Tools to study splicing factor mutations in MDS: Human HSC isolation and CRISPR/Cas9 genome editing

Student Name: Bennett Caughey

With significant thanks to Yan Song, Albert Perez, Sigrid Katz, Gene Yeo, and Rafael Bejar

ISP Committee Chair: Dr. Rafael Bejar

Abstract:

In recent years, genes encoding components of the RNA splicing machinery and regulatory factors have been characterized as a major class of mutated genes in myelodysplastic syndromes (MDS), with mutations occurring in a heterozygous gain-of-function pattern at recurrent amino acids. However, the effects on splicing patterns and consequent pathogenesis of malignancy remain poorly understood, in part due to challenges in isolating appropriate homogenous cellular populations carrying these mutations. We employed two complementary approaches to more effectively study splicing factors mutated in myeloid malignancies. We developed protocols to isolate bone marrow hematopoietic stem cells or peripheral blood progenitor cells from MDS patients known to carry splicing factor mutations. Our goal was to perform single-cell whole transcriptome RNA sequencing and compare differences in splicing between mutated and non-mutated cells from the same patient, as well as between patients. We also created genome-edited cell lines, using the CRISPR/Cas system to alter single DNA base pairs in cell lines to reproduce the mutations found in patients. We successfully generated K562 cell lines with mutations in SRSF2 and U2AF1, two of the most frequently mutated splicing factors. Importantly, these cell lines demonstrate the viability of this novel technology in hematopoietic cells as well as provide effective tools to study the molecular consequences of splicing factor mutations. Together, these approaches will allow better characterization of the role of splicing factors in the pathogenesis of myeloid malignancies.

Introduction

Myelodysplastic syndromes (MDS) are clinically heterogeneous clonal hematologic disorders, characterized by dysplastic myeloid cells, abnormal differentiation, and cytopenias. A significant number of MDS cases progress to acute myeloid leukemia (AML). In recent years, major efforts have been made to molecularly characterize MDS and other myeloid disorders. Bejar and others have worked to understand how mutations and cytogenetic abnormalities lead to disease phenotypes and predict prognosis and response to therapy. By sequencing 18 genes in over 400 MDS patient samples, he demonstrated that the mutational status of several genes correlated with clinical characteristics and improved our ability to predict patient prognosis.[1]

Soon thereafter, splicing factors were identified as a class of mutations common in myeloid disorders.[2] Splicing factors are proteins involved in processing immature RNA transcribed from DNA into mRNA via the removal of introns. Alternative splicing is a regulated process that allows multiple potential mRNAs (and therefore proteins) to be produced from the same initial transcript. Thus, changes in the cellular splicing machinery or its regulation

could have significant and widespread effects. Mutations of the splicing factor gene *SF3B1* were discovered to have a very strong association with the MDS subtype refractory anemia with ringed sideroblasts (RARS).[3] Shortly thereafter, Yoshida et al. reported mutations in several splicing factors in various myeloid neoplasms.[4] The most frequent mutations in MDS are in *SF3B1*, *U2AF1*, and *SRSF2*. Mutations in all of these genes exhibit characteristic gain-of-function patterns, with recurrent heterozygous mutations found at specific amino acid residues. Bejar et al. further explored splicing factor mutations and their clinical correlations in his MDS cohort, confirming that these mutations are mutually exclusive, suggesting a shared mechanism of action that confers a clonal advantage, despite some prognostic and clinical differences.[5] Figure 1 shows the distribution of common mutations in MDS; splicing factor mutations are both common and mutually exclusive.

Since the initial discovery of the importance of splicing factor mutations, there has been significant effort to characterize the mutational status of splicing factors in cohorts of myeloid and lymphoid malignancies. In addition, work has begun on functional studies and creating transgenic mice to understand the contribution of splicing factor mutations to MDS pathogenesis. For example, *TET2* is one of the most frequently mutated genes in MDS; a previous study identified abnormal splicing of *TET2* in whole bone marrow RNA analysis of patients with *U2AF1* splicing factor mutations[6]. Unfortunately, exploration of the impact of splicing factors in MDS has been challenging and has met with only limited success. Thus there is a significant need for novel approaches. In this study, we began to develop a method for studying these mutations *in vivo* via single cell analysis of the hematopoietic stem cells of patients with MDS. We also employed CRISPR/Cas to generate cell lines bearing splicing factor mutations for use in *in vitro* functional and splicing analysis.

Methods

Gene Editing with CRISPR/Cas9: The CRISPR/Cas9 system is a novel technique that allows the precise introduction of one or more mutations at a single nucleotide with high efficiency. Thus, one can introduce gene mutations that will then be expressed in more physiologic conditions. CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) RNAs and Cas are used in many species of bacteria to combat viral and plasmid invasion.[7] The Cas protein recognizes the CRISPR RNA structure. When that structure recognizes a genomic site, the Cas protein will cleave both strands of the DNA helix at that location (see Figure 2). This feature enables the design of unique genome identifiers that can target a large fraction of the genome. The double-stranded breaks are either repaired with Non-Homologous End Joining (NHEJ), which often introduces small insertions or deletions, or with Homologous Recombination (HR) using a DNA template. One can provide a template containing specific single or multi-nucleotide changes, resulting in a repaired genome containing the desired change.

First, potential sequences of the form $N_{21}GG$ were generated to target 4 genomic sites (SF3B1 K700, SRSF2 P95, U2AF1 S34, and U2AF1 Q157). The targeted residues are the most commonly mutated in these proteins in MDS. Putative sequences were then selected to avoid homology with genomic sites in order to minimize off target effects. 2 selected sequences per site (except SRSF2 P95) were then cloned into plasmids as part of a CRISPR guide RNA (gRNA), driven by a U6 promoter.

For each targeted residue, K562 cells were transiently transfected using electroporation with 3 plasmids expressing a gRNA, Cas9, and a puromycin resistance marker with GFP, in addition to 121 or 161 base pair oligonucleotides complementary to the targeted sequence and containing the desired single base pair change. The cells were grown under puromycin selection for three days, then diluted into single cells and grown for at least 2 weeks. At this point, a sample of DNA was taken and the colonies were screened for mutations at the target loci using a combination of Sanger and next-generation sequencing on a MiSeq (Illumina). Mutation positive colonies and controls were then selected for further growth.

HSC and Progenitor Cell Single Cell Analysis: HSCs reside primarily in the bone marrow and contain the disease propagating population in MDS patients.[8] Because these cells represent the crucial population for the pathogenesis of MDS, they are the ideal cells in which to analyze the relevant effects of mutations in splicing factors. Hematopoietic progenitor cells can be found in peripheral blood, which is more easily obtained, and may provide a viable alternative to isolating HSCs. A protocol was adapted to isolate these rare cells and perform whole transcriptome analysis.

First, mononuclear cells are enriched from bone marrow or peripheral blood using a Ficoll-Paque gradient. These cells are stained with fluorescent conjugated antibodies to CD34, CD38, and CD90 (for HSCs) or just CD34 and CD38 (for progenitors), as well as a cocktail of antibodies against various lineage markers of hematopoietic differentiation. These cells are then sorted using FACS, with the HSC population defined as lineage depleted $CD34^{+}CD38^{-}CD90^{+}$ cells (see Figure 3). This approach has been validated for isolating HSCs and disease propagating cells from MDS patients.[8] Similarly, lineage depleted $CD34^{+}CD38^{-}$ hematopoietic progenitor cells from peripheral blood were sorted using FACS (see Figure 4). These cells are then captured onto a C1 (Fluidigm) microfluidics chip as single cells, and prepped into RNA, then cDNA, according to the manufacturer's protocol. This cDNA from each cell is now available for sequencing and analysis for changes in RNA splicing and expression.

Results:

Single-Cell HSC isolation and RNA preparation

HSCs contain the disease propagating cells of MDS and the crucial changes in splicing and gene expression that confer a clonal advantage are relevant primarily to these primitive cells. Examining larger cell fractions (i.e. unfractionated bone marrow, or less restricted progenitors) may capture some abnormalities such as those seen in SF3B1 mutated MDS that gives rise to ringed sideroblasts. However, this observed phenotype may be an effect that is unrelated to the clonal advantage conferred by an SF3B1 mutation. Furthermore, HSCs are very rare in the bone marrow, making splicing changes specific to these cells more difficult to detect in larger cell fractions. Isolating these cells would allow us to examine the splicing changes most relevant to the pathogenesis of MDS.

To investigate the feasibility of this approach, lineage negative $CD34^{+}CD38^{-}CD90^{+}$ HSCs were isolated from the clinically indicated bone marrow aspirate of a patient with MDS harboring an SF3B1 K700E mutation (Figure 3). Approximately 3000 HSCs were collected;

of these 65 were captured on a C1 chip as single cells and prepped into cDNA. 41 of these cells resulted in high quality cDNA libraries available for sequencing and analysis. These results were encouraging for further exploration. However, this approach was limited by the rarity of patients with MDS with known splicing factor mutations undergoing bone marrow biopsies, even at a large referral center such as the Moore's Cancer Center.

Progenitor cell isolation

Due to difficulty identifying patients requiring bone marrow biopsies who also harbor the relevant mutations, we next explored a protocol to isolate progenitor cells from peripheral blood and validate data from HSCs. Bone marrow biopsies are much more invasive than phlebotomy, and patients with MDS provide peripheral blood much more frequently than they bone marrow biopsies. Peripheral blood was obtained from a patient with MDS who harbors an SRSF2 P95H mutation. Lineage depleted $CD34^{+}CD38^{-}$ cells were isolated using FACS (Figure 4). Unfortunately, despite large numbers of cells processed, progenitor cells were very rare and this approach was not considered practical for further exploration.

Generation of genome edited K562 cell lines

The CRISPR/Cas technology allows for targeted mutations in cells that allow more physiologic analysis of the phenotype associated with these mutations in hematopoietic cells. K562 cells were transfected with plasmids expressing CRISPR gRNAs, Cas9, and selection markers, as well as oligonucleotide templates as described in the methods. *SF3B1* K700, *SRSF2* P95 and *U2AF1* S34 and Q157 residues were targeted. 4/39 (10%) *SRSF2* P95 targeted colonies harbored a P95H mutation; 18/39 (48%) had evidence of CRISPR/Cas9 activity. A representative Sanger sequence demonstrating the presence of this mutation is shown is Figure 5. For one *U2AF1* S34 gRNA, 1/80 (1.25%) colonies harbored an S34F mutation; 25/80 (31%) had evidence of CRISPR/Cas9 activity. Other gRNAs showed no evidence of Cas9 activity, suggesting that the gRNAs were either ineffective or, potentially, toxic to cells such that none grew. These results demonstrate that CRISPR/Cas may be effective for generating models of splicing factor mutations in hematopoietic cell lines, but that not all mutations may be readily amenable to targeting by CRISPR. Further optimization of this strategy is needed.

Discussion:

This project explored two complementary approaches to study the contribution of splicing factor mutations the MDS disease pathogenesis. Importantly, we demonstrated that the Fluidigm C1 system can capture single cell HSCs and convert these to RNA and cDNA. Unfortunately, the capture efficiency is low for these cells, likely due to their small size. Large numbers of input cells are required for adequate capture, which further requires adequate bone marrow biopsies with sufficient cellularity, constraining an already limited supply of patients. Given their rarity, peripheral blood progenitor cells are unlikely to be an viable replacement to HSCs derived from the bone marrow. Efforts are needed to identify more patients who can undergo this process and to increase the HSC yield of the FACS sorting as well as the Fluidigm C1 capture step.

In addition, we showed that CRISPR/Cas can be employed in hematopoietic cell lines successfully. However, the efficiency of successful mutations was low in K562s despite

significant Cas9 activity, and for two residues no mutated colonies at all were obtained in the time span of the project. This result may be due in part to limitations of the specific gRNAs used or other components of the process. It may also be due to challenges posed by the cells (i.e. highly effective repair mechanisms) or toxicity caused by mutations in these genes. The latter explanation seems particularly likely, as splicing factors are core cellular machinery required for basic cell function and cells may tolerate mutations poorly.

Beyond improving the efficiency of the above protocols, we aim to employ these techniques to more effectively study splicing factor mutations. Future work would employ analysis of aberrant RNA splicing from both patient HSCs and *in vitro* cell lines. Patients and lines harboring unique mutations in separate splicing factors could be compared for overlap in splicing abberations to identify changes most likely to be crucial for disease pathogenesis. Mechanistic studies could also be performed on these cell lines. For example, mutations in *SF3B1* have a strong association with ringed sideroblasts; future studies could replicate this finding in erythroblastic cell lines such as TF1s harboring *SF3B1* mutations and explore the mechanism of this phenotype. In addition, mutated cell lines could serve as an effective platform to test and identify splicing factor inhibitors and drugs that target these mutations. It would also be interesting to employ CRISPR/Cas to explore the interaction of specific mutations in MDS by mutating multiple genes in the same cell to reproduce the malignant phenotype. These tools have potentially broad implications for improving our understanding of MDS and other myeloid malignancies as well as identifying novel therapeutic approaches.

Figures:

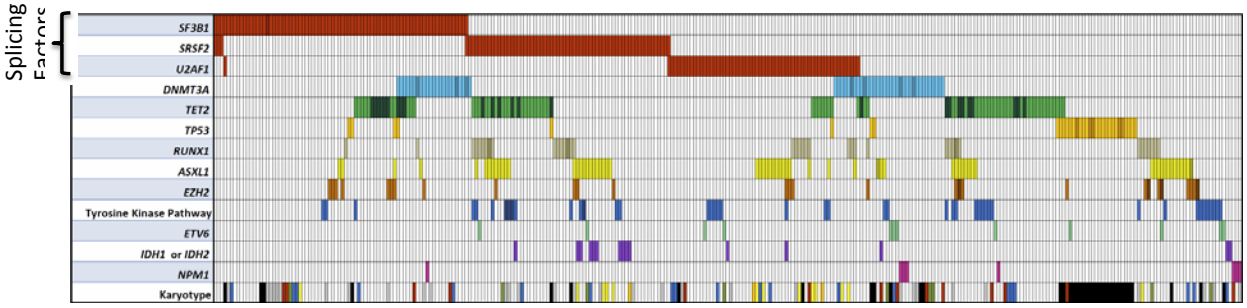


Figure 1: Frequent Mutations in MDS. The distribution and overlap of genes frequently mutated in MDS are shown. Mutations in splicing factors are shown in the first 3 rows; note that they are mutually exclusive with each other, but not with other genes. Data and figure courtesy of Dr. Rafael Bejar.[3,4] TK Pathway Genes = *NRAS*, *KRAS*, *BRAF*, *CBL*, *JAK2.*

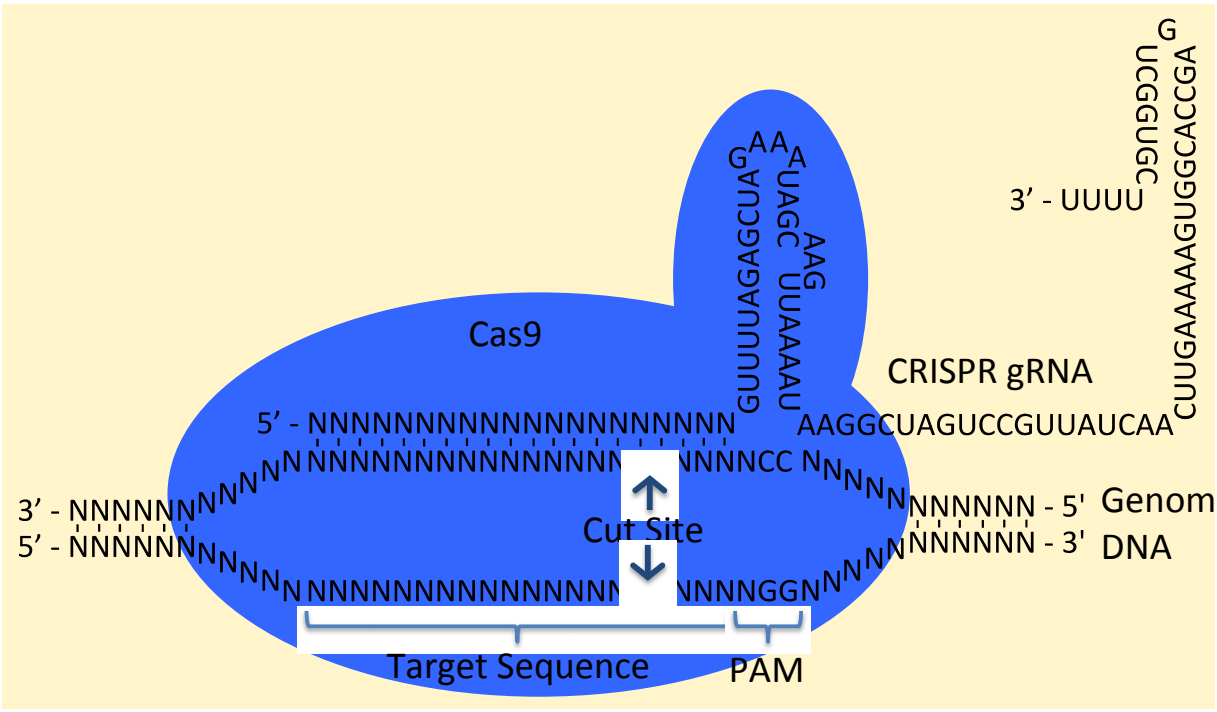


Figure 2: CRISPR/Cas genome editing schematic. The CRISPR gRNA and Cas9 complex recognizes a genomic site, causing a double-stranded break.

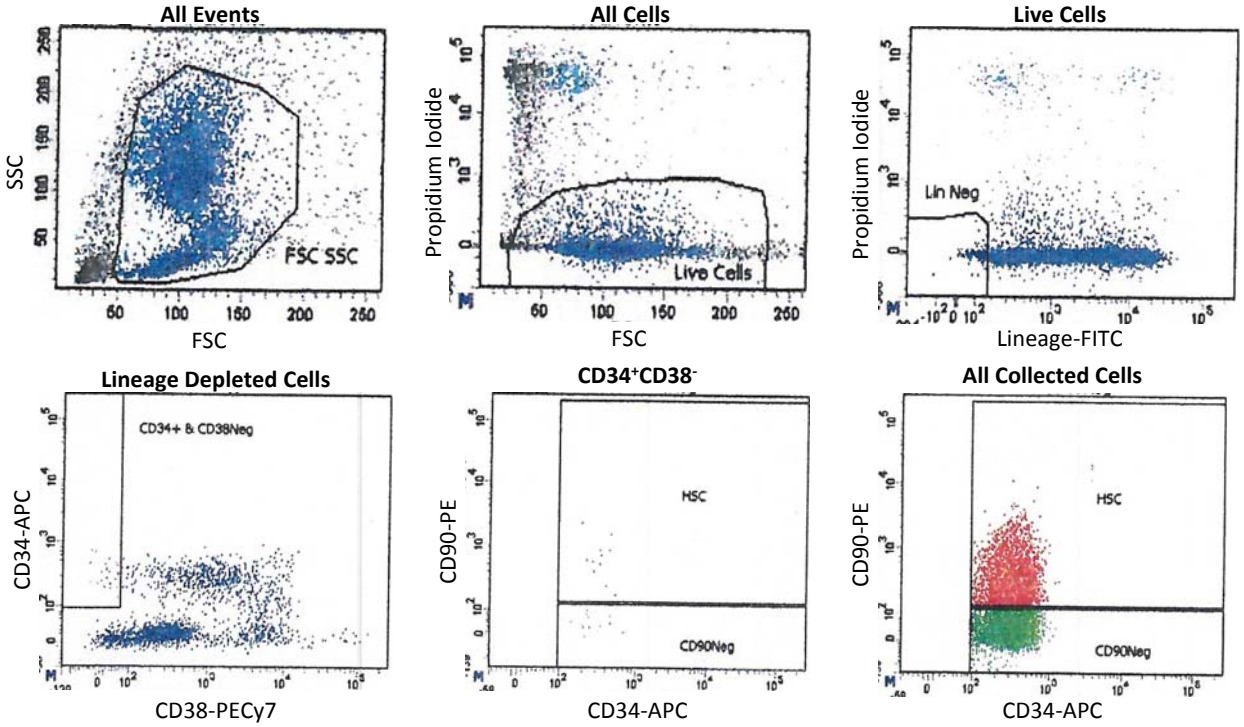


Figure 3: HSC FACS Isolation. A representative FACS plot of HSC sorting. HSCs, defined as lineage depleted $CD34^{+}CD38^{-}CD90^{+}$ bone marrow mononuclear cells, were isolated using the strategy shown. The final panel shows the CD90 expression of all captured cells.

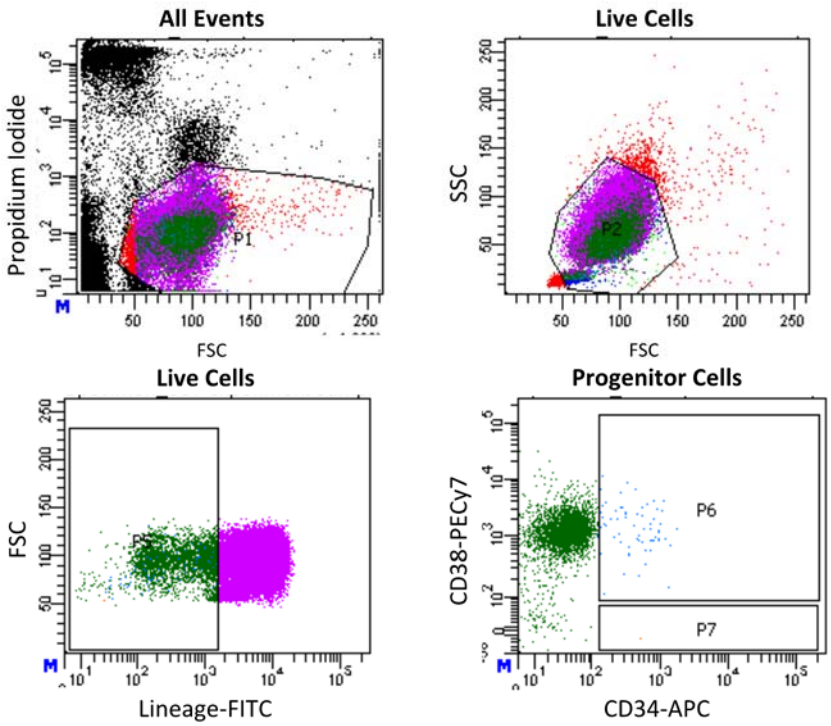
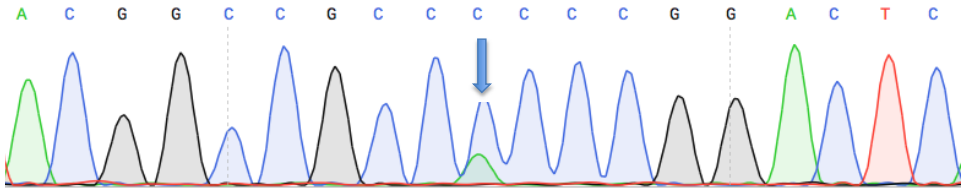


Figure 4: Progenitor Cell FACS Pilot. The strategy for isolation of progenitor cells from the blood of patients with MDS is shown. The $CD34^{+}CD38^{-}$ gate contains the most primitive progenitor cells.

A



B

CRISPR gRNA	Evidence of CRISPR/ Cas9 activity	Successful genome editing
SF3B1 K700 CI	0/54 colonies (0%)	0/54 (0%)
SF3B1 K700 CIII	0/17 (0%)	0/17 (0%)
SRSF2 P95 CI	18/39 (48%)	4/39 (10%)
U2AF1 S34 CIII	0/59 (0%)	0/59 (0%)
U2AF1 S34 CIV	25/80 (31%)	1/80 (1.25%)
U2AF1 Q157 CII	0/40 (0%)	0/40 (0%)
U2AF1 Q157 CIII	0/96 (0%)	0/96 (0%)

Figure 5: CRISPR/Cas genome editing results. A: A representative Sanger sequencing trace of a K562 colony with CRISPR/Cas targeting SRSF2 P95 is shown. The arrow indicates

the C->A mutation, resulting in an *SRSF2* P95H amino acid change. **B:** For each gRNA, the fraction of colonies with Cas9 activity and successful genome editing is shown.

References:

[1] Bejar R, Stevenson K, Abdel-Wahab O, et al. Clinical effect of point mutations in myelodysplastic syndromes. *N Engl J Med.* 2011; 364(26): 2496-2506

[2] Visconte V, Makishima H, Maciejewski J, Tiu R. Emerging roles of the splicesomal machinery in myelodysplastic syndromes and other hematologic disorders. *Leukemia.* 2012; 26: 2447-2454.

[3] Papaemmanuil E, Cazzola M, Boultwood J, et al. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med.* 2011; 365(15):1384-1395.

[4] Yoshida K, Sanada M, Shiraishi Y, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature.* 2011;478(7367):64-69.

[5] Bejar R, Stevenson K, Caughey B, et al. Validation of a Prognostic Model and the Impact of Mutations in Patients With Lower-Risk Myelodysplastic Syndromes. *J Clin Oncol.* 2012; 30:3376-3382.

[6] Makishima H, Visconte V, Sakaguchi H. Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenisis. *Blood.* 2012; 119(14): 3203-3210.

[7] Mali P, Yang L, Esvelt KM, et al. RNA guided human genome engineering via Cas9. *Science*, 2013; 339(6121):823-6.

[8] Tehranchi R, Woll PS, Anderson K, et al. Persistent malignant stem cells in del(5q) myelodysplasia in remission. *N Engl J Med.* 2010;363(11):1025-1037.