博士論文（要約）

骨髄異形成症候群における遺伝子発現と選択的 RNA スプライシング異常の網羅的解析

塩澤　裕介

論文の内容の要旨

論文題目　骨髄異形成症候群における遺伝子発現と
選択的 RNA スプライシング異常の網羅的解析

氏名　　塩澤 裕介

Myelodysplastic syndromes (MDS) are a heterogeneous group of clonal hematopoietic disorders that are characterized by ineffective hematopoiesis, peripheral blood cytopenias, and a risk of leukemic transformation. Little is known about the relationships among genomic, transcriptomic, and hematopoietic abnormalities, especially those between splicing factor mutations and myelodysplastic hematopoiesis. Splicing factors represent a novel class of driver genes in MDS, with *SF3B1*, *SRSF2*, and *U2AF1* being the most frequently mutated genes. Abnormal RNA splicing is thought to play a central role in the pathogenic mechanism of mutated splicing factors. Here we performed a comprehensive transcriptome analysis of MDS using transcriptome sequencing. The purpose of this study was to gain insight into the relationship between genomic and transcriptomic alterations, particularly including the effect of splicing factor mutations on RNA splicing.

We enrolled 214 patients with myeloid neoplasms with myelodysplasia. Genomic DNA was available for 211 patients, for which targeted-capture sequencing was performed. RNA sequencing was performed on bone marrow mononuclear cells (BMMNCs, n=165) and CD34+ cells (n=100) obtained from 214 patients, of whom 51 were analyzed for both cell fractions. Libraries for RNA sequencing were prepared from polyA-selected RNA. The sequencing reads were aligned to the human reference genome (hg19) using the RNA sequencing unified mapper version 2.0.4. Alternative splicing and differential expression analyses were performed using JuncBASE version 0.6 and edgeR, respectively.

Targeted-capture sequencing revealed that 124 patients (58%) had mutations in one or more splicing factors, including *SF3B1* (n=68 [32%]), *SRSF2* (n=39 [18%]), and *U2AF1* (n=14 [6.5%]). RNA sequencing revealed a total of 190,301 and 147,809 alternative splicing events in BMMNCs and CD34+ cells, respectively. Splice junction usage was compared between splicing factor-mutated and unmutated samples to detect alternative splicing events associated with each splicing factor mutation. The analyses were confined to two splicing factors, *SRSF2* and *SF3B1*, for which sufficient numbers of mutated samples were available for both cell fractions. A total of 2,057 and 3,831 events were significantly associated with *SRSF2* and *SF3B1* mutations, respectively, in either BMMNCs or CD34+ cells (q-value <0.01).

Mutant *SRSF2*- and *SF3B1*-associated splicing events showed two different features. First, the patterns of alternative splicing differed substantially. The splicing pattern most frequently associated with

SRSF2 mutation was altered exon usage, which was either inclusion or skipping of a 'cassette' exon. In sharp contrast, *SF3B1* mutation was frequently associated with increased usage of alternative 3' splice sites. Second difference was the specificity of alternative splicing events. Mutant *SRSF2*-associated alternative splicing events were not specific to the *SRSF2*-mutated samples, but were also observed in those without splicing factor mutations, resulting in modest changes in relative abundance of alternative isoforms. By contrast, the mutant *SF3B1*-associated altered 3' splice sites were often observed almost exclusively in the mutated samples, and were barely detected in the controls. The enhanced expression of the abnormal 3' splice sites in the *SF3B1*-mutated samples was validated by RT-PCR for several genes.

We next assessed the significance of the splicing alterations at the transcript level. Alternative splicing events were classified into five groups: (1) truncating alterations, (2) non-truncating alterations, (3) alternative first coding exons, (4) alternative last coding exons, and (5) alterations in non-coding regions. Truncating alterations introduce a frameshift or a premature termination codon into a transcript. Non-truncating alterations yield an in-frame isoform with a nucleotide deletion or insertion. The alternative 3' splice sites associated with *SF3B1* mutation were often clustered around 10–30 bp upstream of canonical 3' splice sites, resulting in nucleotide insertion at authentic exon-exon junctions. Of 671 mutant *SF3B1*-associated alternative 3' splice sites, 373 (56%) yielded an out-of-frame transcript or introduced a premature termination codon, while 192 (29%) were non-truncating events. *SRSF2* mutation was frequently associated with alternative exon usage that resulted in truncated isoforms. This includes usage of cryptic exons harboring a premature termination codon, and inclusion or skipping of out-of-frame exons.

Truncating splicing alterations are likely to have deleterious effects on gene function through reduction of their authentic transcripts. To assess their effects, comprehensive analysis of differential gene expression was performed between the *SRSF2*- and *SF3B1*-mutated CD34+ cells and those without splicing factor mutations. The expression level of transcripts without truncating splicing alterations was estimated from the fraction of reads that showed alternative splicing. Target genes of mutant *SF3B1*-associated alternative 3' splice sites showed a highly consistent reduction in the authentic transcripts; 145 of 290 (50%) target genes expressed in the CD34+ fraction were significantly down-regulated (q-value<0.01). Furthermore, down-regulation associated with aberrant 3' splice sites represented by far the most significant changes in gene expression. Such highly significant changes were not detected for other splicing patterns of the mutant *SF3B1*-associated alterations, or for all the mutant *SRSF2*-associated events.

We next focused on individual target genes of splicing alterations with a moderate to large effect size. These criteria were met in 358 of 3,831 (9.3%) mutant *SF3B1*-associated alterations, which involved 316 genes as functional candidates. *SF3B1* mutation is strongly associated with myelodysplastic syndrome with ring sideroblasts characterized by defects in heme biosynthesis and iron accumulation to mitochondria. The candidates of functional targets included genes related to heme biosynthesis and iron metabolism: *ABCB7*,

PPOX, and *TMEM14C*. *ABCB7* is a causative gene for an inherited sideroblastic anemia, whose haploinsufficiency has been shown to cause iron accumulation to mitochondria. *ABCB7* was one of the most significantly down-regulated genes with an aberrant 3' splice site in the *SF3B1*-mutated samples. The alternative 3' splice site was located at 21 bp upstream from the authentic junction, of which usage led to insertion of a nucleotide sequence harboring a premature termination codon. *PPOX*, a gene encoding the penultimate enzyme of heme biosynthesis that catalyzes the oxidation of protoporphyrinogen IX to form protoporphyrin IX, also had an alternative 3' splice site associated with *SF3B1* mutation at 32 bp upstream of the canonical one, introducing a frameshift. It was also recapitulated by expression of the *SF3B1*^{K700E} mutant. TMEM14C is an inner mitochondrial membrane protein that facilitates the import of protoporphyrinogen IX into the mitochondrial matrix and synthesis of protoporphyrin IX by PPOX. Mutant *SF3B1*-associated transcripts had an alternative 3' splice site located in the 5'-untranslated region, the usage of which neither altered the amino acid sequence, nor had an effect on gene expression as evaluated by luciferase assay. Collectively, *SF3B1* mutation induces usage of aberrant 3' splice sites in two genes related to heme biosynthesis, *ABCB7* and *PPOX*, leading to down-regulation of their authentic transcripts. It should be experimentally assessed whether reduced expression of *ABCB7* and *PPOX* cooperatively contribute to development of anemia.

The candidates of functional targets of mutant *SF3B1*-associated splicing alterations included cancer-related genes: tumor suppressors including *NF1*, *DICER1*, *PML*, *PDS5A*, and *PPP2R5A*, as well as an oncogene *PTPN11*. All of their splicing alterations were usage of alternative 3' splice sites. *NF1*, *DICER1*, *PML*, *PDS5A*, *PPP2R5A*, and *PTPN11* had truncating splicing alterations, associated with reduction in their authentic transcripts.

SRSF2 mutation was mainly associated with modest changes in alternative splicing. However, despite a modest effect of an individual event, we found that *EZH2* had multiple truncating splicing alterations associated with *SRSF2* mutation. *EZH2*, a gene encoding the catalytic component of the polycomb repressive complex 2, acts as a tumor suppressor with recurrent loss-of-function mutations in myeloid malignancies. *SRSF2*-mutated samples showed an increased inclusion of a cryptic exon of *EZH2* harboring a premature termination codon. We also identified an association between *SRSF2* mutation and skipping of exon 11 as another alteration, which also led to a premature termination. Occurring independently from each other, these 2 cassette exons cooperatively contributed to down-regulation of the authentic transcript of *EZH2*. We also identified an additional targets of mutant *SRSF2*-associated splicing alterations potentially involved in the development of myelodysplasia: *CASP8*, and *CDK10*. Mutant *SRSF2*-associated exon skipping generated a prematurely terminating transcript, leading to a significant reduction of the authentic transcripts.

High frequency of splicing factor mutations is a cardinal feature of MDS. Although myelodysplasia is a common clinicopathological presentation of splicing factor-mutated neoplasms, different

mutations are associated with specific disease subtypes, such as those showing increased ring sideroblasts with *SF3B1* mutations and myeloproliferative subtypes with mutated *SRSF2*. Through a comprehensive transcriptome analysis of a large cohort, we demonstrated that different splicing factor mutations had distinct effects on RNA splicing. We also revealed precise targets of splicing alterations across a wide variety of genes that did not largely overlap among different splicing factor mutations. These results provide insights into the molecular basis of the distinct phenotypes for different splicing factor mutations, highlighting potential oncogenic activity of mutant splicing factors through multiple splicing alterations in cancer-related genes.

We also performed clustering analysis of gene expression profiles. Gene expression-based classification of MDS has not been established. Clustering analysis was first performed on gene expression profiles of bone marrow CD34+ cells from 100 cases. Gene expression data were filtered to 3,142 genes with large dispersion than those with similar mean expression level (trended dispersion). Consensus clustering using expression levels of 3,142 genes identified two robust subgroups, with clustering stability decreasing for the number of clusters ≥3. In other words, K-means clustering on the basis of global gene expression profiles was most stable when the samples were broadly classified into two subgroups. The first subgroup was characterized by a lower blast count and the up-regulation of genes specifically expressed in erythroid lineages, while the second one was significantly associated with increased expression of the genes related to stem/progenitor cells. Clinical outcomes also differed substantially; compared to the first subgroup, the second subgroup was significantly associated with survival in either univariate (hazard ratio 5.0 [95% confidence interval (CI), 1.8–14], $P<0.001$) or multivariate analysis (hazard ratio 6.8 [95% CI, 1.5–32], $P=0.015$). Notably, no leukemic transformation occurred in the first subgroup, which contrasted with the high leukemic transformation rate in the second subgroup (38%). The prognostic significance of this gene expression-based classification was validated in an independent cohort. Finally, we constructed a logistic regression model to predict these subgroups based on gene expression of unfractionated BMMNCs. Predicted high-risk subgroup was again significantly associated with poor prognosis in either univariate or multivariate analysis. This newly developed molecular classification will improve risk prediction and treatment stratification of MDS.