Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis

Hideki Makishima¹, Valeria Visconte¹, Hirotoshi Sakaguchi², Anna M. Jankowska¹, Sarah Abu Kar¹, Andres Jerez¹, Bartlomiej Przychodzen¹, Manoj Bupathi¹, Kathryn Guinta¹, Manuel G. Afable¹, Mikkael A. Sekeres¹, Richard A. Padgett³, Ramon V. Tiu¹, Jaroslaw P. Maciejewski¹

¹ Department of Translational Hematology and Oncology Research, Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH, USA.

² Department of Pediatrics, Nagoya University Graduate School of Medicine, Nagoya, Japan.

³ Department of Molecular Genetics, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA.

Key words: Spliceosome, U2AF1, SF3B1, SRSF2, MDS, AML

Running title: Spliceosomal mutations in MDS and AML

Corresponding author: Jaroslaw P. Maciejewski MD., Ph.D., FACP.

Taussig Cancer Institute/R40
9500 Euclid Avenue
Cleveland OH USA 44195
Phone: 216-445-5962
FAX: 216-636-2498
E-mail: maciejj@ccf.org
Myelodysplastic syndromes (MDS) are chronic and often progressive myeloid neoplasms associated with remarkable heterogeneity in the histomorphology and clinical course. Various somatic mutations are involved in the pathogenesis of MDS. Recently, mutations in a gene encoding a spliceosomal protein, \textit{SF3B1}, were discovered in a distinct form of MDS with ring sideroblasts. Whole exome sequencing of 15 patients with myeloid neoplasms was performed and somatic mutations in spliceosomal genes were identified. Sanger sequencing of 310 patients was carried out to assess phenotype/genotype associations. To determine the functional impact of spliceosomal mutations, we evaluated pre-mRNA splicing profiles by RNA deep sequencing. We identified additional somatic mutations in spliceosomal genes, including \textit{SF3B1}, \textit{U2AF1} and \textit{SRSF2}. These mutations alter pre-mRNA splicing patterns. \textit{SF3B1} mutations are prevalent in low risk MDS with ring sideroblasts, while \textit{U2AF1} and \textit{SRSF2} mutations are frequent in chronic myelomonocytic leukemia and advanced forms of MDS. \textit{SF3B1} mutations are associated with a favorable prognosis, while \textit{U2AF1} and \textit{SRSF2} mutations are predictive for shorter survival. Mutations affecting spliceosomal genes that result in defective splicing are a new leukemogenic pathway. Spliceosomal genes are likely tumor suppressors, and their mutations may constitute diagnostic biomarkers that could potentially serve as therapeutic targets.
Introduction

The myelodysplastic syndromes (MDS) are characterized by clonal hematopoiesis, a variety of chromosomal abnormalities, bone marrow failure and a propensity for evolution to acute myeloid leukemia (AML). Because of their often protracted course, MDS recapitulate the stages of acquisition of a malignant phenotype, thereby offering insights into leukemogenesis. While traditionally, histomorphology-based schemes have been applied to sub-classify MDS patients,\(^1,2\) this approach is unlikely to be reflective of the underlying pathogenesis. Instead, a better molecular characterization of MDS on the genomic, epigenetic and genetic levels is more likely to objectively diagnose patients, determine their prognosis and, based on the underlying molecular defects, direct the application of targeted therapies. The emerging realization of the molecular diversity of MDS parallels the clinical and phenotypic heterogeneity of this disease. Moreover, molecular defects have the potential to serve as biomarkers and are more likely to be suitable for the identification of therapy targets and responsiveness/refractoriness to treatment.

The application of high-throughput molecular technologies, including high density single nucleotide polymorphism arrays (SNP-A)\(^3\) and new sequencing technologies\(^4,5\) has led to the improved characterization of genomic lesions such as chromosomal aberrations and of somatic mutations affecting specific classes of genes,\(^5\) including signal transducers (e.g., \(CBL\))\(^7-10\) apoptotic genes (e.g., \(TP53\) and \(RAS\))\(^11-13\) genes involved in epigenetic regulation of DNA (e.g., \(DNMT3A\), \(IDH1/2\) and \(TET2\))\(^14-18\) and histone modifiers (e.g., \(EZH2\), \(UTX\) and \(ASXL1\)).\(^19-24\) While some mutations in these factors are activating, most are loss of function or hypomorphic mutations and affect \textit{bona fide} tumor suppressor genes (TSG). Of greatest diagnostic impact are recurrent mutations found in specific genes. Most TSG mutations are not canonical, though, making systematic clinical diagnostics more difficult.

We and other groups recently identified a variety of new mutations present at distinct frequencies in subgroups of MDS patients.\(^25,26\) Analogous to previously applied strategies that identified \(TET2\), \(CBL\), \(EZH2\) and other mutations,\(^7,18,21\) we have identified mutations affecting, in a recurrent fashion, genes of the spliceosome machinery, using a combination of targeted search approaches and unbiased mass sequencing. Such mutations likely constitute a new class of TSG ubiquitously involved in leukemogenesis.
Methods

Patient population.

Bone marrow aspirates or blood samples were collected from 315 patients with MDS (N=88), MDS/myeloproliferative neoplasms (MDS/MPN, N=66), MPN (N=52), secondary AML (sAML) (N=54) that evolved from these conditions and primary AML (pAML) (N=55) seen at Cleveland Clinic or Nagoya University between 2003 and 2010 (Table 1). Informed consent for sample collection was obtained according to protocols approved by the institutional IRBs and in accordance with the Declaration of Helsinki. Diagnosis was confirmed and assigned according to World Health Organization (WHO) classification criteria. Low-risk MDS was defined as patients having <5% myeloblasts. Patients with >5% myeloblasts constituted those with advanced disease. Serial samples were obtained for 38 patients. To study the germline genotype, immunoselected CD3+ lymphocytes were used. Cytogenetic analysis was performed according to standard banding techniques based on 20 metaphases. Clinical parameters studied included age, sex, overall survival (OS), blood counts, and metaphase cytogenetics. The median follow up of the cohort was 18 months (1-168 months).

Cytogenetics and single nucleotide polymorphism array (SNP-A) analyses.

Technical details regarding sample processing for SNP-A assays were previously described. Affymetrix 250K and 6.0 Kit (Affymetrix, Santa Clara, CA) were used. A stringent algorithm was applied for the identification of SNP-A lesions. Patients with SNP-A lesions concordant with metaphase cytogenetics or typical lesions known to be recurrent required no further analysis. Changes reported in our internal or publicly-available (Database of Genomic Variants; http://projects.tcag.ca/variation) copy number variation (CNV) databases were considered non-somatic and excluded. Results were analyzed using CNAG (v3.0) or Genotyping Console (Affymetrix). All other lesions were confirmed as somatic or germline by analysis of CD3-sorted cells.

Whole exome sequencing.

Genomic DNA was extracted from bone marrow or peripheral blood using standard methods and subjected agarose gel and optical density ratio tests to confirm the purity and concentration prior to Covaris (Covaris, Inc., Woburn, MA) fragmentation. 0.5-2.5 μg of fragmented genomic DNA was tested for size distribution and concentration using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). Illumina libraries were made from qualified fragmented gDNA using NEBNext reagents (New England Biolabs, Ipswich, MA) and the resulting libraries were subjected to exome enrichment using NimbleGen SeqCap EZ Human Exome Library v2.0 (Roche NimbleGen, Inc., Madison, WI) following the manufacturer’s instructions. Enriched libraries were tested for enrichment by qPCR and for size distribution and
concentration by an Agilent 2100 Bioanalyzer. The samples were then sequenced on an Illumina HiSeq2000 (Illumina, San Diego, CA) which generated paired-end reads of 100 nucleotides. Paired bone marrow mononuclear cells and CD3+ peripheral blood lymphocytes were used as germline controls. DNAnexus software (DNAnexus, Inc, Mountain View, CA) (https://dnanexus.com) was used to visualize single nucleotide changes, insertions and/or deletions at the gene, exon and base pair levels. A rational bioanalytic algorithm was applied to identify candidate non-synonymous alterations. Multiple steps were performed to reduce the false positive rate within reported results. First, whole exome assembly was non-redundantly mapped using the reference genome hg19. Next, the analytic algorithm within DNAnexus called all the positions that vary from a reference genome. Each potential mutation was compared against databases of known SNPs, including Entrez Gene (http://www.ncbi.nlm.nih.gov/gene) and the Ensembl Genome Browser (http://useast.ensembl.org/index.html). These candidate alterations were subtracted by the results of CD3+ peripheral blood DNA and subsequently validated using Sanger sequencing (see below). Moreover, spliceosome-associated gene mutations were screened using whole exome sequencing results available through The Cancer Genome Atlas (TCGA; http://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp).

**Sanger sequencing analysis.**

All exons of selected genes were amplified and underwent direct genomic sequencing by standard techniques on the ABI 3730xl DNA analyzer (Applied Biosystems, Foster City, CA) as previously described. All mutations were detected by bidirectional sequencing and scored as pathogenic if not present in non-clonal paired CD3-derived DNA. Frameshift mutations were validated by cloning and sequencing individual colonies (TOPO TA cloning, Invitrogen, Carlsbad, CA). For confirmation of the somatic nature of the mutations, exons containing mutations were tested in non-clonal control DNA.

**Whole RNA deep sequencing.**

Total RNA was extracted from bone marrow mononuclear cells using the Nucleospin RNA II Kit (Macherey-Nagel, Bethlehem, PA) with DNAase treatment. The integrity and purity of total RNA were assessed using Agilent Bioanalyzer. 1-2 μg of cDNA was generated using Clontech SmartPCR cDNA kit (Clontech Laboratories, Inc., Mountain View, CA) from 100ng of total RNA. cDNA was fragmented using Covaris, profiled using Agilent Bioanalyzer, and subjected to Illumina library preparation using NEBNext reagents (New England Biolabs, Ipswich, MA). The quality and quantity and the size distribution of the Illumina libraries were determined using an Agilent Bioanalyzer. The libraries were then submitted for Illumina HiSeq2000 sequencing according to standard procedures. Paired-end 90 base pair reads were generated and subjected to data analysis using the platform provided by DNAnexus. DNAnexus software allowed
visualization of reads derived from spliced mRNA and those that completely match the genome, including both sense and antisense.

Statistical analysis of clinical data.

The Kaplan-Meier method was used to analyze survival outcomes (OS) of subgroups characterized by the presence of mutant vs. wild type (WT) variants of specific spliceosome-associated gene mutations with the log-rank test (JMP9; SAS, Cary, NC). Significance was determined at a two-sided alpha level of 0.05.
Results
Detection of somatic spliceosomal mutations in myeloid malignancies.

Initially, we performed whole exome sequencing of 15 index cases with various forms of myeloid malignancies and identified distinct somatic mutations in genes encoding components of the spliceosomal machinery (Supplemental Table 1). A heterozygous missense mutation in U2AF1 (Q157R) was found in a patient with refractory cytopenia with unilineage dysplasia (RCUD) and uniparental disomy of 2q (UPD2q) (Figure 1A). Analysis of DNA from CD3 positive cells showed a much lower frequency of the base change (2 out of 15 reads), highlighting the somatic nature of this alteration. The 2 reads originated from contamination by a small amount of the mutated clone. The contamination was confirmed by flow cytometric analysis after magnetic bead sorting. Similarly, heterozygous, somatic SF3B1 mutations (E622D) were detected in a patient with refractory anemia with ring sideroblasts (RARS) associated with marked thrombocytosis (RARS-T) (Figure 1B). In a patient with sAML, we identified a somatic mutation (M1307I) in PRPF8 and a heterozygous mutation (R27X) in LUC7L2, respectively (Figure 1C). Based on these findings, we screened other rationally selected genes within the spliceosomal machinery in patients with MDS and related disorders (N=120). Using this targeted approach, we identified mutations in ZRSR2 (W153X) and SRSF2 (P95R) (Figure 1D-E). All of these mutations were somatic as confirmed by sequencing of the corresponding germline-derived DNA. Moreover, screening of whole exome sequencing results for AML available through the Cancer Genome Atlas revealed the presence of mutations in genes encoding other components of the spliceosome, including mutations in HCFC1 (P72L), SAP130 (T247I), SRSF6 (W123S), SON (M1024V) and U2AF26 (C116X) (Supplemental Figure 1). We found that among the initially screened cohort of patients with myeloid malignancies, mutations were highly recurrent in U2AF1, SF3B1 and SRSF2, while mutations of other spliceosomal genes were each detected in one out of 120 cases, indicating a frequency of <1%. While we identified 2 mutant copies of U2AF1 (21q22.3) in a sAML case with trisomy 21 (Supplemental Figure 2), all other 19 cases of trisomy 21 screened for this mutation were negative. Moreover, no homozygous mutations were found in patients with UPD21q (N=8). Similarly, all mutations of SF3B1 and SRSF2 were heterozygous, and patients with somatic UPD2q33.1 or UPD17q25 (regions containing SF3B1 and SRSF2, respectively) did not harbor homozygous mutations of the associated genes.

Clinical associations and frequencies of spliceosomal mutations in myeloid malignancies.

We subsequently screened a large cohort of patients (N=310) with MDS and related disorders (including the initially screened 120 cases) in a stepwise fashion to determine the frequency of spliceosomal mutations discovered in the index cases. Based on the initial screen, we noted that mutations in U2AF1, SF3B1 and SRSF2 were the most frequent. Consequently, we sequenced 120 cases for ZRSR2, LUC7L2 and PRPF8 and a total 310 for SF3B1, U2AF1, and SRSF2. All SF3B1 mutations were located in exon 14 or 15, with
the K700 mutation being the most recurrent (Figure 1B). Similarly, all mutations in SRSF2 affected position P95 (Figure 1D). In contrast, mutations in U2AF1 affected exons 2 and 6, corresponding to the 2 zinc finger domains of this protein (Figure 1A, Supplemental Figure 3). The extended cohort of patients was used to identify phenotype/genotype associations (Supplemental Table 2). To avoid any misunderstanding, we included information as to which cases were analyzed by whole exome sequencing in Supplemental Table 2. Additionally, in Supplemental Table 1, we clarified the mutational status of spliceosomal genes. In low-risk MDS, mutations of any one of these three genes were found in 39% of patients, and further analysis revealed that mutations in SF3B1 were highly associated with RARS. Among patients with MDS/MPN, SF3B1 mutations were not common in chronic myelomonocytic leukemia (CMML), but they were frequent in patients with RARS-T and thus the presence of RS was found to correlate highly with SF3B1 mutations, irrespective of other clinical or morphologic features (Figure 2A). In contrast, U2AF1 mutations were most frequent in the high risk MDS/AML cohort (11%), while SRSF2 was most frequently mutated in MDS/MPN (24%) particularly in CMML (28%) (Figure 2A).

Impact of spliceosomal mutations on clinical outcomes.

Subsequently, we studied the impact of the most common spliceosomal mutations on clinical outcomes. We first analyzed the entire cohort of patients (Table 1) and determined the survival of patients in whom the three most common spliceosomal mutations were present. We included the information as to SF3B1 mutation status of 39 cases reported in our previous paper.32 When 310 patients genotyped for these mutations were analyzed, the presence of SF3B1 mutations was associated with longer survival, U2AF1 mutations with shorter survival, while SRSF2 mutations had no effect on survival. We then analyzed the impact of these mutations in more clinically uniform subgroups to more precisely determine their clinical consequences. As expected, in sAML and pAML, due to overall poor prognosis, the presence of spliceosomal mutations did not further affect survival (Supplemental Figure 4). However, in low-risk MDS, patients with SF3B1 mutations showed a tendency towards better prognosis (p= .09), while those with SRSF2 mutations had worse survival. In MDS/MPN, SRSF2 mutations were more common than U2AF1; however, U2AF1 mutations were associated with shorter survival (Figure 2B). When serial samples were analyzed, we found that an U2AF1 mutation detected at the sAML stage were present from initial MDS presentations, suggesting an ancestral origin of this mutation (Supplemental Figure 5). Overall, SF3B1 mutations were less prevalent in patients with advanced forms of MDS, indicating that mutation of this factor does not contribute to progression (Figure 2A).

Effects of spliceosomal mutations on spliceosomal function.

Conceptually, mutations of spliceosomal proteins could result in defective splicing, including intron retention, altered splice site recognition or altered alternative splicing. To determine the functional consequences of
spliceosomal mutations on splicing, we performed whole mRNA deep sequencing. In the presence of functional spliceosomal machinery, sequencing reads are expected not to cross the intron/exon boundaries, and therefore should not contain any intronic sequences. We analyzed RNA sequencing results in patients with mutations in \textit{U2AF1} (N=3), \textit{SRSF2} (N=2), \textit{SF3B1} (N=2) and \textit{U2AF26} (N=1), as well as in a healthy controls and one MDS patient with a WT configuration of these genes. There was no genome-wide increase in intron retention observed in the mutant patients. However, we found a number of specific genes in which the splicing pattern was altered. For instance, \textit{U2AF1} mutations were associated with defective splicing of intron 5 of \textit{TET2} at both splice sites (Figure 3A, Supplemental Figure 6), while splicing of other \textit{TET2} introns were less affected. Another gene in which splicing was affected was \textit{RUNX1}. At both the 3’ and 5’ splice sites of \textit{RUNX1} intron 6, unspliced reads were more frequent than spliced reads (Figure 3B). Such a splicing abnormality was more prominent in cases with \textit{SRSF2} mutations but not detected in the cases with \textit{SF3B1} mutations or WT spliceosomal genes. Similarly, \textit{U2AF26} mutations resulted in an alteration of \textit{RUNX1} splicing (Supplemental Figure 7). Moreover, alternative splicing analysis showed that exon 9 of \textit{FECH} was skipped in \textit{U2AF1} mutant cases but not in \textit{U2AF1} WT cases, including those with \textit{SF3B1} mutations (Supplemental Figure 8).
Discussion

In recent years, a number of somatic mutations have been associated with various myeloid malignancies, including MDS. Systematic screening studies revealed that such mutations do not only occur as sole abnormalities, but are often found in combination, likely contributing to the phenotypic heterogeneity found in MDS. During our search for new molecular lesions associated with MDS, we identified somatic mutations affecting various components of the spliceosomal machinery, similar to previous reports.

Until these recent results were obtained, spliceosomal dysfunction was only rarely associated with malignant transformation.

In our studies in myeloid malignancies, the most commonly affected spliceosomal genes included SF3B1, U2AF1 and SRSF2. While these mutations were common and found in a wide spectrum of myeloid diseases, some are strongly associated with specific phenotypic features, such as SF3B1 mutations with MDS or MDS/MPN with RS and SRSF2 mutations with CMML and advanced forms of MDS such as sAML and RA with excess blasts. Serial studies performed on patients from the initial diagnosis of low-risk MDS through subsequent transformation indicate that U2AF1 mutations may represent early ancestral events. A higher cross-sectional prevalence among advanced cases suggests that both U2AF1 and SRSF2 constitute high-risk defects. Indeed, U2AF1 and SRSF2 mutations were associated with worse survival in CMML and low-risk MDS, respectively. In contrast, SF3B1 mutations are associated with generally good prognosis, compatible with their association with RARS, which has a protracted clinical course. That these mutations were not overlapping and also not found in cases with other spliceosomal defects suggests that the effect of a spliceosomal mutation on myeloid malignancies might be inhibited by mutation in another factor or that they might not have a synergetic impact on leukemogenesis.

The prognostic impact of SF3B1 mutations has been investigated in several studies. While initial reports (354 patients) suggested SF3B1 mutations convey significantly longer OS, leukemia-free survival (LFS), and event-free survival in their cohort, subsequent analysis of 323 MDS cases, patients carrying an SF3B1 mutation showed a significantly better OS and no effect on LFS in RARS and RCMD-RS; the “protective effect” of SF3B1 mutation on OS was lost in multivariate analysis including clinical variables such as International Prognostic Scoring System (IPSS) score. Our results in the whole cohort of 310 MDS patients were generally consistent with these studies. However, we could not detect a significant difference in OS when subgroups with RS were analyzed separately, perhaps due to low number of WT cases. More recently, a large series of MDS patients (N=317) showed that the SF3B1 mutation status was not associated with time to AML progression or OS, regardless if all patients or only the subgroup of patients with RS were included in the analyses. Similar results were reported for myelofibrosis. When the survival benefit of SF3B1 mutations was examined in patients with MDS with increased RS, the
presence of SF3B1 mutations was associated with better OS and LFS in univariate analysis; however, significance was completely accounted for by WHO morphologic risk categorization and there was no additional prognostic value from the presence or absence of SF3B1 mutations when RARS and RCMD-RS were analyzed separately. In sum, most of the available results suggest that SF3B1 mutations convey better survival in the cohort of whole myeloid malignancies including RARS and RCMD-RS. However, such significant effects seem less in each background-matched specific disease phenotype by multivariate analysis including, for example, IPSS score.

Most of the mutations we detected were heterozygous, indicating that homozygous mutations may lead to cell death or that some of the functional consequences are related to dominant negative effects. Other mutations, including those in LUC7L2, ZRSR2 and PRPF8, appear to be less prevalent and, due to the small number of positive cases within the already large cohort studied, we were unable to establish whether these mutations correlated with characteristic phenotypic features and survival. It is possible that the mutations lead to distinct phenotypes by affecting different stages of splicing and/or by causing defective splicing of specific gene transcripts. It is also possible that the phenotypic features are related to the transcriptional spectrum of cells in which the mutations occur. For instance, SF3B1 mutations may affect splicing of transcripts coding for proteins associated with iron handling in erythroid precursors, leading to RS. Such phenotypes would then be absent in myeloid cells lacking the corresponding transcripts.

The precise pathogenetic mechanisms associated with facilitation of clonal evolution remain unclear. Mechanistically, defective splicing of specific genes may have similar consequences to loss of function mutations through the retention of introns (Supplemental Figure 9). We have performed deep sequencing of mRNA in hematopoietic cells derived from mutant cases and compared the results obtained in cells with a corresponding WT form of the gene. For U2AF1 and U2AF26 we demonstrated that accumulation of unspliced transcripts rather than abnormal alternative splicing is the main consequence of the mutations. We did not observe widespread defective splicing, but rather distinct changes in specific introns, as seen in RUNX1, that would functionally have the same effect as RUNX1 mutations. This splicing abnormality of RUNX1 is also observed in cases with SRSF2 mutations but not in SF3B1 mutants. U2AF1 and SRSF2 are frequently mutated in high risk MDS and CMML where RUNX1 mutations are relevant, which explains our theory that spliceosomal mutations results in similar phenotype to the corresponding loss of function mutations. Such an indirect alteration of key proteins known to be involved in malignant transformation could explain the leukemogenic effects of spliceosomal mutations, in particular how spliceosomal mutations may result in a phenocopy of features associated with known TSG mutations. The most important results of this study of the impact of spliceosomal mutations on the transcriptome are that the associated splicing defects affect a specific subset of mRNAs (Figure 3, Supplemental Figure 8, 9). Some of these RNAs code for key proteins involved in malignant transformation. Thus, there is a functional convergence of diverse
spliceosomal mutations toward effects on specific genes, a phenomenon which explains the similar phenotypes of some of the different mutations. In contrast, mutations in other genes can lead to unsplicing of distinct genes and thereby result in particular phenotypic features.

According to a recent report, knock down of \textit{U2AF1} expression leads to the skipping of a specific exon of \textit{FECH} during splicing.\textsuperscript{40} Based on the same phenomenon seen in alternative splicing analysis using mRNA deep sequencing of \textit{U2AF1} mutants, among the three most frequently mutated spliceosomal genes, \textit{U2AF1} mutations might result in loss of function. Further studies will clarify the whole alteration of mRNA splicing profiles induced by spliceosomal gene mutations.

Interestingly, spliceosome mutations, specific in \textit{SF3B1}, were reported in chronic lymphocytic leukemia.\textsuperscript{41-43} The discovery of recurrent somatic mutations in various genes encoding spliceosomal proteins indicates that spliceosomal defects constitute an important and ubiquitous pathway in malignant transformation. Further mutational screening in other malignancies might uncover the pathogenesis of common or unique tumorgenesis in various tissues.

The most relevant question, both biologically and clinically, is which other gene mutations are associated with the mutation status of individual spliceosomal genes. According to our preliminary results, \textit{U2AF1} mutations seem to be most commonly associated with \textit{ASXL1} and \textit{TET2} mutations, whereas \textit{SF3B1} can occur in the context of \textit{RUNX1} mutations (data not shown). Some presentations by other groups in ASH 2011 meeting suggested a correlation between \textit{SF3B1} and \textit{DNMT3A} mutations\textsuperscript{44,45} and a favorable survival impact of \textit{RUNX1} mutations in CMML with \textit{SRSF2}.\textsuperscript{46} Due to the large numbers of diverse mutations, very large studies will be needed to fully evaluate the correlation between the mutation status of spliceosomal genes and other genes in a proper fashion.

In sum, our studies demonstrate the wide-spread presence of mutations in genes involved in splicing in myeloid neoplasms. Such mutations may lead to distinct phenotypes and, because there is impact on survival, their detection may have future diagnostic utility.

\textbf{Acknowledgments}
We would like to thank the TCGA for access to the whole genome sequencing results described in the text. This work was supported by National Institutes of Health (Bethesda, MD; NIH) grants RO1HL-082983 (J.P.M.), U54 RR019391 (J.P.M.), K24 HL-077522 (J.P.M.), and a grant from the AA & MDS International Foundation (Rockville, MD), and the Robert Duggan Charitable Fund (Cleveland, OH; J.P.M.).
Authorship
H.M. designed research, performed research, collected data, performed statistical analysis and wrote the manuscript. V.V. performed research, collected data, and wrote the manuscript. H.S., A.M.J., S.A., A.J., B.P. performed research. M.B., K.G., M.G.A. collected data. M.A.S., R.V.T. collected data, analyzed and interpreted data, and wrote the manuscript. R.A.P. contributed analytical tools, collected data, analyzed and interpreted data, and wrote the manuscript. J.P.M. designed research, contributed analytical tools, collected data, analyzed and interpreted data, and wrote the manuscript.
Conflict of interest disclosure: The authors declare no competing financial interests.
Corresponding author: Jaroslaw P. Maciejewski. Taussig Cancer Institute/R40, 9500 Euclid Avenue, Cleveland, OH 44195 e-mail: maciejji@ccf.org
REFERENCES

<table>
<thead>
<tr>
<th>TABLES AND FIGURE LEGENDS</th>
</tr>
</thead>
</table>

**Table 1. Clinical characteristics of patients participating in this study.**

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MDS</strong></td>
<td>88</td>
</tr>
<tr>
<td>Low risk</td>
<td>58</td>
</tr>
<tr>
<td>RCUD / RCMD / 5q- / MDS-U*</td>
<td>38</td>
</tr>
<tr>
<td>RARS</td>
<td>20</td>
</tr>
<tr>
<td>High risk</td>
<td></td>
</tr>
<tr>
<td>RAEB 1/2*</td>
<td>30</td>
</tr>
<tr>
<td><strong>MDS/MPN</strong></td>
<td>66</td>
</tr>
<tr>
<td>CMML/aCML/JMML</td>
<td>48</td>
</tr>
<tr>
<td>MDS/MPN-U* (RARS-T)</td>
<td>18 (11)</td>
</tr>
<tr>
<td><strong>MPN</strong></td>
<td>52</td>
</tr>
<tr>
<td>PV/PMF/ET</td>
<td>16</td>
</tr>
<tr>
<td>CML</td>
<td>36</td>
</tr>
<tr>
<td><strong>AML</strong></td>
<td>109</td>
</tr>
<tr>
<td>Primary AML</td>
<td>55</td>
</tr>
<tr>
<td>Secondary AML*</td>
<td>54</td>
</tr>
</tbody>
</table>

* include 8 cases with therapy related myeloid malignancies.

MDS, myelodysplastic syndromes; RCUD, refractory cytopenia with unilineage dysplasia; RCMD, refractory cytopenia with multilineage dysplasia; MDS-U, MDS unclassifiable; RARS, refractory anemia with ring sideroblasts; RAEB, refractory anemia with excess blasts; MDS/MPN, MDS/myeloproliferative neoplasms; CMML, chronic myelomonocytic leukemia; aCML, atypical chronic myeloid leukemia; JMML, juvenile myelomonocytic leukemia; RARS-T, RARS associated with marked thrombocytosis; PV, polycythemia vera; PMF, primary myelofibrosis; ET, essential thrombocythemia; AML, acute myeloid leukemia.
Figure 1 Somatic spliceosomal gene (U2AF1, SF3B1, SRSF2, LUC7L2, PRPF8 and ZRSR2) mutations as detected by NGS and Sanger sequencing technologies.

(A) Using a NGS-based whole exome sequencing analysis of whole bone marrow DNA from a patient with RCUD (left), a mutation of U2AF1 (21q22.3) at position 44,514,777 (T>C) was detected in 13 of 18 reads. Analysis of DNA from CD3 positive cells showed a much lower frequency of the base change (2 out of 15 reads, right), highlighting the somatic nature of this alteration. The finding was confirmed by Sanger sequencing. Arrows and bars indicate the specific nucleotide and predicted codon, respectively. It should be noted that U2AF1 is expressed from the minus strand, and therefore the NGS presentation (upper panels) is complementally reversed in comparison to the Sanger sequencing results (middle panels). This heterozygous somatic mutation results in the predicted nucleotide change 470 A>G in exon 6 of the coding region, which lead to the amino acid change Q157R in the second zinc finger domain. In the entire cohort, 27 mutations were observed in 26 patients, including a whole gene deletion. All 26 missense mutations were located in one of the 2 zinc finger domains (ZNF); 2 residues, S34 or Q157, were frequently affected (lower panels). RRM indicates RNA recognition motif. (B) Using a NGS- analysis of CMML patient (middle left), a mutation of SF3B1 (2q33.1) at position 198,267,491(C>G) was detected in 9/ 24 reads. The somatic nature of this alteration was confirmed by an analogous analysis of the CD3+ fraction, with the change being less frequent (2/23; middle right). The mutation was confirmed by Sanger sequencing (bottom). This heterozygous somatic mutation results in the nucleotide change 1866 G>T in exon 14 of SF3B1, resulting in the amino acid change E622D in the HSH155 domain. Analysis of the entire cohort identified mutations in 33 patients, including a case with a whole gene deletion. (C) Further screening by NGS led to the detection of a nonsense mutation (R27X) in LUC7L2 (7q34) (top) which participates in the recognition of splice donor sites in association with the U1 snRNP spliceosomal subunit, and a missense mutation (M1307I) in PRPF8 (17p13.3) (bottom) which is a large U5 snRNP-specific protein essential for pre-mRNA splicing. RS, U5 2-snRNA bdg and MPN indicate serine/arginine-rich domain, U5-snRNA binding site 2, and Mpr1p, Pad1p N-terminal domain, respectively. (D) Mutations of SRSF2, an arginine/serine-rich splicing factor, were detected in 29 cases among the entire cohort, including 2 whole gene deletions and a microdeletion within the gene (top). All mutations were heterozygous and affected P95. The somatic nature of the P95R mutation was confirmed using whole bone marrow and T-cell rich fraction DNAs (bottom). (E) A nonsense mutation (W153X) was found in ZRSR2, another arginine/serine-rich splicing regulatory factor, in a case of CMML. ZRSR2 is located at Xp22.2 and the nonsense mutation was hemizygous in this male case (bone marrow).

Figure 2 Frequency and phenotypic association of spliceosomal mutations in myeloid malignancies.
(A) In the entire cohort (N=310), a total of 88 mutations in the spliceosome pathway components \textit{U2AF1}, \textit{SF3B1} and \textit{SRSF2} were observed in every subtype of myeloid malignancies, except for MPN. In low risk MDS, \textit{SF3B1} mutations were most frequent among the 3 genes. In particular, \textit{SF3B1} was mutated in 15/20 cases of RARS (60%). In the high risk MDS and AML group, \textit{U2AF1} mutations were most frequent (15/139; 10.8%). In the MDS/MPN group, \textit{SRSF2} was most frequently mutated (13/46; 28.2%), while \textit{SF3B1} is mutated at a high frequency in RARS-T (10/11; 90.1%). (B) Impact of spliceosomal mutations on clinical outcomes. In the entire cohort, patients with \textit{U2AF1} mutations (MT) had worse overall survival (OS), compared with wild type (WT), but \textit{SF3B1} mutations made OS significantly shorter. In low risk MDS, mutation of \textit{SF3B1} was a good prognostic factor but \textit{SRSF2} mutations are associated with worse prognosis. In MDS/MPN, patients with mutated \textit{U2AF1} had a shorter OS, but \textit{SF3B1} mutations were associated with significantly better prognosis. In addition, \textit{SRSF2} mutations did not affect outcomes.

\textbf{Figure 3 Unsplicing of specific genes due to spliceosomal mutations as detected by deep RNA sequencing.}

Next generation-based-RNA deep sequencing was used to quantitatively study splicing patterns. (A) The upper panel shows the intron 5 and exon 6 boundary of \textit{TET2} (dotted line). Five reads correspond to transcripts which were not spliced (unspliced; black circle) and 4 were spliced (white circle) at this boundary. The lower panel shows read counts at the 5’ and 3’ splice sites of each intron (3-10) of \textit{TET2}. White and black bars indicate the number of spliced and unspliced reads, respectively. In a case of AML with a \textit{U2AF1} mutation, more unspliced than spliced reads were observed at the 3’ splice site of intron 5 (left panel), likely due to a loss of spliceosome function. However, unspliced RNAs were less frequent than spliced RNAs in WT RNA sequencing (right panel). (B) At both the 3’ and 5’ splice sites of \textit{RUNX1} intron 6, unspliced reads were more frequent than spliced reads in AML cases with \textit{U2AF1} and \textit{SRSF2} mutations. However, there were fewer unspliced transcripts at the same site in WT and \textit{SF3B1} mutant samples. Splicing abnormalities in the selected genes are summarized (bottom right), including the results presented in detail in Supplemental Figures 7 and 8.
Figure 1

A  U2AF1 (Ch21q22.3)
Whole bone marrow
T-cell rich fraction

B  SF3B1 (Ch2q33.1)
Whole bone marrow
T-cell rich fraction

Isoform 1
Y623C  R625L  R625C
K666N  K666M  K666T
Q699,  K700 del

Isoform 2
E622D

For personal use only on July 16, 2017. For personal use only.
Figure 2

A

% of patients with mutation

0 25 50 75 100

0 25 50 75 100

0 25 50 75 100

SR SF2
SF3B1
U2AF1

Low risk MDS
High risk MDS/AML
MDS/MPN

For personal use only.
B

Whole cohort

U2AF1 WT vs MT
p=0.006

SF3B1 WT vs MT
p=0.015

Low risk MDS

SF3B1 WT vs MT
p=0.0921

SRSF2 WT vs MT
p=0.0245

MDS/MPN

U2AF1 WT vs MT
p=0.0005

SF3B1 WT vs MT
p=0.0047
Figure 3

A

TET2

Intron 5 Exon 6

U2AF1 mutant

U2AF1 wild type
Introns

**RUNX1**

**U2AF1 mutant**

**SF3B1 mutant**

**SRSF2 mutant**

Spliced genes

TET2

RUNX1

TP53

FECH

GAPDH

<table>
<thead>
<tr>
<th>Mutant genes</th>
<th>U2AF1</th>
<th>U2AF26</th>
<th>SRSF2</th>
<th>SF3B1</th>
</tr>
</thead>
<tbody>
<tr>
<td>TET2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RUNX1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP53</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FECH</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAPDH</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For personal use only. on July 16, 2017. For personal use only.
Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis

Hideki Makishima, Valeria Visconte, Hirotoshi Sakaguchi, Anna M. Jankowska, Sarah Abu Kar, Andres Jerez, Bartlomiej Przychodzen, Manoj Bupathi, Kathryn Guinta, Manuel G. Afable, Mikkail A. Sekeres, Richard A. Padgett, Ramon V. Tiu and Jaroslaw P. Maciejewski