HHS Public Access

Author manuscript

Leuk Lymphoma. Author manuscript; available in PMC 2019 December 01.

Published in final edited form as:

Leuk Lymphoma. 2018 December ; 59(12): 2952–2962. doi:10.1080/10428194.2018.1452210.

Distinct Transcriptomic and Exomic Abnormalities within Myelodysplastic Syndrome Marrow Cells

Hogune Im, PhD[#1], **Varsha Rao, PhD**[#1], **Kunju Sridhar, PhD**[#2], **Jason Bentley, MScStat, MBiostat**[3], **Tejaswini Mishra, PhD**[1], **Rui Chen, PhD**[1], **Jeff Hall, PhD**[4], **Armin Graber, PhD**[4], **Yan Zhang, MD**[5], **Li Xiao, MD**[5], **George I. Mias, PhD**[6], **Michael P Snyder, PhD**[1], and **Peter L Greenberg, MD**[2]

[1]Department of Genetics, Stanford University School of Medicine, Stanford, California,

[2]Hematology Division, Stanford University School of Medicine, Stanford Cancer Institute, Stanford, California,

[3]Quantitative Science Unit, Stanford University, Stanford, CA,

[4]Genoptix Inc., Carlsbad, CA,

[5]Hematology, Jiaotong University, 6th Hospital, Shanghai, China,

[6]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan.

[#] These authors contributed equally to this work.

Abstract

To provide biologic insights into mechanisms underlying myelodysplastic syndromes (MDS) we evaluated the CD34+ marrow cells transcriptome using high-throughput RNA sequencing (RNA-Seq). We demonstrated significant differential gene expression profiles (GEPs) between MDS and normal and identified 41 disease classifier genes. Additionally, two main clusters of GEPs distinguished patients based on their major clinical features, particularly between those whose disease remained stable vs patients who transformed into acute myeloid leukemia within 12 months. The genes whose expression was associated with disease outcome were involved in functional pathways and biologic processes highly relevant for MDS. Combined with exomic analysis we identified differential isoform usage of genes in MDS mutational subgroups, with consequent dysregulation of distinct biological functions. This combination of clinical,

Corresponding author: Peter L Greenberg, peterg@stanford.edu, Address: Stanford Cancer Institute, 875 Blake Wilbur Drive, Stanford, CA 94305, Phone: 650-725-8355 Fax: 650-723-1269.

Authorship:

Contribution: HI, VR, KS, MPS and PLG designed the study; HI, VR, KS performed the experiments; HI, VR, KS, JMH, AG, MPS, JB and PLG analyzed and interpreted data; TM, RC, JMH, YZ, LX, GIM contributed to discussions and reviewed the manuscript; JMH and AG performed DNA sequencing and data retrieval; HI, VR and PLG and wrote the manuscript which was reviewed and edited by all authors.

Footnote:

An earlier version of the study data was presented at the 56th annual meeting of the American Society of Hematology (San Francisco, CA, December 6–9, 2014), abstract #1894, Blood 2014; 124 (21): 1894. After completion of our manuscript, a multi-center study that also describes the impact of altered gene expression in MDS using RNA-Seq technology has been reported online[52].

Supplementary Information accompanies this paper on the Lymphoma and Leukemia website

Conflict-of-interest disclosure: J.M.H. and A.G. are employees of Genoptix, Inc. The other authors declare no conflict of interest.

transcriptomic and exomic findings provides valuable understanding of mechanisms underlying MDS and its progression to a more aggressive stage and also facilitates prognostic characterization of MDS patients.

Introduction

The myelodysplastic syndromes (MDS) are a spectrum of clonal myeloid hemopathies with inherent hematopoietic precursor cell (HPC) anomalies, abnormal hematopoietic regulation and clinical heterogeneity.[1–3]. Use of these features has provided methods (e.g., the International Prognostic Scoring System [IPSS] and Revised IPSS [IPSS-R] and WPSS) to help define patients' prognoses, including their relative risk of developing acute myeloid leukemia (AML) and shortened survival.[3–5] However, despite their clinical utility additional relevant biologic and molecular data are needed to enhance the precision of these approaches for more accurately predicting patients' clinical courses and aid disease management.

Disease-specific gene expression profiles (GEPs) and cellular pathways have been identified using microarray platforms and have provided insights into the molecular biology of AML and its subtypes.[6–8] However, in contrast to the relatively homogeneous marrow population of blasts present in AML, the MDS marrow contains heterogeneous populations of cells with various degrees of cellular differentiation. Thus, for representative analysis of HPCs prior studies using microarray platforms from our lab and others have used CD34+ cells to assess GEP alterations in MDS marrow[9–13], in conjunction with clinical outcomes.[14,15] Given the increased sensitivity and accuracy of high-throughput RNA sequencing (RNA-Seq)[16–18] for detecting and quantifying mRNA transcripts, we applied this methodology for evaluation of differential gene expression between MDS and normal CD34+ marrow cells.

Studies using RNA-Seq have shown that >90% of human protein-coding genes produce multiple mRNA isoforms with subsequent analysis demonstrating that alternative splicing switches in tumors reveal novel signatures of cancer[19]. Distinct isoforms have specific roles in hematopoietic lineage generation.[20] The mechanisms for RNA isoform selection, i.e., alternative splicing and transcriptional start site usage, are often dysfunctional in diseases such as MDS and AML where aberrant isoform expression has been reported. [21,22] In this study we have additionally evaluated the differential isoforms usage within the transcriptomes of our MDS patient cohort.

Pathogenic mutations in genes, including those encoding transcription factors, epigenetic modifiers and the components of the spliceosome machinery have described the mutational landscape of MDS.[23–26] In our study, we have linked mutational analyses of the exome with the transcriptomic data of this patient cohort to describe the contributions of clinical, transcriptomic and exomic data that may be potentially valuable for assessing prognosis and pathogenetic features in MDS patients.

Materials/Subjects and Methods

Sample collection and clinical categorization of MDS patients

Marrow samples and clinical information from 44 primary MDS patients and 23 healthy control individuals were obtained and processed within one center (Stanford Cancer Center) with informed consent in accordance with the Declaration of Helsinki, with the approval of the Stanford Investigational Review Board. The healthy controls were either bone marrow transplantation donors or paid healthy volunteers. The MDS patients were categorized by their FAB, IPSS and NCCN classifications[2,3], thus >30% marrow blasts indicated AML transformation.

Isolation of cells and RNA; RNA amplification and library generation

RNA was extracted using the RNeasy kit (Qiagen, Valencia, CA) from magnetic bead affinity-enriched CD34+ cells (>94% by flow cytometry) obtained from marrow aspirates (Miltenyi Biotec, Auburn, CA) as previously described.[14] Total RNA was amplified by using the SMARTer Ultra Low RNA Kit for Illumina Sequencing (Clontech Laboratories, Inc., Mountain View, CA)[16] after testing for the fidelity of the protocol on HeLa cell derived RNA (Supplementary Methods, Suppl Figure 1). cDNA library synthesis and sequencing was performed as previously described (Supplementary Methods).[27] The libraries were sequenced on the Illumina HiSeq 2000 platform at the Stanford Sequencing Center. The sequencing data analyzed in this publication are deposited in NCBI's Gene Expression Omnibus (Accession GSE111085).

RNA-Seq data processing and analysis

For gene-level expression analysis, read mapping using default parameters in TopHat (**v2.0.9**) to the Human hg19 reference genome, read quantification and quality filtering were determined as previously described (**See** Supplementary Methods). Read mapping statistics are provided in Suppl Table 1. All statistical analyses unless otherwise stated were performed in R (**v3.1.3**) and packages from Bioconductor.[28] The term "significant" implies statistically significant at FDR 5%, unless otherwise stated throughout the text.

Raw read counts were obtained using HTseq (**v0.5.4p4**)[29] and differential expression analysis was performed using EdgeR (**v3.12.1**)[30],[31] and Limma (**v3.26.9**)[32]. After filtering the samples for genes with low counts across samples, 12,323 genes were used for subsequent evaluation. Batch correction was performed using the Surrogate Variable Analysis (sva) package (**v3.24.4**)[33]. Supervised hierarchical clustering analysis of the differentially expressed genes was performed as a part of the EdgeR package (**See** Supplementary Methods).

Classification

Classification analysis, based on the nearest shrunken centroid characterization to classify patients as MDS vs healthy, was performed using the prediction analysis of microarrays (PAM) – pamr package (**v2.23**)[34] adapted for analysis of RNA-Seq data. Based on this approach, the subset of genes that best classified each disease class was identified for all

MDS individuals and those whose disease transformed to AML within 12 months (tMDS) vs normals.

Mutation analysis, Sequencing Methods

Genomic DNA was extracted from frozen cell pellets obtained from CD34- bone marrow cells and sheared on a Covaris instrument (Covaris, Inc., Woburn, MA). Libraries were created using KAPA Hyper DNA Library preparation kits (Kapa Biosystems, Wilmington, MA). Selected exons and flanking sequences of 173 genes (Suppl Table 2) were enriched with target specific capture probes (Genoptix, Inc., Carlsbad, CA). Enriched libraries were sequenced on the Illumina NextSeq platform (Illumina, San Diego, CA). Genomic alterations were determined using target capture based next-generation sequencing (NGS) technology, see Supplementary Methods.[35]

Progression to AML within 12 months of marrow sample

For MDS patients univariable and multivariable Cox-regression analysis of time to AML within 12 months of bone-marrow sample was performed using the phreg procedure in SAS 9.4 (SAS Institute, Cary NC) to evaluate associations with clinical features (age, gender, IPSS), gene expression cluster, under or over-expression of genes (dichotomized as below or above median expression, respectively) identified from PAM analysis as well as the number of mutations (0–2 versus >2) for those MDS patients with a mutation analysis. Patients without progression to AML were right censored at the earliest of death, end of follow-up or 12 months. We reported hazard ratios with 95% profile-likelihood confidence intervals along with likelihood-ratio test p-values (a p-value of 0.05 was considered statistically significant) and used the exact method for ties.

Principal component analyses and multidimensional scaling analyses

To evaluate the relationship between the mutational subgroups i.e., with or without spliceosome mutations, principal component analysis was performed using the prcomp function and multidimensional scaling analysis was performed using the plotMDS function within the limma package.[32] (See Supplementary Methods)

Differential isoform usage analysis

For isoform-level analysis on the patients with mutational information, RNA-seq reads were mapped to the hg19 Human Reference Genomes using the STAR-aligner with exon-exon junction spanning settings (**v2.4.0g1**).[37] Differential Isoform Usage (DIU) per gene was evaluated between mutational subgroups using the Isoform Usage Two-step Analysis – IUTA (**v1.0**) package.[38,39] Differential isoform expression datasets between mutational subgroups were also generated with the Tuxedo suite (**v3.1.0**).[38]

Functional analysis

Biological functions, canonical pathways and known disease associations were annotated based on a) differentially expressed genes, b) differentially expressed isoforms and c) genes with DIU using the Qiagen Ingenuity Pathway Analysis (IPA) software which performs causal network analysis based on the Ingenuity Knowledge base and proprietary algorithms.

The statistical significance of the enriched pathways is calculated based on the Fisher exact test (QIAGEN Redwood City,www.qiagen.com/ingenuity)[40].

Comparative quantification of gene expression

The multiplexed probe-based method (nCounter® technology, NanoString Technologies, Seattle, WA) was used for gene expression validation.[41] For quantification, 1–5ng RNA isolated from snap frozen CD34+ MDS and normal marrow cells were used. Probe sets for each gene were designed and synthesized by nCounter®. Manufacturer's protocol was followed for sample preparation and hybridization.[42]

Results

Patient characterization

Clinical and gene expression cluster characteristics of the 44 MDS patients are indicated in Table 1 and Suppl Table 3. In addition to FAB, NCCN, IPSS and IPSS cytogenetic classifications, patients were specifically categorized for progression to AML within 12 months after sample collection (tMDS) (>30% marrow blasts) or whether they remained stable (sMDS). Details of the patients' clinical and molecular characteristics are shown in Suppl Table 3. Marrow samples were generally obtained within 6 months of diagnosis. Approximately half of the patients in the sMDS and tMDS groups received hypomethylating agents after the marrow sample was obtained. These features are also indicated in Figure 1. All tMDS patients who died, did so due to leukemia. As expected, an increased proportion of IPSS higher risk patients resided within the tMDS group (p=0.0003, Table 1).

Gene expression profiling using RNA-Seq

Differential expression analysis was performed on the RNA-seq data obtained from CD34+ marrow cells derived from either MDS or normal individuals. As shown in Figure 1 and Suppl Figure 2, 4148 differentially expressed genes (DEGs) were identified at 5% FDR, i.e., 2048 with elevated expression and 2100 with lower expression in MDS. Two major gene clusters were generated by supervised hierarchical clustering of the DEGs between the normal and MDS patients, with the top cluster (Cluster 1, 2048 genes) demonstrating elevated expression predominantly in the tMDS patients and lower expression in the normal and sMDS individuals (p=0.0079, Table 1, Figure 1). The opposite pattern was observed in the bottom cluster (Cluster 2, 2100 genes). sMDS individuals clustered closer to the healthy controls compared to those with tMDS. A higher proportion of tMDS patients also exhibited Cluster 1 DEG pattern (p=0.05, Table 1). More distinct cluster patterns were demonstrated when analyzing differential GEPs from tMDS patients vs normal (Suppl Figure 3). No clear segregation was noted between other clinical features, including hypomethylating-agent treatment.

Disease Classification

To identify genes that best distinguished MDS from normal we performed PAM on our DEG data and demonstrated that 41 significant genes were capable of distinguishing between the two groups (Suppl Figure 4A, Suppl Table 4A). PAM of tMDS vs normal identified a

minimum number of 46 genes significant for classification (Suppl Figure 4B, Suppl Table 4B).

Mutation analysis

DNA from CD34- marrow samples from 29 patients and 6 control subjects from our original cohort were evaluated for myeloid mutations. In 25 of the 29 patient samples, pathogenic (and presumed) somatic mutations were found across 23 pathogenic genes, the vast majority of which were previously reported to be frequently mutated in MDS patients (Figure 2, Suppl Table 3). Suppl Figure 5A shows the number of patients with specific pathogenic mutations and the potential of these patients to remain stable or transform to AML. Suppl Figure 5B and Suppl Table 3 demonstrates the positive association of mutational burden (ie, ≥2 vs < 2 mutations) with tMDS (p=0.0013, Fisher's exact test).

For these 29 patients, the heatmap clustering of DEGs vs normal was similar to that of the entire MDS (n=44) cohort, showed similar patient grouping between the two gene expression clusters (Suppl Figure 6), indicating that this subset was representative of the entire cohort. Among patients assigned to gene expression Cluster 1, 47% (9/19) had >2 mutations compared with 20% (2/10) of patients assigned to Cluster 2 (Fisher's exact test, p=0.23). Also, no clear association was seen between specific mutations and DEG Cluster (Suppl Table 3).

Association of clinical and biologic variables with progression to AML within 12 months of marrow sample

MDS patient subgroups within our cohort were analyzed to determine the impact of clinical and biologic features on progression to AML. PAM analyses showed that CD38 under-expression (values below median) was associated with an increased risk of progression to AML within 12 months (Hazard Ratio [HR]=2.63, 95% CI: 0.97, 8.27; p=0.057) among MDS (n=44) patients. High or intermediate-2 IPSS category (vs low or intermediate-1) was significantly associated with an increased risk of progression to AML (HR=8.49, 95% CI: 2.96, 30.5; p<0.001) while GEP cluster 1 (vs 2) showed a trend toward this association (HR=3.36, 95% CI: 0.95, 21.3; p=0.062) (Table 2). A multivariable Cox model with only IPSS (HR=8.69, 95% CI: 3.00, 31.4; p<0.001) and GEP cluster (HR=3.43, 95% CI: 0.96, 21.8; p=0.059) provided similar results to the univariable models. Among MDS patients with mutation analysis (n=29) (Table 2), GEP Cluster 1 (vs 2) was significantly associated with increased risk of progression to AML (HR=5.33, 95% CI: 1.02, 97.8; p=0.047), whereas having >2 mutations (vs 0–2) had a trend of association (HR=3.05, 95% CI: 0.92, 11.7; p=0.069).

Differential isoform usage in mutational subgroups

We identified two major DGE clusters between normal and MDS albeit closer examination showed subtle differences within each cluster. As these differences may have arisen due to differences at the transcript level we assessed aberrant isoform expression. In the group of individuals with mutation information (n=29), 25 had at least one pathogenic mutation in genes that regulate isoform expression (Figure 2), i.e. in the spliceosome, epigenetic factors or transcription factors.

Principal component and multidimensional scaling analysis showed distinct segregation of the patient subgroups with spliceosome mutations (SM) or with non-spliceosome mutations (NSM) even at the gene level (Suppl Fig 7A&B). The range of mutant allele frequency was between 37.5–100% for the individuals that expressed the mutant allele in the spliceosome (Supplementary Materials), suggesting the presence of dysfunctional components within the spliceosome machinery of our patient cohort. Transcript-level expression analyses identified numerous differentially expressed isoforms between normal individuals and MDS and its mutational subgroups (Suppl Tables 5&6).

The effect of dysfunctional spliceosome was specifically evaluated by estimating differential isoform usage (DIU). We identified distinct sets of genes that showed DIU between the two mutational subgroups and normal (Figure 3A, Suppl Tables 7 and 8). The NSM group was more similar to all MDS, with a greater number of genes showing DIU and higher overlap whereas the SM group had fewer genes showing DIU (Figure 3A). Eighteen genes showed DIU between the two mutation subgroups (Suppl Tables 7 & 8). The subset of genes that demonstrated DIU exclusively between SM and NSM are shown in Figure 3B.

Functional and pathway analysis of significantly differentially expressed genes

We performed functional and pathway enrichment analysis of the significantly DE genes and isoforms using the Ingenuity Pathway Analysis (IPA) software. Distinct biological functions and canonical pathways were enriched by Cluster 1 vs Cluster 2 DEGs for all MDS (Table 3A) and for tMDS patients (Table 3B).

Functional analyses of the genes with differential isoform expression and usage between the normal vs mutation subgroups demonstrated that each subgroup enriched distinct biological processes: the SM group enriched for pyrimidine biosynthesis pathways and replication and apoptosis, the NSM group enriched for metabolic and immune response pathways (Suppl Figure 9). IPA analysis of the differentially expressed isoforms using the Isoprofiler tool also identified specific isoforms that play a role in other cancers and diseases (Suppl Table 9).

Gene expression quantification and validation

Independent quantification of gene expression for a representative panels of genes using the probe-based nCounter® technology (NanoString) demonstrated good concordance of expression levels and directionality (Pearson correlation coefficient=0.809, $p<0.001$) with RNA-Seq (Suppl Figure 10, Suppl Methods).

Discussion

Our RNA-Seq data demonstrated differential gene expression between CD34+ marrow cells from MDS and normal individuals and identified two distinct gene clusters associated with disease outcome in MDS patients. GEPs of patients who transformed into AML within 12 months (tMDS) were predominantly associated with gene Cluster 1 whereas the GEPs of the more stable MDS (sMDS) patients and normal individuals were associated with gene Cluster 2 (Figure 1, Suppl Table 3, Suppl Figure 3). Numerous novel genes were discerned with these clusters (Suppl Table 4A,B). However, many genes, including the top 150 DEGs within our cohort, have been reported in prior MDS/AML microarray gene expression

studies: elevated expression - *CLK4, HOXB2, RAB27B, SLC15A2, DPYSL3* or lower expression - *AKAP12, ARPP21, MME, PMP22, PRG2, RAG2, TLR2, VPREB3* (Suppl Table 4).[12,14,43–45]

Functional categorization demonstrated deregulation of a number of relevant and disparate functional biologic pathways in MDS vs normal and between sMDS and tMDS. Genes with elevated expression in MDS (mainly Cluster 1) were involved in AML signaling, stem cell pluripotency, oxidative stress and immunologic activation related pathways, whereas those with lower expression in MDS (mainly Cluster 2) were engaged in DNA damage repair, cell cycle and checkpoint regulation, proliferative signaling, tumor suppressor and metabolism-related pathways (Table 3A). Similar pathway enrichment pattern was observed in the tMDS vs normal analyses in addition to lower enrichment of apoptosis-signaling pathways (Table 3B). Although novel altered pathways were discerned, some have previously been reported in MDS patients.[13,14,27,46]

Prediction analysis (PAM) identified 41 and 46 genes as disease classifiers between MDS vs Normal and tMDS vs Normal, respectively (Suppl Figure 4A&B). Of these, lower expression of *CD38* was significantly associated with time to AML. In contrast to its presence on CD34+ HPCs, CD38 expression is absent on the CD34+ hematopoietic stem cells, indicative of a more immature phenotype; findings also demonstrated in MDS.[47,48] As our analyzed cell population is CD34+, this finding supports existing reports that the CD34+CD38- cell subset has enhanced susceptibility to leukemic transforming events.[49]

The DGE clusters (Cluster 1 vs 2) and IPSS categories were associated with the patients' freedom from AML transformation within 12 months (Tables 1**and** 2). Exomic analysis in a subset of our patients detected mutations in genes known to be most frequently mutated in MDS[23–26,50] and resided in distinct functional categories (Figure 2). As previously shown[25], poorer overall outcome with a higher transformation rate occurred in our patients with ≥2 mutations. These findings support the hypothesis that the biological impact of altered gene expression as well as mutations contributes to the prognosis of MDS patients.[51]

Our RNA-Seq study extends prior microarray work[51] in which, using a multivariable model, mutations, gene expression and clinical diagnostic variables, were all found to contain relevant information for predicting clinical outcomes, albeit their analysis was for survival, whereas ours assessed time to AML evolution. Using AML evolution rather than survival as an endpoint may be more relevant for evaluating clinical outcomes, given the multiple non-hematologic causes of mortality in MDS patients. Recently, a multi-institutional study using RNA-Seq also reported distinct DGE patterns predictive of AML transformation[52].

Despite the presence of multiple gene mutations and clinical variables only two major DGE clusters were generated. However, closer observation revealed additional DGE sub-clusters (Figure 1). These subtle changes in overall gene-expression can potentially occur due to differential isoform expression or usage. RNA and protein isoforms play an important role in determining gene function in hematopoiesis and may influence disease outcome in MDS and

AML.[20,53] Alternative splicing (predominantly) and modulation of transcriptional activity regulate spatio-temporal isoform expression.[54–56] The genes that modulate these mechanisms are frequently mutated or have altered function in MDS[54]. Aberrant isoform expression occurs due to defects in spliceosome gene expression or function.[57] Thus, we also evaluated differential isoform expression for these patients, comparing normal individuals to subgroups with spliceosome mutations (SM) (i.e. in SRSF2, SF3B1 and U2AF1) and those without (NSM). Gene level differential expression showed distinct clustering between the two mutational subgroups (p-value <0.01) (Suppl Figure 7). At the transcript level we identified a substantial number of differentially expressed isoforms between normals and the mutational subgroups (FDR 5%, Suppl Table 5&6). Moreover, differential isoform usage (DIU) was observed in distinct sets of genes between the mutational subgroups and normals (Figure 3A & B, Suppl Figure 8, Suppl Table 7). Previous differential isoform expression studies in MDS have either focused on specific genes[58,59] or on the effects of specific spliceosome mutations.[60,61] However, our study demonstrated that the commonly mutated spliceosome genes promoted distinct global transcriptomic changes in isoform selection/usage in comparison to the non-spliceosome genes. This alteration could consequently contribute to unique manifestations of the MDS phenotype within this subgroup, with ramifications for clinical outcome and disease treatment. Our study confirms and extends a recent report that also demonstrated MDS-linked splice gene signatures were associated with converging cancer-related pathways.[61] It should be noted that the individuals in the SM group often had mutations in additional determinants of isoform expression i.e., in transcription factors and epigenetic modifiers (Figure 2).

In our global gene expression profiling and functional pathway analysis of marrow cells of MDS vs normal, these transcriptomic and exomic studies demonstrated that, despite having multiple gene mutations and heterogeneous clinical features, only two distinctive marrow CD34+ marrow cell GE clusters in MDS occurred and were also associated with patients' clinical outcomes. In addition, we showed mutation group-specific altered proportions of isoform expression in MDS and that MDS mutational subgroups (SM vs NSM) were associated with differential isoform usage (DIU) patterns with consequent dysregulated biological functions.

This study combines transcriptomic and exomic data to identify key contributors to AML progression. These findings can be used to determine outcome of the disease in different MDS clinical (s/tMDS) or mutational (NSM/SM) subgroups. Prediction analyses have shown that a subset of DE genes (i.e., 41 between all MDS vs normal and 46 between tMDS vs normal) can be used as biomarkers for prognostically classifying MDS patients (i.e., AML progression within 12 months). The exomic and isoform analyses show that differences in underlying mutations dysregulate distinct biological functions. Together this information can help in real-life assessment of MDS patient prognosis and thus possibly alter the timing and treatment modalities recommended for the patient mutational subgroups.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by the Stanford MDS Center Research Fund, the William E. Walsh Leukemia Research Fund, the Eugene, Elizabeth and Christine Cronkite Hematology Fund and the Leukemia and Lymphoma Society MDS SCOR grant 38363. This work used the Genome Sequencing Service Center by Stanford Center for Genomics and Personalized Medicine Sequencing Center, supported by the grant award NIH S10OD02014. G.I.M. is supported by the National Human Genome Research Institute of the National Institutes of Health under Award Number R00 HG007065. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Support Disclosure: This work was supported by the Stanford MDS Center Research Fund, the William E. Walsh Leukemia Research Fund, the Eugene, Elizabeth and Christine Cronkite Hematology Fund and the Leukemia and Lymphoma Society MDS SCOR grant 38363. G.I.M. is supported by the National Human Genome Research Institute of the National Institutes of Health under Award Number R00 HG007065. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Greenberg PL. The multifaceted nature of myelodysplastic syndromes: clinical, molecular, and biological prognostic features. J Natl Compr Canc Netw 2013;11:877–884. [PubMed: 23847221]
2. Greenberg PL, Stone RM, Bejar R, et al. Myelodysplastic syndromes, version 2.2015. J Natl Compr Canc Netw 2015;13:261–272. [PubMed: 25736003]
3. Greenberg P, Cox C, LeBeau MM, et al. . International scoring system for evaluating prognosis in myelodysplastic syndromes. Blood 1997;89:2079–2088. [PubMed: 9058730]
4. Greenberg PL, Tuechler H, Schanz J, et al. . Revised international prognostic scoring system for myelodysplastic syndromes. Blood 2012;120:2454–2465. [PubMed: 22740453]
5. Della Porta MG, Tuechler H, Malcovati L, et al. . Validation of WHO classification-based Prognostic Scoring System (WPSS) for myelodysplastic syndromes and comparison with the revised International Prognostic Scoring System (IPSS-R). A study of the International Working Group for Prognosis in Myelodysplasia (IWG-PM). Leukemia 2015;29:1502–1513. [PubMed: 25721895]
6. Ebert BL, Golub TR. Genomic approaches to hematologic malignancies. Blood 2004;104:923–932. [PubMed: 15155462]
7. Bullinger L, Dohner K, Bair E, et al. . Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. N Engl J Med 2004;350:1605–1616. [PubMed: 15084693]
8. Marcucci G, Maharry K, Radmacher MD, et al. . Prognostic significance of, and gene and microRNA expression signatures associated with, CEBPA mutations in cytogenetically normal acute myeloid leukemia with high-risk molecular features: a Cancer and Leukemia Group B Study. J Clin Oncol 2008;26:5078–5087. [PubMed: 18809607]
9. Miyazato A, Ueno S, Ohmine K, et al. . Identification of myelodysplastic syndrome-specific genes by DNA microarray analysis with purified hematopoietic stem cell fraction. Blood 2001;98:422–427. [PubMed: 11435312]
10. Hofmann WK, de Vos S, Komor M, Hoelzer D, Wachsman W, Koeffler HP. Characterization of gene expression of CD34+ cells from normal and myelodysplastic bone marrow. Blood 2002;100:3553–3560. [PubMed: 12411319]
11. Ueda M, Ota J, Yamashita Y, et al. . DNA microarray analysis of stage progression mechanism in myelodysplastic syndrome. Br J Haematol 2003;123:288–296. [PubMed: 14531911]
12. Pellagatti A, Cazzola M, Giagounidis AA, et al. . Gene expression profiles of CD34+ cells in myelodysplastic syndromes: involvement of interferon-stimulated genes and correlation to FAB subtype and karyotype. Blood 2006;108:337–345. [PubMed: 16527891]
13. Pellagatti A, Cazzola M, Giagounidis A, et al. . Deregulated gene expression pathways in myelodysplastic syndrome hematopoietic stem cells. Leukemia 2010;24:756–764. [PubMed: 20220779]

14. Sridhar K, Ross DT, Tibshirani R, Butte AJ, Greenberg PL. Relationship of differential gene expression profiles in CD34+ myelodysplastic syndrome marrow cells to disease subtype and progression. Blood 2009;114:4847–4858. [PubMed: 19801443]
15. Pellagatti A, Benner A, Mills KI, et al. . Identification of gene expression-based prognostic markers in the hematopoietic stem cells of patients with myelodysplastic syndromes. J Clin Oncol 2013;31:3557–3564. [PubMed: 24002510]
16. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 2008;5:621–628. [PubMed: 18516045]
17. Pepke S, Wold B, Mortazavi A. Computation for ChIP-seq and RNA-seq studies. Nat Methods 2009;6:S22–32. [PubMed: 19844228]
18. Soon WW, Hariharan M, Snyder MP. High-throughput sequencing for biology and medicine. Mol Syst Biol 2013;9:640. [PubMed: 23340846]
19. Joshi P, Halene S, Abdel-Wahab O. How do messenger RNA splicing alterations drive myelodysplasia? Blood 2017;129:2465–2470. [PubMed: 28348147]
20. Grech G, Pollacco J, Portelli M, et al. . Expression of different functional isoforms in haematopoiesis. International Journal of Hematology 2014;99:4–11. [PubMed: 24293279]
21. Makishima H, Visconte V, Sakaguchi H, et al. . Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. Blood 2012;119:3203–3210. [PubMed: 22323480]
22. Crews LA, Balaian L, Delos Santos NP, et al. . RNA Splicing Modulation Selectively Impairs Leukemia Stem Cell Maintenance in Secondary Human AML. Cell Stem Cell 2016;19:599–612. [PubMed: 27570067]
23. Bejar R, Stevenson K, Abdel-Wahab O, et al. . Clinical effect of point mutations in myelodysplastic syndromes. N Engl J Med 2011;364:2496–2506. [PubMed: 21714648]
24. Yoshida K, Sanada M, Shiraishi Y, et al. . Frequent pathway mutations of splicing machinery in myelodysplasia. Nature 2011;478:64–69. [PubMed: 21909114]
25. Papaemmanuil E, Gerstung M, Malcovati L, et al. . Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood 2013;122:3616–3627; quiz 3699. [PubMed: 24030381]
26. Haferlach T, Nagata Y, Grossmann V, et al. . Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia 2014;28:241–247. [PubMed: 24220272]
27. Im H, Rao V, Sridhar KJ, et al. . Transcriptomic Evaluation of CD34+ Marrow Cells from Myelodysplastic Syndrome (MDS) Patients. Blood 2014;124:1894. [PubMed: 25122610]
28. R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing 2015.
29. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 2015;31:166–169. [PubMed: 25260700]
30. Nikolayeva O, Robinson MD. edgeR for differential RNA-seq and ChIP-seq analysis: an application to stem cell biology. Methods Mol Biol 2014;1150:45–79. [PubMed: 24743990]
31. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139–140. [PubMed: 19910308]
32. Ritchie ME, Phipson B, Wu D, et al. . limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47. [PubMed: 25605792]
33. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet 2007;3:1724–1735. [PubMed: 17907809]
34. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A 2002;99:6567–6572. [PubMed: 12011421]
35. Kwok B, Hall JM, Witte JS, et al. . MDS-associated somatic mutations and clonal hematopoiesis are common in idiopathic cytopenias of undetermined significance. Blood 2015;126:2355–2361. [PubMed: 26429975]
36. Irizarry Rafael A I LM. Rafalib: Convenience Functions for Routine Data Exploration. . 2015;R package version 1.0.0.

37. Dobin A, Davis CA, Schlesinger F, et al. . STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15–21. [PubMed: 23104886]
38. Trapnell C, Roberts A, Goff L, et al. . Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 2012;7:562–578. [PubMed: 22383036]
39. Niu L, Huang W, Umbach DM, Li L. IUTA: a tool for effectively detecting differential isoform usage from RNA-Seq data. BMC Genomics 2014;15:862. [PubMed: 25283306]
40. Kramer A, Green J, Pollard J, Jr., Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics 2014;30:523–530. [PubMed: 24336805]
41. Reis PP, Waldron L, Goswami RS, et al. . mRNA transcript quantification in archival samples using multiplexed, color-coded probes. BMC Biotechnol 2011;11:46. [PubMed: 21549012]
42. Vukmirovic M, Herazo-Maya JD, Blackmon J, et al. . Identification and validation of differentially expressed transcripts by RNA-sequencing of formalin-fixed, paraffin-embedded (FFPE) lung tissue from patients with Idiopathic Pulmonary Fibrosis. BMC Pulm Med 2017;17:15. [PubMed: 28081703]
43. Boultwood J, Pellagatti A, Cattan H, et al. . Gene expression profiling of CD34+ cells in patients with the 5q- syndrome. Br J Haematol 2007;139:578–589. [PubMed: 17916100]
44. Majeti R, Becker MW, Tian Q, et al. . Dysregulated gene expression networks in human acute myelogenous leukemia stem cells. Proc Natl Acad Sci U S A 2009;106:3396–3401. [PubMed: 19218430]
45. Ng SW, Mitchell A, Kennedy JA, et al. . A 17-gene stemness score for rapid determination of risk in acute leukaemia. Nature 2016;540:433–437. [PubMed: 27926740]
46. Mills KI, Kohlmann A, Williams PM, et al. . Microarray-based classifiers and prognosis models identify subgroups with distinct clinical outcomes and high risk of AML transformation of myelodysplastic syndrome. Blood 2009;114:1063–1072. [PubMed: 19443663]
47. Notta F, Doulatov S, Laurenti E, Poeppl A, Jurisica I, Dick JE. Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. Science 2011;333:218–221. [PubMed: 21737740]
48. Woll PS, Kjallquist U, Chowdhury O, et al. . Myelodysplastic syndromes are propagated by rare and distinct human cancer stem cells in vivo. Cancer Cell 2014;25:794–808. [PubMed: 24835589]
49. Jan M, Snyder TM, Corces-Zimmerman MR, et al. . Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. Sci Transl Med 2012;4:149ra118.
50. Walter MJ, Shen D, Ding L, et al. . Clonal architecture of secondary acute myeloid leukemia. N Engl J Med 2012;366:1090–1098. [PubMed: 22417201]
51. Gerstung M, Pellagatti A, Malcovati L, et al. . Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes. Nat Commun 2015;6:5901. [PubMed: 25574665]
52. Shiozawa Y, Malcovati L, Galli A, et al. . Gene expression and risk of leukemic transformation in myelodysplasia. Blood 2017 [11 2]. 10.1182/blood-2017-05-783050
53. Kramarzova K, Stuchly J, Willasch A, et al. . Real-time PCR quantification of major Wilms' tumor gene 1 (WT1) isoforms in acute myeloid leukemia, their characteristic expression patterns and possible functional consequences. Leukemia 2012;26:2086–2095. [PubMed: 22430637]
54. Yip BH, Dolatshad H, Roy S, Pellagatti A, Boultwood J. Impact of Splicing Factor Mutations on Pre-mRNA Splicing in the Myelodysplastic Syndromes. Curr Pharm Des 2016;22:2333–2344. [PubMed: 26916023]
55. Grech G, Pollacco J, Portelli M, et al. . Expression of different functional isoforms in haematopoiesis. Int J Hematol 2014;99:4–11. [PubMed: 24293279]
56. Wang X, Hou J, Quedenau C, Chen W. Pervasive isoform-specific translational regulation via alternative transcription start sites in mammals. Mol Syst Biol 2016;12:875. [PubMed: 27430939]
57. Lee Y, Rio DC. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annu Rev Biochem 2015;84:291–323. [PubMed: 25784052]
58. Sakurai H, Harada Y, Ogata Y, et al. . Overexpression of RUNX1 short isoform has an important role in the development of myelodysplastic/myeloproliferative neoplasms. Blood Advances 2017;1:1382–1386. [PubMed: 29296779]

59. Hopfer O, Komor M, Koehler IS, et al. . Aberrant promotor methylation in MDS hematopoietic cells during in vitro lineage specific differentiation is differently associated with DNMT isoforms. Leuk Res 2009;33:434–442. [PubMed: 18829110]

60. Dolatshad H, Pellagatti A, Fernandez-Mercado M, et al. . Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells. Leukemia 2015;29:1798. [PubMed: 26242354]

61. Qiu J, Zhou B, Thol F, et al. . Distinct splicing signatures affect converged pathways in myelodysplastic syndrome patients carrying mutations in different splicing regulators. RNA 2016;22:1535–1549. [PubMed: 27492256]

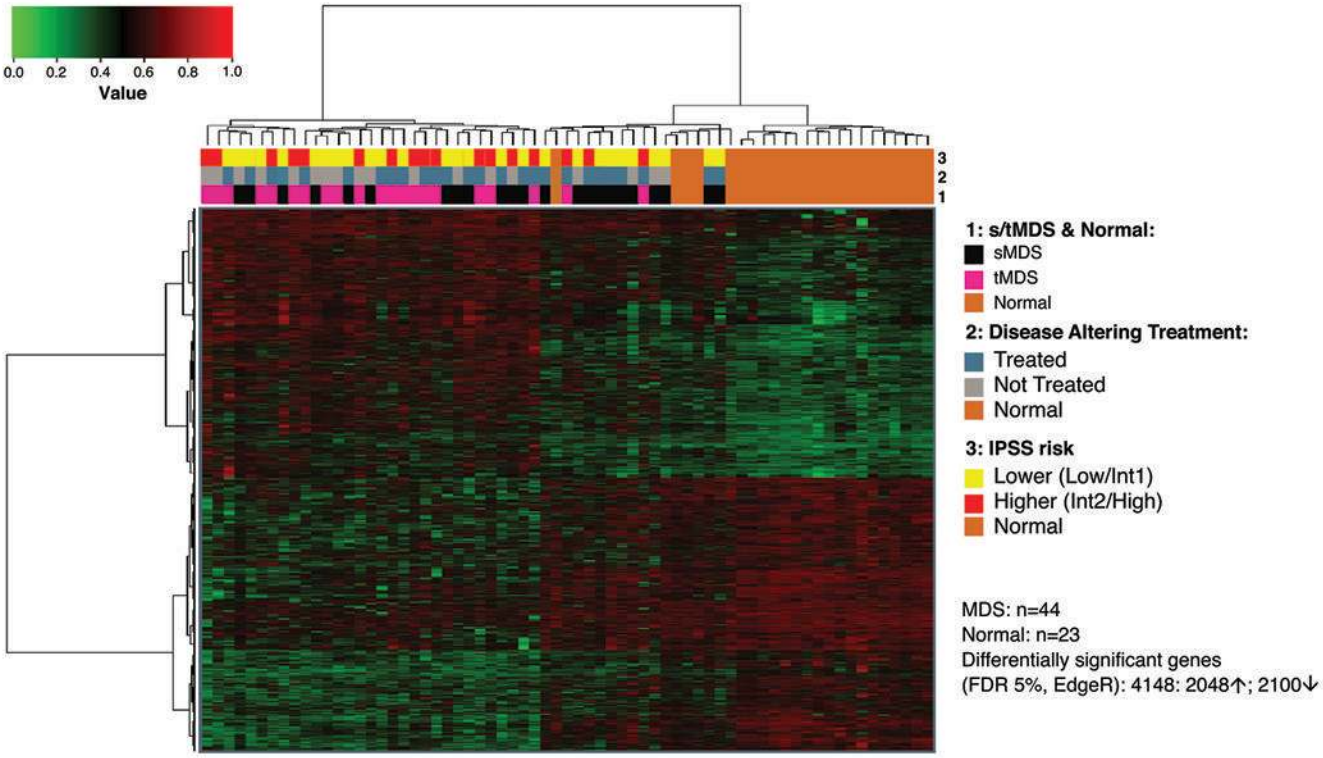

Figure 1. Gene expression profiles of MDS vs Normal CD34+ marrow cells.

Hierarchical cluster dendrogram based on differential expression analysis of RNA-Seq generated distinct clusters of genes between Normal (n=23) and MDS (n=44), sMDS (n=23) patients being more closely associated with Normal than tMDS (n=21) at 5% FDR. Two major gene clusters were noted between the normal and MDS patients, with the top cluster (Cluster 1, 2048 genes) demonstrating elevated expression predominantly in the tMDS patients and lower expression in the normal and sMDS individuals (p=0.0079, Table 1, Figure 1). The opposite pattern was observed in the bottom cluster (Cluster 2, 2100 genes). The clinical and cytogenetic characteristics of these patients are indicated as well as whether they subsequently developed AML within 12 months (tMDS, pink) or remained stable (sMDS, black). Normal individuals are indicated by brown. In the heatmap, red indicates elevated expression, green lower expression.

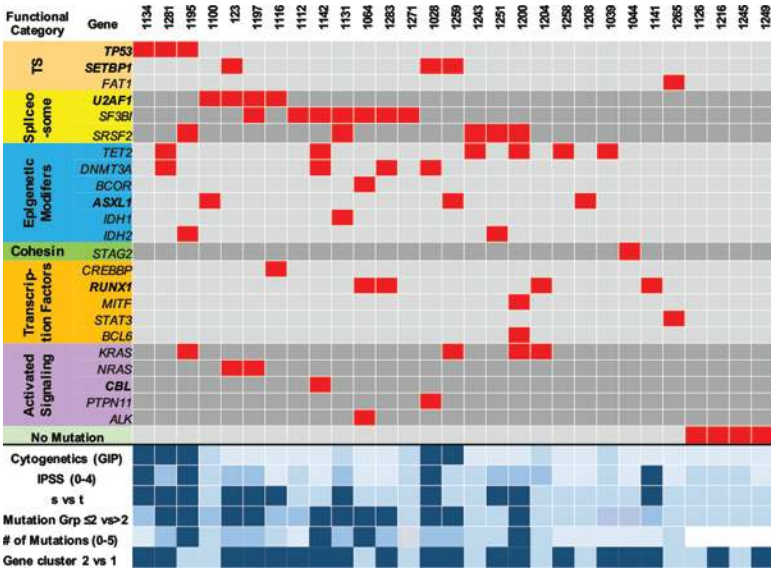

Figure 2. Pathogenic mutation patterns in MDS patients (n=29) for whom gene expression (GE) analysis was also performed.

Mutated genes (n=23) within differing functional groups are shown as well as associated clinical features for the specific patients. Isolated as well as co-expressed mutations are shown amongst the patients. Blue boxes below the mutation chart, with darker blue indicating worse prognosis, demonstrate: Cytogenetic groups [IPSS Good (G), Intermediate (I) Poor (P)], IPSS clinical category 1–4 (Low, Int-1, Int-2, High), AML transformation status (s/tMDS], Mutation number (0–5), gene expression cluster (1 or 2). TS = tumor suppressor.

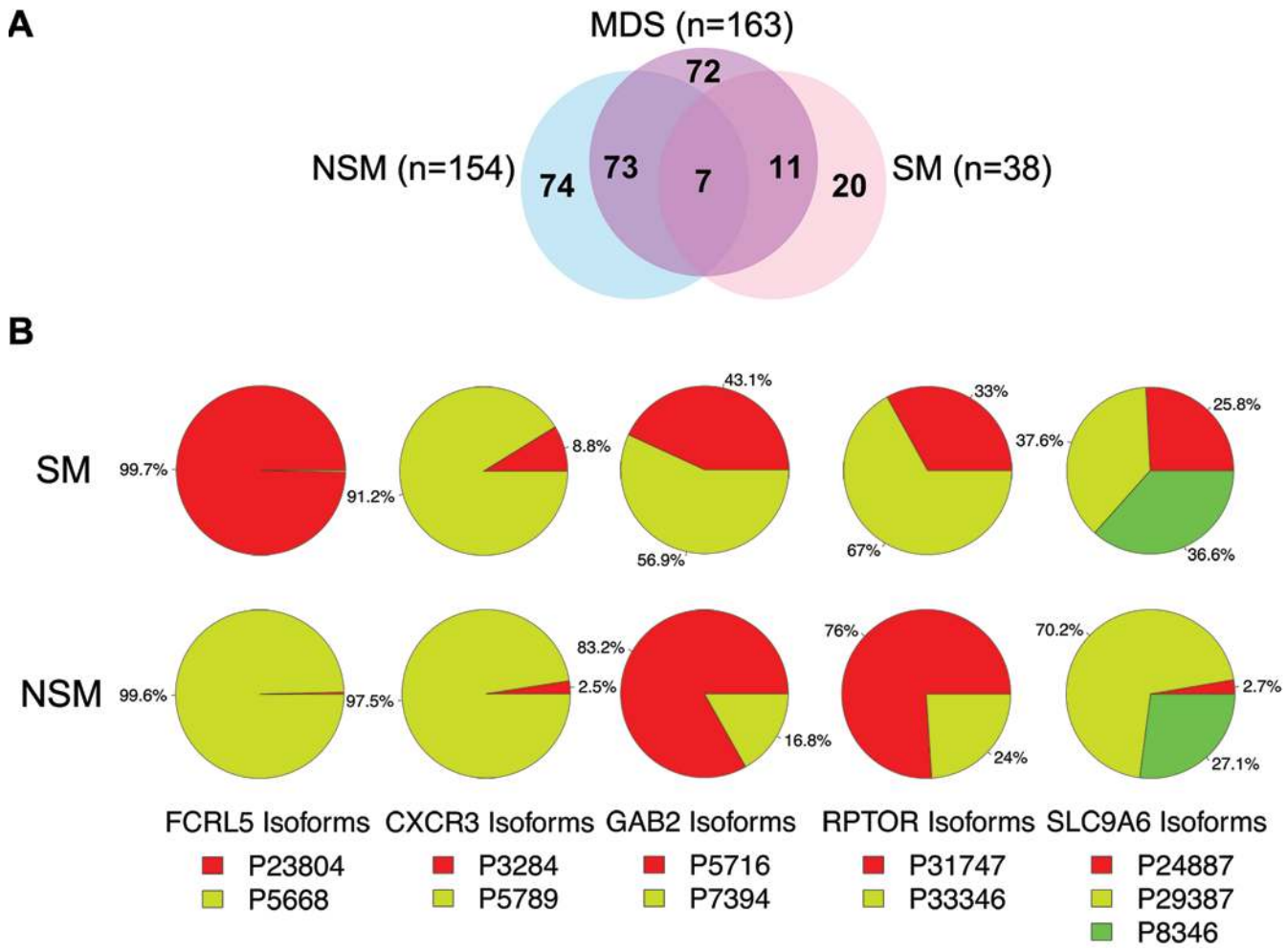

Figure 3. Differential Isoform Usage (DIU) in mutational sub-groups.

(A) Overlap of genes with DIU in pairwise comparison of Normal vs All MDS, Normal vs SM and Normal vs NSM groups, the total number of showing significant DIU are given in parentheses and listed in Suppl Table 7. (B) Pie charts showing the comparative isoform usage estimates for a subset of genes (n=5) with significant DIU exclusively between SM and NSM groups. SM = spliceosome mutations, NSM = Non-spliceosome mutations.

Table 1.

Clinical Characteristics of MDS patients

Clinical Features	Total Patients	Stable MDS (sMDS)	Transforming MDS (tMDS)*	sMDS vs tMDS, p value
Patient Number	44	27	17	
Age, years, median (range)	69 (42–85)	69.5 (50–85)	69 (42–80)	0.50[a]
Males, n (%)	25 (57)	15 (56)	10 (59)	1[b]
Cytogenetic risk, n(%)				0.65[b]
Good	29 (66)	19 (70)	10 (59)	
Intermediate/Poor	15 (34)	8 (30)	7 (41)	
IPSS status:				0.0003[b]
Lower risk: Low/Intermediate1	26 (59)	22 (82)	4 (24)	
Higher risk: Intermediate2/High	18 (41)	5 (18)	13 (76)	
Treated, n/N (%) ***				
Lower risk: Low/Intermediate1	15/25 (60)	12/21 (57)	3/4 (75)	0.63[b]
Higher risk: Intermediate2/High	10/19 (53)	2/6 (33)	9/13 (69)	0.32[b]
Gene Expression Cluster^, n (%):				0.05[b]
Cluster 1	31 (71)	16 (59)	15 (88)	
Cluster 2	13 (29)	11 (41)	2 (12)	
Cluster 1 vs Cluster 2, n_1/n_2 (%n_1):				
Lower risk pts	17/9 (65)	13/9 (59)	4/0 (100)	0.26[b]
Higher risk pts	14/4 (78)	3/2 (60)	11/2 (85)	0.53[b]

*AML transformation within 12 months,

**IPSS Risk cytogenetics for MDS3,

***Treated with hypomethylating agents,

^See Figure 1

[a]T-test

[b]Fisher's exact text

Table 2:

Association of clinical and biologic variables with Time to AML within 12 months of marrow sample[1]

Variable	HR (95% CI)[2]	LRT p-value[3]
All MDS patients (n=44)		
Age, years	0.98 (0.94, 1.03)	0.46
Gender, Male	1.19 (0.46, 3.29)	0.72
IPSS Category		<.0001
Low/int-1	1 (reference)	
High/Int-2	8.49 (2.96, 30.5)	
Gene expression cluster		0.062
1	3.36 (0.95, 21.3)	
2	1 (reference)	
Mutation analysis patients (n=29)		
IPSS Category		0.0047
Low/int-1	1 (reference)	
High/Int-2	5.87 (1.73, 22.8)	
Gene expression cluster		0.047
1	5.33 (1.02, 97.8)	
2	1 (reference)	
Number of mutations		0.069
0–2	1 (reference)	
>2	3.05 (0.92, 11.7)	

1. Cox-regression using the exact method for ties.

2. Hazard Ratio with 95% profile-likelihood confidence intervals.

3. LRT= Likelihood ratio test p-value.

Table 3.

Functional Pathways Associated with Differentially Expressed Gene Clusters in MDS vs Normal

Table 3A. MDS vs Normal

Biological Pathways	p-value
ELEVATED EXPRESSION IN MDS (Cluster 1)	
Stem cell-related:	
Role of NANOG in Mammalian Embryonic Stem Cell Pluripotency	0.01
Embryonic Stem Cell Pluripotency	0.003
Oxidative Stress Response:	
NRF-2	0.03
Immunologic:	
TNFR signaling	0.04
AML signaling- related:	
AML	0.04
Ephrin signaling	0.04
PTEN	0.05
LOWER EXPRESSION IN MDS (Cluster 2)	
Cell cycle regulation:	
Cyclins	0.0001
Mitosis	4.17E-05
Chromosomal Replication	3.63E-09
Checkpoint Regulation	0.008
DNA Damage:	
Mismatch Repair in Eukaryotes	2.04E-08
Role of BRCA1 in DNA Damage Response	2.04E-07
Nucleotide Excision Repair Pathway	0.005
Metabolism-related:	
Oxidative Phosphorylation	2.88E-04
Glycolysis	6.31E-05
Gluconeogenesis	3.72E-04
Nucleotide biosynthesis	1.91E-05
Tumor suppressor signaling:	
p53	0.0011
ATM	4.01E-04
PI3K/AKT signaling	0.002
Proliferative signaling:	
Myc-mediated apoptotic signaling	0.003
Telomerase signaling	0.007
Transcription:	
Aryl Hydrocarbon receptor signaling	0.002

Table 3B. tMDS vs Normal

Biological Pathways	p-value
ELEVATED EXPRESSION IN tMDS (Cluster 1)	
Stem Cell-related:	
Embryonic Stem Cell Pluripotency	0.05
AML signaling-related:	
Acute Myeloid Leukemia	0.04
JAK/Stat	0.04
Apoptosis Signaling	0.02
Immunologic:	
B Cell Receptor Signaling	0.03
PI3K Signaling in B Lymphocytes	0.04
LOWER EXPRESSION IN tMDS (Cluster 2)	
Cell Cycle:	
Assembly of RNA Polymerase III Complex	0.005
Assembly of RNA Polymerase I Complex	0.004
Metabolism-related:	
TCA Cycle II	0.04
Signaling-related:	
Apoptosis pathways	0.02
PI3K/AKT pathway	0.04