



Myelodysplastic Syndrome-Associated SRSF2 Mutations Cause Splicing Changes by Altering Binding Motif Sequences

So Masaki^{1,2*}, Shun Ikeda¹, Asuka Hata¹, Yusuke Shiozawa³, Ayana Kon³, Seishi Ogawa³, Kenji Suzuki², Fumihiko Hakuno⁴, Shin-Ichiro Takahashi⁴ and Naoyuki Kataoka^{1,4*}

¹ Laboratory for Malignancy Control Research, Medical Innovation Center, Kyoto University Graduate School of Medicine, Kyoto, Japan, ² Laboratory of Molecular Medicinal Science, Department of Pharmaceutical Sciences, Ritsumeikan University, Shiga, Japan, ³ Department of Pathology and Tumor Biology, Kyoto University Graduate School of Medicine, Kyoto, Japan, ⁴ Laboratory of Cell Regulation, Departments of Applied Animal Sciences and Applied Biological Chemistry, Graduate School of Agriculture and Life Sciences, The University of Tokyo, Tokyo, Japan

OPEN ACCESS

Edited by:

Philipp Kapranov,
Huaqiao University, China

Reviewed by:

Zhixiang Lu,
Harvard Medical School,
United States
Woan-Yuh Tarn,
Academia Sinica, Taiwan

*Correspondence:

So Masaki
sasaki@fc.ritsumei.ac.jp
Naoyuki Kataoka
akataoka@mail.ecc.u-tokyo.ac.jp

Specialty section:

This article was submitted to
RNA,
a section of the journal
Frontiers in Genetics

Received: 15 November 2018

Accepted: 29 March 2019

Published: 16 April 2019

Citation:

Masaki S, Ikeda S, Hata A,
Shiozawa Y, Kon A, Ogawa S,
Suzuki K, Hakuno F, Takahashi S-I and
Kataoka N (2019) Myelodysplastic
Syndrome-Associated SRSF2
Mutations Cause Splicing Changes by
Altering Binding Motif Sequences.
Front. Genet. 10:338.
doi: 10.3389/fgene.2019.00338

Serine/arginine-rich splicing factor 2 (SRSF2) is a member of the SR protein family that is involved in both constitutive and alternative mRNA splicing. Mutations in SRSF2 gene are frequently reported in myelodysplastic syndromes (MDS) and acute myeloid leukemia (AML). It is imperative to understand how these mutations affect SRSF2-mediated splicing and cause MDS. In this study, we characterized MDS-associated SRSF2 mutants (P95H, P95L, and P95R). We found that those mutants and wild-type SRSF2 proteins showed nuclear localization in HeLa cells. *In vitro* splicing reaction also revealed that mutant proteins associated with both precursor and spliced mRNAs, suggesting that the mutants directly participate in splicing. We established the human myeloid leukemia K562 cell lines that stably expressed myc-tagged wild-type or mutant SRSF2 proteins, and then performed RNA-sequence to analyze the splicing pattern of each cell line. The results revealed that both wild-type and mutants affected splicing of approximately 3,000 genes. Although splice site sequences adjacent to the affected exons showed no significant difference compared to the total exons, exonic motif analyses with both inclusion- and exclusion-enhanced exons demonstrated that wild-type and mutants have different binding sequences in exons. These results indicate that mutations of SRSF2 in MDS change binding properties of SRSF2 to exonic motifs and this causes aberrant splicing.

Keywords: myelodysplastic syndrome, splicing, SRSF2, exonic splicing enhancer, aberrant splicing, EZH2 (enhancer of zeste homolog 2)

INTRODUCTION

Serine/Arginine rich (SR) proteins are essential splicing factors that also confer regulatory activity of alternative splicing (Manley and Krainer, 2010; Howard and Sanford, 2015; Kataoka, 2017). SR protein family consists of 11 proteins in human. SR proteins contain one or two RNA binding domain (RBD) at amino-terminus and multiple repeats of

Serine-Arginine dipeptides at carboxy-terminus. SRSF2, which was originally called SC35, is a member of the SR protein family (Fu and Maniatis, 1990; Fu and Ares, 2014; Kataoka, 2017). During splicing, SRSF2 promotes exon recognition by binding to exonic splicing enhancer (ESE) motifs in precursor of mRNA (pre-mRNA) through its RBD. This promotes both the binding of U2AF heterodimer and U1 snRNP to the upstream 3' splice site and to the downstream 5' splice site, respectively (Chen and Manley, 2009; Fu and Ares, 2014; Kataoka, 2017).

Recently SRSF2 was found to be one of the major responsible genes of myelodysplastic syndrome (MDS). MDS is a heterogeneous group of chronic myeloid neoplasms characterized by many symptoms, such as ineffective hematopoiesis, peripheral blood cytopenia and a high risk of progression to acute myeloid leukemia (Fu and Maniatis, 1990; Cazzola et al., 2013). Mutations in splicing factors represent a novel class of driver mutations in human cancers and affect about 50% of patients with myelodysplasia (Graubert et al., 2011; Papaemmanuil et al., 2011, 2013; Yoshida et al., 2011; Walter et al., 2013; Haferlach et al., 2014). Somatic mutations are frequently found in genes SF3B1, SRSF2, U2AF1 and ZRSR2 (Graubert et al., 2011; Papaemmanuil et al., 2011; Yoshida et al., 2011). Interestingly, the common feature of those gene products in pre-mRNA splicing is a 3' splice site recognition, implicating that MDS pathogenesis is caused most likely by abnormal pre-mRNA splicing. The mutations are heterozygous and clustered in specific amino acid residues for SF3B1, SRSF2 and U2AF1 (Yoshida et al., 2011; Darman et al., 2015; Ilagan et al., 2015; Kim et al., 2015). These 'hotspot' mutations are predicted to be a 'gain-of-function' mutations by affecting protein structure (Yoshida et al., 2011; Darman et al., 2015; Ilagan et al., 2015; Kim et al., 2015). Among them, the hotspot mutations in SRSF2 were found in proline residue residing slightly outside of its RNA binding domain. Changes in alternative splicing patterns with SRSF2 mutations were reported in culture cells, mouse models and primary human samples (Przychodzen et al., 2013; Brooks et al., 2014; Shao et al., 2014; Ilagan et al., 2015; Kim et al., 2015; Komeno et al., 2015; Shirai et al., 2015; Zhang et al., 2015; Obeng et al., 2016; Mupo et al., 2017). In spite of these important findings, the precise mechanism for aberrant splicing in the cells carrying MDS mutations still remains largely unclear.

To address the mechanism of aberrant splicing in MDS, we have prepared expression plasmids for both wild-type and MDS-causing mutants of SRSF2, P95H, P95L and P95R. Transcriptome analyses using total RNAs recovered from those cell lines revealed that both wild-type and mutants affected splicing of approximately 3,000 genes. Motif analyses with both inclusion- and exclusion-enhanced exons demonstrated that the mutants have different binding sequences in exons compared to wild-type protein. These results strongly suggest that MDS mutations in SRSF2 alter binding properties of SRSF2 to exonic motifs and this results in aberrant splicing.

RESULTS

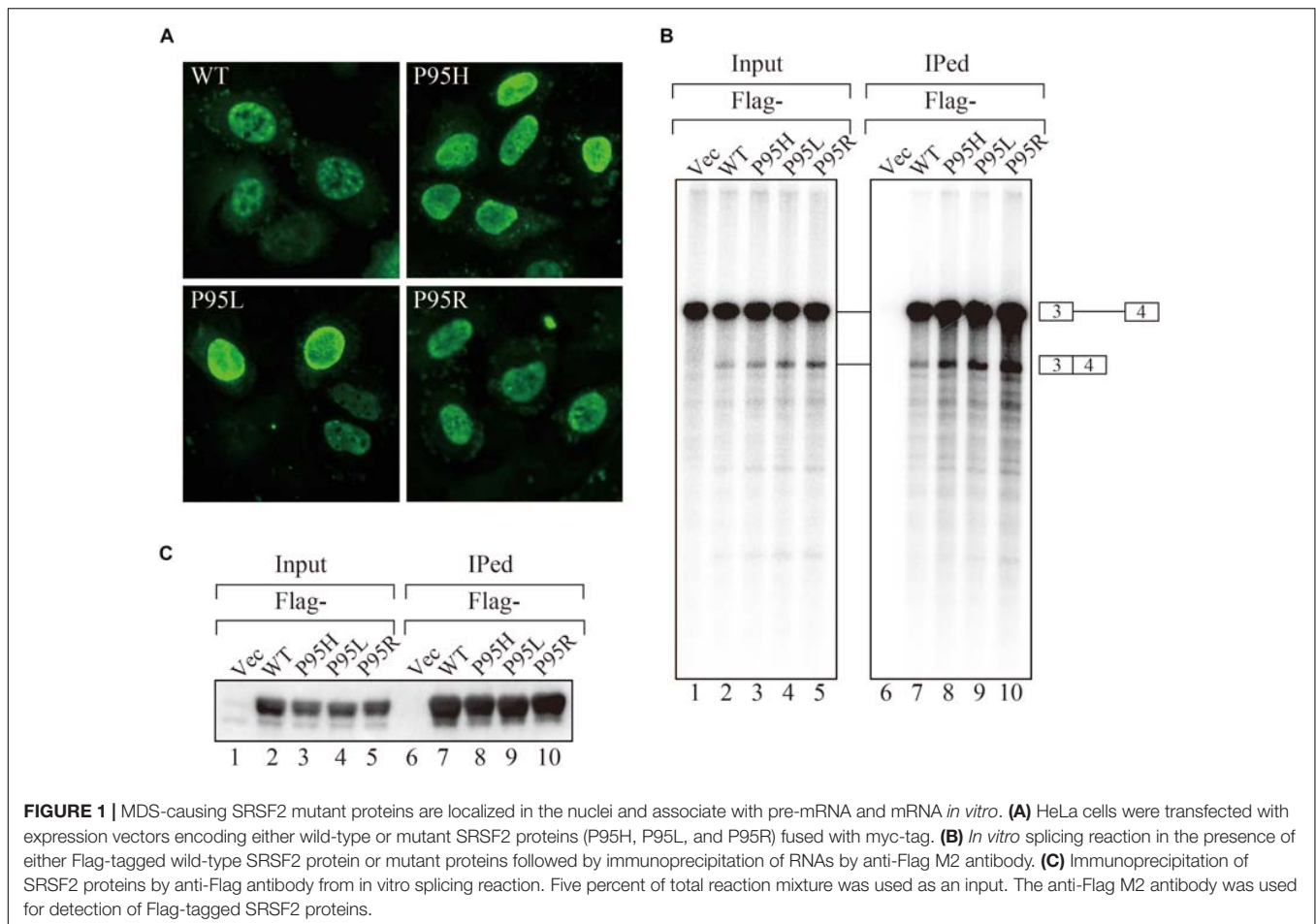
Localization of Wild-Type and Mutant SRSF2 Proteins in HeLa Cells and Association With Pre-mRNA and mRNA *in vitro*

In order to gain insights on how mutations in SRSF2, one of the splicing factors mutated in MDS patients, affect splicing, we have prepared the SRSF2 mutant cDNAs carrying three kinds of mutations at 95th position of Proline residue (P95H, L and R) found in MDS patients. We first determined their subcellular localization in HeLa cells. Both wild-type and mutants of SRSF2 cDNAs were transfected into HeLa cells and those proteins were expressed as fusions with a myc-tag. When the myc-SRSF2 wild-type protein was expressed, it was localized in both speckles and nucleoplasm, likely due to overexpression (Figure 1A, panel a). The subcellular localization of mutant SRSF2 proteins was similar to that of the wild-type, although mutant proteins exhibited slightly less numbers of nuclear speckles (Figure 1A).

We have also carried out *in vitro* splicing reaction with both Flag-tagged wild-type and mutant SRSF2 proteins, followed by immunoprecipitation of RNA from the mixture in order to test whether mutant proteins were able to support splicing or not. For this assay, we used the immunoglobulin μ chain (IgM) pre-mRNA, whose splicing is SRSF2-dependent (Mayeda et al., 1999). HeLa cell extract was mixed with cell lysates from HEK293T cells that express either wild-type or each mutant SRSF2 protein. The Flag-vector transfected cell lysate was used as a negative control. The results of splicing reaction are shown in Figure 1B. With a negative control, splicing did not take place (lane 1), indicating the dependency of IgM pre-mRNA to SRSF2. As shown in Figure 1B (lanes 2–5), both wild-type and each mutant supported IgM splicing *in vitro*, indicating mutant SRSF2 proteins have a splicing supporting activity even stronger than wild-type protein. Immunoprecipitation was carried out with the reaction mixtures using anti-Flag tag antibody. Flag-SRSF2 wild-type precipitated both pre-mRNA and mRNA (Figure 1B, lane 7). Mutant SRSF2 proteins also precipitated pre-mRNA and mRNA more efficiently than wild-type protein (Figure 1B, lane 8–10). Immunoprecipitated SRSF2 proteins were also detected by western blotting with anti-Flag M2 antibody (Figure 1C), and it turned out the comparable amounts of both wild-type and mutant SRSF2 proteins were precipitated (Figure 1C, lanes 7–10). Take all the results together, we concluded that mutant SRSF2 proteins are able to cause change of splicing pattern in cells, and the mutants likely have higher affinity to SRSF2-mediated ESE sequences.

Detection of Splicing Pattern Changes in K562 Cells Expressing Mutant SRSF2 Proteins

We generated to generate cell lines that stably express either wild-type or SRSF2 mutants with K562 cells, a myelogenous leukemia cell line. After establishment, we checked the protein expression level by western blotting by using anti-myc tag antibody. As



shown in **Figure 2A**, all cells expressed myc-tagged SRSF2 proteins except myc-vector transfected cells. During selection of the stable cell lines, cells expressing relatively large amount of mutant SRSF2 proteins tend to die, likely due to toxicity of mutant proteins and cells expressing high amount of mutant proteins may have been eliminated during selection of the clones. Total RNAs derived from those cell lines were applied to RNA sequence analyses. The reads were mapped to human genome with more than 90% efficiency (**Figure 2B**). We determined splicing changes by comparing wild-type and mutants splicing patterns with vector transfected one. It turned out that more than 5000 splicing events in more than 3000 genes were detected with both wild-type and mutants (**Figure 2C**). Drawing Venn diagram demonstrated that overlapping genes for all of them are 1406 genes, and there are also many genes either specific to each protein (**Figure 2D**).

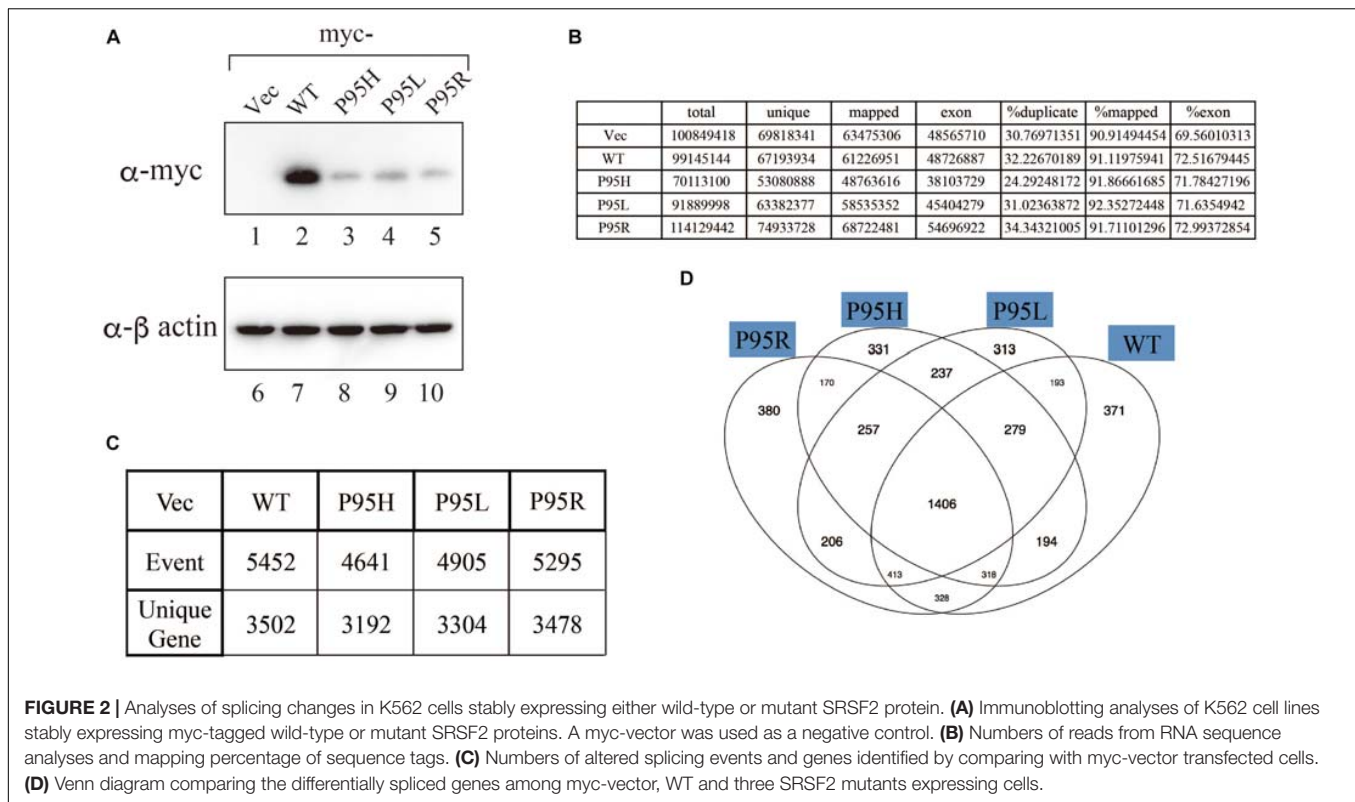
ESE-Like Motifs Were Enriched in Exons Skipped by SRSF2 Mutant Proteins

Since we identified many genes whose splicing patterns were affected by the expression of both wild-type and mutant SRSF2 proteins, we investigated the motifs of exons enriched in both wild-type and mutants regulated genes. Specific sequence features

in the exon were searched by multiple expectation-maximization for motif elicitation (MEME) (Bailey et al., 2006) algorithm to find the enrichment or depletion of novel sequence motifs. Surprisingly, enriched motifs for all mutants are purine-rich sequences (**Figure 3A**). The motif for wild-type SRSF2-excluded exons is AGGTRAG (R indicates purine residue), in which the purine stretch is separated by T residue (**Figure 3A**). It has been shown that SRSF2 proteins are able to bind purine-rich ESEs (Cavaloc et al., 1999). These results strongly suggested that MDS-causing mutations in proline residue of SRSF2 cause reduction of the affinity to purine-rich ESEs.

Mutant SRSF2 Proteins Prefer CCWG Motif for Exon Inclusion

We also carried out motif analysis for included exons, and we found that a purine-rich motif appeared as the most frequently appeared motif with a wild-type protein (**Figure 3B**), consistent with the previous findings that SRSF2 binds to purine-rich ESE to promote exon inclusion (Cavaloc et al., 1999). With SRSF2 P95L, a similar A/G rich motif was also found (**Figure 3B**), suggesting this mutation has slight effect on recognizing sequence of SRSF2 protein. In contrast, with two other mutant proteins, SRSF2 P95H and P95R, CCWG (W: weak as T or A) containing



motif was identified in inclusion-promoted exon at the top (**Figure 3B**). These results strongly suggested that purine-rich motif and CCWG-containing motif function as ESE for wild-type and mutants, respectively, and MDS-causing mutations alter the high affinity of SRSF2 from purine-rich motif to CCWG motif.

We have confirmed the splicing change with wild-type and mutant proteins by choosing several genes to determine splicing changes by RT-PCR. Among them, the splicing change of EZH2 gene is shown in **Figure 3D**. EZH2 is also known as one of the responsible genes for causing MDS (Ernst et al., 2010). With RNA sequencing analysis, the number of reads for exon 9.5, which has two CCWG motifs and premature termination codons (**Figure 3C**), was reduced with SRSF2 wild-type expression. In contrast, mutant protein expression increased the numbers of exon 9.5 reads, suggesting that mutant proteins promote exon 9.5 inclusion whereas wild-type enhances skipping of this exon. To test this possibility, we carried out RT-PCR analysis by amplifying Exon9–Exon10 region of EZH2 mRNA. The results indicate that wild-type protein expression reduced the ratio of exon 9.5 included mRNA (**Figure 3D**, lane 2). In contrast, all of the mutant proteins increased exon 9.5-included mRNA ratio (**Figure 3D**, lanes 3–5). These results indicate that SRSF2 wild-type and mutants have an opposite effect on CCWG-containing exon splicing.

DISCUSSION

In this manuscript we have analyzed the splicing changes in K562 cells stably expressing either wild-type or mutant SRSF2

proteins. As expected, both wild-type and mutants affected many splicing events of various genes, and CCWG motif was found in inclusion promoted exons with mutant proteins. CCNG or GGNG motif was previously reported as the binding sequences of SRSF2 protein by SELEX (Cavaloc et al., 1999). Indeed, three mutant proteins promote CCWG-motif containing pseudo exon of EZH2, which is one of the responsible genes for MDS (**Figure 3D**) (Ernst et al., 2010). This mis-splicing may affect EZH2 protein level in cells and the reduction of EZH2 protein level would result in mis-regulation of epigenetics (Kim et al., 2015; Shirahata-Adachi et al., 2017; Kon et al., 2018; Shiozawa et al., 2018). All three SRSF2 mutants were able to support SRSF2-dependent substrate (IgM) splicing *in vitro* more efficiently and associate with IgM pre-mRNA and mRNA more strongly than wild-type protein (**Figure 1B**). There are several CCWG motifs in 3' exon of IgM pre-mRNA (Mayeda et al., 1999). It is highly likely mutant proteins bind to these motifs more strongly than wild-type (**Figure 4**). On the other hand, the AGGTRAG motif was identified in exons excluded by wild-type SRSF2 (**Figure 3A**). In this motif, purine-stretch is separated by T (U in RNA). It was demonstrated that U residue splitting purine-stretch in ESE abolishes splicing-promoting activity *in vitro* (Watakabe et al., 1989). This motif may bind to SRSF2 with low affinity. Alternatively, it was demonstrated that SR protein overexpression can cause exon skipping which depends on their prevalent actions on a flanking constitutive exon, and it requires collaboration of more than one SR protein (Han et al., 2011). The AGGTRAG sequence might be a binding motif of 'weak' SR protein, not a direct binding motif for SRSF2 protein, to support exon inclusion. When overexpressed, wild-type SRSF2 protein can

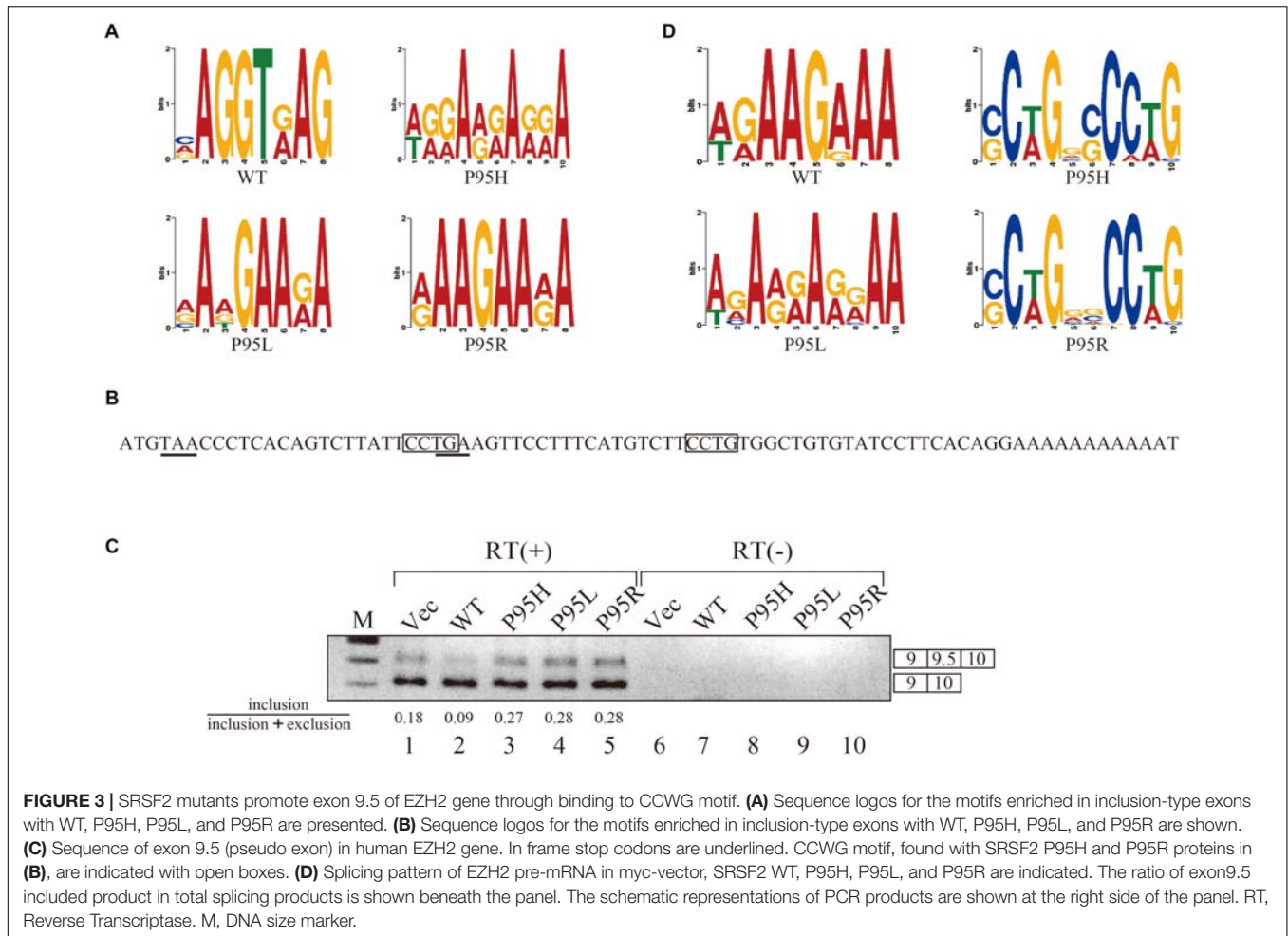


FIGURE 3 | SRSF2 mutants promote exon 9.5 of EZH2 gene through binding to CCWG motif. **(A)** Sequence logos for the motifs enriched in inclusion-type exons with WT, P95H, P95L, and P95R are presented. **(B)** Sequence logos for the motifs enriched in inclusion-type exons with WT, P95H, P95L, and P95R are shown. **(C)** Sequence of exon 9.5 (pseudo exon) in human EZH2 gene. In frame stop codons are underlined. CCWG motif, found with SRSF2 P95H and P95R proteins in **(B)**, are indicated with open boxes. **(D)** Splicing pattern of EZH2 pre-mRNA in myc-vector, SRSF2 WT, P95H, P95L, and P95R are indicated. The ratio of exon9.5 included product in total splicing products is shown beneath the panel. The schematic representations of PCR products are shown at the right side of the panel. RT, Reverse Transcriptase. M, DNA size marker.

bind to the flanking exon as a 'strong' SR protein more efficiently. The purine-rich ESE like sequence motifs were identified in exclusion promoted exons with mutant proteins (**Figure 3A**). Mutant proteins may have lower affinity to purine-rich sequences than wild-type (**Figure 4**). It is assumed that MDS mutations in Proline residue affect the conformation of RNA binding domain of SRSF2, although this proline residue is outside of RNA binding domain consensus. Indeed, 3D clustering analysis with protein structure predicted that this Proline residue is able to contact with RNA (Kamburov et al., 2015). Therefore, this residue is a part of RNA binding domain. By changing Proline residue to Histidine, Arginine or Leucine, the affinity with specific RNA sequence can be altered. Comparing crystal structures of these mutants-RNA complexes with that of wild-type would reveal the mechanism for the recognition of different RNA sequences. We believe this also uncover why mutant SRSF2 proteins have lower affinity to purine-rich ESEs.

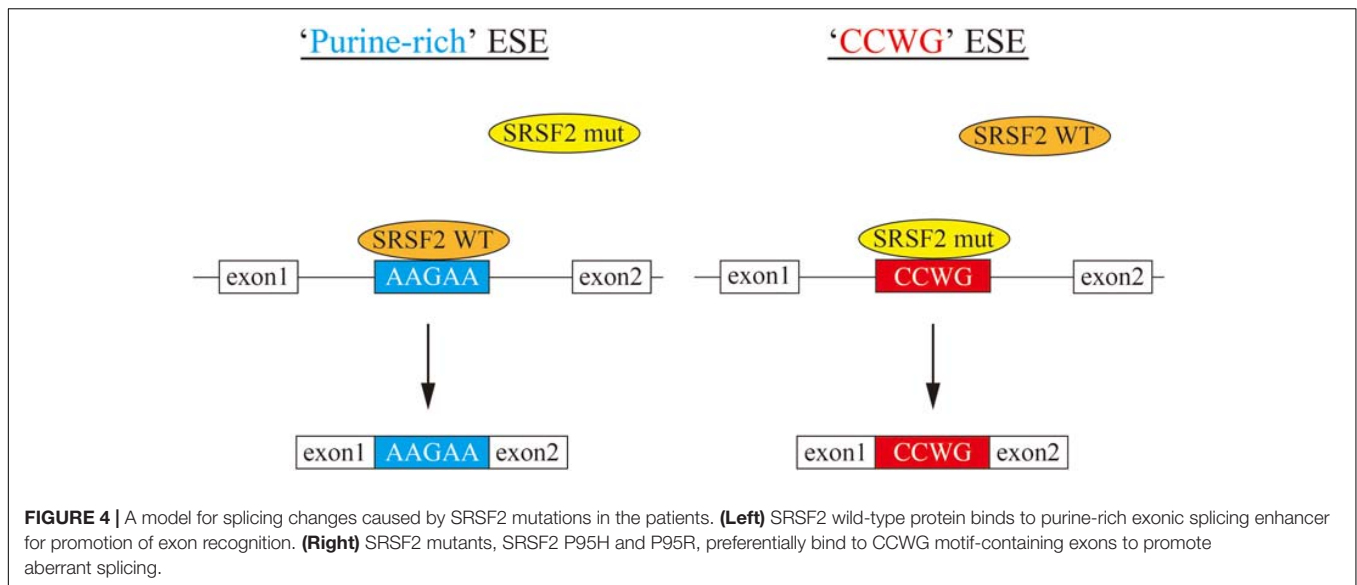
Despite of the common features described above, each mutant contains peculiar subset of genes whose splicing patterns were specifically changed (**Figure 2D**). These splicing changes in certain genes may confer the pathological difference among the MDS patients. Further analyses of specific splicing changes in each mutant are also required.

Most recently several groups demonstrated that MDS responsible mutations in SRSF2 and U2AF1 cause augment of R loop (Chen et al., 2018; Nguyen et al., 2018). Enhanced R loop formation activate the ataxia telangiectasia and Rad3-related protein (ATR)-Chk1 pathway, which likely contributes to MDS phenotype (Nguyen et al., 2018). Efficient formation of R loop can occur by slowing down rearrangement of mRNA-protein complexes during/after splicing. The different binding affinity of SRSF2 mutants to RNA may be involved in this step. It would be of a great interest in which portions of genes form R loops and whether enrichment of CCWG motif can be observed in those regions or not.

MATERIALS AND METHODS

Plasmid Construction

The cDNA of human SRSF2/SC35 was amplified by Reverse Transcription and Polymerase Chain Reaction (RT-PCR). The cDNA was cloned between BamHI and XhoI sites of either myc- or Flag-pCDNA3 vector. The resultant plasmid was used as a template in order to prepare mutant cDNAs that harbor MDS mutations, such as P95H, P95L and P95R. Point Mutations were



introduced by QuikChange™ Site-Directed Mutagenesis Kit (STRATAGENE) with myc-SRSF2 plasmid in accordance with the manufacturer's recommendations.

Cell Culture and Establishment of Stable Cell Lines

K562 cells were cultured at 37°C with 5% CO₂ in RPMI1640 supplemented with 10% (v/v) fetal bovine serum (Sigma-Aldrich) and 1% (v/v) penicillin/streptomycin antibiotics (standard medium). K562 cells were transfected with pcDNA3-myc-SRSF2(WT), pcDNA3-myc-SRSF2(P95H), pcDNA3-myc-SRSF2(P95L), pcDNA3-myc-SRSF2(P95R) or empty vector plasmid, using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instruction. Transfected K562 cells were selected with 100 mg/mL G418, and then resistant cells were isolated with limiting dilution methods in 96-well plates. Obtained each stable clone-cells were maintained in 50 mg/mL G418 RPMI medium.

For HEK293T cells, DMEM supplemented with 10% (v/v) fetal bovine serum (Sigma-Aldrich) and 1% (v/v) Antibiotics Antimycotics (SIGMA) was used for cell culture and Lipofectamine 2000 was also used for transfection. Total cell lysates from transfected HEK293T cells were prepared as previously described (Kataoka and Dreyfuss, 2004, 2008; Kataoka, 2016).

In vitro Transcription, Splicing and Immunoprecipitation of RNAs

For *in vitro* transcription, p_μC3-C4 plasmid linearized by HindIII was used as a template and performed as described previously (Watakabe et al., 1989, 1993). *In vitro* splicing reaction with HeLa cell nuclear extracts complemented with Flag-SRSF2 protein expressing HEK293T total cell lysates was carried out as previously described (Kataoka et al., 2001, 2011; Kim et al., 2001; Kataoka and Dreyfuss, 2004; Kawano et al., 2004; Kataoka, 2016).

Immunoprecipitation of RNAs from splicing reaction was accomplished by using anti-Flag M2 agarose beads (SIGMA) by the protocol described previously (Kataoka et al., 2001, 2011; Kataoka and Dreyfuss, 2004; Kataoka, 2016).

Antibodies, Western Blotting and Immunostaining of HeLa Cells

The antibodies used for immunoblotting and immunostaining are as follows: anti-myc (MC045, Nacalai Tesque, Japan), anti-Flag M2 (Sigma-Aldrich), fluorescein isothiocyanate-conjugated goat anti-mouse F(ab')₂ (Cappel Laboratories, Durham, NC, United States), peroxidase-conjugated goat anti-mouse IgG antibodies (Jackson Immuno Research Laboratories, West Grove, PA, United States). For western blotting, the cells were lysed in CelLytic M Cell Lysis Reagent (Sigma-Aldrich) containing a protease inhibitor cocktail (Roche). The lysates were boiled with SDS-sample buffer at 95°C for 3 min. The samples were subjected to SDS-PAGE, transferred to PVDF membranes by iBlot system (Invitrogen), and incubated with primary antibodies. The membranes were washed and incubated with horseradish peroxidase-conjugated secondary antibody. Finally, chemiluminescence was detected using Chemi-Lumi One Super kit (Nacalai Tesque), and luminescence images were analyzed by LAS 4000 (GE Lifesciences). Immunostaining of HeLa cells with anti-myc antibody was performed as described previously (Kataoka et al., 2011).

RNA Recovery, RNA Sequence and Alternative Splicing Analysis

Total RNAs from K562 stable cells were performed by using RNeasy Mini Kit (QIAGEN). The synthesis and amplification of complementary DNA were performed using SMARTer Ultra Low Input RNA Kit for Sequencing, version 3 (Clontech). Each sample applied illumina GA. The reads were trimmed to 99 bases and were mapped on hg19 genomes and gencode v7 protein

coding transcripts by tophat (v2.0.9). To compute junctions in each sample, we processed the gencode GTF file by eval package (v2.2.8) and applied the mapped reads to juncBASE (v0.6) packages by following the options: ‘-min overhang = 6 -l 99 -c 3.’ For comparing each sample, we calculated “Percent Splicing index” (PSI) and corrected *p*-value and Benjamini & Hochberg multiple test using pairwise fisher’s test getASEvents w reference.py by following options: ‘-jcn seq len = 186 -method = BH’ comparing with vector and wild-type.

Motif Analysis

To search RNA binding motif of mutated SRSF2, we collected 100 base exon side sequences of exclusion junctions separating upper or lower than vector’s PSI in each sample and the upper and lower sequences was applied MEME v4.10.0 as following that options: ‘-minw 4-maxw 10 -maxsize 1000000.’

Reverse Transcription Polymerase Chain Reaction (RT-PCR)

RT-PCR reaction was accomplished as described previously (Wang et al., 2017). Briefly, 1 µg of total RNA was used for reverse transcription with prime Script Reverse Transcriptase (TAKARA, Japan). The produced cDNA was used for PCR reaction by using the following primers and PCR cycles. Cycle conditions were as follows: 94°C for 2 min; followed by 33 cycles of 94°C denaturation for 10 s, 58°C annealing for 15 s, and 72°C elongation for 30 s; with a final incubation at 72°C for 2 min in a PCR Thermal Cycler (BIOMETRA). PCR products were separated by electrophoresis and stained with ethidium bromide. The primers for PCR are as follows; EZH2 exon9 F, 5’-AAGCGGAAGAACACAGAAAC-3’, EZH2 exon10 R, 5’-CAGAGGAGCTCGAAGTTTCA-3’. For quantitation analysis

of alternative splicing products, the signals were measured by ImageJ software [U.S. National Institutes of Health, Bethesda (Schneider et al., 2012)].

AUTHOR CONTRIBUTIONS

SM and NK started this project, conceived and designed the experiments. SM, AH, YS, AK, and NK performed the experiments including RNA sequencing. SI and NK analyzed RNA sequencing results. SM, SI, AH, YS, AK, SO, FH, S-IT, and NK analyzed the data. SM, SI, AH, YS, AK, KS, FH, S-IT, and NK contributed reagents, materials, and analysis tools. SM and NK wrote the manuscript. NK took the primary responsibility for the final content. SM, SI, AH, YS, AK, SO, KS, FH, S-IT, and NK read and approved the final manuscript.

FUNDING

This work was supported by Grants-in-Aid for Scientific Research (23112706 to SM, 18K06012 to NK). This work was also supported in part by research funding from Dainippon Sumitomo Pharma Co., Ltd.

ACKNOWLEDGMENTS

We would like to thank the members of Takahashi lab and Dr. Aiko Sugiyama for helpful discussion and support for experiments. We are grateful to Dr. Gideon Dreyfuss (University of Pennsylvania) for kind gifts of myc- and Flag-pCDNA3 vectors.

REFERENCES

- Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 34, W369–W373. doi: 10.1093/nar/gkl198
- Brooks, A. N., Choi, P. S., de Waal, L., Sharifnia, T., Imielinski, M., Saksena, G., et al. (2014). A pan-cancer analysis of transcriptome changes associated with somatic mutations in U2AF1 reveals commonly altered splicing events. *PLoS One* 9:e87361. doi: 10.1371/journal.pone.0087361
- Cavaloc, Y., Bourgeois, C. F., Kister, L., and Stevenin, J. (1999). The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA* 5, 468–483. doi: 10.1017/s1355838299981967
- Cazzola, M., Della Porta, M. G., and Malcovati, L. (2013). The genetic basis of myelodysplasia and its clinical relevance. *Blood* 122, 4021–4034. doi: 10.1182/blood-2013-09-381665
- Chen, L., Chen, J. Y., Huang, Y. J., Gu, Y., Qiu, J., Qian, H., et al. (2018). The augmented R-loop is a unifying mechanism for myelodysplastic syndromes induced by high-risk splicing factor mutations. *Mol. Cell.* 69:e6. doi: 10.1016/j.molcel.2017.12.029
- Chen, M., and Manley, J. L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.* 10, 741–754. doi: 10.1038/nrm2777
- Darman, R. B., Seiler, M., Agrawal, A. A., Lim, K. H., Peng, S., Aird, D., et al. (2015). Cancer-associated SF3B1 hotspot mutations induce cryptic 3’ splice site selection through use of a different branch point. *Cell Rep.* 13, 1033–1045. doi: 10.1016/j.celrep.2015.09.053
- Ernst, T., Chase, A. J., Score, J., Hidalgo-Curtis, C. E., Bryant, C., Jones, A. V., et al. (2010). Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat. Genet.* 42, 722–726. doi: 10.1038/ng.621
- Fu, X. D., and Ares, M. Jr. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. *Nat. Rev. Genet.* 15, 689–701. doi: 10.1038/nrg3778
- Fu, X. D., and Maniatis, T. (1990). Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature* 343, 437–441. doi: 10.1038/343437a0
- Graubert, T. A., Shen, D., Ding, L., Okeyo-Owuor, T., Lunn, C. L., Shao, J., et al. (2011). Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nat. Genet.* 44, 53–57. doi: 10.1038/ng.1031
- Haferlach, T., Nagata, Y., Grossmann, V., Okuno, Y., Bacher, U., and Nagae, G. (2014). Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* 28, 241–247. doi: 10.1038/leu.2013.336
- Han, J., Ding, J. H., Byeon, C. W., Kim, J. H., Hertel, K. J., Jeong, S., et al. (2011). SR proteins induce alternative exon skipping through their activities on the flanking constitutive exons. *Mol. Cell. Biol.* 4, 793–802. doi: 10.1128/MCB.01117-10
- Howard, J. M., and Sanford, J. R. (2015). The RNAissance family: SR proteins as multifaceted regulators of gene expression. *Wiley Interdiscip. Rev. RNA* 6, 93–110. doi: 10.1002/wrna.1260
- Ilagan, J. O., Ramakrishnan, A., Hayes, B., Murphy, M. E., Zebari, A. S., Bradley, P., et al. (2015). U2AF1 mutations alter splice site recognition in hematological malignancies. *Genome Res.* 25, 14–26. doi: 10.1101/gr.181016.114
- Kamburov, A., Lawrence, M. S., Polak, P., Leshchiner, I., Lage, K., Golub, T. R., et al. (2015). Comprehensive assessment of cancer missense mutation clustering in

- protein structures. *Proc. Natl. Acad. Sci. U.S.A.* 112, E5486–E5495. doi: 10.1073/pnas.1516373112
- Kataoka, N. (2016). Purification of RNA-protein splicing complexes using a tagged protein from in vitro splicing reaction mixture. *Methods Mol. Biol.* 1421, 45–52. doi: 10.1007/978-1-4939-3591-8_5
- Kataoka, N. (2017). Modulation of aberrant splicing in human RNA diseases by chemical compounds. *Hum. Genet.* 136, 1237–1245. doi: 10.1007/s00439-017-1789-4
- Kataoka, N., Diem, M. D., Kim, V. N., Yong, J., and Dreyfuss, G. (2001). Magoh, a human homolog of *Drosophila mago nashi* protein, is a component of the splicing-dependent exon-exon junction complex. *EMBO J.* 20, 6424–6433. doi: 10.1093/emboj/20.22.6424
- Kataoka, N., Diem, M. D., Yoshida, M., Hatai, C., Dobashi, I., and Dreyfuss, G. (2011). Specific Y14 domains mediate its nucleocytoplasmic shuttling and association with spliced mRNA. *Sci. Rep.* 1:92. doi: 10.1038/srep00092
- Kataoka, N., and Dreyfuss, G. (2004). A simple whole cell lysate system for in vitro splicing reveals a stepwise assembly of the exon-exon junction complex. *J. Biol. Chem.* 279, 7009–7013. doi: 10.1074/jbc.M307692200
- Kataoka, N., and Dreyfuss, G. (2008). Preparation of efficient splicing extracts from whole cells, nuclei, and cytoplasmic fractions. *Methods Mol. Biol.* 488, 357–365. doi: 10.1007/978-1-60327-475-3_23
- Kawano, T., Kataoka, N., Dreyfuss, G., and Sakamoto, H. (2004). Ce-Y14 and MAG-1, components of the exon-exon junction complex, are required for embryogenesis and germline sexual switching in *Caenorhabditis elegans*. *Mech. Dev.* 121, 27–35. doi: 10.1016/j.mod.2003.11.003
- Kim, E., Ilagan, J. O., Liang, Y., Daubner, G. M., Lee, S. C., and Ramakrishnan, A. (2015). SRSF2 mutations contribute to myelodysplasia by mutant-specific effects on exon recognition. *Cancer Cell* 27, 617–630. doi: 10.1016/j.ccell.2015.04.006
- Kim, V. N., Kataoka, N., and Dreyfuss, G. (2001). Role of the nonsense-mediated decay factor hUpf3 in the splicing-dependent exon-exon junction complex. *Science* 293, 1832–1836. doi: 10.1126/science.1062829
- Komeno, Y., Huang, Y. J., Qiu, J., Lin, L., Xu, Y., and Zhou, Y. (2015). SRSF2 is essential for hematopoiesis, and its myelodysplastic syndrome-related mutations dysregulate alternative pre-mRNA splicing. *Mol. Cell. Biol.* 35, 3071–3082. doi: 10.1128/MCB.00202-15
- Kon, A., Yamazaki, S., Nannya, Y., Kataoka, K., Ota, Y., and Nakagawa, M. M. (2018). Physiological Srsf2 P95H expression causes impaired hematopoietic stem cell functions and aberrant RNA splicing in mice. *Blood* 131, 621–635. doi: 10.1182/blood-2017-01-762393
- Manley, J. L., and Krainer, A. R. (2010). A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes Dev.* 24, 1073–1074. doi: 10.1101/gad.1934910
- Mayeda, A., Srean, G. R., Chandler, S. D., Fu, X. D., and Krainer, A. R. (1999). Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements. *Mol. Cell. Biol.* 19, 1853–1863. doi: 10.1128/mcb.19.3.1853
- Mupo, A., Seiler, M., Sathiaselan, V., Pance, A., Yang, Y., and Agrawal, A. A. (2017). Hemopoietic-specific Sf3b1-K700E knock-in mice display the splicing defect seen in human MDS but develop anemia without ring sideroblasts. *Leukemia* 31, 720–727. doi: 10.1038/leu.2016.251
- Nguyen, H. D., Leong, W. Y., Li, W., Reddy, P. N. G., Sullivan, J. D., Walter, M. J., et al. (2018). Spliceosome mutations induce r loop-associated sensitivity to ATR inhibition in myelodysplastic syndromes. *Cancer Res.* 78, 5363–5374. doi: 10.1158/0008-5472.CAN-17-3970
- Obeng, E. A., Chappell, R. J., Seiler, M., Chen, M. C., Campagna, D. R., Schmidt, P. J., et al. (2016). Physiologic expression of Sf3b1(K700E) causes impaired erythropoiesis, aberrant splicing, and sensitivity to therapeutic spliceosome modulation. *Cancer Cell* 30, 404–417. doi: 10.1016/j.ccell.2016.08.006
- Papaemmanuil, E., Cazzola, M., Boultonwood, J., Malcovati, L., Vyas, P., Bowen, D., et al. (2011). Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N. Engl. J. Med.* 365, 1384–1395. doi: 10.1056/NEJMoa1103283
- Papaemmanuil, E., Gerstung, M., Malcovati, L., Tauro, S., Gundem, G., Van Loo, P., et al. (2013). Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* 122, 3616–3627. doi: 10.1182/blood-2013-08-518886
- Przychodzen, B., Jerez, A., Guinta, K., Sekeres, M. A., Padgett, R., Maciejewski, J. P., et al. (2013). Patterns of missplicing due to somatic U2AF1 mutations in myeloid neoplasms. *Blood* 122, 999–1006. doi: 10.1182/blood-2013-01-480970
- Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. doi: 10.1038/nmeth.2089
- Shao, C., Yang, B., Wu, T., Huang, J., Tang, P., Zhou, Y., et al. (2014). Mechanisms for U2AF to define 3' splice sites and regulate alternative splicing in the human genome. *Nat. Struct. Mol. Biol.* 21, 997–1005. doi: 10.1038/nsmb.2906
- Shiozawa, Y., Malcovati, L., Galli, A., Sato-Otsubo, A., Kataoka, K., Sato, Y., et al. (2018). Aberrant splicing and defective mRNA production induced by somatic spliceosome mutations in myelodysplasia. *Nat. Commun.* 9:3649. doi: 10.1038/s41467-018-06063-x
- Shirahata-Adachi, M., Iriyama, C., Tomita, A., Suzuki, Y., Shimada, K., and Kiyoi, H. (2017). Altered EZH2 splicing and expression is associated with impaired histone H3 lysine 27 tri-Methylation in myelodysplastic syndrome. *Leuk. Res.* 63, 90–97. doi: 10.1016/j.leukres.2017.10.015
- Shirai, C. L., Ley, J. N., White, B. S., Kim, S., Tibbitts, J., Shao, J., et al. (2015). Mutant U2AF1 expression alters hematopoiesis and pre-mRNA splicing in vivo. *Cancer Cell* 27, 631–643. doi: 10.1016/j.ccell.2015.04.008
- Walter, M. J., Shen, D., Shao, J., Ding, L., White, B. S., Kandoth, C., et al. (2013). Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. *Leukemia* 27, 1275–1282. doi: 10.1038/leu.2013.58
- Wang, D. O., Ninomiya, K., Mori, C., Koyama, A., Haan, M., Kitabatake, M., et al. (2017). Transport granules bound with nuclear cap binding protein and exon junction complex are associated with microtubules and spatially separated from eIF4E granules and p bodies in human neuronal processes. *Front. Mol. Biosci.* 4:93. doi: 10.3389/fmolb.2017.00093
- Watakabe, A., Inoue, K., Sakamoto, H., and Shimura, Y. (1989). A secondary structure at the 3' splice site affects the in vitro splicing reaction of mouse immunoglobulin mu chain pre-mRNAs. *Nucleic Acids Res.* 17, 8159–8169. doi: 10.1093/nar/17.20.8159
- Watakabe, A., Tanaka, K., and Shimura, Y. (1993). The role of exon sequences in splice site selection. *Genes Dev.* 7, 407–418. doi: 10.1101/gad.7.3.407
- Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., et al. (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* 478, 64–69. doi: 10.1038/nature10496
- Zhang, J., Lieu, Y. K., Ali, A. M., Penson, A., Reggio, K. S., Rabadan, R., et al. (2015). Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. *Proc. Natl. Acad. Sci. U.S.A.* 112, E4726–E4734. doi: 10.1073/pnas.1514105112

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Masaki, Ikeda, Hata, Shiozawa, Kon, Ogawa, Suzuki, Hakuno, Takahashi and Kataoka. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.