

# Epigenetic Regulation in Myelodysplastic Syndrome as A Consequence of A MicroRNA Dysregulation.

**Daniele Avenoso** (✉ [d.avenoso@nhs.net](mailto:d.avenoso@nhs.net))

King's College Hospital

**Margherita Squillario**

University of Genoa

**Annalisa Barla**

University of Genoa

**Alessandro Verri**

University of Genoa

**Enrico Carminati**

University of Genoa

**Clara Nurra**

University of Genoa

**Elisabetta Todisco**

European Institute of Oncology IRCCS

**Federica Gigli**

European Institute of Oncology IRCCS

**Francesco Bertolini**

European Institute of Oncology IRCCS

**Giuliana Gregato**

European Institute of Oncology IRCCS

**Claudia Poletti**

European Institute of Oncology IRCCS

**Maurizio Miglino**

University of Genoa

---

## Research Article

**Keywords:** Haemato-Oncology, Myelodysplastic Syndrome, hemopoiesis, isolation

**Posted Date:** April 26th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-371877/v1>



# Abstract

The myelodysplastic syndromes are a group of diseases characterized by impairment in hemopoiesis and variable tendency to progress to acute myeloid leukemia. Recent data regarding the MDS pathogenesis highlight alterations in several compartments of DNA replication, cell cycle control, and gene expression machinery. As previously reported, miRNA families are involved in MDS pathogenesis. Also, different miRNAs have been characterized in AML and MDS and allow the identification of prognostic risk categories independent from the revised international prognostic scoring system (R-IPSS). Herein we report the results and the possible scenarios regarding pathogenesis and disease progression secondary to the evidence of considerable gene network derangement following miRNA evaluation in a cohort of patients.

## Introduction

The myelodysplastic syndromes (MDS) are a group of diseases characterized by impairment in hemopoiesis and variable tendency to progress to acute myeloid leukemia (AML).

The World Health Organization (WHO) classification divides the MDS into different categories such as: refractory anemia with excess of blast (RAEB), refractory anemia (RA), refractory cytopenia with multilineage dysplasia (RCMD)<sup>1</sup>. Chronic myelomonocytic leukemia (CMML) is a myeloid entity between MDS and myeloproliferative neoplasm; it can have dysplastic features, and it has also tendency to transform in AML<sup>2</sup>.

Recent data regarding the MDS pathogenesis highlight the presence of alterations in several compartments of DNA replication, cell cycle control, and genes expression machinery<sup>3,4</sup>, such as:

1. RNA splicing (i.e., SF3B1, SRSF2, U2AF35, and ZRSR2)
2. Epigenetic regulation (i.e., TET2, DNMT3A, IDH1, IDH2, ASXL1)
3. Post-transcriptional regulation of gene expression (i.e., miR10, miR181, mir155, miR29, and miR221)

In the early 1990s the discovery of microRNAs (miRNAs) emerged as important regulators of gene expression<sup>5</sup>. Thus a potential role in oncogenesis has been evocated recently<sup>6</sup>.

Several reports showed an abnormal expression of miRNAs in several hematological neoplasms<sup>7</sup>.

Considering that miRNAs act as post-transcriptional regulators of several proteins, their role in the pathological epigenetic regulation of MDS is becoming progressively relevant<sup>8</sup>.

The micro-array-based platform and SNP-A studies showed alteration in miRNA genes (obtained from either methylation of their promoter or genetic loss of material or gene mutation) that allow pathologic expression of miRNA patterns associated with distinct phenotypes of disease.

As previously reported, miRNA families are involved in MDS pathogenesis. Indeed the deletion of miRNA-145 and miRNA-146 located on chromosome 5 leads to the 5q- syndrome<sup>3,9</sup>.

Also, different miRNAs have been characterized in AML and MDS and allow the identification of prognostic risk categories independent from the revised international prognostic scoring system (R-IPSS)<sup>10</sup>.

miRNAs alteration may have oncogenic properties or act as a tumor suppressor and have an active role in the onset of myeloid disorders.

In the last 5–10 years, next-generation sequencing (NGS) has been introduced in most hematologic laboratories, with various myeloid NGS panels now being commercially available.

These panels are based on targeted resequencing and usually analyze 25–50 genes. The genes tested within these panels can be classified into several functional categories, including the splicing machinery (i.e., U2AF1, SF3B1, SRSF2, ZRSR2), epigenetic modifiers (i.e., TET2, DNMT3A, BCOR, ASXL1, IDH1, IDH2), cohesins (i.e., STAG2, RAD21, and SMC3), transcription factors (i.e., RUNX1, WT1, ETV6), signaling molecules (e.g., NF1, NRAS, CBL, PTPN11, JAK2, FLT3), and chromatin modifiers (i.e., EZH2, ASXL1)<sup>11</sup>.

The detection of these mutation can help to exclude benign causes of cytopenias in cases with non-diagnostic morphology. Considered the relatively high frequency of clonal haematopoiesis in elderly patients, the presence of mutations in MDS-associated genes should be interpreted with caution and does not establish a diagnosis of MDS in isolation<sup>12</sup>.

A prospective study-based miRNA profiling was launched at Policlinico San Martino in 2017.

The aim of this paper is to report the results and the possible speculation regarding pathogenesis and disease progression.

## Methods

### Patients selection and bone marrow processing

Twenty-eight patients were considered eligible for this study because they met our inclusion criteria: (i) availability of bone marrow sample at diagnosis, (ii) IR-PSS, and (iii) clinical follow-up.

Patients were allocated to a different type of MDS according to the WHO 2008 classification as per investigator choice.

Eleven patients were affected by RAEB, nine by RA, five by RCMD, and three by CMML. The median age was 66 years old (48 - 85). Eight patients evolved into acute myeloid leukemia.

The median follow-up is 36 months (range 2 – 120).

Six patients were alive, 17 died due to disease progression, six were lost to follow up.

The population was also stratified according to R-IPSS at diagnosis as it follows: 6 low, three intermediate-1, six intermediate-2, 13 high.

NGS analysis was performed on bone marrow aspirate at diagnosis with Myeloid Oncomine Panel Thermo fisher; miRNA expression was quantified on bone marrow aspirate with TaqMan Advanced miRNA Cards. Data analysis was censored in January 2018. The study was approved by the local ethical committee at Policlinico San Martino; all the procedures were in accordance with local guidelines and regulations. All the patients provided written informed consent regarding the utilization of bone marrow samples for research purposes.

**miRNAs selection within the WHO miRNAs signature**

To identify which genes are influenced by the miRNAs signature, we proceeded with labeling those miRNAs that are expressed. The average expression of each miRNA was calculated, and the expression value of each miRNA in each patient was compared with its average expression: those miRNAs having expression values above their average were considered expressed. In contrast, the remaining were considered not expressed.

After this labeling phase, for each WHO class, we proceed with selecting those miRNAs that are expressed in most patients. Since the number of patients is not evenly distributed among the four WHO classes, we applied the thresholds shown in Table 1.

**Table 1 – miRNAs thresholds and number of subjects in the 4 WHO classes.** For each WHO class it is shown the number of subjects belonging to that class and the thresholds applied for considering a miRNA expressed in that class. For example a miRNA is considered expressed in the RAEB class if it is found expressed in at least the 75% of the subjects. The same threshold is applied for the RA class, while the two remaining classes have more stringent thresholds since characterized by a lower number of subjects.

WHO	Threshold	# subjects
MDS-EB	>= 75%	12
MDS-SLD	>= 75%	9
CMML	100%	3
MDS-MLD	>= 80%	5

**Statistical analysis & miRNA enrichment**

In the context of a machine learning supervised analysis, significant differences in the miRNAs expression among the patients classified based on the WHO 2008 classification were evaluated. This analysis was conducted using l1l2<sup>13</sup>-penalized regularization method within PALLADIO<sup>14</sup>, a machine-learning framework that can provide robust variable selection in high-dimensional problems.

When a signature of miRNAs able to discriminate the four WHO classes were identified, a phase of miRNAs target prediction was performed. We considered six prediction tools and all the target genes resulting from this phase. The lists of genes targeted by the miRNAs signature were characterized through an enrichment analysis using the webtoolkit WebGestalt. Within this tool, we considered the Kyoto Encyclopedia of Genes and Genomes (KEGG), a database of manually curated biological pathways.

Among the pathways found enriched, we selected the pathways characterizing one specific WHO class and those in common to more classes. Successively we selected those common pathways to all four MDS classes, which were modulated/regulated by most of our signature miRNAs. Considering these two criteria, we highlighted 5 KEGG pathways that played significant roles across the four WHO classes of MDS. These pathways are: AMPK signaling pathway<sup>15</sup>, RAS signaling pathway<sup>16</sup>, EGFR tyrosine kinase inhibitor resistance<sup>17</sup>, longevity regulating pathway, RAP1 signaling pathway<sup>18</sup>.

Further experimental procedures are described in the supplemental material.

## Results

### miRNA expression and WHO subtypes.

Each patient carried a distinctive miRNA signature composed of a median of 650 miRNAs.

The supervised analysis identified a signature of 24 miRNAs (Table 2) able to discriminate the 4 WHO classes with 62% of accuracy and p-value < 0.001.

Also, the analysis allowed to select 14 miRNAs frequently expressed in most of the patients of one or more WHO classes (Fig. 1).

**Table 2 – Signature of the twenty-four most frequent miRNAs.** Freq is the frequency that indicates the number of times each miRNAs was selected over the number of repetitions of regular and permutation experiments (see Materials and Methods). Here we chose 100 repetitions subdivided in 50 regular and 50 permutation experiments.

Freq.	miRNAs
96.0	miR-1203
95.0	miR-1292-5p
95.0	miR-183-5p
95.0	miR-1197
92.0	miR-1270
92.0	miR-1178-3p
91.0	miR-141-3p
90.0	miR-100-5p
90.0	let-7a-3p
90.0	miR-10b-5p
89.0	miR-1224-3p
86.0	miR-10a-3p
84.0	miR-222-5p
84.0	miR-1233-3p
84.0	miR-1179
83.0	miR-23a-5p
82.0	miR-1236-3p
80.0	miR-122-5p
79.0	miR-129-2-3p
79.0	miR-145-3p
79.0	miR-216a-5p
79.0	miR-1302
77.0	miR-141-5p
75.0	miR-1183

### NGS on bone marrow sample at diagnosis

NGS identified mutations in 43% of patients affected by RAEB, RCMD, and CMML in genes known to be pathogenic and prognostic for MDS such as: TET2, IKZF1, WT1, SH2B3, NF1, ASXL1, RUNX1, BCOR, U2A1F, ABL1, PRPF8, SETBP1, ZRSR2, SH2B3, MFSD11, CSDE1 (figure 2 summarizes their frequency.)

## Target Gene Prediction

Based on the miRNA expression result, prediction tools for miRNA target gene prediction were used as explained in the supplemental material.

The target genes identified were compared to the complete list of genes known to be involved in MDS from the related page of the Online Mendelian Inheritance in Man (OMIM).

This latter list is composed of the following three groups of genes: TET2, SF3B1, ASXL1, and GNB1 belong to the group of genes characterized by somatic mutations; GATA2, TERC, and TERT belong to the group of genes characterized by heterozygous germline mutations that predispose patients to MDS and other myeloid disorders and the last group of genes, which are DNMT3A, U2AF1, and JAK2, are characterized by heterozygous somatic mutations.

When we compared the lists of targets, coming from the miRNAs signature identified in the multiclass task, with the above-described OMIM list, we found the first and the last group of genes be involved in disease expression in our cohort of patients.

Our miRNA signature did not identify the genes belonging to germline predisposition.

## Functional characterization of the target genes list

Aiming to identify which gene pathways were involved (mostly silenced) by the miRNA signature detected, a functional analysis using the webtoolkit WebGestalt was used<sup>19</sup>.

Within this tool, we considered KEGG a database of pathways<sup>20</sup>.

The miRNA enrichment analysis allowed the identifications of genes involved in (i) pathways (*some in common and some specific for a WHO class*) likely involved in the onset of MDS, (ii) in pathways of diseases related to MDS and disease evolution like the acute myeloid leukemia, as summarized in figure 3.

Through the miRNA signature and its enrichment, it was possible to detect some known gene pathways involved in disease pathogenesis and progression, such as: cAMP signaling, mTOR, hedgehog, TGF-beta signaling, hypoxia-inducible factor, and phospholipase D signaling.

It was also possible to identify other gene pathways that may contribute to some of the symptoms reported by MDS patient; the miRNA signature is suggesting likely the silencing of different genes involved in the T-cell receptor signaling<sup>21</sup> (i.e., AKT serine/threonine kinase 1-2-3, CD3e, CD4, CD8a, CD20, and CD40 ligand), FC gamma receptor, calcium signaling, adherens junction, and CD46<sup>22</sup>.

## Functional characterization in RAEB patients

A specific focus was directed to those patients affected by RAEB because they had detectable blasts and likely had more DNA damage.

Among the miRNAs expressed in most of the patients belonging to the RAEB class, miR-1270 and miR-1179 emerged because the functional characterization of their target genes achieved good enrichment results. Both are targeting TET2, likely as suppressors of its expression.

The enrichment in KEGG gave a statistically relevant association of 16 genes pathways targeted by miR-1270, whereas miR1179 can influence 19 genes pathways (table 3 and table 4 show which gene pathways are controlled).

**Table 3 – Result of the enrichment analysis in KEGG of the miR-1270 target genes list.** The genes predicted to be regulated by miR-1270 were functionally characterized by enrichment analysis in KEGG database. The sixteen pathways shown in this table derived from a selection of the most relevant in the MDS context. FDR, False Discovery Rate. #Genes, number of genes belonging to our list of genes predicted to be regulated by miR-1270.

Name	#Genes	FDR
Signaling pathways regulating pluripotency of stem cells	85	2.35e-05
ErbB signaling pathway	57	3.47e-05
Proteoglycans in cancer	112	8.59e-05
Phospholipase D signaling pathway	82	1.52e-04
Ras signaling pathway	121	1.52e-04
EGFR tyrosine kinase inhibitor resistance	51	1.52e-04
AMPK signaling pathway	72	1.52e-04
Sphingolipid signaling pathway	69	2.91e-04
Focal adhesion	107	3.13e-04
FoxO signaling pathway	75	3.52e-04
Choline metabolism in cancer	59	4.52e-04
Rap1 signaling pathway	110	4.90e-04
AGE-RAGE signaling pathway	58	8.59e-04
mTOR signaling pathway	82	1.24e-03
T cell receptor signaling pathway	59	1.54e-03
Calcium signaling pathway	92	4.09e-03

**Table 4 – Result of the enrichment analysis in KEGG of the miR-1179 target genes list.** The genes predicted to be regulated by miR-1179 were functionally characterized by enrichment analysis in KEGG database. The nineteen pathways shown in this table derived from a selection of the most relevant in the MDS context. The pathways in bold are in common with the ones found enriched in the miR-1270 functional analysis. FDR, False Discovery Rate. #Genes, number of genes belonging to our list of genes predicted to be regulated by miR-1179.

Name	#Genes	FDR
<b>Proteoglycans in cancer</b>	100	2.25e-04
<b>AMPK signaling pathway</b>	65	4.12e-04
<b>Longevity regulating pathway</b>	52	4.12e-04
<b>Choline metabolism in cancer</b>	82	8.56e-04
<b>Sphingolipid signaling pathway</b>	62	8.56e-04
<b>P53 signaling pathway</b>	39	1.3e-03
<b>EGFR tyrosine kinase inhibitor resistance</b>	44	1.55e-03
<b>mTOR signaling pathway</b>	74	1.77e-03
<b>Ubiquitin mediated proteolysis</b>	66	2.72e-03
<b>Signaling pathways regulating pluripotency of stem cells</b>	68	2.72e-03
<b>Phosphatidylinositol signaling system</b>	50	3.47e-03
<b>Ras signaling pathway</b>	101	4.38e-03
<b>PI3K-AKT signaling pathway</b>	143	4.67e-03
<b>Rap1 signaling pathway</b>	94	4.84e-03
<b>Transcriptional misregulation in cancer</b>	81	6.48e-03
<b>Fc gamma R-mediated phagocytosis</b>	46	8.21e-03
<b>Adherens junction</b>	38	8.77e-03
<b>Central carbon metabolism in cancer</b>	35	8.77e-03
<b>AGE-RAGE signaling pathway</b>	49	8.77e-03

Among these 16 genes pathways, NF1, ABL1, and TP53 were also detected with our NGS analysis (not all patients with RAEB).

Since the TET2 gene is essential in the pathogenesis of MDS and can be prognostic in terms of therapy<sup>23</sup>, it was assessed the possible interactions of this gene with those belonging to some interesting KEGG pathways listed in Table 3 and 4 using the web tool STRING.

It was possible to detect 85 protein-protein interactions between TET2 and genes of the “Signaling pathways regulating pluripotency of stem cells” (influenced by miR1270 and miR1179) (Figure 4).

This network shows that the interaction between TET2 and TCF3 is experimentally known<sup>24</sup> (pink edge), while the interactions between TET2 and the other three proteins, that are JARID2, KRAS, and JAK2, derived from co-expression experiments in homologs proteins in other species.

JARID2 encodes for a transcriptional repressor that interacts with the Polycomb repressive complex 2 (PRC2), which plays an essential role in regulating gene expression during embryonic development. This protein facilitates the recruitment of the PRC2 complex to target genes. Mutations in this gene are associated with myeloid malignancies<sup>25</sup>.

### **Shared miRNAs and early steps in MDS pathogenesis.**

The miRNA signature and its enrichment also allowed to identify of a list of miRNAs whose targets are involved in regulating some genetic pathways that are in common between the different WHO classes.

This list comprehends eight miRNAs that are miR-10b-5P, miR-100-5p, miR-183-5p, miR-1179, miR-1270, miR-141-3p, miR-1292-5p, miR-1197. AMPK, EGFR, RAS, longevity, RAP1 represent those gene pathways modulated by the list of miRNAs mentioned above.

This result suggests that several genes in different gene pathways may be affected by these eight miRNAs. Table 5 is summarizing which biological processes are influenced.

**Table 5 – List of significant GO terms belonging to biological process and molecular function and of significant KEGG pathways related to STRING network shown in Figure 5.. This table show the enrichment in Gene Ontology (GO) and in KEGG related to the gene list gave in input to STRING in order to identify the network shown in figure 5. FDR, False Discovery Rate.**

Biological Process	Count in gene set	FDR
hematopoietic or lymphoid organ development	10 of 573	2.56e-05
response to oxygen-containing compound	13 of 1427	6.26e-05
hemopoiesis	9 of 526	6.26e-05
intracellular signal transduction	13 of 1528	9.07e-05
regulation of primary metabolic process	23 of 5982	0.00025
Molecular Function	Count in gene set	FDR
pre-mRNA binding	4 of 36	0.00012
heterocyclic compound binding	21 of 5305	0.00040
organic cyclic compound binding	21 of 5382	0.00040
phosphoprotein binding	4 of 80	0.00063
nucleic acid binding	16 of 3332	0.00063
KEGG pathways	Count in gene set	FDR
RAS signaling pathway	12 of 228	1.65e-13
Glioma	6 of 68	1.34e-07
EGFR tyrosine kinase inhibitor resistance	6 of 78	1.93e-07
Rap1 signaling pathway	7 of 203	1.29e-06

Since these miRNAs are present in all patients, it is reasonable to suggest that these may be the ones to be deranged likely at the onset of disease. Also, they are likely working as post-transcriptional suppressor in gene pathways that have been described to be fundamental for normal haemopoiesis<sup>26,27,28</sup>.

STRING tool allowed to detect protein-protein interactions between the genes controlled by the miRNA signature and the genes mutated detected in NGS (figure 5).

Three genes are outliers in STRING: RAPGEF5, CSDE1, MSFD1. This result is suggesting that even if the miRNA signature may negatively influence the expression of these genes, they could be not involved neither in MDS pathogenesis or disease progression.

## Discussion

MDS are a group of diseases with different clinical behavior arising from a complex pathogenesis.

Several mechanisms of DNA-damages can be detected through SNP karyotyping, NGS, and miRNA assessment.

The present study evaluated the presence of miRNA expression in a cohort of patients affected by MDS.

As described by other groups, we identified different mutations via NGS in known genes involved in pathogenesis and disease progression<sup>29,30</sup>; anyway, a 43% rate of gene mutations underestimates the actual mutation rate in MDS. Our result may be a consequence of old samples tested for NGS; indeed, DNA degradation is a well-known mechanism for failure of NGS analysis<sup>31</sup>.

We decided to focus on miRNA analysis for several reasons: (i) miRNA are well known mechanism of post-transcriptional regulation, (ii) some miRNAs have been associated with some specific type of MDS.

Our analysis identified a signature able to discriminate different types of MDS according to the WHO classification ( $p\text{-value} < 0.001$ ).

It's worth to highlight that the system of MDS classification is based on cells morphology, degree of pancytopenia and karyotyping. For the exception of SF3B1 and deletion chromosome 5q, which are findings defining a specific subgroup of MDS, there is no test in molecular biology that could support a specific type of MDS diagnosis. This analysis gives a different contribution in defining specific types of MDS.

Also, the miRNA enrichment allowed a deepest analysis of miRNA de-arrangement in our population: the gene target prediction of our miRNA signature was able to detect genes involved in RNA splicing and epigenetic mutation but not those involved in germ line predisposition to myeloid malignancies. This finding suggests that MDS arising in patient with the latter group of mutations may have a different pathogenesis even if some mechanism of disease onset and progression are shared with cases de novo.

Further, the miRNA enrichment was able to identify in our population genes involved in MDS pathogenesis and disease evolution to AML. It is interesting that the miRNA signature in this study influences genes involved in the immune system; this result may be the starting point to speculate about infective risk, opportunistic infection and autoimmune events in MDS. Anyway, this aspect was not explored due to the lack of clinical data about infective events and because our study population is relatively restricted.

The patient's functional characterization affected by RAEB suggested a disarranged gene expression in this population. We detected higher rate of NGS mutation in our population, but we identified two miRNAs (mir-1270 and mir-1179) with an interesting enrichment: it suggests a possible broad genetic lesion as suggested in Fig. 5.

The STRING analysis showed several protein-protein interactions likely influenced by our miRNA signature.

Even if there is no mutation all genes illustrated in Fig. 5 we think they are silenced by the miRNA signature identified in our study contributing to the onset of RAEB.

It is well known that mutation in TET2 are causing loss of function; this study showed the miR-1270 and miR-1179 are targeting TET2, causing its silencing.

It is reasonable to suggest that this miRNA expression and gene mutations contribute to the clinical/morphological aspect of the disease.

The availability of a broad miRNA signature allowed also a different point of view: it was possible to identify 8 miRNAs that focus on specific gene pathways common to all patients.

The involvement of AMPK, EGFR, RAS, longevity, RAP1 pathways is in specific genes (Table 6) and some of them (i.e. IGF1R<sup>32</sup>, PRKAA2<sup>33</sup>, TSC1<sup>34</sup>) have key role in all the pathways mentioned above; this suggests a redundant mechanism of disease progression in keeping with the effect of miRNAs on different gene pathways.

**Table 6 – Gene lists regulated by 7 or 8 miRNAs involved in five KEGG pathways common to all the WHO classes.** The 8-focused miRNAs are targeting common genes within this common list of pathways, suggesting a redundant mechanism of disease progression

AMPK	RAS	EGFR	Longevity	RAP1
IGF1R	GRIN2A	IGF1R	IGF1R	IGF1R
PRKAA2	IGF1R	NF1	PRKAA2	PDGFRA
TSC1	KSR2		TSC1	RAPGEF5
	MAPK1			
	NF1			
	PDGFRA			
	PIK3R3			
	PLA2G2D			
	PRKCB			
	RAPGEF5			
	RASAL2			

progression. Anyway, our results are not showing if the 8-focusing miRNAs are the cause of altered epigenetic in MDS or are consequence of it.

It is reasonable to think that these early miRNAs are influencing genes fundamental in normal cell processes and in regulation of other miRNAs; this would justify a domino effect in MDS patient, where there is common miRNAs and genes dysregulation that become patient specific with disease progression due the involvement of other genes and miRNAs. This is an interesting theory that needs further in vitro validation, but a similar mechanism was already described in pediatric acute lymphoblastic leukemia<sup>35</sup>.

Anyway, some limitations can be detected in this study: small and retrospective nature of the study and non-selection of CD34 + cells. Considering the presence of difference clones in MDS population, our miRNA signature may not represent the expression of only one clone<sup>36</sup>. The dominant one may influence the miRNA signature hiding the ones from minor clones or the latter may “contaminate” the miRNA signature detected.

Also, the small population is not allowing the identification of other miRNAs WHO specific and it is not possible to predict a miRNA signature that can be prognostic in term of AML transformation or sensitiveness to hypomethylating therapy.

In conclusion, the miRNA signature and its enrichment may become potential tools for defining the diagnosis of each MDS subtypes; the integration of this strategy with NGS needs a larger and homogenous population to further define which genetic pathways are involved and which miRNAs may be useful to further understand the pathogenesis of MDS.

Also, this study is suggesting that the epigenetic regulation in MDS is not only a consequence of altered methylation but it is also a consequence of a miRNAs dysregulation causing a broad genetic damage.

## Declarations

**Acknowledgments:** AIRC, Italian Ministry of Health

**Funding:** PRIN – Progetti di Rilevante Interesse Nazionale

**Author's contributions:** DA and MM designed the study and wrote the manuscript. MS, AB, AV performed the statistical analysis and contributed to writing the manuscript. EC, CN, DA and MM performed miRNA evaluation. ET, FG contributed to writing the manuscript. FB, GG, CP performed NGS evaluation and contributed to writing the manuscript

## References

1. Arber, D. A. *et al.* The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. **127**, 2391–2406 (2016).

2. Patnaik, M. M. & Tefferi, A. Chronic myelomonocytic leukemia: 2016 update on diagnosis, risk stratification, and management. *Am. J. Hematol.* **91**, 631–642 (2016).
3. Bejar, R. & Steensma, D. P. Recent developments in myelodysplastic syndromes. *Blood.* **124**, 2793–2803 (2014).
4. O'Connell, R. M., Zhao, J. L. & Rao, D. S. MicroRNA function in myeloid biology. *Blood.* **118**, 2960–2969 (2011).
5. Moss, E. G. MicroRNAs: hidden in the genome. *Curr. Biol.* **12**, R138–40 (2002).
6. Zhang, B., Pan, X., Cobb, G. P. & Anderson, T. A. microRNAs as oncogenes and tumor suppressors. *Dev. Biol.* **302**, 1–12 (2007).
7. Gordon, J. E. A., Wong, J. J. L. & Rasko, J. E. J. MicroRNAs in myeloid malignancies. *Br. J. Haematol.* **162**, 162–176 (2013).
8. Sokol, L. *et al.* Identification of a risk dependent microRNA expression signature in myelodysplastic syndromes. *Br. J. Haematol.* **153**, 24–32 (2011).
9. Song, S. J. & Pandolfi, P. P. MicroRNAs in the pathogenesis of myelodysplastic syndromes and myeloid leukaemia. *Curr. Opin. Hematol.* **21**, 276–282 (2014).
10. Bravo, G. M., Lee, E., Merchan, B., Kantarjian, H. M. & García-Manero, G. Integrating genetics and epigenetics in myelodysplastic syndromes: advances in pathogenesis and disease evolution. *Br. J. Haematol.* **166**, 646–659 (2014).
11. Bacher, U. *et al.* Challenges in the introduction of next-generation sequencing (NGS) for diagnostics of myeloid malignancies into clinical routine use. *Blood Cancer Journal.* **8**, 1–10 (2018).
12. Al-Kali, A. *et al.* Continue NCCN Guidelines Panel Disclosures NCCN Guidelines Version 2.2020 Myelodysplastic Syndromes(2020).
13. De Mol, C., Mosci, S., Traskine, M. & Verri, A. A Regularized Method for Selecting Nested Groups of Relevant Genes from Microarray Data. *J. Comput. Biol.* **16**, 677–690 (2009).
14. Matteo Barbieri, S., Fiorini, F., Tomasi, A. B. & PALLADIO A Parallel Framework for Robust Variable Selection in High-Dimensional Data | Semantic Scholar. Available at: <https://www.semanticscholar.org/paper/PALLADIO%3A-A-Parallel-Framework-for-Robust-Variable-Barbieri-Fiorini/26931b26c09600ed41891fc6d748153407ccbdb6>. (Accessed: 4th May 2020)
15. Jacquél, A., Luciano, F., Robert, G. & Auberger, P. Implication and regulation of AMPK during physiological and pathological myeloid differentiation. *International Journal of Molecular Sciences* **19**, (2018).
16. Jhanwar, S. C. Genetic and epigenetic pathways in myelodysplastic syndromes: A brief overview. *Advances in Biological Regulation.* **58**, 28–37 (2015).
17. Bachegowda, L. *et al.* Signal transduction inhibitors in treatment of myelodysplastic syndromes. *Journal of Hematology and Oncology.* **6**, 50 (2013).
18. Gyan, E. *et al.* Mutation in RAP1 is a rare event in myelodysplastic syndromes [5]. *Leukemia.* **19**, 1678–1680 (2005).

19. Wang, J., Vasaikar, S., Shi, Z., Greer, M. & Zhang, B. WebGestalt 2017: A more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* **45**, W130–W137 (2017).
20. Kanehisa, M. K. E. G. G. Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
21. Gaud, G., Lesourne, R. & Love, P. E. Regulatory mechanisms in T cell receptor signalling. *Nature Reviews Immunology.* **18**, 485–497 (2018).
22. CD46 - an overview | ScienceDirect Topics. Available at: <https://www.sciencedirect.com/topics/neuroscience/cd46>. (Accessed: 27th March 2020)
23. Delhommeau, F. *et al.* Mutation in TET2 in Myeloid Cancers. *N. Engl. J. Med.* **360**, 2289–2301 (2009).
24. Lin, X. C. *et al.* Molecular dysfunctions in acute myeloid leukemia revealed by integrated analysis of microRNA and transcription factor. *Int. J. Oncol.* **48**, 2367–2380 (2016).
25. Celik, H. *et al.* JARID2 Functions as a Tumor Suppressor in Myeloid Neoplasms by Repressing Self-Renewal in Hematopoietic Progenitor Cells. *Cancer Cell.* **34**, 741–756 (2018).
26. Saito, Y. & Nakada, D. The role of the Lkb1/AMPK pathway in hematopoietic stem cells and Leukemia. *Crit. Rev. Oncog.* **19**, 383–397 (2014).
27. Wee, P. & Wang, Z. Epidermal growth factor receptor cell proliferation signaling pathways. *Cancers* **9**, (2017).
28. Chung, E. & Kondo, M. Role of Ras/Raf/MEK/ERK signaling in physiological hematopoiesis and leukemia development. *Immunol. Res.* **49**, 248–268 (2011).
29. Papaemmanuil, E. *et al.* Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood.* **122**, 3616–3627 (2013).
30. Della Porta, M. G. *et al.* Clinical Effects of Driver Somatic Mutations on the Outcomes of Patients With Myelodysplastic Syndromes Treated With Allogeneic Hematopoietic Stem-Cell Transplantation. *J. Clin. Oncol.* **34**, 3627–3637 (2016).
31. Ivanov, M. *et al.* Towards standardization of next-generation sequencing of FFPE samples for clinical oncology: Intrinsic obstacles and possible solutions. *J. Transl. Med.* **15**, (2017).
32. IGF1R gene - Genetics Home Reference - NIH. Available at: <https://ghr.nlm.nih.gov/gene/IGF1R>. (Accessed: 27th March 2020)
33. PRKAA2 protein kinase AMP-activated catalytic subunit alpha 2 [Homo sapiens (human)] - Gene - NCBI. Available at: <https://www.ncbi.nlm.nih.gov/gene/5563>. (Accessed: 27th March 2020)
34. TSC1 gene - Genetics Home Reference - NIH. Available at: <https://ghr.nlm.nih.gov/gene/TSC1>. (Accessed: 27th March 2020)
35. Ma, Y. *et al.* Developmental timing of mutations revealed by whole-genome sequencing of twins with acute lymphoblastic leukemia. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 7429–7433 (2013).
36. Visconte, V., Tiu, R. V. & Rogers, H. J. Pathogenesis of myelodysplastic syndromes: An overview of molecular and non-molecular aspects of the disease. *Blood Research.* **49**, 216–227 (2014).

# Figures

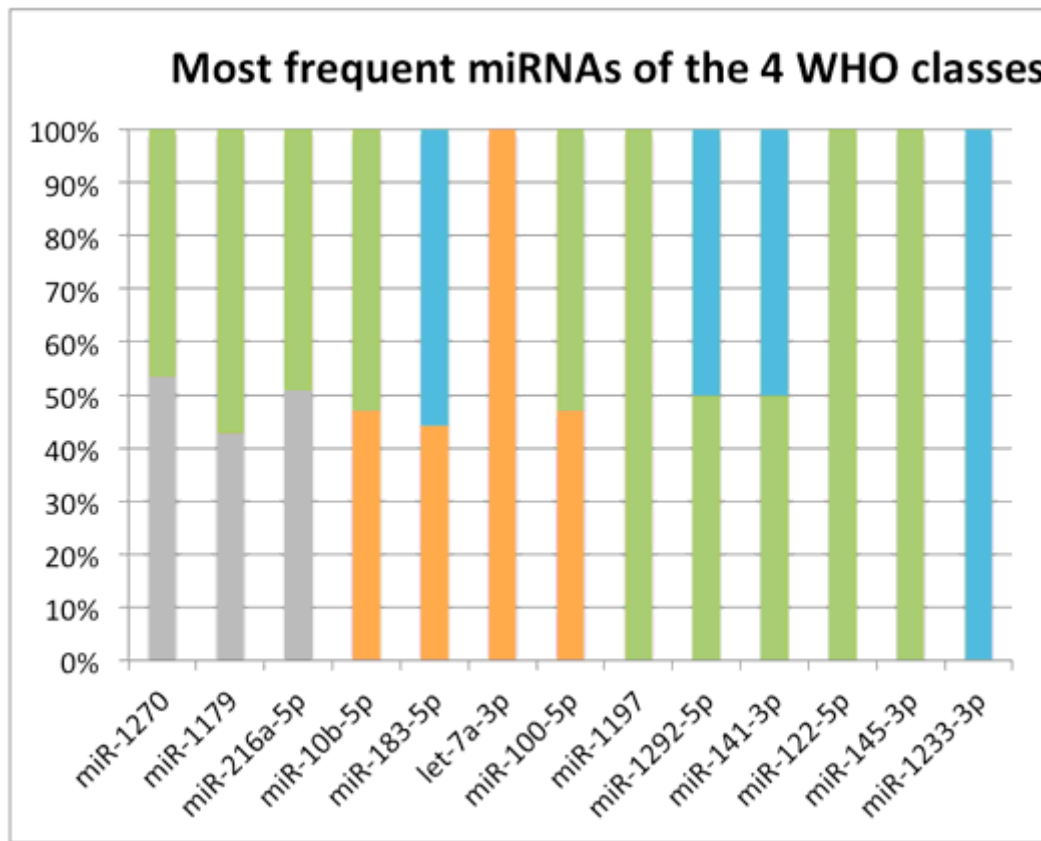
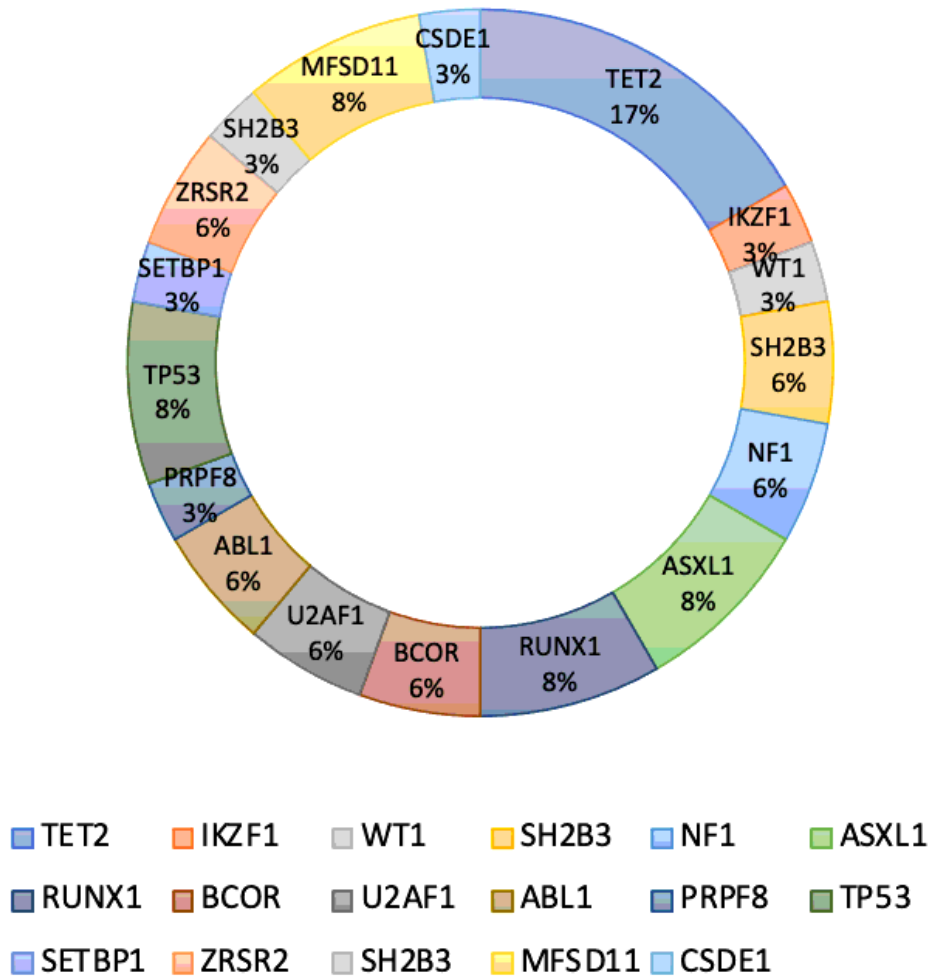


Figure 1

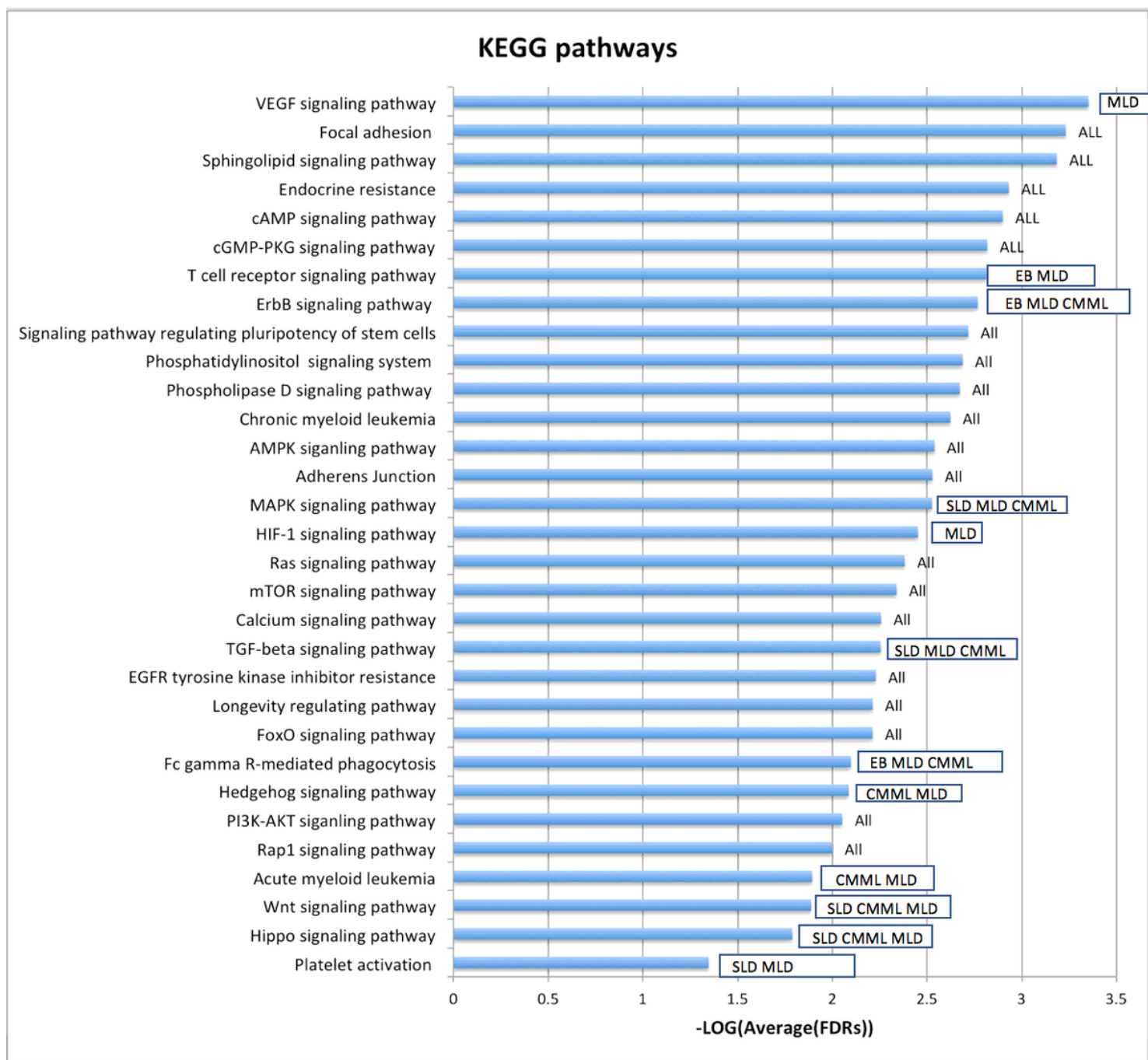
Stacked bar plot of the most frequent miRNAs. This plot shows the frequency of each miRNA in a specific WHO class. The miRNAs shown in this plot belongs to the identified miRNA signature. Blue is CMML, green is MDS-MLD, orange is MDS-SLD, grey is MDS-EB.

## Gene mutations detected with NGS



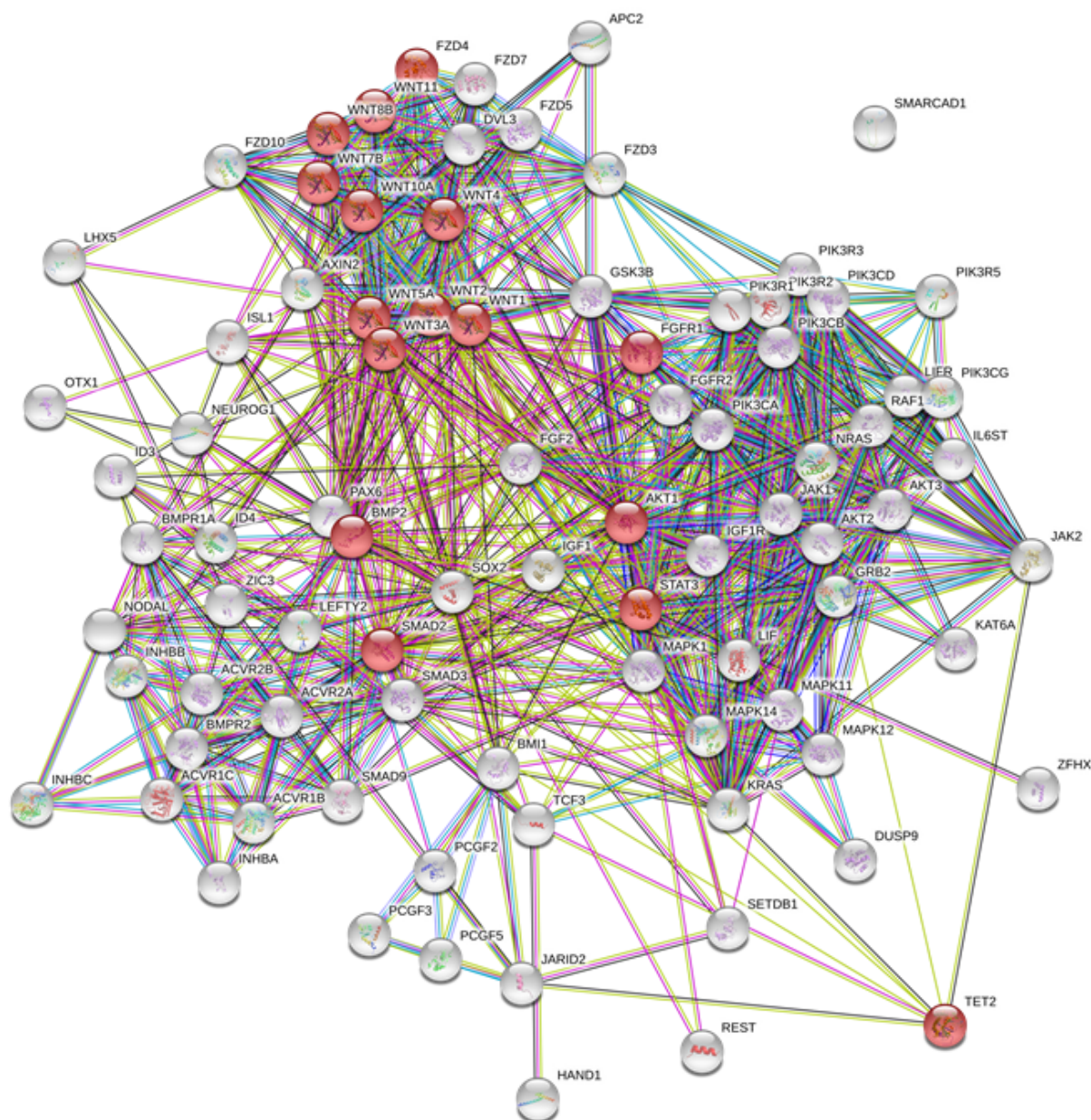
**Figure 2**

Frequency of genes mutations detected by NGS. All these genes where found to be involved in MDS by both NGS and miRNA enrichment analyses. The percentages indicate the number of time each gene has been found mutated or enriched with respect to the total number of analysed subjects.



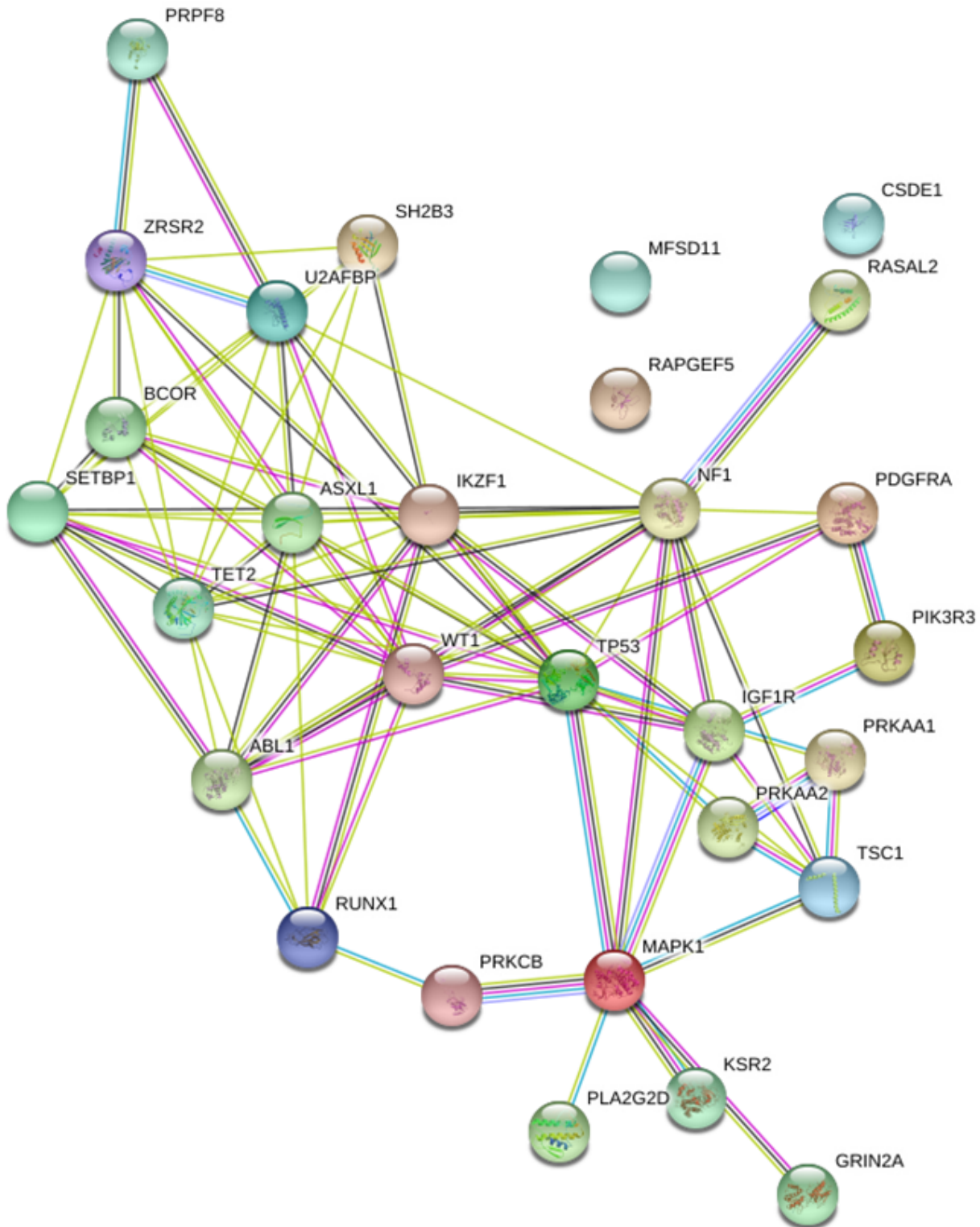
**Figure 3**

Altered cellular pathways involved in our MDS population, detected through miRNA enrichment. Each bar is characterized by the name of the KEGG pathway and by the complete list of the WHO classes in which we found enriched that specific pathway. On the x axis there is the log value of the mean of all the FDRs (False Discovery Rates) associated to each enriched pathway related to the expressed miRNAs specific for one or more WHO classes.



**Figure 4**

STRING network of the Hematopoietic stem cell self-renewal ability highlighted in red. We obtained this network using STRING database. In this context TET2 interact with TCF3, JARID2, KRAS and JAK2 proteins. In red are highlighted those proteins that are mentioned together in one article (PMID: 23675967) focused on the therapeutic potential of the mechanism underneath the HSC self-renewal ability which could be targeted to prevent leukemic stem cell self-renewal in myeloid malignancies.



**Figure 5**

STRING network between the genes common to all the 5 pathways and the genes measured in our NGS experiment. In this PPI analysis we want to observe the possible connections between the genes we identified to be regulated by the 8 miRNAs involved in the 5 pathways common to all the WHO classes and the panel gene list measured in our NGS experiment. The PPI enrichment p-value of this network is significant ( $< 1.0e-16$ ). The pink and the blue sky edges represents the strongest connections because

experimentally determined and derived from curated databases respectively. The green connections are the weakest connections because derived from textmining and the meaning of the remaining colors is here explained: black, co-expression in other homologous proteins; light blue, protein homology; the green, red and blue edges are predicted connections (gene neighborhood, gene fusions and gene co-occurrence).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementalmaterial.pdf](#)