUNX1 KD
SRSF2 P95H/+/+
Double
RUNX1
β-Actin
RUNX1 protein fold change
D
SRSF2 P95H allele frequency (%)
E
wt
SRSF2P95H/+/+
RUNX1 KD
Double
PC1
PC2

Figure 2-11. The co-existence of SRSF2 P95H mutation and RUNX1 deficiency exacerbated global splicing alterations. (A) Unsupervised principal component analysis of the events with splicing alternations in three biological replicates of each genotype in K562 cells. Splicing events identified by rMATS pipeline, with cutoff as FDR<0.05, |PSI|> 0.05. (B) Non-hierarchical clustering based on splicing profiles in K562 cells. (C) Volcano plot of identified differential spliced events in each genotype in K562 cells. (D) Volcano plots of aberrant splicing events in each genotype of mouse LK cells.


A
RUNX1 KD
wt
Double
SRSF2P95H/+/+
PC2
PC1
B
wt
SRSF2P95H/+/+
Double
RUNX1 KD
C
SRSF2P95H/+/+
RUNX1 KD
Double
328
225
891
543
2149
930
−log10 (FDR)
delta inclusion ratio
D
SRSF2P95H/+
RUNX1 KO
Double
−log10 (FDR)
delta inclusion ratio

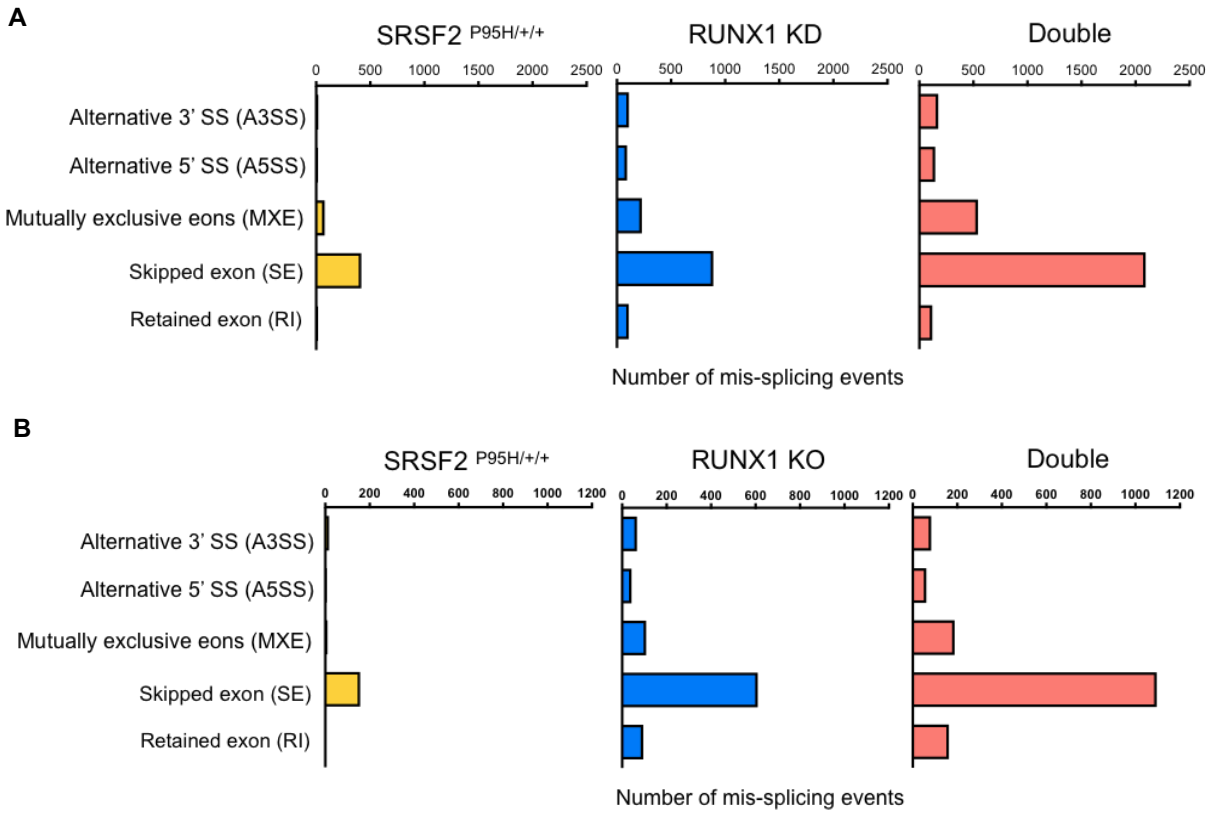
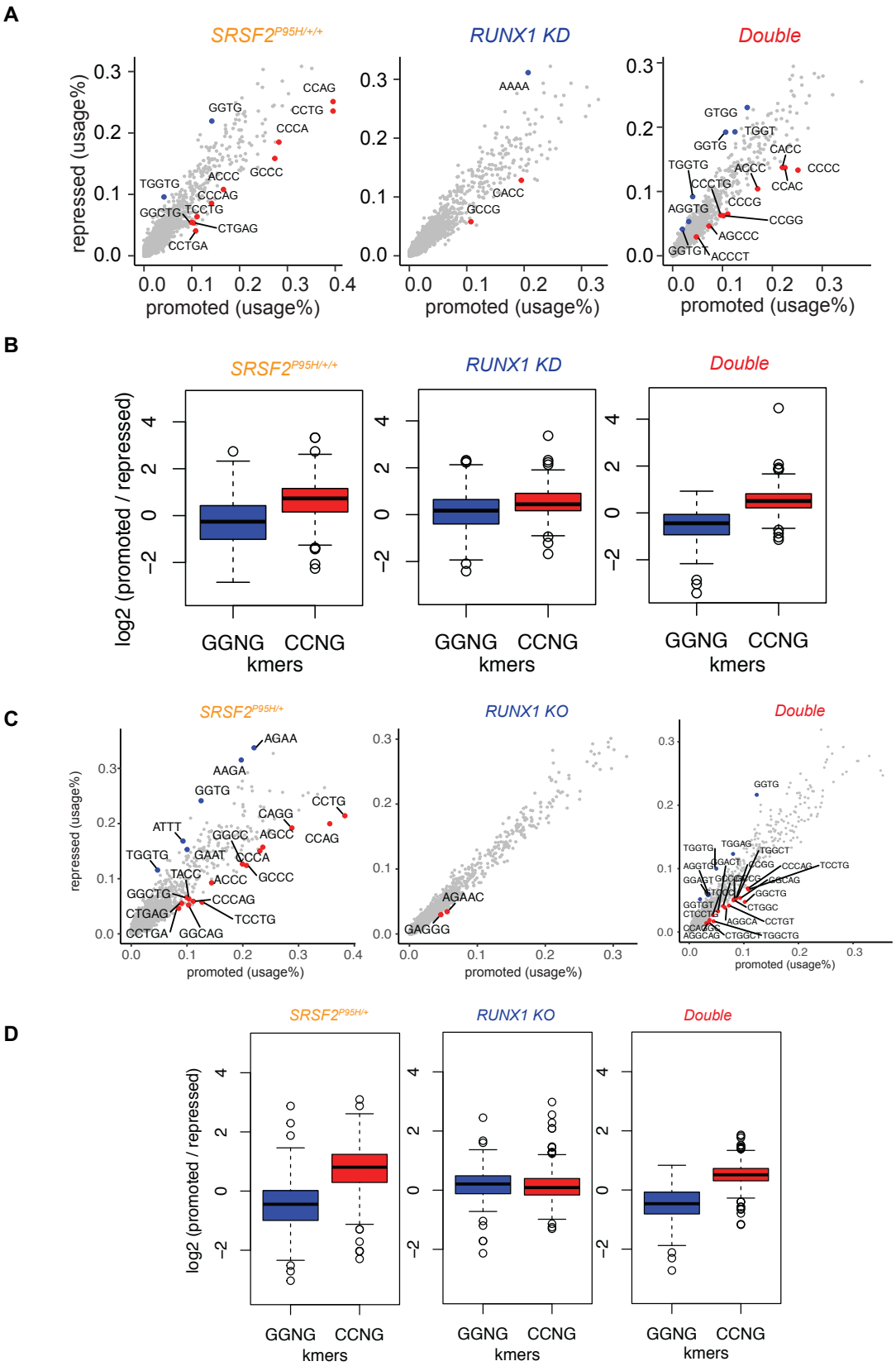


Figure 2-12. Single and double mutations cooperated to predominately affect exon skipping events. (A) The number and types of alternative splicing events in single/double mutation K562 cells compared with WT control cells. (B) Number and types of alternative splicing events in single/double mutation mouse LK cells compared with control.

Figure 2-13. SRSF2 P95H caused enrichment of motifs of CCNG regardless of the presence of RUNX1. (A) Scatter plots illustrating promoted/repressed motif usages in each genotype in K562 cells. (B) Bar plots illustrating enrichment of different variants of the CCNG/GGNG (N = any nucleotide) exonic splicing enhancer adjacent to differentially spliced cassette exons promoted versus repressed in single/double mutation K562 cells relative to WT K562 cells. (p-value of each genotype: Srsf2 P95H <2.2e-16; Runx1 KD= 1.18e-6; Double mt <2.2e-16) (C) Scatter plots illustrating promoted/repressed motif usages in single/double mutation mouse LK cells compared with control. (D) Bar plots illustrating enrichment of different variants of the CCNG/GGNG exonic splicing enhancer in single/double mutation mouse LK cells relative to WT cells. (p-value of each genotype: Srsf2 P95H <2.2e-16; Runx1 KO=0.1443; Double mt <2.2e-16)


A
SRSF2P95H/+/+
RUNX1 KD
Double
repressed (usage%)
promoted (usage%)
CCAG
GGTG
CCTG
CCCA
GCCC
ACCC
TGGTG
CCCAG
GGCTG
TCCTG
CTGAG
CCTGA
AAAA
CACC
GCCG
GTGG
TGGT
GGTG
CACC
TGGTG
ACCC
CCCC
CCCTG
CCAC
AGGTG
CCCG
CCGG
AGCCC
GGTGT
ACCCT
B
SRSF2P95H/+/+
RUNX1 KD
Double
log2 (promoted / repressed)
GGNG
CCNG
kmers
C
SRSF2P95H/+
RUNX1 KO
Double
AGAA
AAGA
GGTG
CCTG
CAGG
ATTT
GGCC
AGCC
CCAG
TGGTG
GAAT
CCCA
TACC
ACCC
GCCC
GGCTG
CCCAG
CTGAG
TCCTG
CCTGA
GGCAG
AGAAC
GAGGG
GGTG
TGGTG
TGGAG
TGGCT
AGGTG
GGACT
GCGG
CCCAG
TCCTG
GGAGT
GGCAG
GTCCC
GGCTG
GGTGT
CTGGC
CTCCTG
AGGCA
CCTGT
CCAGGG
AGGCAG
CTGGCT
TGGCTG
D
SRSF2P95H/+
RUNX1 KO
Double
log2 (promoted / repressed)
GGNG
CCNG
kmers

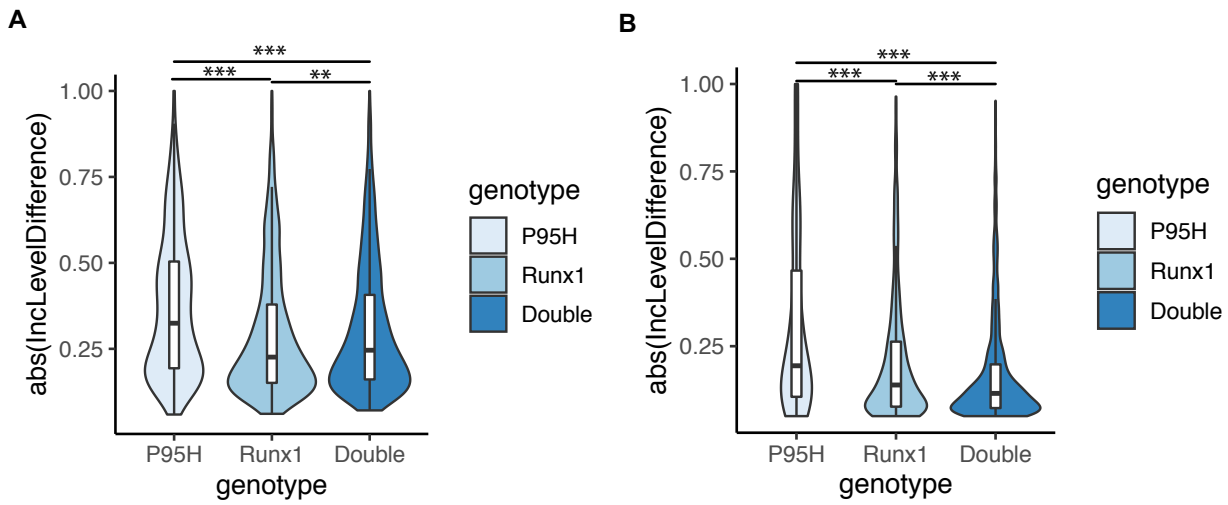


Figure 2-14. The loss of RUNX1 attenuated the level of splicing caused by SRSF2 P95H. (A-B) Splicing inclusion ratio differences of each genotypes in single/double mutation K562 cells (A) and (B) mouse LK cells. **p<0.01, ***p<0.001.

2.3.4 The cooperation of SRSF2 P95H mutation and RUNX1 deficiency results in aberrant splicing of regulators in DNA damage, cell cycle checkpoint and RNA processing

In both human and mouse conditions, the greatest alterations in splicing events were observed in double mutation context compared to single mutant controls, and a combined effect of the two mutations on splicing was shown (Fig. 2-15A, C). Gene Ontology (GO) enrichment analysis was performed on the genes showing significant aberrant splicing events in single and double mutant K562 cells. The clusterProfiler is used to calculate enriched functional categories of each gene clusters and refined overlapping GOs[105]. In single mutant scenario, there is no GO cluster enriched using dysregulated splicing events from SRSF2 P95H mutation alone K562 cells. For RUNX1 deficient K562 cells, GO clusters related to RNA splicing, processing, location, DNA replication and cell cycle checkpoint were revealed. In double mutant K562 context, the analysis also showed strong enrichment of GOs appeared in RUNX1 deficiency cells, while GO cluster associated with DNA damage checkpoint were emerged (Fig. 2-15B). For the individual GO terms enriched in both RUNX1 deficiency and double mutation K562 cells, double mutation cells demonstrated more significant trend (Fig. 2-15E). In mouse LK cells, the enriched GO clusters of mis-spliced genes from double mutant cells highly overlapped with K562 double mutant context (Fig. 2-15E-F), RNA splicing, DNA damage and cell cycle checkpoint again significantly emerged.

We next examined double mutant-induced splicing changes of several target genes in the abovementioned pathways by RT-PCR. Fanconi anemia J gene BRIP1 (also known as FANCJ and BACH1) encodes a DEAH helicase that interacts with the BRCT domain of BRCA1 and has BRCA1-dependent DNA repair and checkpoint functions[106, 107]. The inactivating truncating mutations of BRIP1 were shown to cause Fanconi anemia[108-110], and the point mutations were found to be disease-causing as well[111, 112]. Exclusion of the cassette exon results in an mRNA

encoding a transcript with frameshift (Fig. 2-16A). FANCD2/FANCI-associated nuclease 1 is encoded by FAN1 gene, which was shown to play critical role in DNA interstrand cross link repair[113-115], and missing this exon also cause frameshift (Fig. 2-16B). NABP1 encodes Sensor of Single-Strand DNA Complex Subunit B2 (SOSS-B2), which is component of SOSS complex, a multiprotein complex that functions downstream of the MRN complex to promote DNA repair and G2/M cell cycle checkpoint, and acts as a sensor of single-stranded DNA that binds to single-stranded DNA[116, 117]. The exclusion of the exon leads to frameshift as well (Fig. 2-16C). Targets involved in cell cycle progression were also examined, including TBRG4 (Fig. 2-16D) and AKAP8L (Fig. 2-16E). TBRG4, also known as cell cycle progression restoration protein 2, exclusion of either one of the two adjacent exons in TBRG4 leads to frameshift, and downregulation of its expression contributes to arresting cell cycle progression[118]. The altered isoform ratio can also be observed in EXOSC9 (Fig. 2-16G), which involved in mRNA stability and RNA surveillance pathways[119-121]. In both of the single mutant context, the similar level of splicing alternation can be observed, while in double mutant scenario the altered level is furthered increased, indicating certain cumulative effect in double mutation context to contribute to aberrant splicing (Fig. 2-16H). Multiple altered splicing events were observed in the same gene in double mutant cells. For example, PFKM, a key regulatory enzyme of glycolysis, which loss results in hemolytic anemia[122-124]. There were two skipped exon events of this gene only occurred in double mutant context (Fig. 2-16F). These data results demonstrated the convergent effect of the two genetic alternations impact on the same categories of biological pathways in both human and mouse context, suggesting the importance of these pathways in the disease development process.

Figure 2-15. The cooperation of SRSF2 P95H mutation and RUNX1 deficiency results in mis-splicing targets involved in disease related pathways. (A) Venn diagram showing the overlap between splicing events that were significantly dysregulated in single/double mutation relative to WT controls in K562 cells. (B) Gene Ontology (GO) enrichment cluster analysis showing the dysregulated splicing events in double mutant K562 cells. The color of the circles indicating p-value, and the size of the circles indicating number of genes in each GO terms. Top 30 GO terms are shown. (C) Venn diagram showing the overlap between splicing events that were significantly dysregulated in single/double mutation relative to WT controls in mouse LK cells. (D) GO enrichment cluster analysis showing the dysregulated splicing events in double mutant mouse LK cells. (E-F) Heatmap demonstrating p-value of the GO terms enriched from the mis-spliced genes identified in single and double mutation K562 cells (E) and mouse LK cells (F).

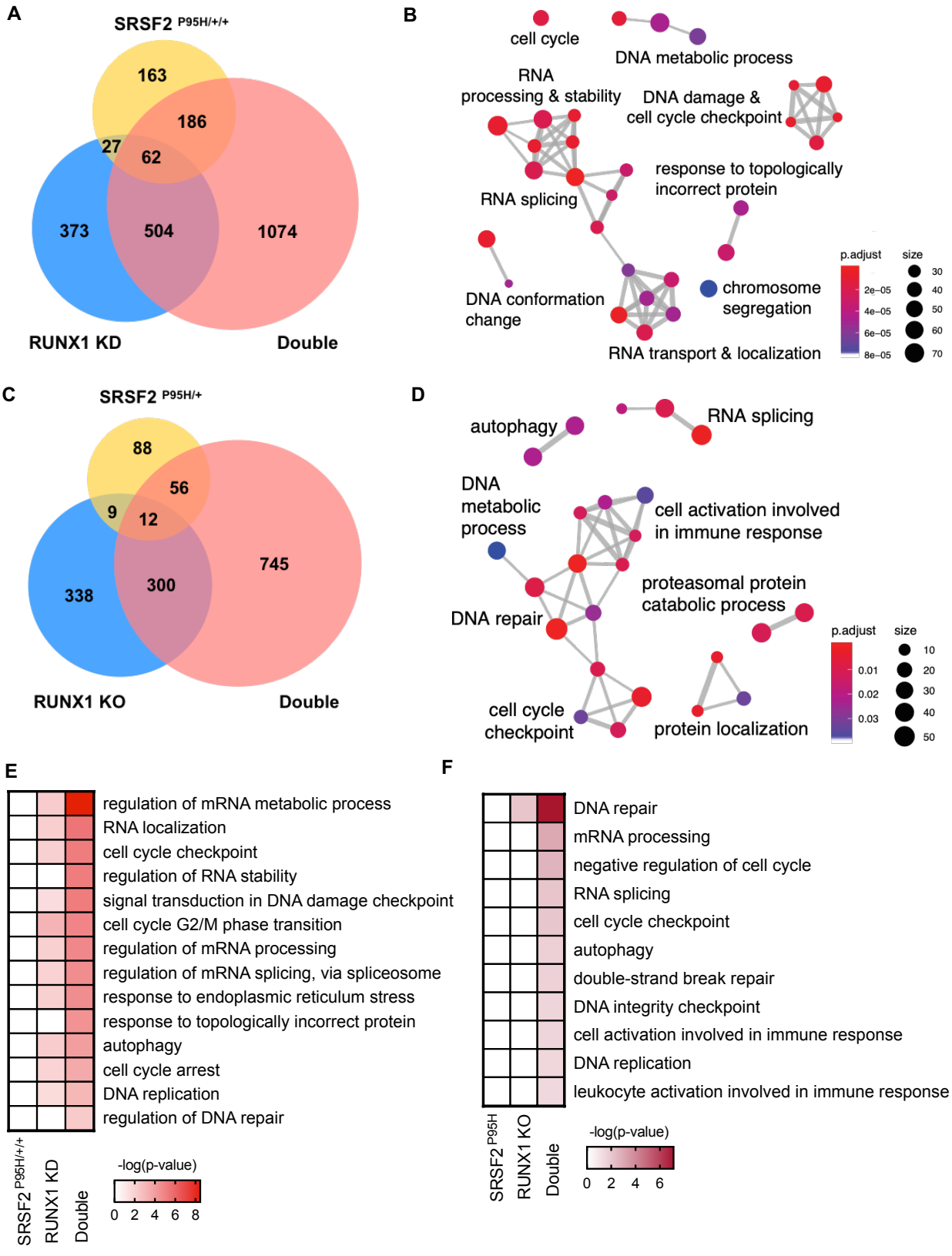
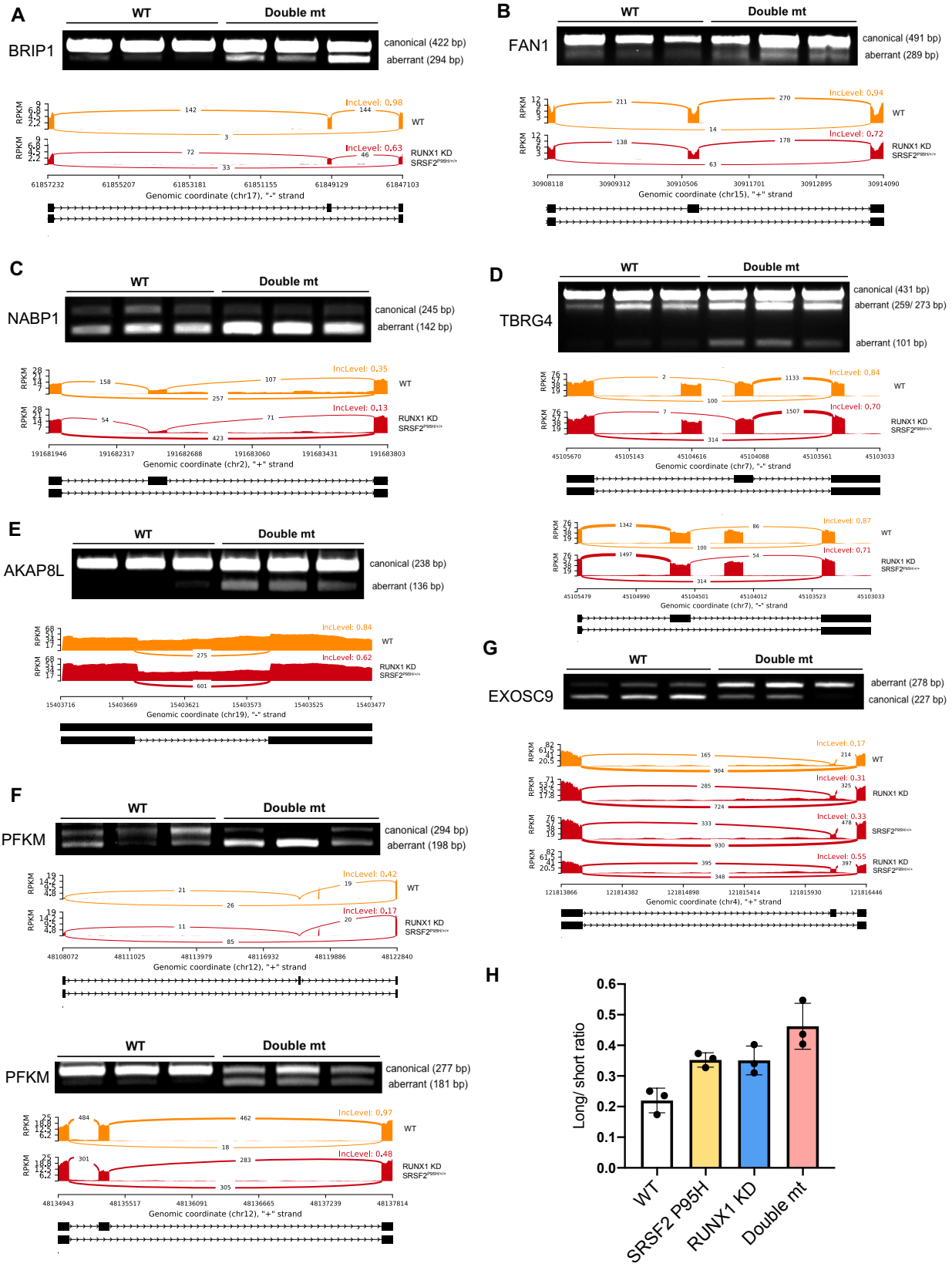

A
SRSF2 P95H/+/+
163
186
27
62
373
504
1074
RUNX1 KD
Double
B
cell cycle
DNA metabolic process
RNA processing & stability
DNA damage & cell cycle checkpoint
response to topologically incorrect protein
RNA splicing
DNA conformation change
chromosome segregation
RNA transport & localization
p.adjust
2e−05
4e−05
6e−05
8e−05
size
30
40
50
60
70
C
SRSF2 P95H/+
88
56
9
12
745
338
300
RUNX1 KO
Double
D
autophagy
RNA splicing
DNA metabolic process
cell activation involved in immune response
proteasomal protein catabolic process
DNA repair
cell cycle checkpoint
protein localization
p.adjust
0.01
0.02
0.03
size
10
20
30
40
50
E
regulation of mRNA metabolic process
RNA localization
cell cycle checkpoint
regulation of RNA stability
signal transduction in DNA damage checkpoint
cell cycle G2/M phase transition
regulation of mRNA processing
regulation of mRNA splicing, via spliceosome
response to endoplasmic reticulum stress
response to topologically incorrect protein
autophagy
cell cycle arrest
DNA replication
regulation of DNA repair
SRSF2 P95H/+/+
RUNX1 KD
Double
-log(p-value)
0 2 4 6 8
F
DNA repair
mRNA processing
negative regulation of cell cycle
RNA splicing
cell cycle checkpoint
autophagy
double-strand break repair
DNA integrity checkpoint
cell activation involved in immune response
DNA replication
leukocyte activation involved in immune response
SRSF2 P95H
RUNX1 KO
Double
-log(p-value)
0 2 4 6

Figure 2-16. Genes related to DNA damage checkpoint, cell cycle, glycolysis and RNA catabolic process were mis-spliced. (A-E) RT-PCR validation and sashimi plot of the genes related to DNA damage checkpoint: BRIP1(A), FAN1(B), NABP1(C); cell cycle: TBRG4 (D), AKAP8L (E); glycolysis: PFKM (F) and RNA catabolic process: EXOSC9 (G) with abnormal alternative splicing identified by RNA sequencing in K562 cells. (H) Bar graph showing increased level of mis-splicing between single and double mutant in EXOSC9 gene.


A
BRIP1
WT
Double mt
canonical (422 bp)
aberrant (294 bp)
IncLevel: 0.98
IncLevel: 0.63
RUNX1 KD
Genomic coordinate (chr17), "-" strand
B
FAN1
canonical (491 bp)
aberrant (289 bp)
IncLevel: 0.94
IncLevel: 0.72
Genomic coordinate (chr15), "+" strand
C
NABP1
canonical (245 bp)
aberrant (142 bp)
IncLevel: 0.35
IncLevel: 0.13
Genomic coordinate (chr2), "+" strand
D
TBRG4
canonical (431 bp)
aberrant (259/ 273 bp)
aberrant (101 bp)
IncLevel: 0.84
IncLevel: 0.70
IncLevel: 0.87
IncLevel: 0.71
Genomic coordinate (chr7), "-" strand
E
AKAP8L
canonical (238 bp)
aberrant (136 bp)
IncLevel: 0.84
IncLevel: 0.62
Genomic coordinate (chr19), "-" strand
F
PFKM
canonical (294 bp)
aberrant (198 bp)
IncLevel: 0.42
IncLevel: 0.17
Genomic coordinate (chr12), "+" strand
PFKM
canonical (277 bp)
aberrant (181 bp)
IncLevel: 0.97
IncLevel: 0.48
Genomic coordinate (chr12), "+" strand
G
EXOSC9
aberrant (278 bp)
canonical (227 bp)
IncLevel: 0.17
IncLevel: 0.31
IncLevel: 0.33
IncLevel: 0.55
RUNX1 KD
SRSF2
Genomic coordinate (chr4), "+" strand
H
Long/ short ratio
WT
SRSF2 P95H
RUNX1 KD
Double mt

2.3.5 SRSF2 P95H mutation and RUNX1 deficiency have additive effect in gene dysregulation and the co-existence worsen the viability outcomes *in vitro*

Next we evaluate how these two genetic alternations affect global gene expression. Unsupervised principal component analysis of differentially expressed genes in the biological replicates of each genotype in K562 cells clearly separated single and double mutant groups, while double mutant cells demonstrating greatest distance from control cells (Fig. 2-17A). In total, 1,221 1,391 and 2,721 genes were dysregulated in human K562 SRSF2 P95H cells, RUNX1 deficiency cells and double mutant cells, respectively (Fig. 2-17C, E). The greatest alternations in gene expression were seen in double mutant scenario in K562 cells, and double mutant cells recapitulated much of the gene expression dysregulation pattern (70.6% in SRSF2 P95H cell and 54.2% in RUNX1 deficiency cells) in single mutant K562 cells (Fig. 2-17C). Interestingly, within the events in the double mutant LK cells, 68% of the differentially expressed genes were also observed in RUNX1 single mutant cells and 31.2% of the differentially expressed genes were uniquely presented in the double mutant cells (Fig. 2-17D). These results suggest that the gene expression program is heavily affected by loss of RUNX1 and the coexistence of an SRSF2 mutation contributes to certain synergistic effects in transcriptional regulation in mouse LK cells context. GO cluster analysis of up-regulated genes in double mutant cells revealed signatures of myeloid leukocyte migration, ERK1 and ERK2 regulation, negative regulation of growth and acute inflammatory response (Fig. 2-18A, C). Analysis of down-regulation genes showed enrichment of leukocyte homeostasis, B cell receptor signaling, lymphocyte proliferation/differentiation and platelet activation (Fig. 2-18B, D), which explained the phenotypes of decreasing lymphocyte numbers, B cell percentage and platelet numbers in double mutant mice at gene expression level. In the single mutant K562 cells all demonstrated trend of slower growth, while double mutant cells

demonstrated the most serve effect on growth retardation (Fig. 2-19A) and increased of apoptosis (Fig. 2-19B), which also reflected the gene expression enrichment in upregulation of GO terms in negative regulation of growth and activation of ERK cascade that has been associated with the intrinsic apoptotic pathway (Fig. 2-18E)[103, 125]. Interesting, both RUNX1 KD and double mutant both dysregulates decent numbers of genes encoding RNA binding protein. Moreover, among the 28 RBPs shared between RUNX1 KD and double mutant cells (Fig. 2-20A), 10 of them have direct roles in splicing, and 90% of these RBP are downregulated (Fig. 2-20B), which conceivably elicit secondary defects that may explain the exacerbated splicing defects in double-mutant cells.

Together, these results suggest that the deficiency of compound RUNX1 and SRSF2 P95H mutations impairs multi-lineage hematopoiesis and exacerbates the disease phenotypes caused by single mutations alone. At the genome-wide level, loss of the transcription factor RUNX1 itself dysregulates splicing outcomes and cooperates with the splicing factor SRSF2 P95H mutation to further perturb the expression and splicing of key regulators involved in DNA damage and cell cycle checkpoint, RNA processing and hematopoietic cell development.

Figure 2-17. SRSF2 P95H mutation and RUNX1 deficiency have additive effect in gene dysregulation. (A) Unsupervised principal component analysis of differentially expressed genes in three biological replicates of each genotype in K562 cells. (B) Non-hierarchical clustering based on gene expression profiles in K562 cells. (C) Venn diagram showing the overlap between differentially expressed genes of each genotypes in K562 cells. (D) Venn diagram showing the overlap between differentially expressed genes of each genotypes in mouse LK cells. (E)Volcano plot of differentially expressed genes in each genotype in K562 cells.

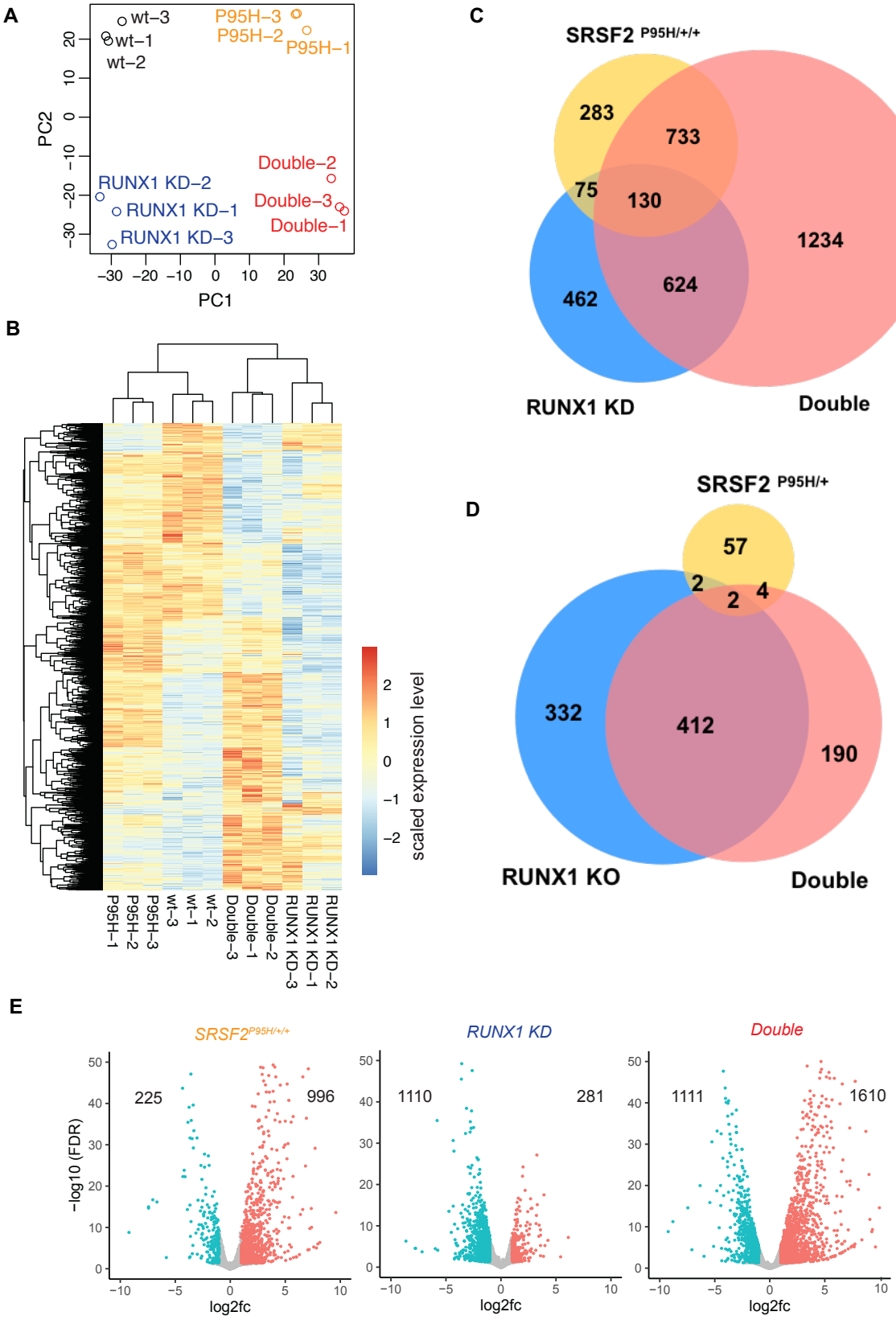
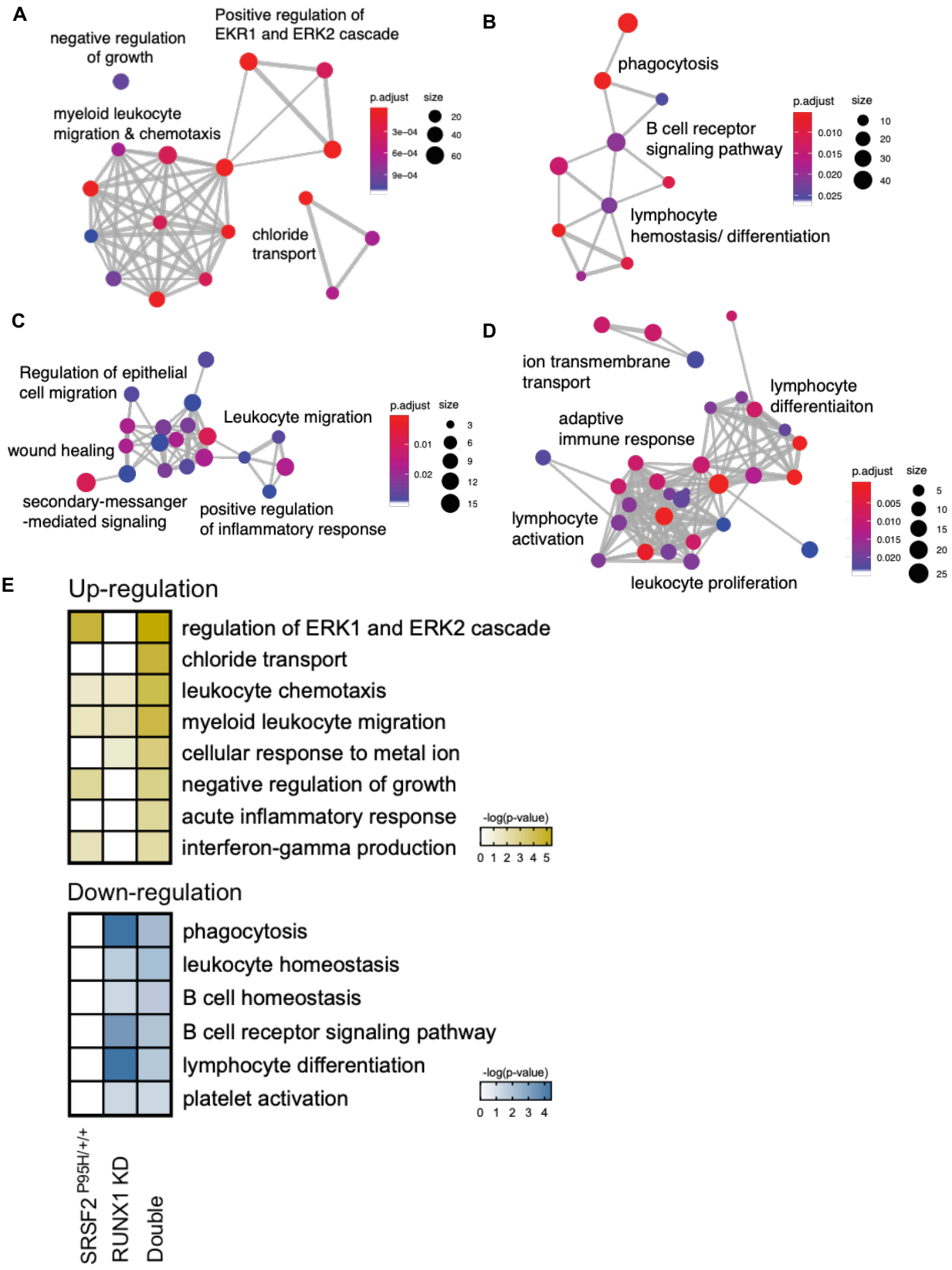

A
wt–3
wt–1
wt–2
P95H–3
P95H–2
P95H–1
Double–2
Double–3
Double–1
RUNX1 KD–2
RUNX1 KD–1
RUNX1 KD–3
PC2
PC1
B
scaled expression level
P95H–1
P95H–2
P95H–3
wt–3
wt–1
wt–2
Double–3
Double–1
Double–2
RUNX1 KD–3
RUNX1 KD–1
RUNX1 KD–2
C
SRSF2 P95H/+/+
283
733
75
130
1234
462
624
RUNX1 KD
Double
D
SRSF2 P95H/+
57
2
2
4
332
412
190
RUNX1 KO
Double
E
SRSF2P95H/+/+
RUNX1 KD
Double
225
996
1110
281
1111
1610
−log10 (FDR)
log2fc

Figure 2-18. The cooperation of double mutation further dysregulated pathways in blood cell development at gene expression level. (A-B) GO enrichment cluster analysis showing the dysregulated expression genes (up-regulation (A), down-regulation (B) in double mutant K562 cells. (C-D) GO enrichment cluster analysis showing the dysregulated expression genes (up-regulation (C), down-regulation (D) in double mutant mouse LK cells. (E) Heatmap demonstrating p-value of the GO terms enriched from the up/down-regulated genes identified in single and double mutation K562 cells.


A
negative regulation of growth
Positive regulation of EKR1 and ERK2 cascade
myeloid leukocyte migration & chemotaxis
chloride transport
p.adjust
size
3e−04
6e−04
9e−04
20
40
60
B
phagocytosis
B cell receptor signaling pathway
lymphocyte hemostasis/ differentiation
p.adjust
size
0.010
0.015
0.020
0.025
10
20
30
40
C
Regulation of epithelial cell migration
Leukocyte migration
wound healing
secondary-messanger -mediated signaling
positive regulation of inflammatory response
p.adjust
size
0.01
0.02
3
6
9
12
15
D
ion transmembrane transport
lymphocyte differentiaiton
adaptive immune response
lymphocyte activation
leukocyte proliferation
p.adjust
size
0.005
0.010
0.015
0.020
5
10
15
20
25
E
Up-regulation
regulation of ERK1 and ERK2 cascade
chloride transport
leukocyte chemotaxis
myeloid leukocyte migration
cellular response to metal ion
negative regulation of growth
acute inflammatory response
interferon-gamma production
-log(p-value)
0 1 2 3 4 5
Down-regulation
phagocytosis
leukocyte homeostasis
B cell homeostasis
B cell receptor signaling pathway
lymphocyte differentiation
platelet activation
-log(p-value)
0 1 2 3 4
SRSF2 P95H/+/+
RUNX1 KD
Double

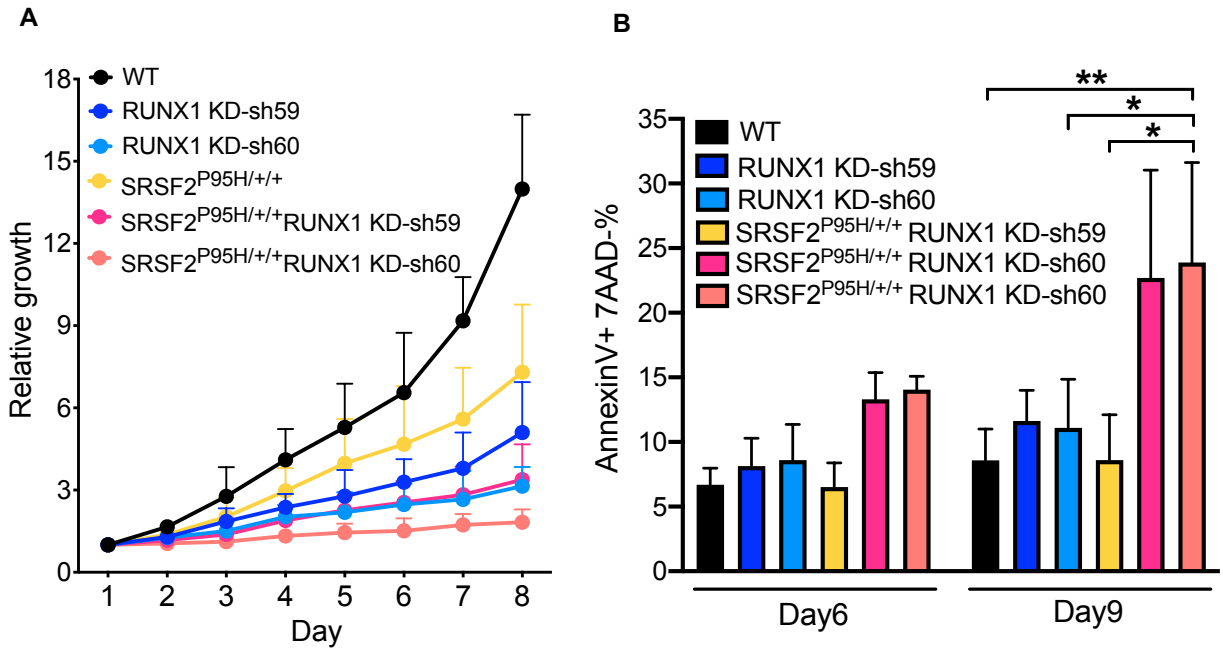


Figure 2-19. The co-existence of RUNX1 loss and SRSF2 P95H caused growth arrest increased apoptosis in K562 cells. (A) Growth curve of single/double mutant K562 cells. Cells were seeded in triplicate in three independent experiments and counted with Trypan blue. (B) Apoptosis assay. Cells were seeded in duplicate in three independent experiments. Early apoptotic cells are defined as AnnexinV+ 7AAD- cells. *p<0.05, **p<0.01.

A

B

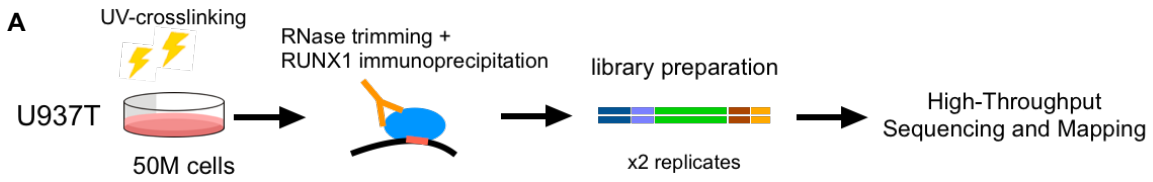
Total RBPs		Total dysregulated genes	dysregulated RBPs		Down-regulated genes	Down-regulated RBPs
1542	Double	2771	86	Double	1111	53
	RUNX1 KD	1391	46	RUNX1 KD	1110	39
	SRSF2 P95H/+/+	1221	29	SRSF2 P95H/+/+	225	6

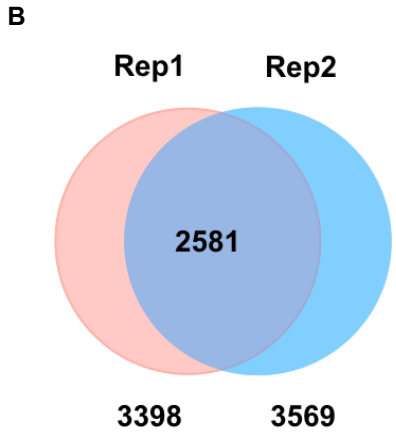
Figure 2-20. Loss of RUNX1 dysregulated expression of RNA binding proteins. (A) Numbers of dysregulated genes encoding RNA binding protein in RUNX1 KD and double mutant K562 cells. (B) Table showing dysregulated and down-regulated RBPs in single and double mutation conditions.

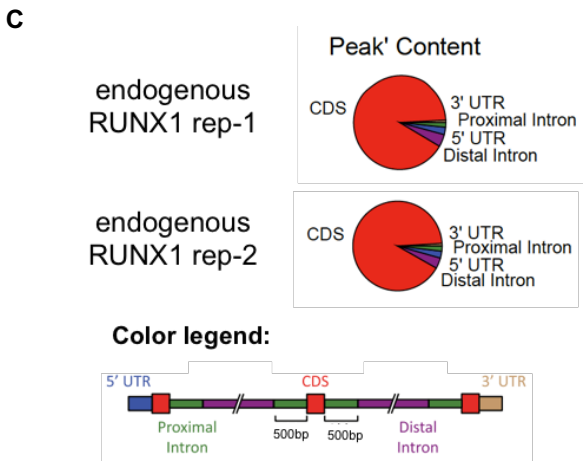
2.3.6 Identify RUNX1 as potential RNA binding protein by eCLIP-seq

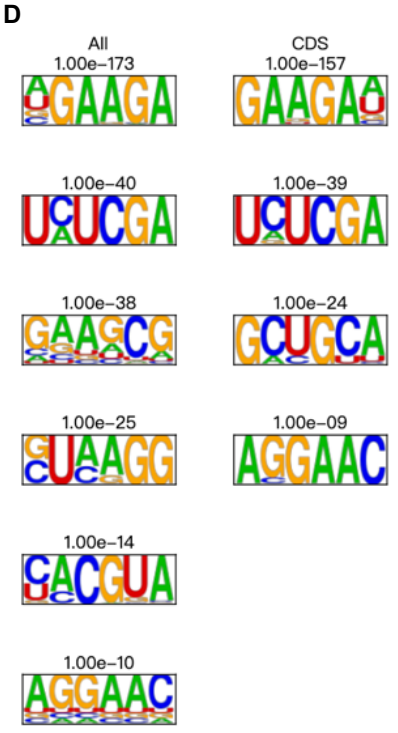
Due to the splicing alternations caused by loss of RUNX1 and the ability of RUNX1 to bind RNA aptamer[102], we further explore the possibility of RUNX1 as a RNA binding protein. We utilized enhanced UV crosslinking and immunoprecipitation (eCLIP) for transcriptome-wide discovery of RNA-binding protein binding targets of RUNX1[126]. We prepared the eCLIP library through immunoprecipitation of endogenous RUNX1 in U937T cells (Fig. 2-21A). The two replicates are highly reproducible, yielding 3398 and 3569 target genes, while 2581 of them are shared between two samples (Fig. 2-21B). We next analyzed the differentially bound region of the peaks. Interestingly, the peaks predominately located in the region of coding sequences (Fig. 2-21C). The exon junction binding indicated its potential role in regulating splicing. The motif analysis also revealed significant motifs such as GAAGA and UCUCGA (Fig. 2-21D). To gain functional insight of the target genes, we performed pathway analysis using the 2581 shared events between two replicates. The enriched pathways including RNA processing, DNA replication and DNA repair. Though the results of RUNX1 eCLIP-seq provides us certain evidence in RUNX1 as an RNA binding protein, more biochemical validation is required for future studies.

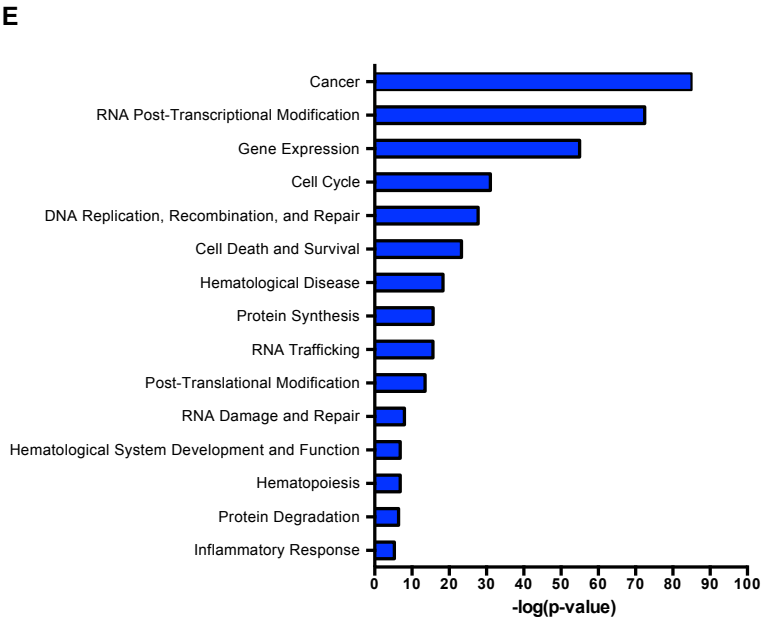
Figure 2-21. Identification of potential RUNX1 RNA targets through eCLIP-seq. (A) Schema for eCLIP-seq protocol. (B) Venn diagram showing overlap for two replicates of RUNX1 eCLIP-seq targets (FC >2; $p < 0.001$) (C) Distribution of annotated features near eCLIP peak of two replicates. (D) Top motif derived from all eCLIP peaks. (E) Pathway analysis of RUNX1 targets in U937T cells.


A
UV-crosslinking
U937T
50M cells
RNase trimming +
RUNX1 immunoprecipitation
library preparation
x2 replicates
High-Throughput
Sequencing and Mapping


B
Rep1
Rep2
2581
3398
3569


C
Peak' Content
endogenous
RUNX1 rep-1
CDS
3' UTR
Proximal Intron
5' UTR
Distal Intron
endogenous
RUNX1 rep-2
CDS
3' UTR
Proximal Intron
5' UTR
Distal Intron
Color legend:
5' UTR
CDS
3' UTR
Proximal
Intron
500bp
500bp
Distal
Intron


D
All
1.00e-173
CDS
1.00e-157
1.00e-40
1.00e-39
1.00e-38
1.00e-24
1.00e-25
1.00e-09
1.00e-14
1.00e-10


E
Cancer
RNA Post-Transcriptional Modification
Gene Expression
Cell Cycle
DNA Replication, Recombination, and Repair
Cell Death and Survival
Hematological Disease
Protein Synthesis
RNA Trafficking
Post-Translational Modification
RNA Damage and Repair
Hematological System Development and Function
Hematopoiesis
Protein Degradation
Inflammatory Response
0
10
20
30
40
50
60
70
80
90
100
-log(p-value)

2.4 Discussion

Mutations in RNA splicing factors are commonly identified in hematological malignancies, and the positive (coexistence) and negative (mutually exclusive) correlations between this category with other subset of mutations are found, indicating the potential functional interactions between these mutations in disease etiology. Significant coexisting mutation pairs in MDS context have identified in multiple patient studies, including combinations of SRSF2 with IDH1/IDH2/TET2/RUNX1/ASXL1/STAG2/CUX1 mutations, U2AF1 with ASXL1/ DNMT3A mutations, SF3B1 with DNMT3A mutations and ZRSR2 with TET2 mutations[28, 30, 31, 89, 91, 92]. However, how these various combinations of genetic interactions participating in pathogenesis biologically yet to be fully explored experimentally in isogenic manners. Here we report mouse and human cell isogeneic models harboring SRSF2 P95H mutations and RUNX1 deficiency to elucidate the cooperative effect of spliceosome and transcription perturbation. The double mutant mice developed MDS-related phenotypes related with both SRSF2 P95H (eg: leukopenia, anemia, neutrophils with dysplasia) and RUNX1 deficiency (eg: leukopenia, thrombocytopenia, increased percentage of myeloid cells), resulting in more severe MDS symptoms were developed though no leukemia transformation was observed. At the global splicing level, in both of the isogenic human myeloid cell line and primary mouse progenitor models, the most profound changes are in the double mutation context, and both of the mutations mainly contributed to the increase in cassette exon events.

Lately, several blood malignancies models with spliceosome genes related coexisted mutations have been established as well. For example, the mutation combination with splicing and epigenetic regulation factors were demonstrated by Srsf2-Idh2 and Srsf2-Tet2 double mutant mouse models[127]. The expression of either IDH2 or SRSF2 mutations caused distinct splicing

changes, and the co-expression resulted in more dramatic splicing alternations than single mutation in both mouse models and AML patients due to increased DNA methylation mediated by IDH2 mutations thus altered RNA splicing. The co-expression of these two mutants also resulted in lethal myelodysplasia and promote leukemogenesis. To the contrast, in cases with SRSF2 and TET2 mutations have much fewer splicing changes when compared to SRSF2 mutations alone and the loss of Tet2 in the Srsf2 mutant mice was insufficient to rescue the impaired self-renewal capacity of Srsf2 single mutant cells. Coexisted mutations with splicing and DNA damage regulator was also investigated through the Sf3b1-Atm double mutant chronic lymphocytic leukemia (CLL) mouse model[128], and the deletion of Atm overcame Sf3b1 mutation induced senescence and caused CLL in 3 of 33 double mutant mice. The deletion of Atm induced most changes in splicing events when compared to Sf3b1 mutant alone or double mutant cells, and a combined effect of the two mutations on splicing was observed. Chromosomal abnormities such as inversions between chromosome 3q21 and 3q26 ("inv(3)/ t(3;3)") also found to be co-mutated with SF3B1 [129], the gene expression alternations driven by inv(3) in the double mutant mouse cells is further augmented by the presence of Sf3b1 mutations, and caused alterations in group of genes in RNA processing and splicing. The above co-mutated isogenic models demonstrated the diverse cooperative outcomes of spliceosome mutations with other types of genetic alternations.

The combination of splicing and transcription factor mutations was also explored in U2af1-Runx1 double mutation mouse model[130]. The double mutant cells demonstrated less numbers of gene with altered expression and splicing when compared to single mutants, implicating certain counteract effect mediated by these two genetic interactions at gene regulation level. However, the differences of the mis-splicing types, the mis-spliced target genes and the level of splicing changes between single and double mutant were not further evaluated. Our double mutant mouse and cell

line models don't indicate transformation under normal hematopoiesis without exposure to genotoxic agents, and the U2af1-Runx1 mouse model also showed mild MDS phenotypes under steady state condition, which agrees with our observation. These results indicate the tertiary mutation which could drive proliferation may be critical to overcome the growth defect of splicing factor and Runx1 double mutation to promote the leukemogenesis. The U2af1-Runx1 double mutant models had 1 out of 16 mice developed primary AML after exposure to a mutagen, N-ethyl-N-nitrosourea (ENU). The whole exome sequencing of the leukemia samples identified somatic mutations in Tet2, Idh1 and Gata2 in the transformation process. The tertiary co-existed mutations identified in MDS and CMML patients with SRSF2-RUNX1 double mutations include mutations such as TET2 and NRAS which promote proliferation, suggesting such mutation could be critical to facilitate the leukemic transformation process[28, 90], and the pre-deposited splicing and gene dysregulation caused by double mutations may be the basis for pathogenesis when cells further acquire additional mutations and/or encounter environmental stimuli.

Intriguingly, in both of our isogenic models in human and mouse, RUNX1 deficiency imparts profound aberrant splicing. The underlying mechanisms can include direct and indirect effects. The possible direct RNA binding of RUNX1 was suggested based on studies that RNA aptamers selected against the RUNX1 RHD–CBFβ complex were able to efficiently compete with DNA *in vitro* for RUNX1 binding[102]. To explore the possibility of direct RNA binding of RUNX1, we performed eCLIP-seq and identified potential RNA targets and significant motifs, though the underlying mechanism and the RNA binding sites required further experimental validation. In our study, we also found downregulation of RNA binding protein in RUNX1 knockdown context, indicating the additional in-direct mechanism. Other mechanism potentially may be involved

includes co-transcriptional splicing of exons[131, 132], and which relied to be further dissected under the transcriptional and splicing factor mutant conditions.

In conclusion, the present study described the *in vivo* cooperative effects of the Srsf2-Runx1 double mutant on the MDS phenotypes and the global genomics alternations caused by the interaction of these two mutations. Lately, the increased mis-splicing in cancer transcriptomes relative to normal tissues was shown, on average, tumor tissues harbor around 20% more alternative spliced events than normal counterpart[132]. In MDS, it was also shown that 26.9% of the expressed genes were aberrantly spliced in MDS patients compared with healthy donors. The higher aberrant AS score was associated with shorter overall survival and served as unfavorable prognostic factor[133]. Due to the heterogeneous nature of MDS, various mutations coexist in patient samples. It will be valuable to further investigate additional mutation combinations in isogenic manners for aberrant AS outcomes to understand their potential synergistic, additive or counteractive effect in global splicing. Due to the increased recognition that genetic alternations disturbing RNA splicing is common in cancer, the understanding of the splicing outcomes may propel generation of novel therapeutics by modulating RNA splicing globally[123]. Together, these data engendered insights to understand the molecular pathogenesis of MDS mutations and demonstrated the importance of additive effects in aberrant splicing caused by non-spliceosome gene mutations.

2.5 Materials and Methods

Mice

C57BL/6 (CD45.2) and congenic strain B6.SJL-Ptprca Pep3b/BoyJ (PEP3, CD45.1) mice were obtained from Jackson Laboratory. Conditional Srsf2 P95H/WT mice were provided by Dr.

Omar Abdel-Wahab[64]. Runx1f/f mice were provided by Dr. Nancy Speck[97]. Polyinosine-polycytosine (pIpC) (Sigma) was injected intraperitoneally (i.p.) to mice at 12μg/g every other day for seven injections. Genotyping PCR was performed using the primers described previously. All the procedures were approved by the institutional animal care and use committee.

Flow cytometry

Primary cells were collected from mice, treated in ammonium-chloridepotassium buffer, stained. The following antibodies were used: PerCP-Cy5.5 or PE-Cy5-conjugated lineage antibodies (CD3, CD4, CD8a, B220, CD19, Gr1, CD11b, and Ter119), APC or PE-Cy5-conjugated Sca-1, PE-Cy7 or APC conjugated c-Kit, PE-conjugated CD48, biotin-conjugated CD150 with streptavidin-conjugated APC-Cy7, PE-conjugated FcγR, FITC-conjugated CD34, and PerCP-Cy5.5 or PE-conjugated IL7Rα (all antibodies from Biolegend). After staining, cells were analyzed on a BD FACS Canto flow cytometer.

RNA-seq and data analysis

The RNA from mouse LK cells and K562 cells were extracted via Trizol reagent, and poly-A selected RNA was used for library preparation using TruSeq RNA Sample Preparation Kit (Illumina) according to the manufacturer's protocol and sequenced on the Illumina platforms by HiSeq4000 at UCSD IGM core (mouse samples) and NovaSeq6000 at UCSF core (K562 samples). RNA-seq reads of LK and K562 cells were mapped against the GRCm38 and GRCh38 respectively with STAR software. The DEseq2 package was used to perform differential gene expression analysis. Fold changes > 2 or < 1/2 and adjusted p-value < 0.05 were used to select differentially expressed genes (DEGs) in K562 cells. Altered splicing events in K562 cells were identified with rMATS by requiring at least 5% inclusion ratio difference (-cstat) under the FDR <0.05. Similar stringent cutoffs were adopted in LK cells to get a comparable number of

genes/events with K562 cells. DEGs in LK cells were selected if fold changes > 1.5 or < 2/3 and adjusted p-value < 0.05, and altered splicing events were selected if inclusion ratio difference >5% and FDR <0.01. Gene list enrichment analyses were all done using R package clusterProfiler. Top 30 enriched GO terms if exist were visualized with R package enrichplot.

Cell culture

293T cells were cultured in DMEM containing 10% bovine calf serum, 10 mM glutamine and penicillin (100 IU)/streptomycin (100 μg/ml). CRISPR-edited K562 cells were provided by Dr. Rafael Bajar[65] and cultured in RPMI with 10% fetal bovine serum, 10 mM glutamine and penicillin (100 IU)/streptomycin (100 μg/ml).

Bone marrow transplantation

For competitive and noncompetitive bone marrow transplantation, total BM cells were harvested from indicated donor mice without treatment. In noncompetitive settings, 2 million cells in 200 μl PBS were intravenously injected (i.v.) into lethally irradiated (900rad) CD45.2 recipient mice. In competitive settings, test cells were mixed with competitor cells (CD45.1 cells) at a 1:1 ratio, and 2 million cells in 200 μl PBS were i.v. injected into lethally irradiated CD45.1 recipient mice.

Western blotting

Protein samples were denatured in 1× loading buffer [10% glycerol, 2% SDS, 10 mM DTT and 50 mM Tris-HCl (pH 6.8)]. Protein concentration was adjusted, and protein samples were loaded on SDS polyacrylamide gels after adding bromophenol blue (0.05%). Primary anti-RUNX1 antibody (4434S; Cell Signaling) and anti-β-actin antibody (A1978; Sigma-Aldrich) were used. Signals from fluorophore-conjugated secondary antibodies were detected with the Odyssey system (LI-COR).

Lentivirus preparation and infection

Lentiviral vector MISSION pLKO.1-shRNA-puro constructs targeting human RUNX1 (TRCN0000013659 and TRCN0000013660) were obtained from the Functional Genomics Center at La Jolla Institute for Immunology. For virus production, 293T were transfected with pLKO.1 lentivirus vectors, psPAX2, pMD2.G using polyethylenimine (Polysciences Inc) for 10 hours and then the medium was changed from DMEM to RPMI. 48 hours later, supernatant was collected and filtered through 0.45μm filter and added to K562 cells, along with 1%HEPES and 0.1% polybrene (final concentration 4μg/ml). The cells were spinoculated at 1200g for 3 hr at 32°C. Infections were performed twice on consecutive days.

Reverse transcription and polymerase chain reaction (RT-PCR)

Total RNA was extracted by using Trizol reagent (ThermoFisher Scientific, #15596026) according to manufacturer's instructions. RT reactions were carried out by using the First Strand cDNA Synthesis Kit (MCLab). PCR was performed using KOD Hot Start DNA Polymerase (EMD Millipore). Primers used for PCR are as follows.

BRIP1(F): 5'-GGAAACCAGCAGATGAGGGCGT-3'

BRIP1(R): 5'-TTTCTGTGGCGAAAAGGAGTTT-3'

FAN1(F): 5'-GGCTCTTTCAACGTAAATTAAGCTGG-3'

FAN1(R): 5'-CTCCATTCGGCCGAGGTTGACC-3'

NABP1(F): 5'-TCCATGTGGAAAGGATGTCTGACA-3'

NABP1 (R): 5'-CCCTTTGTTCTGCTGTCCTCGA-3'

TBRG4(F): 5'-TCGGAGCCACTAATGAACCGCC-3'

TBRG4 (R): 5'-CGCGCAAAAGCCAGAAGTACGC-3'

PFKM-1(F): 5'-CTGGGGAAGCTTCTACTTCC-3'

PFKM-1 (R): 5'-CTTGGGCATCTCCACCAGAG-3'

PFKM-2(F): 5'-TGAGGGTGCAATTGACAAGA-3'

PFKM-2 (R): 5'-GCCTTGGTCACATCTTTGGT-3'

EXOSC9 (F): 5'-TGGAACTGCCCAAATTGGAGAGGG-3'

EXOSC9 (R): 5'-TCTTTGGATTCTTGTCTGGTTCCA-3'

AKAP8L (F): 5'-GCTACGAGGGCTATGGCTATGGC-3'

AKAP8L (R): 5'-TCTCCACCTGAGCCGTACACGC-3'

eCLIP-seq library preparation and sequencing

eCLIP library preparation and data analysis performed as described previously[126]. Two replicate experiments in U937T cells were performed with 50×10^6 cells and N-terminal RUNX1 antibody[134] for endogenous RUNX1 recognition was used. Samples were sequenced on the Illumina HiSeq 4000 platform as SR75 reads at UCSD IGM core.

Statistics

Results from repeated experiments were represented as mean ± standard deviation (SD), and Student's t tests were performed. Whenever asterisks are used to indicate the statistical significances, *stands for $p < 0.05$; **for $p < 0.01$, and ***for $p < 0.001$.

2.6 Acknowledgements

Chapter 2, in full is currently being prepared for submission for publication of the material by Yi-Jou Huang, Jia-Yu Chen, Ming Yan, Sayuri Miyauchi, Amanda Davis, Rafael Bejar, Liang Chen, Xiang-Dong Fu and Dong-Er Zhang. "RUNX1 deficiency cooperates with SRSF2 mutation to further disrupt RNA splicing program and exacerbate myelodysplastic syndromes phenotypes." Yi-Jou Huang was the primary investigator and author of this material.

References

1. Ma X. Epidemiology of myelodysplastic syndromes. *Am J Med*. 2012;125:S2-5.

2. Cogle CR, Kurtin SE, Bentley TG, Broder MS, Chang E, Megaffin S, Fruchtman S, Petrone ME and Mukherjee S. The Incidence and Health Care Resource Burden of the Myelodysplastic Syndromes in Patients in Whom First-Line Hypomethylating Agents Fail. *Oncologist*. 2017;22:379-385.

3. Ma X, Does M, Raza A and Mayne ST. Myelodysplastic syndromes: incidence and survival in the United States. *Cancer*. 2007;109:1536-42.

4. Sekeres MA. Epidemiology, natural history, and practice patterns of patients with myelodysplastic syndromes in 2010. *J Natl Compr Canc Netw*. 2011;9:57-63.

5. Troy JD, Atallah E, Geyer JT and Saber W. Myelodysplastic syndromes in the United States: an update for clinicians. *Ann Med*. 2014;46:283-9.

6. Gangat N, Patnaik MM, Begna K, Kourelis T, Al-Kali A, Elliott MA, Hogan WJ, Letendre L, Litzow MR, Knudson RA, Ketterling RP, Hodnefield JM, Hanson CA, Pardanani AD and Tefferi A. Primary Myelodysplastic Syndromes: The Mayo Clinic Experience With 1000 Patients. *Mayo Clin Proc*. 2015;90:1623-38.

7. Steensma DP. Myelodysplastic Syndromes: Diagnosis and Treatment. *Mayo Clin Proc*. 2015;90:969-83.

8. Greenberg PL, Tuechler H, Schanz J, Sanz G, Garcia-Manero G, Sole F, Bennett JM, Bowen D, Fenaux P, Dreyfus F, Kantarjian H, Kuendgen A, Levis A, Malcovati L, Cazzola M, Cermak J, Fonatsch C, Le Beau MM, Slovak ML, Krieger O, Luebbert M, Maciejewski J, Magalhaes SM, Miyazaki Y, Pfeilstocker M, Sekeres M, Sperr WR, Stauder R, Tauro S, Valent P, Vallespi T, van de Loosdrecht AA, Germing U and Haase D. Revised international prognostic scoring system for myelodysplastic syndromes. *Blood*. 2012;120:2454-65.

9. Kantarjian H, O'Brien S, Ravandi F, Cortes J, Shan J, Bennett JM, List A, Fenaux P, Sanz G, Issa JP, Freireich EJ and Garcia-Manero G. Proposal for a new risk model in myelodysplastic syndrome that accounts for events not considered in the original International Prognostic Scoring System. *Cancer*. 2008;113:1351-61.

10. Malcovati L, Della Porta MG, Strupp C, Ambaglio I, Kuendgen A, Nachtkamp K, Travaglino E, Invernizzi R, Pascutto C, Lazzarino M, Germing U and Cazzola M. Impact of the degree of anemia on the outcome of patients with myelodysplastic syndrome and its integration into the WHO classification-based Prognostic Scoring System (WPSS). *Haematologica*. 2011;96:1433-40.

11. Bejar R, Stevenson K, Abdel-Wahab O, Galili N, Nilsson B, Garcia-Manero G, Kantarjian H, Raza A, Levine RL, Neuberg D and Ebert BL. Clinical effect of point mutations in myelodysplastic syndromes. *N Engl J Med*. 2011;364:2496-506.

12. Bejar R, Stevenson KE, Caughey BA, Abdel-Wahab O, Steensma DP, Galili N, Raza A, Kantarjian H, Levine RL, Neuberg D, Garcia-Manero G and Ebert BL. Validation of a prognostic model and the impact of mutations in patients with lower-risk myelodysplastic syndromes. *J Clin Oncol*. 2012;30:3376-82.

13. Malcovati L, Hellstrom-Lindberg E, Bowen D, Ades L, Cermak J, Del Canizo C, Della Porta MG, Fenaux P, Gattermann N, Germing U, Jansen JH, Mittelman M, Mufti G, Platzbecker U, Sanz GF, Selleslag D, Skov-Holm M, Stauder R, Symeonidis A, van de Loosdrecht AA, de Witte T, Cazzola M and European Leukemia N. Diagnosis and treatment of primary myelodysplastic syndromes in adults: recommendations from the European LeukemiaNet. *Blood*. 2013;122:2943-64.

14. Ades L, Itzykson R and Fenaux P. Myelodysplastic syndromes. *Lancet*. 2014;383:2239-52.

15. Walter MJ, Shen D, Shao J, Ding L, White BS, Kandoth C, Miller CA, Niu B, McLellan MD, Dees ND, Fulton R, Elliot K, Heath S, Grillot M, Westervelt P, Link DC, DiPersio JF, Mardis E, Ley TJ, Wilson RK and Graubert TA. Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. *Leukemia*. 2013;27:1275-82.

16. Kulasekararaj AG, Mohamedali AM and Mufti GJ. Recent advances in understanding the molecular pathogenesis of myelodysplastic syndromes. *Br J Haematol*. 2013;162:587-605.

17. Bejar R and Steensma DP. Recent developments in myelodysplastic syndromes. *Blood*. 2014;124:2793-803.

18. Graubert TA, Shen D, Ding L, Okeyo-Owuor T, Lunn CL, Shao J, Krysiak K, Harris CC, Koboldt DC, Larson DE, McLellan MD, Dooling DJ, Abbott RM, Fulton RS, Schmidt H, Kalicki-Veizer J, O'Laughlin M, Grillot M, Baty J, Heath S, Frater JL, Nasim T, Link DC, Tomasson MH, Westervelt P, DiPersio JF, Mardis ER, Ley TJ, Wilson RK and Walter MJ. Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nat Genet*. 2011;44:53-7.

19. 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 WK, Miyawaki S, Sugano S, Haferlach C, Koeffler HP, Shih LY, Haferlach T, Chiba S, Nakauchi H, Miyano S and Ogawa S. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*. 2011;478:64-9.

20. Yoshida K and Ogawa S. Splicing factor mutations and cancer. *Wiley Interdiscip Rev RNA*. 2014;5:445-59.

21. Papaemmanuil E, Cazzola M, Boultwood J, Malcovati L, Vyas P, Bowen D, Pellagatti A, Wainscoat JS, Hellstrom-Lindberg E, Gambacorti-Passerini C, Godfrey AL, Rapado I, Cvejic A, Rance R, McGee C, Ellis P, Mudie LJ, Stephens PJ, McLaren S, Massie CE, Tarpey PS, Varela I, Nik-Zainal S, Davies HR, Shlien A, Jones D, Raine K, Hinton J, Butler AP, Teague JW, Baxter EJ, Score J, Galli A, Della Porta MG, Travaglino E, Groves M, Tauro S, Munshi NC, Anderson KC, El-Naggar A, Fischer A, Mustonen V, Warren AJ, Cross NC, Green AR, Futreal PA, Stratton

MR, Campbell PJ and Chronic Myeloid Disorders Working Group of the International Cancer Genome C. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med*. 2011;365:1384-95.

22. Visconte V, Makishima H, Jankowska A, Szpurka H, Traina F, Jerez A, O'Keefe C, Rogers HJ, Sekeres MA, Maciejewski JP and Tiu RV. SF3B1, a splicing factor is frequently mutated in refractory anemia with ring sideroblasts. *Leukemia*. 2012;26:542-5.

23. Zhang SJ, Rampal R, Manshouri T, Patel J, Mensah N, Kayserian A, Hricik T, Heguy A, Hedvat C, Gonen M, Kantarjian H, Levine RL, Abdel-Wahab O and Verstovsek S. Genetic analysis of patients with leukemic transformation of myeloproliferative neoplasms shows recurrent SRSF2 mutations that are associated with adverse outcome. *Blood*. 2012;119:4480-5.

24. Damm F, Nguyen-Khac F, Fontenay M and Bernard OA. Spliceosome and other novel mutations in chronic lymphocytic leukemia and myeloid malignancies. *Leukemia*. 2012;26:2027-31.

25. Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, Werner L, Sivachenko A, DeLuca DS, Zhang L, Zhang W, Vartanov AR, Fernandes SM, Goldstein NR, Folco EG, Cibulskis K, Tesar B, Sievers QL, Shefler E, Gabriel S, Hacohen N, Reed R, Meyerson M, Golub TR, Lander ES, Neuberg D, Brown JR, Getz G and Wu CJ. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med*. 2011;365:2497-506.

26. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, Bloomfield CD, Cazzola M and Vardiman JW. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016;127:2391-405.

27. Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P, Yoon CJ, Ellis P, Wedge DC, Pellagatti A, Shlien A, Groves MJ, Forbes SA, Raine K, Hinton J, Mudie LJ, McLaren S, Hardy C, Latimer C, Della Porta MG, O'Meara S, Ambaglio I, Galli A, Butler AP, Walldin G, Teague JW, Quek L, Sternberg A, Gambacorti-Passerini C, Cross NC, Green AR, Boultwood J, Vyas P, Hellstrom-Lindberg E, Bowen D, Cazzola M, Stratton MR, Campbell PJ and Chronic Myeloid Disorders Working Group of the International Cancer Genome C. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood*. 2013;122:3616-27; quiz 3699.

28. 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 HP, Klein HU, Dugas M, Aburatani H, Kohlmann A, Miyano S, Haferlach C, Kern W and Ogawa S. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia*. 2014;28:241-7.

29. Lindsley RC, Mar BG, Mazzola E, Grauman PV, Shareef S, Allen SL, Pigneux A, Wetzler M, Stuart RK, Erba HP, Damon LE, Powell BL, Lindeman N, Steensma DP, Wadleigh M, DeAngelo DJ, Neuberg D, Stone RM and Ebert BL. Acute myeloid leukemia ontogeny is defined by distinct somatic mutations. *Blood*. 2015;125:1367-76.

30. Damm F, Kosmider O, Gelsi-Boyer V, Renneville A, Carbuccia N, Hidalgo-Curtis C, Della Valle V, Couronne L, Scourzic L, Chesnais V, Guerci-Bresler A, Slama B, Beyne-Rauzy O, Schmidt-Tanguy A, Stamatoullas-Bastard A, Dreyfus F, Prebet T, de Botton S, Vey N, Morgan MA, Cross NC, Preudhomme C, Birnbaum D, Bernard OA, Fontenay M and Groupe Francophone des M. Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes. *Blood*. 2012;119:3211-8.

31. Thol F, Kade S, Schlarmann C, Loffeld P, Morgan M, Krauter J, Wlodarski MW, Kolking B, Wichmann M, Gorlich K, Gohring G, Bug G, Ottmann O, Niemeyer CM, Hofmann WK, Schlegelberger B, Ganser A and Heuser M. Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. *Blood*. 2012;119:3578-84.

32. Vannucchi AM, Lasho TL, Guglielmelli P, Biamonte F, Pardanani A, Pereira A, Finke C, Score J, Gangat N, Mannarelli C, Ketterling RP, Rotunno G, Knudson RA, Susini MC, Laborde RR, Spolverini A, Pancrazzi A, Pieri L, Manfredini R, Tagliafico E, Zini R, Jones A, Zoi K, Reiter A, Duncombe A, Pietra D, Rumi E, Cervantes F, Barosi G, Cazzola M, Cross NC and Tefferi A. Mutations and prognosis in primary myelofibrosis. *Leukemia*. 2013;27:1861-9.

33. Quesada V, Conde L, Villamor N, Ordonez GR, Jares P, Bassaganyas L, Ramsay AJ, Bea S, Pinyol M, Martinez-Trillos A, Lopez-Guerra M, Colomer D, Navarro A, Baumann T, Aymerich M, Rozman M, Delgado J, Gine E, Hernandez JM, Gonzalez-Diaz M, Puente DA, Velasco G, Freije JM, Tubio JM, Royo R, Gelpi JL, Orozco M, Pisano DG, Zamora J, Vazquez M, Valencia A, Himmelbauer H, Bayes M, Heath S, Gut M, Gut I, Estivill X, Lopez-Guillermo A, Puente XS, Campo E and Lopez-Otin C. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat Genet*. 2011;44:47-52.

34. Dolnik A, Engelmann JC, Scharfenberger-Schmeer M, Mauch J, Kelkenberg-Schade S, Haldemann B, Fries T, Kronke J, Kuhn MW, Paschka P, Kayser S, Wolf S, Gaidzik VI, Schlenk RF, Rucker FG, Dohner H, Lottaz C, Dohner K and Bullinger L. Commonly altered genomic regions in acute myeloid leukemia are enriched for somatic mutations involved in chromatin remodeling and splicing. *Blood*. 2012;120:e83-92.

35. Zhou Z and Fu XD. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma*. 2013;122:191-207.

36. Graveley BR and Maniatis T. Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. *Mol Cell*. 1998;1:765-71.

37. Liu HX, Chew SL, Cartegni L, Zhang MQ and Krainer AR. Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. *Mol Cell Biol*. 2000;20:1063-71.

38. Schaal TD and Maniatis T. Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. *Mol Cell Biol*. 1999;19:261-73.

39. Zahler AM, Damgaard CK, Kjems J and Caputi M. SC35 and heterogeneous nuclear ribonucleoprotein A/B proteins bind to a juxtaposed exonic splicing enhancer/exonic splicing silencer element to regulate HIV-1 tat exon 2 splicing. *J Biol Chem*. 2004;279:10077-84.

40. Ding JH, Xu X, Yang D, Chu PH, Dalton ND, Ye Z, Yeakley JM, Cheng H, Xiao RP, Ross J, Chen J and Fu XD. Dilated cardiomyopathy caused by tissue-specific ablation of SC35 in the heart. *EMBO J*. 2004;23:885-96.

41. Xiao R, Sun Y, Ding JH, Lin S, Rose DW, Rosenfeld MG, Fu XD and Li X. Splicing regulator SC35 is essential for genomic stability and cell proliferation during mammalian organogenesis. *Mol Cell Biol*. 2007;27:5393-402.

42. Wang HY, Xu X, Ding JH, Bermingham JR, Jr. and Fu XD. SC35 plays a role in T cell development and alternative splicing of CD45. *Mol Cell*. 2001;7:331-42.

43. Zhou T, Hasty P, Walter CA, Bishop AJ, Scott LM and Rebel VI. Myelodysplastic syndrome: an inability to appropriately respond to damaged DNA? *Exp Hematol*. 2013;41:665-74.

44. Chan YA, Hieter P and Stirling PC. Mechanisms of genome instability induced by RNA-processing defects. *Trends Genet*. 2014;30:245-53.

45. Li X and Manley JL. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell*. 2005;122:365-78.

46. Shirai CL, Ley JN, White BS, Kim S, Tibbitts J, Shao J, Ndonwi M, Wadugu B, Duncavage EJ, Okeyo-Owuor T, Liu T, Griffith M, McGrath S, Magrini V, Fulton RS, Fronick C, O'Laughlin M, Graubert TA and Walter MJ. Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. *Cancer Cell*. 2015;27:631-43.

47. Te Raa GD, Derks IA, Navrkalova V, Skowronska A, Moerland PD, van Laar J, Oldreive C, Monsuur H, Trbusek M, Malcikova J, Loden M, Geisler CH, Hullein J, Jethwa A, Zenz T, Pospisilova S, Stankovic T, van Oers MH, Kater AP and Eldering E. The impact of SF3B1 mutations in CLL on the DNA-damage response. *Leukemia*. 2015;29:1133-42.

48. Makishima H, Yoshizato T, Yoshida K, Sekeres MA, 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, Gomez-Segui I, Husseinzadeh HD, Thota S, Guinta KM, Dienes B, Nakamaki T, Miyawaki S, Sauntharajah Y, Chiba S, Miyano S, Shih LY, Haferlach T, Ogawa S and Maciejewski JP. Dynamics of clonal evolution in myelodysplastic syndromes. *Nat Genet*. 2017;49:204-212.

49. Martincorena I and Campbell PJ. Somatic mutation in cancer and normal cells. *Science*. 2015;349:1483-9.

50. Santos-Pereira JM and Aguilera A. R loops: new modulators of genome dynamics and function. *Nat Rev Genet*. 2015;16:583-97.

51. Huertas P and Aguilera A. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell*. 2003;12:711-21.

52. Tuduri S, Crabbe L, Conti C, Tourriere H, Holtgreve-Grez H, Jauch A, Pantesco V, De Vos J, Thomas A, Theillet C, Pommier Y, Tazi J, Coquelle A and Pasero P. Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat Cell Biol*. 2009;11:1315-24.

53. Aguilera A and Garcia-Muse T. R loops: from transcription byproducts to threats to genome stability. *Mol Cell*. 2012;46:115-24.

54. Sollier J and Cimprich KA. Breaking bad: R-loops and genome integrity. *Trends Cell Biol*. 2015;25:514-22.

55. Makishima H, Visconte V, Sakaguchi H, Jankowska AM, Abu Kar S, Jerez A, Przychodzen B, Bupathi M, Guinta K, Afable MG, Sekeres MA, Padgett RA, Tiu RV and Maciejewski JP. Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. *Blood*. 2012;119:3203-10.

56. Komeno Y, Huang YJ, Qiu J, Lin L, Xu Y, Zhou Y, Chen L, Monterroza DD, Li H, DeKelver RC, Yan M, Fu XD and Zhang DE. SRSF2 Is Essential for Hematopoiesis, and Its Myelodysplastic Syndrome-Related Mutations Dysregulate Alternative Pre-mRNA Splicing. *Mol Cell Biol*. 2015;35:3071-82.

57. Lopato S, Kalyna M, Dorner S, Kobayashi R, Krainer AR and Barta A. atSRp30, one of two SF2/ASF-like proteins from Arabidopsis thaliana, regulates splicing of specific plant genes. *Genes Dev*. 1999;13:987-1001.

58. Kraus ME and Lis JT. The concentration of B52, an essential splicing factor and regulator of splice site choice in vitro, is critical for Drosophila development. *Mol Cell Biol*. 1994;14:5360-70.

59. Labourier E, Bourbon HM, Gallouzi IE, Fostier M, Allemand E and Tazi J. Antagonism between RSF1 and SR proteins for both splice-site recognition in vitro and Drosophila development. *Genes Dev*. 1999;13:740-53.

60. Matsuoka A, Tochigi A, Kishimoto M, Nakahara T, Kondo T, Tsujioka T, Tasaka T, Tohyama Y and Tohyama K. Lenalidomide induces cell death in an MDS-derived cell line with deletion of chromosome 5q by inhibition of cytokinesis. *Leukemia*. 2010;24:748-55.

61. Drexler HG, Dirks WG and Macleod RA. Many are called MDS cell lines: one is chosen. *Leuk Res*. 2009;33:1011-6.

62. Zhou Z, Qiu J, Liu W, Zhou Y, Plocinik RM, Li H, Hu Q, Ghosh G, Adams JA, Rosenfeld MG and Fu XD. The Akt-SRPK-SR axis constitutes a major pathway in transducing EGF signaling to regulate alternative splicing in the nucleus. *Mol Cell*. 2012;47:422-33.

63. Chen L, Chen JY, Huang YJ, Gu Y, Qiu J, Qian H, Shao C, Zhang X, Hu J, Li H, He S, Zhou Y, Abdel-Wahab O, Zhang DE and Fu XD. The Augmented R-Loop Is a Unifying Mechanism for Myelodysplastic Syndromes Induced by High-Risk Splicing Factor Mutations. *Mol Cell*. 2018;69:412-425 e6.

64. Kim E, Ilagan JO, Liang Y, Daubner GM, Lee SC, Ramakrishnan A, Li Y, Chung YR, Micol JB, Murphy ME, Cho H, Kim MK, Zebari AS, Aumann S, Park CY, Buonamici S, Smith PG, Deeg HJ, Lobry C, Aifantis I, Modis Y, Allain FH, Halene S, Bradley RK and Abdel-Wahab O. SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. *Cancer Cell*. 2015;27:617-30.

65. Pollyea DA, Harris C, Rabe JL, Hedin BR, De Arras L, Katz S, Wheeler E, Bejar R, Walter MJ, Jordan CT, Pietras EM and Alper S. Myelodysplastic syndrome-associated spliceosome gene mutations enhance innate immune signaling. *Haematologica*. 2019;104:e388-e392.

66. Cerritelli SM and Crouch RJ. Cloning, expression, and mapping of ribonucleases H of human and mouse related to bacterial RNase HI. *Genomics*. 1998;53:300-7.

67. Nowotny M, Gaidamakov SA, Ghirlando R, Cerritelli SM, Crouch RJ and Yang W. Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription. *Mol Cell*. 2007;28:264-76.

68. de Boer J, Williams A, Skavdis G, Harker N, Coles M, Tolaini M, Norton T, Williams K, Roderick K, Potocnik AJ and Kioussis D. Transgenic mice with hematopoietic and lymphoid specific expression of Cre. *Eur J Immunol*. 2003;33:314-25.

69. Montecucco A and Biamonti G. Pre-mRNA processing factors meet the DNA damage response. *Front Genet*. 2013;4:102.

70. Li H, Qiu J and Fu XD. RASL-seq for massively parallel and quantitative analysis of gene expression. *Curr Protoc Mol Biol*. 2012;Chapter 4:Unit 4 13 1-9.

71. Daubner GM, Clery A, Jayne S, Stevenin J and Allain FH. A syn-anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well. *EMBO J*. 2012;31:162-74.

72. MacArthur MW and Thornton JM. Influence of proline residues on protein conformation. *J Mol Biol*. 1991;218:397-412.

73. Ozawa T, Okazaki K and Kitaura K. Importance of CH/pi hydrogen bonds in recognition of the core motif in proline-recognition domains: an ab initio fragment molecular orbital study. *J Comput Chem*. 2011;32:2774-82.

74. Colla S, Ong DS, Ogoti Y, Marchesini M, Mistry NA, Clise-Dwyer K, Ang SA, Storti P, Viale A, Giuliani N, Ruisaard K, Ganan Gomez I, Bristow CA, Estecio M, Weksberg DC, Ho YW, Hu B, Genovese G, Pettazzoni P, Multani AS, Jiang S, Hua S, Ryan MC, Carugo A, Nezi L, Wei Y, Yang H, D'Anca M, Zhang L, Gaddis S, Gong T, Horner JW, Heffernan TP, Jones P, Cooper LJ, Liang H, Kantarjian H, Wang YA, Chin L, Bueso-Ramos C, Garcia-Manero G and DePinho RA. Telomere dysfunction drives aberrant hematopoietic differentiation and myelodysplastic syndrome. *Cancer Cell*. 2015;27:644-57.

75. Chen PC, Dudley S, Hagen W, Dizon D, Paxton L, Reichow D, Yoon SR, Yang K, Arnheim N, Liskay RM and Lipkin SM. Contributions by MutL homologues Mlh3 and Pms2 to DNA mismatch repair and tumor suppression in the mouse. *Cancer Res*. 2005;65:8662-70.

76. Chini CC and Chen J. Claspin, a regulator of Chk1 in DNA replication stress pathway. *DNA Repair (Amst)*. 2004;3:1033-7.

77. Nguyen HD, Leong WY, Li W, Reddy PNG, Sullivan JD, Walter MJ, Zou L and Graubert TA. Spliceosome Mutations Induce R Loop-Associated Sensitivity to ATR Inhibition in Myelodysplastic Syndromes. *Cancer Res*. 2018;78:5363-5374.

78. Shalini Singh DA, Hamid Dolatshad, Dharamveer Tatwavedi, Ulrike Schulze, Andrea Sanchi, Sarah Ryley, Ashish Dhir, Lee Carpenter, Suzanne M. Watt, Amal M. Abdel-Aal, Sohair K. Sayed, Somia A. Mohamed, Anna Schuh, Nicholas J. Proudfoot, Andriana G. Kotini, Eirini P Papapetrou, Daniel Wiseman, Andrea Pellagatti, Jacqueline Boultwood. The SF3B1 K700E Mutation Induces R-Loop Accumulation and Associated DNA Damage. Paper presented at: American Society of Hematology; 2019; Orlando.

79. Awasthi P, Foiani M and Kumar A. ATM and ATR signaling at a glance. *J Cell Sci*. 2015;128:4255-62.

80. Cimprich KA and Cortez D. ATR: an essential regulator of genome integrity. *Nat Rev Mol Cell Biol*. 2008;9:616-27.

81. Komeno Y, Yan M, Matsuura S, Lam K, Lo MC, Huang YJ, Tenen DG, Downing JR and Zhang DE. Runx1 exon 6-related alternative splicing isoforms differentially regulate hematopoiesis in mice. *Blood*. 2014;123:3760-9.

82. Nguyen HD, Yadav T, Giri S, Saez B, Graubert TA and Zou L. Functions of Replication Protein A as a Sensor of R Loops and a Regulator of RNaseH1. *Mol Cell*. 2017;65:832-847 e4.

83. Walter MJ, Shen D, Ding L, Shao J, Koboldt DC, Chen K, Larson DE, McLellan MD, Dooling D, Abbott R, Fulton R, Magrini V, Schmidt H, Kalicki-Veizer J, O'Laughlin M, Fan X, Grillot M, Witowski S, Heath S, Frater JL, Eades W, Tomasson M, Westervelt P, DiPersio JF, Link DC, Mardis ER, Ley TJ, Wilson RK and Graubert TA. Clonal architecture of secondary acute myeloid leukemia. *N Engl J Med*. 2012;366:1090-8.

84. Pellagatti A, Roy S, Di Genua C, Burns A, McGraw K, Valletta S, Larrayoz MJ, Fernandez-Mercado M, Mason J, Killick S, Mecucci C, Calasanz MJ, List A, Schuh A and Boultwood J. Targeted resequencing analysis of 31 genes commonly mutated in myeloid disorders in serial samples from myelodysplastic syndrome patients showing disease progression. *Leukemia*. 2016;30:247-50.

85. Sperling AS, Gibson CJ and Ebert BL. The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. *Nat Rev Cancer*. 2017;17:5-19.

86. Kon A, Yamazaki S, Nannya Y, Kataoka K, Ota Y, Nakagawa MM, Yoshida K, Shiozawa Y, Morita M, Yoshizato T, Sanada M, Nakayama M, Koseki H, Nakauchi H and Ogawa S.

Physiological Srsf2 P95H expression causes impaired hematopoietic stem cell functions and aberrant RNA splicing in mice. *Blood*. 2018;131:621-635.

87. Smeets MF, Tan SY, Xu JJ, Anande G, Unnikrishnan A, Chalk AM, Taylor SR, Pimanda JE, Wall M, Purton LE and Walkley CR. Srsf2(P95H) initiates myeloid bias and myelodysplastic/myeloproliferative syndrome from hemopoietic stem cells. *Blood*. 2018;132:608-621.

88. Zhang J, Lieu YK, Ali AM, Penson A, Reggio KS, Rabadan R, Raza A, Mukherjee S and Manley JL. Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. *Proc Natl Acad Sci U S A*. 2015;112:E4726-34.

89. Wu SJ, Kuo YY, Hou HA, Li LY, Tseng MH, Huang CF, Lee FY, Liu MC, Liu CW, Lin CT, Chen CY, Chou WC, Yao M, Huang SY, Ko BS, Tang JL, Tsay W and Tien HF. The clinical implication of SRSF2 mutation in patients with myelodysplastic syndrome and its stability during disease evolution. *Blood*. 2012;120:3106-11.

90. Meggendorfer M, Roller A, Haferlach T, Eder C, Dicker F, Grossmann V, Kohlmann A, Alpermann T, Yoshida K, Ogawa S, Koeffler HP, Kern W, Haferlach C and Schnittger S. SRSF2 mutations in 275 cases with chronic myelomonocytic leukemia (CMML). *Blood*. 2012;120:3080-8.

91. Kar SA, Jankowska A, Makishima H, Visconte V, Jerez A, Sugimoto Y, Muramatsu H, Traina F, Afable M, Guinta K, Tiu RV, Przychodzen B, Sakaguchi H, Kojima S, Sekeres MA, List AF, McDevitt MA and Maciejewski JP. Spliceosomal gene mutations are frequent events in the diverse mutational spectrum of chronic myelomonocytic leukemia but largely absent in juvenile myelomonocytic leukemia. *Haematologica*. 2013;98:107-13.

92. Mian SA, Smith AE, Kulasekararaj AG, Kizilors A, Mohamedali AM, Lea NC, Mitsopoulos K, Ford K, Nasser E, Seidl T and Mufti GJ. Spliceosome mutations exhibit specific associations with epigenetic modifiers and proto-oncogenes mutated in myelodysplastic syndrome. *Haematologica*. 2013;98:1058-66.

93. Gaidzik VI, Teleanu V, Papaemmanuil E, Weber D, Paschka P, Hahn J, Wallrabenstein T, Kolbinger B, Kohne CH, Horst HA, Brossart P, Held G, Kundgen A, Ringhoffer M, Gotze K, Rummel M, Gerstung M, Campbell P, Kraus JM, Kestler HA, Thol F, Heuser M, Schlegelberger B, Ganser A, Bullinger L, Schlenk RF, Dohner K and Dohner H. RUNX1 mutations in acute myeloid leukemia are associated with distinct clinico-pathologic and genetic features. *Leukemia*. 2016;30:2282.

94. Speck NA, Stacy T, Wang Q, North T, Gu TL, Miller J, Binder M and Marin-Padilla M. Core-binding factor: a central player in hematopoiesis and leukemia. *Cancer Res*. 1999;59:1789s-1793s.

95. Speck NA and Gilliland DG. Core-binding factors in haematopoiesis and leukaemia. *Nat Rev Cancer*. 2002;2:502-13.

96. Ichikawa M, Asai T, Saito T, Seo S, Yamazaki I, Yamagata T, Mitani K, Chiba S, Ogawa S, Kurokawa M and Hirai H. AML-1 is required for megakaryocytic maturation and lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis. *Nat Med*. 2004;10:299-304.

97. Growney JD, Shigematsu H, Li Z, Lee BH, Adelsperger J, Rowan R, Curley DP, Kutok JL, Akashi K, Williams IR, Speck NA and Gilliland DG. Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. *Blood*. 2005;106:494-504.

98. Chen CY, Lin LI, Tang JL, Ko BS, Tsay W, Chou WC, Yao M, Wu SJ, Tseng MH and Tien HF. RUNX1 gene mutation in primary myelodysplastic syndrome--the mutation can be detected early at diagnosis or acquired during disease progression and is associated with poor outcome. *Br J Haematol*. 2007;139:405-14.

99. Tsai SC, Shih LY, Liang ST, Huang YJ, Kuo MC, Huang CF, Shih YS, Lin TH, Chiu MC and Liang DC. Biological Activities of RUNX1 Mutants Predict Secondary Acute Leukemia Transformation from Chronic Myelomonocytic Leukemia and Myelodysplastic Syndromes. *Clin Cancer Res*. 2015;21:3541-51.

100. Ito Y, Bae SC and Chuang LS. The RUNX family: developmental regulators in cancer. *Nat Rev Cancer*. 2015;15:81-95.

101. Sood R, Kamikubo Y and Liu P. Role of RUNX1 in hematological malignancies. *Blood*. 2017;129:2070-2082.

102. Barton JL, Bunka DH, Knowling SE, Lefevre P, Warren AJ, Bonifer C and Stockley PG. Characterization of RNA aptamers that disrupt the RUNX1-CBFbeta/DNA complex. *Nucleic Acids Res*. 2009;37:6818-30.

103. Cagnol S and Chambard JC. ERK and cell death: mechanisms of ERK-induced cell death--apoptosis, autophagy and senescence. *FEBS J*. 2010;277:2-21.

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

105. Yu G, Wang LG, Han Y and He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16:284-7.

106. Cantor S, Drapkin R, Zhang F, Lin Y, Han J, Pamidi S and Livingston DM. The BRCA1-associated protein BACH1 is a DNA helicase targeted by clinically relevant inactivating mutations. *Proc Natl Acad Sci U S A*. 2004;101:2357-62.

107. Cantor SB, Bell DW, Ganesan S, Kass EM, Drapkin R, Grossman S, Wahrer DC, Sgroi DC, Lane WS, Haber DA and Livingston DM. BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. *Cell*. 2001;105:149-60.

108. Levitus M, Waisfisz Q, Godthelp BC, de Vries Y, Hussain S, Wiegant WW, Elghalbzouri-Maghrani E, Steltenpool J, Rooimans MA, Pals G, Arwert F, Mathew CG, Zdzienicka MZ, Hiom K, De Winter JP and Joenje H. The DNA helicase BRIP1 is defective in Fanconi anemia complementation group J. *Nat Genet*. 2005;37:934-5.

109. Levran O, Attwooll C, Henry RT, Milton KL, Neveling K, Rio P, Batish SD, Kalb R, Velleuer E, Barral S, Ott J, Petrini J, Schindler D, Hanenberg H and Auerbach AD. The BRCA1-interacting helicase BRIP1 is deficient in Fanconi anemia. *Nat Genet*. 2005;37:931-3.

110. Seal S, Thompson D, Renwick A, Elliott A, Kelly P, Barfoot R, Chagtai T, Jayatilake H, Ahmed M, Spanova K, North B, McGuffog L, Evans DG, Eccles D, Breast Cancer Susceptibility C, Easton DF, Stratton MR and Rahman N. Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat Genet*. 2006;38:1239-41.

111. Guo M, Vidhyasagar V, Ding H and Wu Y. Insight into the roles of helicase motif Ia by characterizing Fanconi anemia group J protein (FANCJ) patient mutations. *J Biol Chem*. 2014;289:10551-65.

112. Wu Y, Sommers JA, Suhasini AN, Leonard T, Deakyne JS, Mazin AV, Shin-Ya K, Kitao H and Brosh RM, Jr. Fanconi anemia group J mutation abolishes its DNA repair function by uncoupling DNA translocation from helicase activity or disruption of protein-DNA complexes. *Blood*. 2010;116:3780-91.

113. Smogorzewska A, Desetty R, Saito TT, Schlabach M, Lach FP, Sowa ME, Clark AB, Kunkel TA, Harper JW, Colaiacovo MP and Elledge SJ. A genetic screen identifies FAN1, a Fanconi anemia-associated nuclease necessary for DNA interstrand crosslink repair. *Mol Cell*. 2010;39:36-47.

114. Trujillo JP, Mina LB, Pujol R, Bogliolo M, Andrieux J, Holder M, Schuster B, Schindler D and Surralles J. On the role of FAN1 in Fanconi anemia. *Blood*. 2012;120:86-9.

115. Yoshikiyo K, Kratz K, Hirota K, Nishihara K, Takata M, Kurumizaka H, Horimoto S, Takeda S and Jiricny J. KIAA1018/FAN1 nuclease protects cells against genomic instability induced by interstrand cross-linking agents. *Proc Natl Acad Sci U S A*. 2010;107:21553-7.

116. Huang J, Gong Z, Ghosal G and Chen J. SOSS complexes participate in the maintenance of genomic stability. *Mol Cell*. 2009;35:384-93.

117. Li Y, Bolderson E, Kumar R, Muniandy PA, Xue Y, Richard DJ, Seidman M, Pandita TK, Khanna KK and Wang W. HSSB1 and hSSB2 form similar multiprotein complexes that participate in DNA damage response. *J Biol Chem*. 2009;284:23525-31.

118. Edwards MC, Liegeois N, Horecka J, DePinho RA, Sprague GF, Jr., Tyers M and Elledge SJ. Human CPR (cell cycle progression restoration) genes impart a Far- phenotype on yeast cells. *Genetics*. 1997;147:1063-76.

119. Mukherjee D, Gao M, O'Connor JP, Raijmakers R, Pruijn G, Lutz CS and Wilusz J. The mammalian exosome mediates the efficient degradation of mRNAs that contain AU-rich elements. *EMBO J*. 2002;21:165-74.

120. van Dijk EL, Schilders G and Pruijn GJ. Human cell growth requires a functional cytoplasmic exosome, which is involved in various mRNA decay pathways. *RNA*. 2007;13:1027-35.

121. West S, Gromak N, Norbury CJ and Proudfoot NJ. Adenylation and exosome-mediated degradation of cotranscriptionally cleaved pre-messenger RNA in human cells. *Mol Cell*. 2006;21:437-43.

122. Garcia M, Pujol A, Ruzo A, Riu E, Ruberte J, Arbos A, Serafin A, Albella B, Feliu JE and Bosch F. Phosphofructo-1-kinase deficiency leads to a severe cardiac and hematological disorder in addition to skeletal muscle glycogenosis. *PLoS Genet*. 2009;5:e1000615.

123. Obeng EA, Stewart C and Abdel-Wahab O. Altered RNA Processing in Cancer Pathogenesis and Therapy. *Cancer Discov*. 2019;9:1493-1510.

124. van Wijk R and van Solinge WW. The energy-less red blood cell is lost: erythrocyte enzyme abnormalities of glycolysis. *Blood*. 2005;106:4034-42.

125. Wang X, Martindale JL and Holbrook NJ. Requirement for ERK activation in cisplatin-induced apoptosis. *J Biol Chem*. 2000;275:39435-43.

126. Van Nostrand EL, Pratt GA, Shishkin AA, Gelboin-Burkhart C, Fang MY, Sundararaman B, Blue SM, Nguyen TB, Surka C, Elkins K, Stanton R, Rigo F, Guttman M and Yeo GW. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods*. 2016;13:508-14.

127. Yoshimi A, Lin KT, Wiseman DH, Rahman MA, Pastore A, Wang B, Lee SC, Micol JB, Zhang XJ, de Botton S, Penard-Lacronique V, Stein EM, Cho H, Miles RE, Inoue D, Albrecht TR, Somervaille TCP, Batta K, Amaral F, Simeoni F, Wilks DP, Cargo C, Intlekofer AM, Levine RL, Dvinge H, Bradley RK, Wagner EJ, Krainer AR and Abdel-Wahab O. Coordinated alterations in RNA splicing and epigenetic regulation drive leukaemogenesis. *Nature*. 2019;574:273-277.

128. Yin S, Gambe RG, Sun J, Martinez AZ, Cartun ZJ, Regis FFD, Wan Y, Fan J, Brooks AN, Herman SEM, Ten Hacken E, Taylor-Weiner A, Rassenti LZ, Ghia EM, Kipps TJ, Obeng EA, Cibulskis CL, Neuberg D, Campagna DR, Fleming MD, Ebert BL, Wiestner A, Leshchiner I, DeCaprio JA, Getz G, Reed R, Carrasco RD, Wu CJ and Wang L. A Murine Model of Chronic Lymphocytic Leukemia Based on B Cell-Restricted Expression of Sf3b1 Mutation and Atm Deletion. *Cancer Cell*. 2019;35:283-296 e5.

129. Daichi Inoue SCL, Akihide Yoshimi, Justin Taylor, Lillian E Bitner, Atsushi Tanaka, Yasutaka Hayashi, Hana Cho, Chie Fukui, Ruth Saganty, Alexander V Penson, Omar Abdel-Wahab. Aberrant RNA Splicing Contributes to the Pathogenesis of EVI-Rearranged Myeloid Leukemias. Paper presented at: American Society of Hematology; 2019; Orlando.

130. Fei DL, Zhen T, Durham B, Ferrarone J, Zhang T, Garrett L, Yoshimi A, Abdel-Wahab O, Bradley RK, Liu P and Varmus H. Impaired hematopoiesis and leukemia development in mice with a conditional knock-in allele of a mutant splicing factor gene U2af1. *Proc Natl Acad Sci U S A*. 2018;115:E10437-E10446.

131. Kornblihtt AR, de la Mata M, Fededa JP, Munoz MJ and Nogues G. Multiple links between transcription and splicing. *RNA*. 2004;10:1489-98.

132. Pandya-Jones A and Black DL. Co-transcriptional splicing of constitutive and alternative exons. *RNA*. 2009;15:1896-908.

133. Yang YT, Chiu YC, Kao CJ, Hou HA, Lin CC, Tsai CH, Tseng MH, Chou WC and Tien HF. The prognostic significance of global aberrant alternative splicing in patients with myelodysplastic syndrome. *Blood Cancer J*. 2018;8:78.

134. Yan M, Kanbe E, Peterson LF, Boyapati A, Miao Y, Wang Y, Chen IM, Chen Z, Rowley JD, Willman CL and Zhang DE. A previously unidentified alternatively spliced isoform of t(8;21) transcript promotes leukemogenesis. *Nat Med*. 2006;12:945-9.