Univerzita Karlova

1. lékařská fakulta

Studijní program: Molekulární a buněčná biologie, genetika a virologie

Mgr. Monika Kaisrlíková

Včasná detekce progrese onemocnění u pacientů
s myelodysplastickým syndromem

Early Detection of Disease Progression in Patients
with Myelodysplastic Syndromes

Disertační práce

Školitel: RNDr. Monika Beličková, Ph.D.
Konzultant: MUDr. Markéta Šťastná Marková, CSc.

Praha, 2022

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem řádně uvedla a citovala všechny použité prameny a literaturu. Současně prohlašuji, že práce nebyla využita k získání jiného nebo stejného titulu.

Souhlasím s trvalým uložením elektronické verze mé práce v databázi systému meziuniverzitního projektu Theses.cz za účelem soustavné kontroly podobnosti kvalifikačních prací.

V Praze, 21.11.2022

………………………..

Mgr. Monika Kaisrlíková

Identifikační záznam:

KAISRLÍKOVÁ, Monika. *Včasná detekce progrese onemocnění u pacientů s myelodysplastickým syndromem. [Early Detection of Disease Progression in Patients with Myelodysplastic Syndromes].* Praha, 2022. 193 s., 0 příl. Disertační práce. Univerzita Karlova, 1. lékařská fakulta, Ústav klinické a experimentální hematologie 1. LF UK a ÚHKT. Školitel Beličková, Monika.

ACKNOWLEDGEMENTS

ABSTRACT

Myelodysplastic syndromes (MDS) are a heterogeneous group of clonal hematopoietic disorders with a risk of transformation into acute myeloid leukemia (AML). The International Prognostic Scoring Systems integrate clinical data and cytogenetics to determine the risk of AML transformation for individual patients. Precise risk assessment is crucial for treatment decision-making.

The aim of this thesis was to identify molecular markers for the early detection of disease progression in MDS patients. Using cDNA microarrays and next-generation sequencing, we targeted long noncoding RNAs (lncRNAs) and recurrently mutated genes in bone marrow cells. In addition, we focused on the identification of pathways related to the progression of MDS and understanding how the identified biomarkers participate.

In the transcriptome study, we identify 4 candidate lncRNAs that may serve as prognostic biomarkers of the adverse course of MDS: *H19*, *WT1-AS*, *TCL6*, and *LEF1-AS*. Using various statistical approaches, we determined the level of *H19* to be a strong independent prognostic marker. Furthermore, our data showed that disruption of transcriptional coregulation of the imprinting locus *H19/IGF2* and miR-675, which directly regulates *H19* and plays a role in tumorigenesis, accompanies disease progression.

In the genomic study aimed at lower-risk MDS patients, we identified mutated *RUNX1*, *SETBP1*, *STAG2*, *TP53*, and *U2AF1* genes to have a significant effect on progression-free survival by univariate analysis. In multivariate analysis, the mutated *RUNX1* gene was determined to be the strongest predictive marker of rapid progression. We showed how the implementation of the *RUNX1* mutational status into the Revised International Prognostic Scoring System may improve patient stratification. We described an association of *RUNX1* with the DNA damage response (DDR) and cellular senescence and that its loss-of-function mutations lead to escape from these cellular protection barriers and to progression. The deregulation of DDR and cellular senescence in *RUNX1*-mutated patients was verified at the functional level by the detection of γH2AX protein expression and senescence-associated β-galactosidase activity.

In conclusion, we identified mutated and deregulated genes that can be used as predictive markers of rapid progression in MDS. Our results may contribute to the early detection of patients at risk of disease progression and the initiation of appropriate treatment. Simultaneously, we described cellular processes in which the biomarkers are involved and suggested their role in disease pathogenesis.

ABSTRAKT

Myelodysplastický syndrom (MDS) je heterogenní skupina onemocnění charakterizována klonální poruchou krvetvorby s rizikem transformace do akutní myeloidní leukémie (AML). Na základě vyšetření krevního obrazu, kostní dřeně a cytogenetiky je podle mezinárodních prognostických skórovacích systémů určováno riziko transformace do AML. Přesné určení rizika je klíčové pro zvolení správné léčby.

Cílem této práce byla identifikace molekulárních markerů pro včasnou detekci progrese onemocnění. Pomocí cDNA čipů a sekvenování nové generace byly analyzovány dlouhé nekódující RNA (lncRNA) a rekurentně mutované geny v buňkách kostní dřeně. Zároveň bylo naším cílem popsání signálních drah, které se podílí na progresi onemocnění, a vysvětlení, jak dané biomarkery k progresi přispívají.

V transkriptomové studii jsme identifikovali 4 kandidátní lncRNA, které by mohly sloužit jako prognostické biomarkery horšího průběhu MDS, a to *H19*, *WT1-AS*, *TCL6* a *LEF1-AS1*. Na základě několika statistických přístupů jsme prokázali, že hladina transkriptu *H19* může sloužit jako velmi silný nezávislý prognostický marker. Navíc naše data ukázala, že progrese je doprovázena poruchou transkripční regulace imprintovaného lokusu *H19/IGF2* a miR-675, která přímo reguluje *H19* a hraje významnou roli v tumorigenezi.

V genomické studii zaměřené na pacienty s nižším rizikem jsme pomocí univariantní analýzy identifikovali mutované geny *RUNX1*, *SETBP1*, *STAG2*, *TP53* a *U2AF1* jako geny se signifikantním vlivem na délku přežití bez progrese. V multivariantní analýze byl mutovaný gen *RUNX1* určen jako nejsilnější prediktivní marker časné progrese. Ukázali jsme, jak inkorporace mutačního statutu genu *RUNX1* do Revidovaného mezinárodního prognostické skórovacího systému může zlepšit stratifikaci pacientů. Popsali jsme asociaci tohoto genu s dráhou odpovědi na DNA poškození (DDR) a buněčnou senescencí, a že ztráta jeho funkce způsobená mutací vede k překonání buňku chránící bariéry a k progresi onemocnění. Deregulace dráhy DDR a buněčné senescence u pacientů s mutací v genu *RUNX1* byla pozorována i na funkční úrovni sledováním exprese proteinu γH2AX a aktivity β-galaktosidázy asociované se senescencí.

Identifikovali jsme geny, které, ať mutované nebo s deregulovanou expresí, mohou být využity jako prediktivní markery progrese MDS. Tyto poznatky mohou přispět k včasné identifikaci pacientů v riziku progrese onemocnění a vést k zahájení optimální léčby. Zároveň jsme popsali buněčné procesy asociované s danými biomarkery a navrhli jejich možné zapojení v patogenezi onemocnění.

Klíčová slova: myelodysplastický syndrom, patogeneze, progrese, lncRNA, *RUNX1*

LIST OF ABBREVIATIONS

AML	acute myeloid leukemia
AML-MRC	acute myeloid leukemia with myelodysplasia-related changes
ANC	absolute neutrophil count
BM	bone marrow
DDR	DNA damage response
del(5q)	deletion of the long arm of chromosome 5
FC	fold change
FDR	false discovery rate
GO	Gene Ontology
GO BP	Gene Ontology Biological Processes
GSEA	Gene Set Enrichment Analysis
Hb	hemoglobin
HR	high + very-high risk patients
HR-MDS	higher-risk myelodysplastic syndromes
HSCs	hematopoietic stem cells
HSCT	hematopoietic stem cell transplantation
IHBT	Institute of Hematology and Blood Transfusion
INT-1	intermediate-1 (IPSS category)
INT-2	intermediate-2 (IPSS category)
IPSS	International Prognostic Scoring System
IPSS-R	Revised International Prognostic Scoring System
IPSS-M	Molecular International Prognostic Scoring System
KEGG	Kyoto Encyclopedia of Genes and Genomes
lncRNA	long non-coding RNA
LR	very low + low-risk MDS patients
LR-MDS	lower-risk myelodysplastic syndromes
MDS	myelodysplastic syndromes
MDS-EB	MDS with excess blasts
MDS-MLD	MDS with multilineage dysplasia
MDS-RS	MDS with ring sideroblasts
MDS-SLD	MDS with single lineage dysplasia
MDS-U	MDS unclassifiable
miRNA	microRNA

mutR-LR	LR-MDS with *RUNX1* mutations
ncRNA	noncoding RNA
NGS	next-generation sequencing
nt	nucleotide
OS	overall survival
PB	peripheral blood
PCGs	protein-coding genes
PFS	progression-free survival
piRNA	piwi-interacting RNA
PLT	platelets
RT-qPCR	reverse transcription quantitative real-time PCR
SA-b-Gal	senescence-associated β-galactosidase
SASP	senescence-associated secretory phenotype
SC	stem cell
TE	transposable element
Th17	IL-17-producing CD4+ T cells
VAF	variant allele frequency
WHO	World Health Organization
wtR-LR	LR-MDS without *RUNX1* mutations
wt*RUNX1*	wild-type *RUNX1*

TABLE OF CONTENTS

1. INTRODUCTION

1.1. Myelodysplastic syndromes

Myelodysplastic syndromes (MDS) are a heterogeneous group of clonal disorders of hematopoietic stem cells (HSCs) characterized by ineffective hematopoiesis, cytopenias, and risk of transformation to acute myeloid leukemia (AML). Typically, it is a disease of the elderly (with a median age of 71-74 years at diagnosis) (Dinmohamed et al., 2014; Neukirchen et al., 2011). The incidence was described to be 2-4 cases per 100,000 people (Dinmohamed et al., 2014; Zeidan et al., 2017) and increases dramatically with age. In various studies, the incidence was described to be 20-50 cases per 100,000 after 60 years of age (reviewed by Malcovati et al., 2013). Risk factors for MDS are older age, male sex, smoking, chronic exposure to cancer-causing chemicals, prior chemotherapy or radiation therapy, family history of hematopoietic cancer, and inherited disorders such as Schwachman-Diamond syndromes, Fanconi anemia, and familial platelet disorder (Sekeres, 2010).

The management of MDS is most often intended to slow the disease, ease symptoms, and prevent complications. There is no cure except for hematopoietic stem cell transplantation (HSCT), but medications can help slow the progression of the disease.

1.1.1. MDS classification

The World Health Organization (WHO) 2016 classification (Arber et al., 2016) identifies several morphological subtypes of MDS according to the findings in peripheral blood (PB) and bone marrow (BM): MDS with single lineage dysplasia (MDS-SLD), MDS with multilineage dysplasia (MDS-MLD), MDS with ring sideroblasts (MDS-RS, further divided into MDS-RS-SLD and MDS-RS-MLD), MDS with isolated del(5q), MDS with excess blasts (MDS-EB-1 and MDS-EB-2), and MDS unclassifiable (MDS-U), further divided into subtypes with 1% peripheral blood blasts, with SLD and pancytopenia, or with defining cytogenetic abnormality (Table 1a). Refractory cytopenia of childhood represents the childhood MDS. At the time of writing this thesis, the 5th edition of the WHO Classification

of Haematolymphoid Tumours has been published (Khoury et al., 2022). Myelodysplastic syndromes have been replaced by the term myelodysplastic neoplasms, and the classification has been changed by distinguishing MDS with defining genetic abnormalities and morphologically defined MDS (Table 1b). Childhood MDS has become an independent entity with its own classification. Furthermore, MDS-U has been removed due to a new category named clonal hematopoiesis.

For prognostic purposes, MDS patients are classified according to their risk of transformation to AML using the International Prognostic Scoring System (IPSS) (Greenberg et al., 1997) or the Revised International Prognostic Scoring System (IPSS-R) (Greenberg et al., 2012). The IPSS stratifies MDS patients into 4 categories according to the percentage of BM blasts, number of cytopenias and cytogenetic category: low, intermediate-1 (INT-1), intermediate-2 (INT-2), and high risk (Figure 1). The IPSS-R defines 5 risk categories by the percentage of BM blasts, depth of cytopenias, and revised cytogenetic prognostic subgroups: very low, low, intermediate, high, and very high risk (Figure 1). The revised system should predict clinical outcomes more precisely.

Patients with low and INT-1, very low, low and intermediate risk up to 3.5 points, respectively, are considered to have lower-risk MDS (LR-MDS), while patients with INT-2 and high or intermediate risk with more than 3.5 points, high, and very high, respectively, are considered to have higher-risk MDS (HR-MDS) (Mufti et al., 2018; Pfeilstöcker et al., 2016). Scoring systems are essential for treatment decision-making.

Table 1: Criteria for MDS classification according to the WHO guidelines. A) WHO 2016 classification (Arber et al., 2016), B) WHO 2022 classification (Khoury et al., 2022).

A

WHO 2016

Type	Subtype	Dys-plastic lineages	Cytope-nias[1]	Ring sideroblasts in erythroid elements of BM	Blasts	Cytogenetics
MDS-SLD		1	1 or 2	RS< 15% (or < 5%[2])	PB < 1% BM < 5% No AR	Any, unless fulfills criteria for isolated del(5q)
MDS-MLD		2 or 3	1-3	RS < 15% (or < 5%[2])	PB < 1% BM < 5% No AR	Any, unless fulfills criteria for isolated del(5q)
MDS-RS						
	MDS-RS-SLD	1	1 or 2	RS ≧ 15% (or ≧ 5%[2])	PB < 1% BM < 5% No AR	Any, unless fulfills criteria for isolated del(5q)
	MDS-RS-MLD	2 or 3	1-3	RS ≧ 15% (or ≧ 5%[2])	PB < 1% BM < 5% No AR	Any, unless fulfills criteria for isolated del(5q)
MDS with isolated del(5q)		1-3	1 or 2	None or any	PB < 1% BM < 5% No AR	del(5q) alone or with 1 additional abnormality except -7 or del(7q)
MDS-EB						
	MDS-EB-1	0-3	1-3	None or any	PB 2 ∼ 4% or BM 5 ∼ 9%, no AR	Any
	MDS-EB-2	0-3	1-3	None or any	PB 5 ∼ 19% or BM 10% ∼ 19% or AR	Any
MDS-U						
	With 1% PB blast	1-3	1-3	None or any	PB = 1%[3], BM < 5%, AR	Any
	With SLD and pancyto-penia	1	3	None or any	PB < 1% BM < 5% No AR	Any
	With defining cytoge-netic abnorma-lity	0	1-3	< 15%[4]	PB < 1% BM < 5% No AR	MDS defining abnormality
RCC		1-3	1-3	None	PB < 2% BM < 5% No AR	Any

16

B

Type	Subtype	Blasts	Cytogenetics	Mutations
MDS with defining genetic abnormalities				
	MDS with low blasts and isolated del(5q) (MDS-5q)	BM < 5% and PB < 2%	Del(5q) alone, or with 1 other abnormality other than monosomy 7 or 7q deletion	
	MDS with low blasts and *SF3B1* mutation[a] (MDS-*SF3B1*)		Absence of del(5q), monosomy 7, or complex karyotype	*SF3B1*
	MDS with biallelic *TP53* inactivation (MDS-bi*TP53*)	BM and PB < 20%	Usually complex	Two or more *TP53* mutations, or 1 mutation with evidence of *TP53* copy number loss or cnLOH
MDS, morphologically defined				
	MDS with low blasts (MDS-LB)	BM < 5% and PB < 2%		
	MDS, hypoplastic[b] (MDS-h)			
MDS with increased blasts				
	MDS-IB1	5–9% BM or 2–4% PB		
	MDS-IB2	10-19% BM or 5–19% PB or AR		
	MDS with fibrosis (MDS-f)	5–19% BM; 2–19% PB		

[1]Cytopenias MDS-defining: Hb < 100g/L, PLT < 100 × 10^9/L, ANC < 1.8 × 10^9/L; absolute monocytes count < 1.0 × 10^9/L; [2]with *SF3B1* mutation; [3]1% PB blasts must be recorded on at least two separate observations; [4]If with >15% ring sideroblasts and significant erythroid dysplasia, and SLD are classified as MDS-RS-SLD; [a]Detection of ≥15% ring sideroblasts may substitute for *SF3B1* mutation. Acceptable related terminology: MDS with low blasts and ring sideroblasts; [b]By definition, ≤ 25% bone marrow cellularity, age adjusted. WHO: World Health Organization; MDS: myelodysplastic syndromes/neoplasms; AR: Auer rods; BM: bone marrow; PB: peripheral blood; RS: ring sideroblasts; cnLOH copy neutral loss of heterozygosity; MDS-SLD: MDS with single lineage dysplasia; MDS-MLD: MDS with multilineage dysplasia; MDS-EB: MDS with excess blasts; MDS-U: MSD, unclassifiable; RCC: refractory cytopenia of childhood; MDS-IB: MDS with increased blasts.

Because outcomes of the part of patients do not correspond to their risk estimation, efforts have been made to develop more accurate predictors with a focus on molecular data, such as mutations and gene expression. Several new predictors have been proposed, but none have been implemented in clinical practice (Bejar et al., 2012; Bersanelli et al., 2021; Jiang et al., 2020; Mills et al., 2009; Nazha et al., 2021, 2016; Shiozawa et al., 2017).
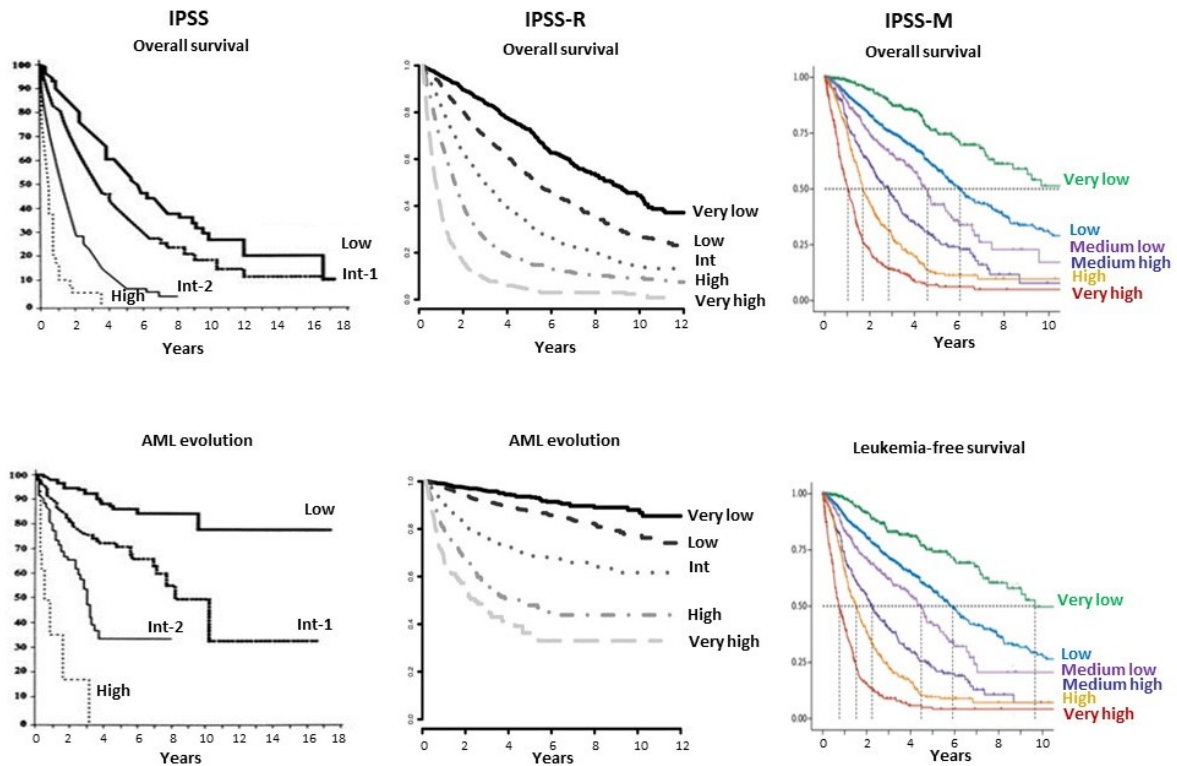
Figure 1. Comparison of overall survival and AML evolution/leukemia-free survival in IPSS, IPSS-R, and IPSS-M. Figures adapted from Bernard et al., 2022; Greenberg et al., 1997 and 2012.

To address the impact of specific gene mutations on survival and disease progression in MDS, the International Working Group for Prognosis in Myelodysplastic Syndromes recently presented the IPSS-molecular (IPSS-M) (Bernard et al., 2022) (Figure 1). It is an innovation of the IPSS-R that includes not only hematological parameters and cytogenetics but also mutational data. The system considers somatic mutations in 16 main genes. Moreover, it distinguishes *TP53* single- and multihit mutations, mutations in *SF3B1* with del(5q) or its comutation with one of 6 certain genes. In addition, the other 15 residual genes were grouped and the number of mutated genes within this group

was incorporated into the risk score calculator. Patients are therefore classified into 6 risk categories. However, this system was published after the publications that form the basis of this thesis, so the IPSS and the IPSS-R are used throughout the thesis.

1.1.2. Genetic factors of the pathogenesis of MDS

The heterogeneity of MDS is well characterized at the morphological and molecular levels. MDS are clonal diseases that affect HSCs. Genetic and nongenetic factors are involved in the pathogenesis of MDS; however, the exact mechanism has not yet been fully elucidated. Due to the topic of the thesis, only genetic factors will be described below.

1.1.2.1.   Cytogenetic aberrations

The first genetic insights into MDS classification and pathogenesis were obtained by cytogenetics. Cytogenetic aberrations are present in half of MDS patients (reviewed in Haase, 2008, and Zahid et al., 2017). Loss of genetic information is much more common than gain. Deletion, monosomies, and unbalanced translocations are common in MDS. The most frequent abnormalities are 5q-, 7-, +8, 20q-, and Y-. In MDS samples, we can see isolated cytogenetic aberrations as well as complex karyotypes (3 and more abnormalities). A complex karyotype is a common feature of an unfavorable prognosis. On the other hand, the presence of isolated 5q-, 20q- aberration or loss of Y chromosome is usually associated with better prognosis. Nevertheless, patients with 5q- involved in complex karyotypes have an extremely poor prognosis (Zemanova et al., 2014).

Currently, conventional karyotyping has begun to be combined with molecular cytogenomic methods. Thus, it is possible to detect cryptic chromosomal changes. In a study by Svobodova et al. (2020), 18% of 214 chromosomal aberrations were cryptic abnormalities. Cryptic abnormalities significantly affect overall survival (OS) (Starczynowski et al., 2008; Svobodova et al., 2020). The implementation of molecular cytogenomic methods into the current cytogenetical procedure should allow more accurate outcome prediction and disease classification (Stevens-Kroef et al., 2017; Svobodova et al., 2020).

The 5q- aberration is the most common cytogenetic abnormality in MDS patients (Haase, 2008; Zahid et al., 2017). It is incorporated into the WHO 2016 and 2022 classifications and the prognostic systems. The deletions vary in length but always include the 5q31 band. However, within this band, there are two different areas that can be deleted (Ebert, 2011). The one with a more telomeric location is connected with a better prognosis and is related to the 5q- syndrome. 5q- syndrome is more common in women and is usually associated with refractory macrocytic anemia, elevated platelets, and mild leukocytopenia. The clinical course is mild and has a very low risk for leukemic transformation. The *RPS14* gene, encoding ribosomal subunit protein, lies within this band and was identified as the causal gene for 5q-syndrome (Boultwood et al., 2007; Ebert et al., 2008). Patients with 5q- have a defect in ribosomal biogenesis and protein translation, and 5q- syndrome thus represents a disorder of impaired ribosomal biogenesis (Pellagatti et al., 2008).

1.1.2.2.   Somatic mutations

Somatic DNA mutations are present in 70-80% of MDS patients, and the most frequently mutated genes encode spliceosomal factors, epigenetic regulators, transcription factors, tumor suppressor *TP53*, or parts of the signal transduction and cohesin complex (Haferlach et al., 2014; Papaemmanuil et al., 2013; Platzbecker et al., 2021) (Figure 2). The heterogeneity of the disease is also reflected in the spectrum of mutations. More than 50 genes are recurrently mutated in MDS (Haferlach et al., 2014; Papaemmanuil et al., 2013); however, no gene is mutated in more than a third of MDS patients (Bersanelli et al., 2021; Papaemmanuil et al., 2013). Early driver mutations (typically affecting genes for splicing and epigenetic regulators) determine the future trajectory of disease evolution with distinct clinical phenotypes (Bersanelli et al., 2021; X. Li et al., 2020; Papaemmanuil et al., 2013). Mutations in transcription factors and signaling molecules usually occur later as passenger mutations and are typically associated with a worse outcome.

The most common mutations occur in genes coding parts of a spliceosomal complex (*SF3B1*, *SRSF2*, *U2AF1*, *ZRSR2*) (Haferlach et al., 2014; Papaemmanuil et al., 2013; Platzbecker et al., 2021). These mutations are considered to result in gain-of-function or neomorphic phenotypes and are mutually exclusive (Dvinge et al., 2016; Haferlach et al., 2014; Papaemmanuil et al., 2013; Walter et al., 2013). *SF3B1* is the most commonly mutated gene in MDS and is typically associated with

the presence of ring sideroblasts in BM (Malcovati et al., 2015). Mutated *SF3B1* has been included in the WHO 2016 classification as the first gene defining a separate entity (Arber et al., 2016). *SF3B1* mutations are also generally described as a predictor of a favorable outcome. However, the position of mutations is probably crucial (Kanagal-Shamanna et al., 2021). Mutations in *SRSF2* and *U2AF1* are, on the other hand, related to shorter survival with an increased risk of progression (Thol et al., 2012; S. J. Wu et al., 2013). Interestingly, mutations in spliceosomal genes, while mutually exclusive with one another, show a strong tendency to co-occur with mutations of specific epigenetic modifiers in MDS, suggesting that abnormalities of these processes may cooperate to give the MDS phenotype (Pellagatti and Boultwood, 2015).

Mutations in genes for epigenetic regulators (regulators of DNA methylation and histone function) usually cause a loss-of-function phenotype (Heuser et al., 2018). Mutations in *DNMT3A* and *TET2* are often present at the onset of disease. *TET2* seems to have no prognostic significance, while *DNMT3A* mutations correlate with an adverse course of MDS (Z. Guo et al., 2017; Hou et al., 2018; Lin et al., 2018, 2017). *ASXL1* mutations are associated with a worse prognosis (Bejar et al., 2011; Chen et al., 2014; Gelsi-Boyer et al., 2012; Hou et al., 2018; Lin et al., 2016). Mutated *EZH2* is associated with a worse outcome in LR-MDS patients (Bejar et al., 2012; Jiang et al., 2020). Mutations in *IDH1* and *IDH2* genes are more common in HR-MDS patients and affect DNA methylation and mitochondrial function (Di Nardo et al., 2016; Patnaik et al., 2012).

Mutated genes encoding signal transduction molecules, transcription factors, and cohesion complexes are usually associated with an adverse course of the disease (Liu et al., 2021; Makishima et al., 2017; Pellagatti and Boultwood, 2015). Recurrent mutations in the signal transduction pathway occur in the *NRAS*, *JAK2*, and *FLT3* genes. Typical mutated genes for transcription factors are *RUNX1* and *ETV6*, and the main representative of genes of the cohesion complex is *STAG2* (Haferlach et al., 2014; Pellagatti and Boultwood, 2015).

The adverse effect of *RUNX1* mutations on the outcome of MDS patients has been widely described (Bejar et al., 2012; Chen et al., 2007; He et al., 2020; Jiang et al., 2020). *RUNX1* encodes a transcription factor critical for embryonic hematopoiesis

and the development of megakaryocytes and platelets in adult hematopoiesis (Ichikawa et al., 2013) and is frequently mutated in hematologic malignancies (Branford et al., 2018; Ichikawa et al., 2013; Sood et al., 2017). In AML and acute lymphocytic leukemia, translocations including this gene are common. *RUNX1* fusion oncoproteins and dysregulated expression of *RUNX* genes are linked to premature senescence (Anderson et al., 2018). Additionally, *RUNX* genes participate in the DNA damage response (DDR) (Ozaki et al., 2013; Tay et al., 2018; D. Wu et al., 2013). In MDS, lower RUNX1 activity predicts the risk of AML transformation and chronic myelomonocytic leukemia (Tsai et al., 2015); thus, the most pathogenic mutations are those that reduce *RUNX1* biological activity.

The tumor suppressor *TP53* is frequently mutated in MDS, and mutations in this gene are often predictive of worse outcome. The adverse effect of these mutations led to more detailed studies on the importance of variant allele frequency (VAF) and allelic state (Belickova et al., 2016; Bernard et al., 2020; Sallman et al., 2015). The prognosis of patients with monoallelic mutations as well as those with mutations with low VAF did not differ from the prognosis of wild-type *TP53* patients.

Similar studies comparing VAF and OS or risk of transformation to AML were performed on other genes, such as *SF3B1* (Malcovati et al., 2011), *RUNX1*, *SRSF2*, and *ZRSR2* (Sallman et al., 2015). In general, a higher VAF leads to a more severe phenotype.

1.1.2.3. *Epigenetics*

Epigenetics includes mechanisms that can affect gene expression without the change in nucleotide sequence, such as DNA methylation and histone modification. The effect can also be caused by mutations in genes encoding proteins involved in these mechanisms. Moreover, noncoding RNAs (ncRNAs) play an important role in epigenetics.
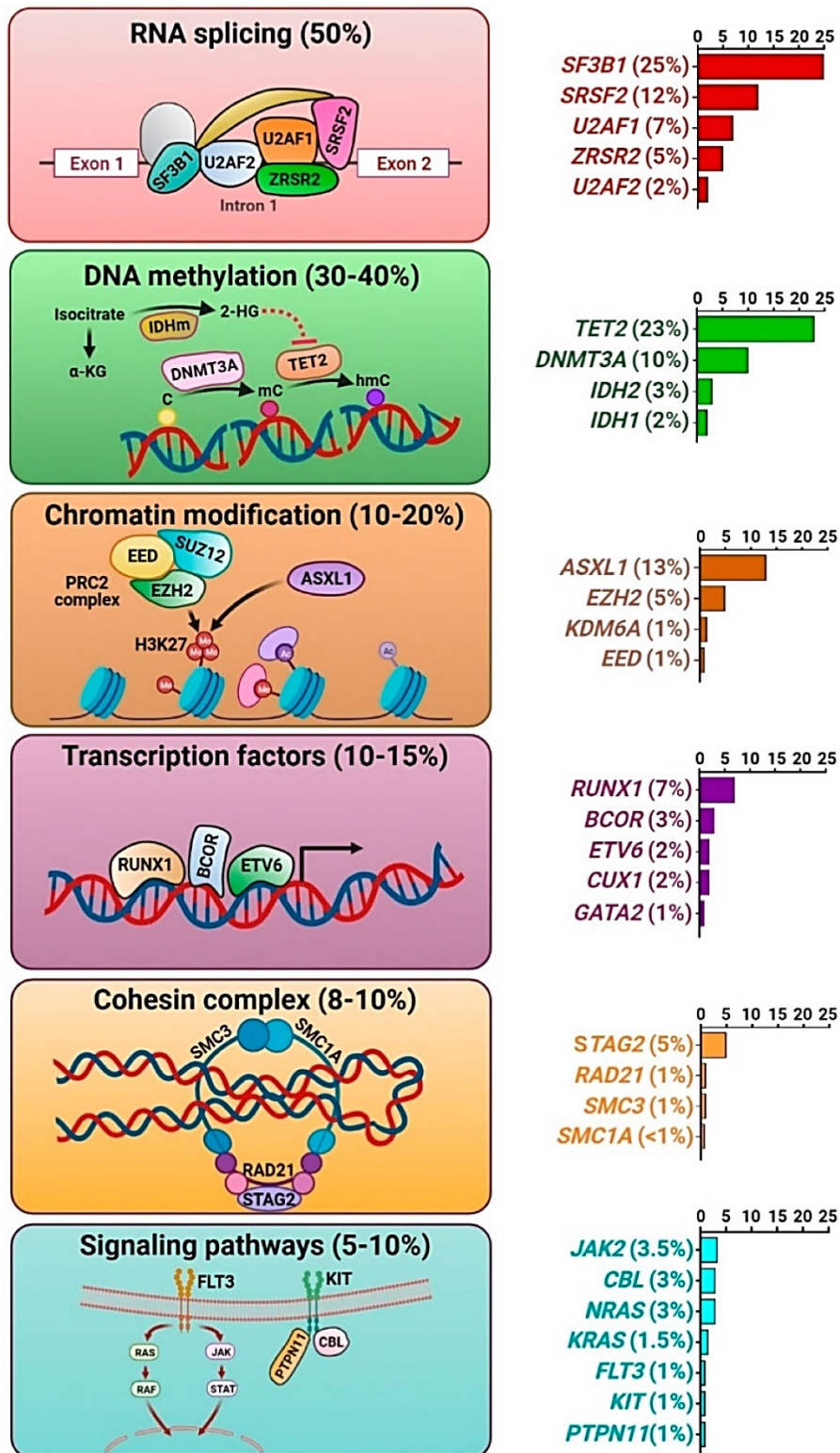
Figure 2. Recurrently mutated genes in MDS. The figure depicts the functional groups of mutations and their frequency in MDS patients. Figure adapted from Saygin and Godley, 2021.

1.1.2.3.1. Aberrant methylation

Aberrant methylation of DNA, especially hypermethylation, is a common feature of MDS cells (Jiang et al., 2009; Zhou et al., 2020). Hypermethylation of CpG promotor islands of tumor-suppressor genes was described as one of the pathogenic mechanisms (Aggerholm et al., 2006; Figueroa et al., 2009). The level of aberrant methylation correlates with the stage of MDS (Jiang et al., 2009). In LR-MDS, 356 differentially methylated regions were described between patients with stable MDS and those with progressive MDS (worsening of the disease within 18 months of diagnosis) (Qin et al., 2019). Furthermore, the most frequently mutated genes encode epigenetic regulators (*DNMT3A*, *TET2*, *ASXL1*, *EZH2*, *IDH1*, and *IDH2*) (Haferlach et al., 2014; Xu et al., 2017). The efficacy of hypomethylating agents in MDS treatment also supports an important role of epigenetic modifications in MDS pathogenesis.

1.1.2.3.2. Noncoding RNAs

Noncoding RNAs are a group of RNAs that are not translated into proteins. The main classes contributing to MDS pathogenesis are microRNAs (miRNAs), long noncoding RNAs (lncRNAs), and piwi-interacting RNAs (piRNAs).

1.1.2.3.2.1. miRNAs

miRNAs are the most characterized group of ncRNAs. They are 21-25 nucleotides (nt) long single-stranded RNA molecules (Bartel, 2004). Based on the degree of miRNA–mRNA complementarity, miRNAs regulate gene expression through cleavage or translational suppression of target messenger RNA (mRNA). They affect many cellular pathways, such as signaling, proliferation, and apoptosis, and their expression has been proven to be deregulated in various types of tumors (Peng and Croce, 2016).

miRNAs are involved in the regulation of all stages of normal hematopoiesis, and thus, their abnormal expression contributes to the development of hematologic malignancies: the tumor suppressor potential of some miRNAs was proposed in chronic lymphocytic leukemia (Calin et al., 2002; Cimmino et al., 2005), and the expression levels of selected miRNAs correlated with cytogenetic aberration and molecular alterations in AML (Cammarata et al., 2010; Garzon et al., 2008; Jongen-Lavrencic et al., 2008; Li et al., 2008).

In MDS, many deregulated miRNAs have been described (reviewed in Veryaskina et al., 2021). They participate in MDS pathogenesis by regulating hematopoiesis, leukocyte migration, and apoptotic processes (Votavova et al., 2011; Wan et al., 2020). For example, the miR-145 gene is located in the common deletion region of 5q- syndrome and affects megakaryocyte and erythroid differentiation through repression of the expression of the protein-coding gene *Fli-1* (Kumar et al., 2011). Notably, miRNA signatures are able to distinguish MDS patients from controls and stratify patients into risk categories (Y. Guo et al., 2017; Merkerova et al., 2011; Pons et al., 2009). Additionally, miRNA profiles and levels of specific miRNAs may be used as predictive markers of response to azacytidine therapy (Krejcik et al., 2018).

Furthermore, miRNAs cooperate with the epigenetic machinery. miRNAs may be silenced by aberrant epigenetic silencing; on the other hand, miRNAs can regulate the expression of MDS-important epigenetic regulators and modifiers, such as *DNMT3A* (Garzon et al., 2009).

1.1.2.3.2.2. LncRNAs

LncRNAs are RNA molecules longer than 200 nt that do not encode proteins. They are transcribed by RNA polymerase II or III and can be spliced and polyadenylated at the 3' end and capped at the 5' end. LncRNAs operate at the transcriptional, translational, and posttranslational levels. XIST is one of the most well-known representatives of lncRNAs. It inactivates the X chromosome in female cells and may also play a role in hematopoiesis (Savarese et al., 2006; Yildirim et al., 2013). Generally, lncRNAs are involved in normal hematopoiesis. They regulate the development of various hematologic cell types and are needed for their normal function. Their dysregulated expression has been reported in various cancers (Chi et al., 2019). Many aberrantly expressed lncRNAs have been recently described in AML (Huang et al., 2019; Hughes et al., 2015; Ma et al., 2020; Wu et al., 2015; Zhang et al., 2014); some lncRNAs have been shown to act as oncogenes, while others act as tumor suppressors. They regulate the cell cycle, apoptosis, proliferation, maturation and differentiation and the expression of transcription factors, tumor suppressors, and oncogenes. The role of lncRNAs alone has been described, as well as lncRNA profiles associated with AML subgroups, cytogenetic aberrations, and mutations. The various mechanisms by which lncRNAs can

regulate cell processes in AML are well reviewed in Ng et al. (2019). In the cytoplasm, lncRNAs can compete with other RNAs, for example, sequestering miRNAs away from their targets or binding to initiation factors of translation and thus suppressing translation. In the nucleus, lncRNAs can recruit transcription factors and epigenetic modifiers, act as scaffolds for the assembly of transcription machinery, and assist in the formation of chromatin structures such as enhancer-promoter loops.

Although many studies have focused on deregulated lncRNAs in AML, few have investigated their effect in MDS. For example, *MEG3* shows tumor-suppressor activity (Benetatos et al., 2010), and its low expression is associated with a poor prognosis in MDS. On the other hand, *KIAA0125* is overexpressed in HR-MDS and is an independent unfavorable prognostic factor for OS and leukemia-free survival (Hung et al., 2021). In 176 MDS patients, high expression of 4 lncRNAs, TC07000551.hg.1, TC08000489.hg.1, TC02004770.hg.1, and TC03000701.hg.1, predicted poor OS (Yao et al., 2017).

Although an increasing number of lncRNAs are being described, they have not been functionally characterized. Few studies have focused on the construction of regulatory networks, including information on lncRNAs in MDS and their targets, genes and miRNAs (Liu et al., 2017; Zhao et al., 2019). These studies showed that aberrantly expressed lncRNAs in MDS are involved in cancer-associated signaling pathways and cellular processes, such as cell proliferation, cell migration, and immune response.

1.1.2.3.2.3. piRNAs

piRNAs are small RNA molecules that are 24-31 nucleotides in length. They evolved as a protective tool to protect the genome against mobilization of transposable elements. In cooperation with piwi proteins, they target actively transcribed transposable elements (TEs), heterochromatinize the region, and degrade TEs (reviewed in Dostalova Merkerova and Krejcik, 2022). Moreover, piRNAs play a role in the viral defense pathway and the silencing of damaged DNA fragments.

Abnormal piRNA expression has been analyzed in hematologic malignancies, such as multiple myeloma, Hodgkin lymphoma, and AML (reviewed in Dostalova

Merkerova and Krejcik, 2022); however, piRNA expression in MDS has been described in only a limited number of studies. Although piRNA expression increases in advanced stages of malignancies, it has been described conversely in MDS: increased levels of piRNAs in early stages of MDS and their decrease during progression (Beck et al., 2011).

Based on a review of previous studies, Dostalova Merkerova and Krejcik (2021) suggested the possible protective mechanism of piRNA activation in MDS. In LR-MDS, the transcription of TEs and the piRNA pathway is activated. This induces the viral defense pathway, which may be responsible for the clearance of leukemic blasts. However, after acquiring additional somatic mutations and developing HR-MDS, TE and piRNA expression is suppressed, leukemic cells may escape immune control and proliferate, and the disease may progress.

1.1.3. Mechanisms of the progression

The development and progression of MDS to AML is suggested to be a consequence of the sequential acquisition of somatic mutations in HSCs (Nolte and Hofmann, 2010). The progression from LR-MDS to HR-MDS and to AML is a continuum. Multiple clones, including the founding clone and subclones derived from the founding clone, are present in MDS BM. The acquisition of favorable mutations and expansion of a subclone during progression is a common phenomenon (Da Silva-Coelho et al., 2017; Kim et al., 2017; Liu et al., 2021; Stosch et al., 2018; Walter et al., 2012). Usually, mutations in subclones associated with progression may be detected months before progression is observed clinically.

The effect of mutations on progression has been intensively studied, and trends in early and later mutated genes have been described (X. Li et al., 2020; Papaemmanuil et al., 2013). Early mutations were identified in genes coding epigenetic regulators (*DNMT3A*, *TET2*, and *ASXL1*) and usually emerge before the clinical phenotypes of MDS. On the other hand, mutations in genes associated with signal transduction and transcription factors occur later during disease and progression, suggesting their involvement in disease evolution. These mutations probably provide a significant proliferative advantage to the cell. In particular,

mutations in the *TP53*, *GATA2*, *KRAS*, *RUNX1*, STAG2, *ASXL1*, *ZRSR2,* and *TET2* genes were described as more frequent in HR-MDS than in LR-MDS. Mutations in *FLT3*, *PTPN11*, *WT1*, *IDH1*, *NPM1*, *IDH2,* and *NRAS* genes were associated with a faster evolution of AML (Kim et al., 2017; X. Li et al., 2020; Makishima et al., 2017; Menssen and Walter, 2020; Nolte and Hofmann, 2010; Shiozawa et al., 2017).

The former model of linear clonal evolution (Walter et al., 2012) is being replaced by a nonlinear model (Chen et al., 2019) (Figure 3). Linear clonal evolution has been proposed on bulk BM sequencing and has described the evolution of the premyelodysplastic stem cell (SC) into MDS SC, which further evolves into leukemic SC. A recent study showed that in MDS as well as AML, SCs had significantly higher subclonal diversity than blasts. This finding indicates that SCs leading to the generation of MDS blasts may be different from those contributing to the progression to AML.

As reviewed by Zhou et al. (2013), MDS is a disease of genomic instability. Therefore, the origin of mutations that lead to progression may be caused by altered DNA damage recognition and repair mechanisms. Furthermore, aberrant DNA methylation accompanies disease progression (Jiang et al., 2009; Nolte and Hofmann, 2010; Stosch et al., 2018; Zhou et al., 2020). Hypermethylation of tumor suppressors may be one of the mechanisms of progression.

By measuring pro-apoptotic (Bax/Bad) and anti-apoptotic (Bcl-2/Bcl-X) Bcl-2-related protein ratios, it was reported that MDS progression arises more likely through inhibition of apoptosis rather than excessive cell growth (Parker et al., 2000). Additionally, deregulation of the immune system may contribute to clonal immune escape and drive progression (Montes et al., 2019).

Despite these discoveries and new technologies, the precise nature of disease progression remains to be elucidated.
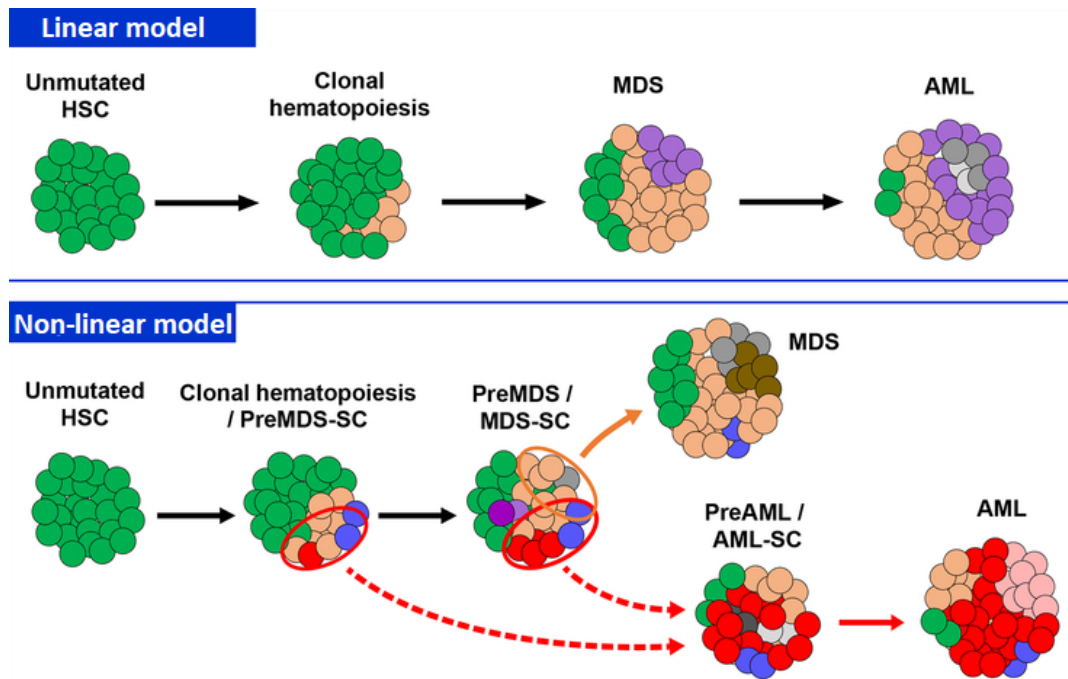
Figure 3. Models of linear and nonlinear clonal evolution during MDS development and AML progression. According to the linear model, HSCs acquire mutations, and the accumulation of mutations subsequently leads to AML progression. On the other hand, the nonlinear model shows HSCs accumulating a variety of mutations, giving rise to a diverse subclonal architecture and generation of MDS blasts different from those contributing to the progression of AML. Mutations that drive MDS blast formation and mutations that drive AML progression originate in parallel in stem cell pool. Figure adapted from Chen et al., 2019.

1.1.4. LR-MDS vs. HR-MDS

Stratification of MDS patients according to their risk of AML transformation is crucial for treatment decision-making and patient management. The hardest goal is to recognize LR-MDS patients who may have a higher likelihood of progression and should be treated appropriately. Few studies have tried to identify features of LR-MDS at risk of rapid progression. The poor prognostic features, including somatic mutations and chromosomal abnormalities, clinical and laboratory parameters, such as age, sex, degree of anemia, neutrophil and platelet counts, transfusion requirement, certain immunophenotypes, and short telomers or germline predispositions, are reviewed in DeZern and Dalton (2022) and Mittelman et al. (2010).

Approximately two-thirds of MDS patients have LR-MDS. LR-MDS are characterized by increased apoptosis, deregulated immunity, and ineffective

myelopoiesis, whereas HR-MDS are characterized by increased cell survival and proliferation (Parker et al., 2000, 1998; Pellagatti et al., 2010).

Lower- and higher-risk patients have distinct cytokine profiles (Kordasti et al., 2009; Lopes et al., 2013). IL-17-producing CD4+ T cells (Th17) are increased in LR-MDS, and the ratio of Th17:regulatory T cells is significantly higher in LR-MDS than in HR-MDS and correlates with increased BM cell apoptosis (Kordasti et al., 2009). In LR-MDS, the proinflammatory state prevails, whereas in HR-MDS, immunosuppression and escape from immune surveillance are dominant (Kordasti et al., 2009; Lopes et al., 2013). Immune deregulation is also connected with apoptosis. Increased expression of cytokines, such as TNF-alpha, triggers apoptosis of the cell through the Fas receptor and its ligand (Gersuk et al., 1998; Shetty et al., 1996).

1.2.   DDR and cellular senescence

DDR represents the cellular reaction to DNA damage. It is a cascade of DNA damage sensors, mediators, transducers, and effectors resulting in a cellular response (Figure 4). The response may be cell-cycle arrest, chromatin remodeling, changes in transcription, repair or bypass of DNA damage, or apoptosis (reviewed in Jackson and Bartek, 2009, and O'Connor, 2015).

DNA damage signaling in precancerous lesions may be the result of replication stress, oxidative damage, or may be induced by oncogene or dysfunctional telomeres (Bartkova et al., 2005; Gorgoulis et al., 2005) (Figure 5). DDR in precancerous cells provides a barrier to uncontrolled cell growth. However, this supports the selective pressure for DDR inactivation. Aberrations in the DDR pathway, which are characterized by genomic instability, accompany tumor development and progression (Bartkova et al., 2005; Gorgoulis et al., 2005) (Figure 5).
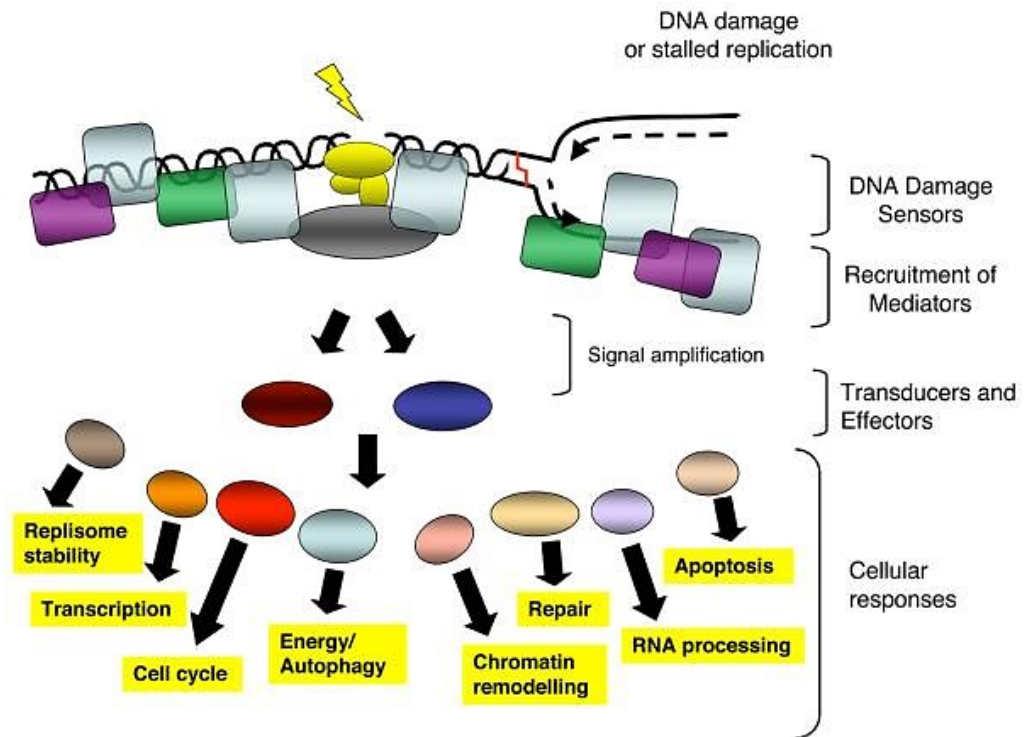
Figure 4. DNA damage response cascade. Lesions in DNA caused by DNA damage or stalled replication forks lead to DDR activation. The DNA lesion is recognized by DNA damage sensors and their presence recruits mediators that activate signal amplification. The DDR signaling pathway affects a variety of cellular processes. Figure adapted from Jackson and Bartek, 2009.

Despite increasing DNA damage from MDS to AML, the DDR is reduced in AML compared to MDS (Boehrer et al., 2009; Popp et al., 2017). Variants in DNA repair genes were described to be associated with an increased risk of MDS (Belickova et al., 2013), and polymorphisms in DNA repair genes contribute to genomic instability in MDS (Ribeiro et al., 2016). The expression of DNA repair genes is deregulated in CD34+ BM cells of MDS patients and presents a specific expression pattern between LR-MDS and HR-MDS (Valka et al., 2017). The most deregulated pathways are excision repair and homologous recombination. The highest levels of *RAD51* and *XRCC2* expression were observed in LR-MDS, and the lowest levels were observed in HR-MDS. These genes belong to the homologous recombination pathway. Poly(ADP-ribose) polymerase 1 inhibitors are designed as targeted therapy for DNA repair-defective tumors. It was shown that they may have therapeutic potential in MDS patients (Faraoni et al., 2019).
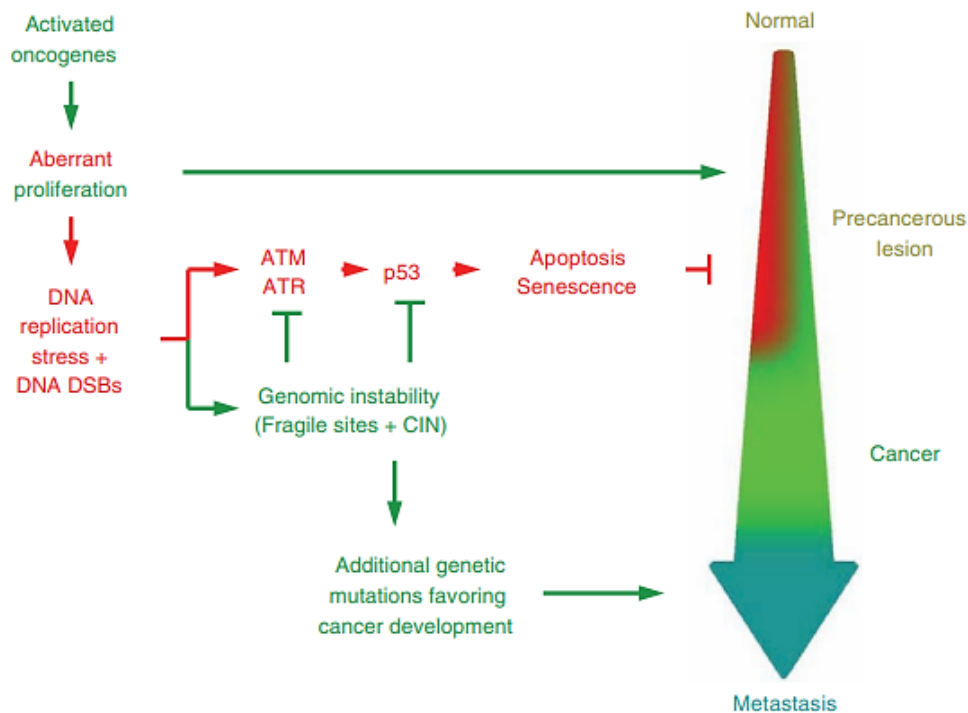
Figure 5. Cancer development and progression according to the oncogene-induced DNA damage model. Activation of oncogenes leads to aberrant proliferation and the emergence of DNA replication stress and DNA damage. DDR and consequentially activated apoptosis and senescence are cellular tumor barrier mechanisms. However, DNA replication stress as well as DNA damage cause genomic instability, inducing additional DNA damage and subsequent cancer development. Figure adapted from Halazonetis et al., 2008.

Long-lasting DDR signaling may result in cellular senescence (Coppé et al., 2008; Feringa et al., 2018). Cellular senescence is a complex mechanism protecting the organism against damage that accumulates in the cell during its life and is closely related to aging (Baker et al., 2011; Hernandez-Segura et al., 2018; Schosserer et al., 2017). It was shown to be upregulated in various preneoplastic lesions and serves as a barrier in tumor development (Acosta et al., 2008; Braig et al., 2005; Chen et al., 2005). Senescence can arise from various causes: shortening of telomeres, DNA damage, mitochondrial dysfunction, activation of oncogenes, chemotherapy, epigenetic stimuli, or oxidative stress. Senescent cells irreversibly stop their cell cycle, although they undergo changes in metabolism, organization, and structure and gain a new metabolic and secretory phenotype (SASP) (Coppé et al., 2008). Senescence-associated interleukins also positively influence DDR signaling through a feedback loop (Acosta et al., 2008). Although senescence has been shown to protect the organism against the emergence of malignant clones,

it can promote chronic inflammation and subsequently cancer or age-associated diseases due to the secretory activity of senescent cells (Coppé et al., 2008; Georgilis et al., 2018; Ortiz-Montero et al., 2017).

Senescence, as well as DDR signaling, has been described to increase in mononuclear cells or CD34+ MDS cells compared to AML (Wang et al., 2009) and decrease with a higher risk score according to the IPSS. In addition, senescence increases in mesenchymal stem cells of patients with MDS (Fei et al., 2014; Ferrer et al., 2013; Geyh et al., 2013) and influences the cell cycle activity of CD34+ HSCs (Geyh et al., 2013).

Senolytics, a new emerging drug class, induce apoptosis of senescent cells or inhibit SASP and could be used in age-associated diseases, chronic degenerative disorders, neurodegenerative diseases, precancerous cells, and chemotherapy-induced senescent cells (reviewed in Gasek et al., 2021; Robbins et al., 2021; Saleh and Carpenter, 2021). Senolytics target senescent cell features, including unique surface markers, specific signaling pathways, biochemical changes, and organelle alterations typical for senescence. The problem is the heterogeneity of senescent cells. However, some senolytics are already being tested in clinical trials.

1.3. Treatment of MDS

The only potentially curative option for MDS patients is HSCT. However, this involves many risks, and for now, only patients with poor prognostic features and good fitness obtain the greatest benefit (De Witte et al., 2017; Jain and Elmariah, 2022). The risk of complications from HSCT over conservative treatment must always be considered.

LR-MDS patients are characterized by single-lineage or multilineage dysplasia and a low BM blast percentage. For patients without symptoms, treatment is not necessary immediately. For symptomatic LR-MDS, treatment is limited to supportive care, including transfusions, growth factors, and very few drugs: lenalidomide, luspatercept, and hypomethylating agents (azacitidine or decitabine; originally intended for HR-MDS, approved for LR-MDS only in the United States) (Carraway and Saygin, 2020; Santini, 2022; Toprak, 2022). Patients with higher-

risk features (poor genetic and cytogenetic characteristics, life-threatening cytopenias, high transfusion burden, persistent blast increase, significant fibrosis, or HMA failure) may be candidates for HSCT (De Witte et al., 2017; Jain and Elmariah, 2022).

For fit HR-MDS patients, intensive chemotherapy followed by HSCT is the main choice (Bewersdorf et al., 2020; Sekeres and Cutler, 2014). For elderly patients and patients without a suitable donor, hypomethylating agents are the only choice, although the treatment has limitations in response rates, and the response is usually only temporary. However, clinical studies have shown new options, e.g., new hypomethylating agents, combinations of hypomethylating agents with the BCL-2 inhibitor venetoclax or with immune checkpoint inhibitors (reviewed in Bewersdorf, Carraway, and Prebet, 2020), or haploidentical lymphocyte infusion together with decitabine-based chemotherapy (Ma et al., 2018).

Targeted therapies are emerging for small subsets of MDS patients with specific somatic mutations (e.g., *TP53*, *IDH1/2*, *FLT3*, spliceosomal genes) and cytogenetic aberrations (Carraway and Saygin, 2020; Pagliuca et al., 2021; Santini, 2022) (Figure 6). However, there is currently only one approved: lenalidomide for patients with del(5q). Other promising targeted therapies in clinical studies for LR-MDS patients target the HIF pathway (roxadustat) and telomerase (imetelstat).
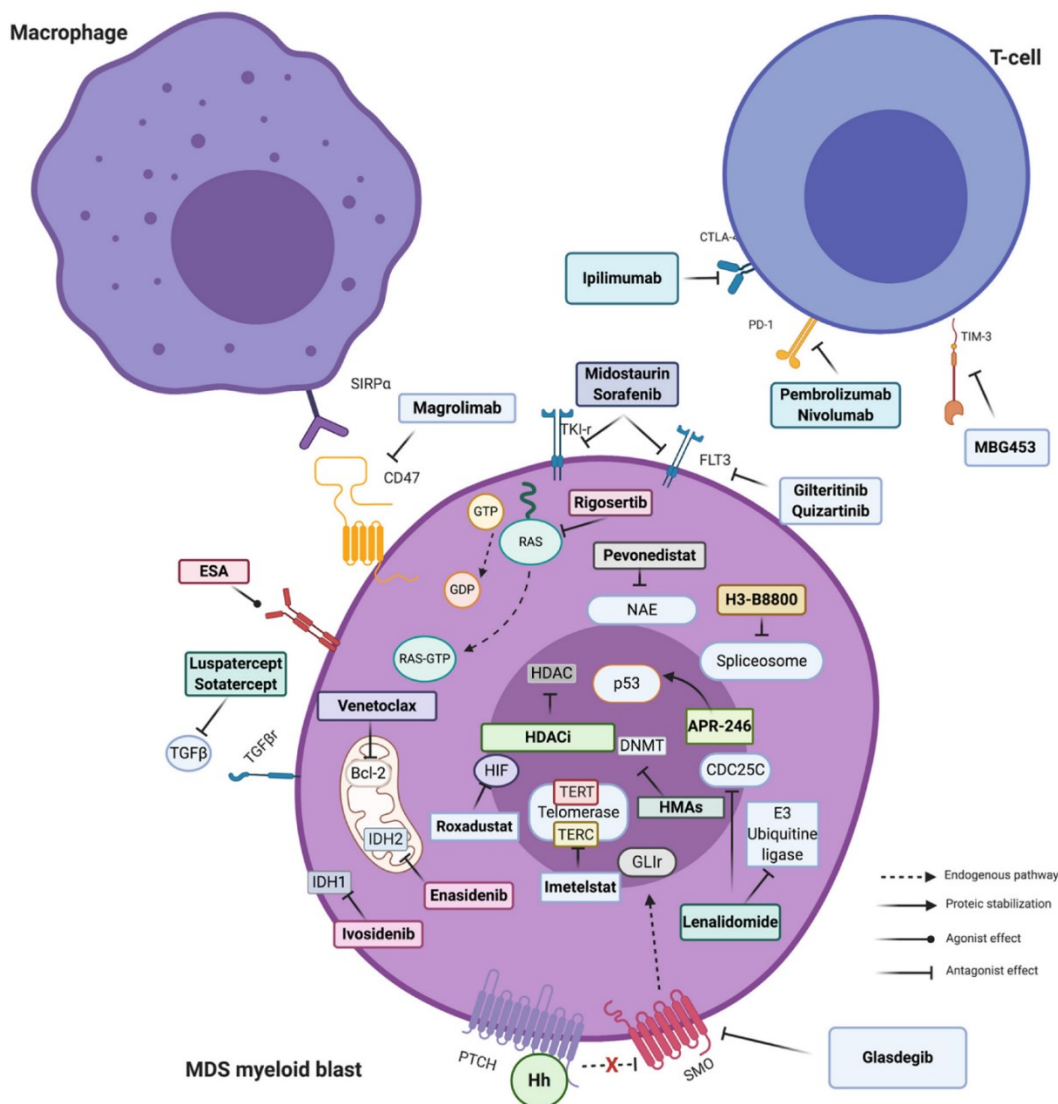
Figure 6. The main treatment targets in MDS. The figure shows ongoing targeted therapies as well as new promising approaches. The mechanism of the effect is indicated by the arrow and connector style. SIRPα: Signal regulatory protein alpha; ESA: Erythropoietin stimulating agent; TGFβ: Transforming growth factor beta; IDH: Isocitrate dehydrogenase; HDAC(i): Histone deacetylases (inhibitor); Hh: Hedgehog polypeptides; PTCH: Protein patched homolog; SMO: Smoothened; HMA: Hypomethylating agents; DNMT: DNA methyl transferase; NAE: Neural Precursor Cell Expressed, Developmentally Down-Regulated 8 (NEDD8)-activating enzyme; GTP: guanosine triphosphate; GDP: Guanosine 5′-diphosphate; FLT3: FMS-like tyrosine kinase 3; PD-1: Programmed cell death 1; CTLA4: cytotoxic T-lymphocyte-associated protein 4; HIF: Hypoxia inducible factor; TERT: Telomerase reverse transcriptase; and TERC: Telomerase RNA component. Figure adapted from Pagliuca et al., 2021.

2.   AIMS OF THE THESIS

Due to the highly variable clinical course of MDS, it is crucial to determine reliable markers of patient outcomes. Especially in low-risk categories, it is very important to identify patients at risk of rapid progression and to choose the most appropriate treatment. The identification of markers associated with rapid progression may further provide deeper insights into the molecular nature of MDS progression and may suggest novel candidates for targeted therapy.

Major aims:

- To identify novel potential biomarkers of adverse outcomes in MDS patients at the DNA and RNA levels
    - To identify deregulated lncRNAs predicting adverse outcomes in MDS patients
    - To identify somatic mutations acting as molecular markers of rapid progression in LR-MDS patients
- To describe the role of these biomarkers in disease development and rapid progression
- To identify the main biological pathways whose deregulation plays a role in rapid progression

3. MATERIAL AND METHODS

The description of material including the patients' samples is detailed in the published papers included in this thesis. Here, I present the list of methods used in the published papers. A detailed description is provided in the Methods and Supplementary Methods sections of the published papers.

DNA/RNA isolation – Publication I and II

Microarray profiling – Publication I

Reverse transcription quantitative PCR (RT-qPCR) – Publication I

Next-generation sequencing (NGS)

– Targeted gene sequencing – Publication I and II

– RNA sequencing – Publication II

Sanger sequencing – Publication II

Flow cytometry – Publication II

Immunohistochemistry – Publication II

Bioinformatics – NGS data processing pipeline – Publication I and II

– LncRNA-PCG coexpression network analysis (network-based lncRNA co-module function annotation method) – Publication I

– Machine learning – Publication II

Statistical analysis – Publication I and II

4. RESULTS

4.1. List of publications and author contribution

Publication I

Katarina Szikszai, Zdenek Krejcik, Jiri Klema, Nikoleta Loudova, Andrea Hrustincova, Monika Belickova, **Monika Hruba***, Jitka Vesela, Viktor Stranecky, David Kundrat, Pavla Pecherkova, Jaroslav Cermak, Anna Jonasova, Michaela Dostalova Merkerova. **LncRNA Profiling Reveals That the Deregulation of H19, WT1-AS, TCL6, and LEF1-AS1 Is Associated with Higher-Risk Myelodysplastic Syndrome.** *Cancers (Basel).* 2020;12(10):2726.
doi:10.3390/cancers12102726

$IF_{2020} = 6.639$

Monika Kaisrlikova (*Hruba is maiden name) performed and interpreted the NGS experiments.

Publication II

Monika Kaisrlikova, Jitka Vesela, David Kundrat, Hana Votavova, Michaela Dostalova Merkerova, Zdenek Krejcik, Vladimir Divoky, Marek Jedlicka, Jan Fric, Jiri Klema, Dana Mikulenkova, Marketa Stastna Markova, Marie Lauermannova, Jolana Mertova, Jacqueline Soukupova Maaloufova, Anna Jonasova, Jaroslav Cermak & Monika Belickova. **RUNX1 mutations contribute to the progression of MDS due to disruption of antitumor cellular defense: a study on patients with lower-risk MDS.** *Leukemia*. 2022;36:1898-1906. doi:10.1038/s41375-022-01584-3

$IF_{2021} = 11.528$

Monika Kaisrlikova performed the NGS experiments, Sanger sequencing, statistical analyses, interpreted the results, and wrote the manuscript.

4.2. Summary of results

The molecular pathogenesis of MDS is a very complex process, and whole genome approaches have enabled us to obtain a comprehensive picture of the MDS genomic landscape and to reveal cellular pathways involved in disease development, progression, and relapse. In this thesis, we focused on two emerging directions, the application of microarray and NGS technologies, for the identification of potential biomarkers of adverse outcomes in MDS patients. We targeted both protein-coding and noncoding regions of the MDS genome to detect pathogenic variants in recurrently mutated genes and deregulated expression of lncRNAs.

The first study aimed to identify lncRNAs predictive of adverse outcomes in a cohort of MDS patients, to characterize lncRNAs deregulated in MDS and particular MDS subtypes and to determine their function in the pathogenesis of the disease. The second study focused on LR-MDS patients as the group that most benefits from an early identification of the risk of rapid progression. The biomarkers of rapid progression were searched within the mutated genes. Furthermore, both studies identified deregulated biological pathways associated with adverse outcomes and suggested possible mechanisms of disease development and progression.

In Publication I, we examined CD34+ BM cells of 54 MDS patients, 14 patients with AML with myelodysplasia-related changes (AML-MRC), and 9 healthy controls as a discovery cohort for microarray profiling and 79 MDS, 14 AML-MRC, and 13 healthy controls as a testing cohort for RT-qPCR experiments. Differentially expressed lncRNAs and protein-coding genes (PCGs) were analyzed in relation to MDS, its subtypes and risk categories, and gene mutations. Functional changes were assessed performing Gene Set Enrichment Analysis (GSEA). LncRNA-PCG coexpression network analysis was performed, and the extracted modules were functionally annotated to Gene Ontology (GO) terms. LncRNAs whose expression correlated with the expression of the core PCGs were suggested to be related to deregulated processes associated with these PCGs. Thus, we were able to recognize the potential association of lncRNAs and several deregulated pathways.

In MDS patients compared to controls, 32 lncRNAs and 87 PCGs were significantly deregulated (|logFC| > 1, FDR < 0.05). MDS samples showed enrichment in gene sets of hemoglobin complex and oxygen transport, immune response, epigenetic modifications, and regulation of gene expression. We recognized lncRNAs potentially associated with pathways of oxygen transport machinery and other lncRNAs potentially associated with metabolic cellular processes such as transcription, translation, and protein catabolism.

The expression profiles of the lncRNAs were also compared between MDS subtypes. Interestingly, the lncRNA profiles of MDS-SLD patients resembled those of healthy controls more than those of patients with other MDS subtypes. MDS-EB-1 clustered with early MDS subtypes, such as MDS-MLD, MDS-RS, and MDS with del(5q). On the other hand, MDS-EB-2 clustered together with AML-MRC. A gradual widening of the disparities was observed in the expression profiles of healthy individuals to the advanced stages of MDS.

We aimed to find associations between lncRNA expression and somatic mutations. Mutational screening was performed in 64 patients from the discovery cohort. We analyzed differences by the presence of mutations in each of the 5 most frequently mutated genes (*SF3B1*, *TET2*, *TP53*, *DNMT3A*, and *RUNX1*) in our cohort. The number of deregulated lncRNAs and PCGs (106 lncRNAs and 646 PCGs) was exceptionally higher in *RUNX1*-mutated samples than in patients with the other abovementioned mutations. Therefore, we focused on *RUNX1*-mutated patients (n = 9). We observed deregulation of signaling pathways, immune response, and cell death pathways compared to MDS patients without *RUNX1* mutations. Furthermore, genes in the coexpression network were enriched in translational regulation, RNA splicing, cell cycle, DNA repair, and DNA recombination. The deregulated PCGs *LEF1* and *RAG1* and the lncRNAs *LEF1-AS1* and *TCL6* in *RUNX1*-mutated patients were related to DNA repair, DNA recombination, and the p53 pathway.

The main goal of this work was to determine lncRNAs associated with worse outcomes. First, we compared the transcriptomic data of 31 patients with short survival (OS < 18 months) and 25 patients with long survival. Eight lncRNAs and 29 PCGs were significantly deregulated (|logFC| > 1, FDR < 0.05) in patients with

short survival. Two well-known tumorigenic lncRNAs, *H19* and *WT1-AS*, were significantly upregulated in patients with short survival. Secondly, we compared LR (very low, low) (n = 26) with HR (high, very high) (n = 19) patients; 16 lncRNAs and 82 PCGs with differential expression were detected in HR. Among them, *TCL6*, a lncRNA with a known oncogenic association, and *LEF1-AS*, which is associated with hematopoiesis, were downregulated. Between LR and HR patients, gene silencing, immune response, cell differentiation and proliferation, motility, and angiogenesis pathways showed differential expression. Three modules were identified in the coexpression network. The IPSSR_1 module was associated with cell differentiation, growth, adhesion, and migration. All genes within this module were downregulated in HR, suggesting that the maintenance and differentiation processes of HSCs are attenuated in HR. IPSSR-2 and 3 were related mainly to epigenetic modification and chromatin structure and included *LEF1*, *LEF1-AS1*, and *TCL6*. Genes from all modules were also involved in the immune system.

For prognostic purposes, we chose 4 candidate lncRNAs, *H19*, *WT1-AS*, *TCL6*, and *LEF-AS1*, as possible prognostic biomarkers and performed analysis in two independent cohorts (discovery and testing cohorts). The expression of these four lncRNAs gradually increased (*H19*, *WT1-AS1*) or decreased (*LEF1-AS*, *TCL6*) from healthy controls to HR-MDS patients. The expression levels of these lncRNAs were significant for OS and PFS. In multivariate analysis, high blast count, high level of *H19*, and the presence of somatic mutation in *TP53* were independent prognostic variables for MDS outcome. Furthermore, age, platelet count, *TCL6* and *LEF1-AS1* levels added prognostic value to these main predictors. Focusing on lower-intermediate patients (IPSS-R < 4.5), only the *H19* expression level was highly significant for OS and PFS. Age, platelet count, *LEF1-AS1,* and *TCL6* levels were less significant.

Finally, we focused on the *cis* and *trans* regulatory activities of these four prognosis-related lncRNAs. *Cis* regulatory activity affects the expression of neighboring genes, and *trans* regulatory activity regulates distant genes. We observed disrupted regulation of the expression of *IGF2*, a PCG located in proximity to *H19*, and miR-675, which uses *H19* as a primary template, in HR-MDS patients. Discordant expression was associated with a worse outcome.

To address the *trans* regulatory mechanisms, we constructed coexpression networks with these four lncRNAs as the central nodes. Only two modules were generated: *LEF1-AS1/TCL6* enriched in cell adhesion and differentiation processes and *H19/WT1-AS1* associated with chromatin modification, cytokine response, cell proliferation, and cell death. Our results suggested that these two pairs might be functionally related.

In Publication II, we focused on LR-MDS patients who generally have a good prognosis. This very group needs enhancing of the risk stratification and to identify the patients at risk of rapid progression. We examined the mutational profile of genes associated with hematologic malignancies in these patients at diagnosis and tested their prognostic value. Furthermore, we analyzed the transcriptome of MDS patients bearing somatic mutations associated with unfavorable prognosis to define the molecular mechanisms that contribute to rapid progression.

We applied NGS targeted sequencing using the TruSight Myeloid Sequencing Panel (Illumina), focusing on genes frequently mutated in hematological malignancies. The cohort consisted of 214 LR-MDS patients according to the IPSS. We sequenced DNA from bone marrow or peripheral blood diagnostic samples.

At least one mutation was detected in 64% of patients. The most commonly mutated gene was *SF3B1*, although the most frequently mutated genes in terms of functional categories were epigenetic regulators. The number of mutations per patient varied from 0 to 9, and the number of mutations significantly affected OS and PFS. In univariate analysis for OS and PFS, platelet count, male sex, age, and the presence and total number of mutations were significant ($p < 0.05$). For OS, mutations in *DNMT3A*, *RUNX1*, *SETBP1*, *STAG2*, and *TP53* were significant, while for PFS, significantly mutated genes were *RUNX1*, *SETBP1*, *STAG2*, *TP53*, and *U2AF1*. The effect of mutational data on the prediction of survival was also confirmed by machine learning using two independent methods: stepwise backward feature selection and elastic network models. These methods identified genes responsible for shorter OS and PFS. Both methods identified *SETBP1*, *TP53*, and *RUNX1* as the genes most responsible for shorter PFS when mutated. In the multivariate analysis for PFS, platelet count, age, and mutated *RUNX1* were the most significant independent prognostic factors. The effect of *RUNX1* mutations

on shortened PFS indicated its potential significance as a marker of rapid progression.

RUNX1 was also the most commonly mutated gene in patients who progressed within 5 years (n = 41) compared to those who did not progress and were followed for at least 5 years (n = 53). Additionally, the mutational load in the range of all analyzed genes was higher in the patients who progressed rapidly (median 3 mutations, range 0-8 versus median 0, range 0-5). Mutational burden also increased during progression. We observed a higher VAF and number of mutations when comparing paired samples of 36 patients who progressed within 5 years.

Furthermore, the implementation of the *RUNX1* mutational status significantly improved the prognostic discrimination by IPSS-R. Patients with *RUNX1* mutations had a significantly different percentage of BM blasts and platelet levels, as well as a significantly different median number of mutations compared to the other patients in our cohort.

Following these results, we further studied the impact of *RUNX1* mutations on the regulation of cellular pathways. We aimed to understand the changes at the molecular level underlying the progression in *RUNX1*-mutated patients. We compared the transcriptomes of CD34+ cells from 8 *RUNX1*-mutated LR-MDS patients (mutR-LR) and 29 LR-MDS patients without *RUNX1* mutations (wtR-LR). A total of 2235 genes were significantly (FDR < 0.05) upregulated and 2094 were significantly downregulated in mutR-LR according to the differential expression analysis. According to the GO BP and KEGG databases, the pathways of chromatin and gene silencing, nucleosome assembly, chromatin organization, regulation of megakaryocyte differentiation, myeloid cell differentiation, and hemopoiesis, telomere organization and capping, cellular metabolic processes, DDR, cellular response to stress, cellular senescence, aging, chronic inflammation, and oxidative stress were downregulated in mutR-LR. These pathways play a crucial role in cellular tumor protection. Pathways upregulated in mutR-LR were related to cancer and leukemia.

Next, we performed GSEA on our custom dataset consisting of 88 gene sets related to DDR, DNA repair, cellular senescence, apoptosis, and hypoxia. Eighty-two gene sets (93%) were significantly enriched in wtR-LR (FDR < 0.1).

When comparing the expression profiles of LR-MDS to HR-MDS (n = 20), we observed a greater resemblance of mutR-LR with HR-MDS than with wtR-LR at diagnosis.

Finally, we aimed to validate the suppression of DDR and cellular senescence at the protein level by immunohistochemical staining of γH2AX protein on BM formalin-fixed paraffin-embedded sections and fluorescence detection of senescence-associated β-galactosidase (SA-β-gal) activity in BM sorted cells. γH2AX is a marker of DNA damage and repair, whereas SA-β-gal activity is a marker of senescent cells. We observed a higher level of γH2AX in wtR-LR (n = 4) than in mutR-LR (n = 3). Moreover, we detected significantly higher SA-β-gal activity in CD14+ monocytes of wtR-LR (n = 6) compared to those of HR-MDS (n = 6). Although mutR-LR samples were not available for this assay, based on the highly similar expression profiles of senescence-associated pathways in mutR-LR and HR-MDS described above, we anticipated similar results in mutR-LR.

4.3. Publication I

cancers

Article

LncRNA Profiling Reveals That the Deregulation of H19, WT1-AS, TCL6, and LEF1-AS1 Is Associated with Higher-Risk Myelodysplastic Syndrome

Katarina Szikszai [1], Zdenek Krejcik [1], Jiri Klema [2], Nikoleta Loudova [1], Andrea Hrustincova [1,3], Monika Belickova [1], Monika Hruba [1,4], Jitka Vesela [1], Viktor Stranecky [4], David Kundrat [1], Pavla Pecherkova [1], Jaroslav Cermak [1], Anna Jonasova [5] and Michaela Dostalova Merkerova [1,*]

[1] Institute of Hematology and Blood Transfusion, U Nemocnice 1, 128 20 Prague 2, Czech Republic; katarina.szikszai@uhkt.cz (K.S.); zdenek.krejcik@uhkt.cz (Z.K.); nikoleta.loudova@uhkt.cz (N.L.); andrea.mrhalkova@uhkt.cz (A.H.); monika.belickova@uhkt.cz (M.B.); monika.hruba@uhkt.cz (M.H.); jitka.vesela@uhkt.cz (J.V.); david.kundrat@uhkt.cz (D.K.); pavla.pecherkova@uhkt.cz (P.P.); jaroslav.cermak@uhkt.cz (J.C.)

[2] Department of Computer Science, Czech Technical University, 121 35 Prague, Czech Republic; klema@fel.cvut.cz

[3] Faculty of Science, Charles University, 128 00 Prague, Czech Republic

[4] First Faculty of Medicine, Charles University, 121 08 Prague, Czech Republic; vstra@lf1.cuni.cz

[5] General University Hospital, 128 08 Prague, Czech Republic; Anna.Jonasova@vfn.cz

* Correspondence: michaela.merkerova@uhkt.cz; Tel.: +420-221-977-231

check for updates

Simple Summary: Although lncRNAs have been increasingly recognized as regulators of hematopoiesis, only several studies addressed their role in myelodysplastic syndrome (MDS). By genome-wide profiling, we identified lncRNAs deregulated in various groups of MDS patients. We computationally constructed lncRNA-protein coding gene networks to associate deregulated lncRNAs with cellular processes involved in MDS. We showed that expression of H19, WT1-AS, TCL6, and LEF1-AS1 lncRNAs associate with higher-risk MDS and proposed processes related with these transcripts.

Abstract: Background: myelodysplastic syndrome (MDS) is a hematopoietic stem cell disorder with an incompletely known pathogenesis. Long noncoding RNAs (lncRNAs) play multiple roles in hematopoiesis and represent a new class of biomarkers and therapeutic targets, but information on their roles in MDS is limited. Aims: here, we aimed to characterize lncRNAs deregulated in MDS that may function in disease pathogenesis. In particular, we focused on the identification of lncRNAs that could serve as novel potential biomarkers of adverse outcomes in MDS. Methods: we performed microarray expression profiling of lncRNAs and protein-coding genes (PCGs) in the CD34+ bone marrow cells of MDS patients. Expression profiles were analyzed in relation to different aspects of the disease (i.e., diagnosis, disease subtypes, cytogenetic and mutational aberrations, and risk of progression). LncRNA-PCG networks were constructed to link deregulated lncRNAs with regulatory mechanisms associated with MDS. Results: we found several lncRNAs strongly associated with disease pathogenesis (e.g., H19, WT1-AS, TCL6, LEF1-AS1, EPB41L4A-AS1, PVT1, GAS5, and ZFAS1). Of these, downregulation of LEF1-AS1 and TCL6 and upregulation of H19 and WT1-AS were associated with adverse outcomes in MDS patients Multivariate analysis revealed that the predominant variables predictive of survival are blast count, H19 level, and TP53 mutation. Coexpression network data suggested that prognosis-related lncRNAs are predominantly related to cell adhesion and differentiation processes (H19 and WT1-AS) and mechanisms such as chromatin modification, cytokine response, and cell proliferation and death (LEF1-AS1 and TCL6). In addition, we observed that transcriptional regulation in the H19/IGF2 region is disrupted in higher-risk MDS, and discordant expression in this locus is associated with worse outcomes. Conclusions: we identified specific

lncRNAs contributing to MDS pathogenesis and proposed cellular processes associated with these transcripts. Of the lncRNAs associated with patient prognosis, the level of H19 transcript might serve as a robust marker comparable to the clinical variables currently used for patient stratification.

Keywords: myelodysplastic syndrome; lncRNA; progression; outcome; pathogenesis; coexression network

1. Introduction

Myelodysplastic syndrome (MDS) is a heterogeneous group of clonal hematopoietic stem cell (HSC) disorders characterized by bone marrow (BM) dysplasia with ineffective hematopoiesis, peripheral blood cytopenia, and an increased tendency for transformation to acute myeloid leukemia (AML). Pathogenesis of MDS is a multifactorial process in which cytogenetic aberrations, gene mutations, and epigenetic changes are involved. Using the WHO classification criteria [2], MDS patients are classified into various diagnostic subtypes based on the number of affected hematopoietic lineages, the percentage of BM blasts, cytogenetics and the presence of ring sideroblasts. Patient prognosis is evaluated based on the Revised International Prognostic Scoring System (IPSS-R) depending on similar clinicopathological criteria [2]. For higher-risk MDS, treatment with the hypomethylating agent azacitidine (AZA) is currently considered standard therapy, which prolongs patient survival, improves clinical outcomes and quality of life, and delays the disease progression in a proportion of patients [3].

Long noncoding RNAs (lncRNAs) are a group of RNAs that are defined as non-protein-coding transcripts longer than 200 nucleotides. The properties of lncRNAs, such as stability and tissue specificity, make them highly promising diagnostic and prognostic markers as well as interesting therapeutic targets. Although lncRNAs are increasingly recognized as regulators of normal and aberrant hematopoiesis, only several studies have addressed their expression and function in relation to MDS. For example, Liu et al. profiled lncRNA expression and identified several lncRNAs (linc-ARFIP1-4, linc-TAAR9-1, lincC2orf85, linc-RNFT2-1 and linc-RPIA) deregulated in MDS with excess blasts II (MDS-EB2) [4]. Further, Yao et al. established a 4-lncRNA risk score significantly associated with patient survival [5].

Although increasing numbers of deregulated lncRNAs are currently being described in MDS, their functional characterization is still difficult. Transcriptomic data may be used to construct coexpression networks of similarly regulated lncRNAs and protein-coding genes (PCGs), which enables the functional analysis of lncRNAs with unknown functions [4]. This approach uses the "guilt-by-association" strategy working with the principle that genes with related functions tend to have similar expression profiles. Thus, the PCGs in a coexpression module are associated with signaling pathways and Gene Ontology terms, attributing the same functions to the unknown lncRNAs in the network.

In this study, we used a microarray platform to profile lncRNA and PCG expression in parallel in CD34+ BM cells of MDS patients with an emphasis on the identification of lncRNAs with altered levels in various groups of MDS patients. In particular, we aimed to characterize the lncRNAs that could serve as novel potential biomarkers of adverse outcomes in MDS. Moreover, a computational approach for constructing lncRNA-PCG networks was applied to associate these lncRNAs with regulatory mechanisms associated with MDS.

2. Patients and Methods

The study included CD34+ BM cells of 183 patient or control samples randomly divided into a discovery cohort (54 MDS patient samples, 14 AML with myelodysplasia-related changes (AML-MRC) patient samples, and 9 healthy control samples) used for the microarray profiling and a testing cohort (79 MDS patient samples, 14 AML-MRC patient samples, and 13 healthy control samples) used for reverse transcription quantitative PCR (RT-qPCR). Informed consent was obtained from all individuals and the study was approved by the Institutional Scientific Board and the IHBT ethic committee on

16/06/2016 (ethic code: EK3/AZV/06/2016) and performed in accordance with the ethical standards of the Declaration of Helsinki. The detailed clinical and laboratory characteristics of both cohorts, including the classification of MDS patients into subgroups, IPSS-R categories, bone marrow features and blood counts, are summarized in Table S1.

Expression profiles were determined using Agilent Human GENCODE Custom lncRNA Expression Microarray Design [6], consisting of probes for 22,001 lncRNA transcripts and 17,535 PCG mRNAs. Bioinformatical analyses were performed with the Bioconductor project in the R statistical environment using the limma package. The raw and normalized data have been deposited in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database under accession number GSE145733.

RT-qPCR was applied using the TaqMan gene expression system (Thermo Fisher Scientific, Waltham, MA, USA) to measure individual transcript levels (lncRNAs: CHRM3-AS2, EPB41L4A-AS1, H19, LEF1-AS1, PVT1, TCL6, and WT1-AS; PCGs: IGF2, LEF1, WT1, TCL1A, and TCL1B; miRNAs: miR-675 and RNU48 as a reference). For normalization of lncRNA data, several reference genes were tested using the RefFinder tool [7] (Figure S1) and the data were finally normalized to the level of HPRT1.

The TruSight Myeloid Sequencing Panel Kit (Illumina, San Diego, CA, USA) containing 568 amplicons in 54 genes was used for mutational screening. Variants were detected by LoFreq and annotated using Variant Effect Predictor, and their clinical significance was verified in several genomic databases (UCSC, COSMIC, ExAC, and PubMed). The arbitrary cut-off was set at 5% of the variant allele frequency (VAF).

The functional changes related to deregulations in gene expression were assessed using gene set enrichment analysis (GSEA) [8]. The lncRNA-PCG coexpression network analysis was carried out as introduced in [4]. Briefly, we identified differentially expressed lncRNAs and PCGs (FDR < 0.05) and constructed a correlation matrix for these transcripts. Then, non-negative matrix factorization (NMF) was used to extract modules from the correlation matrix and each module was functionally annotated by mapping of PCGs to GO terms. The representative cores of the individual modules were visually plotted (for only 4 lncRNAs and 13 PCGs with the highest module membership).

All other statistical analyses were performed using GraphPad Prism 7 (GraphPad Software, La Jolla, CA, USA) and SPSS software (IBM, Armonk, NY, USA).

The detailed version of the Methods is included in the manuscript as a supplementary file (Supplementary Methods).

3. Results

3.1. MDS-Specific Transcriptome

The gene expression profiles of PCGs and lncRNAs were examined in the CD34+ BM cells of MDS patients. The discovery cohort used for the microarray profiling included 54 patients with MDS, 14 patients with AML-MRC, and 9 healthy donors (Table S1). In summary, we detected a signal of 29,604 transcript probes (out of 61,538 probes spotted on the array). After merging sequence duplicates, probes for 12,444 PCGs and 14,518 lncRNAs were detected. To compare the effect of PCGs and lncRNAs on the disease, we analyzed the data for both categories of transcripts separately.

First, we evaluated gene expression changes between MDS patients and healthy individuals. In MDS, we found 32 lncRNAs (28 upregulated/4 downregulated) and 87 PCGs (83 upregulated/4 downregulated) significantly deregulated compared to those in controls ($|logFC| > 1$, FDR < 0.05) (Table S2). Of the functionally described lncRNAs, we detected the upregulation of H19, EMCN-IT1, WT1-AS, MEG8, and PVT1 and the downregulation of ST6GAL2-IT1 and U3. Second, the expression of 11 lncRNAs (all 11 downregulated, e.g., VPS9D1-AS1, PVT1, and CXADRP3) and 161 PCGs (2 upregulated/159 downregulated) was significantly changed in AML-MRC compared to that in MDS (Table S3).

Gene set enrichment analysis (GSEA) identified 12 gene sets enriched (Figure 1A) in MDS CD34+ cells compared to healthy cells. Overall, the enriched processes were mainly related to four mechanisms:

(i) hemoglobin complex and oxygen transport, (ii) immune response, (iii) epigenetic modifications, and (iv) regulation of gene expression. To assess the functions of the deregulated lncRNAs, we built the lncRNA-PCG coexpression network. Within this network, we defined individual modules (assigned as MDS modules) and recognized those associated with the abovementioned processes enriched in MDS. In the cores of these modules, we identified lncRNAs whose expression was substantially correlated with the expression of core PCG nodes and might therefore be linked to deregulated processes. Table 1 summarizes enriched processes and core PCGs and lncRNAs within selected modules and Figure 1B shows an illustrative module of the lncRNA-PCG coexpression network (i.e., MDS_1 module).

A. GSEA: MDS patients *vs.* healthy donors

Gene sets	NES	*p*
Enriched in MDS		
HDAC targets silenced by methylation UP (Heller *et al.*)	1.87	<0.001
Heme metabolism (Hallmark)	1.85	<0.001
Immune response (GO)	1.79	0.003
Immune system process (GO)	1.67	0.005
Transition metal ion binding (GO)	1.71	0.006
Defense response (GO)	1.71	0.007
Response to cytokine (GO)	1.63	0.016
Cellular response to organic substance (GO)	1.54	0.025
EZH2 targets UP (Nuytten *et al.*)	1.52	0.032
Apoptosis by doxorubicin UP (Graessmann *et al.*)	1.52	0.034
Transporter activity (GO)	1.53	0.036
Enriched in controls		
Regulation of protein modification process (GO)	−1.60	0.018

GO term	*p* value
Hemoglobin complex	3.62×10^{-4}
Oxygen transport	4.82×10^{-4}
Iron ion binding	6.81×10^{-5}
Solute sodium symporter activity	7.52×10^{-3}

Figure 1. Pathway analysis of genes deregulated between MDS patients and healthy donors. (**A**) Gene set enrichment analysis (GSEA) of differentially expressed PCGs and four enrichment plots for selected enriched gene sets. NES - normalized enrichment score. References: Heller et al. [9], Nuytten et al. [10], and Graessmann et al. [11]. (**B**) MDS_1 module from the lncRNA-PCG coexpression network. The Gene Ontology (GO) terms significantly associated with these modules are listed in the corresponding table. Square-PCG, circle-lncRNA, red-upregulated in MDS.

Table 1. Characteristics of selected modules in lncRNA-PCG coexpression networks based on differentially expressed genes between (**A**) MDS patients and healthy controls, (**B**) patients with isolated del(5q) and patients with normal karyotype, (**C**) RUNX1-mutated and RUNX1-wild type patients, and (**D**) MDS patients with lower- and higher-risk IPSS-R. Enrichment analysis was done to associate module PCGs with GO terms ($p < 0.01$). The representative core PCGs and lncRNAs with the highest module membership are listed.

DEA	Module	Associated GO Terms	Core Nodes
A. MDS vs. CTR	MDS_1	Hemoglobin complex, oxygen transport, iron ion binding, solute sodium symporter activity	PCGs: A_33_P3261428, A_33_P3328284, AC131097, ARG2, C20orf108, GABRP, LOC100128548, MOSPD1, RUNDC3A, SERPING1, SLC6A8, TRIM10 lncRNAs: AC131097, PRKAR2A-AS1, RP4-669L17
	MDS_2	Hemoglobin complex, oxygen transport, ganglioside biosynthetic process, sialylation	PCGs: DIAPH1, EPB42, GYPB, HBD, HBBP1, OSBP2, PQLC1, SUB1, TCM2, USE1, WDFY2 lncRNAs: CTD-2319112, P11-640M9.1, RP11-640M9.1, RP11-556A11
	MDS_3	Golgi apparatus, protein polyubiquitination, proteasomal protein catabolic process	PCGs: AK8, C11orf2, EBP1, FBXO6, GABPB2, GLT8CR2, LGALS8, LOC644285, M54A6A, RAB20, RPL34, THC2496362, TMEM104 lncRNAs: AC021188, ZFAS1
	MDS_4	Translational initiation, mRNA metabolic process, cytosolic ribosome, protein localization to endoplasmic reticulum, nuclear transcribed mRNA catabolic process nonsense mediated decay, SMAD protein signal transduction, poly-A-RNA binding, cytoplasmic translation, aminoacid activation, ribonucleoprotein complex biogenesis	PCGs: A_24_P24724, ANKRD42, C11orf2, EIF4B, GRMIN5, GLT8CR2, HARS, LARS, RBM27, RPL29, RPS14, SLC26A2 lncRNAs: CSNK1A1P1, EPB41L4A-AS1
	del(5q)_1	RNA binding, translational initiation, ribonucleoprotein complex, protein localization to endoplasmatic reticulum, nuclear transcribed mRNA catabolic process nonsense mediated decay, protein targeting to membrane, ribosome, large ribosomal subunit	PCGs: A_24_P59247, A_33_P3358856, DNAJC18, HARS, HNRNPA0, HSPA9, IMPDH2, MRPL45, NOA1, RPL29P2, RPS14, TLR3, ZCCHC10 lncRNAs: EPB41L4A-AS1
B. Patients with isolated del(5q) vs. patients with normal karyotype	Del(5q)_2	Intracellular receptor signaling pathway, STAT cascade, positive regulation of immune system process, regulation of VEGFR signaling pathway, tyrosine phosphorylation of STAT protein, regulation of cell activation, cytokine activity, JAK/STAT cascade involved in growth hormone signaling pathway	PCGs: A_24_P143683, BRD8, C5orf56, FBXL17, FPGT, MYADM, RBM10, RNF139, SLC23A1, SFTLC1, UQCR10, YTHDC2 lncRNAs: NUTM2A-AS1, RP11-506M13.3, STARD4-AS1
	del(5q)_3	Phosphatase inhibitor activity, platelet morphogenesis, platelet activation, hemostasis, activation of MAPK activity	PCGs: ANK1, C18orf10, CTNNBL1, DNAJC6, EPOR, KLF1, MINPP1, NMU, PCYT1B, UROD, ZFPM1 lncRNAs: BOLA3-AS1, ENST433198.2, MIR4435-2HG

Table 1. *Cont.*

DEA	Module	Associated GO Terms	Core Nodes
	del(5q)_4	Erythrocyte homeostasis, homeostasis of number of cells, myeloid cell differentiation, myeloid cell homeostasis, platelet morphogenesis, platelet activation	PCGs: ADAMTS14, DKNND5A, GYPB, ITLN1, LOC100128881, MFSD2B, PGF, PPAPDC1A, SPTA1, TRIB2, UBAC1; lncRNAs: PVT1, RP11-558A11.3, RP11-797H7.5
	del(5q)_5	Oxidative phosphorylation, cellular respiration, mitochondrial envelope, organelle inner membrane, mitochondrion, mitochondrial electron transport ubiquinol to cytochrome c	PCGs: ANAPC11, ARHGAP19, CASP3, EFTUD2, FGFR1OP, GATC, GK, METTL8, NF2, PSMA6, SAE1, SENP1; lncRNAs: OIP5-AS1, POFUT1-006
C. RUNX1-mutated vs. RUNX1-wild type patients	RUNX1_1	RNA binding, RNA processing, RNA localization, RNA splicing, termination of RNA polymerase II transcription, RNA 3'-end processing, protein sumoylation, protein folding, spliceosomal complex, cajal body	PCGs: A_33_P3268147, C11orf31, FIP1L1, FOXN2, LOC100216545, MYL12B, NEDD8, SRSF4, STK36, TMEM199, TOMM6, TPM3; lncRNAs: CTD-2540L5.5, SNHG17, Z83851.3
	RUNX1_2	Translational initiation, RNA binding, protein localization to endoplasmatic reticulum, ribosome biogenesis, rRNA metabolic process, posttranscriptional regulation of gene expression	PCGs: A_24_P350307, A_33_P3210702, CSNK2A2, EIF3F, GNB2L1, NOA1, PAIP2, RPL3, RPL5, RPL7, RPL21, RPL22, RPS14; lncRNAs: GAS5
	RUNX1_3	Cell cycle, immunoglobulin complex, antimicrobial humoral response, 3' to 5' DNA helicase activity, beta-catenin-TCF complex formation, double-strand break repair, V(D)J recombination, p53 binding, chromatin remodeling, recombinational repair, somatic cell DNA recombination, negative regulation of signal transduction by p53 class mediator	PCGs: A_33_P3397473, AKAP12, DUSP26, EBF1, LEF1, LOC283454, NPY, RAG1, SH2D4B; lncRNAs: LEF1-AS1, LINC01013-003, TCL6
D. MDS patients with lower- vs. higher-risk IPSS-R	IPSSR_1	Locomotion, regulation of cell adhesion, blood vessel morphogenesis, regulation of cell proliferation, regulation of cell growth, positive regulation of cell differentiation, Golgi cisterna, circulatory system development, regulation of immune system process, angiogenesis, substrate-dependent cell migration, response to growth factor	PCGs: BCAR1, BTNL9, CLDN5, CLEC4G, EFNB2, GNAZ, NR2F2, NUAK1, RBP7, STAB2, TFF3; lncRNAs: RP11-401P9.5, RP11-474N8.5, RP11-879F14.2
	IPSSR_2	Chromatin organization, immune response, chromatin silencing, negative regulation of gene expression epigenetic, regulation of cell adhesion, histone H4 acetylation	PCGs: A_24_P401601, BAI1, DLG5, LEF1, LOC116437, LOXL4, N4BP2, NPY, PTH2R, RNF19B, STAT4; lncRNAs: LEF1-AS1, RP11-897M7.1
	IPSSR_3	Chromatin organization, histone H3K4 trimethylation, chromatin silencing, negative regulation of gene expression epigenetic, Immune system process, leukocyte homeostasis	PCGs: A_33_P3333327, CCDC106, CLEC4C, DDX60L, F11R, FLJ36777, HIST2H2AC, IFITM3, LOC100288884, SSBP4, ZNF746; lncRNAs: AC012181, LINC00963, ODC1-DT, TCL6

Two modules (MDS_1 and MDS_2) were assigned to hemoglobin and oxygen transport. Both of them included mostly upregulated genes (Table 1), suggesting that the ineffective erythropoiesis found in MDS can result in the increased transcription of multiple factors associated with oxygen transport. Based on significant coexpression, the following lncRNAs were recognized as potentially associated with alterations in oxygen transport machinery: e.g., PRKAR2A-AS1, AC131097, and RP4-669L17 in MDS_1 module, and P11-640M9.1, RP11-640M9.1, CTD-2319I12, and RP11-558A11 in MDS_2 module.

Further, we found that two other modules (MDS_3 and MDS_4) were enriched for genes involved mainly in protein metabolism processes (Table 1). MDS_3 module was primarily associated with protein catabolism and the ubiquitin-proteasome pathway and in the core of this module, we identified AC021188 and ZFAS1 lncRNAs. Two other lncRNAs (EPB41L4A-AS1 and CSNK1A1P1) were found in the core of the MDS_4 module that was associated with transcriptional and translational processes. Interestingly, all the genes that were included in the core of MDS_4 module were significantly downregulated, suggesting that the mechanisms of protein expression can be globally suppressed in dysplastic cells.

3.2. LncRNA Expression in Relation to the Diagnostic Subtypes of MDS

Based on the WHO diagnostic criteria [1], eight groups of samples were defined (i.e., CTR, MDS-SLD, MDS-MLD, MDS-RS, MDS with del(5q), MDS-EB1, MDS-EB2, and AML-MRC) and their expression profiles were compared using one-way ANOVA (Figure 2). We identified three clusters of samples. First, MDS-SLD samples had comparable expression profiles to those of healthy controls. Second, the other remaining early disease subtypes (MDS-MLD, MDS-RS, and MDS with del(5q)) surprisingly clustered with samples from the patients with MDS-EB1 disease (i.e., an advanced subtype of MDS). Third, patients with MDS-EB2 showed similar expression to those with AML-MRC. This distribution shows that there are no specific expression profiles for particular disease subtypes but rather that there is a gradual shift in expression from a healthy state to an advanced myelodysplasia. Interestingly, disease progression can be detected at the molecular level at different point (i.e., between MDS-EB1 and MDS-EB2 subtypes) compared to the classical progression scheme created on the basis of clinical variables (i.e., between MDS-EB2 and AML-MRC).
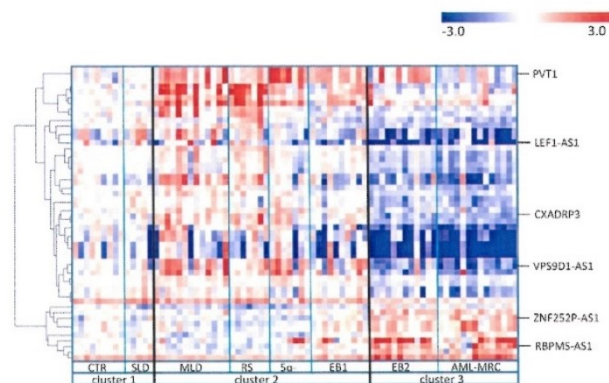
Figure 2. Heatmap of differentially expressed lncRNAs among MDS subtypes. Only the lncRNAs identified as significantly deregulated (FDR < 0.05) in one-way ANOVA were plotted. Samples were divided into eight groups (CTR, MDS-SLD, MDS-MLD, MDS-RS, MDS with del(5q), MDS-EB1, MDS-EB2, AML-MRC), and the analysis defined three clusters of samples with comparable expression profiles. The expression level is calculated as the binary logarithm of fold change (logFC) compared to the mean expression of controls. The heatmap uses a color gradient intensity scale to visually express the logFC values in a range of colors (blue-downregulation, red-upregulation, white-unchanged expression).

3.3. LncRNAs in MDS with Isolated del(5q)

The interstitial deletion of the long arm of chromosome 5, del(5q), is the most common cytogenetic aberration in myelodysplasia, and MDS with isolated del(5q) forms a distinct subtype of this disease [1]. In the discovery cohort, isolated del(5q) was found in 12 MDS/AML-MRC patients, and 20 patients had a normal karyotype. The gene expression profiles of these two groups were compared, and deregulation of 31 lncRNAs (16 upregulated/15 downregulated) and 160 PCGs (106 upregulated/54 downregulated) were identified (|logFC| > 1, FDR < 0.05). The list of deregulated genes is included in Table S4.

Significant deregulation of several hematopoiesis/oncology-related lncRNAs was detected in patients with isolated del(5q), e.g., upregulation of EMCN-IT1, CHRM3-AS2, and PVT1 and downregulation of ZFAS1, EPB41L4A-AS1, and GAS5. The expression levels of CHRM3-AS2, PVT1, and EPB41L4A-AS1 were validated in the testing cohort using RT-qPCR (Figure S2), and the data showed a high level of concordance, indicating the accuracy of the microarray results.

GSEA showed that del(5q) mainly affected genes involved in ribosome formation, translational regulation, STAT5 signaling, and hematopoietic cell lineages including stem/progenitor cells, platelets, and erythrocytes (Figure S3). Further, we constructed a lncRNA-PCG coexpression network, and among the generated modules (assigned as del(5q) modules), we identified those associated with the abovementioned pathways. Based on these modules, we proposed association of several lncRNAs with ribosome formation and translational regulation (del(5q)_1 module, lncRNAs: EPB41L4A-AS1), JAK/STAT cascade (del(5q)_2 module, lncRNAs: NUTM2A-AS1, STARD4-AS1, and RP11-506M13.3) development of blood lineages such as platelet, erythrocyte and myeloid cells (del(5q)_3 module, lncRNAs: BOLA3-AS1, MIR4435-2HG, and ENST433198.2; and del(5q)_4 module, lncRNAs: PVT1, RP11-797H7.5, and RP11-558A11.3), and mitochondria-related processes (del(5q)_5 module, lncRNAs: OIP5-AS1 and POFUT1-006). Table 1 summarizes enriched processes and core PCGs and lncRNAs within these modules.

3.4. Association Between lncRNA Expression and Somatic Mutations

Somatic mutations in multiple genes have recently been associated with MDS [12], rapidly becoming the most frequently discussed aberrations in MDS. Here, we investigated the relationship between the presence of somatic mutations and the expression of lncRNAs. Mutational screening was performed in 64 out of 68 patients and in 8 out of 9 controls in the discovery cohort (due to DNA availability). The results showed that 81% of patients bore at least one somatic mutation (VAF > 5%) with 1.9 mutational events per patient on average (range 0–7). The five most frequently mutated genes included SF3B1 (14 patients, 22%), TET2 (10 patients, 16%), TP53 (10 patients, 16%), DNMT3A (9 patients, 14%), and RUNX1 (9 patients, 14%). In contrast, we found no mutations in healthy controls, excluding the presence of clonal hematopoiesis of indeterminate potential (CHIP). The distribution of the detected mutations within the cohort is shown in Figure S4.

The transcriptional effects of somatic mutations were analyzed in the five most frequently mutated genes (SF3B1, TET2, TP53, DNMT3A, and RUNX1). Within differential expression analyses, we searched for the transcripts with differential levels between patients with and without the given mutation. However, the analysis identified only a few transcripts with standard settings (|logFC| > 1, FDR < 0.05). Therefore, we moderately refined the cut-off of fold change values and reanalyzed the data (|logFC| > 0.3, FDR < 0.05). Interestingly, the numbers of differentially expressed genes substantially varied among the mutations tested (SF3B1: 18 lncRNAs and 20 PCGs; TET2: 13 lncRNAs and 5 PCGs; TP53: 8 lncRNAs and no PCGs; DNMT3A: 1 lncRNA and no PCGs; and RUNX1: 106 lncRNAs and 646 PCGs). At the level of individual transcripts related to hematopoiesis/oncology, we observed the downregulation of the ABCB7 PCG in SF3B1-mutated patients, the downregulation of WT1-AS in TET2-mutated patients, and the upregulation of GAS5 lncRNA, the downregulation of LEF1-AS and TCL6 lncRNAs, and the downregulation of LEF1 and RAG1 PCGs in RUNX1-mutated patient. The full lists of significantly deregulated genes in the patients with the five studied mutations are included in Tables S5–S9.

Because RUNX1-mutated patients displayed the most distinct expression profile, we further focused on this particular gene. A descriptive heatmap (Figure 3A) proved that patients with RUNX1 mutations had substantially different expression profiles. GSEA showed that RUNX1-mutated patients had specifically affected MYC and MAPK signaling pathways, regulation of immune response and cell death, etc. (Figure S5). The lncRNA-PCG coexpression network defined several key modules (assigned as RUNX1 modules) of similarly regulated genes that were enriched in the processes of translational regulation and RNA splicing. Based on these data, we associated Z83851.3, SNHG17, and CTD-2540L5.5 lncRNAs with RNA processing (RUNX1_1 module) and GAS5 lncRNA with protein translation (RUNX1_2 module). Interestingly, we identified an additional module (RUNX1_3 module) whose core nodes included several important genes, namely, LEF1 and RAG1 PCGs and LEF1-AS1 and TCL6 lncRNAs (Figure 3B). Enrichment analysis suggested that these genes are related to DNA repair, DNA recombination, and the p53 pathway. Table 1 summarizes enriched processes and core PCGs and lncRNAs within these modules.

3.5. LncRNAs Related to MDS Prognosis

One of the major goals of this work was the identification of lncRNAs linked to MDS progression potentially serving as prognostic markers of patient outcomes. By differential expression analyses, we evaluated microarray expression profiles according to (i) the overall survival (OS) of patients and (ii) their prognosis based on the IPSS-R system.

To identify genes associated with OS, we established an arbitrary cut-off to 18 months and categorized MDS/AML patients as those with short survival (deceased within 18 months, $N = 31$) or long survival (surviving more than 18 months, $N = 25$). Only 8 lncRNAs (5 upregulated/3 downregulated) and 29 PCGs (5 upregulated/24 downregulated) were significantly changed in the patients with short survival (FDR < 0.05, |logFC| > 1; Table S10). Importantly, two well-known tumorigenic lncRNAs, H19 and WT1-AS, were significantly upregulated in patients with adverse outcomes.

Then, we analyzed differential expression in MDS patients stratified according to the IPSS-R system. For the analysis, the patients were grouped into lower-risk (very low and low IPSS-R scores) and a higher-risk (high and very high IPSS-R scores) categories, while MDS with intermediate risk and AML-MRC patients were excluded from this analysis. We identified 16 lncRNAs (2 upregulated/14 downregulated) and 82 PCGs (15 upregulated/67 downregulated) with significantly changed expression in the higher-risk patients (|logFC| > 1, FDR < 0.05; Table S11). Among the lncRNAs, TCL6 and LEF1-AS were downregulated.

To explore the associations of genes related to disease progression with cellular processes, we performed GSEA on the differentially expressed genes between the lower-risk and higher-risk IPSS-R patient categories. The results showed that the most affected mechanisms include gene expression silencing through chromatin modifications, immune response, cell differentiation and proliferation, adhesion, motility, and angiogenesis (Figure S6).

The coexpression network based on these data identified modules (assigned as IPSS-R modules) that contained similarly regulated lncRNAs and PCGs. Pathway analysis further showed that the majority of these modules were related to processes involved in the immune system. However, more specific enrichment was found for IPSSR_1 module, which was associated with GO terms related to cell differentiation, growth, adhesion and migration, i.e., with the processes that may be linked with specific features of HSCs present in the BM niche. The core lncRNAs found in this module were RP11-474N8.5, RP11-879F14.2, and RP11-401P9.5. Interestingly, the expression of all the genes within the IPSSR_1 module was downregulated, suggesting that the processes of the maintenance and/or differentiation of HSCs in the BM niche may be substantially impaired in higher-risk MDS patients.

Additional interesting network modules of the IPSS-R coexpression network were IPSSR_2 and IPSSR_3 modules, both related mainly to the epigenetic modification (methylation and acetylation) and chromatin structure of DNA. Remarkably, a pair of PCG-lncRNA counterparts, LEF1 and LEF1-AS1, was found in the core of IPSSR_2 module, and TCL6 lncRNA was one of the core nodes in IPSSR_3

53

module, suggesting their involvement in chromatin structure in MDS. Other lncRNAs found in these modules were RP11-897M7.1 (IPSSR_2 module), ODC1-DT, LINC00963, and AC012181 (IPSSR_3 module). Table 1 summarizes enriched processes and core PCGs and lncRNAs within all these modules.
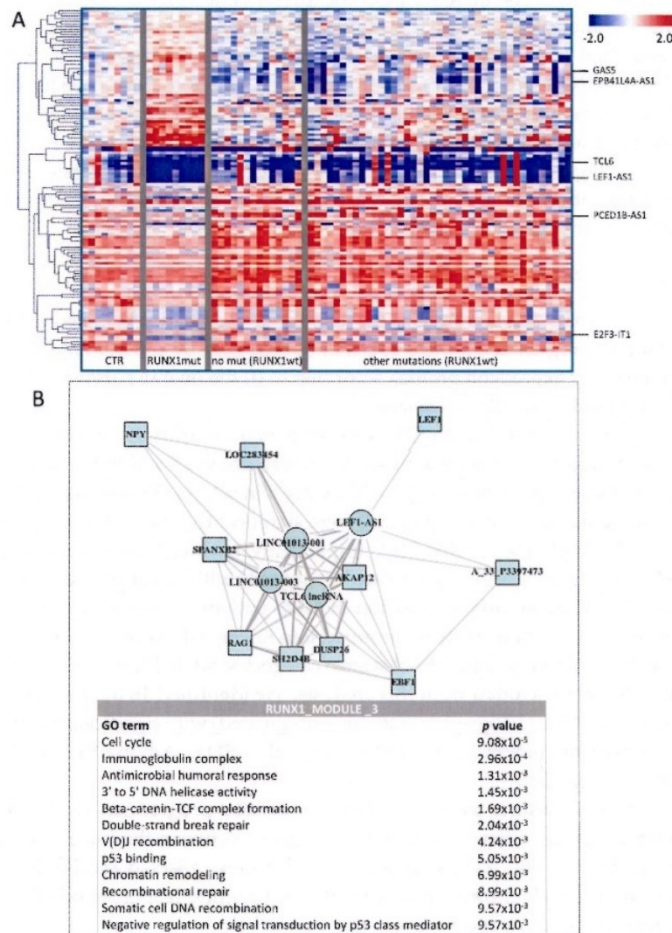
Figure 3. Deregulation of gene expression between the RUNX1-mutated (RUNX1mut) vs. RUNX1-wild type (RUNX1wt) patient samples. (**A**) Heatmap showing the difference in lncRNA expression in MDS/AML-MRC patients stratified according to the presence/absence of the RUNX1 mutation. Only the lncRNAs with significantly changed levels (FDR < 0.05) are plotted. The expression level is calculated as the binary logarithm of fold change (logFC) compared to the mean expression of controls. The heatmap uses a color gradient intensity scale to visually express the logFC values in a range of colors (blue-downregulation, red-upregulation, white-unchanged expression). (**B**) RUNX1_3 module from the lncRNA-PCG coexpression network constructed based on differentially expressed genes. The results of pathway enrichment analysis for this module are included. Square-PCG, circle-lncRNA, red-upregulated in RUNX1mut, blue-downregulated in RUNX1mut.

3.6. Individual lncRNAs as Potential Prognostic Markers of MDS

Based on the microarray results, we chose four candidate lncRNAs applicable as prognostic biomarkers and performed a series of subsequent analyses on two independent cohorts of patients (i.e., discovery and testing cohorts; Figure S1). This set included H19, WT1-AS, TCL6, and LEF1-AS1. Initially, we reanalyzed their expression by RT-qPCR in the testing cohort and proved that the levels of all these transcripts gradually changed from healthy controls to higher-risk patients (Figure 4A), showing a strong concordance with the microarray data.
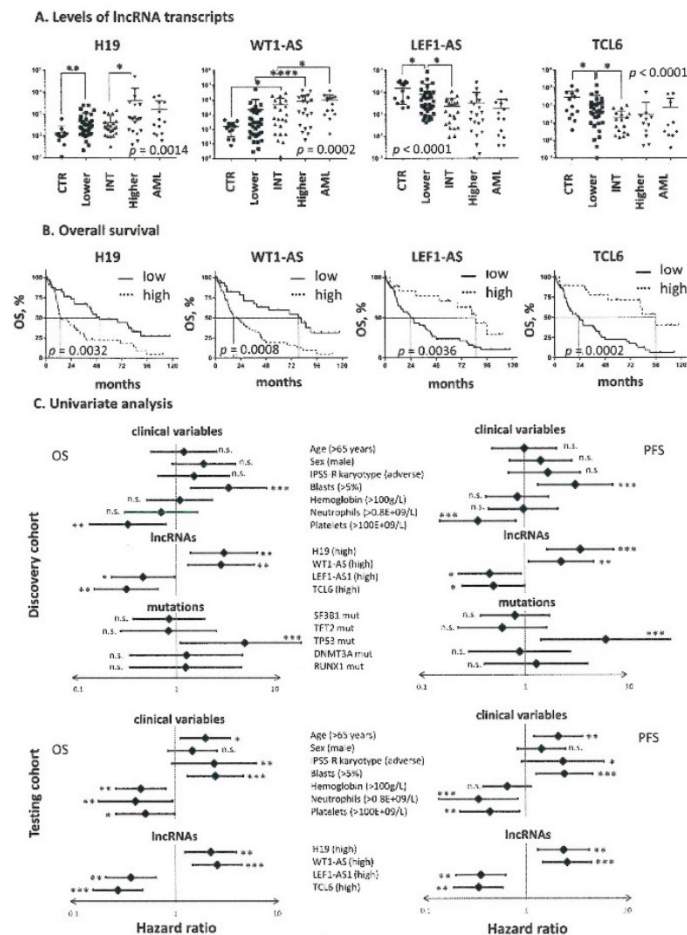
Figure 4. (**A**) Expression of H19, WT1-AS, TCL6, and LEF1-AS lncRNAs in CD34+ BM cells measured by RT-qPCR. The data were logarithmically scaled. The MDS patients (testing cohort) were grouped according to IPSS-R. (**B**) Kaplan-Meier curves for overall survival of MDS patients stratified based on lncRNA levels. (**C**) Forest plots of univariate analysis performed for overall survival (OS) and progression-free survival (PFS) by the log-rank test in both cohorts of MDS patients. Hazard ratios including 95% confidence intervals are plotted and the significance of the results is included (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, n.s.—nonsignificant). The data from mutational screening were available only for the discovery cohort.

To address the question of whether the levels of these four lncRNAs are really associated with patient survival, we performed numerous log-rank tests in both cohorts of patients. Kaplan-Meier curves (Figure 4B) demonstrated a clear difference in survival between MDS patients with low vs. high levels of these lncRNAs. In both cohorts, an adverse outcome was significantly associated with high levels of H19 and WT1-AS and low levels of LEF1-AS and TCL6 (Figure 4C). Further univariate analyses examined the impact of other clinicopathological characteristics on patient survival and revealed that OS and PFS were most significantly associated with blast count, number of platelets, presence of TP53 mutation, and levels of the four lncRNAs (Figure 4C). Other variables, such as age, hemoglobin level, neutrophil count, and karyotype, also showed significant results in at least one of the tested cohorts.

To test the possible dependency between clinical factors and lncRNA alterations, we performed a multitude of Spearman correlation tests between each pair of variables (lncRNAs: H19, WT1-AS, LEF1-AS1, and TCL6; TP53 mutation; clinicopathological features: age, blast count, hemoglobin level, and numbers of neutrophils and platelets) in both cohorts of samples (Table S12). Interestingly, the percentage of marrow blasts significantly correlated with the levels of WT1-AS, TCL6, and LEF1-AS ($p < 0.01$) but was independent of the level of H19. Further, we found a strong positive correlation between the levels of LEF1-AS1 and TCL6 ($p < 0.001$), suggesting their coregulation. The presence of somatic mutations in the TP53 gene was associated with hemoglobin level ($r = -0.316$, $p < 0.05$) and WT1-AS expression ($r = 0.445$, $p < 0.01$). Interestingly, the only almost independent molecular variable was the level of the H19 transcript (with one exception of a slight negative correlation with platelet count specifically in the testing cohort; $r = -0.254$, $p < 0.05$).

To finally determine whether any of the selected lncRNAs might serve as prognostic markers for MDS outcome, we performed Cox multivariate analysis and applied the backward variable selection method to retain only the independent variables significantly contributing to the predictive power of the resulting model. Although the results from both cohorts slightly varied, the analysis revealed that predominant variables predictive of OS and PFS in MDS patients are high blast counts, high levels of the H19 transcript, and presence of somatic mutations in the TP53 gene. To a lesser extent, additional variables such as platelet count, age and TCL6 and LEF1-AS1 levels might add some prognostic value to these three major predictors (Table 2).

To stratify MDS patients based on their prognosis, the IPSS-R system is used in routine clinical practice. However, a proportion of patients scored as having lower to intermediate risk still suffer from an early progression of the disease. Therefore, we tested whether some of the selected lncRNAs can be predictive of adverse outcomes in these patients. In both cohorts, we specifically selected the patients with IPSS-R < 4.5 (i.e., lower-intermediate risk MDS patients) and reanalyzed the data using Cox multivariate regression. Neither blast count nor TP53 mutation remained informative for these patients; the only highly significant variable associated with OS and PFS was H19 expression. Additionally, some other variables such as age, platelet count, and LEF1-AS1 and TCL6 levels, were found to be predictive of patient outcome, but with less significance (Table 2).

Table 2. Multivariate Cox-regression analysis of the overall survival and progression-free survival of MDS patients. Only the variables that remained significant after backward variable selection are listed and sorted according to their descending *p*-values.

Variable	Discovery Cohort			Testing Cohort		
	HR	95% CI	*p*	HR	95% CI	*p*
A. All MDS Patients Overall Survival						
Blast count	19.70	3.83–101.26	<0.001	5.65	1.844–17.30	0.002
H19 level	16.92	1.73–165.12	0.015	54.35	13.10–225.62	<0.001
TP53 mutation	4.86	1.64–14.43	0.004		n.a.	
Platelet count		n.s.		0.19	0.06–0.62	0.006

Table 2. *Cont.*

Variable	Discovery Cohort			Testing Cohort		
LEF1-AS1 level	n.s.			0.01	0.01–0.09	0.006
TCL6 level	n.s.			104.56	1.30–8403.98	0.038
Age	27.80	1.52–507.17	0.025	n.s.		
Progression-free survival						
Blast count	7.60	1.96–29.44	0.003	7.53	2.48–22.79	<0.001
TP53 mutation	6.99	2.75–17.75	<0.001	n.a.		
H19 level	n.s.			76.13	16.97–341.42	<0.001
LEF1-AS1 level	n.s.			0.01	0.01–0.01	0.001
Platelet count	n.s.			0.13	0.03–0.50	0.003
TCL6 level	n.s.			2169.18	$6.11–77.05 \times 10^4$	0.010
B. Lower/intermediate-risk MDS patients (IPSS-R < 4.5)						
Overall survival						
H19 level	n.s.			14.20	3.19–64.65	0.001
TCL6 level	n.s.			0.01	0.01–0.12	0.022
Platelet count	n.s.			0.19	0.04–1.01	0.050
Age	580.70	$1.39–24.28 \times 10^4$	0.039	n.s.		
TP53 mutation	16.13	0.10–260.52	0.050	n.a.		
Progression-free survival						
H19 level	175.43	$4.15–74.17 \times 10^4$	0.007	16.37	3.79–70.67	<0.001
Age	10114.11	$14.58–70.16 \times 10^5$	0.006	n.s.		
Platelet count	n.s.			0.10	0.02–0.62	0.013
LEF1-AS1 level	n.s.			0.01	0.01–0.22	0.023

HR—hazard ratio, CI—confidence interval, n.s.—nonsignificant, n.a.—not analyzed.

3.7. Cis and Trans Correlations of Prognosis-Related lncRNAs

Perturbations of antisense RNAs can regulate the expression of their sense gene counterparts (i.e., PCGs) and vice versa via *cis*-acting regulatory elements. Therefore, we analyzed the expression of the PCG/antisense RNA pairs WT1/WT1-AS and LEF1/LEF1-AS1 by RT-qPCR in the testing cohort of samples. We found that the levels of both pairs of transcripts were highly correlated (WT1: $r = 0.865$, $p < 0.0001$; LEF1: $r = 0.924$, $p < 0.0001$) across all samples (both controls and MDS/AML-MRC patients), suggesting that their direct interactions could be necessary for WT1 and LEF1 functions (Figure S7).

Although TCL6 and H19 lncRNA do not have direct PCG counterparts, they are located in close proximity to some PCGs. TCL6 colocalizes with the TCL1A and TCL1B genes. We found only a very low level of expression for both PCGs; however, if detectable, their expression correlated with the expression of TCL6 (TCL1A: $r = 0.645$, $p < 0.0001$; TCL1B: $r = 0.738$, $p < 0.0001$; Figure S7).

H19 is located adjacent to IGF2. Moreover, H19 functions as a primary template for miR-675. Therefore, we compared the expression levels of H19 to those of IGF2 and miR-675. Although we detected an upregulation of H19 and downregulation of IGF2 and miR-675 in higher-risk MDS compared to lower-risk disease (Figure 5A), the levels of H19 did not directly correlate with either IGF2 or miR-675 when analyzed in the whole cohort. However, an analysis performed separately on the samples from healthy controls and lower-risk IPSS-R patients identified a moderate correlation of H19 expression with both IGF2 ($r = 0.321$, $p = 0.034$) and miR-675 ($r = 0.342$, $p = 0.023$), whereas this concordance was disturbed in intermediate/higher-risk MDS (Figure 5B). Interestingly, the correlation between IGF2 and miR-675 remained unchanged in all samples ($r = 0.307$, $p = 0.006$; Figure 5B). Therefore, we compared the survival of patients with concordant regulation in the H19/IGF2 region with those who did not present this concordance (discordant expression was defined as a > 10-fold change in the ratio between the expression of the tested transcripts). The patients with discordant levels of H19/IGF2 or H19/miR-675 had inferior OS and PFS compared to those with concordant levels of these transcripts (univariate analyses: $p < 0.05$; Figure 5C).

A. Expression levels in IPSS-R categories

B. Correlation of expression levels

	H19 vs. IGF2	H19 vs. miR-675	IGF2 vs. miR-675
all samples	$r = 0.135$ $p = 0.168$	$r = -0.049$ $p = 0.664$	$r = 0.307$ $p = 0.006*$
CTR/lower-risk MDS	$r = 0.321$ $p = 0.034*$	$r = 0.342$ $p = 0.023*$	$r = 0.339$ $p = 0.030*$
int/higher-risk MDS	$r = -0.174$ $p = 0.298$	$r = -0.027$ $p = 0.876$	$r = 0.340$ $p = 0.042*$

C. Survival analysis

Figure 5. Expression of H19, IGF2, and miR-675. (**A**) Relative expression levels in MDS patients (testing cohort) grouped according to IPSS-R. (**B**) Correlation of expression levels (Spearman test) in different sample groups. (**C**) Kaplan-Meier curves for overall survival (OS) and progression-free survival (PFS) in MDS patients with concordant vs. discordant expression of the H19/IGF2 and H19/miR-675 pairs. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

To address the *trans* regulatory mechanisms of the four lncRNAs related to MDS prognosis (H19, WT1-AS, LEF1-AS1, and TCL6), we constructed coexpression networks in which these lncRNAs formed the central nodes. Interestingly, the network-computing strategy generated only two modules for H19/WT1-AS and LEF1-AS1/TCL6 lncRNA pairs (assigned as the H19/WT1-AS_module and LEF1-AS1/TCL6_module, respectively), suggesting that these two pairs of genes might be functionally related. Enrichment analysis suggested that the H19/WT1-AS pair was predominantly associated with cell adhesion and differentiation processes whereas the LEF1-AS1/TCL6 pair might function in

diverse mechanisms, such as chromatin modification, cytokine response, or cell proliferation and death (Figure 6).
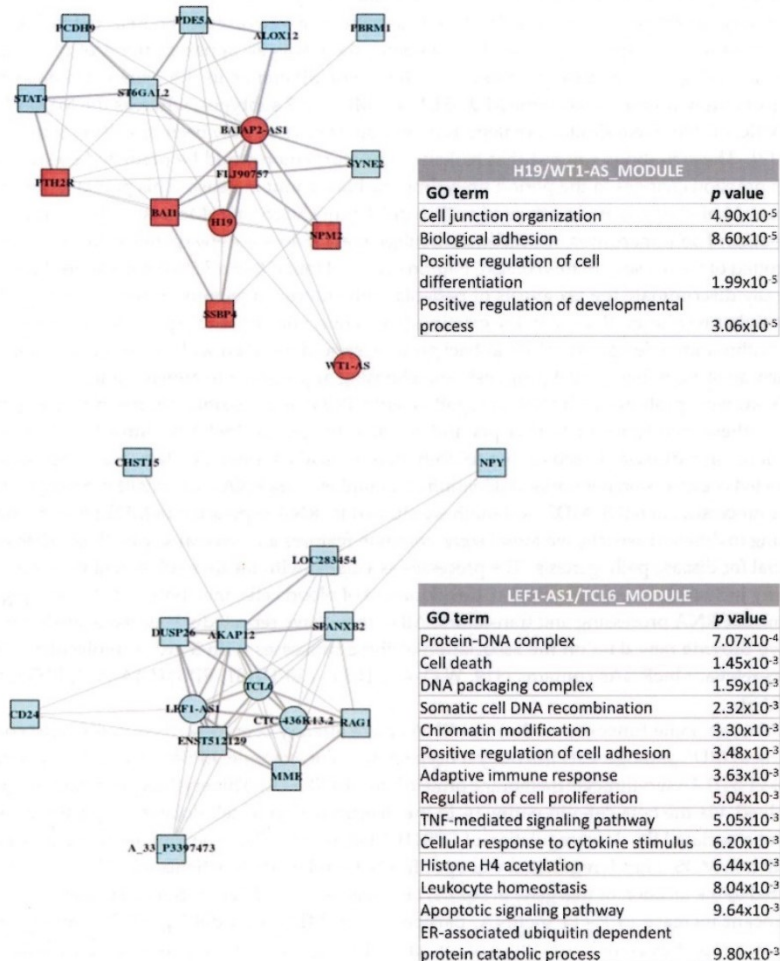

H19/WT1-AS_MODULE	
GO term	***p* value**
Cell junction organization	4.90x10⁻⁵
Biological adhesion	8.60x10⁻⁵
Positive regulation of cell differentiation	1.99x10⁻⁵
Positive regulation of developmental process	3.06x10⁻⁵

LEF1-AS1/TCL6_MODULE	
GO term	***p* value**
Protein-DNA complex	7.07x10⁻⁴
Cell death	1.45x10⁻³
DNA packaging complex	1.59x10⁻³
Somatic cell DNA recombination	2.32x10⁻³
Chromatin modification	3.30x10⁻³
Positive regulation of cell adhesion	3.48x10⁻³
Adaptive immune response	3.63x10⁻³
Regulation of cell proliferation	5.04x10⁻³
TNF-mediated signaling pathway	5.05x10⁻³
Cellular response to cytokine stimulus	6.20x10⁻³
Histone H4 acetylation	6.44x10⁻³
Leukocyte homeostasis	8.04x10⁻³
Apoptotic signaling pathway	9.64x10⁻³
ER-associated ubiquitin dependent protein catabolic process	9.80x10⁻³

Figure 6. Coexpression network formed around H19, WT1-AS, LEF1-AS1, and TCL6 lncRNAs. The computational process generated two modules for the H19/WT1-AS and LEF1-AS1/TCL6 lncRNA pairs.

4. Discussion

MDS patients display heterogeneous clinicopathological features accompanied by distinct genetic characteristics. Until recently, proteins acting through complex signaling pathways were considered exclusive vehicles linking these genomic abnormalities to clinical phenotypes. However, aside from the proteins, various classes of noncoding RN As have been shown to contribute to the variability of the disease [4,5,13,14]. In this work, we screened the lncRNA background in MDS with relation to different

59

characteristics of the disease (i.e., diagnosis, disease subtypes, cytogenetic/mutational aberrations, and risk of progression). In addition to the description of lncRNA profiles specific for this disease, we aimed to identify lncRNAs with potential prognostic capability.

Using a genome-wide approach, we showed that samples from patients with early MDS had substantially different expression profiles than those with advanced disease. We detected the significant induction of multiple genes in early MDS, whereas there was an apparent trend of reduced gene expression along with disease progression. The major difference in the gene expression profile was surprisingly found between the MDS-EB1 and MDS-EB2 subtypes, whereas the MDS-EB2 and AML-MRC profiles were similar. Analogous observations have already been described at the miRNA level [13]. These findings suggest that leukemic transformation could be predicted by monitoring gene expression changes in the period preceding the blast count increase. This points to the issue of the accuracy of blast count cut-offs used within the diagnostic criteria. Moreover, the morphological examination of bone marrow is, to some extent, subjective and does not always reflect the true molecular background of the disease. In this context, Padron et al. [15] found that a 7.5% bone marrow blast cut-off point may discriminate the prognosis of patients with chronic myelomonocytic leukemia (CMML) with a higher resolution than with the existing 10%. Thus, clinicopathological classification criteria alone, without knowledge of molecular background, should be taken with some caution, mainly for the purpose of assessing patient prognosis and choosing appropriate treatment strategies.

Expression profiling of lncRNAs together with PCGs is a feasible approach to compare the levels of these two types of transcripts and to identify specific lncRNAs linked with pathways deregulated in a disease. Based on the combination of lncRNA and PCG data, we computationally constructed coexpression networks and identified a number of lncRNAs that might function in various cellular processes altered in MDS. Although we studied lncRNA expression in MDS patients stratified according to different aspects, we found some common features and several key lncRNAs that seem to be crucial for disease pathogenesis. The processes associated with the deregulation of gene expression primarily included immune regulation, development of blood cells, metabolism of heme, epigenetic mechanisms, RNA processing and translation. All of them have repeatedly been associated with MDS; here, we provide new data on the association of these processes with lncRNA molecules. The list of MDS-relevant lncRNAs contains H19, WT1-AS, TCL6, LEF1-AS1, EPB41L4A-AS1, PVT1, GAS5, and ZFAS1.

Although some functions of these lncRNAs have already been described in different contexts, their link to MDS pathogenesis has not yet been shown. For example, EPB41L4A-AS1 is an antisense RNA to EPB41L4A (erythrocyte membrane protein band 4.1 like 4A). This erythrocyte membrane protein is involved, via the beta-catenin pathway, in the determination of cell polarity or proliferation [16]. Here, we associated the downregulation of EPB41L4A-AS1 with ribosome formation and translational regulation in MDS. This deregulation was specifically found in MDS patients with del(5q) and can be attributed to the location of this gene in 5q22.1, near the common deleted region. Further, we detected a significant increase in PVT1 lncRNA, particularly in MDS with del(5q). PVT1 usually confers oncogenic properties on different types of cancer, including AML, and functions as a mediator of the tumor-suppressive functions of p53 [17].

Currently, the molecular testing of somatic mutations is increasingly being applied in routine practice in MDS diagnostics. The impact of somatic mutations on clinical variables and patient outcomes naturally depends on their manifestation through gene expression. To provide an understanding of how genomic variations interfere with the noncoding transcriptome, we combined data from the expression profiling with information on the mutational status of MDS patients. However, we identified only a few deregulated transcripts in the patients with vs. without particular mutations, and their numbers substantially varied among the mutations tested. SF3B1 is an RNA splicing factor, TP53 functions as a tumor suppressor inducing apoptosis, and TET2 and DNMT3A are involved in epigenetic modifications of the genome [12]; therefore, it is surprising that we detected remarkably small numbers of affected transcripts instead of a pervasive effect on the whole transcriptome. To define the reasons

for this lack of expressional difference, we have to consider several aspects. A single nucleotide change in one gene might not be sufficient to induce such strong expressional change, at least when compared with the effects of del(5q) causing the haploinsufficiency of tens of genes. Moreover, diverse variants found in one gene or the cooccurrence of several mutations or cytogenetic aberrations in one patient can have variable effects on gene transcription. Therefore, larger cohorts of patients with isolated mutations are necessary to better define the transcriptional effects of somatic mutations.

RUNX1 was the gene whose mutations had the strongest transcriptional impact in our dataset, even comparable to the effects of del(5q). RUNX1 is a hematopoietic transcription factor whose somatic mutations are considered one of the most prognostically unfavorable mutational events, even in lower-risk MDS [18]. Given the adverse outcomes of RUNX1-mutated patients it is not surprising that these patients had substantially distinct expression profiles. In the patients with RUNX1 mutations, we identified the deregulation of RAG1 PCG, a core node in the DNA repair and recombination module. RAG1 (recombination activating gene 1) is an RUNX1-associated recombinase involved in antibody and T-cell receptor recombination [19]. Network modeling linked the RUNX1-RAG1 axis with LEF1/LEF1-AS and TCL6, the same transcripts whose deregulations were significantly associated with poor prognosis.

A vast body of literature exists on the deregulated expression of various PCGs and their potential applicability as prognostic markers in MDS (e.g., [20–22]). For example, Pellagatti et al. identified several PCGs (e.g., LEF1, CDH1, WT1, and MN1), the expression of which was significantly associated with the survival of MDS patients [20]. Transcripts of PCGs, however, are not the final effectors in the cells, unlike proteins and noncoding RNAs. Therefore, it can be assumed that lncRNA expression should be a more reliable prognostic marker than the expression of PCGs. In our study, four important lncRNAs (H19, WT1-AS, TCL6, and LEF1-AS1) were significantly associated with the outcome of MDS patients. A series of statistical tests proved that monitoring lncRNA transcription may have a highly significant potential for the prediction of outcomes in MDS patients, and the only other molecular method able to compete with them is TP53 mutational screening.

Of the four abovementioned lncRNAs, H19 (H19 imprinted maternally expressed transcript) was the most promising MDS marker in our dataset. Its increased level was associated with the rapid progression of the disease and short patient survival. Further, we associated the upregulation of H19 in higher-risk MDS with altered cell adhesion and differentiation processes in CD34+ BM cells. It has already been shown that H19 overexpression promotes leukemogenesis and predicts unfavorable prognosis in AML through its proliferative and antiapoptotic effects. Moreover, H19 overexpression correlated with a lower complete remission rate of induction therapy in AML [23]. Importantly, we showed that the level of H19 is independent of the majority of clinical and molecular variables and that its increase has strong predictive value comparable to increased blast count and the presence of TP53 mutation. Moreover, increased H19 level remained informative even in lower/intermediate-risk patients unlike blast count and TP53 mutation.

Although the H19 gene does not have a direct protein-coding counterpart, it is located in close proximity to the IGF2 gene in a region denoted as the H19/IGF2 locus. H19 and IGF2 are mutually imprinted genes, sharing one imprinting control region. In most tissues, H19 is expressed from the maternal allele, whereas IGF2 is expressed from the paternal allele [24]. H19 also functions as a primary template for miR-675, which plays an important role in tumorigenesis and the development of various cancers [25,26]. Our data suggest that the transcriptional coregulation of H19/IGF2/miR-675 seen in healthy donors and low-risk MDS becomes disrupted along with disease progression. Moreover, the discordant expression of these genes is associated with worse outcomes in MDS patients. Given to imprinting described in H19/IGF2 locus, we hypothesize that the disruption of transcriptional coregulation of H19/IGF2/miR-675 may be linked to abnormal methylation in the imprinting control region. However, this hypothesis has to be verified in an ongoing study.

Other lncRNAs whose transcription levels were strongly related to the outcomes of MDS patients were TCL6, WT1-AS, and LEF1-AS1. TCL6 (T-cell leukemia/lymphoma 6) is a lncRNA whose specific

61

expression was initially reported in T-cell leukemia with t (14; 14) (q11; q32.1) translocation [27]. Decreased TCL6 levels have been associated with poor prognosis in patients with clear cell renal cell carcinoma [28]. WT1-AS and LEF1-AS1 are antisense transcripts of two PCGs, WT1 (Wilms tumor 1) and LEF1 (lymphoid enhancer binding factor 1), which belong among the strongest candidate genes showing an association with the prognosis of MDS patients [20]. WT1 plays a role in cell differentiation and apoptosis, and monitoring of the WT1 transcript is useful for estimating minimal residual disease and predicting outcomes in AML and MDS [29–31]. On the other hand, the repression of LEF1 inhibits proliferation, induces the apoptosis of CD34$^+$ progenitors, and plays a critical role in the defective maturation program of myeloid progenitors [32]. Additionally, our data suggest a novel link of these two lncRNAs to chromatin modification, cytokine response, or cell proliferation and death.

Furthermore, we found that WT1-AS and LEF1-AS1 were strongly transcriptionally coregulated with their sense PCG counterparts. WT1-AS colocalizes with WT1 RNA and forms RNA:RNA duplexes, indicating a possible RNA stabilization role for WT1-AS transcripts [33]. Congrains-Castillo et al. [34] demonstrated a correlation between LEF1 and LEF1-AS1 expression in BM cells from MDS/AML patients. Upon overexpression of LEF1-AS1, they observed an inhibition of cell proliferation. However, they did not detect any alteration in LEF1 expression, suggesting that LEF1-AS1 affects cell proliferation in a LEF1-independent manner [34].

Although significantly associated with patient prognosis, the transcription of TCL6, WT1-AS, and LEF1-AS correlated with the percentage of marrow blasts. In this context, Nagasaki et al. previously reported that elevated WT1 levels may be related to increased blast cell numbers and to the presence of preleukemic MDS clones with poor prognostic chromosomal rearrangements [29]. Our correlation analyses revealed additional associations among the WT1-AS, TCL6, and LEF1-AS levels and the presence of TP53 mutations. These relations warrant that routine measurements of gene expression may be potentially confused by various reliant factors or even be redundant, at least in specific subgroups of patients (especially those with high blast counts, unfavorable cytogenetics or TP53 mutations). Thus, measurements of WT1-AS, TCL6, and LEF1-AS levels seem to provide only limited information to the current prognostic systems.

Besides description of new prognostic markers, another important aspect of this study is identification of new options for targeted therapy of the disease. Some of the examined lncRNAs, such as H19, WT1-AS, TCL6, and LEF1-AS, might serve as new druggable targets especially in higher-risk MDS. However, careful examination of functional aspects of their deregulation is required to bring necessary information for proper design of new efficient targeted therapies.

To conclude, our findings provide novel information on particular lncRNAs contributing to MDS pathogenesis and propose cellular processes associated with these transcripts. Moreover, we found that H19, WT1-AS, TCL6, and LEF1-AS1 lncRNAs are particularly associated with the outcome of MDS patients. Based on a series of statistical tests, we demonstrated that the level of H19 transcript might serve as a robust independent prognostic marker comparable to clinical variables currently used for patient stratification. Based on our data, we encourage further, larger-scale studies that will suggest a novel prognostic scoring system combining clinical variables with several genetic markers of diverse characteristics, including somatic mutations and both PCG and lncRNA expression.

30 transcripts are listed, Table S3: List of significantly deregulated transcripts in AML-MRC compared to MDS patients (|logFC| > 1, FDR < 0.05). Of the 159 downregulated PCGs, only the top 30 transcripts are listed, Table S4: List of significantly deregulated transcripts in MDS/AML-MRC patients with isolated del(5q) vs. those with a normal karyotype (|logFC| > 1, FDR < 0.05). Of the 106 upregulated and 54 downregulated PCGs, only the top 30 transcripts are listed in each category, Table S5: List of significantly deregulated transcripts in MDS patients with vs. without a SF3B1 mutation (|logFC| > 0.3, FDR < 0.05), Table S6: List of significantly deregulated transcripts in MDS patients with vs. without a TET2 mutation (|logFC| > 0.3, FDR < 0.05), Table S7: List of significantly deregulated transcripts in MDS patients with vs. without a TP53 mutation (|logFC| > 0.3, FDR < 0.05), Table S8: List of significantly deregulated transcripts in MDS patients with vs. without a DNMT3A mutation (|logFC| > 0.3, FDR < 0.05), Table S9: List of significantly deregulated transcripts in MDS patients with vs. without RUNX1 mutation (|logFC| > 0.3, FDR < 0.05), Table S10: List of significantly deregulated transcripts in patients with long vs. short survival (|logFC| > 1, FDR < 0.05), Table S11: List of significantly deregulated transcripts in MDS patients with lower- vs. higher-risk IPSS-R (|logFC| > 1, FDR < 0.05), Table S12: Correlations of lncRNA expression with clinical variables of MDS patients.

References

1. Arber, D.A.; Orazi, A.; Hasserjian, R.; Thiele, J.; Borowitz, M.J.; Le Beau, M.M.; Bloomfield, C.D.; Cazzola, M.; Vardiman, J.W. The 2016 Revision to the World Health Organization Classification of Myeloid Neoplasms and Acute Leukemia. *Blood* **2016**, *127*, 2391–2405. [CrossRef]

2. Greenberg, P.L.; Tuechler, H.; Schanz, J.; Sanz, G.; Garcia-Manero, G.; Solé, F.; Bennett, J.M.; Bowen, D.; Fenaux, P.; Dreyfus, F.; et al. Revised International Prognostic Scoring System for Myelodysplastic Syndromes. *Blood* **2012**, *120*, 2454–2465. [CrossRef]

3. Scott, L.J. Azacitidine: A Review in Myelodysplastic Syndromes and Acute Myeloid Leukaemia. *Drugs* **2016**, *76*, 889–900. [CrossRef]

4. Liu, K.; Beck, D.; Thoms, J.A.I.; Liu, L.; Zhao, W.; Pimanda, J.E.; Zhou, X. Annotating Function to Differentially Expressed LincRNAs in Myelodysplastic Syndrome using a Network-Based Method. *Bioinformatics* **2017**, *33*, 2622–2630. [CrossRef]

5. Yao, C.; Chen, C.; Huang, H.; Hou, H.; Lin, C.; Tseng, M.; Kao, C.; Lu, T.; Chou, W.; Tien, H. A 4-lncRNA Scoring System for Prognostication of Adult Myelodysplastic Syndromes. *Blood Adv.* **2017**, *1*, 1505–1516. [CrossRef] [PubMed]

6. Derrien, T.; Johnson, R.; Bussotti, G.; Tanzer, A.; Djebali, S.; Tilgner, H.; Guernec, G.; Martin, D.; Merkel, A.; Knowles, D.G.; et al. The GENCODE V7 Catalog of Human Long Noncoding RNAs: Analysis of their Gene Structure, Evolution, and Expression. *Genome Res.* **2012**, *22*, 1775–1789. [CrossRef]

7. Xie, F.; Xiao, P.; Chen, D.; Xu, L.; Zhang, B. miRDeepFinder: A miRNA Analysis Tool for Deep Sequencing of Plant Small RNAs. *Plant Mol. Biol.* **2012**, *80*, 75–84. [CrossRef] [PubMed]

8. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [CrossRef] [PubMed]

9. Heller, G.; Schmidt, W.M.; Ziegler, B.; Holzer, S.; Müllauer, L.; Bilban, M.; Zielinski, C.C.; Drach, J.; Zöchbauer-Müller, S. Genome-Wide Transcriptional Response to 5-Aza-2'-Deoxycytidine and Trichostatin a in Multiple Myeloma Cells. *Cancer Res.* **2008**, *68*, 44–54. [CrossRef]

10. Nuytten, M.; Beke, L.; Van Eynde, A.; Ceulemans, H.; Beullens, M.; Van Hummelen, P.; Fuks, F.; Bollen, M. The Transcriptional Repressor NIPP1 is an Essential Player in EZH2-Mediated Gene Silencing. *Oncogene* **2008**, *27*, 1449–1460. [CrossRef]

11. Graessmann, M.; Berg, B.; Fuchs, B.; Klein, A.; Graessmann, A. Chemotherapy Resistance of Mouse WAP-SVT/T Breast Cancer Cells is Mediated by Osteopontin, Inhibiting Apoptosis Downstream of Caspase-3. *Oncogene* **2007**, *26*, 2840–2850. [CrossRef] [PubMed]

12. Papaemmanuil, E.; Gerstung, M.; Malcovati, L.; Tauro, S.; Gundem, G.; Van Loo, P.; Yoon, C.J.; Ellis, P.; Wedge, D.C.; Pellagatti, A.; et al. Clinical and Biological Implications of Driver Mutations in Myelodysplastic Syndromes. *Blood* **2013**, *122*, 3616–3627. [CrossRef] [PubMed]

13. Dostalova Merkerova, M.; Krejcik, Z.; Votavova, H.; Belickova, M.; Vasikova, A.; Cermak, J. Distinctive microRNA Expression Profiles in CD34+ Bone Marrow Cells from Patients with Myelodysplastic Syndrome. *Eur. J. Hum. Genet.* **2011**, *19*, 313–319. [CrossRef]

14. Votavova, H.; Grmanova, M.; Dostalova Merkerova, M.; Belickova, M.; Vasikova, A.; Neuwirtova, R.; Cermak, J. Differential Expression of microRNAs in CD34+ Cells of 5q- Syndrome. *J. Hematol. Oncol.* **2011**, *4*, 1. [CrossRef]

15. Padron, E.; Garcia-Manero, G.; Patnaik, M.M.; Itzykson, R.; Lasho, T.; Nazha, A.; Rampal, R.K.; Sanchez, M.E.; Jabbour, E.; Al Ali, N.H.; et al. An International Data Set for CMML Validates Prognostic Scoring Systems and Demonstrates a Need for Novel Prognostication Strategies. *Blood Cancer J.* **2015**, *5*, e333. [CrossRef]

16. Ishiguro, H.; Furukawa, Y.; Daigo, Y.; Miyoshi, Y.; Nagasawa, Y.; Nishiwaki, T.; Kawasoe, T.; Fujita, M.; Satoh, S.; Miwa, N.; et al. Isolation and Characterization of Human NBL4, a Gene Involved in the Beta-Catenin/Tcf Signaling Pathway. *Jpn. J. Cancer Res.* **2000**, *91*, 597–603. [CrossRef] [PubMed]

17. Boloix, A.; Masanas, M.; Jiménez, C.; Antonelli, R.; Soriano, A.; Roma, J.; Sánchez de Toledo, J.; Gallego, S.; Segura, M.F. Long Non-Coding RNA PVT1 as a Prognostic and Therapeutic Target in Pediatric Cancer. *Front. Oncol.* **2019**, *9*, 1173. [CrossRef]

18. Hruba, M.; Vesela, J.; Votavova, H.; Dostalova Merkerova, M.; Kundrat, D.; Szikszai, K.; Lauermanova, M.; Zemanova, Z.; Jonasova, A.; Cermak, J.; et al. RUNX1 Mutation Accompanied with Dysregulated Cellular Senescence in Lower-Risk Myelodysplastic Syndrome Patients is Associated with Disease Progression. *Blood* **2019**, *134*, 4230. [CrossRef]

19. Cieslak, A.; Le Noir, S.; Trinquand, A.; Lhermitte, L.; Franchini, D.; Villarese, P.; Gon, S.; Bond, J.; Simonin, M.; Vanhille, L.; et al. RUNX1-Dependent RAG1 Deposition Instigates Human TCR-Δ Locus Rearrangement. *J. Exp. Med.* **2014**, *211*, 1821–1832. [CrossRef]

20. Pellagatti, A.; Benner, A.; Mills, K.I.; Cazzola, M.; Giagounidis, A.; Perry, J.; Malcovati, L.; Della Porta, M.G.; Jädersten, M.; Verma, A.; 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. [CrossRef]

21. Prall, W.C.; Czibere, A.; Grall, F.; Spentzos, D.; Steidl, U.; Giagounidis, A.A.N.; Kuendgen, A.; Otu, H.; Rong, A.; Libermann, T.A.; et al. Differential Gene Expression of Bone Marrow-Derived CD34+ Cells is Associated with Survival of Patients Suffering from Myelodysplastic Syndrome. *Int. J. Hematol.* **2009**, *89*, 173–187. [CrossRef] [PubMed]

22. Minetto, P.; Guolo, F.; Clavio, M.; De Astis, E.; Colombo, N.; Grasso, R.; Fugazza, G.; Sessarego, M.; Lemoli, R.M.; Gobbi, M.; et al. Combined Assessment of WT1 and BAALC Gene Expression at Diagnosis may Improve Leukemia-Free Survival Prediction in Patients with Myelodysplastic Syndromes. *Leuk. Res.* **2015**, *39*, 866–873. [CrossRef] [PubMed]

23. Zhang, T.; Zhou, J.; Zhang, W.; Lin, J.; Ma, J.; Wen, X.; Yuan, Q.; Li, X.; Xu, Z.; Qian, J. H19 Overexpression Promotes Leukemogenesis and Predicts Unfavorable Prognosis in Acute Myeloid Leukemia. *Clin. Epigenetics* **2018**, *10*, 47. [CrossRef] [PubMed]

24. Nordin, M.; Bergman, D.; Halje, M.; Engström, W.; Ward, A. Epigenetic Regulation of the Igf2/H19 Gene Cluster. *Cell Prolif.* **2014**, *47*, 189–199. [CrossRef] [PubMed]

25. He, D.; Wang, J.; Zhang, C.; Shan, B.; Deng, X.; Li, B.; Zhou, Y.; Chen, W.; Hong, J.; Gao, Y.; et al. Down-Regulation of miR-675-5p Contributes to Tumor Progression and Development by Targeting Pro-Tumorigenic GPR55 in Non-Small Cell Lung Cancer. *Mol. Cancer* **2015**, *14*, 73. [CrossRef]

26. Vennin, C.; Spruyt, N.; Dahmani, F.; Julien, S.; Bertucci, F.; Finetti, P.; Chassat, T.; Bourette, R.P.; Le Bourhis, X.; Adriaenssens, E. H19 Non Coding RNA-Derived miR-675 Enhances Tumorigenesis and Metastasis of Breast Cancer Cells by Downregulating C-Cbl and Cbl-B. *Oncotarget* **2015**, *6*, 29209–29223. [CrossRef]

64

27. Saitou, M.; Sugimoto, J.; Hatakeyama, T.; Russo, G.; Isobe, M. Identification of the TCL6 Genes within the Breakpoint Cluster Region on Chromosome 14q32 in T-Cell Leukemia. *Oncogene* **2000**, *19*, 2796–2802. [CrossRef]

28. Su, H.; Sun, T.; Wang, H.; Shi, G.; Zhang, H.; Sun, F.; Ye, D. Decreased TCL6 Expression is Associated with Poor Prognosis in Patients with Clear Cell Renal Cell Carcinoma. *Oncotarget* **2017**, *8*, 5789–5799. [CrossRef]

29. Nagasaki, J.; Aoyama, Y.; Hino, M.; Ido, K.; Ichihara, H.; Manabe, M.; Ohta, T.; Mugitani, A. Wilms Tumor 1 (WT1) mRNA Expression Level at Diagnosis is a Significant Prognostic Marker in Elderly Patients with Myelodysplastic Syndrome. *Acta Haematol.* **2017**, *137*, 32–39. [CrossRef]

30. Galimberti, S.; Ghio, F.; Guerrini, F.; Ciabatti, E.; Grassi, S.; Ferreri, M.I.; Petrini, M. WT1 Expression Levels at Diagnosis could Predict Long-Term Time-to-Progression in Adult Patients Affected by Acute Myeloid Leukaemia and Myelodysplastic Syndromes. *Br. J. Haematol.* **2010**, *149*, 451–454. [CrossRef]

31. Inoue, K.; Sugiyama, H.; Ogawa, H.; Nakagawa, M.; Yamagami, T.; Miwa, H.; Kita, K.; Hiraoka, A.; Masaoka, T.; Nasu, K. WT1 as a New Prognostic Factor and a New Marker for the Detection of Minimal Residual Disease in Acute Leukemia. *Blood* **1994**, *84*, 3071–3079. [CrossRef] [PubMed]

32. Skokowa, J.; Cario, G.; Uenalan, M.; Schambach, A.; Germeshausen, M.; Battmer, K.; Zeidler, C.; Lehmann, U.; Eder, M.; Baum, C.; et al. LEF-1 is Crucial for Neutrophil Granulocytopoiesis and its Expression is Severely Reduced in Congenital Neutropenia. *Nat. Med.* **2006**, *12*, 1191–1197. [CrossRef] [PubMed]

33. Dallosso, A.R.; Hancock, A.L.; Malik, S.; Salpekar, A.; King-Underwood, L.; Pritchard-Jones, K.; Peters, J.; Moorwood, K.; Ward, A.; Malik, K.T.A.; et al. Alternately Spliced WT1 Antisense Transcripts Interact with WT1 Sense RNA and show Epigenetic and Splicing Defects in Cancer. *RNA* **2007**, *13*, 2287–2299. [CrossRef] [PubMed]

34. Congrains-Castillo, A.; Niemann, F.S.; Santos Duarte, A.S.; Olalla-Saad, S.T. LEF1-AS1, Long Non-Coding RNA, Inhibits Proliferation in Myeloid Malignancy. *J. Cell. Mol. Med.* **2019**, *23*, 3021–3025. [CrossRef] [PubMed]
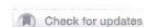
4.4. Publication II

ARTICLE   **OPEN**

Check for updates

RUNX1 mutations contribute to the progression of MDS due to disruption of antitumor cellular defense: a study on patients with lower-risk MDS

Monika Kaisrlikova [1,2], Jitka Vesela [1], David Kundrat [1], Hana Votavova[1], Michaela Dostalova Merkerova [1], Zdenek Krejcik[1], Vladimir Divoky [3], Marek Jedlicka[1,4], Jan Fric[1,5], Jiri Klema[6], Dana Mikulenkova[1], Marketa Stastna Markova[1], Marie Lauermannova[1], Jolana Mertova[1], Jacqueline Soukupova Maaloufova[1], Anna Jonasova [7], Jaroslav Cermak[1] and Monika Belickova [1,2 ✉]

Patients with lower-risk myelodysplastic syndromes (LR-MDS) have a generally favorable prognosis; however, a small proportion of cases progress rapidly. This study aimed to define molecular biomarkers predictive of LR-MDS progression and to uncover cellular pathways contributing to malignant transformation. The mutational landscape was analyzed in 214 LR-MDS patients, and at least one mutation was detected in 137 patients (64%). Mutated *RUNX1* was identified as the main molecular predictor of rapid progression by statistics and machine learning. To study the effect of mutated *RUNX1* on pathway regulation, the expression profiles of CD34 + cells from LR-MDS patients with *RUNX1* mutations were compared to those from patients without *RUNX1* mutations. The data suggest that *RUNX1*-unmutated LR-MDS cells are protected by DNA damage response (DDR) mechanisms and cellular senescence as an antitumor cellular barrier, while *RUNX1* mutations may be one of the triggers of malignant transformation. Dysregulated DDR and cellular senescence were also observed at the functional level by detecting γH2AX expression and β-galactosidase activity. Notably, the expression profiles of *RUNX1*-mutated LR-MDS resembled those of higher-risk MDS at diagnosis. This study demonstrates that incorporating molecular data improves LR-MDS risk stratification and that mutated *RUNX1* is associated with a suppressed defense against LR-MDS progression.

INTRODUCTION

Myelodysplastic syndromes (MDS) are a heterogeneous group of diseases with clonal hematopoiesis [1]. MDS patients are usually stratified into four risk groups according to their risk of transformation to acute myeloid leukemia (AML) by the International Prognostic Scoring System (IPSS) [2] or 5 risk groups by the Revised International Prognostic Scoring System (IPSS-R) [3]. Low- and intermediate-1 (INT-1) risk groups of IPSS and very low-risk, low-risk, and part of the intermediate-risk groups of IPSS-R are considered lower-risk MDS (LR-MDS) [4, 5]. Despite the more favorable prognosis, some LR-MDS patients progress rapidly [6].

Early identification of LR-MDS patients at risk of rapid progression is crucial for the initiation of effective treatment. In this context, numerous studies have mapped the genomic landscape in MDS patients to improve risk stratification and prognosis estimation. There has been a long-lasting effort to upgrade scoring systems by incorporating molecular features to give rise to IPSS-molecular [7–14]. However, no unified results have been generally accepted yet. The sole mutated gene included in the MDS classification by the World Health Organization is *SF3B1*, which is related to the percentage of ring sideroblasts in erythroid elements of bone marrow (BM) [15].

RUNX1 is a frequently mutated gene in hematological malignancies and is associated with an adverse course of disease. This gene encodes a transcription factor that is critical for embryonic hematopoiesis and the development of megakaryocytes and platelets in adult hematopoiesis [16]. Mutations in this gene are related to thrombocytopenia. Somatic mutations were identified in MDS, AML, chronic myelomonocytic leukemia, acute lymphoblastic leukemia, and chronic myeloid leukemia [17, 18].

This study aimed to identify molecular markers at diagnosis that indicate the risk of disease progression in LR-MDS patients. Transcriptome analysis was used to uncover signaling pathways involved in malignant transformation. We identified mutated *RUNX1* as the main molecular marker of rapid progression and described its effect on the disruption of the antitumor cellular response.

MATERIALS AND METHODS
Patient cohort
The study cohort consisted of 214 patients with de novo LR-MDS according to the IPSS. Forty-one patients (19%) progressed within 5 years. Progression was defined according to the revised International Working Group criteria [19]. All patients whose samples were used in this study

[1]Institute of Hematology and Blood Transfusion, Prague, Czech Republic. [2]First Faculty of Medicine, Charles University, Prague, Czech Republic. [3]Department of Biology, Faculty of Medicine and Dentistry, Palacky University, Olomouc, Czech Republic. [4]Faculty of Science, Charles University, Prague, Czech Republic. [5]International Clinical Research Center, St. Anne's University Hospital, Brno, Czech Republic. [6]Czech Technical University, Prague, Czech Republic. [7]First Department of Medicine, First Faculty of Medicine, Charles University and General University Hospital, Prague, Czech Republic. ✉email: monika.belickova@uhkt.cz

provided signed informed consent forms. The study was approved by the Institutional Scientific Board and the IHBT Ethics Committee (EK 4/AZV CR/ 06/2017) and was performed in accordance with the ethical standards of the Declaration of Helsinki. The median age of the cohort was 65 years (range, 20.8–86.5 years). The median follow-up period was 33.4 months (range, 0.2–183.0 months), and 133 (62%) patients were still alive. Twenty-eight patients underwent hematopoietic stem cell transplantation (HSCT), and for the purposes of this study, they were followed until the date of HSCT. The patient characteristics are summarized in SI 1.

Sequencing
Samples of BM or peripheral blood from diagnosis and, if available, from progression (90% of patients who progressed) were processed. Specific protocols for DNA and RNA isolation and detailed descriptions of targeted gene sequencing, Sanger sequencing, and RNA sequencing are reported in the Supplementary Methods.

Targeted gene sequencing. The sequencing library was prepared by the TruSight Myeloid Sequencing Panel Kit (Illumina, San Diego, CA, USA), which targets certain regions of 54 genes involved in hematological malignancies. NextGene software (SoftGenetics, State College, PA, USA) and an in-house pipeline were used for analysis of the output data. Variants were selected for further analysis if they met the following criteria: minimal coverage of 500x, Phred score greater than 35, and variant allele frequency (VAF) of ≥0.05. Variants were analyzed using 1000 Genomes, dbSNP, Varsome, ExAc, and other databases.

Sanger sequencing. Sanger sequencing was used to determine whether the mutations in *RUNX1* present at both diagnosis and progression with a VAF close to 0.5 were somatic or germline. We designed primer pairs for the amplification of exons 5–7, where these mutations were found by next-generation sequencing (NGS). Primer sequences are described in SI 2.

RNA sequencing. Seventy samples were sequenced (detailed in SI 3). For library preparation, the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA) was used. The processed data were analyzed by DAVID 6.8 and String 11.0 online tools using functional enrichment and analysis of protein–protein interaction networks. Furthermore, the data were analyzed using Gene Set Enrichment Analysis (GSEA) in GSEA software 3.0.

Machine learning
Two different techniques for the feature selection method applicable to Cox hazard models were used: stepwise backward feature selection and elastic network. Two different datasets were used: data1—binary mutational data and data2—the number of distinct mutations per gene. The details of the methods are given in the Supplementary Methods.

Immunohistochemistry
BM formalin-fixed paraffin-embedded (FFPE) sections (from four LR-MDS patients without *RUNX1* mutation and three LR-MDS patients with *RUNX1* mutation) were stained with rabbit anti-human γH2AX primary antibody (phosphoSer139, polyclonal; Cell Signaling, Danvers, MA, USA) as described in [20].

β-galactosidase detection
Six LR-MDS and six HR-MDS cryopreserved BM samples were thawed and washed in PBS with anti-clumping agent according to the manufacturer's instructions (Gibco, Waltham, MA, United States). The cells were washed twice in autoMACS rinsing buffer (Miltenyi Biotec, Bergisch Gladbach, Germany) and incubated for 1 hour (37 °C, 5% CO_2) with the β-galactosidase stain FITC (Senescence assay kit; Abcam, Cambridge, UK). Then, the cells were washed in PBS and stained for 30 min in a cocktail of antibodies specified in the Supplementary Methods. The cells were washed and directly measured at Cytek Aurora (Cytek, Fremont, CA, USA). The data were analyzed with the FlowJo software (BD, Franklin Lakes, NJ, USA).

Statistical analysis
MedCalc (MedCalc Software Ltd, Ostend, Belgium) was used to perform a Kaplan–Meier survival analysis, Cox proportional hazard regression (for univariate and multivariate analyses), the Mann–Whitney test, Fisher's exact test, and the chi-squared test. Graphs were created in GraphPad Prism 7 (GraphPad Software, La Jolla, CA, USA). Statistical level of significance was set at 0.05. Data were assumed to be non-normal (tested by Shapiro-Wilk test).

RESULTS
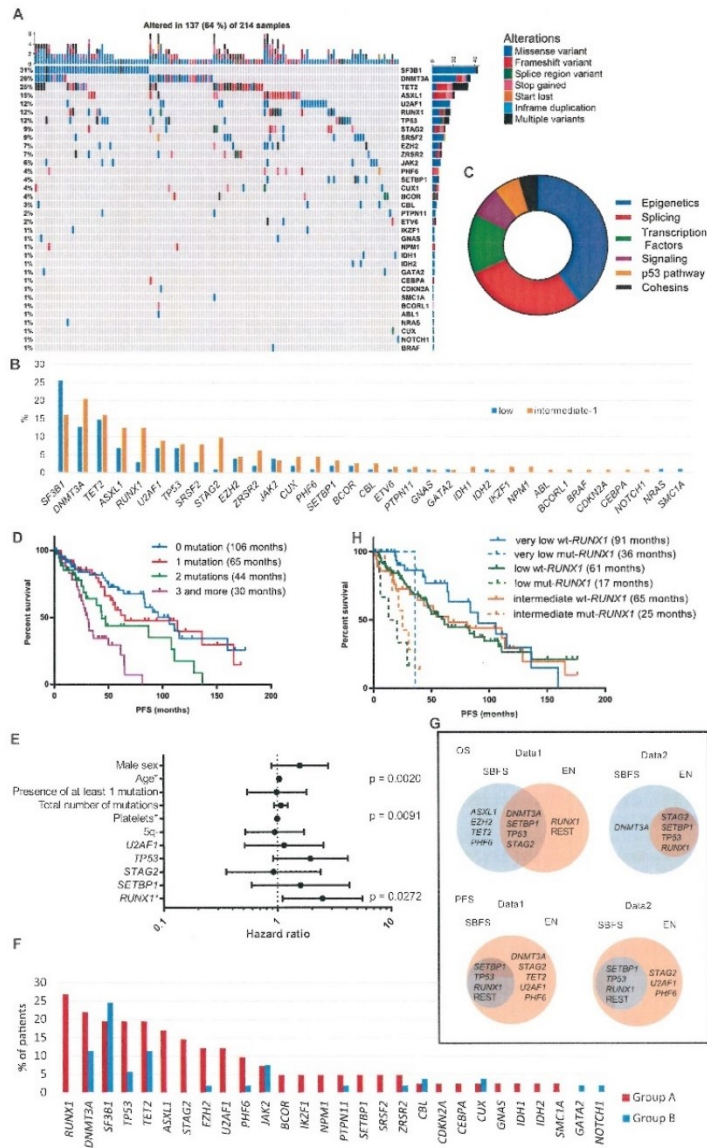### Mutational landscape of LR-MDS patients and survival analyses
We characterized the mutational landscape of 54 tested genes in the LR-MDS patient cohort at diagnosis (Fig. 1A). At least one pathogenic mutation was found in 137 patients (64%); in greater detail, pathogenic mutations were found in 53% of low-risk patients and 74% of INT-1. The number of mutations ranged from 0 to 9. The mutational complexity of co-occurrences is depicted in a Circos plot (SI 4A). The most common mutated gene was *SF3B1*, which was identified in 21% of patients, followed by *DNMT3A* in 17% of patients. The mutational profiles of the low-risk group and INT-1 group are depicted in Fig. 1B. In terms of functional categories, the most frequently mutated genes were epigenetic regulators (42%) (Fig. 1C) classified according to Sperling, Gibson, & Ebert [21].

Univariate analyses for overall survival (OS) and progression-free survival (PFS) (time from diagnosis until progression or death) were performed for BM blast count, cytopenias, IPSS and IPSS-R score, male sex, age, and presence of a 5q deletion and mutated genes (detected in more than five patients) (SI 5A). The significant variables in both analyses ($p < 0.05$) were platelet count, male sex, age, and the presence and total number of mutations. Significantly mutated genes for OS were *DNMT3A*, *RUNX1*, *SETBP1*, *STAG2*, and *TP53*, while mutated *RUNX1*, *SETBP1*, *STAG2*, *TP53*, and *U2AF1* were significant for PFS. OS and PFS decreased as the number of mutations increased (Fig. 1D). The presence of the deletion of 5q was significant for PFS and, in contrast to other variables, increased PFS. Neither IPSS nor IPSS-R showed significant differences between groups in our cohort (SI 6). However, adding information on the mutational status of genes that were significant in the univariate analysis led to great diversification of the OS and PFS curves among the groups (SI 7). Platelet count, age, and mutated *TP53* and *DNMT3A* were the most significant variables for OS in multivariate analysis of all significant variables from the univariate analysis (SI 5B). Considering a recent report on the effect of allelic status of *TP53* mutations on MDS prognosis [22], out of 16 patients carrying *TP53* mutations, 11 seemed to carry a monoallelic mutation. However, we could consider the allelic status only according to the number of identified mutations and their VAF. The median VAF of *TP53* mutations at diagnosis was 10% (range, 1–52%). Platelet count, age, and mutated *RUNX1* were the most significant independent prognostic factors in the multivariate analysis for PFS (Figs. 1E, SI 5C). Thus, the effect of *RUNX1* mutations on shortened PFS indicates its potential significance as a marker of rapid progression. Detailed statistical data are available in Supplement (SI 5).

The mutational landscape is different between patients with and without rapid progression
We compared the baseline characteristics of patients who progressed within 5 years (group A) to those without progression (or who progressed later than 5 years) (group B). We censored the patients who were not monitored for at least 5 years and patients who underwent HSCT up to 5 years from diagnosis. Therefore, 41 patients who progressed rapidly (group A) and 53 patients who did not progress (group B) were compared. The median time to progression in group A was 19.8 months.

Between these groups, significant differences were observed in the median age ($p = 0.0030$), male sex ($p = 0.0197$), and platelet count ($p = 0.0003$). The median OS was 33 months for group A and 136 months for group B ($p < 0.0001$) (SI 8). More detailed information on the patients is described in SI 9.

A — Altered in 137 (64 %) of 214 samples

Alterations
- Missense variant
- Frameshift variant
- Splice region variant
- Stop gained
- Start lost
- Inframe duplication
- Multiple variants

C — Epigenetics / Splicing / Transcription Factors / Signaling / p53 pathway / Cohesins

B — low / intermediate-1

D — 0 mutation (106 months) / 1 mutation (65 months) / 2 mutations (44 months) / 3 and more (30 months)

H — very low wt-RUNX1 (91 months) / very low mut-RUNX1 (36 months) / low wt-RUNX1 (61 months) / low mut-RUNX1 (17 months) / intermediate wt-RUNX1 (65 months) / intermediate mut-RUNX1 (25 months)

E — Male sex / Age* / Presence of at least 1 mutation / Total number of mutations* / Platelets* / 5q / U2AF1 / TP53 / STAG2 / SETBP1 / RUNX1*
p = 0.0020, p = 0.0091, p = 0.0272

G — OS Data1 / OS Data2 / PFS Data1 / PFS Data2

F — Group A / Group B

Eighty-five percent of the patients in group A and 47% of the patients in group B carried at least one mutation. The median number of mutations in group A was 3 (range 0–8), while it was 0 (range 0–5) in group B. The landscape of mutated genes was very different between the groups (Fig. 1F). The most commonly mutated gene in group A was RUNX1 (27%); in contrast, this gene was not mutated in group B at all. The most commonly mutated gene in group B was SF3B1 (25%), and this gene was mutated in 20% of patients in group A. Highly mutated genes in group A that were wild-type in group B included ASXL1, STAG2, and U2AF1.

4

Fig. 1 The landscape of mutated genes in the cohort of 214 LR-MDS patients. A Distribution, cooccurrence, and type of mutations in 137 of 214 LR-MDS patients. Each column represents an individual sample. The colored cells indicate a mutation in the gene described in the row on the right. The color indicates the type of alteration. The percentage on the left indicates the representation of mutated genes in 137 patients with mutations. The upper columns illustrate the number of mutations in the samples. The right stripes demonstrate the number of mutations of the gene throughout our cohort. **B** The most frequently mutated genes grouped by low and intermediate-1 IPSS risk groups. The Y-axis indicates the percent representation in the cohort. **C** Mutated genes grouped by functional categories. The most represented categories were epigenetic regulators (blue) and splicing regulators (red). **D** Effect of the number of mutations on PFS, p < 0.0001, with the median PFS in parentheses. **E** Multivariate analysis of mutational and clinical variables that were significant in univariate analysis of PFS depicted in a forest plot (hazard ratio, confidence intervals). Details are listed in SI 5C. * indicates significant independent prognostic factors. **F** The mutational landscape at the time of diagnosis in two groups of patients according to their progression within 5 years. Group A included patients who progressed within 5 years, and group B included patients who did not progress and were followed for at least 5 years. **G** Results of both machine learning methods (multivariate Cox regression with stepwise backward feature selection (SBFS) and elastic networks (EN)) applied to OS and PFS in datasets 1 (data1: binary mutational data) and 2 (data2: the number of distinct mutations per gene) depicted in Venn diagrams. The results of SBFS are depicted in blue circles, and the results of EN are depicted in orange circles. Common results are shown in overlaps. **H** Kaplan–Meier survival curves of patients stratified by IPSS-R and mutational status of the *RUNX1* gene, p < 0.0001, with the median OS in parentheses. wt-*RUNX1*, patients without *RUNX1* mutations, mut-*RUNX1*, patients with *RUNX1* mutations.

The mutational burden is higher during progression

We compared the mutational landscapes of paired samples from 36 patients who progressed within 5 years (before vs. after progression). We identified 24 new mutations in samples after progression. The greatest increase in the total number of mutations (114%) was observed in genes involved in signaling pathways (SI 10). Generally, the VAF of mutations increased from diagnosis to progression with few exceptions. Examples of VAF changes in paired samples are shown in SI 11.

Machine learning applied to mutational data confirms the significant effect of mutations on survival

According to the multivariate Cox regression with stepwise backward feature selection (SBFS), the mutated gene responsible for the shortest OS was *STAG2* in dataset 1 (binary mutational data) and *RUNX1* in dataset 2 (the number of distinct mutations per gene) (SI 12A). For the shortest PFS, *RUNX1* was mutated in both datasets.

According to the cross-validation experiments for SBFS and elastic network (EN) models (Supplementary Results), the optimal number of genes responsible for a shorter OS and PFS was greater than 1. The most significant genes are listed in Tables SI 12B and SI 13B for the individual datasets. Both methods identified mutated *DNMT3A*, *SETBP1*, *TP53*, and *STAG2* as significant for OS in dataset 1 and mutated *STAG2*, *SETBP1*, *TP53*, and *RUNX1* as significant for dataset 2. In both datasets, significant genes for shorter PFS identified by both methods were *SETBP1*, *TP53*, and *RUNX1*. The complex results from both methods are depicted in Figs. 1G and SI 14.

When the SBFS model was extended with comutational data (SI 15A-B), the presence of mutated *RUNX1* and *EZH2* together had the strongest impact on OS and PFS. In the EN approach, including gene interactions in the model did not improve its quality.

Because the initial number of independent variables was too large with respect to the number of events, the full models led to overfitting. The regularized models with smaller feature sets outperformed the full models. At the same time, they were significantly better than random, which confirms our hypothesis that risk stratification in MDS may be improved by including molecular data. The predicted hazards for the individual subjects could be used to assume their survival (described in the Supplementary Results).

RUNX1 mutational status can improve risk stratification of LR-MDS

We identified 25 unique mutations in *RUNX1* in 17 patients at diagnosis and in 2 patients who developed *RUNX1* mutations during progression (SI 16). Eighteen of the identified *RUNX1* mutations (75%) were located in the Runt homology domain (RUNT), which is responsible for DNA binding and interaction with CBFβ (SI 16). Overall, most mutations remove residues that are important for RUNX1 activity, suggesting a loss of RUNX1 function in these mutants [23]. Some mutations are likely dominant-negative [18], and in some mutants, the effect could not be predicted without functional assays

[24]. All *RUNX1* mutations were proven to be somatic (except for one presented in a patient whose CD3 + cells were not available). Most mutations were present at a lower VAF (<10%). *ASXL1*, *EZH2*, and *STAG2* were most frequently comutated with *RUNX1* (SI 4B).

RUNX1 mutational status significantly affected the IPSS-R scores (Fig. 1H). After adding information on *RUNX1* mutational status to the IPSS-R scoring system, the survival curves divided patients into two groups: i) patients with prolonged PFS from the three risk classes without any *RUNX1* mutation (wt-*RUNX1*) and ii) patients with shortened PFS with *RUNX1* mutations (mut-*RUNX1*).

Comparing the clinical features between *RUNX1*-mutated patients and others in our cohort revealed significant differences in the BM blast and platelet counts and the median number of mutations (SI 17).

The antitumor cellular response is downregulated in *RUNX1*-mutated LR-MDS

Because mutated *RUNX1* showed the greatest impact on rapid progression, we aimed to analyze the mechanism by which mutations in this gene contribute to rapid progression. We compared the transcriptomes of CD34 + cells between 8 *RUNX1*-mutated lower-risk patients (mutR-LR) and 29 lower-risk patients without *RUNX1* mutations (wtR-LR) (SI 3).

Hierarchical clustering (Fig. 2A) and principal component analysis (Fig. 2B) of RNA-seq data showed differences in the expression profiles of mutR-LR from those of wtR-LR. In the differential expression analysis of mutR-LR versus wtR-LR, 2235 genes were significantly (FDR < 0.05) upregulated and 2094 were significantly downregulated (Fig. 2C). Differentially expressed genes were enriched in 641 GO biological processes. GO enrichment analysis (GOrilla) [25] reduced this number to 103. The main pathways that had significant FDR values were chromatin and gene silencing, nucleosome assembly, chromatin organization, regulation of megakaryocyte differentiation and myeloid cell differentiation and hemopoiesis, telomere organization and capping, cellular metabolic processes, DNA damage response (DDR) and DNA repair, and cellular response to stress. The top 10 up- and down-regulated terms in GO biological processes are visualized in the Supplementary Material (SI 18A, B).

In the KEGG database, 47 pathways were significantly enriched. The top 10 upregulated KEGG pathways in mutR-LR were related to cancer and leukemia (SI 18C). The top 10 downregulated KEGG pathways were pathways of neurodegenerative diseases, inflammatory response, and cell cycle (Fig. 2D). These pathways are tightly connected to DDR and DNA repair, cellular senescence, aging, chronic inflammation, oxidative stress, and apoptosis [26–30], which all play a role in cellular tumor protection.

In our custom dataset consisting of 88 gene sets connected to DDR, DNA repair, cellular senescence, apoptosis, and hypoxia, 82 gene sets were significantly enriched in wtR-LR (FDR < 0.1). Enrichment plots and the heatmap of the top 50 genes are depicted in Fig. 2E, F.

To better understand the differences between mutR-LR and wtR-LR, we supplemented the cohort with 20 higher-risk patients (HR) and
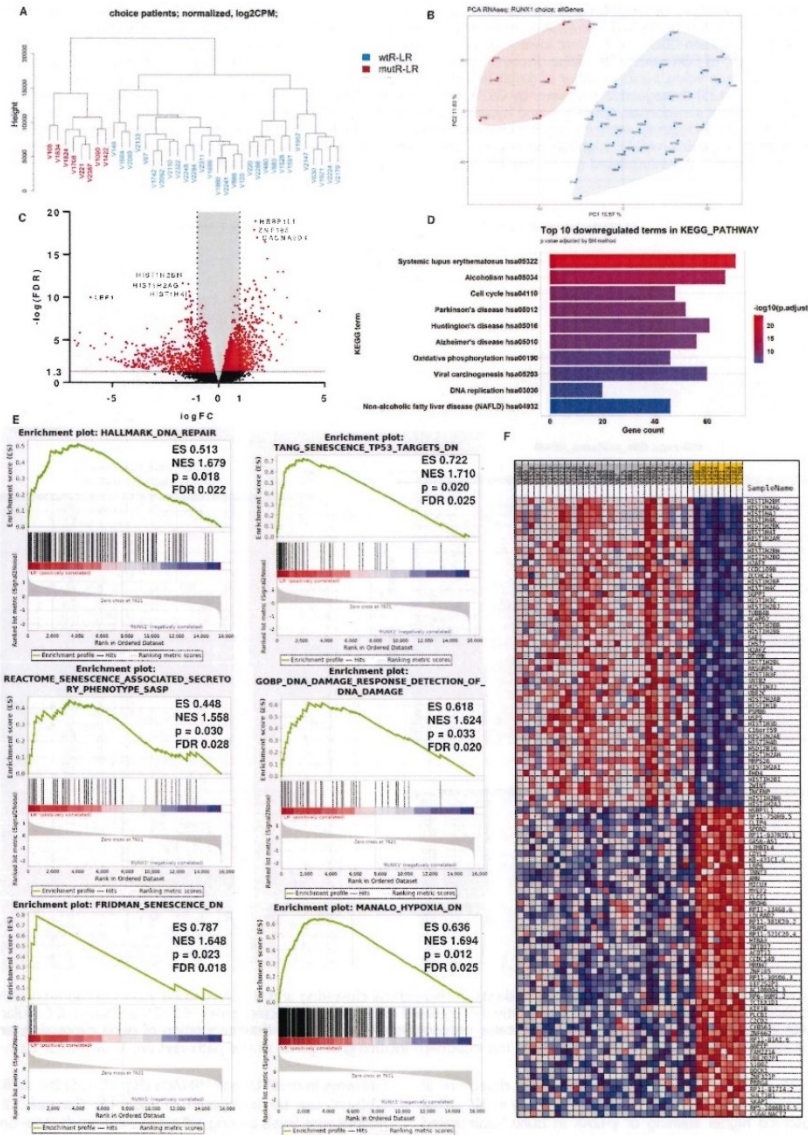
13 healthy controls (median age 41 years) (SI 3) and compared the expression profiles of CD34 + cells. Interestingly, mutR-LR patients clustered with HR patients (Fig. 3A, B). In dysregulated GSEA pathways, mutR-LR CD34 + cells transcriptionally resembled HR cells, indicating transcriptional similarity with HR patient cells already at diagnosis (Fig. 3C–E).

Markers of senescence are dysregulated in *RUNX1*-mutated LR-MDS and HR-MDS cells

To validate the suppression of DDR and senescence in cells of LR-MDS patients with *RUNX1* mutations and HR-MDS patients compared to that in cells of LR-MDS patients without *RUNX1* mutations, we performed two types of analysis: i) immunohistochemical staining of

Fig. 2 Transcriptome analysis of mutR-LR and wtR-LR RNA-seq data. A Hierarchical clustering and **B** PCA of *RUNX1*-mutated (mutR-LR) and *RUNX1*-unmutated LR-MDS patients (wtR-LR). **C** Differentially expressed genes depicted in a volcano plot. The red points indicate significantly dysregulated genes between CD34 + cells from lower-risk MDS patients with and without *RUNX1* mutations. x-axis: logFC, logarithm of fold-changes; y-axis: −log10 of FDR value; FDR - false discovery rate, red points: FDR < 0.05. **D** Top 10 downregulated KEGG pathways in mutR-LR compared to wtR-LR by p value. x-axis: number of genes in the pathway; color depicts adjusted p value (the highest values are red). **E** Six of 82 significantly (FDR < 0.25) dysregulated pathways by GSEA in the custom dataset consisting of 88 gene sets linked to the DNA repair, DNA damage response, cellular senescence, apoptosis, and hypoxia pathways. ES, enrichment score; NES, normalized enrichment score; p, p value; FDR, false discovery rate. **F** Heatmap representing the expression profiles of the top 50 up- and down-regulated genes in the custom dataset. mutR-LR highlighted in yellow, wtR-LR highlighted in gray. Gene expression levels are represented by colors; red represents upregulated genes and blue represents downregulated genes. The intensity indicates the level of differential expression.
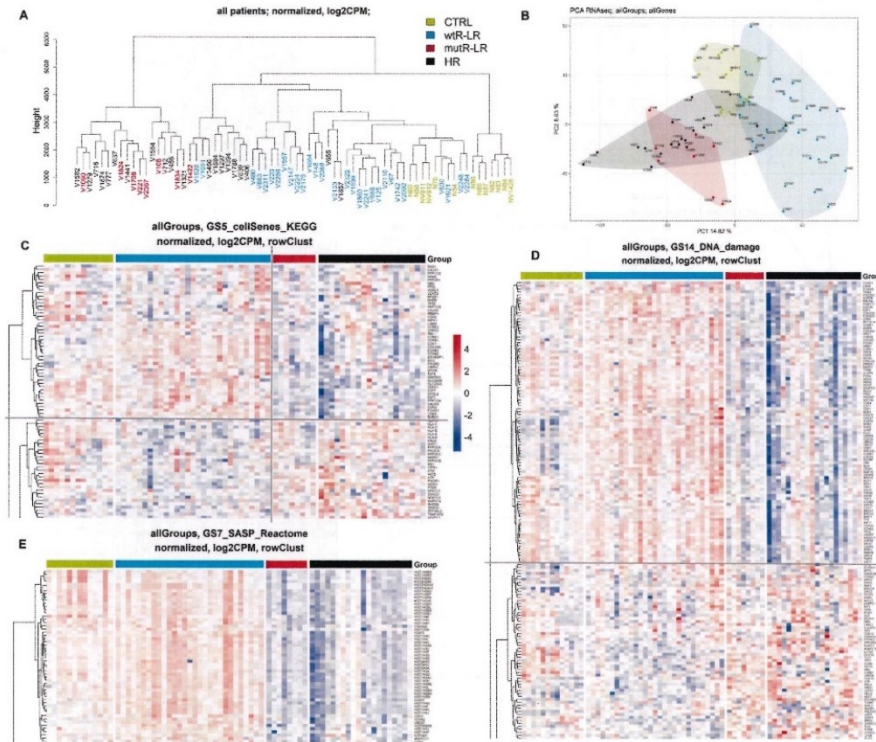

Fig. 3 Transcriptome analysis of LR- and HR-MDS patients. A Hierarchical clustering and **B** PCA of CD34 + cells of mutR-LR, wtR-LR, HR and healthy controls (CTRL). Heatmaps displaying significantly dysregulated expression in selected genes of GSEA pathways: **C** Cellular senescence (KEGG), **D** DNA damage, **E** SASP (Reactome). Red indicates upregulation, blue indicates downregulation of gene expression, and the color intensity indicates the level of differential expression. The heatmaps including all genes of the pathways are shown in SI 20.

γH2AX on BM FFPE sections and ii) fluorescence detection of senescence-associated β-galactosidase (SA-β-gal) activity in BM sorted cells. We observed higher staining of γH2AX in *RUNX1*-unmutated samples than in *RUNX1*-mutated samples, where the marker was very low or undetectable (Fig. 4A, B, SI 19). Furthermore, significantly higher SA-β-gal activity, indicating a higher percentage of senescent cells, was observed in CD14 + monocytes of LR-MDS compared to HR-MDS (Fig. 4C, D). The CD34 + cell results had to be omitted in statistical analyses because of the low number of CD34 + cells in samples, and mutR-LR samples were not available for this assay. However, based on the expression profiles of senescence-associated

pathways in mutR-LR and HR-MDS (Fig. 3C–E; SI 20), similar results for SA-β-gal can be anticipated in mutR-LR. Generally, the detected fluorescence levels among cell types were much more uniform in HR-MDS samples than in LR-MDS (Fig. 4E). A gating strategy example is depicted in SI 21.

DISCUSSION
This study aimed to describe the MDS mutational landscape using NGS technology, which is unique for a cohort composed exclusively of LR-MDS patients. To our knowledge, the only study exclusively
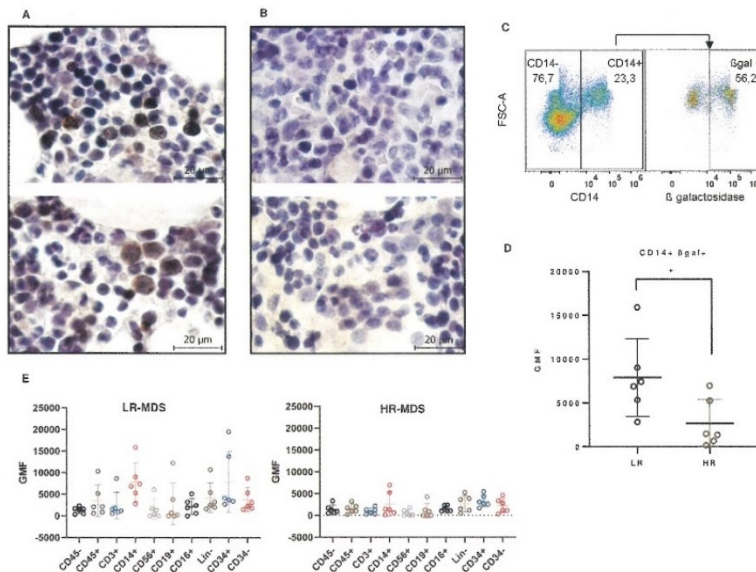
Fig. 4 Detection of markers of cellular senescence. Immunohistochemical staining of γH2AX protein in BM FFPE sections of **A** wtR-LR patients and **B** mutR-LR patients. Images of the two wtR-LR patients show highly positive zones of the sections that were not present in the mutR-LR patients. **C** Representative example of the gating strategy of CD14 + cells and SA-β-gal expression. The numbers in the plot indicate the percentage of gated cells. **D** Significant difference in activity (GMF) of SA-β-gal between LR- and HR-MDS CD14 + cells (median, interquartile range), $p = 0.026$, calculated using two-sided Mann-Whitney test. **E** Geometric mean fluorescence (GMF) represents the level of SA-β-gal activity for LR-MDS and HR-MDS patients in immune cell subsets from bone marrow (median, interquartile range).

targeting LR-MDS patients and aiming to enhance the prognostic system with molecular data thus far was published in 2012 and consisted of 288 LR-MDS patients [9]; however, few genes were sequenced, and the prognosis was based only on OS, not PFS. In this context, our study describes a novel, unique strategy for MDS stratification based on molecular markers and machine learning methods.

In our cohort, at least one pathogenic mutation was detected in 64% of patients. One of the most frequently mutated genes was *SF3B1*, which is in line with other studies [11, 12, 31]. This gene did not have a significant effect on OS, as reported earlier; however, these previous studies evaluated the effect in the entire spectrum of patients with MDS from low- to high-risk. Because *SF3B1* is predominantly mutated in lower-risk patients, its effect on survival may appear greater in the unstratified MDS cohort than in the lower-risk cohort.

Incorporation of the mutational status of genes affecting OS or PFS into IPSS-R significantly improved risk stratification. In multivariate analysis, age, platelet count, mutated *TP53* and *DNMT3A* were significant for OS, and age, platelets, and mutated *RUNX1* were significant for PFS. We have previously reported platelet count as well as mutated *TP53* as one of the strongest independent prognostic factors for OS in LR-MDS [32]. Unfavorable outcomes related to *RUNX1* mutations were described in a 16-study meta-analysis of MDS patients without risk stratification [33].

Machine learning is an emerging approach for risk stratification in various disorders, including MDS [13, 34, 35]. Nevertheless, to date, no algorithm has been used globally to stratify patients or predict the disease course. In our cohort, machine learning showed that mutated

RUNX1, *TP53*, and *SETBP1* are significant predictors of rapid progression, with *RUNX1* being the main factor.

Due to the strong effect of mutated *RUNX1* on PFS, we further aimed to investigate this gene and its role in progression. According to the VAF of *RUNX1* mutations and other commutated genes in *RUNX1*-mutated patients, we suppose that *RUNX1* mutations are not founder mutations but rather subsequent events in clonal evolution contributing to cell transformation. Similar conclusions were drawn by earlier studies [12, 36].

To determine the dysregulated molecular pathways associated with mutated *RUNX1*, we compared the expression profiles of CD34 + cells of LR-MDS patients with (mutR-LR) and without (wtR-LR) *RUNX1* mutations. Overall, data from differential expression analysis and GSEA showed suppression of pathways associated with antitumor cellular response—DDR, cellular senescence, chromatin and gene silencing, apoptosis, cellular response to stress, telomere maintenance, and hypoxia—in mutR-LR patients.

These data indicate the role of *RUNX1* as a tumor suppressor in LR-MDS and suggest a functional impact of *RUNX1* mutations, direct or indirect, in eliminating a biological anticancer barrier against accelerated progression in LR-MDS patients. We found that wtR-LR CD34 + cells activate the DDR and attain hallmarks of senescence, resulting in delayed progression. Indeed, senescence has been described as a part of the tumorigenesis barrier in premalignant lesions [37–39]. With the assumption that DDR and senescence are activated in the vicinity of senescent cells by senescence-associated secretory phenotype (SASP) [40, 41], we measured SA-β-gal expression in several BM sorted cell types and showed its significantly higher level, particularly in CD14 + monocytes of wtR-

LR-MDS. Our transcriptional comparison of SASP genes also suggests that senescence-associated inflammatory cytokine secretion (as described by Rodier et al. [42]) serves as a local microenvironmental mediator of the LR-MDS cellular state, contributing to the barrier against malignant progression and enforcing DDR activation, a phenomenon we proposed to be a barrier counteracting the progression of preleukemia to leukemia [43]. Our data also suggest that while some wtR-LR BM progenitors activate the DDR (marked by γH2AX), including increased DNA repair capacity consistent with proliferation, some wtR-LR BM cells suffer more DNA damage and undergo senescence. Thus, wtRUNX1 is functionally intertwined with DDR in LR-MDS in our cohort, and RUNX1 mutations are associated with elimination of the DDR-mediated senescence barrier and accelerated disease progression.

Several studies have shown that wtRUNX1 contributes to the protection of cells against oncogenesis. It is necessary for the p53 response to DNA damage [44], and knockdown of this gene may cause escape from senescence and enhance apoptosis suppression [45]. RUNX1 also interacts with a subunit of HIF1, HIF-1$_D$, and inhibits its transcriptional activity [46]. Overexpression of HIF-1$_D$ may result in tumor angiogenesis and tumor progression [47].

In our cohort, HIF1 and hypoxia cellular response pathways were significantly dysregulated in mutR-LR, which may impact the origin of senescence [48]. HIF1 and hypoxia are known to have an antisenescent effect [48–50]; however, they can induce the transcription of SASP genes and thus promote senescence in a paracrine fashion [48]. The dysregulation of HIF1 and hypoxia cellular response pathways has been described in various types of tumors [47, 51, 52].

Our data also show that mutR-LR cell expression profiles are more similar to those of HR-MDS cells than to those of wtR-LR cells at the time of diagnosis. We previously demonstrated that CD34 + cells of patients with early MDS show significant overexpression of genes involved in the cell cycle, DDR and DNA repair compared to those from advanced MDS patients [53]. Suppression of the DDR in AML cells versus MDS cells [54] and downregulation of homologous recombination gene expression in high-risk compared to low-risk MDS patients [55] have been reported. Similarly, a decrease in the expression of DNA damage checkpoints and dysregulation of the cell cycle were described in advanced MDS [56].

To conclude, this study shows that MDS risk stratification may be improved by including molecular data. Based on these data, we can identify patients at risk of rapid progression and choose proper follow-up and treatment strategies. LR-MDS patients with a RUNX1 mutation at diagnosis should be intensively monitored despite the lower-risk group. Transcriptome data suggest that RUNX1 mutations disrupt the fail-safe mechanism in hematopoietic stem cells and contribute to rapid progression in LR-MDS.

DATA AVAILABILITY
Raw data were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database (accession number PRJNA797993).

REFERENCES
1. Platzbecker U, Kubasch AS, Homer-Bouthiette C, Prebet T. Current challenges and unmet medical needs in myelodysplastic syndromes. Leukemia. 2021;35:2182–98.
2. Greenberg P, Cox C, LeBeau MM, Fenaux P, Morel P, Sanz G, et al. International scoring system for evaluating prognosis in myelodysplastic syndromes. Blood. 1997;89:2079–88.
3. Greenberg PL, Tuechler H, Schanz J, Sanz G, Garcia-Manero G, Solé F, et al. Revised international prognostic scoring system for myelodysplastic syndromes. Blood. 2012;120:2454–65.
4. Mufti GJ, McLornan DP, van de Loosdrecht AA, Germing U, Hasserjian RP. Diagnostic algorithm for lower-risk myelodysplastic syndromes. Leukemia. 2018;32:1679–96.
5. Giagounidis A. Current treatment algorithm for the management of lower-risk MDS. Hematol Am Soc Hematol Educ Progr. 2017;2017:453.
6. DeZern AE. Lower risk but high risk. Hematol Am Soc Hematol Educ Progr. 2021;2021:428–34.
7. Hou HA, Tsai CH, Lin CC, Chou WC, Kuo YY, Liu CY, et al. Incorporation of mutations in five genes in the revised International Prognostic Scoring System can improve risk stratification in the patients with myelodysplastic syndrome. Blood Cancer J. 2018;8:39.
8. Bejar R, Stevenson K, Abdel-Wahab O, Galili N, Nilsson B, Garcia-Manero G, et al. Clinical effect of point mutations in myelodysplastic syndromes. N Engl J Med. 2011;364:2496–506.
9. Bejar R, Stevenson KE, Caughey BA, Abdel-Wahab O, Steensma DP, Galili N, et al. Validation of a prognostic model and the impact of mutations in patients with lower-risk myelodysplastic syndromes. J Clin Oncol. 2012;30:3376–82.
10. Nazha A, Narkhede M, Radivoyevitch T, Seastone DJ, Patel BJ, Gerds AT, et al. Incorporation of molecular data into the Revised International Prognostic Scoring System in treated patients with myelodysplastic syndromes. Leukemia. 2016;30:2214–20.
11. Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia. 2014;28:241–7.
12. Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood. 2013;122:3616–27.
13. Nazha A, Komrokji RS, Barnard J, Al-Issa K, Padron E, Madanat YF, et al. A personalized prediction model to risk stratify patients with myelodysplastic syndromes (MDS). Blood. 2017;130:160–160.
14. Bersanelli M, Travaglino E, Meggendorfer M, Matteuzzi T, Sala C, Mosca E, et al. Classification and personalized prognostic assessment on the basis of clinical and genomic features in myelodysplastic syndromes. J Clin Oncol. 2021;39:1223–33.
15. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood Am Soc Hematol. 2016;127:2391–405.
16. Ichikawa M, Yoshimi A, Nakagawa M, Nishimoto N, Watanabe-Okochi N, Kurokawa M. A role for RUNX1 in hematopoiesis and myeloid leukemia. Int J Hematol. 2013;97:726–34.
17. Branford S, Wang P, Yeung DT, Thomson D, Purins A, Wadham C, et al. Integrative genomic analysis reveals cancer-associated mutations at diagnosis of CML in patients with high-risk disease. Blood. 2018;132:948–61.
18. Sood R, Kamikubo Y, Liu P. Role of RUNX1 in hematological malignancies. Blood. 2017;129:2070–82.
19. Cheson BD, Greenberg PL, Bennett JM, Lowenberg B, Wijermans PW, Nimer SD, et al. Clinical application and proposal for modification of the International Working Group (IWG) response criteria in myelodysplasia. Blood. 2006;108:419–25.
20. Stetka J, Vyhlidalova P, Lanikova L, Koralkova P, Gursky J, Hlusi A, et al. Addiction to DUSP1 protects JAK2V617F-driven polycythemia vera progenitors against inflammatory stress and DNA damage, allowing chronic proliferation. Oncogene. 2019;38:5627–42.
21. Sperling AS, Gibson CJ, Ebert BL. The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. Nat Rev Cancer. 2017;17:5–19.
22. Bernard E, Nannya Y, Hasserjian RP, Devlin SM, Tuechler H, Medina-Martinez JS, et al. Implications of TP53 allelic state for genome stability, clinical presentation and outcomes in myelodysplastic syndromes. Nat Med. 2020;26:1549–56.
23. Yokota A, Huo L, Lan F, Wu J, Huang G. The clinical, molecular, and mechanistic basis of RUNX1 mutations identified in hematological malignancies. Mol Cells. 2020;43:145.
24. Tsai SC, Shih LY, Liang ST, Huang YJ, Kuo MC, Huang CF, et al. Biological activities of RUNX1 mutants predict secondary acute leukemia transformation from chronic myelomonocytic leukemia and myelodysplastic syndromes. Clin Cancer Res. 2015;21:3541–51.
25. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinform. 2009;10:48.
26. Maiuri T, Suart CE, Hung CLK, Graham KJ, Barba Bazan CA, Truant R. DNA damage repair in Huntington's disease and other neurodegenerative diseases. Neurotherapeutics. 2019;16:948–56.
27. Merlo D, Mollinari C, Racaniello M, Garaci E, Cardinale A. DNA double strand breaks: a common theme in neurodegenerative diseases. Curr Alzheimer Res. 2016;13:1208–18.
28. Martínez-Cué C, Rueda N. Cellular senescence in neurodegenerative diseases. Front Cell Neurosci. 2020;14:16.
29. Radi E, Formichi P, Battisti C, Federico A. Apoptosis and oxidative stress in neurodegenerative diseases. J Alzheimer's Dis. 2014;42:S125–52.
30. Labadorf A, Choi SH, Myers RH. Evidence for a pan-neurodegenerative disease response in Huntington's and Parkinson's disease expression profiles. Front Mol Neurosci. 2018;10:430.

31. Malcovati L, Papaemmanuil E, Bowen DT, Boultwood J, Della Porta MG, Pascutto C, et al. Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. Blood. 2011;118:6239–46.
32. Belickova M, Vesela J, Jonasova A, Pejsova B, Votavova H, Merkerova MD, et al. TP53 mutation variant allele frequency is a potential predictor for clinical outcome of patients with lower-risk myelodysplastic syndromes. Oncotarget. 2016;7:36266–79.
33. He W, Zhao C, Hu H. Prognostic effect of RUNX1 mutations in myelodysplastic syndromes: a meta-analysis. Hematology. 2020;25:494–501.
34. Nagata Y, Zhao R, Awada H, Kerr CM, Mirzaev I, Kongkiatkamon S, et al. Machine learning demonstrates that somatic mutations imprint invariant morphologic features in myelodysplastic syndromes. Blood. 2020;136:2249–62.
35. Radakovich N, Meggendorfer M, Malcovati L, Hilton CB, Sekeres MA, Shreve J, et al. A geno-clinical decision model for the diagnosis of myelodysplastic syndromes. Blood Adv. 2021;5:4361–9.
36. Harada H, Harada Y. Recent advances in myelodysplastic syndromes: molecular pathogenesis and its implications for targeted therapies. Cancer Sci. 2015;106:329–36.
37. Campisi J. Cellular senescence as a tumor-suppressor mechanism. Trends Cell Biol. 2001;11:27–31.
38. Bartkova J, Rezaei N, Liontos M, Karakaidos P, Kletsas D, Issaeva N, et al. Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. Nature. 2006;444:633–7.
39. Bartkova J, Hořejší Z, Koed K, Krämer A, Tort F, Zieger K, et al. DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. Nature. 2005;434:864–70.
40. Bartek J, Hodny Z, Lukas J. Cytokine loops driving senescence. Nat Cell Biol. 2008;10:887–9.
41. Hubackova S, Krejcikova K, Bartek J, Hodny Z. IL1- and TGFβ-Nox4 signaling, oxidative stress and DNA damage response are shared features of replicative, oncogene-induced, and drug-induced paracrine "bystander senescence. Aging. 2012;4:932–51.
42. Rodier F, Coppé JP, Patil CK, Hoeijmakers WAM, Muñoz DP, Raza SR, et al. Persistent DNA damage signalling triggers senescence-associated inflammatory cytokine secretion. Nat Cell Biol. 2009;11:973–9.
43. Takacova S, Slany R, Bartkova J, Stranecky V, Dolezel P, Luzna P, et al. DNA damage response and inflammatory signaling limit the MLL-ENL-induced leukemogenesis in vivo. Cancer Cell. 2012;21:517–31.
44. Wu D, Ozakis T, Yoshiharas Y, Kubos N, Nakagawara A. Runt-related transcription factor 1 (RUNX1) stimulates tumor suppressor p53 protein in response to DNA damage through complex formation and acetylation. J Biol Chem. 2013;288:1353–64.
45. Motoda L, Osato M, Yamashita N, Jacob B, Chen LQ, Yanagida M, et al. Runx1 protects hematopoietic stem/progenitor cells from oncogenic insult. Stem Cells. 2007;25:2976–86.
46. Peng ZG, Zhou MY, Huang Y, Qiu JH, Wang LS, Liao SH, et al. Physical and functional interaction of Runt-related protein 1 with hypoxia-inducible factor-1α. Oncogene. 2008;27:839–47.
47. Poon E, Harris AL, Ashcroft M. Targeting the hypoxia-inducible factor (HIF) pathway in cancer. Expert Rev Mol Med. 2009;11(e26).
48. Welford SM, Giaccia AJ. Hypoxia and senescence: the impact of oxygenation on tumor suppression. Mol Cancer Res. 2011;9:538–44.
49. Eren MK, Tabor V. The role of hypoxia-inducible factor-1 alpha in bypassing oncogene-induced senescence. PLoS One. 2014;9:e101064.
50. van Vliet T, Varela-Eirin M, Wang B, Borghesan M, Brandenburg SM, Franzin R, et al. Physiological hypoxia restrains the senescence-associated secretory phenotype via AMPK-mediated mTOR suppression. Mol Cell. 2021;81:2041–2052.e6.
51. Simon F, Bockhorn M, Praha C, Baba HA, Broelsch CE, Frilling A, et al. Deregulation of HIF-1-alpha and hypoxia-regulated pathways in hepatocellular carcinoma and corresponding non-malignant liver tissue-influence of a modulated host stroma on the prognosis of HCC. Langenbeck's Arch Surg. 2010;395:395–405.
52. Schito L, Semenza GL. Hypoxia-inducible factors: master regulators of cancer progression. Trends Cancer. 2016;2:758–70.
53. Vasikova A, Belickova M, Budinska E, Cermak J. A distinct expression of various gene subsets in CD34+ cells from patients with early and advanced myelodysplastic syndrome. Leuk Res. 2010;34:1566–72.
54. Boehrer S, Adès L, Tajeddine N, Hofmann WK, Kriener S, Bug G, et al. Suppression of the DNA damage response in acute myeloid leukemia versus myelodysplastic syndrome. Oncogene. 2009;28:2205–18.
55. Valka J, Vesela J, Votavova H, Dostalova-Merkerova M, Horakova Z, Campr V, et al. Differential expression of homologous recombination DNA repair genes in the early and advanced stages of myelodysplastic syndrome. Eur J Haematol. 2017;99:323–31.
56. Pellagatti A, Cazzola M, Giagounidis A, Perry J, Malcovati L, Della Porta MG, et al. Deregulated gene expression pathways in myelodysplastic syndrome hematopoietic stem cells. Leukemia. 2010;24:756–64.

AUTHOR CONTRIBUTIONS
MK performed the data analysis and interpretation and drafted the manuscript. MB arranged funding and supervised the project. MB, HV, and VD reviewed the manuscript. MK, JV, MJ, and ZK processed the samples and performed the experiments. DK performed the bioinformatic analyses. JF, VD, and MDM processed and interpreted the data. JK performed machine learning analyses. DM, ML, JM, MSM, JSM, AJ, and JC were responsible for the patients' treatment and selected samples. All authors contributed to the article and approved the final version.

COMPETING INTERESTS
The authors declare no competing interests.

ADDITIONAL INFORMATION
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41375-022-01584-3.

Correspondence and requests for materials should be addressed to Monika Belickova.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

5. DISCUSSION

MDS are a group of highly heterogeneous diseases, and the molecular mechanisms underlying the disease pathogenesis are now in the center of interest. Using high-throughput technologies such as microarray assays and next-generation sequencing, we aimed to identify the molecular markers predictive of disease development at the level of lncRNA expression and recurrently mutated genes and to interpret their effect on disease biology through differential expression profiling. Throughout these studies, we engaged emerging computational techniques to support the power of our results, to prioritize candidate genes and to link specific lncRNAs to MDS-specific pathways.

ncRNAs play various roles in hematologic malignancies (Bhat et al., 2020; Ghafouri-Fard et al., 2020). They have regulatory functions in hematopoiesis, immune response, and apoptosis. They have tumor-suppressor or oncogenic potential, can serve as prognostic markers of disease evolution, and contribute to disease variability. However, only a small portion of all ncRNAs that contribute to hematologic malignancies have been discovered. Because deregulated expression of miRNAs has been comprehensively described in MDS, we focused on lncRNAs in this study.

To our knowledge, only a few studies have targeted BM lncRNAs in MDS. One of them studied lncRNAs in MDS to connect them with the outcome (Yao et al., 2017).This study showed that 4 lncRNAs together may have a prognostic effect, but it did not link lncRNAs to their biological functions. Another study presented the network-based lncRNA comodule function annotation method, which we also used in this publication (Liu et al., 2017). They identified a number of differentially expressed lncRNAs in MDS; however, they did not evaluate lncRNA expression in relation to patient outcome, diseases subtypes, or genetic abnormalities. Differentially expressed lncRNAs and PCGs between MDS patients and healthy controls have also been analyzed in one recent study (Wen et al., 2020).

Herein, we identified 32 deregulated lncRNAs and 87 deregulated PCGs between MDS patients and healthy controls. The main deregulated pathways were related to the hemoglobin complex and oxygen transport, immune response, epigenetic

modifications, and regulation of gene expression. Similarly, Pellagatti et al. (2010) showed that immune-related pathways were the most deregulated in MDS compared to healthy controls, and Wen et al. (2020) reported deregulation in necroptosis, apoptosis, immunodeficiency, p53, and FoxO signaling pathways in MDS.

Although many lncRNAs have been identified recently, their function has not been clarified. That is why we constructed coexpression networks and connected lncRNAs to MDS-associated cellular processes. For example, *EPB41L4A-AS1* has been reported to function as a repressor of the Warburg effect in cancer cells and a regulator of the cell cycle (Liao et al., 2019; Samdal et al., 2021). In our MDS patients, we associated the downregulation of *EPB41L4A-AS1* with ribosome formation and translational regulation.

We showed that lncRNA expression profiles differ between early and advanced stages of MDS. The expression profiles gradually changed from a healthy state to advanced myelodysplasia. Interestingly, MDS-EB1 clustered closer to early stages than to MDS-EB2. This may be a surprising result because MDS with excess blasts are usually related to poorer survival and a higher risk of AML transformation (Hasserjian, 2018), but at the level of miRNAs, similar results were published (Merkerova et al., 2011). In contrast, the lncRNA expression profile of MDS-EB2 imitated the AML-MRC profile, indicating that disease progression can be detected at the molecular level between MDS-EB1 and MDS-EB2, which is different from the classical scheme based on clinical variables (between MDS-EB2 and AML-MRC).

Numerous somatic mutations are found in MDS patients, and they play an important role in the pathogenesis of MDS. Therefore, we combined expression profiling data with the information on the mutational status of the five most often mutated genes in our cohort (*SF3B1*, *TP53*, *TET2*, *DNMT3A*, and *RUNX1*). In samples with mutated one from the first four genes, we found only a small number of affected transcripts. One could expect a larger impact of mutations in genes encoding spliceosomal factor, tumor suppressor, or epigenetic factor; genes with a wide range of targets. However, it is possible that a single nucleotide change might not be strong enough to induce a larger expression change. However, *RUNX1* mutations

caused high transcriptional impact, similar to the effect of cytogenetic aberrations. Mutations in *RUNX1* are related to worse outcomes in MDS (Chen et al., 2007; He et al., 2020). We connected the deregulation of the hematopoiesis and oncology-related *RAG1*, *LEF1* PCGs and *GAS5*, *LEF1-AS1,* and *TCL6* lncRNAs with *RUNX1* mutations. *RAG1* is a *RUNX1*-associated recombinase involved in T-cell receptor recombination (Cieslak et al., 2014). *RUNX1* upregulates the expression of *RAG1*, and the *RUNX1* DNA-binding domain is involved in this regulation (Jakobczyk et al., 2022). We found *RAG1* to be a core node in the DNA repair and recombination module of *RUNX1*-mutated deregulated genes and linked the *RUNX1-RAG1* axis with *LEF1/LEF1-AS* and *TCL6*, the same transcripts whose deregulation was significantly associated with poor prognosis.

Several transcriptomic studies have identified genes whose expression can be used to predict the outcome and sensitivity or resistance to treatment (Kim et al., 2020, 2021, Pellagatti et al., 2013, 2010; Prall et al., 2009; Shiozawa et al., 2017). However, PCG transcripts are not the final effectors in cells, unlike proteins and noncoding RNAs. Therefore, we assume that lncRNA expression should be a more reliable prognostic marker than PCG expression.

Herein, we identified four lncRNAs, *H19*, *WT1-AS*, *TCL6*, and *LEF1-AS1*, with a significant effect on outcomes. One of these four lncRNAs, *H19*, was the most promising prognostic marker. We demonstrated that an increased level of *H19* has strong prognostic value comparable to an increased blast count and the presence of *TP53* mutation, and it remained informative also in LR-MDS when the other variables did not. We associated the upregulation of *H19* with rapid progression, short OS, and altered cell adhesion and differentiation processes in CD34+ BM cells.

The aberrant expression of *H19* is associated with tumors; however, it has not yet been described in MDS. According to a review from 2015, *H19* is actively involved in all stages of tumorigenesis and is expressed in almost every human cancer (Raveh et al., 2015). It is involved in proliferation and differentiation. In a review from 2020, the expression of *H19* was connected with inflammation and was recognized as an age-related factor (B. Wang et al., 2020). *H19* seems to be a promising therapeutic target in various cancers (Raveh et al., 2015; J. Wang et al., 2020).

In AML, *H19* overexpression is linked to leukemogenesis and an unfavorable prognosis through its proliferative and antiapoptotic effects (Zhang et al., 2018).

H19 is only expressed maternally. Its counterpart is *IGF2*, which is expressed only from the paternal allele, and these two genes share one imprinting control region (Thorvaldsen et al., 1998). *H19* also functions as a primary template for miR-675, which plays an important role in tumorigenesis and the development of various cancers (He et al., 2015; Vennin et al., 2015). We identified transcriptional coregulation of *H19/IGF2*/miR-675 in healthy donors and LR-MDS, but disruption of this axis in HR-MDS. Deregulated expression of the *H19/IGF2* locus is presumably due to abnormal methylation of the locus that results in imprinting disruption, as described in other cancers (Kanduri et al., 2002; Park et al., 2017)

Downregulation of *TCL6* has been associated with a poor prognosis in patients with various cancers (Kulkarni et al., 2021; Luo et al., 2020; Yaqiong Zhang et al., 2020). It has been reported that *TCL6* behaves as a tumor suppressor and, through cooperation with miRNAs, regulates key signaling pathways in hepatocellular carcinoma and renal cell carcinoma (Kulkarni et al., 2021; Luo et al., 2020).

WT1-AS and *LEF1-AS1* are antisense transcripts of two PCGs, *WT1* and *LEF1*, which have an association with the prognosis of MDS patients (Pellagatti et al., 2013). *WT1-AS* participates in the regulation of tumor cell proliferation, the cell cycle and apoptosis and is also involved in tumor invasion and metastasis (Ye Zhang et al., 2020). *WT1* plays a role in cell differentiation and apoptosis, and *WT1* transcript monitoring is used to estimate minimal residual disease and predict outcomes in AML and MDS (Galimberti et al., 2010; Inoue et al., 1994; Nagasaki et al., 2017). *LEF1-AS1* probably has a tumor-suppressive function – inhibits proliferation and activates other tumor suppressors; thus, its level is decreased in myeloid malignancies (Congrains-Castillo et al., 2019). On the other hand, *LEF1-AS1* promotes the metastasis of prostate cancer by promoting proliferation, migration, and invasion (W. Li et al., 2020). *LEF1* participates in the proliferation and apoptosis of CD34+ progenitors and hematopoiesis (Skokowa et al., 2006). Its downregulation is related to a worse prognosis and progression of MDS (Pellagatti et al., 2009). We found that *LEF1-AS1* was transcriptionally coregulated

with *LEF1*; however, Congrains-Castillo et al. (2019) suggested that *LEF1-AS1* affects cell proliferation in a *LEF1*-independent manner.

Finally, our data functionally linked *WT1-AS* to *H19* and *LEF1-AS1* to *TCL6*. The *WT1-AS/H19* pair was associated with cell adhesion and differentiation, while the *LEF1-AS1/TCL6* pair participated in chromatin modification, cytokine response, and cell proliferation and death.

In the second study, our objective was to describe the mutational profile of LR-MDS patients and to identify markers of rapid progression. It is necessary to identify LR-MDS patients at a higher risk of rapid progression to ensure proper treatment. Many studies have described mutational profiles of MDS; however, very few have exclusively targeted LR-MDS patients. When analyzing this subgroup, slight but important differences can be distinguished. The study of Bejar et al. (2012) targeted LR-MDS patients to enhance the prognostic system with molecular data. However, few genes were sequenced, and the prognosis was based only on OS, not PFS. After publishing our manuscript, the IPSS-molecular was established and mutated genes associated with worse outcome have been proposed promising more accurate risk stratification (Bernard et al., 2022). However, in this context, our study provides new insights into the molecular pathogenesis of MDS in LR patients not only by molecular profiling supported by machine learning but also by studying the molecular changes in patients at risk of rapid progression.

At least one pathogenic mutation was detected in 64% of LR-MDS patients. One of the most frequently mutated genes was *SF3B1*, which corresponds to other studies (Haferlach et al., 2014; Malcovati et al., 2011; Papaemmanuil et al., 2013). However, this gene did not show a significant effect on OS, as previously reported. We suggest that this is a result of the individual study of LR-MDS. Previous studies have always evaluated the effect on the entire spectrum of MDS patients, and *SF3B1* is predominantly mutated in LR-MDS. Therefore, its effect on survival may appear greater in the unstratified MDS cohort than in the LR-MDS cohort.

In univariate analyses, mutated *DNMT3A*, *RUNX1*, *SETBP1*, *STAG2*, and *TP53* genes were significant for OS, and mutated *RUNX1*, *SETBP1*, *STAG2*, *TP53*, and *U2AF1* genes were significant for PFS. Additionally, a higher number of mutations decreased OS and PFS.

We supported our results by using a machine learning approach. It is an emerging methodology, and several studies show its potential for risk stratification and disease course prediction in various disorders, including MDS (Nagata et al., 2020; Nazha et al., 2017; Radakovich et al., 2021). Despite this, no algorithm has been included in MDS clinical practice. Herein, mutated *RUNX1*, *TP53*, and *SETBP1* genes were significant predictors of rapid progression according to machine learning. The mutated gene *RUNX1* was the strongest factor.

However, neither IPSS nor IPSS-R showed a distribution with significant differences in our cohort, and incorporation of the mutational status of genes affecting OS or PFS significantly improved risk stratification. Even incorporating only *RUNX1* mutational status significantly improved patient stratification.

In multivariate analysis, age, platelet count, mutated *TP53* and *DNMT3A* were significant for OS, and age, platelets, and mutated *RUNX1* were significant for PFS. Platelet count and mutated *TP53* have been previously reported as one of the strongest independent prognostic factors for OS in LR-MDS (Belickova et al., 2016). *RUNX1* mutations related to unfavorable outcomes were described in a 16-study meta-analysis of MDS patients without risk stratification (He et al., 2020). According to the VAF of *RUNX1* mutations and other commutated genes in *RUNX1*-mutated patients, we suppose that *RUNX1* mutations are not founder mutations but rather subsequent events in clonal evolution contributing to cell transformation. Previous studies have drawn similar conclusions (Harada and Harada, 2015; Papaemmanuil et al., 2013).

All our analyses showed that *RUNX1* is the strongest independent molecular prognostic factor for rapid progression. Therefore, we decided to analyze the impact of *RUNX1* mutations on transcriptional regulation. As we showed in Publication I, mutated *RUNX1* has a great impact on the transcriptome in the unstratified MDS cohort. In the cohort of LR-MDS patients, we observed an even greater number of deregulated genes.

In patients with rapid progression, we observed downregulation of pathways of chromatin and gene silencing, regulation of megakaryocyte differentiation and myeloid cell differentiation and hemopoiesis, telomere organization and capping, cellular metabolic processes, the DDR, cellular response to stress, cellular

senescence, apoptosis, aging, chronic inflammation, and hypoxia. On the other hand, pathways of leukemia and cancer were upregulated.

All the downregulated pathways mentioned above play a role in cellular tumor protection. These data suggest that wild-type *RUNX1* (wt*RUNX1*), a master regulator of hematopoiesis, is a tumor suppressor in LR-MDS and plays a role in eliminating a biological anticancer barrier against accelerated progression in LR-MDS patients. According to the literature, wt*RUNX1* is necessary for the p53 response to DNA damage (D. Wu et al., 2013); its knockdown may cause escape from senescence and enhance apoptosis suppression (Motoda et al., 2007). In AML, the tumor-suppressor function of *RUNX1* has been indicated due to analysis of homozygous mutations on *RUNX1* function (Silva et al., 2003). However, the dual role of *RUNX1* in myeloid leukemogenesis has been suggested (Goyama et al., 2013). It is possible that wt*RUNX1* is necessary for maintaining the cancer barrier, but the decreased level is needed for tumor growth. The oncogenic role of *RUNX1* was also suggested in T-cell acute lymphoblastic leukemia (Choi et al., 2017).

According to our data, wtR-LR CD34+ cells activate the DDR and attain hallmarks of senescence. Senescence has been described to be part of the tumorigenesis barrier in premalignant lesions (Bartkova et al., 2006, 2005; Campisi, 2001). One of the features of senescent cells is a senescence-associated secretory phenotype (SASP); senescent cells produce a variety of molecules that promote the inflammatory microenvironment and induce senescence in the vicinity. Transcriptional profiles of SASP genes showed a great increase in wtR-LR CD34+ cells. Thus, we detected senescence-associated β-galactosidase activity, one of the commonly used markers of cellular senescence, in BM sorted cell types and observed significantly higher senescence-associated β-galactosidase activity, particularly in CD14+ monocytes of wtR-LR.

Cellular senescence is closely associated with DNA damage. Excessive permanent DNA damage induces senescence in affected cells (Zglinicki et al., 2005), and the DDR probably plays a role in SASP production (Rodier et al., 2009). Based on our data, we hypothesize that while some wtR-LR BM progenitors activate the DDR and increase the DNA repair capacity consistent with proliferation, some

wtR-LR BM cells suffer more from DNA damage and undergo senescence. DDR activation plays an essential role in cellular protection against the progression of preleukemia to leukemia (Takacova et al., 2012). We used the protein expression of one DDR marker, γH2AX, to observe where the DDR is activated. We observed higher staining of the marker in *RUNX1*-unmutated samples than in *RUNX1*-mutated samples. This shows that *RUNX1* is functionally linked to the DDR in LR-MDS and its mutations are associated with elimination of the DDR-mediated senescence barrier and accelerate disease progression.

We also observed deregulation of *HIF1* and hypoxia cellular response pathways in mutR-LR. *HIF1* and hypoxia have been described to have an antisenescence effect (Eren and Tabor, 2014; van Vliet et al., 2021; Welford and Giaccia, 2011); however, they can promote the expression of SASP genes and thus induce senescence by paracrine signaling (Welford and Giaccia, 2011). The deregulation of *HIF1* and the hypoxia cellular response pathway has been described in various types of tumors (Poon et al., 2009; Schito and Semenza, 2016; Simon et al., 2010). RUNX1 inhibits the transcriptional activity of HIF1 and therefore protects against tumor angiogenesis and tumor progression (Peng et al., 2008).

Surprisingly, when supplementing our cohort with HR-MDS cases and healthy controls, we observed high transcriptional similarity of *RUNX1*-mutated LR-MDS cells and HR-MDS cells already at diagnosis, suggesting a possible efficacy of using a similar approach in clinical practice. The early and advanced stages of MDS have been previously reported to be transcriptionally different; early MDS show overexpression of genes involved in the cell cycle and DDR compared to advanced MDS (Pellagatti et al., 2010; Valka et al., 2017; Vasikova et al., 2010).

In both studies, we demonstrated the enormous impact of *RUNX1* mutations on MDS patient outcomes and the regulation of gene expression. We showed that pathways of immune response, cell death, and signaling pathways, especially the MAPK signaling pathway, translational regulation, RNA splicing, DNA repair, and p53 pathway, are critical, and their deregulation in *RUNX1*-mutated samples is associated with a higher risk of progression in LR-MDS as well as in the unstratified MDS cohort. We thus deduce that *RUNX1* mutations disrupt

the fail-safe mechanism in hematopoietic stem cells and contribute to rapid progression.

Taken together, the results presented in these two publications may provide novel potential therapeutic approaches based on four identified lncRNAs with the highest impact on prognosis, *H19, WT1-AS, TCL6*, and *LEF1-AS, RUNX1*, and senescent cells. Although cellular senescence protects cells from malignant transformation and tumor progression, it can also promote cancer itself. The main process involved in this process is the SASP, which induces an inflammatory environment, fertile soil for malignant transformation and tumor progression. By using senolytics, senescent cells can be eliminated. It may be useful in LR-MDS, where senescent cells are present at a higher density, or in MDS patients who underwent senescence-inducing treatment. This therapeutic approach could potentially reduce the risk of disease progression. The first reports of the elimination of premalignant lesions have been recently published (Kolodkin-Gal et al., 2022; Saleh and Carpenter, 2021). However, more studies are needed to support this hypothesis.

Based on our data, we indicate that LR-MDS patients with a *RUNX1* mutation at diagnosis should be intensively monitored despite being in the lower-risk group. Fortunately, the new IPSS-M includes information on *RUNX1* mutational status; thus, *RUNX1*-mutated patients should be stratified into higher-risk categories than in previous scoring systems.

To conclude, our findings provide novel information on particular lncRNAs and mutated genes contributing to MDS progression and propose cellular pathways involved in progression. It is worth emphasizing that the level of the *H19* transcript and mutated *RUNX1* gene may serve as robust independent prognostic markers comparable to clinical variables currently used for prognostication in MDS. Overall, we showed that molecular data could be used to identify patients at risk of rapid progression, and these findings could help to choose proper follow-up and treatment strategies.

6. CONCLUSIONS AND EVALUATION OF THE AIMS

This study identified lncRNAs and mutated genes associated with worse outcome and rapid progression in MDS patients. This finding enlightened new functions of these markers in MDS pathogenesis and progression. This knowledge may contribute to the accurate prognosis necessary for treatment decision-making. Additionally, the deepening of knowledge of MDS pathogenesis may point to promising therapeutic targets.

The aims were met, and the results are as follows:

We found novel biomarkers of adverse outcomes in MDS. At RNA level, we identified 4 lncRNAs, *H19*, *WT1-AS*, *LEF1-AS1*, and *TCL6*, associated with worse outcome. In particular, the level of *H19* was an independent prognostic factor for shorter OS and PFS. At DNA level, we identified genes associated with rapid progression in LR-MDS. Mutated *RUNX1*, *SETBP1*, *STAG2*, *TP53*, and *U2AF1* were significant for PFS by univariate analysis, and *SETBP1*, *TP53*, and *RUNX1* were significant for PFS by machine learning. The strongest independent prognostic factor was mutated *RUNX1*. We showed that molecular data improve the risk stratification and identify patients at risk of rapid progression.

We linked deregulated lncRNAs to cellular pathways with a lncRNA-PCG coexpression network and predicted their role in disease development. *WT1-AS* and *H19* are associated with cell adhesion and differentiation processes, and *LEF1-AS1* and *TCL6* are related to chromatin modification, cytokine response, and cell proliferation and death. Moreover, we reported disrupted transcriptional regulation in the *H19/IGF2* region in HR-MDS, suggesting the importance of this locus for disease development.

At the transcriptome level, we showed that *RUNX1* has a tumor-suppressive function in LR-MDS. In LR-MDS CD34+ cells, pathways of antitumor cellular defense are upregulated. However, mutations in the *RUNX1* gene probably disrupt this DDR-mediated senescence barrier and contribute to disease progression. LR-MDS patients with *RUNX1* mutations are thus at risk of rapid progression. Notably,

the expression profiles of these patients were more similar to those of HR-MDS than to those of other LR-MDS already at diagnosis. Based on our data, we suppose that rapid progression may be associated with the loss of cellular tumor barrier pathways in hematopoietic stem cells.

In both studies, we showed that pathways of immune response, cell death, signaling pathways, especially MAPK signaling pathway, translational regulation, RNA splicing, DNA repair, and p53 pathway are critical and their deregulation play a role in rapid progression of MDS patients.

7. ZÁVĚR A ZHODNOCENÍ CÍLŮ

Identifikovali jsme lncRNA a mutované geny asociované se závažnějším průběhem a časnou progresí u MDS pacientů. Také jsme popsali nové funkce těchto markerů v patogenezi a progresi MDS. Tyto znalosti mohou přispět jak k přesnějšímu určení prognózy, což je nezbytné ke zvolení správné léčby, tak i poukázat na potenciální terapeutické cíle.

Cíle byly splněny a výsledky jsou následující:

Objevili jsme nové biomarkery závažnějšího průběhu MDS. Na úrovní RNA jsme identifikovali 4 lncRNA, *H19*, *WT1-AS*, *LEF1-AS1* a *TCL6*, asociované s horším průběhem. Zvlášť hladina *H19* se ukázala být nezávislým prognostickým faktorem pro kratší celkové přežití i přežití do progrese. Na úrovní DNA jsme identifikovali geny asociované s časnou progresí u pacientů s nižším rizikem transformace do AML (LR-MDS). Mutované geny *RUNX1*, *SETBP1*, *STAG2*, *TP53* a *U2AF1* vyšly v univariantní analýze signifikantní pro přežití bez progrese. Geny *SETBP1*, *TP53*, *RUNX1* vyšly signifikantně i pomocí strojového učení. Nejsilnější nezávislý prognostický faktor byl mutovaný gen *RUNX1*. Prokázali jsme, že molekulární data mohou zlepšit stratifikaci podle rizika a identifikovat pacienty v riziku časné progrese.

Následně jsme propojili deregulované lncRNA s buněčnými drahami pomocí metody koexpresních sítí a predikovali jsme jejich roli ve vývoji onemocnění. *WT1-AS* a *H19* byly asociované s buněčnou adhezí a diferenciačními procesy, *LEF1-AS1* a *TCL6* s modifikací chromatinu, cytokinovou odpovědí a buněčnou proliferací i smrtí. Navíc jsme poukázali na narušenou transkripční regulaci v lokusu *H19/IGF2* u MDS pacientů s vyšším rizikem transformace, což naznačuje, že tento lokus hraje roli ve vývoji onemocnění.

Na úrovni transkriptomu jsme ukázali, že gen *RUNX1* plní tumor-supresorovou funkci u LR-MDS. V CD34+ buňkách LR-MDS jsou zvýšené dráhy buněčné protinádorové ochrany a mutace v genu *RUNX1* pravděpodobně tuto protinádorovou bariéru narušují, a přispívají tak k progresi onemocnění. LR-MDS pacienti s mutací v genu *RUNX1* jsou tedy v riziku časné progrese. Expresní

profily těchto pacientů jsou navíc podobné spíše těm s vysokým rizikem než ostatním s nižším rizikem. Na základě našich výsledků lze předpokládat, že časná progrese je asociována se ztrátou buněčné protinádorové bariéry v hematopoetických kmenových buňkách.

Výsledky obou publikací ukazují, že dráhy imunitní odpovědi, buněčné smrti, regulace translace, RNA sestřihu, DNA oprav, ale i dráha p53 a signální dráhy, především MAPK signální dráha, jsou klíčové a jejich deregulace hraje roli v časné progresi MDS pacientů.

REFERENCES

Acosta, J.C., O'Loghlen, A., Banito, A., Guijarro, M. V., Augert, A., Raguz, S., Fumagalli, M., Da Costa, M., Brown, C., Popov, N., Takatsu, Y., Melamed, J., d'Adda di Fagagna, F., Bernard, D., Hernando, E., Gil, J., 2008. Chemokine Signaling via the CXCR2 Receptor Reinforces Senescence. Cell 133, 1006–1018. https://doi.org/10.1016/J.CELL.2008.03.038

Aggerholm, A., Holm, M.S., Guldberg, P., Olesen, L.H., Hokland, P., 2006. Promoter hypermethylation of p15INK4B, HIC1, CDH1, and ER is frequent in myelodysplastic syndrome and predicts poor prognosis in early-stage patients. Eur. J. Haematol. 76, 23–32. https://doi.org/10.1111/J.1600-0609.2005.00559.X

Anderson, G., Mackay, N., Gilroy, K., Hay, J., Borland, G., McDonald, A., Bell, M., Hassanudin, S.A., Cameron, E., Neil, J.C., Kilbey, A., 2018. RUNX-mediated growth arrest and senescence are attenuated by diverse mechanisms in cells expressing RUNX1 fusion oncoproteins. J. Cell. Biochem. 119, 2750–2762. https://doi.org/10.1002/jcb.26443

Arber, D.A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M.J., Le Beau, M.M., Bloomfield, C.D., Cazzola, M., Vardiman, J.W., 2016. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood 127, 2391–2405. https://doi.org/10.1182/blood-2016-03-643544

Baker, D.J., Wijshake, T., Tchkonia, T., Lebrasseur, N.K., Childs, B.G., Van De Sluis, B., Kirkland, J.L., Van Deursen, J.M., 2011. Clearance of p16 Ink4a-positive senescent cells delays ageing-associated disorders. Nature 479, 232–236. https://doi.org/10.1038/nature10600

Bartel, D.P., 2004. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. Cell 116, 281–297. https://doi.org/10.1016/S0092-8674(04)00045-5

Bartkova, J., Hořejší, Z., Koed, K., Krämer, A., Tort, F., Zleger, K., Guldberg, P., Sehested, M., Nesland, J.M., Lukas, C., Orntoft, T., Lukas, J., Bartek, J., 2005. DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. Nature 434, 864–870. https://doi.org/10.1038/NATURE03482

Bartkova, J., Rezaei, N., Liontos, M., Karakaidos, P., Kletsas, D., Issaeva, N., Vassiliou, L.-V.F., Kolettas, E., Niforou, K., Zoumpourlis, V.C., Takaoka, M., Nakagawa, H., Tort, F., Fugger, K., Johansson, F., Sehested, M., Andersen, C.L., Dyrskjot, L., Ørntoft, T., Lukas, J., Kittas, C., Helleday, T., Halazonetis, T.D., Bartek, J., Gorgoulis, V.G., 2006. Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. Nature 444, 633–637. https://doi.org/10.1038/nature05268

Beck, D., Ayers, S., Wen, J., Brandl, M.B., Pham, T.D., Webb, P., Chang, C.C., Zhou, X., 2011. Integrative analysis of next generation sequencing for small non-coding RNAs and transcriptional regulation in Myelodysplastic Syndromes. BMC Med. Genomics 4, 19. https://doi.org/10.1186/1755-8794-4-19

Bejar, R., Stevenson, K., Abdel-Wahab, O., Galili, N., Nilsson, B., Garcia-Manero, G., Kantarjian, H., Raza, A., Levine, R.L., Neuberg, D., Ebert, B.L., 2011. Clinical Effect of Point Mutations in Myelodysplastic Syndromes. N. Engl. J. Med. 364, 2496–2506. https://doi.org/10.1056/nejmoa1013343

Bejar, R., Stevenson, K.E., Caughey, B.A., Abdel-Wahab, O., Steensma, D.P., Galili, N., Raza, A., Kantarjian, H., Levine, R.L., Neuberg, D., Garcia-Manero, G., Ebert, B.L., 2012. Validation of a prognostic model and the impact of mutations in patients with lower-risk myelodysplastic syndromes. J. Clin. Oncol. 30, 3376–3382. https://doi.org/10.1200/JCO.2011.40.7379

Belickova, M., Merkerova, M.D., Stara, E., Vesela, J., Sponerova, D., Mikulenkova, D., Brdicka, R., Neuwirtova, R., Jonasova, A., Cermak, J., 2013. DNA repair gene variants are associated with an increased risk of myelodysplastic syndromes in a Czech population. J. Hematol. Oncol. 6, 1–6. https://doi.org/10.1186/1756-8722-6-9

Belickova, M., Vesela, J., Jonasova, A., Pejsova, B., Votavova, H., Merkerova, M.D., Zemanova,

Z., Brezinova, J., Mikulenkova, D., Lauermannova, M., Valka, J., Michalova, K., Neuwirtova, R., Cermak, J., Belickova, M., Vesela, J., Jonasova, A., Pejsova, B., Votavova, H., Merkerova, M.D., Zemanova, Z., Brezinova, J., Mikulenkova, D., Lauermannova, M., Valka, J., Michalova, K., Neuwirtova, R., Cermak, J., 2016. TP53 mutation variant allele frequency is a potential predictor for clinical outcome of patients with lower-risk myelodysplastic syndromes. Oncotarget 7, 36266–36279. https://doi.org/10.18632/ONCOTARGET.9200

Benetatos, L., Hatzimichael, E., Dasoula, A., Dranitsaris, G., Tsiara, S., Syrrou, M., Georgiou, I., Bourantas, K.L., 2010. CpG methylation analysis of the MEG3 and SNRPN imprinted genes in acute myeloid leukemia and myelodysplastic syndromes. Leuk. Res. 34, 148–153. https://doi.org/10.1016/J.LEUKRES.2009.06.019

Bernard, E., Nannya, Y., Hasserjian, R.P., Devlin, S.M., Tuechler, H., Medina-Martinez, J.S., Yoshizato, T., Shiozawa, Y., Saiki, R., Malcovati, L., Levine, M.F., Arango, J.E., Zhou, Y., Solé, F., Cargo, C.A., Haase, D., Creignou, M., Germing, U., Zhang, Y., Gundem, G., Sarian, A., van de Loosdrecht, A.A., Jädersten, M., Tobiasson, M., Kosmider, O., Follo, M.Y., Thol, F., Pinheiro, R.F., Santini, V., Kotsianidis, I., Boultwood, J., Santos, F.P.S., Schanz, J., Kasahara, S., Ishikawa, T., Tsurumi, H., Takaori-Kondo, A., Kiguchi, T., Polprasert, C., Bennett, J.M., Klimek, V.M., Savona, M.R., Belickova, M., Ganster, C., Palomo, L., Sanz, G., Ades, L., Della Porta, M.G., Elias, H.K., Smith, A.G., Werner, Y., Patel, M., Viale, A., Vanness, K., Neuberg, D.S., Stevenson, K.E., Menghrajani, K., Bolton, K.L., Fenaux, P., Pellagatti, A., Platzbecker, U., Heuser, M., Valent, P., Chiba, S., Miyazaki, Y., Finelli, C., Voso, M.T., Shih, L.-Y., Fontenay, M., Jansen, J.H., Cervera, J., Atsuta, Y., Gattermann, N., Ebert, B.L., Bejar, R., Greenberg, P.L., Cazzola, M., Hellström-Lindberg, E., Ogawa, S., Papaemmanuil, E., 2020. Implications of TP53 allelic state for genome stability, clinical presentation and outcomes in myelodysplastic syndromes. Nat. Med. 2020 2610 26, 1549–1556. https://doi.org/10.1038/s41591-020-1008-z

Bernard, E., Tuechler, H., Greenberg, P.L., Hasserjian, R.P., Arango Ossa, J.E., Nannya, Y., Devlin, S.M., Creignou, M., Pinel, P., Monnier, L., Gundem, G., Medina-Martinez, J.S., Domenico, D., Jädersten, M., Germing, U., Sanz, G., van de Loosdrecht, A.A., Kosmider, O., Follo, M.Y., Thol, F., Zamora, L., Pinheiro, R.F., Pellagatti, A., Elias, H.K., Haase, D., Ganster, C., Ades, L., Tobiasson, M., Palomo, L., Della Porta, M.G., Takaori-Kondo, A., Ishikawa, T., Chiba, S., Kasahara, S., Miyazaki, Y., Viale, A., Huberman, K., Fenaux, P., Belickova, M., Savona, M.R., Klimek, V.M., Santos, F.P.S., Boultwood, J., Kotsianidis, I., Santini, V., Solé, F., Platzbecker, U., Heuser, M., Valent, P., Ohyashiki, K., Finelli, C., Voso, M.T., Shih, L.-Y., Fontenay, M., Jansen, J.H., Cervera, J., Gattermann, N., Ebert, B.L., Bejar, R., Malcovati, L., Cazzola, M., Ogawa, S., Hellström-Lindberg, E., Papaemmanuil, E., 2022. Molecular International Prognostic Scoring System for Myelodysplastic Syndromes. NEJM Evid. 1. https://doi.org/10.1056/evidoa2200008

Bersanelli, M., Travaglino, E., Meggendorfer, M., Matteuzzi, T., Sala, C., Mosca, E., Chiereghin, C., Di Nanni, N., Gnocchi, M., Zampini, M., Rossi, M., Maggioni, G., Termanini, A., Angelucci, E., Bernardi, M., Borin, L., Bruno, B., Bonifazi, F., Santini, V., Bacigalupo, A., Teresa Voso, M., Oliva, E., Riva, M., Ubezio, M., Morabito, L., Campagna, A., Saitta, C., Savevski, V., Giampieri, E., Remondini, D., Passamonti, F., Ciceri, F., Bolli, N., Rambaldi, A., Kern, W., Kordasti, S., Sole, F., Palomo, L., Sanz, G., Santoro, A., Platzbecker, U., Fenaux, P., Milanesi, L., Haferlach, T., Castellani, G., Della Porta, M.G., 2021. Classification and Personalized Prognostic Assessment on the Basis of Clinical and Genomic Features in Myelodysplastic Syndromes. J. Clin. Oncol. 39, 1223–1233. https://doi.org/10.1200/JCO.20.01659

Bewersdorf, J.P., Carraway, H., Prebet, T., 2020. Emerging treatment options for patients with high-risk myelodysplastic syndrome. Ther. Adv. Hematol. 11, 1–22. https://doi.org/10.1177/2040620720955006

Bhat, A.A., Younes, S.N., Raza, S.S., Zarif, L., Nisar, S., Ahmed, I., Mir, R., Kumar, S., Sharawat, S.K., Hashem, S., Elfaki, I., Kulinski, M., Kuttikrishnan, S., Prabhu, K.S., Khan, A.Q., Yadav, S.K., El-Rifai, W., Zargar, M.A., Zayed, H., Haris, M., Uddin, S., 2020. Role of non-coding RNA networks in leukemia progression, metastasis and drug resistance. Mol. Cancer. https://doi.org/10.1186/s12943-020-01175-9

Boehrer, S., Adès, L., Tajeddine, N., Hofmann, W.K., Kriener, S., Bug, G., Ottmann, O.G., Ruthardt, M., Galluzzi, L., Fouassier, C., Tailler, M., Olaussen, K.A., Gardin, C., Eclache, V., De Botton, S., Thepot, S., Fenaux, P., Kroemer, G., 2009. Suppression of the DNA damage response in acute myeloid leukemia versus myelodysplastic syndrome. Oncogene 28, 2205–2218. https://doi.org/10.1038/onc.2009.69

Boultwood, J., Pellagatti, A., Cattan, H., Lawrie, C.H., Giagounidis, A., Malcovati, L., Porta, M.G.D., Jädersten, M., Killick, S., Fidler, C., Cazzola, M., Hellström-Lindberg, E., Wainscoat, J.S., 2007. Gene expression profiling of CD34 + cells in patients with the 5q-syndrome. Br. J. Haematol. 139, 578–589. https://doi.org/10.1111/j.1365-2141.2007.06833.x

Braig, M., Lee, S., Loddenkemper, C., Rudolph, C., Peters, A.H.F.M., Schlegelberger, B., Stein, H., Dörken, B., Jenuwein, T., Schmitt, C.A., 2005. Oncogene-induced senescence as an initial barrier in lymphoma development. Nature 436, 660–665. https://doi.org/10.1038/nature03841

Branford, S., Wang, P., Yeung, D.T., Thomson, D., Purins, A., Wadham, C., Shahrin, N.H., Marum, J.E., Nataren, N., Parker, W.T., Geoghegan, J., Feng, J., Shanmuganathan, N., Mueller, M.C., Dietz, C., Stangl, D., Donaldson, Z., Altamura, H., Georgievski, J., Braley, J., Brown, A., Hahn, C., Walker, I., Kim, S.H., Choi, S.Y., Park, S.H., Kim, D.W., White, D.L., Yong, A.S.M., Ross, D.M., Scott, H.S., Schreiber, A.W., Hughes, T.P., 2018. Integrative genomic analysis reveals cancer-associated mutations at diagnosis of CML in patients with high-risk disease. Blood 132, 948–961. https://doi.org/10.1182/blood-2018-02-832253

Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K., Rassenti, L., Kipps, T., Negrini, M., Bullrich, F., Croce, C.M., 2002. Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc. Natl. Acad. Sci. U. S. A. 99, 15524–15529. https://doi.org/10.1073/pnas.242606799

Cammarata, G., Augugliaro, L., Salemi, D., Agueli, C., La Rosa, M., Dagnino, L., Civiletto, G., Messana, F., Marfia, A., Bica, M.G., Cascio, L., Floridia, P.M., Mineo, A.M., Russo, M., Fabbiano, F., Santoro, A., 2010. Differential expression of specific microRNA and their targets in acute myeloid leukemia. Am. J. Hematol. 85, 331–339. https://doi.org/10.1002/ajh.21667

Campisi, J., 2001. Cellular senescence as a tumor-suppressor mechanism. Trends Cell Biol. 11, S27–S31. https://doi.org/10.1016/S0962-8924(01)82148-6

Carraway, H.E., Saygin, C., 2020. Therapy for lower-risk MDS. Hematol. (United States) 20, 426–433. https://doi.org/10.1182/HEMATOLOGY.2020000127

Chen, C.Y., Lin, L.I., Tang, J.L., Ko, B.S., Tsay, W., Chou, W.C., Yao, M., Wu, S.J., Tseng, M.H., Tien, H.F., 2007. RUNX1 gene mutation in primary myelodysplastic syndrome - The mutation can be detected early at diagnosis or acquired during disease progression and is associated with poor outcome. Br. J. Haematol. 139, 405–414. https://doi.org/10.1111/j.1365-2141.2007.06811.x

Chen, J., Kao, Y.R., Sun, D., Todorova, T.I., Reynolds, D., Narayanagari, S.R., Montagna, C., Will, B., Verma, A., Steidl, U., 2019. Myelodysplastic syndrome progression to acute myeloid leukemia at the stem cell level. Nat. Med. 25, 103–110. https://doi.org/10.1038/s41591-018-0267-4

Chen, T.-C., Hou, H.-A., Chou, W.-C., Tang, J.-L., Kuo, Y.-Y., Chen, C.-Y., Tseng, M.-H., Huang, C.-F., Lai, Y.-J., Chiang, Y.-C., Lee, F.-Y., Liu, M.-C., Liu, C.-W., Liu, C.-Y., Yao, M., Huang, S.-Y., Ko, B.-S., Hsu, S.-C., Wu, S.-J., Tsay, W., Chen, Y.-C., Tien, H.-F., 2014. Dynamics of ASXL1 mutation and other associated genetic alterations during disease progression in patients with primary myelodysplastic syndrome. Blood Cancer J. 4, e177. https://doi.org/10.1038/BCJ.2013.74

Chen, Z., Trotman, L.C., Shaffer, D., Lin, H.K., Dotan, Z.A., Niki, M., Koutcher, J.A., Scher, H.I., Ludwig, T., Gerald, W., Cordon-Cardo, C., Pandolfi, P.P., 2005. Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. Nature 436,

725–730. https://doi.org/10.1038/nature03918

Chi, Y., Wang, D., Wang, J., Yu, W., Yang, J., 2019. Long non-coding RNA in the pathogenesis of cancers. Cells. https://doi.org/10.3390/cells8091015

Choi, Ah., Illendula, A., Pulikkan, J.A., Roderick, J.E., Tesell, J., Yu, J., Hermance, N., Zhu, L.J., Castilla, L.H., Bushweller, J.H., Kelliher, M.A., 2017. RUNX1 is required for oncogenic Myb and Myc enhancer activity in T-cell acute lymphoblastic leukemia. Blood 130, 1722–1733. https://doi.org/10.1182/blood-2017-03-775536

Cieslak, A., le Noir, S.L., Trinquand, A., Lhermitte, L., Franchini, D.M., Villarese, P., Gon, S., Bond, J., Simonin, M., Vanhile, L., Reimann, C., Verhoeyen, E., Larghero, J., Six, E., Spicuglia, S., André-Schmutz, I., Langerak, A., Nadel, B., Macintyre, E., Payet-Bornet, D., Asnafi, V., 2014. RUNX1-dependent RAG1 deposition instigates human TCR-δ locus rearrangement. J. Exp. Med. 211, 1821–1832. https://doi.org/10.1084/jem.20132585

Cimmino, A., Calin, G.A., Fabbri, M., Iorio, M. V., Ferracin, M., Shimizu, M., Wojcik, S.E., Aqeilan, R.I., Zupo, S., Dono, M., Rassenti, L., Alder, H., Volinia, S., Liu, C.G., Kipps, T.J., Negrini, M., Croce, C.M., 2005. miR-15 and miR-16 induce apoptosis by targeting BCL2. Proc. Natl. Acad. Sci. U. S. A. 102, 13944–13949. https://doi.org/10.1073/pnas.0506654102

Congrains-Castillo, A., Niemann, F.S., Santos Duarte, A.S., Olalla-Saad, S.T., 2019. LEF1-AS1, long non-coding RNA, inhibits proliferation in myeloid malignancy. J. Cell. Mol. Med. 23, 3021–3025. https://doi.org/10.1111/jcmm.14152

Coppé, J.P., Patil, C.K., Rodier, F., Sun, Y., Muñoz, D.P., Goldstein, J., Nelson, P.S., Desprez, P.Y., Campisi, J., 2008. Senescence-associated secretory phenotypes reveal cell-nonautonomous functions of oncogenic RAS and the p53 tumor suppressor. PLoS Biol. 6. https://doi.org/10.1371/journal.pbio.0060301

Da Silva-Coelho, P., Kroeze, L.I., Yoshida, K., Koorenhof-Scheele, T.N., Knops, R., Van De Locht, L.T., De Graaf, A.O., Massop, M., Sandmann, S., Dugas, M., Stevens-Kroef, M.J., Cermak, J., Shiraishi, Y., Chiba, K., Tanaka, H., Miyano, S., De Witte, T., Blijlevens, N.M.A., Muus, P., Huls, G., Van Der Reijden, B.A., Ogawa, S., Jansen, J.H., 2017. Clonal evolution in myelodysplastic syndromes. Nat. Commun. 8, 1–11. https://doi.org/10.1038/ncomms15099

De Witte, T., Bowen, D., Robin, M., Malcovati, L., Niederwieser, D., Yakoub-Agha, I., Mufti, G.J., Fenaux, P., Sanz, G., Martino, R., Alessandrino, E.P., Onida, F., Symeonidis, A., Passweg, J., Kobbe, G., Ganser, A., Platzbecker, U., Finke, J., Van Gelder, M., Van De Loosdrecht, A.A., Ljungman, P., Stauder, R., Volin, L., Deeg, H.J., Cutler, C., Saber, W., Champlin, R., Giralt, S., Anasetti, C., Kröger, N., 2017. Allogeneic hematopoietic stem cell transplantation for MDS and CMML: recommendations from an international expert panel. Blood 129, 1753–1762. https://doi.org/10.1182/BLOOD-2016-06-724500

DeZern, A.E., Dalton, W.B., 2022. How low risk are low risk myelodysplastic syndromes? Expert Rev. Hematol. 15, 15–24. https://doi.org/10.1080/17474086.2022.2029698

Di Nardo, C.D., Jabbour, E., Ravandi, F., Takahashi, K., Daver, N., Routbort, M., Patel, K.P., Brandt, M., Pierce, S., Kantarjian, H., Garcia-Manero, G., 2016. IDH1 and IDH2 mutations in myelodysplastic syndromes and role in disease progression. Leukemia 30, 980. https://doi.org/10.1038/LEU.2015.211

Dinmohamed, A.G., Visser, O., Van Norden, Y., Huijgens, P.C., Sonneveld, P., Van De Loosdrecht, A.A., Jongen-Lavrencic, M., 2014. Trends in incidence, initial treatment and survival of myelodysplastic syndromes: a population-based study of 5144 patients diagnosed in the Netherlands from 2001 to 2010. Eur. J. Cancer 50, 1004–1012. https://doi.org/10.1016/J.EJCA.2013.12.002

Dostalova Merkerova, M., Krejcik, Z., 2021. Transposable elements and Piwi-interacting RNAs in hemato-oncology with a focus on myelodysplastic syndrome. Int. J. Oncol. 59, 105. https://doi.org/10.3892/ijo

Dvinge, H., Kim, E., Abdel-Wahab, O., Bradley, R.K., 2016. RNA splicing factors as

oncoproteins and tumour suppressors. Nat. Rev. Cancer 16, 413–430. https://doi.org/10.1038/nrc.2016.51

Ebert, B.L., 2011. Molecular Dissection of the 5q Deletion in Myelodysplastic Syndrome. Semin. Oncol. 38, 621–626. https://doi.org/10.1053/j.seminoncol.2011.04.010

Ebert, B.L., Pretz, J., Bosco, J., Chang, C.Y., Tamayo, P., Galili, N., Raza, A., Root, D.E., Attar, E., Ellis, S.R., Golub, T.R., 2008. Identification of RPS14 as a 5q- syndrome gene by RNA interference screen. Nature 451, 335–339. https://doi.org/10.1038/nature06494

Eren, M.K., Tabor, V., 2014. The role of hypoxia inducible factor-1 alpha in bypassing oncogene-induced senescence. PLoS One 9, e101064. https://doi.org/10.1371/journal.pone.0101064

Faraoni, I., Consalvo, M.I., Aloisio, F., Fabiani, E., Giansanti, M., Di Cristino, F., Falconi, G., Tentori, L., Di Veroli, A., Curzi, P., Maurillo, L., Niscola, P., Lo-Coco, F., Graziani, G., Voso, M.T., 2019. Cytotoxicity and differentiating effect of the poly(ADP-ribose) polymerase inhibitor olaparib in myelodysplastic syndromes. Cancers (Basel). 11, 1373. https://doi.org/10.3390/cancers11091373

Fei, C., Zhao, Y., Guo, J., Gu, S., Li, X., Chang, C., 2014. Senescence of bone marrow mesenchymal stromal cells is accompanied by activation of p53/p21 pathway in myelodysplastic syndromes. Eur. J. Haematol. 93, 476–486. https://doi.org/10.1111/ejh.12385

Feringa, F.M., Raaijmakers, J.A., Hadders, M.A., Vaarting, C., Macurek, L., Heitink, L., Krenning, L., Medema, R.H., 2018. Persistent repair intermediates induce senescence. Nat. Commun. 9, 1–10. https://doi.org/10.1038/s41467-018-06308-9

Ferrer, R.A., Wobus, M., List, C., Wehner, R., Schönefeldt, C., Brocard, B., Mohr, B., Rauner, M., Schmitz, M., Stiehler, M., Ehninger, G., Hofbauer, L.C., Bornhäuser, M., Platzbecker, U., 2013. Mesenchymal stromal cells from patients with myelodyplastic syndrome display distinct functional alterations that are modulated by lenalidomide. Haematologica 98, 1677–1685. https://doi.org/10.3324/haematol.2013.083972

Figueroa, M.E., Skrabanek, L., Li, Y., Jiemjit, A., Fandy, T.E., Paietta, E., Fernandez, H., Tallman, M.S., Greally, J.M., Carraway, H., Licht, J.D., Gore, S.D., Melnick, A., 2009. MDS and secondary AML display unique patterns and abundance of aberrant DNA methylation. Blood 114, 3448–3458. https://doi.org/10.1182/blood-2009-01-200519

Galimberti, S., Ghio, F., Guerrini, F., Ciabatti, E., Grassi, S., Ferreri, M.I., Petrini, M., 2010. WT1 expression levels at diagnosis could predict long-term time-to-progression in adult patients affected by acute myeloid leukaemia and myelodysplastic syndromes. Br. J. Haematol. 149, 451–454. https://doi.org/10.1111/J.1365-2141.2009.08063.X

Garzon, R., Garofalo, M., Martelli, M.P., Briesewitz, R., Wang, L., Fernandez-Cymering, C., Volinia, S., Liu, C.G., Schnittger, S., Haferlach, T., Liso, A., Diverio, D., Mancini, M., Meloni, G., Foa, R., Martelli, M.F., Mecucci, C., Croce, C.M., Falini, B., 2008. Distinctive microRNA signature of acute myeloid leukemia bearing cytoplasmic mutated nucleophosmin. Proc. Natl. Acad. Sci. U. S. A. 105, 3945. https://doi.org/10.1073/PNAS.0800135105

Garzon, R., Liu, S., Fabbri, M., Liu, Z., Heaphy, C.E.A., Callegari, E., Schwind, S., Pang, J., Yu, J., Muthusamy, N., Havelange, V., Volinia, S., Blum, W., Rush, L.J., Perrotti, D., Andreeff, M., Bloomfield, C.D., Byrd, J.C., Chan, K., Wu, L.C., Croce, C.M., Marcucci, G., 2009. MicroRNA-29b induces global DNA hypomethylation and tumor suppressor gene reexpression in acute myeloid leukemia by targeting directly DNMT3A and 3B and indirectly DNMT1. Blood 113, 6411–6418. https://doi.org/10.1182/BLOOD-2008-07-170589

Gasek, N.S., Kuchel, G.A., Kirkland, J.L., Xu, M., 2021. Strategies for targeting senescent cells in human disease. Nat. Aging 2021 110 1, 870–879. https://doi.org/10.1038/s43587-021-00121-8

Gelsi-Boyer, V., Brecqueville, M., Devillier, R., Murati, A., Mozziconacci, M.J., Birnbaum, D.,

2012. Mutations in ASXL1 are associated with poor prognosis across the spectrum of malignant myeloid diseases. J. Hematol. Oncol. 5, 1–6. https://doi.org/10.1186/1756-8722-5-12/TABLES/1

Georgilis, A., Klotz, S., Hanley, C.J., Herranz, N., Weirich, B., Morancho, B., Leote, A.C., D'Artista, L., Gallage, S., Seehawer, M., Carroll, T., Dharmalingam, G., Wee, K.B., Mellone, M., Pombo, J., Heide, D., Guccione, E., Arribas, J., Barbosa-Morais, N.L., Heikenwalder, M., Thomas, G.J., Zender, L., Gil, J., 2018. PTBP1-Mediated Alternative Splicing Regulates the Inflammatory Secretome and the Pro-tumorigenic Effects of Senescent Cells. Cancer Cell 34, 85–102.e9. https://doi.org/10.1016/J.CCELL.2018.06.007

Gersuk, G.M., Beckham, C., Loken, M.R., Kiener, P., Anderson, J.E., Farrand, A., Troutt, A.B., Ledbetter, J.A., Deeg, H.J., 1998. A role for tumour necrosis factor-α, Fas and Fas-ligand in marrow failure associated with myelodysplastic syndrome. Br. J. Haematol. 103, 176–188. https://doi.org/10.1046/J.1365-2141.1998.00933.X

Geyh, S., Öz, S., Cadeddu, R.P., Fröbel, J., Brückner, B., Kündgen, A., Fenk, R., Bruns, I., Zilkens, C., Hermsen, D., Gattermann, N., Kobbe, G., Germing, U., Lyko, F., Haas, R., Schroeder, T., 2013. Insufficient stromal support in MDS results from molecular and functional deficits of mesenchymal stromal cells. Leukemia 27, 1841–1851. https://doi.org/10.1038/LEU.2013.193

Ghafouri-Fard, S., Esmaeili, M., Taheri, M., 2020. Expression of non-coding RNAs in hematological malignancies. Eur. J. Pharmacol. 875, 172976. https://doi.org/10.1016/J.EJPHAR.2020.172976

Gorgoulis, V.G., Vassiliou, L.V.F., Karakaidos, P., Zacharatos, P., Kotsinas, A., Liloglou, T., Venere, M., DiTullio, R.A., Kastrinakis, N.G., Levy, B., Kletsas, D., Yoneta, A., Herlyn, M., Kittas, C., Halazonetis, T.D., 2005. Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. Nature 434, 907–913. https://doi.org/10.1038/nature03485

Goyama, S., Schibler, J., Cunningham, L., Zhang, Y., Rao, Y., Nishimoto, N., Nakagawa, M., Olsson, A., Wunderlich, M., Link, K.A., Mizukawa, B., Grimes, H.L., Kurokawa, M., Liu, P.P., Huang, G., Mulloy, J.C., 2013. Transcription factor RUNX1 promotes survival of acute myeloid leukemia cells. J. Clin. Invest. 123, 3876–3888. https://doi.org/10.1172/JCI68557

Greenberg, P., Cox, C., LeBeau, M.M., Fenaux, P., Morel, P., Sanz, G., Sanz, M., Vallespi, T., Hamblin, T., Oscier, D., Ohyashiki, K., Toyama, K., Aul, C., Mufti, G., Bennett, J., 1997. International scoring system for evaluating prognosis in myelodysplastic syndromes. Blood 89, 2079–2088. https://doi.org/10.1182/blood.v89.6.2079

Greenberg, P.L., Tuechler, H., Schanz, J., Sanz, G., Garcia-Manero, G., Solé, F., Bennett, J.M., Bowen, D., Fenaux, P., Dreyfus, F., Kantarjian, H., Kuendgen, A., Levis, A., Malcovati, L., Cazzola, M., Cermak, J., Fonatsch, C., Le Beau, M.M., Slovak, M.L., Krieger, O., Luebbert, M., Maciejewski, J., Magalhaes, S.M.M., Miyazaki, Y., Pfeilstöcker, M., Sekeres, M., Sperr, W.R., Stauder, R., Tauro, S., Valent, P., Vallespi, T., Van De Loosdrecht, A.A., Germing, U., Haase, D., 2012. Revised international prognostic scoring system for myelodysplastic syndromes. Blood 120, 2454–2465. https://doi.org/10.1182/blood-2012-03-420489

Guo, Y., Strickland, S.A., Mohan, S., Li, S., Bosompem, A., Vickers, K.C., Zhao, S., Sheng, Q., Kim, A.S., 2017. MicroRNAs and tRNA-derived fragments predict the transformation of myelodysplastic syndromes to acute myeloid leukemia. Leuk. Lymphoma 58, 1. https://doi.org/10.1080/10428194.2016.1272680

Guo, Z., Zhang, S. kai, Zou, Z., Fan, R. hua, Lyu, X. dong, 2017. Prognostic significance of TET2 mutations in myelodysplastic syndromes: A meta-analysis. Leuk. Res. 58, 102–107. https://doi.org/10.1016/J.LEUKRES.2017.03.013

Haase, D., 2008. Cytogenetic features in myelodysplastic syndromes. Ann. Hematol. 87, 515. https://doi.org/10.1007/S00277-008-0483-Y

Haferlach, T., Nagata, Y., Grossmann, V., Okuno, Y., Bacher, U., Nagae, G., Schnittger, S., Sanada, M., Kon, A., Alpermann, T., Yoshida, K., Roller, A., Nadarajah, N., Shiraishi, Y.,

Shiozawa, Y., Chiba, K., Tanaka, H., Koeffler, H.P., Klein, H.U., Dugas, M., Aburatani, H., Kohlmann, A., Miyano, S., Haferlach, C., Kern, W., Ogawa, S., 2014. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia 28, 241–247. https://doi.org/10.1038/leu.2013.336

Halazonetis, T.D., Gorgoulis, V.G., Bartek, J., 2008. An oncogene-induced DNA damage model for cancer development. Science (80-.). 319, 1352–1355. https://doi.org/10.1126/science.1140735

Harada, H., Harada, Y., 2015. Recent advances in myelodysplastic syndromes: Molecular pathogenesis and its implications for targeted therapies. Cancer Sci. https://doi.org/10.1111/cas.12614

Hasserjian, R.P., 2018. Pathology of Myelodysplastic Syndrome With Excess Blasts: Overview, Epidemiology and Clinical Features, Morphologic Features. Medscape.

He, D., Wang, J., Zhang, C., Shan, B., Deng, X., Li, B., Zhou, Y., Chen, W., Hong, J., Gao, Y., Chen, Z., Duan, C., 2015. Down-regulation of miR-675-5p contributes to tumor progression and development by targeting pro-tumorigenic GPR55 in non-small cell lung cancer. Mol. Cancer 14, 1–14. https://doi.org/10.1186/s12943-015-0342-0

He, W., Zhao, C., Hu, H., 2020. Prognostic effect of RUNX1 mutations in myelodysplastic syndromes: a meta-analysis. Hematol. (United Kingdom) 25, 494–501. https://doi.org/10.1080/16078454.2020.1858598

Hernandez-Segura, A., Nehme, J., Demaria, M., 2018. Hallmarks of Cellular Senescence. Trends Cell Biol. 28, 436–453. https://doi.org/10.1016/j.tcb.2018.02.001

Heuser, M., Yun, H., Thol, F., 2018. Epigenetics in myelodysplastic syndromes. Semin. Cancer Biol. 51, 170–179. https://doi.org/10.1016/j.semcancer.2017.07.009

Hou, H.A., Tsai, C.H., Lin, C.C., Chou, W.C., Kuo, Y.Y., Liu, C.Y., Tseng, M.H., Peng, Y.L., Liu, M.C., Liu, C.W., Liao, X.W., Lin, L.I., Yao, M., Tang, J.L., Tien, H.F., 2018. Incorporation of mutations in five genes in the revised International Prognostic Scoring System can improve risk stratification in the patients with myelodysplastic syndrome. Blood Cancer J. 8, 39. https://doi.org/10.1038/s41408-018-0074-7

Huang, H.H., Chen, F.Y., Chou, W.C., Hou, H.A., Ko, B.S., Lin, C.T., Tang, J.L., Li, C.C., Yao, M., Tsay, W., Hsu, S.C., Wu, S.J., Chen, C.Y., Huang, S.Y., Tseng, M.H., Tien, H.F., Chen, R.H., 2019. Long non-coding RNA HOXB-AS3 promotes myeloid cell proliferation and its higher expression is an adverse prognostic marker in patients with acute myeloid leukemia and myelodysplastic syndrome. BMC Cancer 19, 1–14. https://doi.org/10.1186/s12885-019-5822-y

Hughes, J.M., Legnini, I., Salvatori, B., Masciarelli, S., Marchioni, M., Fazi, F., Morlando, M., Bozzoni, I., Fatica, A., 2015. C/EBPα-p30 protein induces expression of the oncogenic long non-coding RNA UCA1 in acute myeloid leukemia. Oncotarget 6, 18534–18544. https://doi.org/10.18632/ONCOTARGET.4069

Hung, S.Y., Lin, C.C., Hsu, C.L., Yao, C.Y., Wang, Y.H., Tsai, C.H., Hou, H.A., Chou, W.C., Tien, H.F., 2021. The expression levels of long non-coding RNA KIAA0125 are associated with distinct clinical and biological features in myelodysplastic syndromes. Br. J. Haematol. 192, 589–598. https://doi.org/10.1111/bjh.17231

Ichikawa, M., Yoshimi, A., Nakagawa, M., Nishimoto, N., Watanabe-Okochi, N., Kurokawa, M., 2013. A role for RUNX1 in hematopoiesis and myeloid leukemia. Int. J. Hematol. 97, 726–734. https://doi.org/10.1007/s12185-013-1347-3

Inoue, K., Sugiyama, H., Ogawa, H., Nakagawa, M., Yamagami, T., Miwa, H., Kita, K., Hiraoka, A., Masaoka, T., Nasu, K., Kyo, T., Dohy, H., Nakauchi, H., Ishidate, T., Akiyama, T., Kishimoto, T., 1994. WT1 as a new prognostic factor and a new marker for the detection of minimal residual disease in acute leukemia. Blood 84, 3071–3079. https://doi.org/10.1182/blood.v84.9.3071.3071

Jackson, S.P., Bartek, J., 2009. The DNA-damage response in human biology and disease. Nature

461, 1071–1078. https://doi.org/10.1038/nature08467

Jain, A.G., Elmariah, H., 2022. BMT for Myelodysplastic Syndrome: When and Where and How. Front. Oncol. 11, 771614. https://doi.org/10.3389/fonc.2021.771614

Jakobczyk, H., Jiang, Y., Debaize, L., Soubise, B., Avner, S., Sérandour, A.A., Rouger-Gaudichon, J., Rio, A.G., Carroll, J.S., Raslova, H., Gilot, D., Liu, Z., Demengeot, J., Salbert, G., Douet-Guilbert, N., Corcos, L., Galibert, M.D., Gandemer, V., Troadec, M.B., 2022. ETV6-RUNX1 and RUNX1 directly regulate RAG1 expression: one more step in the understanding of childhood B-cell acute lymphoblastic leukemia leukemogenesis. Leukemia 36, 549–554. https://doi.org/10.1038/s41375-021-01409-9

Jiang, L., Luo, Y., Zhu, S., Wang, L., Ma, L., Zhang, H., Shen, C., Yang, W., Ren, Y., Zhou, X., Mei, C., Ye, L., Xu, W., Yang, H., Lu, C., Jin, J., Tong, H., 2020. Mutation status and burden can improve prognostic prediction of patients with lower-risk myelodysplastic syndromes. Cancer Sci. 111, 580–591. https://doi.org/10.1111/cas.14270

Jiang, Y., Dunbar, A., Gondek, L.P., Mohan, S., Rataul, M., O'Keefe, C., Sekeres, M., Saunthararajah, Y., Maciejewski, J.P., O 'keefe, C., Sekeres, M., Saunthararajah, Y., Maciejewski, J.P., 2009. Aberrant DNA methylation is a dominant mechanism in MDS progression to AML. Blood 113, 1315–1325. https://doi.org/10.1182/BLOOD-2008-06-163246

Jongen-Lavrencic, M., Sun, S.M., Dijkstra, M.K., Valk, P.J.M., Löwenberg, B., 2008. MicroRNA expression profiling in relation to the genetic heterogeneity of acute myeloid leukemia. Blood 111, 5078–5085. https://doi.org/10.1182/BLOOD-2008-01-133355

Kanagal-Shamanna, R., Montalban-Bravo, G., Sasaki, K., Darbaniyan, F., Jabbour, E., Bueso-Ramos, C., Wei, Y., Chien, K., Kadia, T., Ravandi, F., Borthakur, G., Soltysiak, K.A., Routbort, M., Patel, K., Pierce, S., Medeiros, L.J., Kantarjian, H.M., Garcia-Manero, G., 2021. Only SF3B1 mutation involving K700E independently predicts overall survival in myelodysplastic syndromes. Cancer 127, 3552–3565. https://doi.org/10.1002/CNCR.33745

Kanduri, C., Kanduri, M., Liu, L., Thakur, N., Pfeifer, S., Ohlsson, R., 2002. The kinetics of deregulation of expression by de novo methylation of the H19 imprinting control region in cancer cells. Cancer Res. 62, 4545–4548.

Khoury, J.D., Solary, E., Abla, O., Akkari, Y., Alaggio, R., Apperley, J.F., Bejar, R., Berti, E., Busque, L., Chan, J.K.C., Chen, W., Chen, X., Chng, W.-J., Choi, J.K., Colmenero, I., Coupland, S.E., Cross, N.C.P., De Jong, D., Elghetany, M.T., Takahashi, E., Emile, J.-F., Ferry, J., Fogelstrand, L., Fontenay, M., Germing, U., Gujral, S., Haferlach, T., Harrison, C., Hodge, J.C., Hu, S., Jansen, J.H., Kanagal-Shamanna, R., Kantarjian, H.M., Kratz, C.P., Li, X.-Q., Lim, M.S., Loeb, K., Loghavi, S., Marcogliese, A., Meshinchi, S., Michaels, P., Naresh, K.N., Natkunam, Y., Nejati, R., Ott, G., Padron, E., Patel, K.P., Patkar, N., Picarsic, J., Platzbecker, U., Roberts, I., Schuh, A., Sewell, W., Siebert, R., Tembhare, P., Tyner, J., Verstovsek, S., Wang, W., Wood, B., Xiao, W., Yeung, C., Hochhaus, A., 2022. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. Leuk. 2022 1–17. https://doi.org/10.1038/s41375-022-01613-1

Kim, K., Park, S., Choi, H., Kim, H.J., Kwon, Y.R., Ryu, D., Kim, M., Kim, T.M., Kim, Y.J., 2020. Gene expression signatures associated with sensitivity to azacitidine in myelodysplastic syndromes. Sci. Reports 2020 101 10, 1–10. https://doi.org/10.1038/s41598-020-76510-7

Kim, S., Shin, D.Y., Kim, D., Oh, S., Hong, J., Kim, I., Kim, E., 2021. Gene expression profiles identify biomarkers of resistance to decitabine in myelodysplastic syndromes. Cells 10. https://doi.org/10.3390/CELLS10123494/S1

Kim, T., Tyndel, M.S., Kim, H.J., Ahn, J.S., Choi, S.H., Park, H.J., Kim, Y.K., Yang, D.H., Lee, J.J., Jung, S.H., Kim, S.Y., Min, Y.H., Cheong, J.W., Sohn, S.K., Moon, J.H., Choi, M., Lee, M., Zhang, Z., Kim, D.D.H., 2017. The clonal origins of leukemic progression of myelodysplasia. Leuk. 2017 319 31, 1928–1935. https://doi.org/10.1038/leu.2017.17

Kolodkin-Gal, D., Roitman, L., Ovadya, Y., Azazmeh, N., Assouline, B., Schlesinger, Y., Kalifa, R., Horwitz, S., Khalatnik, Y., Hochner-Ger, A., Imam, A., Demma, J.A., Winter, E., Benyamini, H., Elgavish, S., Khatib, A.A.S., Meir, K., Atlan, K., Pikarsky, E., Parnas, O., Dor, Y., Zamir, G., Ben-Porath, I., Krizhanovsky, V., 2022. Senolytic elimination of Cox2-expressing senescent cells inhibits the growth of premalignant pancreatic lesions. Gut 71, 345–355. https://doi.org/10.1136/gutjnl-2020-321112

Kordasti, S.Y., Afzali, B., Lim, Z., Ingram, W., Hayden, J., Barber, L., Matthews, K., Chelliah, R., Guinn, B., Lombardi, G., Farzaneh, F., Mufti, G.J., 2009. IL-17-producing CD4(+) T cells, pro-inflammatory cytokines and apoptosis are increased in low risk myelodysplastic syndrome. Br. J. Haematol. 145, 64–72. https://doi.org/10.1111/J.1365-2141.2009.07593.X

Krejcik, Z., Belickova, M., Hrustincova, A., Votavova, H., Jonasova, A., Cermak, J., Dyr, J.E., Merkerova, M.D., 2018. MicroRNA profiles as predictive markers of response to azacitidine therapy in myelodysplastic syndromes and acute myeloid leukemia. Cancer Biomarkers 22, 101–110. https://doi.org/10.3233/CBM-171029

Kulkarni, P., Dasgupta, P., Hashimoto, Y., Shiina, M., Shahryari, V., Tabatabai, Z.L., Yamamura, S., Tanaka, Y., Saini, S., Dahiya, R., Majid, S., 2021. A lncRNA TCL6-miR-155 Interaction Regulates the Src-Akt-EMT Network to Mediate Kidney Cancer Progression and Metastasis. Cancer Res. 81, 1500–1512. https://doi.org/10.1158/0008-5472.CAN-20-0832

Kumar, M.S., Narla, A., Nonami, A., Mullally, A., Dimitrova, N., Ball, B., McAuley, J.R., Poveromo, L., Kutok, J.L., Galili, N., Raza, A., Attar, E., Gilliland, D.G., Jacks, T., Ebert, B.L., 2011. Coordinate loss of a microRNA and protein-coding gene cooperate in the pathogenesis of 5q- syndrome. Blood 118, 4666–4673. https://doi.org/10.1182/BLOOD-2010-12-324715

Li, W., Yang, G., Yang, D., Li, D., Sun, Q., 2020. LncRNA LEF1-AS1 promotes metastasis of prostatic carcinoma via the Wnt/β-catenin pathway. Cancer Cell Int. 20. https://doi.org/10.1186/s12935-020-01624-x

Li, X., Xu, F., Wu, L.Y., Zhao, Y.S., Guo, J., He, Q., Zhang, Z., Chang, C.K., Wu, D., 2020. A genetic development route analysis on MDS subset carrying initial epigenetic gene mutations. Sci. Reports 2020 101 10, 1–8. https://doi.org/10.1038/s41598-019-55540-w

Li, Z., Lu, J., Sun, M., Mi, S., Zhang, H., Luo, R.T., Chen, P., Wang, Y., Yan, M., Qian, Z., Neilly, M.B., Jin, J., Zhang, Y., Bohlander, S.K., Zhang, D.E., Larson, R.A., Le Beau, M.M., Thirman, M.J., Golub, T.R., Rowley, J.D., Chen, J., 2008. Distinct microRNA expression profiles in acute myeloid leukemia with common translocations. Proc. Natl. Acad. Sci. U. S. A. 105, 15535. https://doi.org/10.1073/PNAS.0808266105

Liao, M., Liao, W., Xu, N., Li, B., Liu, F., Zhang, S., Wang, Y., Wang, S., Zhu, Y., Chen, D., Xie, W., Jiang, Y., Cao, L., Yang, B.B., Zhang, Y., 2019. LncRNA EPB41L4A-AS1 regulates glycolysis and glutaminolysis by mediating nucleolar translocation of HDAC2. EBioMedicine 41, 200–213. https://doi.org/10.1016/j.ebiom.2019.01.035

Lin, M.E., Hou, H.A., Tsai, C.H., Wu, S.J., Kuo, Y.Y., Tseng, M.H., Liu, M.C., Liu, C.W., Chou, W.C., Chen, C.Y., Tang, J.L., Yao, M., Li, C.C., Huang, S.Y., Ko, B.S., Hsu, S.C., Lin, C.T., Tien, H.F., 2018. Dynamics of DNMT3A mutation and prognostic relevance in patients with primary myelodysplastic syndrome. Clin. Epigenetics 10. https://doi.org/10.1186/S13148-018-0476-1

Lin, Y., Lin, Z., Cheng, K., Fang, Z., Li, Z., Luo, Y., Xu, B., 2017. Prognostic role of TET2 deficiency in myelodysplastic syndromes: A meta-analysis. Oncotarget 8, 43295. https://doi.org/10.18632/ONCOTARGET.17177

Lin, Y., Zheng, Y., Wang, Z.C., Wang, S.Y., 2016. Prognostic significance of ASXL1 mutations in myelodysplastic syndromes and chronic myelomonocytic leukemia: A meta-analysis. Hematology 21, 454–461. https://doi.org/10.1080/10245332.2015.1106815

Liu, K., Beck, D., Thoms, J.A.I., Liu, L., Zhao, W., Pimanda, J.E., Zhou, X., 2017. Annotating function to differentially expressed LincRNAs in myelodysplastic syndrome using a network-based method. Bioinformatics 33, 2622.

https://doi.org/10.1093/BIOINFORMATICS/BTX280

Liu, M., Wang, F., Zhang, Y., Chen, X., Cao, P., Nie, D., Fang, J., Wang, M., Liu, M., Liu, H., 2021. Gene mutation spectrum of patients with myelodysplastic syndrome and progression to acute myeloid leukemia. Int. J. Hematol. Oncol. 10. https://doi.org/10.2217/ijh-2021-0002

Lopes, M.R., Traina, F., Campos, P. de M., Pereira, J.K.N., Machado-Neto, J.A., Machado, H. da C., Gilli, S.C.O., Saad, S.T.O., Favaro, P., 2013. IL10 inversely correlates with the percentage of CD8[+] cells in MDS patients. Leuk. Res. 37, 541–546. https://doi.org/10.1016/J.LEUKRES.2013.01.019

Luo, L.H., Jin, M., Wang, L.Q., Xu, G.J., Lin, Z.Y., Yu, D.D., Yang, S.L., Ran, R.Z., Wu, G., Zhang, T., 2020. Long noncoding RNA TCL6 binds to miR-106a-5p to regulate hepatocellular carcinoma cells through PI3K/AKT signaling pathway. J. Cell. Physiol. 235, 6154–6166. https://doi.org/10.1002/JCP.29544

Ma, X., Zhang, W., Zhao, M., Li, S., Jin, W., Wang, K., 2020. Oncogenic role of lncRNA CRNDE in acute promyelocytic leukemia and NPM1-mutant acute myeloid leukemia. Cell Death Discov. 6, 1–13. https://doi.org/10.1038/s41420-020-00359-y

Ma, Y., Shen, J., Wang, L.X., 2018. Successful treatment of high-risk myelodysplastic syndrome with decitabine-based chemotherapy followed by haploidentical lymphocyte infusion A case report and literature review. Med. (United States) 97. https://doi.org/10.1097/MD.0000000000010434

Makishima, H., Yoshizato, T., Yoshida, K., Sekeres, M.A., Radivoyevitch, T., Suzuki, H., Przychodzen, B.J., Nagata, Y., Meggendorfer, M., Sanada, M., Okuno, Y., Hirsch, C., Kuzmanovic, T., Sato, Y., Sato-Otsubo, A., Laframboise, T., Hosono, N., Shiraishi, Y., Chiba, K., Haferlach, C., Kern, W., Tanaka, H., Shiozawa, Y., Gómez-Seguí, I., Husseinzadeh, H.D., Thota, S., Guinta, K.M., Dienes, B., Nakamaki, T., Miyawaki, S., Saunthararajah, Y., Chiba, S., Miyano, S., Shih, L.Y., Haferlach, T., Ogawa, S., MacIejewski, J.P., 2017. Dynamics of clonal evolution in myelodysplastic syndromes. Nat. Genet. 49, 204. https://doi.org/10.1038/NG.3742

Malcovati, L., Hellström-Lindberg, E., Bowen, D., Adès, L., Cermak, J., Del Cañizo, C., Della Porta, M.G., Fenaux, P., Gattermann, N., Germing, U., Jansen, J.H., Mittelman, M., Mufti, G., Platzbecker, U., Sanz, G.F., Selleslag, D., Skov-Holm, M., Stauder, R., Symeonidis, A., Van De Loosdrecht, A.A., De Witte, T., Cazzola, M., 2013. Diagnosis and treatment of primary myelodysplastic syndromes in adults: Recommendations from the European LeukemiaNet. Blood 122, 2943–2964. https://doi.org/10.1182/blood-2013-03-492884

Malcovati, L., Karimi, M., Papaemmanuil, E., Ambaglio, I., Jädersten, M., Jansson, M., Elena, C., Gallì, A., Walldin, G., Porta, M.G.D., Raaschou-Jensen, K., Travaglino, E., Kallenbach, K., Pietra, D., Ljungström, V., Conte, S., Boveri, E., Invernizzi, R., Rosenquist, R., Campbell, P.J., Cazzola, M., Lindberg, E.H., 2015. SF3B1 mutation identifies a distinct subset of myelodysplastic syndrome with ring sideroblasts. Blood 126, 233–241. https://doi.org/10.1182/blood-2015-03-633537

Malcovati, L., Papaemmanuil, E., Bowen, D.T., Boultwood, J., Della Porta, M.G., Pascutto, C., Travaglino, E., Groves, M.J., Godfrey, A.L., Ambaglio, I., Gallì, A., Da Vià, M.C., Conte, S., Tauro, S., Keenan, N., Hyslop, A., Hinton, J., Mudie, L.J., Wainscoat, J.S., Futreal, P.A., Stratton, M.R., Campbell, P.J., Hellström-Lindberg, E., Cazzola, M., 2011. Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. Blood 118, 6239–6246. https://doi.org/10.1182/blood-2011-09-377275

Menssen, A.J., Walter, M.J., 2020. Genetics of progression from MDS to secondary leukemia. Blood 136, 50. https://doi.org/10.1182/BLOOD.2019000942

Merkerova, M.D., Krejcik, Z., Votavova, H., Belickova, M., Vasikova, A., Cermak, J., 2011. Distinctive microRNA expression profiles in CD34+ bone marrow cells from patients with myelodysplastic syndrome. Eur. J. Hum. Genet. 19, 313. https://doi.org/10.1038/EJHG.2010.209

Mills, K.I., Kohlmann, A., Williams, P.M., Wieczorek, L., Liu, W.M., Li, R., Wei, W., Bowen, D.T., Loeffler, H., Hernandez, J.M., Hofmann, W.K., Haferlach, T., 2009. Microarray-based classifiers and prognosis models identify subgroups with distinct clinical outcomes and high risk of AML transformation of myelodysplastic syndrome. Blood 114, 1063–1072. https://doi.org/10.1182/BLOOD-2008-10-187203

Mittelman, M., Oster, H.S., Hoffman, M., Neumann, D., 2010. The lower risk MDS patient at risk of rapid progression. Leuk. Res. 34, 1551–1555. https://doi.org/10.1016/j.leukres.2010.05.023

Montes, P., Bernal, M., Campo, L.N., González-Ramírez, A.R., Jiménez, P., Garrido, P., Jurado, M., Garrido, F., Ruiz-Cabello, F., Hernández, F., 2019. Tumor genetic alterations and features of the immune microenvironment drive myelodysplastic syndrome escape and progression. Cancer Immunol. Immunother. 68, 2015–2027. https://doi.org/10.1007/S00262-019-02420-X

Motoda, L., Osato, M., Yamashita, N., Jacob, B., Chen, L.Q., Yanagida, M., Ida, H., Wee, H.-J., Sun, A.X., Taniuchi, I., Littman, D., Ito, Y., 2007. Runx1 Protects Hematopoietic Stem/Progenitor Cells from Oncogenic Insult . Stem Cells 25, 2976–2986. https://doi.org/10.1634/STEMCELLS.2007-0061

Mufti, G.J., McLornan, D.P., van de Loosdrecht, A.A., Germing, U., Hasserjian, R.P., 2018. Diagnostic algorithm for lower-risk myelodysplastic syndromes. Leukemia 32, 1679–1696. https://doi.org/10.1038/s41375-018-0173-2

Nagasaki, J., Aoyama, Y., Hino, M., Ido, K., Ichihara, H., Manabe, M., Ohta, T., Mugitani, A., 2017. Wilms Tumor 1 (WT1) mRNA Expression Level at Diagnosis Is a Significant Prognostic Marker in Elderly Patients with Myelodysplastic Syndrome. Acta Haematol. 137, 32–39. https://doi.org/10.1159/000452732

Nagata, Y., Zhao, R., Awada, H., Kerr, C.M., Mirzaev, I., Kongkiatkamon, S., Nazha, A., Makishima, H., Radivoyevitch, T., Scott, J.G., Sekeres, M.A., Hobbs, B.P., Maciejewski, J.P., 2020. Machine learning demonstrates that somatic mutations imprint invariant morphologic features in myelodysplastic syndromes. Blood 136, 2249–2262. https://doi.org/10.1182/BLOOD.2020005488

Nazha, A., Komrokji, R., Meggendorfer, M., Jia, X., Radakovich, N., Shreve, J., Beau Hilton, C., Nagata, Y., Hamilton, B.K., Mukherjee, S., Al Ali, N., Walter, W., Hutter, S., Padron, E., Sallman, D., Kuzmanovic, T., Kerr, C., Adema, V., Steensma, D.P., Dezern, A., Roboz, G., Garcia-Manero, G., Erba, H., Haferlach, C., Maciejewski, J.P., Haferlach, T., Sekeres, M.A., 2021. Personalized Prediction Model to Risk Stratify Patients With Myelodysplastic Syndromes. J. Clin. Oncol. 39, 3737–3746. https://doi.org/10.1200/JCO.20.02810

Nazha, A., Komrokji, R.S., Barnard, J., Al-Issa, K., Padron, E., Madanat, Y.F., Kuzmanovic, T., Abuhadra, N., Steensma, D.P., DeZern, A.E., Roboz, G.J., Garcia-Manero, G., List, A.F., Maciejewski, J.P., Sekeres, M.A., Meggendorfer, M., Jia, X., Radakovich, N., Shreve, J., Hilton, C.B., Nagata, Y., Hamilton, B.K., Mukherjee, S., Al Ali, N., Walter, W., Hutter, S., Padron, E., Sallman, D., Kuzmanovic, T., Kerr, C., Adema, V., Steensma, D.P., DeZern, A.E., Roboz, G.J., Garcia-Manero, G., Erba, H., Haferlach, C., Maciejewski, J.P., Haferlach, T., Sekeres, M.A., 2017. A Personalized Prediction Model to Risk Stratify Patients with Myelodysplastic Syndromes (MDS). Blood 130, 160–160. https://doi.org/10.1200/jco.20.02810

Nazha, A., Narkhede, M., Radivoyevitch, T., Seastone, D.J., Patel, B.J., Gerds, A.T., Mukherjee, S., Kalaycio, M., Advani, A., Przychodzen, B., Carraway, H.E., Maciejewski, J.P., Sekeres, M.A., 2016. Incorporation of molecular data into the Revised International Prognostic Scoring System in treated patients with myelodysplastic syndromes. Leukemia 30, 2214–2220. https://doi.org/10.1038/leu.2016.138

Neukirchen, J., Schoonen, W.M., Strupp, C., Gattermann, N., Aul, C., Haas, R., Germing, U., 2011. Incidence and prevalence of myelodysplastic syndromes: Data from the Düsseldorf MDS-registry. Leuk. Res. 35, 1591–1596. https://doi.org/10.1016/J.LEUKRES.2011.06.001

Ng, M., Heckl, D., Klusmann, J.-H., 2019. The Regulatory Roles of Long Noncoding RNAs in

Acute Myeloid Leukemia. Front. Oncol. 9, 570. https://doi.org/10.3389/fonc.2019.00570

Nolte, F., Hofmann, W.K., 2010. Molecular mechanisms involved in the progression of myelodysplastic syndrome. Future Oncol. 6, 445–455. https://doi.org/10.2217/FON.09.175

O'Connor, M.J., 2015. Targeting the DNA Damage Response in Cancer. Mol. Cell 60, 547–560. https://doi.org/10.1016/J.MOLCEL.2015.10.040

Ortiz-Montero, P., Londoño-Vallejo, A., Vernot, J.P., 2017. Senescence-associated IL-6 and IL-8 cytokines induce a self- and cross-reinforced senescence/inflammatory milieu strengthening tumorigenic capabilities in the MCF-7 breast cancer cell line. Cell Commun. Signal. 15. https://doi.org/10.1186/s12964-017-0172-3

Ozaki, T., Nakagawara, A., Nagase, H., 2013. RUNX Family Participates in the Regulation of p53-Dependent DNA Damage Response. Int. J. Genomics 2013, 1–12. https://doi.org/10.1155/2013/271347

Pagliuca, S., Gurnari, C., Visconte, V., 2021. Molecular Targeted Therapy in Myelodysplastic Syndromes: New Options for Tailored Treatments. Cancers (Basel). 13, 1–22. https://doi.org/10.3390/CANCERS13040784

Papaemmanuil, E., Gerstung, M., Malcovati, L., Tauro, S., Gundem, G., Van Loo, P., Yoon, C.J., Ellis, P., Wedge, D.C., Pellagatti, A., Shlien, A., Groves, M.J., Forbes, S.A., Raine, K., Hinton, J., Mudie, L.J., McLaren, S., Hardy, C., Latimer, C., Della Porta, M.G., O'Meara, S., Ambaglio, I., Galli, A., Butler, A.P., Walldin, G., Teague, J.W., Quek, L., Sternberg, A., Gambacorti-Passerini, C., Cross, N.C.P., Green, A.R., Boultwood, J., Vyas, P., Hellstrom-Lindberg, E., Bowen, D., Cazzola, M., Stratton, M.R., Campbell, P.J., 2013. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood 122, 3616–3627. https://doi.org/10.1182/blood-2013-08-518886

Park, K.S., Mitra, A., Rahat, B., Kim, K., Pfeifer, K., 2017. Loss of imprinting mutations define both distinct and overlapping roles for misexpression of IGF2 and of H19 lncRNA. Nucleic Acids Res. 45, 12766–12779. https://doi.org/10.1093/nar/gkx896

Parker, J.E., Fishlock, K.L., Mijovic, A., Czepulkowski, B., Pagliuca, A., Mufti, G.J., 1998. "Low-risk" myelodysplastic syndrome is associated with excessive apoptosis and an increased ratio of pro- versus anti-apoptotic bcl-2-related proteins. Br. J. Haematol. 103, 1075–1082. https://doi.org/10.1046/j.1365-2141.1998.01114.x

Parker, J.E., Mufti, G.J., Rasool, F., Mijovic, A., Devereux, S., Pagliuca, A., 2000. The role of apoptosis, proliferation, and the Bcl-2-related proteins in the myelodysplastic syndromes and acute myeloid leukemia secondary to MDS. Blood 96, 3932–3938. https://doi.org/10.1182/blood.v96.12.3932.h8003932_3932_3938

Patnaik, M.M., Hanson, C.A., Hodnefield, J.M., Lasho, T.L., Finke, C.M., Knudson, R.A., Ketterling, R.P., Pardanani, A., Tefferi, A., 2012. Differential prognostic effect of IDH1 versus IDH2 mutations in myelodysplastic syndromes: A Mayo Clinic Study of 277 patients. Leukemia 26, 101–105. https://doi.org/10.1038/leu.2011.298

Peduzzi, P., Concato, J., Feinstein, A.R., Holford, T.R., 1995. Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. J. Clin. Epidemiol. 48, 1503–1510. https://doi.org/10.1016/0895-4356(95)00048-8

Pellagatti, A., Benner, A., Mills, K.I., Cazzola, M., Giagounidis, A., Perry, J., Malcovati, L., Della Porta, M.G., Jädersten, M., Verma, A., McDonald, E.J., Killick, S., Hellström-Lindberg, E., Bullinger, L., Wainscoat, J.S., Boultwood, J., 2013. Identification of gene expression-based prognostic markers in the hematopoietic stem cells of patients with myelodysplastic syndromes. J. Clin. Oncol. 31, 3557–3564. https://doi.org/10.1200/JCO.2012.45.5626

Pellagatti, A., Boultwood, J., 2015. The molecular pathogenesis of the myelodysplastic syndromes. Eur. J. Haematol. 95, 3–15. https://doi.org/10.1111/EJH.12515

Pellagatti, A., Cazzola, M., Giagounidis, A., Perry, J., Malcovati, L., Della Porta, M.G., Jädersten, M., Killick, S., Verma, A., Norbury, C.J., Hellström-Lindberg, E., Wainscoat, J.S.,

Boultwood, J., 2010. Deregulated gene expression pathways in myelodysplastic syndrome hematopoietic stem cells. Leukemia 24, 756–764. https://doi.org/10.1038/leu.2010.31

Pellagatti, A., Hellström-Lindberg, E., Giagounidis, A., Perry, J., Malcovati, L., Della Porta, M.G., Jädersten, M., Killick, S., Fidler, C., Cazzola, M., Wainscoat, J.S., Boultwood, J., 2008. Haploinsufficiency of RPS14 in 5q- syndrome is associated with deregulation of ribosomal- and translation-related genes. Br. J. Haematol. 142, 57–64. https://doi.org/10.1111/j.1365-2141.2008.07178.x

Pellagatti, A., Marafioti, T., Paterson, J.C., Malcovati, L., Della Porta, M.G., Jädersten, M., Pushkaran, B., George, T.I., Arber, D.A., Killick, S., Giagounidis, A., Hellström-Lindberg, E., Cazzola, M., Wainscoat, J.S., Boultwood, J., 2009. Marked downregulation of the granulopoiesis regulator LEF1 is associated with disease progression in the myelodysplastic syndromes. Br. J. Haematol. 146, 86–90. https://doi.org/10.1111/j.1365-2141.2009.07720.x

Peng, Y., Croce, C.M., 2016. The role of MicroRNAs in human cancer. Signal Transduct. Target. Ther. 2016 11 1, 1–9. https://doi.org/10.1038/sigtrans.2015.4

Peng, Z.G., Zhou, M.Y., Huang, Y., Qiu, J.H., Wang, L.S., Liao, S.H., Dong, S., Chen, G.Q., 2008. Physical and functional interaction of Runt-related protein 1 with hypoxia-inducible factor-1α. Oncogene 27, 839–847. https://doi.org/10.1038/sj.onc.1210676

Pfeilstöcker, M., Tuechler, H., Sanz, G., Schanz, J., Garcia-Manero, G., Solé, F., Bennett, J.M., Bowen, D., Fenaux, P., Dreyfus, F., Kantarjian, H., Kuendgen, A., Malcovati, L., Cazzola, M., Cermak, J., Fonatsch, C., Le Beau, M.M., Slovak, M.L., Levis, A., Luebbert, M., Maciejewski, J., Machherndl-Spandl, S., Magalhaes, S.M.M., Miyazaki, Y., Sekeres, M.A., Sperr, W.R., Stauder, R., Tauro, S., Valent, P., Vallespi, T., Van De Loosdrecht, A.A., Germing, U., Haase, D., Greenberg, P.L., 2016. Time-dependent changes in mortality and transformation risk in MDS. Blood 128, 902–910. https://doi.org/10.1182/blood-2016-02-700054

Platzbecker, U., Kubasch, A.S., Homer-Bouthiette, C., Prebet, T., 2021. Current challenges and unmet medical needs in myelodysplastic syndromes. Leuk. 2021 358 35, 2182–2198. https://doi.org/10.1038/s41375-021-01265-7

Pons, A., Nomdedeu, B., Navarro, A., Gaya, A., Gel, B., Diaz, T., Valera, S., Rozman, M., Belkaid, M., Montserrat, E., Monzo, M., 2009. Hematopoiesis-related microRNA expression in myelodysplastic syndromes. Leuk. Lymphoma 50, 1854–1859. https://doi.org/10.3109/10428190903147645

Poon, E., Harris, A.L., Ashcroft, M., 2009. Targeting the hypoxia-inducible factor (HIF) pathway in cancer. Expert Rev. Mol. Med. 11. https://doi.org/10.1017/S1462399409001173

Popp, H.D., Naumann, N., Brendel, S., Henzler, T., Weiss, C., Hofmann, W.K., Fabarius, A., 2017. Increase of DNA damage and alteration of the DNA damage response in myelodysplastic syndromes and acute myeloid leukemias. Leuk. Res. 57, 112–118. https://doi.org/10.1016/j.leukres.2017.03.011

Prall, W.C., Czibere, A., Grall, F., Spentzos, D., Steidl, U., Giagounidis, A.A.N., Kuendgen, A., Otu, H., Rong, A., Libermann, T.A., Germing, U., Gattermann, N., Haas, R., Aivado, M., 2009. Differential gene expression of bone marrow-derived CD34+ cells is associated with survival of patients suffering from myelodysplastic syndrome. Int. J. Hematol. 89, 173–187. https://doi.org/10.1007/S12185-008-0242-9

Qin, T., Sotzen, J., Rampal, R.K., Rapaport, F.T., Levine, R.L., Klimek, V., Nimer, S.D., Figueroa, M.E., 2019. Risk of Disease Progression in Low-Risk MDS is Linked to DistinctEpigenetic Subtypes. Leukemia 33, 2753. https://doi.org/10.1038/S41375-019-0518-5

Radakovich, N., Meggendorfer, M., Malcovati, L., Hilton, C.B., Sekeres, M.A., Shreve, J., Rouphail, Y., Walter, W., Hutter, S., Galli, A., Pozzi, S., Elena, C., Padron, E., Savona, M.R., Gerds, A.T., Mukherjee, S., Nagata, Y., Komrokji, R.S., Jha, B.K., Haferlach, C., Maciejewski, J.P., Haferlach, T., Nazha, A., 2021. A geno-clinical decision model for the diagnosis of myelodysplastic syndromes. Blood Adv. 5, 4361–4369.

https://doi.org/10.1182/bloodadvances.2021004755

Raveh, E., Matouk, I.J., Gilon, M., Hochberg, A., 2015. The H19 Long non-coding RNA in cancer initiation, progression and metastasis - a proposed unifying theory. Mol. Cancer 14, 1–14. https://doi.org/10.1186/s12943-015-0458-2

Ribeiro, H.L., Soares Maia, A.R., Costa, M.B., Farias, I.R., de Paula Borges, D., de Oliveira, R.T.G., de Sousa, J.C., Magalhães, S.M.M., Pinheiro, R.F., 2016. Influence of functional polymorphisms in DNA repair genes of myelodysplastic syndrome. Leuk. Res. 48, 62–72. https://doi.org/10.1016/J.LEUKRES.2016.06.008

Robbins, P.D., Jurk, D., Khosla, S., Kirkland, J.L., LeBrasseur, N.K., Miller, J.D., Passos, J.F., Pignolo, R.J., Tchkonia, T., Niedernhofer, L.J., 2021. Senolytic Drugs: Reducing Senescent Cell Viability to Extend Health Span. Annu. Rev. Pharmacol. Toxicol. 61, 779–803. https://doi.org/10.1146/annurev-pharmtox-050120-105018

Rodier, F., Coppé, J.P., Patil, C.K., Hoeijmakers, W.A.M., Muñoz, D.P., Raza, S.R., Freund, A., Campeau, E., Davalos, A.R., Campisi, J., 2009. Persistent DNA damage signalling triggers senescence-associated inflammatory cytokine secretion. Nat. Cell Biol. 11, 973–979. https://doi.org/10.1038/NCB1909

Saleh, T., Carpenter, V.J., 2021. Potential use of senolytics for pharmacological targeting of precancerous lesions. Mol. Pharmacol. 100, 580–587. https://doi.org/10.1124/molpharm.121.000361

Sallman, D.A., Komrokji, R., Vaupel, C., Cluzeau, T., Geyer, S.M., McGraw, K.L., Al Ali, N.H., Lancet, J., McGinniss, M.J., Nahas, S., Smith, A.E., Kulasekararaj, A., Mufti, G., List, A., Hall, J., Padron, E., 2015. Impact of TP53 mutation variant allele frequency on phenotype and outcomes in myelodysplastic syndromes. Leuk. 2016 303 30, 666–673. https://doi.org/10.1038/leu.2015.304

Samdal, H., Hegre, S.A., Chawla, K., Liabakk, N.B., Aas, P.A., ..., 2021. The lncRNA EPB41L4A-AS1 regulates gene expression in the nucleus and exerts cell type-dependent effects on cell cycle progression. bioRxiv PREPRINT. https://doi.org/10.1101/2021.02.10.430566

Santini, V., 2022. Treatment of Lower Risk Myelodysplastic Syndromes. hemato 3, 153–162. https://doi.org/10.3390/hemato3010013

Savarese, F., Flahndorfer, K., Jaenisch, R., Busslinger, M., Wutz, A., 2006. Hematopoietic precursor cells transiently reestablish permissiveness for X inactivation. Mol. Cell. Biol. 26, 7167–7177. https://doi.org/10.1128/MCB.00810-06

Saygin, C., Godley, L.A., 2021. Genetics of Myelodysplastic Syndromes. Cancers (Basel). 13, 3380. https://doi.org/10.3390/CANCERS13143380

Schito, L., Semenza, G.L., 2016. Hypoxia-Inducible Factors: Master Regulators of Cancer Progression. Trends in Cancer 2, 758–770. https://doi.org/10.1016/j.trecan.2016.10.016

Schosserer, M., Grillari, J., Breitenbach, M., 2017. The Dual Role of Cellular Senescence in Developing Tumors and Their Response to Cancer Therapy. Front. Oncol. 7. https://doi.org/10.3389/fonc.2017.00278

Sekeres, M.A., 2010. The Epidemiology of Myelodysplastic Syndromes. Hematol. Clin. 24, 287–294. https://doi.org/10.1016/J.HOC.2010.02.011

Sekeres, M.A., Cutler, C., 2014. How we treat higher-risk myelodysplastic syndromes. Blood 123, 829–836. https://doi.org/10.1182/BLOOD-2013-08-496935

Shetty, V., Mundle, S., Alvi, S., Showel, M., Broady-Robinson, L.T., Dar, S., Borok, R., Showel, J., Gregory, S., Rifkin, S., Gezer, S., Parcharidou, A., Venugopal, P., Shah, R., Hernandez, B., Klein, M., Alston, D., Robin, E., Dominquez, C., Raza, A., 1996. Measurement of apoptosis, proliferation and three cytokines in 46 patients with myelodysplastic syndromes. Leuk. Res. 20, 891–900. https://doi.org/10.1016/S0145-2126(96)00008-2

Shiozawa, Y., Malcovati, L., Gallì, A., Pellagatti, A., Karimi, M., Sato-Otsubo, A., Sato, Y., Suzuki, H., Yoshizato, T., Yoshida, K., Shiraishi, Y., Chiba, K., Makishima, H., Boultwood, J., Hellström-Lindberg, E., Miyano, S., Cazzola, M., Ogawa, S., 2017. Gene expression and risk of leukemic transformation in myelodysplasia. Blood 130, 2642–2653. https://doi.org/10.1182/BLOOD-2017-05-783050

Silva, F.P.G., Morolli, B., Storlazzi, C.T., Anelli, L., Wessels, H., Bezrookove, V., Kluin-Nelemans, H.C., Giphart-Gassler, M., 2003. Identification of RUNX1/AML1 as a classical tumor suppressor gene. Oncogene 22, 538–547. https://doi.org/10.1038/sj.onc.1206141

Simon, F., Bockhorn, M., Praha, C., Baba, H.A., Broelsch, C.E., Frilling, A., Weber, F., 2010. Deregulation of HIF1-alpha and hypoxia-regulated pathways in hepatocellular carcinoma and corresponding non-malignant liver tissue-influence of a modulated host stroma on the prognosis of HCC. Langenbeck's Arch. Surg. 395, 395–405. https://doi.org/10.1007/s00423-009-0590-9

Skokowa, J., Cario, G., Uenalan, M., Schambach, A., Germeshausen, M., Battmer, K., Zeidler, C., Lehmann, U., Eder, M., Baum, C., Grosschedl, R., Stanulla, M., Scherr, M., Welte, K., 2006. LEF-1 is crucial for neutrophil granulocytopoiesis and its expression is severely reduced in congenital neutropenia. Nat. Med. 12, 1191–1197. https://doi.org/10.1038/nm1474

Sood, R., Kamikubo, Y., Liu, P., 2017. Role of RUNX1 in hematological malignancies. Blood 129, 2070–2082. https://doi.org/10.1182/blood-2016-10-687830

Starczynowski, D.T., Vercauteren, S., Telenius, A., Sung, S., Tohyama, K., Brooks-Wilson, A., Spinelli, J.J., Eaves, C.J., Eaves, A.C., Horsman, D.E., Lam, W.L., Karsan, A., 2008. High-resolution whole genome tiling path array CGH analysis of CD34+ cells from patients with low-risk myelodysplastic syndromes reveals cryptic copy number alterations and predicts overall and leukemia-free survival. Blood 112, 3412–3424. https://doi.org/10.1182/BLOOD-2007-11-122028

Stevens-Kroef, M.J., Olde Weghuis, D., ElIdrissi-Zaynoun, N., van der Reijden, B., Cremers, E.M.P., Alhan, C., Westers, T.M., Visser-Wisselaar, H.A., Chitu, D.A., Cunha, S.M., Vellenga, E., Klein, S.K., Wijermans, P., de Greef, G.E., Schaafsma, M.R., Muus, P., Ossenkoppele, G.J., van de Loosdrecht, A.A., Jansen, J.H., 2017. Genomic array as compared to karyotyping in myelodysplastic syndromes in a prospective clinical trial. Genes, Chromosom. Cancer 56, 524–534. https://doi.org/10.1002/GCC.22455

Stosch, J.M., Heumüller, A., Niemöller, C., Bleul, S., Rothenberg-Thurley, M., Riba, J., Renz, N., Szarc vel Szic, K., Pfeifer, D., Follo, M., Pahl, H.L., Zimmermann, S., Duyster, J., Wehrle, J., Lübbert, M., Metzeler, K.H., Claus, R., Becker, H., 2018. Gene mutations and clonal architecture in myelodysplastic syndromes and changes upon progression to acute myeloid leukaemia and under treatment. Br. J. Haematol. 182, 830–842. https://doi.org/10.1111/BJH.15461

Svobodova, K., Lhotska, H., Hodanova, L., Pavlistova, L., Vesela, D., Belickova, M., Vesela, J., Brezinova, J., Sarova, I., Izakova, S., Lizcova, L., Siskova, M., Jonasova, A., Cermak, J., Michalova, K., Zemanova, Z., 2020. Cryptic aberrations may allow more accurate prognostic classification of patients with myelodysplastic syndromes and clonal evolution. Genes, Chromosom. Cancer 59, 396–405. https://doi.org/10.1002/GCC.22841

Takacova, S., Slany, R., Bartkova, J., Stranecky, V., Dolezel, P., Luzna, P., Bartek, J., Divoky, V., 2012. DNA damage response and inflammatory signaling limit the MLL-ENL-induced leukemogenesis in vivo. Cancer Cell 21, 517–531. https://doi.org/10.1016/J.CCR.2012.01.021

Tay, L.S., Krishnan, V., Sankar, H., Chong, Y.L., Chuang, L.S.H., Tan, T.Z., Kolinjivadi, A.M., Kappei, D., Ito, Y., 2018. RUNX Poly(ADP-Ribosyl)ation and BLM Interaction Facilitate the Fanconi Anemia Pathway of DNA Repair. Cell Rep. 24, 1747–1755. https://doi.org/10.1016/j.celrep.2018.07.038

Thol, F., Kade, S., Schlarmann, C., Löffeld, P., Morgan, M., Krauter, J., Wlodarski, M.W., Kölking, B., Wichmann, M., Görlich, K., Göhring, G., Bug, G., Ottmann, O., Niemeyer, C.M., Hofmann, W.K., Schlegelberger, B., Ganser, A., Heuser, M., 2012. Frequency and

prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. Blood 119, 3578–3584. https://doi.org/10.1182/blood-2011-12-399337

Thorvaldsen, J.L., Duran, K.L., Bartolomei, M.S., 1998. Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. Genes Dev. 12, 3693–3702. https://doi.org/10.1101/GAD.12.23.3693

Toprak, S.K., 2022. Past, present and future in low-risk myelodysplastic syndrome. Front. Med. 9. https://doi.org/10.3389/fmed.2022.967900

Tsai, S.C., Shih, L.Y., Liang, S.T., Huang, Y.J., Kuo, M.C., Huang, C.F., Shih, Y.S., Lin, T.H., Chiu, M.C., Liang, D.C., 2015. Biological activities of RUNX1 mutants predict secondary acute leukemia transformation from chronic myelomonocytic leukemia and myelodysplastic syndromes. Clin. Cancer Res. 21, 3541–3551. https://doi.org/10.1158/1078-0432.CCR-14-2203

Valka, J., Vesela, J., Votavova, H., Dostalova-Merkerova, M., Horakova, Z., Campr, V., Brezinova, J., Zemanova, Z., Jonasova, A., Cermak, J., Belickova, M., 2017. Differential expression of homologous recombination DNA repair genes in the early and advanced stages of myelodysplastic syndrome. Eur. J. Haematol. 99, 323–331. https://doi.org/10.1111/ejh.12920

van Vliet, T., Varela-Eirin, M., Wang, B., Borghesan, M., Brandenburg, S.M., Franzin, R., Evangelou, K., Seelen, M., Gorgoulis, V., Demaria, M., 2021. Physiological hypoxia restrains the senescence-associated secretory phenotype via AMPK-mediated mTOR suppression. Mol. Cell 81, 2041–2052.e6. https://doi.org/10.1016/j.molcel.2021.03.018

Vasikova, A., Belickova, M., Budinska, E., Cermak, J., 2010. A distinct expression of various gene subsets in CD34+ cells from patients with early and advanced myelodysplastic syndrome. Leuk. Res. 34, 1566–1572. https://doi.org/10.1016/j.leukres.2010.02.021

Vennin, C., Spruyt, N., Dahmani, F., Julien, S., Bertucci, F., Finetti, P., Chassat, T., Bourette, R.P., Le Bourhis, X., Adriaenssens, E., 2015. H19 non coding RNA-derived miR-675 enhances tumorigenesis and metastasis of breast cancer cells by downregulating c-Cbl and Cbl-b. Oncotarget 6, 29209–29223. https://doi.org/10.18632/oncotarget.4976

Veryaskina, Y.A., Titov, S.E., Kovynev, I.B., Fedorova, S.S., Pospelova, T.I., Zhimulev, I.F., 2021. MicroRNAs in the Myelodysplastic Syndrome. Acta Naturae 13, 4–15. https://doi.org/10.32607/actanaturae.11209

Votavova, H., Grmanova, M., Dostalova Merkerova, M., Belickova, M., Vasikova, A., Neuwirtova, R., Cermak, J., 2011. Differential Expression of MicroRNAs in CD34+ Cells of 5q- Syndrome. J. Hematol. Oncol. 4, 1. https://doi.org/10.1186/1756-8722-4-1

Walter, M.J., Shen, D., Ding, L., Shao, J., Koboldt, D.C., Chen, K., Larson, D.E., McLellan, M.D., Dooling, D., Abbott, R., Fulton, R., Magrini, V., Schmidt, H., Kalicki-Veizer, J., O'Laughlin, M., Fan, X., Grillot, M., Witowski, S., Heath, S., Frater, J.L., Eades, W., Tomasson, M., Westervelt, P., DiPersio, J.F., Link, D.C., Mardis, E.R., Ley, T.J., Wilson, R.K., Graubert, T.A., 2012. Clonal Architecture of Secondary Acute Myeloid Leukemia. N. Engl. J. Med. 366, 1090–1098. https://doi.org/10.1056/nejmoa1106968

Walter, M.J., Shen, D., Shao, J., Ding, L., White, B., Kandoth, C., Miller, C.A., Niu, B., McLellan, M.D., Dees, N.D., Fulton, R., Elliot, K., Heath, S., Grillot, M., Westervelt, P., Link, D.C., DiPersio, J.F., Mardis, E., Ley, T.J., Wilson, R.K., Graubert, T.A., 2013. Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. Leukemia 27, 1275–1282. https://doi.org/10.1038/leu.2013.58

Wan, C., Wen, J., Huang, Y., Li, H., Wu, W., Xie, Q., Liang, X., Tang, Z., Zhao, W., Cheng, P., Liu, Z., 2020. Microarray analysis of differentially expressed microRNAs in myelodysplastic syndromes. Medicine (Baltimore). 99, e20904. https://doi.org/10.1097/MD.0000000000020904

Wang, B., Suen, C.W., Ma, H., Wang, Y., Kong, L., Qin, D., Lee, Y.W.W., Li, G., 2020. The

Roles of H19 in Regulating Inflammation and Aging. Front. Immunol. 11, 2769. https://doi.org/10.3389/fimmu.2020.579687

Wang, J., Zhao, L., Shang, K., Liu, F., Che, J., Li, H., Cao, B., 2020. Long non-coding RNA H19, a novel therapeutic target for pancreatic cancer. Mol. Med. 26, 30. https://doi.org/10.1186/s10020-020-00156-4

Wang, Y. yuan, Cen, J. nong, He, J., Shen, H. jie, Liu, D. dan, Li, Y., Qi, X. fei, Chen, Z. xing, 2009. Accelerated cellular senescence in myelodysplastic syndrome. Exp. Hematol. 37, 1310–1317. https://doi.org/10.1016/j.exphem.2009.09.002

Welford, S.M., Giaccia, A.J., 2011. Hypoxia and senescence: The impact of oxygenation on tumor suppression. Mol. Cancer Res. 9, 538–544. https://doi.org/10.1158/1541-7786.MCR-11-0065

Wen, J., Luo, Q., Wu, Y., Zhu, S., Miao, Z., 2020. Integrated Analysis of Long Non-Coding RNA and mRNA Expression Profile in Myelodysplastic Syndromes. Clin. Lab. 66, 825–834. https://doi.org/10.7754/CLIN.LAB.2019.190939

Wu, D., Ozaki, T., Yoshihara, Y., Kubo, N., Nakagawara, A., 2013. Runt-related transcription factor 1 (RUNX1) stimulates tumor suppressor p53 protein in response to DNA damage through complex formation and acetylation. J. Biol. Chem. 288, 1353–1364. https://doi.org/10.1074/jbc.M112.402594

Wu, S., Zheng, C., Chen, S., Cai, X., Shi, Y., Lin, B., Chen, Y., 2015. Overexpression of long non-coding RNA HOTAIR predicts a poor prognosis in patients with acute myeloid leukemia. Oncol. Lett. 10, 2410–2414. https://doi.org/10.3892/OL.2015.3552

Wu, S.J., Tang, J.L., Lin, C.T., Kuo, Y.Y., Li, L.Y., Tseng, M.H., Huang, C.F., Lai, Y.J., Lee, F.Y., Liu, M.C., Liu, C.W., Hou, H.A., Chen, C.Y., Chou, W.C., Yao, M., Huang, S.Y., Ko, B.S., Tsay, W., Tien, H.F., 2013. Clinical implications of U2AF1 mutation in patients with myelodysplastic syndrome and its stability during disease progression. Am. J. Hematol. 88. https://doi.org/10.1002/AJH.23541

Xu, Y., Li, Y., Xu, Q., Chen, Y., Lv, N., Jing, Y., Dou, L., Bo, J., Hou, G., Guo, J., Wang, X., Wang, L., Li, Y., Chen, C., Yu, L., 2017. Implications of mutational spectrum in myelodysplastic syndromes based on targeted next-generation sequencing. Oncotarget 8, 82475–82490. https://doi.org/10.18632/oncotarget.19628

Yao, C.Y., Chen, C.H., Huang, H.H., Hou, H.A., Lin, C.C., Tseng, M.H., Kao, C.J., Lu, T.P., Chou, W.C., Tien, H.F., 2017. A 4-lncRNA scoring system for prognostication of adult myelodysplastic syndromes. Blood Adv. 1, 1505–1516. https://doi.org/10.1182/bloodadvances.2017008284

Yao, X., Xue, Y., 2009. LETTER TO THE EDITOR DOG 1.0: illustrator of protein domain structures. Mol. Biol. Cell Mol. Cel-lular Biol. 19, 271–273. https://doi.org/10.1038/cr.2009.6

Yildirim, E., Kirby, J.E., Brown, D.E., Mercier, F.E., Sadreyev, R.I., Scadden, D.T., Lee, J.T., 2013. Xist RNA Is a Potent Suppressor of Hematologic Cancer in Mice. Cell 152, 727–742. https://doi.org/10.1016/J.CELL.2013.01.034

Zahid, M.F., Malik, U.A., Sohail, M., Hassan, I.N., Ali, S., Shaukat, M.H.S., 2017. Cytogenetic abnormalities in myelodysplastic syndromes: An overview. Int. J. Hematol. Stem Cell Res. 11, 232–240.

Zeidan, A.M., Pullarkat, V.A., Komrokji, R.S., 2017. Overcoming barriers to treating iron overload in patients with lower-risk myelodysplastic syndrome. Crit. Rev. Oncol. Hematol. 117, 57–66. https://doi.org/10.1016/j.critrevonc.2017.07.002

Zemanova, Z., Michalova, K., Buryova, H., Brezinova, J., Kostylkova, K., Bystricka, D., Novakova, M., Sarova, I., Izakova, S., Lizcova, L., Ransdorfova, S., Krejcik, Z., Merkerova, M.D., Dohnalova, A., Siskova, M., Jonasova, A., Neuwirtova, R., Cermak, J., 2014. Involvement of deleted chromosome 5 in complex chromosomal aberrations in newly diagnosed myelodysplastic syndromes (MDS) is correlated with extremely adverse

prognosis. Leuk. Res. 38, 537–544. https://doi.org/10.1016/j.leukres.2014.01.012

Zglinicki, T. Von, Saretzki, G., Ladhoff, J., Fagagna, F.D.A. Di, Jackson, S.P., 2005. Human cell senescence as a DNA damage response. Mech. Ageing Dev. 126, 111–117. https://doi.org/10.1016/j.mad.2004.09.034

Zhang, T., Zhou, J., Zhang, W., Lin, J., Ma, J., Wen, X., Yuan, Q., Li, X., Xu, Z., Qian, J., 2018. H19 overexpression promotes leukemogenesis and predicts unfavorable prognosis in acute myeloid leukemia. Clin. Epigenetics 10, 47. https://doi.org/10.1186/s13148-018-0486-z

Zhang, X., Weissman, S.M., Newburger, P.E., 2014. Long intergenic non-coding RNA HOTAIRM1 regulates cell cycle progression during myeloid maturation in NB4 human promyelocytic leukemia cells. RNA Biol. 11, 777–787. https://doi.org/10.4161/rna.28828

Zhang, Y., Fan, L.J., Zhang, Y., Jiang, J., Qi, X.W., 2020. Long Non-coding Wilms Tumor 1 Antisense RNA in the Development and Progression of Malignant Tumors. Front. Oncol. 10, 35. https://doi.org/10.3389/fonc.2020.00035

Zhang, Y., Li, Z., Chen, M., Chen, H., Zhong, Q., Liang, L., Li, B., 2020. lncRNA TCL6 correlates with immune cell infiltration and indicates worse survival in breast cancer. Breast Cancer 27, 573–585. https://doi.org/10.1007/s12282-020-01048-5

Zhao, X., Yin, H., Li, N., Zhu, Y., Shen, W., Qian, S., He, G., Li, J., Wang, X., 2019. An Integrated Regulatory Network Based on Comprehensive Analysis of mRNA Expression, Gene Methylation and Expression of Long Non-coding RNAs (lncRNAs) in Myelodysplastic Syndromes. Front. Oncol. 9. https://doi.org/10.3389/fonc.2019.00200

Zhou, J. dong, Zhang, T. juan, Xu, Z. jun, Deng, Z. qun, Gu, Y., Ma, J. chun, Wen, X. mei, Leng, J. yan, Lin, J., Chen, S. ning, Qian, J., 2020. Genome-wide methylation sequencing identifies progression-related epigenetic drivers in myelodysplastic syndromes. Cell Death Dis. 11, 1–15. https://doi.org/10.1038/s41419-020-03213-2

Zhou, T., Hasty, P., Walter, C.A., Bishop, A.J.R., Scott, L.M., Rebel, V.I., 2013. Myelodysplastic syndrome: An inability to appropriately respond to damaged DNA? Exp. Hematol. 41, 665–674. https://doi.org/10.1016/j.exphem.2013.04.008

APPENDICES

I.   Permission to reprint publications

Publication I: LncRNA Profiling Reveals That the Deregulation of H19, WT1-AS, TCL6, and LEF1-AS1 Is Associated with Higher-Risk Myelodysplastic Syndrome

Author: Katarina Szikszai et al

Cancers (Basel), 2020

© 2020 by the authors

Publication II: RUNX1 mutations contribute to the progression of MDS due to disruption of antitumor cellular defense: a study on patients with lower-risk MDS

Author: Monika Kaisrlikova et al

Leukemia, 2022

© 2022 by the authors

II.   Supplementary Materials of publications

Publication I: LncRNA Profiling Reveals That the Deregulation of H19, WT1-AS, TCL6, and LEF1-AS1 Is Associated with Higher-Risk Myelodysplastic Syndrome

SUPPLEMENTARY MATERIAL

LncRNA profiling revealed that deregulation of H19, WT1-AS, TCL6, and LEF1-AS1 are predictive of adverse outcome in myelodysplastic syndromes

Katarina Szikszai, Zdenek Krejcik, Jiri Klema, Nikoleta Loudova, Andrea Mrhalkova, Monika Belickova, Monika Hruba, Jitka Vesela, Viktor Stranecky, David Kundrat, Pavla Pecherkova, Jaroslav Cermak, Anna Jonasova, and Michaela Dostalova Merkerova

SUPPLEMENTARY METHODS

Patients

The study included 133 patients with various subtypes of MDS, 28 patients with AML-MRC, and 22 healthy donors. The individuals were randomly divided into a discovery cohort (54 MDS, 14 AML-MRC, and 9 healthy controls) and a testing cohort (79 MDS, 14 AML-MRC, and 13 healthy controls). The bone marrow (BM) samples were obtained from the patients during routine clinical assessment at the Institute of Hematology and Blood Transfusion and the First Department of Internal Medicine, General Faculty Hospital, Prague. MDS patient age ranged from 31 to 82 years (average 63) and male/female distribution was 70/63. Similarly, age of AML-MRC patients ranged from 29 to 82 years (average 66) and male/female distribution was 19/9. The study included only the patients with no known history of previous malignancy, chemotherapy or radiation therapy. Moreover, none of the patients had received therapy for their disease or HSC transplantation (HSCT) prior to BM collection. The patient's diagnoses were assessed based on the standard WHO 2016 classification criteria [1] and all the patients were classified according to the IPSS-R categories [2] at the time of sample collection, except of two patients with unavailable cytogenetics. Control groups contained hematological healthy, age matched donors (age ranged from 28 to 70 years, average 56, and male/female distribution was 14/8). Informed consent was obtained from all the patients and the healthy donors for being included in the study. The study was approved by the Institutional Scientific Board and the Local Ethics Committee in accordance with the ethical standards of the Declaration of Helsinki and its later amendments. The detailed clinical and laboratory characteristics of both cohorts, including classification of MDS patients into subgroups, IPSS-R categories, BM features and blood counts, are summarized in SI 1.

Cell separation and nucleic acid extraction

Mononuclear cells (MNCs) and granulocytes were purified from BM aspirates using Ficoll-Histopaque (GE Healthcare, Munich, Germany) density centrifugation. CD34+ cells were subsequently isolated from MNCs using magnetic cell separation according to the manufacturer's instructions (Miltenyi Biotec, Bergisch Gladbach, Germany). DNA was isolated using MagCore Genomic DNA Whole Blood Kit. Total RNA was extracted by the acid-guanidine-phenol-chloroform method and the samples were incubated with DNase I

to prevent genomic DNA contamination. Quantity of DNA/RNA was quantified using Invitrogen Qubit 3 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and the RNA integrity was assessed using the Agilent 4200 TapeStation (Agilent Technologies, Santa Clara, CA, USA).

lncRNA microarrays and data analysis

Genome-wide lncRNA profiles were determined using Agilent Human GENCODE Custom lncRNA Expression Microarray Design v15 developed by the Bioinformatics and Genomics Group at the Centre for Genomic Regulation in Spain [3]. The array contains probes for 22,001 lncRNA transcripts and 17,535 PCG mRNAs. Agilent Low Input Quick Amp Labeling Kit was used for sample preparation (RNA input was set up to 200 ng) according to the manufacturer's recommendation. The hybridized arrays were scanned using Agilent DNA microarray scanner. Microarray probes were initially mapped to GRCh37/hg19 genome using NovoAlign program (Novocraft Technologies, Malaysia) and re-annotated according to UCSC Genome Browser (http://genome.ucsc.edu). Raw data were extracted using the Agilent Feature Extraction Software. Quality control, quantile normalization, and filtering were performed with the Bioconductor project in the R statistical environment using limma package. Differentially expressed lncRNAs and PCGs were identified using empirical Bayesian method implemented in R limma package. Multiple testing correction was performed to compute false discovery rate (FDR) using the Benjamini-Hochberg method. To visualize the differential expression data, expression heatmaps were designed using MeV v4.3.2 software [4] and the hierarchical clustering of the data was done using average linkage and Pearson distance. The raw and normalized data have been deposited in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database under accession number GSE145733.

RT-qPCR

Reverse transcription quantitative PCR (RT-qPCR) was applied to measure transcript levels of individual genes (lncRNAs: CHRM3A82, EPB41L4A-AS1, H19, LEF1-AS1, PVT1, TCL6, and WT1-AS; PCGs: IGF2, LEF1, WT1, TCL1A, and TCL1B; miRNAs: miR-675 and RNU48 as a reference). SuperScript IV VILO Master Mix (Thermo Fisher Scientific, Waltham, MA, USA) was used for cDNA synthesis and TaqMan gene expression assays with TaqMan universal mastermix II with UNG (Thermo Fisher Scientific) were applied for quantitative PCR using StepOnePlus instrument (Thermo Fisher Scientific).

For normalization of raw C_T data of lncRNAs and PCGs, we tested several known reference genes (B2M, GAPDH, GUSB, HPRT1, TUBB, UBC, and YWHAZ). The stability of the genes was compared using web-based tool RefFinder that integrates four major currently available computational programs (geNorm, Normfinder, BestKeeper, and the

comparative delta-Ct method) [5]. Based on the results from this optimization procedure (SI 2), the RT-qPCR data were finally normalized to HPRT1 reference gene and further processed by the $2^{-\Delta\Delta CT}$ method [6].

Mutational screening and data analysis
TruSight Myeloid Sequencing Panel Kit (Illumina, San Diego, CA, USA) containing 568 amplicons in 54 genes associated with myeloid malignancies was used for mutational screening of patients from the discovery cohort. The amplicon library was constructed according to the manufacturer's recommendations. After library purification, and subsequent normalization on beads, quantification was performed by Kapa Library Quantification Kit Illumina Platforms (Kapa Biosystems, Wilmington, Massachusetts, USA). The libraries were pooled and 2x150 bp paired-end sequenced with Rapid SBS Kit V2 chemistry on a HiSeq 2500 instrument. FASTQ files were subjected to initial quality control by FastQC. Adaptor trimming was done by Trimmomatic and low-quality sequences were removed by Illuminaclip. The remaining reads were aligned to the human genome hg19 using BWA-MEM. Variants were detected by LoFreq v2.1.3.1. and annotated using Variant Effect Predictor (Ensembl). Clinical significance of each variant was verified in several genomic databases (UCSC, COSMIC, ExAC, PubMed). The arbitrary cut off was set at 5 % of variant allele frequency (VAF).

Statistical analysis

Statistical analyses were performed using GraphPad Prism 7 (GraphPad Software, La Jolla, CA, USA) and SPSS software (IBM, Armonk, NY, USA). Nonparametric Mann-Whitney test was used to compare transcript levels and clinical parameters between two groups of samples. Spearman rank test was performed to assess the correlation of continuous variables. The survival distributions for overall survival (OS) and progression-free survival (PFS) were estimated using the Kaplan-Meier method, and the differences were compared using the log-rank test. For determination of the optimum cut-off values of transcript levels, we computed the p-values with the log-rank test on a dense net local computation and defined the cut-off points using Gaussian mixture models where the obtained p-values were divided into two components. For multivariate analysis, we estimated a Cox proportional hazards regression model with Min-Max method for normalizing of the data. The backward likelihood method was applied for reduction of variables. The differences were considered statistically significant if $p < 0.05$.

Pathway analysis
Changes in gene expression were related to functional changes using gene set enrichment analysis (GSEA) [7]. As a reference, c2 (c2.all.v7.0.symbols.gmt [Curated]), c5 (c5.all.v7.0.symbols.gmt [Gene Ontology, GO]), and hallmark (h.all.v7.0.symbols.gmt [Hallmarks]) gene sets from the Molecular Signatures Database were utilized. The number

of permutations was set up at 1,000. The enrichment results with p < 0.05 were considered statistically significant.

LncRNA-PCG coexpression networks

The network analysis directly stems from the network-based lncRNA module function annotation method introduced in [8]. Firstly, we identified differentially expressed lncRNAs and PCGs (FDR<0.05) and constructed a correlation matrix for these transcripts. The correlation was calculated for all the lncRNA-PCG pairs and the absolute value of Pearson correlation coefficient represented each pair. Second, a non-negative matrix factorization (NMF) was used to extract modules from the correlation matrix. In particular, the standard factorization based on the Kullback-Leibler divergence was employed [9]. The factorization was run multiple times for different numbers of modules and with different random seeds for computation of initial values for the factor matrices to avoid improper local minima of the objective function. The Frobenius norm of the factorization residual matrix served as the factorization objective function. Then, each module was functionally annotated. All the PCGs were mapped to the corresponding GO terms and the terms with at least two corresponding PCGs were kept. GO enrichment analysis served to annotate individual modules and Fisher exact test was used to calculate the score for the individual terms. Eventually, the representative cores of the individual modules were plotted. In each plot, 4 lncRNAs and 13 PCGs with the highest module membership visually represent the module. Edges connect those nodes whose absolute correlation exceeds the median module correlation (weak) and its third quartile (strong). The network analysis was carried out in the R statistical environment with the packages limma, NMF, GSEABase and iGraph.

SUPPLEMENTARY FIGURES AND TABLES

SI 1. Characteristics of the cohorts. Discovery cohort was examined by microarrays and testing cohort was used for RT-qPCR measurements.

Variable	Discovery cohort	Testing cohort
Number of samples (healthy controls/MDS/AML-MRC)	77 (9/54/14)	106 (13/79/14)
Healthy controls	**9**	**13**
Gender (male/female)	6/3	8/5
Age; mean (range)	61 (45-72)	52 (28-70)
MDS	**54**	**79**
Diagnosis (SLD/MLD/RS-SLD/RS-MLD/5q-/EB1/EB2)	5/13/4/3/7/10/12	8/12/8/8/12/10/21
Gender (male/female)	29/25	41/38
Age; mean (range)	65 (31-82)	62 (29-88)
IPSS-R category (very low/low/intermediate/high/very high/n.a.)	7/19/9/10/9/0	12/25/21/11/8/2
IPSS-R karyotype (very good/good/intermediate/poor/very poor/n.a.)	2/36/6/2/8/0	0/65/4/1/7/2
Cytogenetic features		
normal karyotype	16	38
isolated del(5q)	12	10
isolated del(20q)	4	4
isolated +8	2	4
complex	8	7
other	13	14
n.a.	0	2
Somatic mutations		
no. of patients with detected mutations (%)	39 (76 %)	n.a.
no. of mutations/patient:		
0/1/2/3/4/5/6/7/na	12/17/7/3/8/3/0/1/3	
Marrow blasts [%]: mean (range)	5.6 (0.0 – 19.0)	6.1 (0.2-19.4)

Hemoglobin (g/L): mean (range)	99 (68-159)	98 (67- 139)
Neutrophils (x10^9/L): mean (range)	2.3 (0.1-11.4)	2.4 (0.1-8.0)
Platelets (x10^9/L): mean (range)	184 (26-597)	205 (19-766)
Follow-up, number of patients	53	79
mean follow-up [months] (range)	26 (1-115)	34 (1-118)
i. HSCT (censored), number of patients	2	13
mean time to HSCT [months] (range)	16 (5-27)	18 (1-59)
ii. progression, number of patients	33	50
mean time to progression [months] (range)	23 (1-55)	25 (1-90)
iii deceased, number of patients	28	48
mean time to death [months] (range)	21 (1-78)	30 (1-96)
iv. alive (censored), number of patients	23	18
mean follow-up time [months] (range)	34 (1-115)	60 (1-118)
AML-MRC	**14**	**14**
Gender (male/female)	12/2	7/7
Age; mean (range)	69 (58-77)	63(29-82)
Cytogenetic features		
normal karyotype	4	6
isolated del(5q)	1	0
isolated +8	3	1
complex	3	3
other	3	5
Somatic mutations		
no. of patients with detected mutations (%)	11 (85 %)	n.a.
no. of mutations/patient: 0/1/2/3/4/na	2/4/4/1/2/1	
Marrow blasts [%]: mean (range)	26.5 (20.0-33.0)	35.0 (20.0-77.0)
Hemoglobin (g/L): mean (range)	97 (78-114)	97 (70-132)
Neutrophils (x10^9/L): mean (range)	1.7 (0.06-11.8)	2.1 (0.1-15.7)
Platelets (x10^9/L): mean (range)	93 (13-578)	59 (5-196)

n.a. - not analyzed

SI 2. Stability of selected reference genes potentially applicable for RT-qPCR normalization. The tested genes (B2M, GAPDH, GUSB, HPRT1, TUBB, UBC, and YWHAZ) were ranked using web-based tool RefFinder that integrates four major currently available computational programs (geNorm, Normfinder, BestKeeper, and the comparative delta-Ct method). Based on the rankings from each program, RefFinder assigns an appropriate weight to an individual gene and calculates the geometric mean of their weights for the overall final ranking [5].

Comprehensive gene stability

<== Most stable genes   Least stable genes ==>
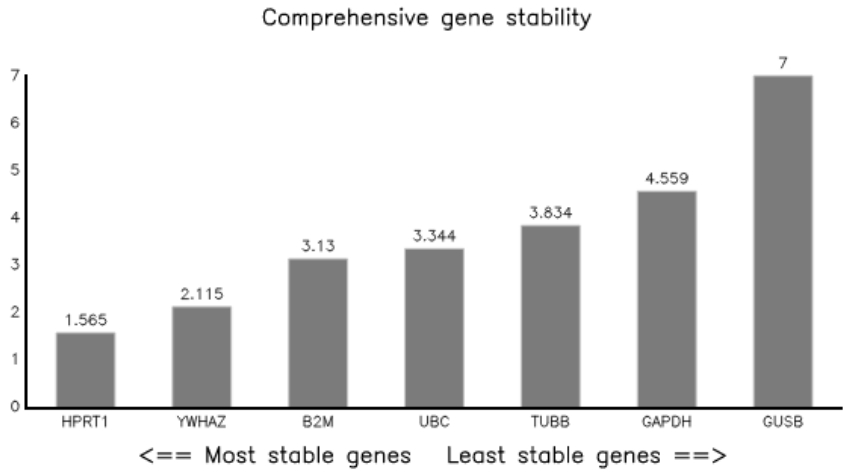
Method	1	2	3	4	5	6	7
Delta CT	HPRT1	YWHAZ	TUBB	B2M	UBC	GAPDH	GUSB
BestKeeper	UBC	B2M	GAPDH	TUBB	YWHAZ	HPRT1	GUSB
Normfinder	HPRT1	YWHAZ	TUBB	B2M	UBC	GAPDH	GUSB
Genorm	HPRT1 | YWHAZ		B2M	GAPDH	UBC	TUBB	GUSB
Recommended ranking	HPRT1	YWHAZ	B2M	UBC	TUBB	GAPDH	GUSB

Ranking Order

SI 3. List of significantly deregulated transcripts in MDS patients compared to healthy controls (|logFC| > 1, FDR < 0.05). Of the 83 upregulated PCGs, only the top 30 transcripts are listed. logFC – binary logarithm of fold change, FDR - false discovery rate.

No.	Transcript	Chromosome	logFC	FDR
lncRNAs increased in MDS				
1	PRKAR2A-AS1	chr3	3.78	2.68E-03
2	RP11-408E5.5	chr13	3.22	3.29E-04
3	H19	chr11	3.18	2.08E-03
4	RP5-867C24.4	chr17	2.75	2.75E-02
5	EMCN-IT1	chr4	2.58	1.41E-05
6	RP11-558A11.3	chr16	2.30	5.01E-06
7	LINC00570	chr2	2.16	1.41E-02
8	WT1-AS	chr11	2.13	4.51E-02
9	RP11-677I18.3	chr11	1.94	3.57E-04
10	RP11-567J20.3	chr8	1.82	3.08E-02
11	RP11-753D20.1	chr14	1.77	1.15E-02
12	LINC00640	chr14	1.66	4.90E-02
13	FAM225A	chr9	1.61	1.51E-02
14	AC131097.3	chr2	1.51	1.90E-02
15	RP4-669L17.2	chr1	1.48	2.59E-03
16	MEG8	chr14	1.36	1.70E-02
17	AL132709.8	chr14	1.28	3.26E-02
18	CTD-2373N4.5	chr8	1.23	1.33E-02
19	RP11-792D21.2	chr4	1.22	3.29E-04
20	RP5-1029F21.4	chr17	1.19	3.55E-02
21	AC017076.5	chr2	1.12	3.25E-02
22	PVT1	chr8	1.12	3.67E-02

23	CTD-2319I12.2	chr17	1.11	5.66E-04
24	RP11-277L2.4	chr1	1.09	4.39E-02
25	AL132709.5	chr14	1.08	2.15E-02
26	AC020571.3	chr2	1.08	2.95E-02
27	RP1-249H1.4	chr6	1.08	2.12E-02
28	LINC00484	chr9	1.04	1.58E-02
lncRNAs reduced in MDS				
1	U3	chr18	-1.40	4.96E-04
2	AC079779.4	chr2	-1.18	2.55E-02
3	ST6GAL2-IT1	chr2	-1.03	3.98E-02
4	RP11-13K12.1	chr17	-1.01	4.12E-02
PCGs increased in MDS				
1	HBG1	chr11	5.45	1.45E-05
2	HBBP1	chr11	4.00	4.70E-05
3	GYPB	chr4	3.93	4.70E-05
4	PGF	chr14	3.41	2.08E-05
5	IFI27	chr14	3.36	1.44E-03
6	OAS1	chr12	3.15	2.24E-02
7	NCAM1	chr11	3.15	9.79E-03
8	SH2D1A	chrX	2.98	8.45E-03
9	GYPB	chr4	2.93	2.95E-07
10	HBA2	chr16	2.91	2.83E-02
11	TMCC2	chr1	2.80	1.79E-03
12	TRIM10	chr6	2.78	1.51E-04
13	SPAG6	chr10	2.73	2.90E-02
14	ALDH1A3	chr15	2.64	2.24E-02

15	EPB42	chr15	2.53	4.30E-02
16	SRMS	chr20	2.43	1.33E-02
17	C20orf108	chr20	2.39	1.25E-03
18	SLC6A9	chr1	2.39	4.75E-03
19	BAI1	chr8	2.36	1.77E-04
20	PABPC4L	chr4	2.32	1.53E-02
21	LOC285758	chr6	2.28	4.70E-05
22	FHDC1	chr4	2.28	2.17E-02
23	ARG2	chr14	2.24	1.45E-02
24	SLC6A8	chr16	2.08	2.51E-02
25	OSBP2	chr22	1.89	1.89E-03
26	RFPL4A	chr19	1.88	2.32E-02
27	ABCC13	chr21	1.80	2.32E-02
28	IL2RA	chr10	1.80	2.70E-03
29	ENST00000515150	chr4	1.78	4.67E-04
30	LOC284561	chr1	1.78	3.17E-02
PCGs reduced in MDS				
1	ECEL1P2	chr2	-1.81	1.26E-03
2	A_33_P3258324	chr19	-1.34	4.58E-02
3	AVP	chr20	-1.21	1.17E-02
4	HLF	chr17	-1.08	1.29E-03

SI 4. List of significantly deregulated transcripts in AML-MRC compared to MDS patients (|logFC| > 1, FDR < 0.05). Of the 159 downregulated PCGs, only the top 30 transcripts are listed. logFC – binary logarithm of fold change, FDR - false discovery rate, ns – none significant transcript.

No.	Transcript	Chromosome	logFC	FDR
lncRNAs increased in AML-MRC				
ns	ns	ns	ns	ns
lncRNAs reduced in AML-MRC				
1	AC004510.3	chr19	-3.28	1.65E-05
2	RP11-489D6.2	chr15	-2.45	4.60E-03
3	VPS9D1-AS1	chr16	-2.15	8.63E-04
4	RP11-96B2.1	chr8	-2.01	8.63E-04
5	RP11-327I22.8	chr9	-1.98	3.80E-02
6	PVT1	chr8	-1.86	1.78E-02
7	RP11-48O20.4	chr1	-1.52	2.74E-02
8	RP11-315A17.1	chr4	-1.39	1.22E-02
9	CTD-2319I12.2	chr17	-1.34	1.22E-02
10	CXADRP3	chr18	-1.25	7.21E-05
11	C1QTNF9B-AS1	chr13	-1.07	2.36E-02
PCGs increased in AML-MRC				
1	LATS2	chr13	1.07	1.57E-02
2	GUCY1A3	chr4	1.04	3.41E-02
PCGs rduced in AML-MRC				
1	DEFA3	chr8	-3.72	1.34E-02
2	PRG3	chr11	-3.64	7.85E-03
3	C21orf67	chr21	-3.44	3.15E-02
4	RBP7	chr1	-3.19	1.12E-03

5	IGHV1-18	chr14	-3.15	1.88E-02
6	IGKV4-1	chr2	-3.15	3.90E-02
7	IGLV3-10	chr22	-3.06	3.58E-02
8	SLC10A4	chr4	-2.87	1.66E-03
9	STAB2	chr12	-2.85	7.07E-03
10	ST6GALNAC1	chr17	-2.83	1.92E-03
11	DUSP26	chr8	-2.72	3.49E-02
12	APOC1	chr19	-2.67	1.00E-02
13	NMU	chr4	-2.62	1.57E-02
14	SELENBP1	chr1	-2.62	2.49E-02
15	EPB42	chr15	-2.55	5.00E-02
16	IGHV1-2	chr14	-2.53	2.75E-02
17	CLEC4G	chr19	-2.51	9.99E-04
18	SEC14L4	chr22	-2.50	7.85E-03
19	C10orf116	chr10	-2.49	6.34E-03
20	EPX	chr17	-2.47	1.28E-02
21	SPAG6	chr10	-2.45	3.87E-02
22	GYPB	chr4	-2.42	2.75E-02
23	ANK1	chr8	-2.40	1.34E-02
24	COL6A5	chr3	-2.38	7.07E-03
25	CYP1B1	chr2	-2.37	1.57E-02
26	GTSF1	chr12	-2.37	2.96E-02
27	AKR1C1	chr10	-2.37	8.07E-03
28	SNX22	chr15	-2.37	5.26E-04
29	KLF1	chr19	-2.35	7.28E-03
30	CHI3L1	chr1	-2.33	3.55E-02

SI 5. Higher resolution image of Figure 1C. Selected modules of the coexpression network designed based on differentially expressed genes between MDS patients and healthy controls. Gene ontology (GO) terms significantly (p < 0.01) associated with these modules are listed in the corresponding tables. Square – PCG, circle – lncRNA, red – upregulation in MDS, blue – downregulation in MDS.

Hemoglobin-related modules
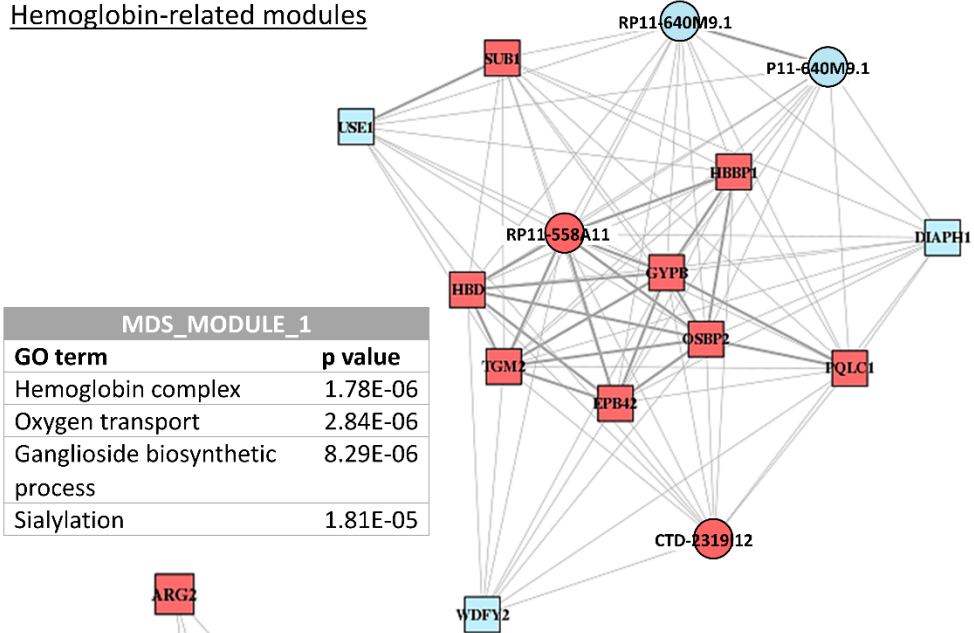
MDS_MODULE_1	
GO term	**p value**
Hemoglobin complex	1.78E-06
Oxygen transport	2.84E-06
Ganglioside biosynthetic process	8.29E-06
Sialylation	1.81E-05
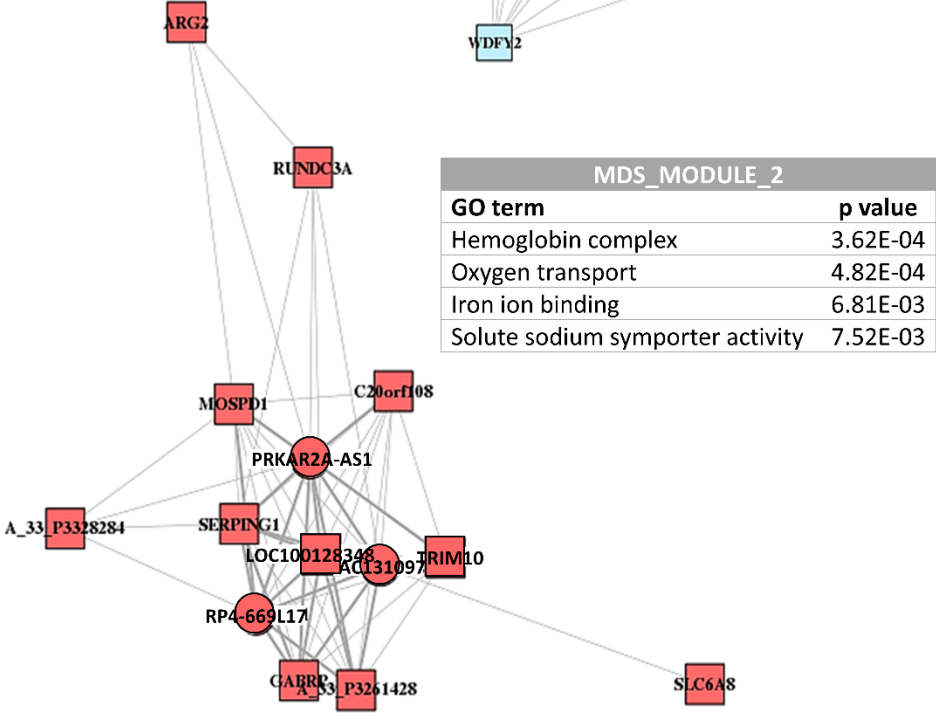
MDS_MODULE_2	
GO term	**p value**
Hemoglobin complex	3.62E-04
Oxygen transport	4.82E-04
Iron ion binding	6.81E-03
Solute sodium symporter activity	7.52E-03

Protein metabolism-related modules

MDS_MODULE_3	
GO term	P value
Golgi apparatus	4.00E-03
Protein polyubiquitination	4.87E-03
Proteasomal protein catabolic process	5.28E-03
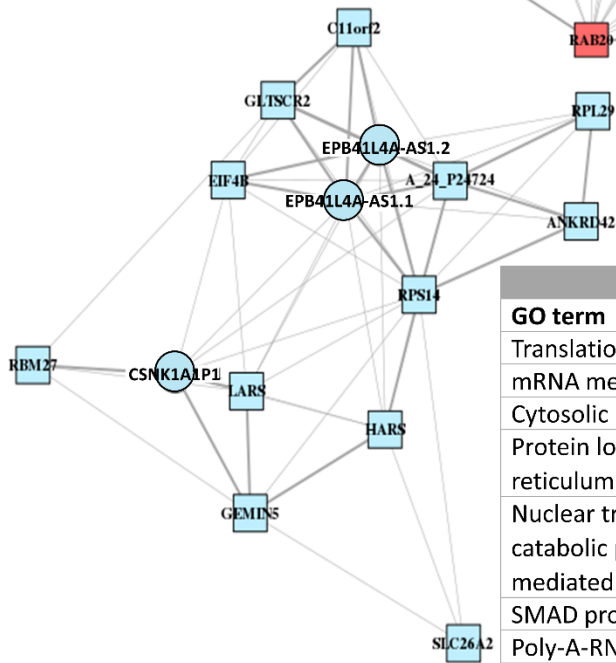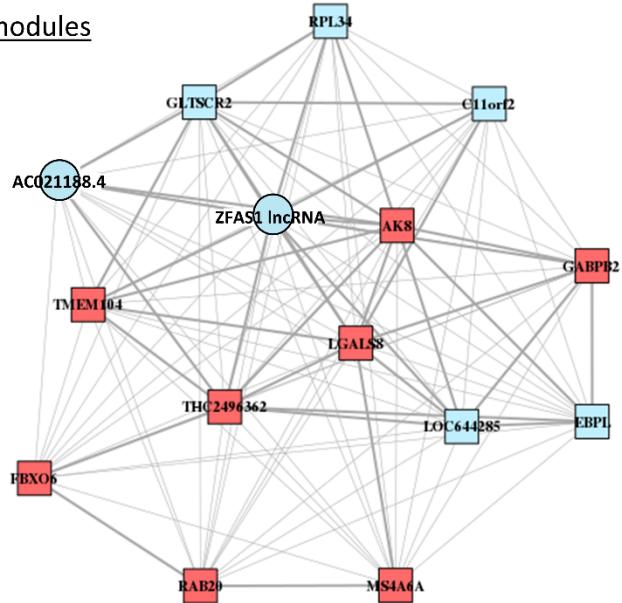
MDS_MODULE_4	
GO term	P value
Translational initiation	9.54E-04
mRNA metabolic process	1.30E-03
Cytosolic ribosome	4.75E-03
Protein localization to endoplasmic reticulum	5.13E-03
Nuclear transcribed mRNA catabolic process nonsense mediated decay	5.13E-03
SMAD protein signal transduction	5.24E-03
Poly-A-RNA binding	5.28E-03
Cytoplasmic translation	5.70E-03
Amino-acid activation	6.67E-03
Ribonucleoprotein complex biogenesis	6.80E-03

SI 6. List of significantly deregulated transcripts in MDS/AML-MRC patients with isolated del(5q) vs. those with normal karyotype (|logFC| > 1, FDR < 0.05). Of the 106 increased and 54 reduced PCGs, only the top 30 transcripts are listed in each category. logFC – binary logarithm of fold change, FDR - false discovery rate.

No.	Transcript	Chromosome	logFC	FDR
lncRNAs increased in the patients with isolated del(5q)				
1	EMCN-IT1	chr4	3.01	4.60E-03
2	RP11-56F10.3	chr9	2.27	3.11E-03
3	AC026806.2	chr19	2.26	2.91E-02
4	RP11-185E8.1	chr3	2.07	3.58E-02
5	RP11-264B17.5	chr16	2.06	3.23E-02
6	CTB-114C7.3	chr5	1.93	1.49E-02
7	CHRM3-AS2	chr1	1.88	6.58E-04
8	MAST4-IT1	chr5	1.78	2.66E-02
9	RP4-773A18.4	chr1	1.70	2.49E-02
10	RP3-510D11.2	chr1	1.59	1.94E-02
11	RP11-496N12.6	chr1	1.56	3.16E-02
12	RP11-797H7.5	chr7	1.38	4.88E-02
13	RP11-83N9.5	chr9	1.31	6.56E-03
14	PVT1	chr8	1.22	3.52E-02
15	CCDC26	chr8	1.10	4.31E-02
16	LINC00534	chr8	1.04	3.80E-02
lncRNAs reduced in the patients with isolated del(5q)				
1	TTN-AS1	chr2	-3.11	5.06E-03
2	RP11-171I2.4	chr2	-1.42	9.60E-03
3	RP11-861E21.1	chr18	-1.30	3.47E-02
4	RP11-434C1.1	chr12	-1.22	3.78E-02
5	ZFAS1	chr20	-1.18	9.61E-03

6	STARD4-AS1	chr5	-1.15	2.81E-02
7	RP1-69M21.2	chr1	-1.14	5.06E-03
8	EPB41L4A-AS1	chr5	-1.10	8.31E-06
9	CTC-345K18.2	chr5	-1.09	6.58E-04
10	AC116366.5	chr5	-1.05	9.61E-03
11	GAS5	chr1	-1.04	3.10E-02
12	CTC-304I17.3	chr17	-1.03	2.71E-02
13	PCBP1-AS1	chr2	-1.02	3.16E-02
14	RP11-493K19.3	chr3	-1.02	4.02E-02
15	RP11-169D4.2	chr11	-1.02	2.85E-02
PCGs increased in the patients with isolated del(5q)				
1	HBBP1	chr11	5.07	1.59E-02
2	CNN1	chr19	3.09	1.81E-02
3	SLC35D3	chr6	2.95	1.67E-02
4	SPOCD1	chr1	2.79	1.27E-02
5	TMEM158	chr3	2.79	2.49E-02
6	ENST00000515150	chr4	2.70	3.57E-03
7	LAT	chr16	2.57	1.05E-02
8	PLIN2	chr9	2.48	5.53E-03
9	CLEC1B	chr12	2.40	4.16E-02
10	CD40LG	chrX	2.39	3.40E-03
11	SPAG6	chr10	2.39	1.78E-02
12	GNAZ	chr22	2.37	1.24E-02
13	THBS1	chr15	2.36	4.41E-02
14	ST6GALNAC1	chr17	2.26	2.22E-03
15	LGALS12	chr11	2.25	3.81E-03

16	LY6G6F	chr6	2.23	2.41E-02
17	CTTN	chr11	2.22	2.91E-02
18	LRRC32	chr11	2.22	4.15E-02
19	NRGN	chr11	2.14	4.72E-02
20	TUBAL3	chr10	2.13	6.38E-03
21	VSTM1	chr19	2.11	3.57E-03
22	CATSPER1	chr11	2.09	4.71E-02
23	DENND2C	chr1	2.08	3.42E-02
24	COL6A5	chr3	2.04	1.12E-02
25	TMEM40	chr3	2.00	3.68E-02
26	ACE2	chrX	2.00	2.91E-02
27	SLC2A14	chr12	1.99	1.15E-02
28	TUBA4A	chr2	1.95	2.83E-02
29	RAB6B	chr3	1.91	4.22E-03
30	PPAPDC1A	chr10	1.89	1.20E-02
PCGs reduced in the patients with isolated del(5q)				
1	ANK3	chr10	-2.78	3.69E-02
2	APBB2	chr4	-2.16	3.83E-02
3	THC2753069	chr17	-2.08	4.54E-02
4	EGR1	chr5	-1.90	5.16E-03
5	USP9Y	chrY	-1.85	4.89E-02
6	ENST00000507296	chr8	-1.79	2.29E-02
7	NCRNA00185	chrY	-1.72	2.56E-02
8	ELFN1	chr7	-1.71	4.46E-02
9	ETV7	chr6	-1.68	1.33E-02
10	C17orf51	chr17	-1.62	3.92E-02

11	PLEKHG5	chr1	-1.61	2.40E-02
12	SLC23A1	chr5	-1.42	4.16E-02
13	C5orf56	chr5	-1.41	1.51E-03
14	EN2	chr7	-1.41	4.12E-02
15	AK125205	chr2	-1.39	2.91E-02
16	MZB1	chr5	-1.37	4.60E-02
17	GIMAP2	chr7	-1.34	1.42E-03
18	IL15	chr4	-1.30	2.24E-02
19	A_33_P3261024	chr6	-1.29	2.94E-02
20	C10orf10	chr10	-1.29	3.75E-02
21	IL28A	chr19	-1.29	5.31E-03
22	C1orf54	chr1	-1.28	3.93E-02
23	CD74	chr5	-1.28	1.41E-04
24	GLTSCR2	chr19	-1.24	5.04E-03
25	ANXA6	chr5	-1.22	1.56E-03
26	KLHL3	chr5	-1.22	1.57E-02
27	LOC100240735	chr12	-1.22	4.85E-02
28	TLR3	chr4	-1.19	8.71E-03
29	GLI4	chr8	-1.17	3.32E-02
30	DNHD1	chr11	-1.17	3.14E-02

SI 7. Expression levels of PVT1, CHRM3-AS2, and EPB41LA-AS1 lncRNAs in MDS/AML-MRC patients with relation to their karyotype. Relative expression was assessed by RT-qPCR. CTR – healthy controls, **p < 0.01, ***p < 0.001, ****p < 0.0001, ns – non-significant.
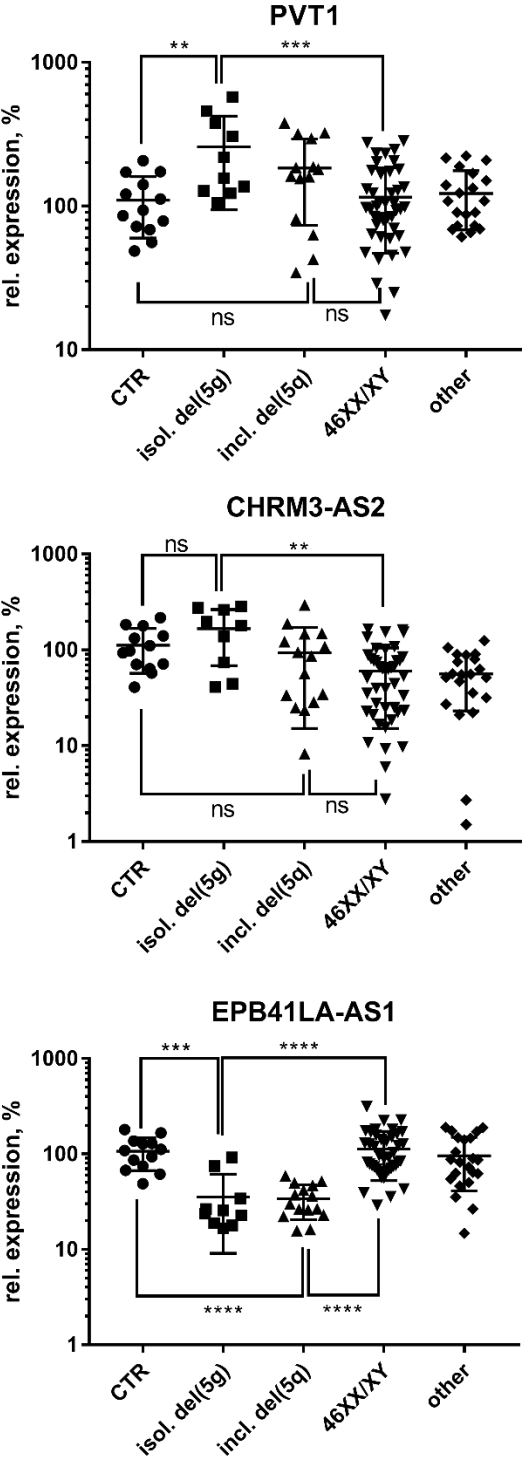
SI 8. Gene set enrichment analysis (GSEA) of differentially expressed PCGs in MDS/AML-MRC patients with isolated del(5q) vs. those with normal karyotype. Four selected enrichment plots are shown. Gene sets with FDR < 0.25 were considered as significantly enriched and only the top eight upregulated gene sets are listed. NES -normalized enrichment score, FDR - false discovery rate. References: Wienerga et al. [10], Jaatinen et al. [11], Ross et al. [12], Eppert et al. [13], Graham et al. [14], Lim et al. [15], and Massarweh et al. [16].

GSEA: del(5q) patients *vs.* normal karyotype patients

Gene sets	NES	p	FDR
Upregulated in del(5q) patients (the top 8 gene sets)			
STAT5 targets DN (Wierenga)	2.79	0.000	0.000
Hematopoietic stem cell DN (Jaatinen)	2.71	0.000	0.000
AML of FAB M7 type (Ross)	2.59	0.000	0.000
Progenitor (Eppert)	2.54	0.000	0.001
Platelet activation, signaling, and aggregation (Reactome)	2.52	0.000	0.000
CML quiescent vs. normal quiescent UP (Graham)	2.46	0.000	0.001
Mammary stem cell UP (Lim)	2.43	0.000	0.001
Heme metabolism (Hallmark)	2.42	0.000	0.001
Downregulated in del(5q) patients			
3´-UTR mediated translational regulation (Reactome)	-2.15	0.003	0.047
Translation (Reactome)	-2.03	0.008	0.109
Translational initiation (GO)	-1.99	0.003	0.115
Ribosomal subunit (GO)	-1.95	0.003	0.122
Tamoxifen resistance DN (Massarweh)	-1.87	0.000	0.200
Nuclear transcribed mRNA catabolic process nonsense mediated decay (GO)	-1.83	0.011	0.229
Hematopoietic stem cell UP (Jaatinen)	-1.82	0.011	0.214
Ribosome (GO)	-1.80	0.015	0.214
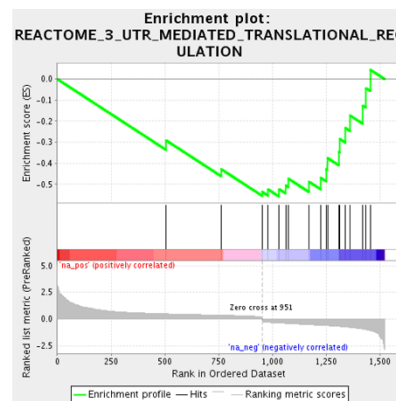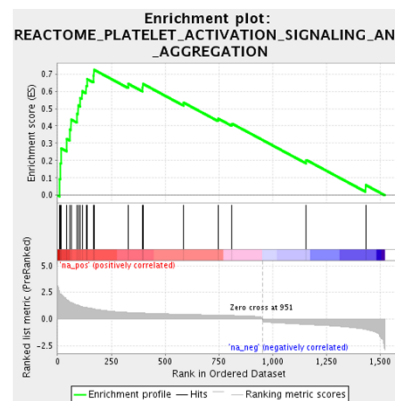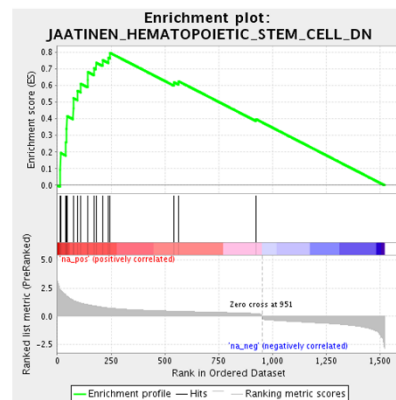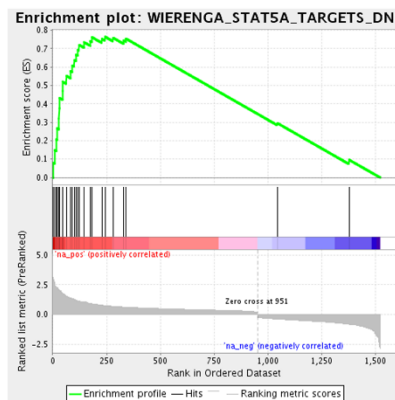
SI 9. Selected modules of the coexpression network designed based on differentially expressed genes between MDS/AML-MRC patients with isolated del(5q) and those with normal karyotype. Gene ontology (GO) terms significantly (p < 0.01) associated with these modules are listed in the corresponding tables. Square – PCG, circle – lncRNA, red – upregulation in del(5q) patients, blue – downregulation in del(5q) patients.

Ribosome & translation-related module

Del(5q)_MODULE_1	
GO term	P value
RNA binding	1.23E-07
Translational initiation	1.49E-07
Ribonucleoprotein complex	2.32E-07
Protein localization to endoplasmatic reticulum	1.09E-06
Nuclear transcribed mRNA catabolic process nonsense mediated decay	9.25E-06
Protein targeting to membrane	1.31E-05
Ribosome	4.51E-05
Large ribosomal subunit	5.47E-05

JAK/STAT-related module
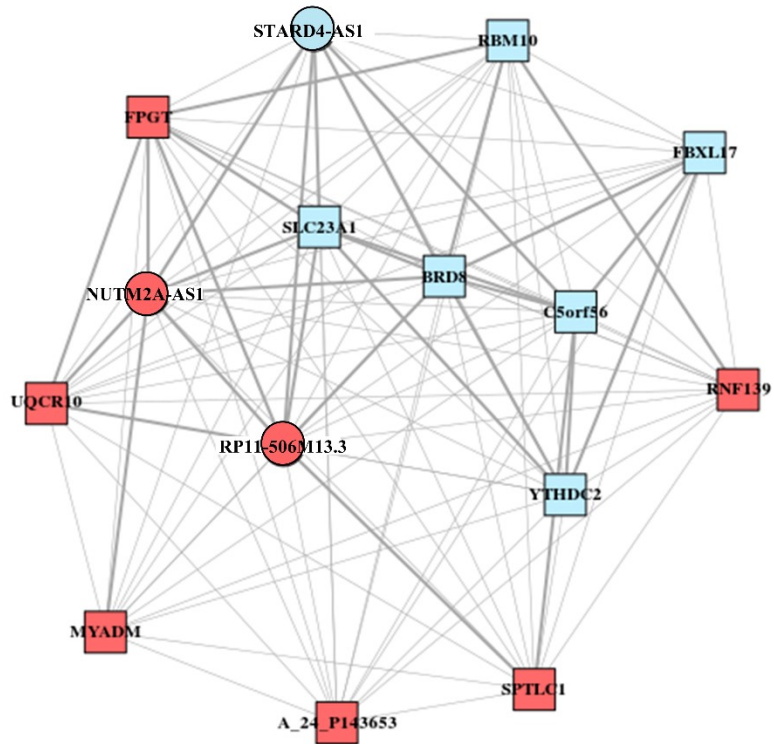
Del(5q)_MODULE_2	
GO term	P value
Intracellular receptor signaling pathway	2.44E-04
STAT cascade	2.50E-04
Positive regulation of immune system process	1.08E-03
Regulation of VEGFR signaling pathway	3.77E-03
Tyrosine phosphorylation of STAT protein	4.48E-03
Regulation of cell activation	7.98E-03
Cytokine activity	8.12E-03
JAK/STAT cascade involved in growth hormone signaling pathway	9.14E-03

Blood cell-related modules
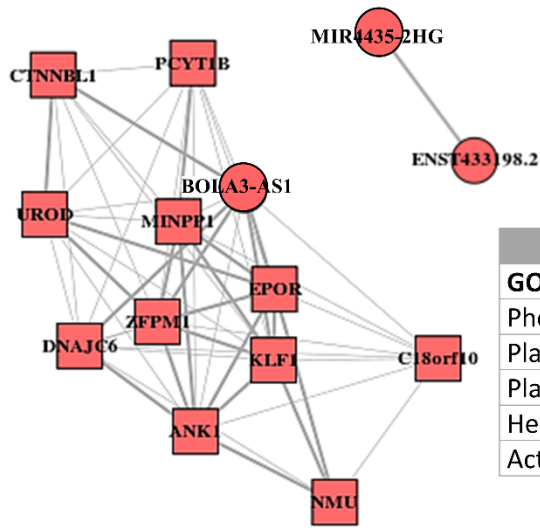
Del(5q)_MODULE_3	
GO term	**P value**
Phosphatase inhibitor activity	1.74E-06
Platelet morphogenesis	3.26E-06
Platelet activation	3.32E-05
Hemostasis	2.50E-04
Activation of MAPK activity	3.40E-04
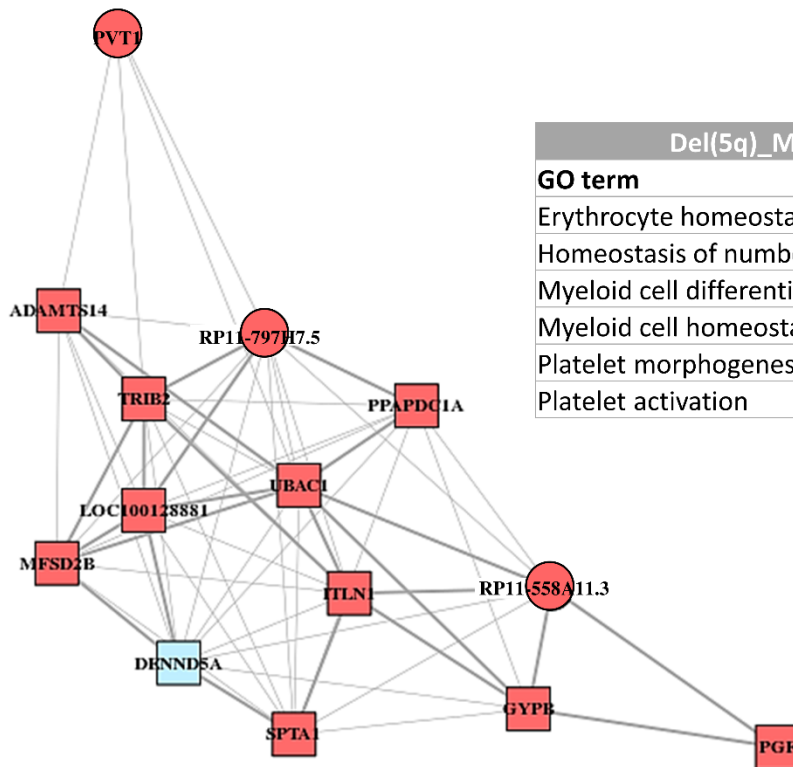
Del(5q)_MODULE_4	
GO term	**P value**
Erythrocyte homeostasis	4.22E-04
Homeostasis of number of cells	4.79E-04
Myeloid cell differentiation	9.27E-04
Myeloid cell homeostasis	1.12E-03
Platelet morphogenesis	1.49E-03
Platelet activation	2.21E-03

Mitochondrion-related module
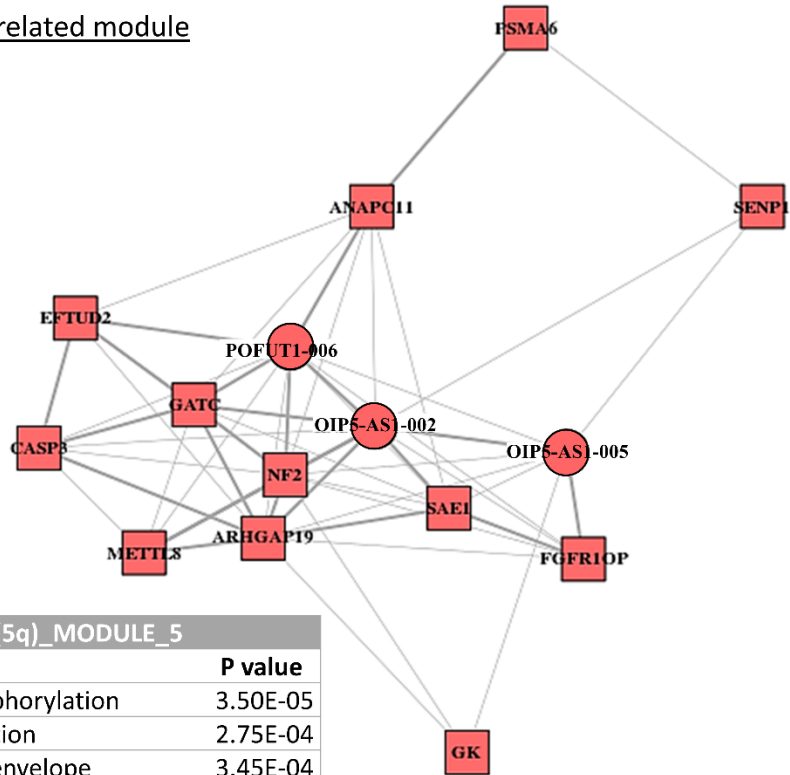
Del(5q)_MODULE_5	
GO term	**P value**
Oxidative phosphorylation	3.50E-05
Cellular respiration	2.75E-04
Mitochondrial envelope	3.45E-04
Organelle inner membrane	3.88E-04
Mitochondrion	8.99E-04
Mitochondrial electron transport ubiquinol to cytochrome c	1.26E-03

SI 10. Characterization of somatic mutations in the discovery cohort.

Pt ID	Dg	Gene	Variant	VAF (%)	AA change	COSMIC
PT1	RS-MLD	n.a.				
PT2	SLD	n.a.				
PT3	EB2	n.a.				
PT4	EB2	TET2	4:106180852_T/TG	16	Y1294*	NA
		RUNX1	21:36259156_A/G	21	p.L112P	COSM3737026
		BCOR	X:39932628_G/GT	34	Y657*	NA
		SRSF2	17:74732962_C/G	39	p.R94P	NA
		STAG2	X:123215311_C/T	38	p.R953*	COSM1114446
		NOTCH1	9:139391061_G/A	52	p.P2377L	COSM4745912
		RAD21	8:117868527_A/C	9	p.M272R	NA
PT5	RS	TET2	4:106157698_T/C	50	p.Y867H	COSM327337
PT6	EB2	SF3B1	2:198267361	9	p.K666E	NA
PT7	SLD	wt				
PT8	MLD	RUNX1	21:36252880_A/G	49	p.L161P	COSM444417
		SRSF2	17:74732959_G/A	50	p.P95L	COSM146288
		IKZF1	7:50468176_T/A	16	p.F471I	NA
		SETBP1	18:42531907_G/A	50	p.D868N	COSM1318400
PT9	AML-MRC	wt				
PT10	SLD	wt				
PT11	EB1	RUNX1	21:36259172_G/A	33	p.R107C	COSM24736
		ASXL1	20:31022366_T/TA	38	p.K618fs*1	COSM96394
		IKZF1	7:50450292_A/G	22	p.N159S	NA
		ETV6	12:12037413_G/C	40	p.L348F	NA
		PTPN11	12:112926887_G/A	6	p.G503R	NA
PT12	EB1	SF3B1	2:198266834_T/C	39	p.K700E	COSM84677
		TET2	4:106197269_C/T	46	p.H1868Y	COSM166837
		CUX1	7:101813749_G/T	49	p.E260D	NA
PT13	MLD	IKZF1	7:50467903_A/T	5	p.K380*	COSM5487650
		U2AF1	21:44524456	43	p.S34F	NA
PT14		DNMT3A	2:25462085_C/T	10	c.2323-1G>A	NA

	AML-	BCOR	X:39921510_G/C	9	p.S1437*	NA
	MRC	IDH2	15:90631838_C/T	10	p.R172K	COSM33733
		PHF6	X:133559286_C/T	24	p.R342*	COSM144563
PT15	MLD	wt				
PT16	EB2	CUX1	7:101921292_A/C	38	p.K546Q	NA
		ZRSR2	X:15841230_C/CAGCCGG	48	p.R448_R449insSR	COSM5762985
PT17	EB1	TP53	17:7578235_T/G	86	p.Y205S	COSM215719
PT18	5q-	wt				
PT19	MLD	wt				
PT20	EB2	TET2	4:106180823_CCTT/C	45	p.F1285del	COSM2270963
		RUNX1	21:36164716_C/CG	45	S388fs*212	NA
		ASXL1	20:31024704_G/A	50	p.G1397S	COSM133033
		EZH2	7:148523591_G/A	49	p.R288*	COSM1000721
		PHF6	X:133527949_C/T	45	p.R129*	COSM4606367
PT21	AML-	SF3B1	2:198266834_T/C	12	p.K700E	COSM84677
	MRC	IDH2	15:90631913_A/C	55	p.V147G	NA
PT22	SLD	wt				
PT23	EB1	TP53	17:7578416_C/A	61	p.V172F	COSM3378354
PT24	EB2	TP53	17:7577548_C/T	10	p.G245S	COSM121035
PT25	EB2	RUNX1	21:36164892_G/GT	26	p.T328fs*272	NA
		EZH2	7:148516697_G/C	93	p.Y330*	NA
		IKZF1	7:50450292_A/G	25	p.N159S	NA
PT26	MLD	CEBPA	19:33792731_G/GGCGGGT	39	p.P196_P197insHP	COSM4170207
PT27	RS	SF3B1	2:198267371	40	p.H662Q	NA
		TET2	4:106197269	34	p.H1868Y	NA
PT28	EB1	TP53	17:7577120_C/A	49	p.R273L	COSM10779
		BCORL1	X:129159288_C/T	15	p.R1338*	NA
		CDKN2A	9:21994269_C/T	46	p.R21K	NA
PT29	EB1	TET2	4:106158343_G/T	96	p.E1082*	NA
		NOTCH1	9:139391061_G/A	54	p.P2377L	COSM4745912
		CUX1	7:101848393_G/T	14	c.3107-1G>T	NA
		ZRSR2	X:15840905_A/G	95	p.H330R	NA
PT30	AML-	JAK2	9:5073770_G/T	19	p.V617F	COSM12600
	MRC	KDM6A	X:44966740_C/CG	43	p.R1322*	NA

PT31	AML-MRC	BCORL1	X:129173198_A/G	61	p.Y1520C	NA
PT32	5q-	DNMT3A	2:25457243	40	p.R882C	NA
		EZH2	7:148523591	37	p.R288*	NA
PT33	MLD	DNMT3A	2:25523096_T/G	50	p.E30A	COSM307361
PT34	EB1	TP53	17:7578271	31	p.H193L	NA
PT35	RS	SF3B1	2:198267491_C/A	45	p.E622D	COSM110693
		DNMT3A	2:25464481	49	p.Q678*	NA
PT36	5q-	wt				
PT37	EB2	RUNX1	21:36171751_G/A	20	p.Q272*	NA
PT38	EB1	TP53	17:7577099_C/A	26	p.R280I	COSM11287
		ASXL1	20:31022397_A/G	7	p.R628G	NA
		STAG2	X:123182927_C/T	58	p.R298C	COSM5565467
		NOTCH1	9:139391061_G/A	45	p.P2377L	COSM4745912
		U2AF1	21:44514777_T/G	11	p.Q157P	COSM1318797
		GATA1	X:48649530_G/A	9	p.G5D	NA
PT39	SLD	JAK2	9:5073770_G/T	20	p.V617F	COSM12600
PT40	MLD	SF3B1	2:198266834_T/C	44	p.K700E	COSM84677
PT41	EB2	TP53	17:7577099_C/A	5	p.R280I	COSM11287
		DNMT3A	2:25523040	13	p.G49R	NA
		U2AF1	21:44514777_T/G	6	p.Q157P	COSM1318797
		CSF3R	1:36932356	6	p.C732Y	NA
PT42	AML-MRC	DNMT3A	2:25457242_C/T	13	p.R882H	COSM52944
PT43	5q-	DNMT3A	2:25468122_C/G	24	p.K518N	NA
PT44	MLD	ASXL1	20:31022877_G/GAATGTGAGTCTGGCACCACTT	29	p.S795*	NA
		SRSF2	17:74732935_CGGCGGCTGTGGTGTGAGTCCGGGG/C	39	p.P95_R102del	COSM146289
		STAG2	X:123159689_G/A	78	c.45-1G>A	NA
		IDH2	15:90631934_C/T	42	p.R140Q	COSM41590
PT45	5q-	TET2	4:106157703_T/G	49	p.F868L	COSM87107
		DNMT3A	2:25457252_T/C	46	p.N879D	COSM1583135
		ASXL1	20:31022413_A/G	18	H633R	NA
		BCOR	X:39933593_A/AG	39	p.S336fs*45	COSM4385748
PT46	RS	SF3B1	2:198267491_C/A	45	p.E622D	COSM110693
		TET2	4:106162585_A/G	47	R1167G	NA

		TP53	17:7578268_A/C	54	p.L194R	COSM117647
		ASXL1	20:31022413_A/G	18	p.H633R	NA
PT47	EB1	SF3B1	2:198266711_T/C	5	p.K741E	COSM5419581
PT48	AML-MRC	JAK2	9:5073770_G/T	41	p.V617F	COSM12600
PT49	MLD	wt				
PT50	5q-	wt				
PT51	EB2	SF3B1	2:198266611_C/T	21	p.G742D	COSM145923
		MLL	11:118353156_G/GA	39	p.V1347fs*24	NA
PT52	AML-MRC	TP53	17:7579458_GT/G	11	p.P77fs*46	
PT53	EB2	ABL1	9:133738340_A/G	52	p.K266R	NA
PT54	AML-MRC	CBL	11:119148936_G/A	8	p.E386K	NA
		NRAS	1:115258747_C/T	8	p.G12D	COSM564
PT55	MLD	PHF6	X:133511650_G/A	7	p.M1I	NA
		CBL	11:119149251_G/A	13	p.R420Q	COSM34077
PT56	AML-MRC	SF3B1	2:198266834_T/C	32	p.K700E	COSM84677
		RUNX1	21:36164838_C/CG	40	p.R346fs*137	COSM36063
		BCOR	X:39932027_C/A	55	p.E858*	NA
PT57	EB2	SF3B1	2:198267705_C/T	50	p.E592K	COSM132936
		RUNX1	21:36206722_G/A	22	p.Q264*	NA
		EZH2	7:148514414_C/T	13	p.W437*	NA
		STAG2	X:123179121_CAT/C	14	p.I191*	NA
PT58	AML-MRC	wt				
PT59	RS-SLD	wt				
PT60	MLD	SF3B1	2:198267371	41	p.H662Q	NA
PT61	EB1	MLL	11:118307399_CCGGCTGTGGCGGCCGCGGCGGCGGCGGCGGGAAGCAGCGGGGCTGGGGTTCCAGGGGGAG/C	12	p.V60_A79del	NA
PT62	AML-MRC	TET2	4:106196821_A/AG	6	p.K1720fs*9	NA
		BCOR	X:39916501_A/G	10	p.L1501P	NA
PT63	MLD	SRSF2	17:74732959_G/A	28	p.P95R	COSM146288
		SETBP1	18:42531907_G/A	30	p.D868N	COSM1318400
PT64		RUNX1	21:36231782_C/T	45	p.R201Q	COSM24805

Pt ID	dg	gene	variant	VAF %	AA change	COSMIC
	AML-MRC	ASXL1	20:31022402_TCACCACTGCCATAGAGAGGCGGC/T	22	p.E635fs*15	COSM51200
		EZH2	7:148506443_C/T	36	p.R690H	COSM52980
		RAD21	8:117866482_A/C	63	c.1161+2T>G	NA
PT65	RS	SF3B1	2:198267483_C/A	28	p.R625L	COSM110695
PT66	5q-	TP53	17:7577566_T/C	10	p.N239D	COSM10777
PT67	MLD	SF3B1	2:198266834_T/C	50	p.K700E	COSM84677
		TET2	4:106164793_T/G	10	p.C1221G	NA
		DNMT3A	2:25458593_C/T	47	p.W860*	COSM4169946
		JAK2	9:5073770_G/T	9	p.V617F	COSM12600
PT68	AML-MRC	n.a.				

Pt ID – patient identification number, dg – diagnosis, VAF % - percentage of variant allele frequency, AA change – amino acid change, n.a. - not analyzed, wt - wild type

SI 11. Frequency and distribution of somatic mutations in the discovery cohort.
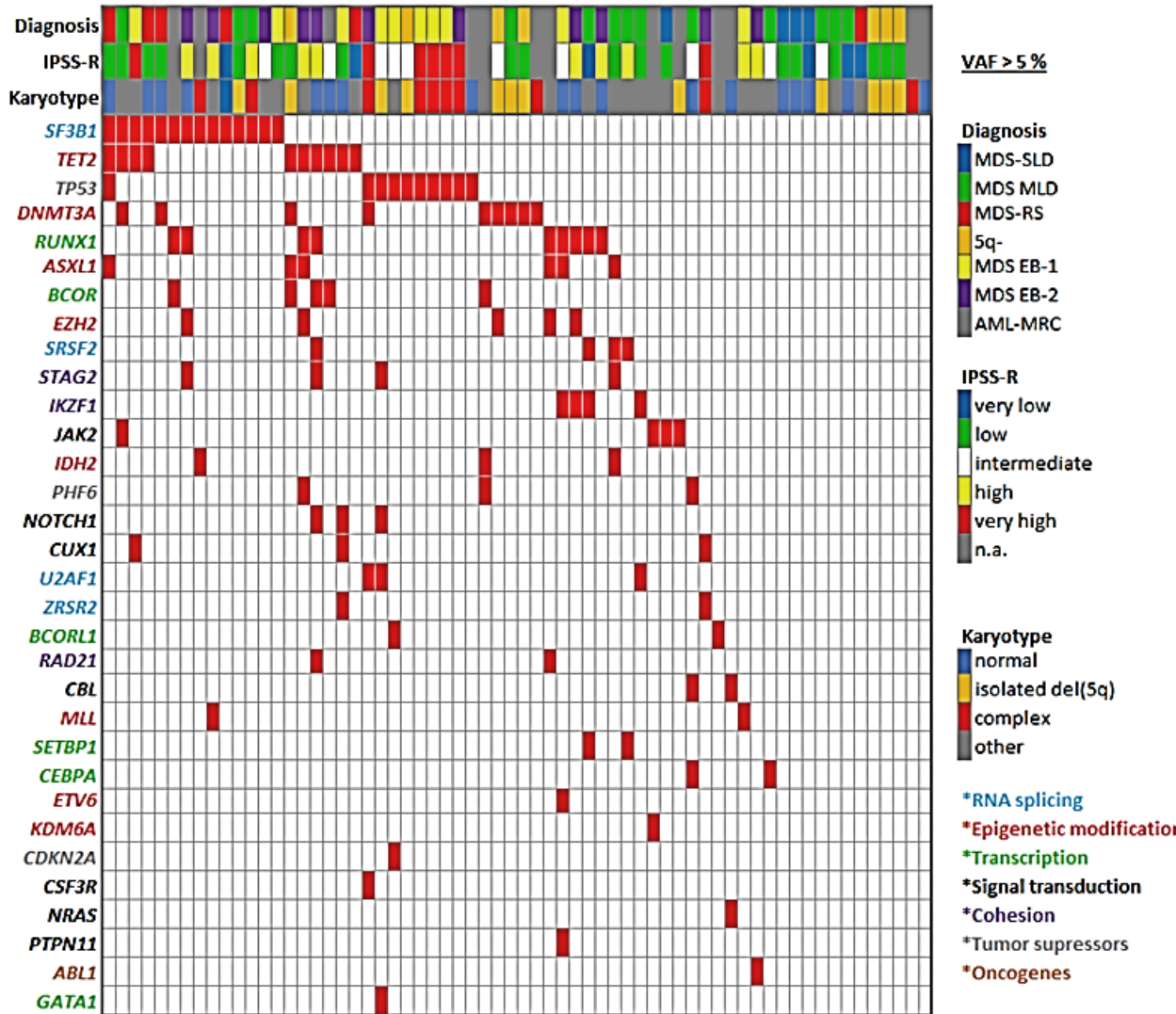
SI 12. List of significantly deregulated transcripts in MDS patients with vs. without a SF3B1 mutation (|logFC| > 0.3, FDR < 0.05). logFC – binary logarithm of fold change, FDR - false discovery rate.

No.	Transcript	Chromosome	logFC	FDR
lncRNAs increased in the patients with mutated SF3B1				
1	RP11-380O24.1	chr3	0.64	3.12E-02
2	AL592435.1	chr1	0.62	2.13E-02
3	MIR1302-11	chr19	0.56	1.61E-02
4	LINC00959	chr10	0.56	1.61E-02
5	AC093415.2	chr3	0.49	2.70E-03
6	LINC00705	chr10	0.45	4.43E-02
7	RP11-692D12.1	chr4	0.44	4.43E-02
8	RP11-809N15.2	chr6	0.42	1.61E-02
9	RP11-211C9.1	chr8	0.35	1.61E-02
10	RP11-446J8.1	chr4	0.35	1.60E-02
11	USP3-AS1	chr15	0.34	1.09E-02
12	AC005786.7	chr19	0.31	4.97E-02
lncRNAs reduced in the patients with mutated SF3B1				
1	RP11-710F7.3	chr4	-0.66	4.71E-02
2	PCBP1-AS1	chr2	-0.60	4.90E-02
3	RP11-872J21.3	chr14	-0.58	1.97E-02
4	RP11-348M17.2	chr5	-0.36	4.97E-02
5	AL163953.3	chr14	-0.34	3.12E-02
6	LINC00877	chr3	-0.34	4.71E-02
PCGs increased in the patients with mutated SF3B1				
1	AB209400	chr20	2.68	1.66E-03

2	TCAM1P	chr17	1.86	4.49E-02
3	ZNF541	chr19	1.34	5.60E-03
4	CLIC2	chrX	0.79	1.95E-02
5	KLF11	chr2	0.54	2.53E-02
6	C15orf40	chr15	0.39	2.39E-02
PCGs reduced in the patients with mutated SF3B1				
1	ZNF883	chr9	-1.49	1.97E-02
2	LMO1	chr11	-1.49	1.04E-02
3	GIPC2	chr1	-1.32	6.35E-03
4	ARHGAP10	chr4	-0.96	4.86E-02
5	RTF1	chr15	-0.93	1.82E-03
6	ABCB7	chrX	-0.88	7.08E-03
7	RECQL	chr12	-0.81	7.83E-03
8	ACD	chr16	-0.81	3.03E-04
9	GAGE1	chrX	-0.57	6.35E-03
10	EAPP	chr14	-0.48	3.96E-02
11	NNT	chr5	-0.46	2.39E-02
12	ATP11C	chrX	-0.44	6.35E-03
13	LOC100129518	chr6	-0.39	2.67E-02
14	POLG	chr15	-0.32	4.86E-02

SI 13. List of significantly deregulated transcripts in MDS patients with *vs.* without a TET2 mutation (|logFC| > 0.3, FDR < 0.05). logFC – binary logarithm of fold change, FDR - false discovery rate.

No.	Transcript	Chromosome	logFC	FDR
lncRNAs increased in the patients with mutated TET2				
1	CTD-2231H16.1	chr5	0.56	4.29E-02
2	VIPR1-AS1	chr3	0.53	2.78E-02
3	LINC00518	chr6	0.46	2.78E-02
4	LINC01193	chr15	0.44	4.20E-02
5	RP5-1109J22.2	chr1	0.40	1.92E-02
6	RP11-325N19.3	chr15	0.39	8.53E-03
7	TBX5-AS1	chr12	0.38	1.92E-02
8	RP11-104J23.2	chr17	0.30	3.45E-02
lncRNAs reduced in the patients with mutated TET2				
1	WT1-AS	chr11	-2.70	1.92E-02
2	EMCN-IT1	chr4	-2.11	1.93E-02
3	RP11-185E8.1	chr3	-1.56	3.64E-02
4	RP4-773A18.4	chr1	-0.80	4.29E-02
5	AC004947.2	chr7	-0.54	1.93E-02
PCGs increased in the patients with mutated TET2				
1	MAP2K3	chr17	0.49	4.20E-02
2	RGS8	chr1	0.43	4.20E-02
3	BPIFA3	chr20	0.36	2.66E-02
lncRNAs reduced in the patients with mutated TET2				
1	PLAC1	chrX	-1.86	2.66E-02
2	FAM83E	chr19	-0.53	4.20E-02

SI 14. List of significantly deregulated transcripts in MDS patients with vs. without a TP53 mutation (|logFC| > 0.3, FDR < 0.05). logFC – binary logarithm of fold change, FDR - false discovery rate, ns – none significant transcript.

No.	Transcript	Chromosome	logFC	FDR
lncRNAs increased in the patients with mutated TP53				
ns	ns	ns	ns	ns
lncRNAs reduced in the patients with mutated TP53				
1	STARD4-AS1	chr5	-1.11	3.83E-02
2	RP5-1050D4.5	chr17	-0.89	2.03E-02
3	RP11-53I6.3	chr18	-0.87	2.03E-02
4	RP11-347I19.8	chr12	-0.81	3.72E-02
5	RP11-325L7.2	chr5	-0.67	4.47E-02
6	RP11-57H14.2	chr10	-0.63	3.88E-02
7	RP11-351M16.3	chr10	-0.60	2.03E-02
8	RP11-169E6.1	chr16	-0.40	4.11E-02
PCGs increased in the patients with mutated TP53				
ns	ns	ns	ns	ns
PCGs reduced in the patients with mutated TP53				
ns	ns	ns	ns	ns

SI 15. List of significantly deregulated transcripts in MDS patients with *vs.* without a DNMT3A mutation (|logFC| > 0.3, FDR < 0.05). logFC – binary logarithm of fold change, FDR - false discovery rate, ns – none significant transcript.

No.	Transcript	Chromosome	logFC	FDR
lncRNAs increased in the patients with mutated DNMT3A				
1	RP11-68L1.1	chr3	0.31	4.06E-03
lncRNAs reduced in the patients with mutated DNMT3A				
ns	ns	ns	ns	ns
PCGs increased in the patients with mutated DNMT3A				
ns	ns	ns	ns	ns
PCGs reduced in the patients with mutated DNMT3A				
ns	ns	ns	ns	ns

SI 16. List of significantly deregulated transcripts in MDS patients with vs. without a RUNX1 mutation (|logFC| > 0.3, FDR < 0.05). Of the 67 increased and 39 reduced lncRNAs, and the 206 increased and 440 reduced PCGs, only the top 30 transcripts in each category are listed. logFC – binary logarithm of fold change, FDR - false discovery rate.

No.	Transcript	Chromosome	logFC	FDR
lncRNAs increased in the patients with mutated RUNX1				
1	AC068057.2	chr2	2.65	8.76E-03
2	C9orf106	chr9	2.00	2.08E-02
3	RP11-66B24.1	chr15	1.62	4.11E-03
4	LINC01071	chr13	1.50	3.96E-03
5	FAM225B	chr9	1.36	3.07E-02
6	AJ271736.10	chrX	1.22	3.38E-03
7	WASIR2	chr16	1.13	2.22E-03
8	RBPMS-AS1	chr8	1.08	4.49E-02
9	RP11-433M22.1	chr17	1.06	2.96E-02
10	RP11-490M8.1	chr2	1.01	1.62E-02
11	HCG9	chr6	1.00	1.89E-02
12	RP11-750H9.5	chr11	0.92	4.95E-05
13	RP11-374M1.4	chr9	0.88	3.78E-02
14	RP11-834C11.3	chr12	0.88	3.63E-02
15	RP11-783L4.1	chr14	0.87	2.48E-02
16	LINC01257	chr12	0.81	3.28E-02
17	RP11-626G11.3	chr16	0.85	4.11E-03
18	RP11-527L4.5	chr17	0.83	5.14E-03
19	RP1-309F20.3	chr20	0.81	3.59E-02
20	RP11-65J3.15	chr9	0.81	4.22E-02
21	PAN3-AS1	chr13	0.81	4.27E-02
22	RP5-1024C24.1	chr11	0.81	1.49E-02

23	PRKAG2-AS1	chr7	0.81	6.36E-03
24	RP11-650P15.1	chr18	0.81	3.90E-02
25	RP1-122P22.2	chr20	0.81	4.27E-02
26	HCP5	chr6	0.78	2.77E-02
27	LA16c-321D4.2	chr16	0.75	4.42E-02
28	FAM95B1	chr9	0.75	2.48E-02
29	GAS5	chr1	0.73	3.64E-03
30	LINC00954	chr2	0.70	1.90E-02
lncRNAs reduced in the patients with mutated RUNX1				
1	AL928768.3	chr14	-5.94	4.49E-02
2	LINC01013	chr6	-4.20	1.13E-03
3	TCL6	chr14	-4.11	4.11E-03
4	RP11-542K23.7	chr9	-4.02	4.11E-03
5	LEF1-AS1	chr4	-2.85	3.39E-03
6	RP11-222A5.1	chr1	-2.12	1.84E-02
7	RP11-161M6.2	chr16	-2.09	2.37E-02
8	RP11-161M6.2	chr16	-2.09	4.19E-02
9	RP11-175K6.1	chr5	-1.94	1.20E-02
10	RP11-558A11.3	chr16	-1.86	3.60E-02
11	RP11-56F10.3	chr9	-1.78	1.13E-03
12	LINC01218	chr4	-1.54	4.84E-02
13	PCED1B-AS1	chr12	-1.37	4.49E-02
14	RP11-78B10.2	chr1	-1.25	4.11E-03
15	RP11-394O9.1	chr9	-1.25	4.87E-02
16	RP11-435B5.4	chr1	-1.20	4.10E-02
17	RP11-384O8.1	chr2	-0.90	2.46E-02

18	AC005307.1	chr19	-0.89	4.68E-02
19	GS1-421I3.2	chrX	-0.83	2.66E-02
20	LINC00226	chr14	-0.82	3.10E-02
21	RP11-83N9.5	chr9	-0.81	2.14E-02
22	RP11-417J8.3	chr1	-0.72	4.05E-02
23	RP11-664D1.1	chr12	-0.70	4.15E-02
24	RP11-584P21.2	chr4	-0.69	4.99E-02
25	AC019221.4	chr2	-0.69	4.05E-02
26	AC007381.3	chr2	-0.63	2.60E-02
27	CTC-444N24.13	chr19	-0.60	2.91E-02
28	RP11-527H14.4	chr18	-0.59	1.89E-02
29	CTD-2384A14.1	chr14	-0.52	4.25E-02
30	RP11-619L12.3	chr5	-0.47	3.10E-02
PCGs increased in the patients with mutated RUNX1				
1	SCARA3	chr8	2.44	4.41E-02
2	BAI1	chr8	1.87	4.64E-02
3	ANKRD65	chr1	1.85	4.97E-02
4	HTRA3	chr4	1.77	2.19E-02
5	PRDM16	chr1	1.56	1.07E-02
6	MAMDC2	chr9	1.53	3.49E-02
7	KBTBD12	chr3	1.37	2.77E-02
8	GIMAP2	chr7	1.22	2.77E-02
9	CPNE7	chr16	1.20	2.49E-02
10	LRP6	chr12	1.17	3.39E-05
11	GPR162	chr12	1.17	4.97E-02
12	LOC100134167	chr9	1.17	1.26E-02

13	AK130024	chr17	1.15	4.44E-06
14	LINC00256B	chr9	1.10	4.56E-02
15	CCDC149	chr4	1.10	2.05E-02
16	PRKCA	chr17	1.09	8.18E-03
17	LRRD1	chr7	1.08	4.18E-02
18	LOC100288911	chr2	1.06	7.15E-03
19	ANPEP	chr15	1.05	4.57E-02
20	HLA-DQB1	chr6	1.05	3.69E-02
21	UBA7	chr3	1.04	2.79E-02
22	HLA-DOA	chr6	1.04	8.18E-03
23	ISG20	chr15	1.04	1.36E-02
24	GNPDA1	chr5	1.03	2.31E-02
25	HSBP1L1	chr18	0.99	4.12E-04
26	GALNT14	chr2	0.98	1.88E-02
27	KCNE3	chr11	0.97	4.54E-03
28	BTG2	chr1	0.95	1.69E-02
29	FBXO15	chr18	0.93	4.16E-02
30	C1orf151-NBL1	chr1	0.93	7.97E-03

PCGs reduced in the patients with mutated RUNX1

1	POU4F1	chr13	-5.83	2.27E-02
2	LEF1	chr4	-5.03	2.75E-06
3	NPY	chr7	-4.73	4.44E-06
4	IGKV116	chr2	-4.69	2.77E-02
5	IGKV1D-43	chr2	-4.39	1.67E-02
6	IGLV1-47	chr22	-4.05	2.19E-02
7	IGKV1D-8	chr2	-3.98	2.23E-02

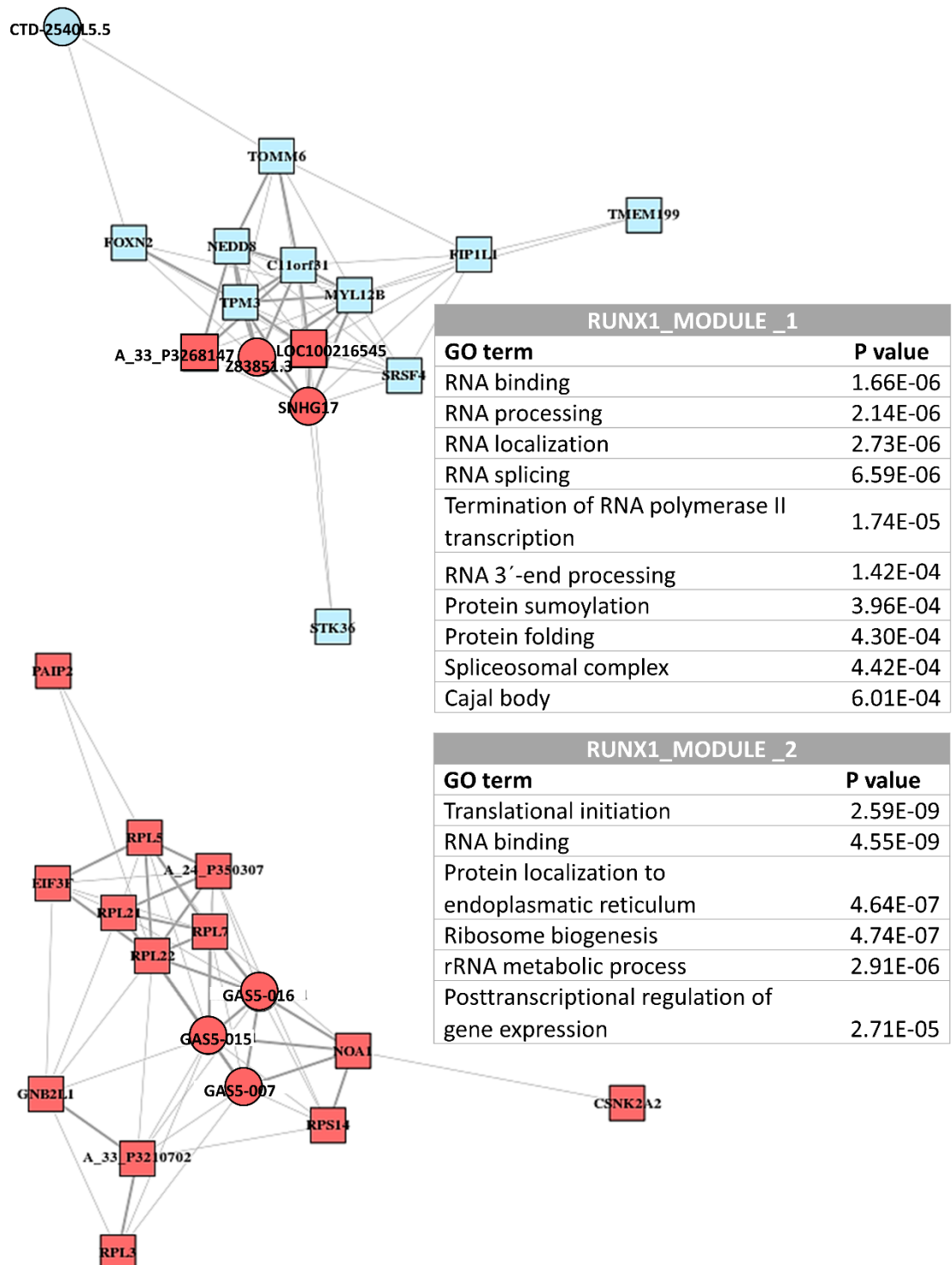|				
8	RAG1	chr11	-3.96	1.02E-04
9	AB363267	chr2	-3.91	2.55E-03
10	IGLV1-44	chr22	-3.87	3.42E-02
11	IGHV1-18	chr14	-3.78	2.75E-06
12	NP113779	chr2	-3.68	2.97E-02
13	LOC100653210	chr2	-3.42	7.87E-03
14	AF194718	chr22	-3.41	2.04E-02
15	IGLV3-10	chr22	-3.38	1.18E-02
16	IRX1	chr5	-3.28	2.96E-02
17	DUSP26	chr8	-3.25	1.55E-02
18	IGLV3-9	chr22	-3.19	2.42E-02
19	COL6A5	chr3	-3.18	9.11E-08
20	IGLV3-25	chr22	-3.17	1.75E-02
21	IGKV1D-16	chr2	-3.17	6.81E-03
22	IGHV1-2	chr14	-3.03	8.80E-06
23	IGKV1D-27	chr22	-2.98	1.36E-02
24	IGKV1D-8	chr2	-2.88	5.22E-03
25	MECOM	chr3	-2.79	2.90E-02
26	TGM2	chr20	-2.79	4.03E-02
27	SH2D4B	chr10	-2.78	1.76E-02
28	IFI27	chr14	-2.77	4.51E-02
29	ECEL1P2	chr2	-2.75	7.51E-04
30	NPTX2	chr7	-2.74	1.71E-02

SI 17. Gene set enrichment analysis (GSEA) of differentially expressed PCGs in MDS/AML-MRC patients with a RUNX1 mutation vs. those with the RUNX1 wildtype. Four selected enrichment plots are shown. Gene sets with FDR < 0.25 were considered as significantly enriched and only the top eight upregulated gene sets are listed. NES -normalized enrichment score, FDR - false discovery rate. References: Wang et al. [17], Haddad et al. [18], Yoshimura et al. [19], Martens et al. [20], and Heller et al. [21].

GSEA: MDS patients with *vs.* without *RUNX1* mutation

Gene sets	NES	p	FDR
Upregulated in RUNX1-mutated MDS			
Golgi apparatus part (GO)	2.23	0.000	0.050
Vesicle membrane (GO)	1.77	0.009	0.224
Response to GSK3 inhibitor SB216763 UP (Wang)	1.71	0.025	0.199
MYC targets V1 (Hallmark)	1.63	0.019	0.244
Protein phosphorylation (GO)	1.54	0.039	0.301
mRNA metabolic process (GO)	1.48	0.044	0.357
Downregulated in RUNX1-mutated MDS			
B lymphocyte progenitor (Haddad)	-1.91	0.001	0.127
MAPK8 targets UP (Yoshimura)	-1.80	0.001	0.272
Tretinoin response UP (Martens)	-1.79	0.001	0.179
Immune response (GO)	-1.68	0.005	0.479
Sequence specific DNA binding (GO)	-1.62	0.012	0.592
Immune system development (GO)	-1.62	0.018	0.555
HDAC targets silenced by methylation UP (Heller)	-1.58	0.030	0.609
Positive regulation of cell death (GO)	-1.57	0.023	0.612
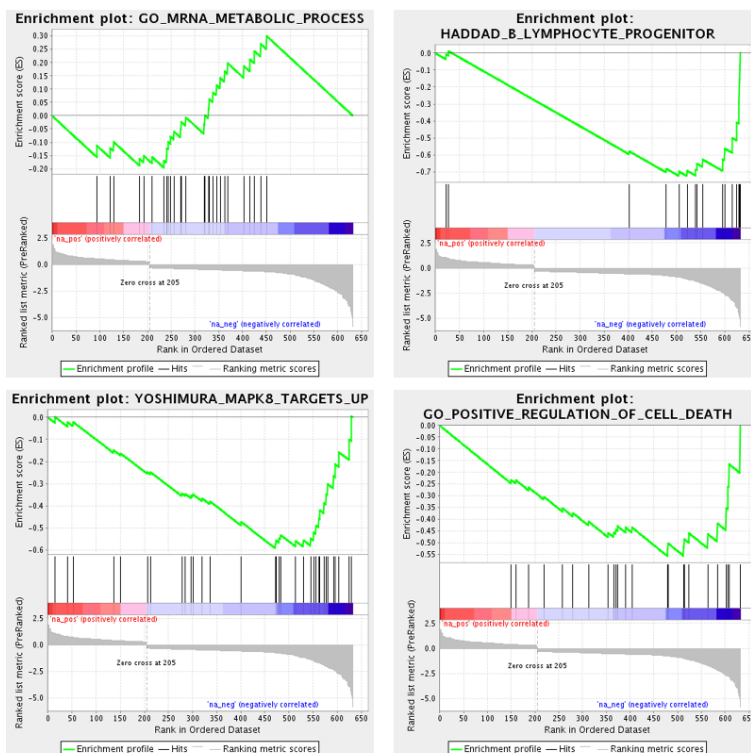
SI 18. Selected modules of the coexpression network designed based on differentially expressed genes between the RUNX1-mutated (mut) and RUNX1-wild type (wt) patients. Gene ontology (GO) terms significantly (p < 0.01) associated with these modules are listed in the corresponding tables. Square – PCG, circle – lncRNA, red – upregulation in RUNX1-mut, blue – downregulation in RUNX1-mut.

RUNX1_MODULE _1	
GO term	P value
RNA binding	1.66E-06
RNA processing	2.14E-06
RNA localization	2.73E-06
RNA splicing	6.59E-06
Termination of RNA polymerase II transcription	1.74E-05
RNA 3´-end processing	1.42E-04
Protein sumoylation	3.96E-04
Protein folding	4.30E-04
Spliceosomal complex	4.42E-04
Cajal body	6.01E-04

RUNX1_MODULE _2	
GO term	P value
Translational initiation	2.59E-09
RNA binding	4.55E-09
Protein localization to endoplasmatic reticulum	4.64E-07
Ribosome biogenesis	4.74E-07
rRNA metabolic process	2.91E-06
Posttranscriptional regulation of gene expression	2.71E-05

RUNX1_MODULE _3	
GO term	**P value**
Cell cycle	9.08E-05
Immunoglobulin complex	2.96E-04
Antimicrobial humoral response	1.31E-03
3' to 5' DNA helicase activity	1.45E-03
Beta-catenin-TCF complex formation	1.69E-03
Double-strand break repair	2.04E-03
V(D)J recombination	4.24E-03
p53 binding	5.05E-03
Chromatin remodeling	6.99E-03
Recombinational repair	8.99E-03
Somatic cell DNA recombination	9.57E-03
Negative regulation of signal transduction by p53 class mediator	9.57E-03
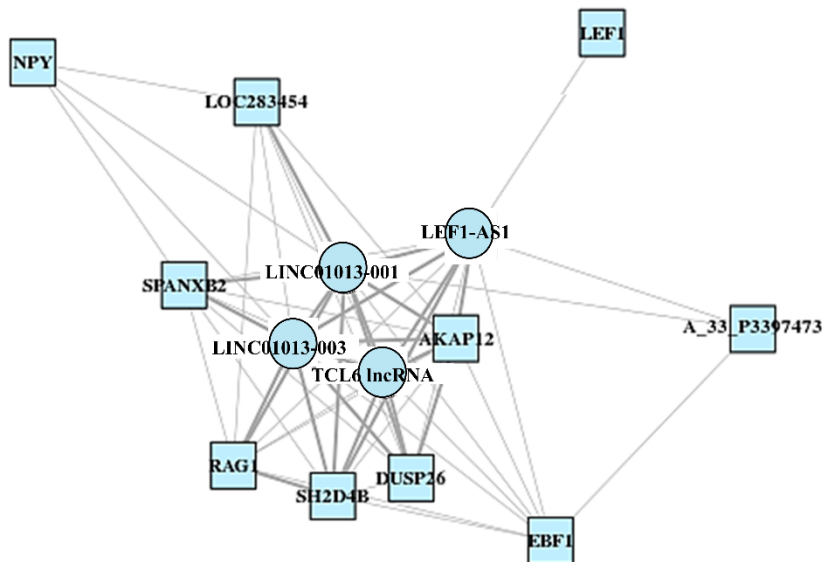
SI 19. List of significantly deregulated transcripts in long vs. short surviving patients (|logFC| > 1, FDR < 0.05). The cut-off for patient stratification was 18 months from the time of sample collection. logFC – binary logarithm of fold change, FDR - false discovery rate.

No.	Transcript	Chromosome	logFC	FDR
lncRNAs increased in short survivals				
1	H19	chr11	3.13	1.10E-02
2	WT1-AS	chr11	1.21	1.03E-02
3	AC093818.1	chr2	1.20	1.25E-02
4	ITGA6-AS1	chr2	1.06	3.12E-02
5	LBX2-AS1	chr2	1.02	3.12E-02
lncRNAs reduced in short survivals				
1	RP11-121P10.1	chr6	-2.48	3.12E-02
2	LINC01122	chr2	-1.48	1.25E-02
3	RP11-120K24.3	chr13	-1.35	2.66E-02
PCGs increased in short survivals				
1	PDE3B	chr11	1.61	2.96E-03
2	GPR124	chr8	1.56	4.41E-02
3	HIC1	chr17	1.45	2.84E-03
4	CD97	chr19	1.05	2.84E-03
5	FLJ90757	chr17	1.01	4.86E-02
PCGs reduced in short survivals				
1	ECEL1P2	chr2	-2.45	3.88E-04
2	TCEAL2	chrX	-2.27	3.87E-04
3	ST6GAL2	chr2	-1.90	1.15E-02
4	SLC1A6	chr19	-1.86	3.95E-03
5	PDZK1IP1	chr1	-1.70	5.69E-03

6	SH3GL3	chr15	-1.69	2.81E-02
7	TM7SF4	chr8	-1.66	4.41E-02
8	HLF	chr17	-1.64	8.92E-03
9	CDH7	chr18	-1.58	5.56E-03
10	CLCN4	chrX	-1.57	4.41E-02
11	NFIB	chr9	-1.52	6.39E-03
12	CXorf57	chrX	-1.45	8.24E-03
13	STAC	chr3	-1.44	5.69E-03
14	C3orf14	chr3	-1.43	8.24E-03
15	TMSB15A	chrX	-1.39	1.33E-02
16	PRKG2	chr4	-1.37	2.18E-02
17	AVP	chr20	-1.34	4.41E-02
18	THC2656240	chrX	-1.34	1.13E-03
19	JAM2	chr21	-1.32	5.69E-03
20	MCF2L-AS1	chr13	-1.30	8.24E-03
21	PCDH9	chr13	-1.21	4.86E-02
22	RP11-551L14.1	chr12	-1.12	2.11E-02
23	C7orf58	chr7	-1.02	3.12E-02
24	VWCE	chr11	-1.00	4.86E-02

SI 20. List of significantly deregulated transcripts in MDS patients with lower- vs. higher-risk IPSS-R (|logFC| > 1, FDR < 0.05). Of the 67 significantly reduced PCGs, only the top 30 transcripts are listed. logFC – binary logarithm of fold change, FDR - false discovery rate.

No.	Transcript	Chromosome	logFC	FDR
lncRNAs – increased in higher-risk MDS				
1	RP11-897M7.1	chr12	1.27	4.31E-02
2	LINC00539	chr13	1.02	4.03E-02
lncRNAs - reduced in higher-risk MDS				

1	TCL6	chr14	-4.48	4.73E-02
2	LINC01013	chr6	-4.33	2.46E-02
3	LEF1-AS1	chr4	-1.96	4.29E-02
4	CTC-436K13.2	chr5	-1.95	4.83E-02
5	RP11-474N8.5	chr12	-1.88	4.29E-02
6	AC096579.7	chr2	-1.84	2.46E-02
7	RP11-879F14.2	chr18	-1.55	4.73E-02
8	RP11-69I8.3	chr6	-1.35	4.04E-02
9	LINC01037	chr1	-1.34	4.91E-02
10	RP11-401P9.5	chr16	-1.26	4.73E-02
11	AC147651.3	chr7	-1.20	4.73E-02
12	RP11-71G12.1	chr1	-1.08	1.47E-02
13	RP3-523C21.2	chr6	-1.05	4.73E-02
14	CTA-250D10.23	chr22	-1.02	4.29E-02
PCGs - increased in higher-risk MDS				
1	BAI1	chr8	2.56	7.51E-03
2	ARC	chr8	2.55	1.32E-02
3	PTH2R	chr2	1.83	2.17E-02
4	MAMDC2	chr9	1.77	4.23E-02
5	LOXL4	chr10	1.67	4.40E-02
6	NPM2	chr8	1.36	1.58E-02
7	A_33_P3387493	chrX	1.30	4.41E-02
8	A_24_P247454	chr2	1.29	3.60E-02
9	KRT18	chr4	1.23	2.77E-02
10	A_24_P230057	chrX	1.22	2.30E-02
11	A_24_P401601	chr19	1.20	9.47E-03

12	KRT18P55	chr17	1.17	2.29E-02
13	A_24_P358131	chr2	1.15	4.34E-03
14	A_33_P3240295	chr2	1.06	8.75E-03
15	FLJ90757	chr17	1.02	1.17E-02
PCGs - reduced in higher-risk MDS				
1	RAG1	chr11	-3.91	1.49E-02
2	NPY	chr7	-3.71	1.34E-02
3	DUSP26	chr8	-3.39	1.20E-02
4	IGHV1-18	chr14	-3.25	2.77E-02
5	IGHV1-2	chr14	-3.01	3.86E-02
6	STAB2	chr12	-2.88	2.31E-02
7	CD24	chrY	-2.79	2.18E-02
8	LEF1	chr4	-2.73	7.96E-03
9	SPANXB2	chrX	-2.59	2.42E-02
10	AKAP12	chr6	-2.50	1.34E-02
11	CHST15	chr10	-2.46	8.75E-03
12	RBP7	chr1	-2.42	2.62E-02
13	SLAMF7	chr1	-2.40	2.36E-02
14	MME	chr3	-2.37	4.44E-02
15	LOC283454	chr12	-2.35	2.95E-02
16	TFF3	chr21	-2.20	2.30E-02
17	CTTN	chr11	-2.19	3.97E-03
18	IGJ	chr4	-2.15	2.30E-02
19	BTNL9	chr5	-2.12	1.41E-02
20	NR2F2	chr15	-2.10	4.87E-02
21	THC2750292	chr6	-2.06	2.25E-02

22	CLEC4G	chr19	-2.05	4.41E-02
23	NGFR	chr17	-1.95	7.51E-03
24	SLC35D3	chr6	-1.87	2.30E-02
25	CTGF	chr6	-1.86	1.17E-02
26	NUAK1	chr12	-1.86	2.77E-02
27	PDE5A	chr4	-1.86	8.75E-03
28	IL28RA	chr1	-1.76	4.42E-03
29	CLCN4	chrX	-1.73	2.12E-02
30	CLDN5	chr22	-1.73	2.77E-02

SI 21. Gene set enrichment analysis (GSEA) of differentially expressed PCGs in MDS patients with higher- vs. lower-risk IPSS-R. Four selected enrichment plots are shown. Gene sets with p < 0.05 were considered as significantly enriched (the top 10 downregulated gene sets are shown out of the list of the 117 significant sets). NES - normalized enrichment score, FDR – false discovery rate. References: Wang et al. [17], Mikkelsen et al. [22], Meissner et al. [23], Diaz et al. [24], Haddad et al. [18], Piccaluga et al. [25], Jaatinen et al. [11], and Kumar et al. [26].

GSEA: lower- *vs.* higher-risk MDS

Gene sets	NES	p	FDR
Upregulated in lower-risk MDS			
Response to GSK3 inhibitor SB216763 UP (Wang)	1.71	0.019	0.889
MEF HCP with H3K27ME3 (Mikkelsen)	1.67	0.020	0.553
NPC HCP with H3 unmethylated (Meissner)	1.51	0.030	0.994
Chronic myelogenous leukemia UP (Diaz)	1.39	0.045	0.667
Downregulated in higher-risk MDS (the top 10 gene sets)			
B-lymphocyte progenitor (Haddad)	-2.21	0.000	0.000
Angioimmunoblastic lymphoma UP (Piccaluga)	-2.04	0.000	0.008
Hematopoietic stem cell DOWN (Jaatinen)	-1.95	0.000	0.022
Blood vessel morphogenesis (GO)	-1.86	0.000	0.085
Positive regulation of cell differentiation (GO)	-1.84	0.000	0.073
Targets of MLL/AF9 fusion (Kumar)	-1.84	0.001	0.065
Locomotion (GO)	-1.83	0.001	0.066
Regulation of response to external stimulus (GO)	-1.81	0.000	0.076
Immune response (GO)	-1.80	0.000	0.077
Perulation of cell adhesion (GO)	-1.76	0.001	0.083

156
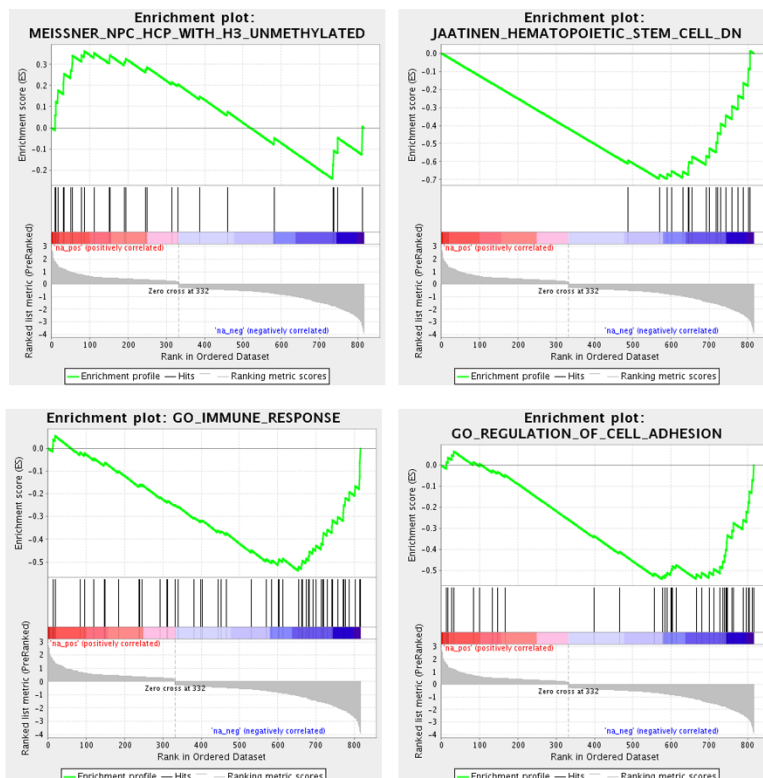
SI 22. Selected modules of the coexpression network designed based on differentially expressed genes between MDS patients with lower- and higher-risk IPSS-R. Gene ontology (GO) terms significantly (p < 0.01) associated with these modules are listed in the corresponding tables. Square – PCG, circle – lncRNA, red – upregulation in higher-risk samples, blue – downregulation in higher-risk samples.


Cell differentiation, adhesion, and migration-related module
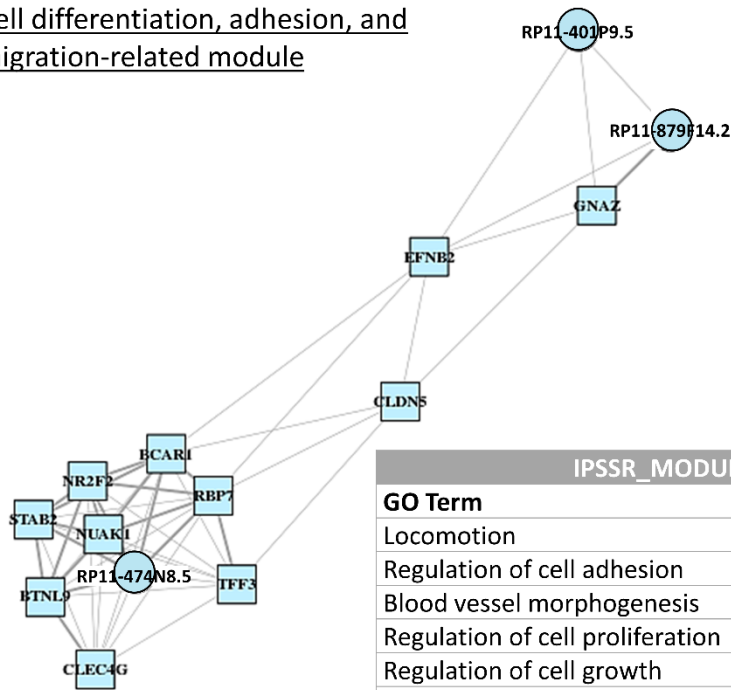
IPSSR_MODULE _1	
GO Term	**P value**
Locomotion	5.88E-05
Regulation of cell adhesion	6.49E-05
Blood vessel morphogenesis	1.72E-04
Regulation of cell proliferation	2.49E-04
Regulation of cell growth	2.93E-04
Positive regulation of cell differentiation	3.99E-04
Golgi cisterna	4.12E-04
Circulatory system development	1.59E-03
Regulation of immune system process	1.59E-03
Angiogenesis	2.41E-03
Substrate-dependent cell migration	4.01E-03
Resonse to growth factor	4.44E-03

Epigenetics-related modules

RUNX1_MODULE _2	
GO term	P value
Chromatin organization	1.07E-03
Immune response	1.77E-03
Chromatin silencing	2.77E-03
Negative regulation of gene expression epigenetic	4.51E-03
Regulation of cell adhesion	7.97E-03
Histone H4 acythylation	9.36E-03
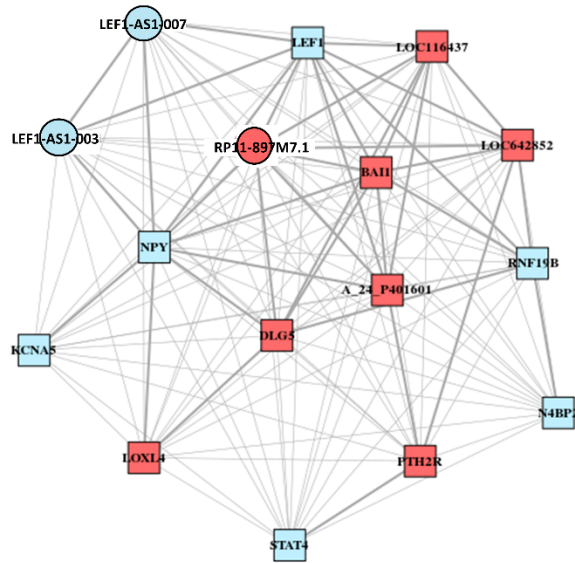
RUNX1_MODULE _3	
GO term	P value
Chromatin organization	6.63E-04
Histone H3K4 trimethylation	7.70E-04
Chromatin silencing	1.07E-03
Negative regulation of gene expression epigenetic	1.76E-03
Immune system process	3.85E-03
Leukocyte homeostasis	6.16E-03
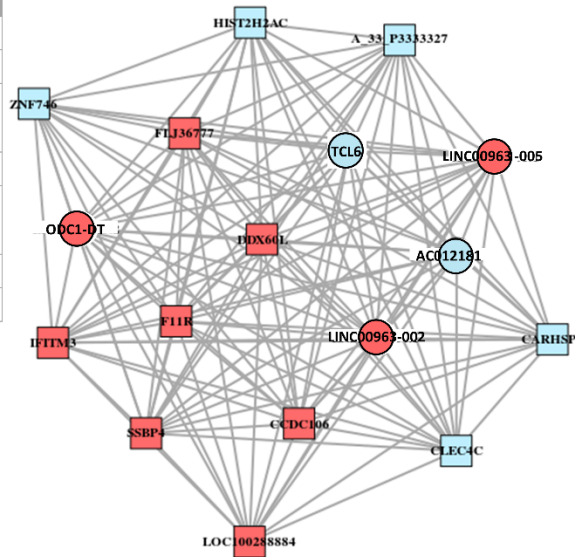
SI 23. Correlations of lncRNA expressions with clinical variables of MDS patients. Spearman correlation coefficients (r) are shown for each pair of variables. (D) discovery cohort, (T) testing cohort. Significant correlations are marked (*p < 0.05, **p < 0.01, ***p < 0.001).

Correlation coefficient (r)	H19 level	WT1-AS level	LEF1-AS1 level	TCL6 level	TP53 mutation
Age	(D) -0.142 (T) 0.073	(D) -0.118 (T) 0.022	(D) -0.134 **(T) -0.288***	(D) 0.060 (T) 0.1415	(D) 0.094 (T) n.a.
Marrow blast count	(D) 0.149 (T) 0.118	**(D) 0.269*** **(T) 0.284****	**(D) -0.383**** **(T) -0.297****	(D) -0.214 **(T) -0.249****	(D) 0.238 (T) n.a.
Hemoglobin level	(D) 0.201 (T) -0.174	(D) 0.006 (T) -0.016	(D) 0.105 (T) 0.149	(D) 0.130 (T) -0.006	**(D) -0.316*** (T) n.a.
Neutrophil count	(D) -0.022 (T) -0.023	(D) -0.176 **(T) -0.309****	(D) -0.087 (T) 0.188	(D) -0.084 (T) -0.098	(D) -0.234 (T) n.a.
Platelet count	(D) -0.163 **(T) -0.254***	(D) -0.180 (T) -0.191	(D) 0.123 (T) 0.176	(D) 0.081 (T) -0.209	(D) -0.214 (T) n.a.
H19 level	-	(D) 0.141 (T) 0.187	(D) -0.102 (T) -0.045	(D) -0.054 (T) -0.112	(D) -0.028 (T) n.a.
WT1-AS level	-	-	(D) -0.034 **(T) -0.322****	(D) -0.058 (T) -0.078	**(D) 0.445**** (T) n.a.
LEF1-AS1 level	-	-	-	**(D) 0.851***** **(T) 0.580*****	(D) -0.130 (T) n.a.
TCL6 level	-	-	-	-	(D) -0.127 (T) n.a.

n.a.- not assessed

SI 24. Correlation of expression levels of WT1 to WT1-AS, LEF1 to LEF1-AS1, and TCL6 to TCL1A/TCL1B.
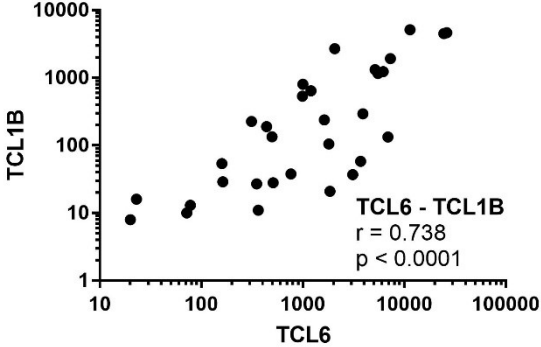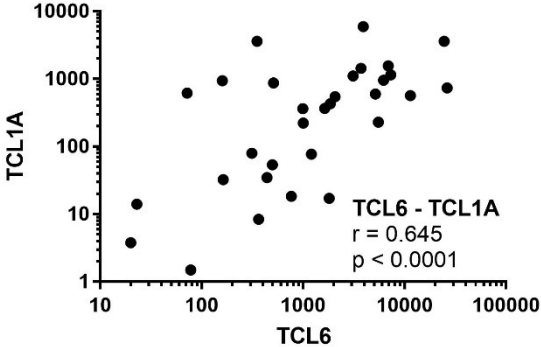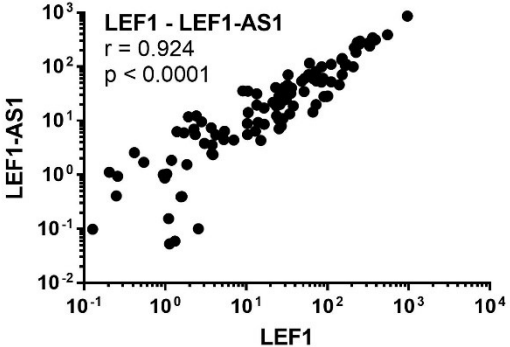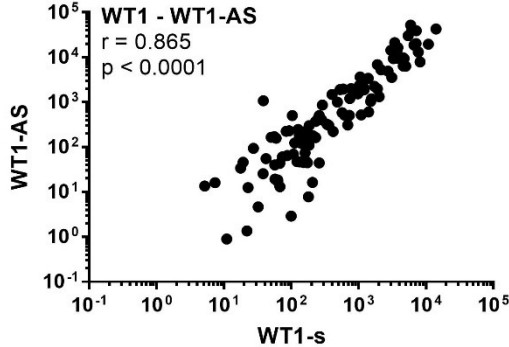
REFERENCES

[1] Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood 2016;127:2391-405.

[2] Greenberg PL, Tuechler H, Schanz J, Sanz G, Garcia-Manero G, Solé F et al. Revised International Prognostic Scoring System for Myelodysplastic Syndromes. Blood 2012;120:2454-65.

[3] Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 2012;22:1775-89.

[4] Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA et al. TM4 microarray software suite. Meth Enzymol 2006;411:134-93.

[5] Xie F, Xiao P, Chen D, Xu L, Zhang B. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. Plant Mol Biol 2012.

[6] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001;25:402-8.

[7] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545-50.

[8] Liu K, Beck D, Thoms JAI, Liu L, Zhao W, Pimanda JE et al. Annotating function to differentially expressed LincRNAs in myelodysplastic syndrome using a network-based method. Bioinformatics 2017;33:2622-30.

[9] Brunet J, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci U S A 2004;101:4164-9.

[10] Wierenga ATJ, Vellenga E, Schuringa JJ. Maximal STAT5-induced proliferation and self-renewal at intermediate STAT5 activity levels. Mol Cell Biol 2008;28:6668-80.

[11] Jaatinen T, Hemmoranta H, Hautaniemi S, Niemi J, Nicorici D, Laine J et al. Global gene expression profile of human cord blood-derived CD133+ cells. Stem Cells 2006;24:631-41.

[12] Ross ME, Mahfouz R, Onciu M, Liu H, Zhou X, Song G et al. Gene expression profiling of pediatric acute myelogenous leukemia. Blood 2004;104:3679-87.

[13] Eppert K, Takenaka K, Lechman ER, Waldron L, Nilsson B, van Galen P et al. Stem cell gene expression programs influence clinical outcome in human leukemia. Nat Med 2011;17:1086-93.

[14] Graham SM, Vass JK, Holyoake TL, Graham GJ. Transcriptional analysis of quiescent and proliferating CD34+ human hemopoietic cells from normal and chronic myeloid leukemia sources. Stem Cells 2007;25:3111-20.

[15] Lim E, Wu D, Pal B, Bouras T, Asselin-Labat M, Vaillant F et al. Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. Breast Cancer Res 2010;12:R21.

[16] Massarweh S, Osborne CK, Creighton CJ, Qin L, Tsimelzon A, Huang S et al. Tamoxifen resistance in breast tumors is driven by growth factor receptor signaling with repression of classic estrogen receptor genomic function. Cancer Res 2008;68:826-33.

[17] Wang Z, Iwasaki M, Ficara F, Lin C, Matheny C, Wong SHK et al. GSK-3 promotes conditional association of CREB and its coactivators with MEIS1 to facilitate HOX-mediated transcription and oncogenesis. Cancer Cell 2010;17:597-608.

[18] Haddad R, Guardiola P, Izac B, Thibault C, Radich J, Delezoide A et al. Molecular characterization of early human T/NK and B-lymphoid progenitor cells in umbilical cord blood. Blood 2004;104:3918-26.

[19] Yoshimura K, Aoki H, Ikeda Y, Fujii K, Akiyama N, Furutani A et al. Regression of abdominal aortic aneurysm by inhibition of c-Jun N-terminal kinase. Nat Med 2005;11:1330-8.

[20] Martens JHA, Brinkman AB, Simmer F, Francoijs K, Nebbioso A, Ferrara F et al. PML-RARalpha/RXR Alters the Epigenetic Landscape in Acute Promyelocytic Leukemia. Cancer Cell 2010;17:173-85.

[21] Heller G, Schmidt WM, Ziegler B, Holzer S, Müllauer L, Bilban M et al. Genome-wide transcriptional response to 5-aza-2'-deoxycytidine and trichostatin a in multiple myeloma cells. Cancer Res 2008;68:44-54.

[22] Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 2007;448:553-60.

[23] Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 2008;454:766-70.

[24] Diaz-Blanco E, Bruns I, Neumann F, Fischer JC, Graef T, Rosskopf M et al. Molecular signature of CD34(+) hematopoietic stem and progenitor cells of patients with CML in chronic phase. Leukemia 2007;21:494-504.

[25] Piccaluga PP, Agostinelli C, Califano A, Carbone A, Fantoni L, Ferrari S et al. Gene expression analysis of angioimmunoblastic lymphoma indicates derivation from T follicular helper cells and vascular endothelial growth factor deregulation. Cancer Res 2007;67:10703-10.

[26] Kumar AR, Hudson WA, Chen W, Nishiuchi R, Yao Q, Kersey JH. Hoxa9 influences the phenotype but not the incidence of Mll-AF9 fusion gene leukemia. Blood 2004;103:1823-8.

RUNX1 **mutations contribute to the progression of MDS due to disruption of antitumor cellular defense: A study on patients with lower-risk MDS**

Supplementary Material

1. **Supplementary Methods**
2. **Supplementary Results**
3. **Supplementary Figures and Tables**

1. SUPPLEMENTARY METHODS

DNA and RNA Isolation

For the preparation of the DNA library, DNA from bone marrow (BM) or, if BM was not available, peripheral blood (PB) was isolated using MagCore according to the manufacturers´ recommendations (RBC Bioscience, New Taipei City, Taiwan). DNA concentration was measured by Qubit 3.0 fluorometer (Life Technologies, Carlsbad, CA, USA) and quality was checked using Nanodrop (Thermo Fisher Scientific, Waltham, MA, USA).

For the preparation of the RNA library, BM CD34+ cells were isolated by magnetic separation on an autoMACS Separator (Miltenyi Biotec, Bergisch Gladbach, Germany). Then, RNA was isolated by acid guanidinium thiocyanate-phenol-chloroform extraction. RNA concentration was measured with a Qubit 3.0 fluorometer (Life Technologies) and RNA quality was checked on an Agilent 2100 bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). Only high-quality RNA with an RNA integrity number (RIN) of at least 7.5 was used.

Targeted gene sequencing

50 ng of DNA were used to prepare the indexed library according to the manufacturer's protocol (TruSight Myeloid Sequencing Panel Kit, Illumina, San Diego, CA, USA). Quantification of the prepared library was done with the KAPA Library Quantification Kit for Illumina sequencing platforms (KAPA Biosystems, Wilmington, MA, USA) also according to the manufacturer's protocol. Sequencing was performed on the MiSeq or NextSeq (Illumina). Analysis was performed with NextGene software (SoftGenetics, State College, PA, USA) and in-house pipeline (described separately in the following paragraph). In paired samples, variants with VAF up to 0.01 were detected if the variant was present in VAF more than 0.05 in one of the paired samples. SIFT and Polyphen were used for the prediction of variant effects. Visualization of NGS results was done with R 4.0.2 software (circlize 0.4.13, ComplexHeatmap 2.8.0).

In-house pipeline for analyzing NGS data from TruSight Myeloid Sequencing Panel:

The quality of raw data obtained from high-throughput sequencing was checked by FastQC version 0.11.8. Reads were then trimmed and filtered using Trimmomatic software version 0.39 and resulting files were quality checked by FastQC again. Cleaned up data from DNA sequencing were then mapped to GRCh19 version of human genome using BWA aligner version 0.7.17. The mapped data was further indexed and sorted using the Samtools suite of tools version 1.10 and the percentage of mapped reads was assessed. Samples with a percentage of mapped reads exceeding 95 % were processed using variant calling software Freebayes version 1.3.1. To discover additional insertions and deletions that span across long sections of the genome, Pindel software version 0.2.5b9 was used. The discovered variants in the form of VCF files were then filtered and annotated using the online interface of Ensembl Variant Effect Predictor (VEP). The annotated variants were formatted in the R software version 4.0.2 and then exported to TSV format.

Sanger sequencing

DNA from CD3+ cells was used in PCR reaction with Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, USA) and the PCR reaction was run according to the manufacturer's protocol. The sequences of the primers are in Table SI 1. 10 µl of the PCR product was run on the 1% TAE gel at 80 V to control the PCR reaction. The PCR product was then cleaned with ExaSap-IT PCR Product Cleanup Reagent (Thermo Fisher Scientific) according to the manufacturer's protocol. The next step was to prepare the sequencing PCR reaction with the BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific) using 0.5 µl of the PCR product. The PCR products were then cleaned with the DyeEx 2.0 Spin Kit (QIAGEN, Venlo, The Netherlands) according to the manufacturer's protocol. Finally, the sample was analyzed on the ABI 3500 (Thermo Fisher Scientific) and the sequences were visualized on Sequencing Analysis Software 5.4 (Thermo Fisher Scientific).

RNA sequencing

100 ng of total RNA from CD34+ cells of 8 LR-MDS patients with *RUNX1* mutation, 29 LR-MDS without *RUNX1* mutation, 20 HR-MDS and 13 healthy controls (SI 3) was ribodepleted with the RiboCop rRNA Depletion Kit (Lexogen, Wien, Austria). Sequencing was performed on HiSeq 2500 or NovaSeq (Illumina). Raw reads in the form of FASTQ files were trimmed and filtered using Trimmomatic 0.39 and their quality was assessed using FastQC 0.11.8. The quantification of gene expression was performed by StringTie2 software 1.3.6. The filtered reads were mapped to human genome GRCh38.p13 using STAR 2.7.2b. The quantification of gene expression was performed by StringTie2 software 1.3.6. For analysis and visualization of expression data, several packages in R software 4.0.2 (e.g. edgeR 3.30.3, pheatmap 1.0.12, ggplot 3.3.2, pcaMethods 1.84.0, ComplexHeatmap 2.8.0) and GraphPad Prism 7 software (GraphPad Software, La Jolla, CA, USA) were used. Databases such as Gene Ontology, KEGG Pathways, Reactome Pathways, and ConsensusPathDB were used for functional enrichment analysis.

Machine Learning

Genes with negligible mutation occurrence (mutated in fewer than 6 subjects) were grouped into one category (REST). The variables were therefore: *ASXL1*, *CUX1*, *DNMT3A*, *EZH2*, *JAK2*, *PHF6*, *RUNX1*, *SETBP1*, *SF3B1*, *SRSF2*, *STAG2*, *TET2*, *TP53*, *U2AF1*, *ZRSR2*, REST. In multivariate Cox regression with stepwise backward feature selection, the Aikake information criterion was used by default (rms R library, fastbw function). It is a heuristic criterion, and its application led to very small models of only one gene. Then, after the cross-validated experiment, the model was adjusted for the optimal number of features according to the D value and the general recommendations for the number of events per variable (EPV) in survival regression models (Peduzzi et al., 1995). In our data, we had 214 subjects, 81 death events; then the EPV should be around 10, which means we should have not worked with more than 8 features. Lasso regression (elastic networks) that worked with L1 norm and minimized the number of features was used. The optimal parameterization/number of features was set in cross-validation again, the model quality was measured with the Harrell's C-index (the concordance index, its value

is between 0 and 1). We got two recommendations, lambda.min (optimum) and lambda.1se (a regularized model near optimum result).

β-galactosidase detection

Cocktail of antibodies: CD3 – Spark Blue 550 (Biolegend, San Diego, CA, USA), CD14 – Alexa Fluor 594 (Biolegend), CD16 – BV650 (Biolegend), CD19 – BV570 (Biolegend), CD34 – eFluor 450 (eBioscience, San Diego, CA, USA), CD45 – Alexa Fluor 647 (Biolegend), CD56 – APC fire 750 (Biolegend), LIVE/DEAD fixable blue dead cell stain kit (Invitrogen, Carlsbad, CA, USA).

2. SUPPLEMENTARY RESULTS

Machine Learning – Cross-Validation

In the cross-validation experiments of SBFS, the maximum D-value was approximately 0.17 for data1 and 0.18 for data2 in OS, respectively. The maximum D-value for both PFS datasets was approximately 0.25. In Table SI 11B, the most significant genes are listed in the number that should be ideal for individual datasets according to the cross-validated value of the D measure. Extending the model with comutational data, the maximal D value increased in OS but decreased in PFS (SI 14C).

In cross-validating EN, the highest C-index was around 0.6 in all analyses (SI 12A) and the number of optimal features is specified within the table of results (SI 12B). Comutational data did not improve the C-index.

Machine Learning – Individual hazard ratio model

We used our data to create a hazard ratio model to count the hazard ratio for individual patients.

The predictions corresponded to logarithmic relative hazards:

$\log(h_k(t)/h_{k'}(t)) = \log(h_0(t)^{(\beta_1 x_{1k} + \beta_2 x_{2k} + \ldots + \beta_p x_{pk})}/\log(h_0(t)^{(\beta_1 x_{1k'} + \beta_2 x_{2k'} + \ldots + \beta_p x_{pk'})}) = \beta_1(x_{1k} - x_{1k'}) + \beta_2(x_{2k} - x_{2k'}) + \ldots + \beta_p(x_{pk} - x_{pk'})$, where x_i denoted the ith covariate (mutations in our case), β_i the ith coefficient (the effect size of the given covariate) and the individual k' represented the baseline (average) individual. Counting the hazard ratios between individuals, $h_0(t)$ became unimportant since it remained

the same for all the individuals. Therefore, we were able to count the relative hazard ratio for individual patients from individual analysis.

3. SUPPLEMENTARY FIGURES AND TABLES

PATIENT CHARACTERISTICS AT THE TIME OF DIAGNOSIS, all patients	
Number of patients	214
Age median (years) (range)	65 (20.8-86.5)
Sex	
Male	107 (50.0%)
Female	107 (50.0%)
Laboratory data	**Median (range)**
BM blasts (%)	2 (0-9.8)
Haemoglobin (g/dL)	10 (5.1-14.9)
ANC (10^9/L)	2 (0.1-9.2)
Platelets (10^9/L)	195 (1.0-1115.0)
Cytogenetics (IPSS)	
Good	188 (87.9%)
Intermediate	22 (10.3%)
Poor	4 (1.9%)
IPSS	
Low	102 (47.7%)
Intermediate I	112 (52.3%)
Intermediate II	0
High	0
IPSS - R	
Very low	45 (21.0%)
Low	119 (55.6%)
Intermediate	46 (21.5%)
High	4 (1.9%)
Very high	0
WHO classification (2016)	
MDS-MLD	113 (52,8%)
MDS-SLD	20 (9,3%)
MDS-del(5q)	37 (17,3%)
MDS-RS	22 (10,3%)
MDS-EB-1	18 (8,4%)
MDS-EB-2	3 (1,4%)
MDS-U	1 (0,5%)

SI 1: Patient characteristics at the time of diagnosis. BM, bone marrow; ANC, absolute neutrophil count.

| Primers for exons 5-7	Sequence 5'-3'	Ta
RUNX1_5F	TCCCTGATGTCTGCATTTGTCC	66
RUNX1_5R	AGACAGACCGAGTTTCTAGGG	
RUNX1_6F	AGCAAAGCCAAAATTCCGGG	67
RUNX1_6R	GGTCCCTGAGTATACCAGCCT	
RUNX1_7F	AGCGAGTCTATGTTGGGGTG	68
RUNX1_7R	AAGGGGAAACCCCAGTTGGT	

SI 2. Primers for PCR and sequencing variants in exons 5-7 of the *RUNX1* gene. Ta, annealing temperature; F, forward primer; R, reverse primer.

Sample	Diagnosis (WHO)	Cyto-genetics (IPSS)	Mutations (VAF %) in total BM DNA	*VAF (%)* of *RUNX1* mutations in cDNA from CD34+
RUNX1-LR-MDS				
V108	MDS-EB-1	good	*RUNX1* (35 and 3 and 4), *SF3B1* (36), *TET2* (9), *IKZF1* (10)	61; 12; 4
V1834	MDS-EB-1	good	*RUNX1* (35), *SF3B1* (37), *EZH2* (7)	46
V1824	MDS-EB-2	good	*RUNX1* (11), *SF3B1* (42), *ASXL1* (25), *ZRSR2* (25), *STAG2* (35)	31
V1708	MDS-EB-2	good	*RUNX1* (44), *SRSF2* (40), *STAG2* (89), *ASXL1* (43)	42
V221	MDS-EB-1	good	*RUNX1* (10), *SETBP1* (4), *STAG2* (12)	23
V2387	MDS-EB-1	good	*RUNX1* (14 and 12), *ASXL1* (30), *STAG2* (14)	17; 19
V1090	MDS-EB-1	good	*RUNX1* (41 and 2), *ASXL1* (21), *GNAS* (50), *PHF6* (43), *EZH2* (39 and 40)	15; 43
V1422	MDS-MLD	good	*RUNX1* (49), *SRSF2* (51), *SETBP1* (48)	53
LR-MDS				
V148	MDS-MLD	good	*U2AF1* (43), *TET2* (3)	
V1664	MDS-MLD	Int	*DNMT3A* (11 and 12)	
V2089	MDS-SLD	good	*U2AF1* (40)	
V2133	MDS-MLD	good	none	
V67	MDS-RS	good	*SF3B1* (46), *DNMT3A* (46)	
V1742	MDS-RS	good	*SF3B1* (48), *DNMT3A* (50), *TET2* (42 and 37), *CUX1* (5)	
V2092	MDS-RS	good	*SF3B1* (29), *TET2* (21)	
V2110	MDS-MLD	good	none	
V2322	MDS-MLD	good	none	
V2248	MDS-MLD	good	none	
V2284	MDS-MLD	good	*U2AF1* (14)	
V2311	MDS-MLD	good	*TP53* (34)	
V1699	MDS-RS	good	*SF3B1* (40), *TET2* (34)	
V1860	MDS-RS	good	none	
V2241	MDS-SLD	good	none	
V888	MDS-RS	good	*SF3B1* (28 and 4)	
V125	MDS-SLD	good	*SF3B1* (27)	
V220	MDS-MLD	int	*SF3B1* (43)	
V2286	MDS-MLD	good	*SF3B1* (38)	
V480	MDS-del(5q)	good	*DNMT3A* (24)	
V883	MDS-RS	good	*SF3B1* (42), *TET2* (19)	
V1528	MDS-del(5q)	good	none	
V1591	MDS-MLD	good	*SF3B1* (10)	
V1957	MDS-EB-1	good	*ASXL1* (29); *PHF6* (82)	
V2147	MDS-MLD	good	*TET2* (30 and 40)	
V630	MDS-MLD	good	none	
V1921	MDS-SLD	good	none	
V2224	MDS-SLD	good	none	
V2179	MDS-SLD	good	*TET2* (32)	
HR-MDS				
V1592	MDS-EB-2	int	*TET2* (45), *RUNX1* (45), *ASXL1* (50), *EZH2* (49), *PHF6* (45)	
V1279	MDS-EB-2	good	*TET2* (19 and 22), *EZH2* (6), *ZRSR2* (68)	
V716	MDS-EB-2	poor	*SF3B1* (18)	
V1874	AML-MRC	N/A	none	
V777	AML-MRC	good	none	
V637	MDS-EB-2	int	none	
V1441	MDS-EB-2	good	*RUNX1* (30), *TET2* (6), *BCOR* (28)	
V1554	AML -MRC	good	*IDH2* (5), *IKZF1* (22), *STAG2* (6)	

V712	MDS-EB-2	good	*SRSF2* (32), *ASXL1* (26), *RUNX1* (30), *BCOR* (6), *STAG2* (6 and 45)
V456	MDS-EB-2	good	none
V1321	MDS-EB-1	good	*SF3B1* (9)
V1456	MDS-MLD	int	*SF3B1* (39), *TET2* (46), *CUX1* (49)
V1884	MDS-EB-2	N/A	*TP53* (36 and 40)
V1297	MDS-EB-2	poor	*TP53* (10)
V1394	MDS-EB-2	N/A	*NRAS* (46), *ETV6* (49), *ASXL1* (38), *STAG2* (90), *PHF6* (96), *GATA2* (43)
V1788	MDS-EB-2	poor	*TP53* (59), *SF3B1* (40)
V839	MDS-EB-2	Int	*DNMT3A* (44), *RUNX1* (26)
V406	MDS-EB-2	int	*SF3B1* (27)
V655	MDS-EB-2	poor	*TP53* (69)
V1857	MDS-EB-2	good	*DNMT3A* (41)

SI 3. The list of patients in the expression study with their cytogenetic and mutation profiles. Thirteen healthy controls are not included. VAF, variant allele frequency; BM, bone marrow; R-LR, *RUNX1*-mutated LR-MDS patients; wtR-LR – LR-MDS without *RUNX1* mutations; HR, higher-risk MDS patients; int, intermediate; N/A, not available.
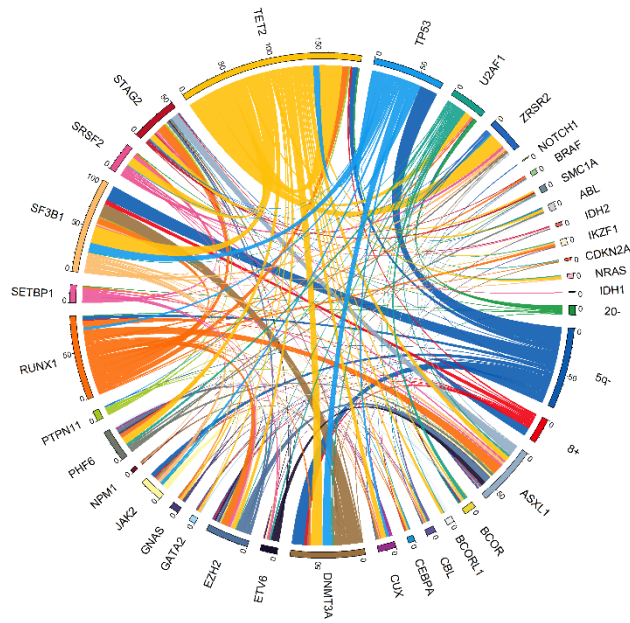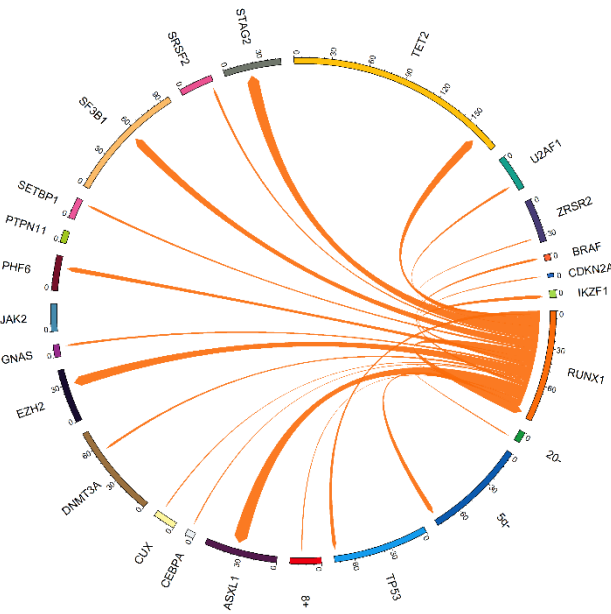
SI 4. Circos plots illustrating pairwise co-occurrences of selected genetic alterations. A) Circos plot of all pairwise co-occurrences identified in the Czech cohort of 214 LR-MDS patients. B) Circos plot of the co-occurrence of molecular aberrations with the mutated *RUNX1* gene. The length of the arch depicts the number of mutations of the first gene comutated with other mutations. The width of the ribbon corresponds to the frequency of co-occurence with the second gene.

A

Univariate analysis	p value	
Variable	OS	PFS
ASXL1	ns	ns
CUX	ns	ns
DNMT3A	0.0286	ns
EZH2	ns	ns
JAK2	ns	ns
PHF6	ns	ns
RUNX1	0.0005	<0.0001
SETBP1	0.0201	0.0225
SF3B1	ns	ns
SRSF2	ns	ns
STAG2	0.0004	0.0019
TET2	ns	ns
TP53	0.0154	0.0487
U2AF1	ns	0.0426
ZRSR2	ns	ns
Male sex	0.0003	<0.0001
Presence of at least 1 mutation	0.0071	0.0016
IPSS	ns	ns
IPSS-R	ns	ns
BM blasts	ns	ns
Platelet counts	0.0015	0.0004
Haemoglobin	ns	ns
ANC	ns	ns
Total number of mutations	0.0001	<0.0001
5q-	ns	0.0348
Age	<0.0001	<0.0001

B

Variable - OS	p value univariate	p value multivariate	HR	95% CI of HR
Male sex	0.0003	0.3125	1.0832	0.9276 to 1.2648
Age	<0.0001	<0.0001	1.0605	1.0344 to 1.0872
Presence of at least 1 mutation	0.0071	0.3926	0.7405	0.3719 to 1.4745
Total number of mutations	0.0001	0.0039	0.9971	0.9951 to 0.9991
Platelet counts	0.0015	0.4022	1.2776	0.7202 to 2.2662
TP53	0.0154	0.0405	2.0931	1.0326 to 4.2424
STAG2	0.0004	0.6950	1.1980	0.4855 to 2.9564
SETBP1	0.0201	0.0649	2.8633	0.9371 to 8.7490
RUNX1	0.0005	0.2680	1.6272	0.6876 to 3.8509
DNMT3A	0.0286	0.0492	1.8803	1.0022 to 3.5280

C

Variable - PFS	p value univariate	p value multivariate	HR	95% CI of HR
Male sex	<0.0001	0.1164	1.5805	0.8926 to 2.7989
Age	<0.0001	0.0002	1.0404	1.0187 to 1.0626
Presence of at least 1 mutation	0.0016	0.9679	0.9877	0.5419 to 1.8002
Total number of mutations	<0.0001	0.3303	1.0724	0.9316 to 1.2343
Platelet counts	0.0004	0.0091	0.9977	0.9959 to 0.9994
5q-	0.0348	0.8576	0.9470	0.5224 to 1.7167
U2AF1	0.0426	0.7427	1.1430	0.5146 to 2.5384
TP53	0.0487	0.0849	1.9462	0.9124 to 4.1516
STAG2	0.0019	0.8639	0.9200	0.3544 to 2.3877
SETBP1	0.0225	0.3584	1.5904	0.5908 to 4.2812
RUNX1	<0.0001	0.0272	2.4782	1.1077 to 5.5443

SI 5. Univariate and multivariate analyses. A) The tested variables and p values for OS and PFS in univariate analysis. Only genes mutated in more than 5 patients were tested. All significant variables of the univariate analysis ($p < 0.05$) were analysed in the multivariate analysis: B) OS, C) PFS. The significant variables in the multivariate analysis ($p < 0.05$) are highlighted. ns, not significant; BM, bone marrow; ANC, absolute neutrophil count; HR, hazard ratio; CI, confidence intervals of the hazard ratios.
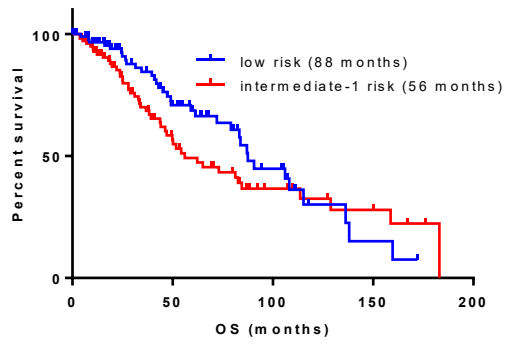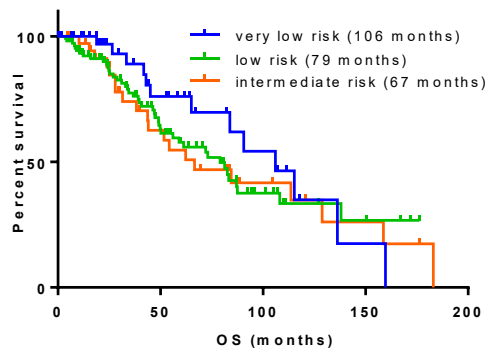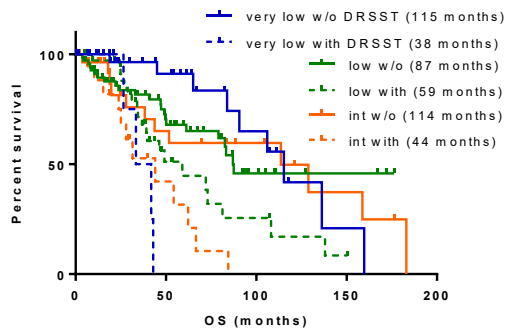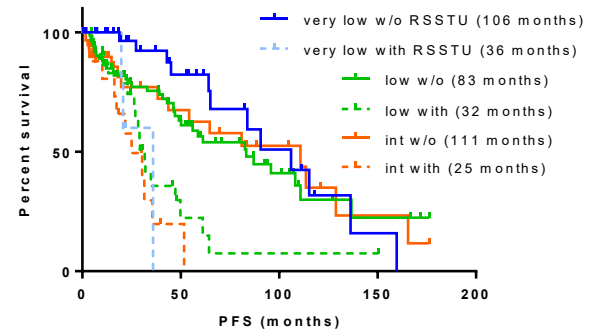
SI 6. Overall survival (OS) of patients according to their IPSS (A) and IPSS-R (B) scores. Neither was significant. Median OS in parentheses.

A

B

SI 7. Integration of the mutational status of significant genes in univariate analysis into the IPSS-R system. A) The graph shows overall survival curves of patients with or without at least one mutated gene of *DNMT3A*, *RUNX1*, *SETBP1*, *STAG2*, and *TP53* (DRSST), p < 0.0001. B) Implementation of the mutational status of *RUNX1*, *SETBP1*, *STAG2*, *TP53*, and *U2AF1* (RSSTU), p < 0.0001. Patients with RSSTU mutations are those with mutations in at least one of these genes. Intermediate, int; without, w/o. Median OS/PFS is in parentheses.
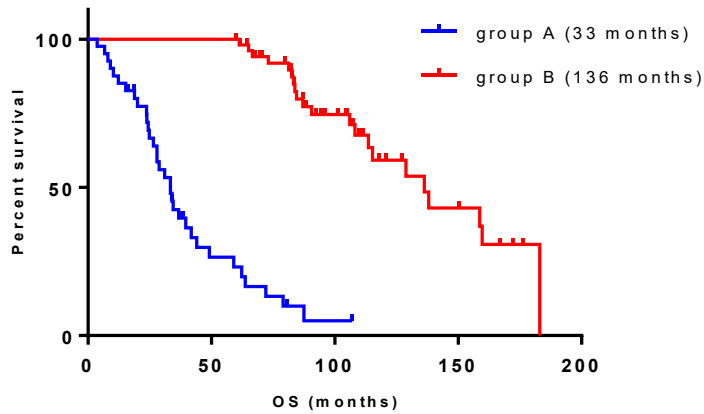
SI 8. OS curves for group A (progressed within 5 years) and group B (did not progress within 5 years), p < 0.0001, median OS in parentheses.

PATIENT CHARACTERISTICS AT THE TIME OF DIAGNOSIS (who progressed within 5 years or were followed at least 5 years)			
	Group A	**Group B**	**P value**
Number of patients	41	53	-
Age median* (years) (range)	68 (28.2-86.5)	58 (20.8-84.4)	0.003
Sex*			0.0197
Male	23 (56.1%)	16 (30.2%)	-
Female	18 (43.9%)	37 (69.8%)	-
Laboratory data	**Median (range)**		
BM blasts (%)	4 (0.4-8.6)	2 (0-7.6)	ns
Hemoglobin (g/dL)	10 (7.5-14.6)	10 (6.1-13.6)	ns
ANC (10^9/L)	2 (0.1-7.5)	2 (0.4-6.9)	ns
Platelets* (10^9/L)	150 (15-406)	284 (25-1115)	0.0003
Cytogenetics (IPSS)			
Good	37 (90.2%)	48 (90.6%)	ns
Intermediate	4 (9.8%)	5 (9.4%)	ns
Poor	0	0	-
IPSS			
Low	15 (36.6%)	26 (49.1%)	ns
Intermediate I	26 (63.4%)	27 (50.9%)	ns
Intermediate II	0	0	-
High	0	0	-
IPSS-R			
Very low	4 (9.8%)	13 (24.5%)	ns
Low	26 (63.4%)	27 (51.0%)	ns
Intermediate	10 (24.4%)	13 (24.5%)	ns
High	1 (2.4%)	0	-
Very high	0	0	-
WHO classification (2016)			
MDS-MLD	21 (51,2%)	19 (35.9%)	ns
MDS-SLD	2 (4,9%)	3 (5.7%)	ns
MDS-del(5q)	10 (24.4%)	20 (37.7%)	ns
MDS-RS	2 (4,9%)	7 (13.2%)	ns
MDS-EB-1	6 (14.6%)	4 (7.5%)	ns
MDS-EB-2	0	0	-
MDS-U	0	0	-

SI 9: Patient characteristics at the time of diagnosis for two groups of patients. Group A – patients who progressed within 5 years, group B – patients who did not progress within 5 years, but were followed at least 5 years. * indicates values significantly different between groups A and B; ns, not significant; BM, bone marrow; ANC, absolute neutrophil count.
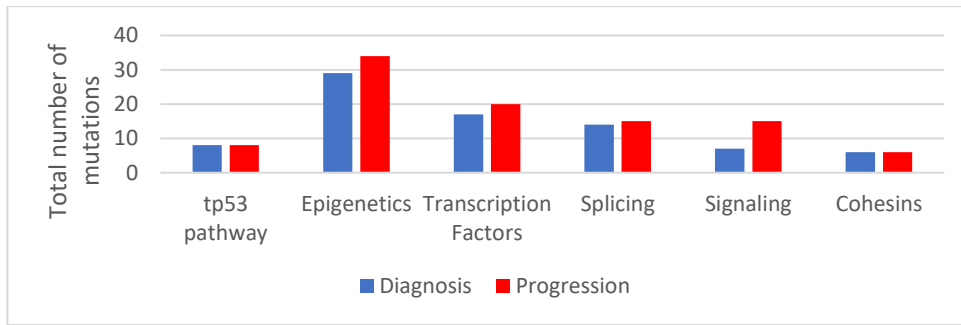
SI 10: Change in the number of mutations by functional categories in paired samples from the time of diagnosis and progression. The blue columns represent the total number of mutations at diagnosis, and the red columns represent the progression.

SI 11: Progression-related changes in VAF of individual samples. Examples of 5 patients sequenced at the time of diagnosis (blue) and in progression (orange). In general, there was an increase in VAF

from diagnosis to progression, but not exclusively. Very often, a novel mutation was found in progression. VAF – variant allele frequency (%).

A

	Coef	S.E.	Wald Z	Pr(>|Z|)
OS_data1:				
STAG2	1.356	0.4172	3.25	0.001156
OS_data2:				
RUNX1	0.9267	0.2518	3.679	0.0002338
PFS_data1:				
RUNX1	1.4463	0.3301	4.38	<0.0001
PFS_data2:				
RUNX1	1.0109	0.2153	4.70	<0.0001

B

|	Coef	S.E.	Wald Z	Pr(>|Z|)
OS_data1:				
ASXL1	0.5253	0.4823	1.09	0.2761
EZH2	-0.8193	0.7089	-1.16	0.2478
TET2	0.4467	0.3035	1.47	0.1412
PHF6	1.0501	0.5309	1.98	0.0479
DNMT3A	0.5144	0.2793	1.84	0.0655
SETBP1	1.6311	0.5795	2.81	0.0049
TP53	0.7984	0.3387	2.36	0.0184
STAG2	0.9723	0.4770	2.04	0.0415
OS_data2:				
DNMT3A	0.3508	0.2178	1.61	0.1073
STAG2	0.6609	0.3664	1.80	0.0713
SETBP1	0.9309	0.4075	2.28	0.0223
TP53	0.3935	0.1802	2.18	0.0290
RUNX1	0.7447	0.2700	2.76	0.0058
PFS_data1				
SETBP1	0.5529	0.4885	1.13	0.2578
TP53	0.5459	0.3296	1.66	0.0977
REST	0.8023	0.2610	3.07	0.0021
RUNX1	1.2577	0.3458	3.64	0.0003
PFS_data2				
SETBP1	0.6228	0.3835	1.62	0.1043
TP53	0.4227	0.1783	2.37	0.0177
REST	0.4304	0.1948	2.21	0.0271
RUNX1	0.9529	0.2198	4.33	<0.0001

SI 12. The results of the stepwise backward feature selection for both datasets of OS and PFS. A) Genes responsible for the shortest OS. B) Optimal number of features responsible for shorter OS and PFS according to cross-validated D-value. Data1, dataset 1, binary mutational data; data2, dataset 2, the number of distinct mutations per gene; coef, coefficient; S.E., standard error; Wald z; Wald test z value; Pr(>|Z|), p value.
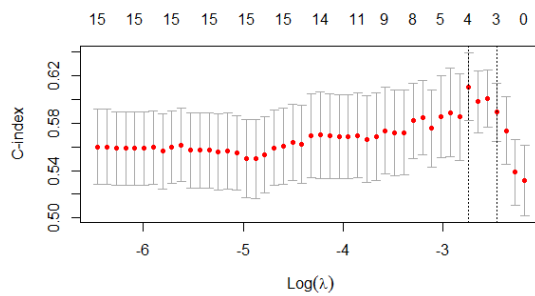
A
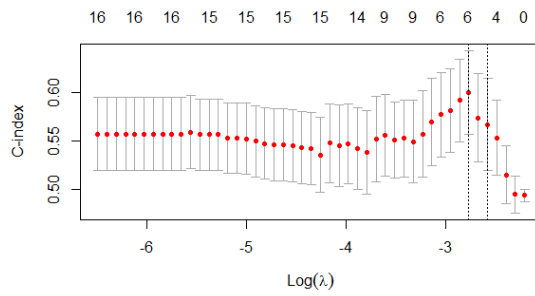
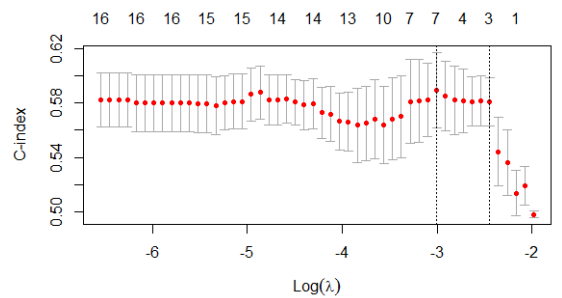| OS data1	PFS data1
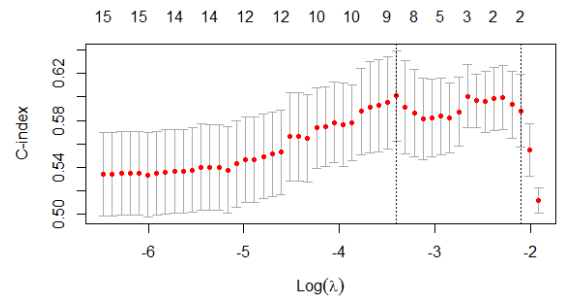
OS data2	PFS data2

B

	Coef lambda.min	Coef lambda.1se
OS_data1 (6 genes with lambda.min, 5 genes with lambda.1se), C-index 0.60:		
DNMT3A	0.1546945	0.06630357
RUNX1	0.4350070	0.36033025
SETBP1	0.1535746	.
STAG2	0.4307018	0.31244418
TP53	0.2333416	0.13650722
REST	0.1072556	0.01689362
OS_data2 (4 genes with lambda.min, 3 genes with lambda.1se), C-index 0.61:		
RUNX1	0.4317455	0.31356625
SETBP1	0.2150566	.
STAG2	0.2770572	0.08685572
TP53	0.2113882	0.11858889
PFS_data1 (9 genes with lambda.min, 2 genes with lambda.1se), C-index 0.6:		
DNMT3A	0.1523	.
PHF6	0.0054	.
RUNX1	1.0552	0.6183
SETBP1	0.3186	.
STAG2	0.1374	.
TET2	0.0498	.
TP53	0.3448	.
U2AF1	0.1662	.
REST	0.5519	0.2047
PFS_data2 (7 genes with lambda.min, 3 genes with lambda.1se), C-index 0.59:		
PHF6	0.0506	.
RUNX1	0.7080	0.5405
SETBP1	0.2790	.
STAG2	0.1228	.
TP53	0.2429	0.0619
U2AF1	0.0968	.
REST	0.2152	0.0670

SI 13. OS and PFS analysis by elastic network approach. A) Cross-validating plots indicating the number of features with the highest C-index for both datasets of OS and PFS. B) Results indicating the most significant genes responsible for shorter OS and PFS. Data1, dataset 1, binary mutational data; data2, dataset 2, the number of distinct mutations per gene; lambda.min,

the minimum mean cross-validated error; lambda.1se, the value when the cross-validated error is within one standard error of the minimum, that gives the most regularized model.
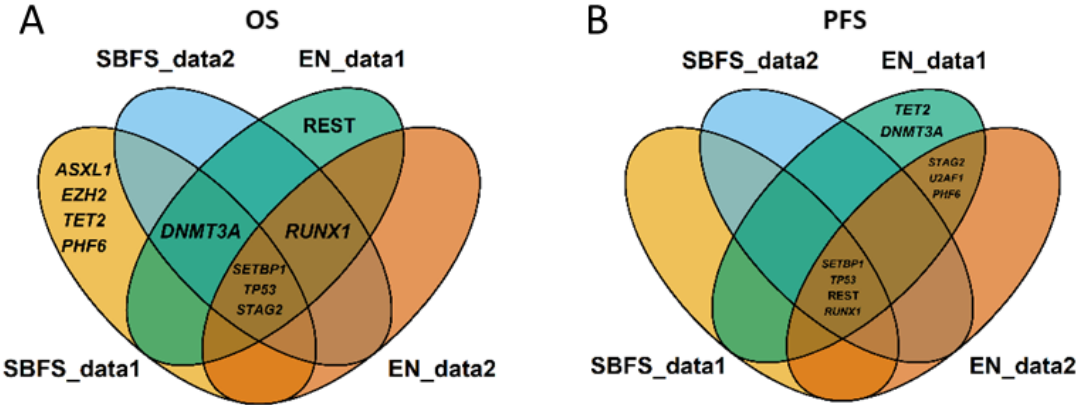
SI 14: Venn diagrams depicting the results of different machine learning methods. Multivariate Cox regression with stepwise backward feature selection and elastic network methods were used on two datasets for A) OS as well as B) PFS. SBFS, multivariate Cox regression with stepwise backward feature selection; EN, elastic networks; data1, dataset 1, binary mutational data; data2, dataset 2, the number of distinct mutations per gene.

A

Interaction	Freq	Fisher exact test p-value
DNMT3A-SF3B1	12	4.415016e-02
SF3B1-TET2	12	1.981671e-02
ASXL1-RUNX1	8	3.054957e-05
ASXL1-STAG2	8	8.356220e-07
RUNX1-STAG2	7	3.355163e-06
ASXL1-EZH2	5	5.515366e-04
TET2-ZRSR2	5	5.218261e-03
EZH2-RUNX1	4	2.735664e-03
SRSF2-TET2	4	9.402812e-02

B

	Coef	S.E.	Wald z	Pr(>|Z|)
OS_data1				
ASXL1	0.7190	0.5148	1.40	0.1625
EZH2	-1.9953	1.1960	-1.67	0.0953
ZRSR2	-1.2313	0.8451	-1.46	0.1451
SRSF2	-0.0308	0.7844	-0.04	0.9687
TET2	0.4733	0.3218	1.47	0.1413
DNMT3A	0.5335	0.2852	1.87	0.0614
SETBP1	1.8196	0.8119	2.24	0.0250
TP53	0.7193	0.3488	2.06	0.0392
RUNX1	0.0983	0.5046	0.19	0.8456
STAG2	0.9471	0.5325	1.78	0.0753
SRSF2*TET2	1.7668	1.4002	1.26	0.2070
EZH2*RUNX1	3.0166	1.5151	1.99	0.0465
PFS_data1				
ASXL1	0.7760	0.5915	1.31	0.1896
RUNX1	1.3547	0.5172	2.62	0.0088
TP53	0.4570	0.3584	1.28	0.2022
EZH2	-0.3875	0.9551	-0.41	0.6850
SETBP1	0.8947	0.6326	1.41	0.1573
U2AF1	0.8584	0.4356	1.97	0.0487
DNMT3A	0.5790	0.2715	2.13	0.0330
TET2	0.1363	0.3145	0.43	0.6646
ZRSR2	-1.0543	1.0765	-0.98	0.3274
STAG2	1.4398	0.6655	2.16	0.0305
ASXL1*RUNX1	-1.4470	1.0875	-1.33	0.1833
ASXL1*EZH2	-1.2378	1.2145	-1.02	0.3081
RUNX1*EZH2	2.6793	1.2394	2.16	0.0306
TET2*ZRSR2	2.3201	1.2924	1.80	0.0726
RUNX1*STAG2	-1.3304	1.0084	-1.32	0.1870

C
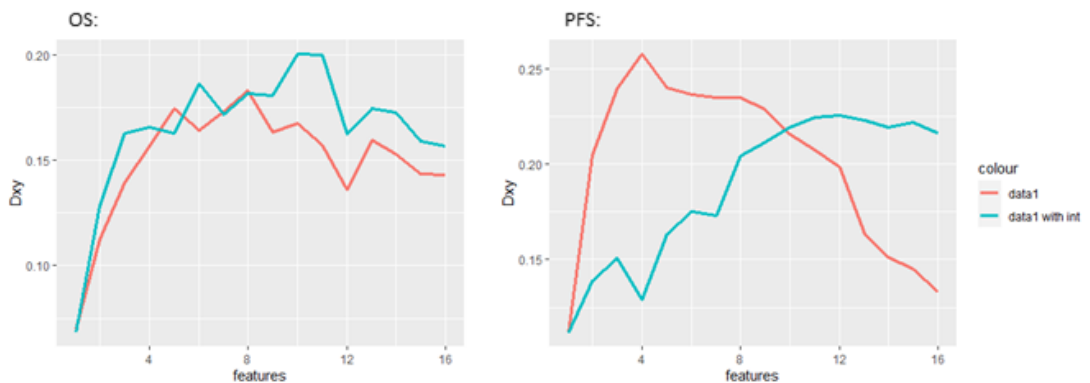
SI 15. Effect of interaction between comutated genes on survival. A) Significant (p<0.05) interactions between comutated genes for dataset 1 according to their effect on survival. Interactions with the REST category were omitted, as well as interactions with frequency less than 4. Freq, frequency of interactions in the cohort. B) Multivariate Cox regression with stepwise backward feature selection (SBFS) model counting with the effect of single mutations and comutations. C) Cross-validation plots of the D value depicting the difference between the SBFS model with and

181

without interactions for OS and PFS. Data1, dataset 1, binary mutational data; coef, coefficient; S.E., standard error; Wald z; Wald test z value; Pr(>|Z|), p value.
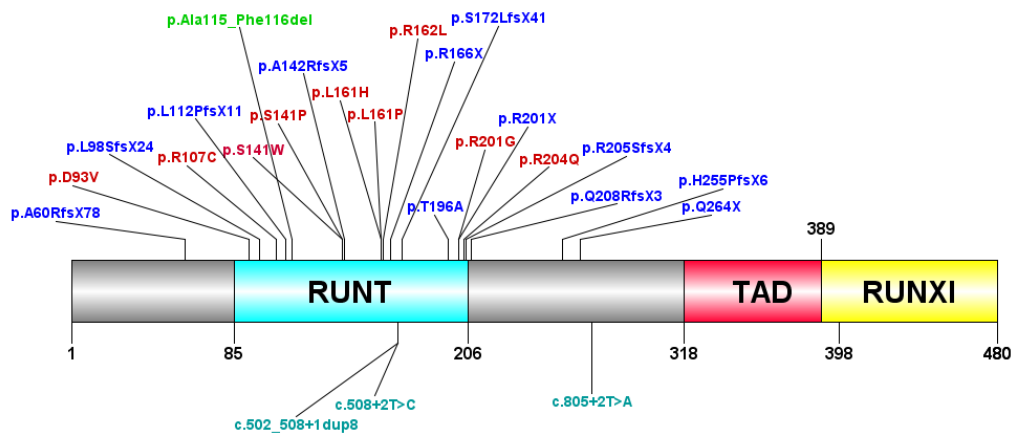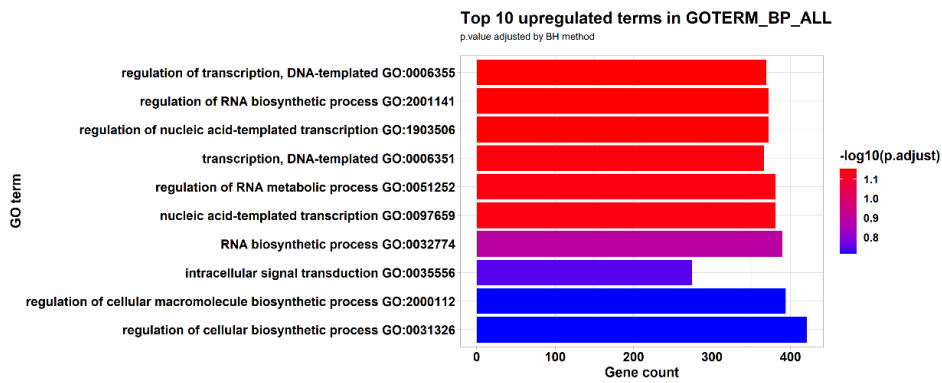

SI 16. Schematic illustration of the *RUNX1* gene and the mutations identified in the cohort of 214 Czech LR-MDS patients. Most mutations lied in the Runt homology domain (RUNT); no mutation was located in two others domains - the transcriptional activation domain (TAD) and the Runx1 inhibition domain (RUNXI). The illustration was created using DOG 2.0 (Yao and Xue, 2009). The mutations are distinguished by colors: red – missense, blue – frameshift and stop gain, green – in-frame deletion, teal – splice region.
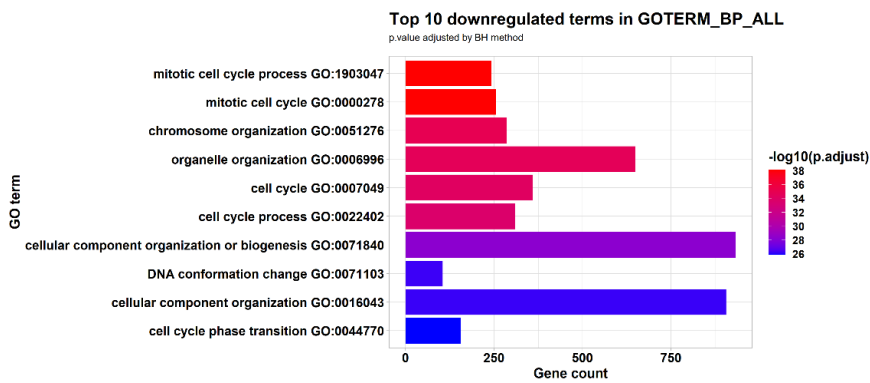
182

PATIENT CHARACTERISTICS AT THE TIME OF DIAGNOSIS			
	RUNX1 mutated	*RUNX1* wildtype	p value
Number of patients	17	197	-
Age median (years) (range)	66 (28-78)	64 (21-87)	ns
Sex			
Male	10 (58.8%)	97 (49.2%)	ns
Female	7 (41.2%)	100 (50.8%)	
Laboratory data	**Median (range)**		
BM blasts* (%)	5.0 (0.8-9.8)	1.8 (0.0-8.8)	<0,001
Hemoglobin (g/dL)	9.5 (7.8-12.8)	9.8 (5.1-14.9)	ns
ANC (10^9/L)	1.6 (0.6-7.5)	1.8 (0.1-9.2)	ns
Platelets* (10^9/L)	103 (15-313)	202.5 (1-1115)	0.010
Cytogenetics (IPSS)			
Good	16 (94.1%)	172 (87.3%)	ns
Intermediate	1 (5.9%)	21 (10.7%)	
Poor	0	4 (2.0%)	
IPSS*			
Low	3 (17.6%)	99 (50.3%)	0.011
Intermediate I	14 (82.4%)	98 (49.7%)	
Intermediate II	0	0	-
High	0	0	-
IPSS-R*			
Very low	1 (5.9%)	44 (22.3%)	0.004
Low	6 (35.3%)	113 (57.4%)	
Intermediate	9 (52.9%)	37 (18.8%)	
High	1 (5.9%)	3 (1.5%)	
Very high	0	0	-
WHO classification (2016)			
MDS-MLD	5 (29.4%)	108 (54.8%)	<0,001
MDS-SLD	0	20 (10.2%)	
MDS-del(5q)	2 (11.8%)	35 (17.8%)	
MDS-RS	1 (5.9%)	21 (10.7%)	
MDS-EB-1	7 (41.2%)	11 (5.6%)	
MDS-EB-2	2 (11.8%)	1 (0.5%)	
MDS-U	0	1 (0.5%)	
Mutation data			
No. of mutations	4 (2-7)	1 (0-9)	<0,001

SI 17. Baseline characteristics of lower-risk MDS patients with and without *RUNX1* mutations. BM, bone marrow; ANC, absolute neutrophil count; ns, not significant; * indicates significantly different values between groups.
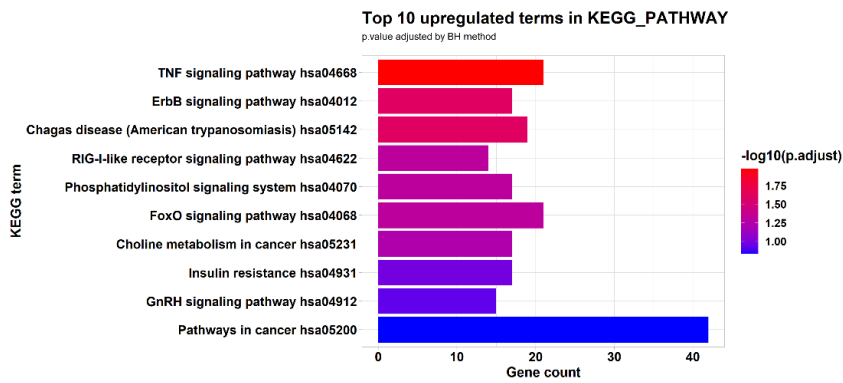
A

Top 10 upregulated terms in GOTERM_BP_ALL
p.value adjusted by BH method

B

Top 10 downregulated terms in GOTERM_BP_ALL
p.value adjusted by BH method

C

Top 10 upregulated terms in KEGG_PATHWAY
p.value adjusted by BH method

SI 18. Top 10 up- and downregulated terms in GO biological processes (A-B) and top 10 upregulated KEGG pathways (C) for mutR-LR compared to wtR-LR by p value. Colors indicate the level of significance.

% of positive cells	4 wtR-LR	3 mutR-LR
Minimum	0	0
25% Percentile	0.25	0
Median	6.5	0
75% Percentile	12.75	2
Maximum	13	2
Mean	6.5	0.6667
Std. Deviation	6.952	1.155
Std. Error of Mean	3.476	0.6667
Lower 95% CI of mean	-4.563	-2.202
Upper 95% CI of mean	17.56	3.535

SI 19: Quantification of the percentage of cells expressing γH2AX in 4 wtR-LR and 3 mutR-LR patients. The γH2AX staining was evaluated in three to five fields of view of individual samples representing different zones of stained sections.
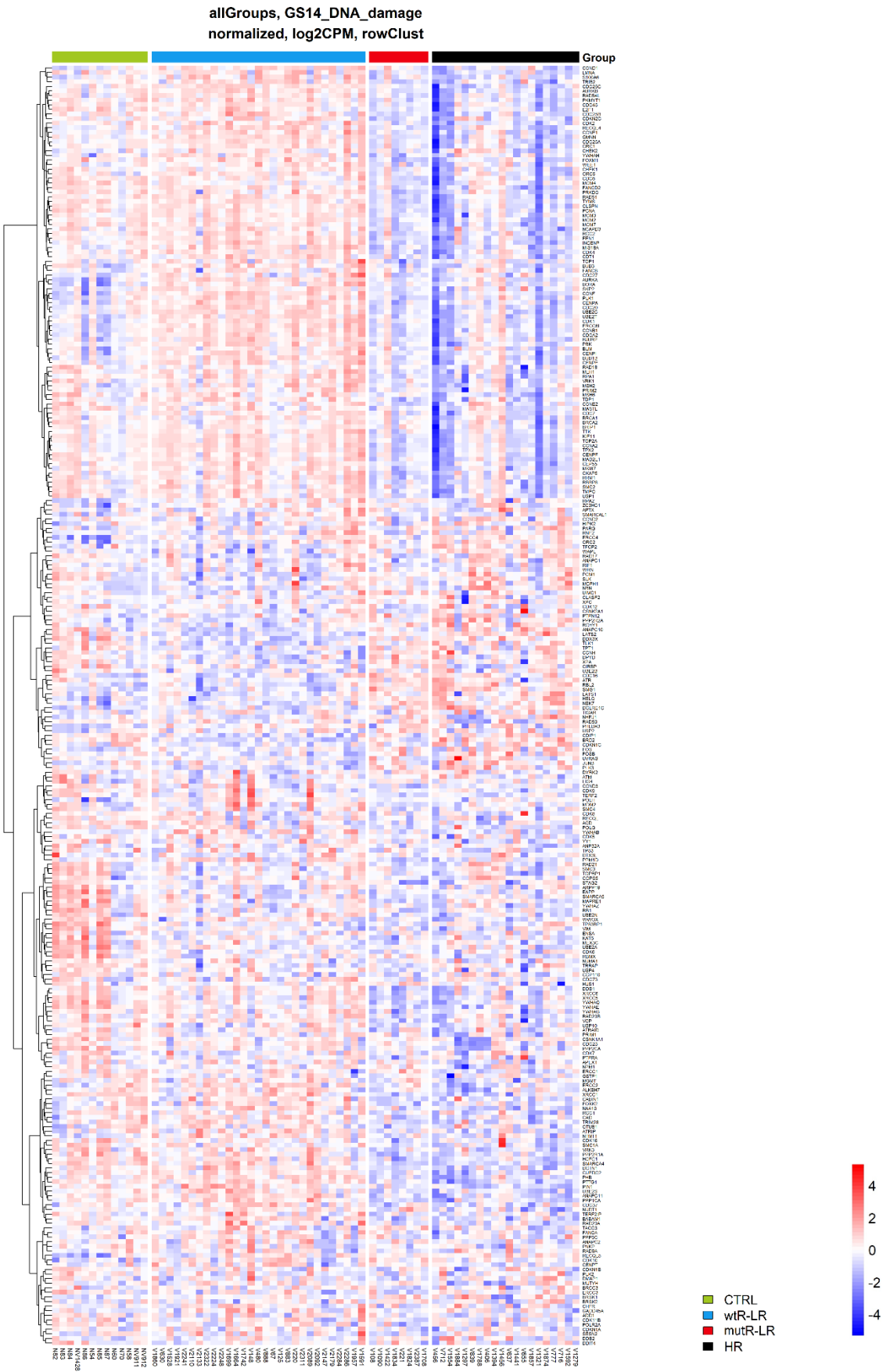
A

allGroups, GS5_cellSenes_KEGG
normalized, log2CPM, rowClust

B

allGroups, GS7_SASP_Reactome
normalized, log2CPM, rowClust

C

allGroups, GS14_DNA_damage
normalized, log2CPM, rowClust

188
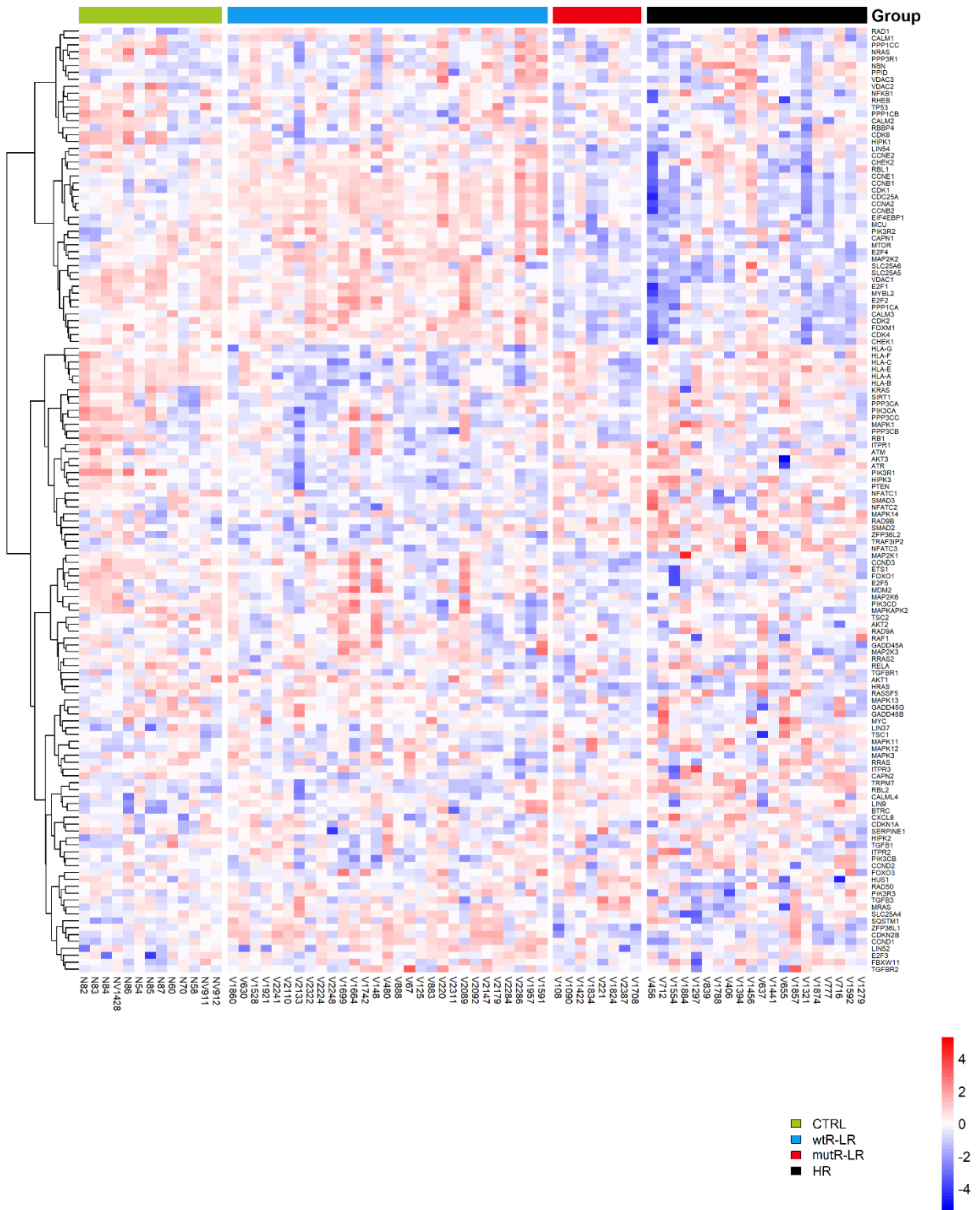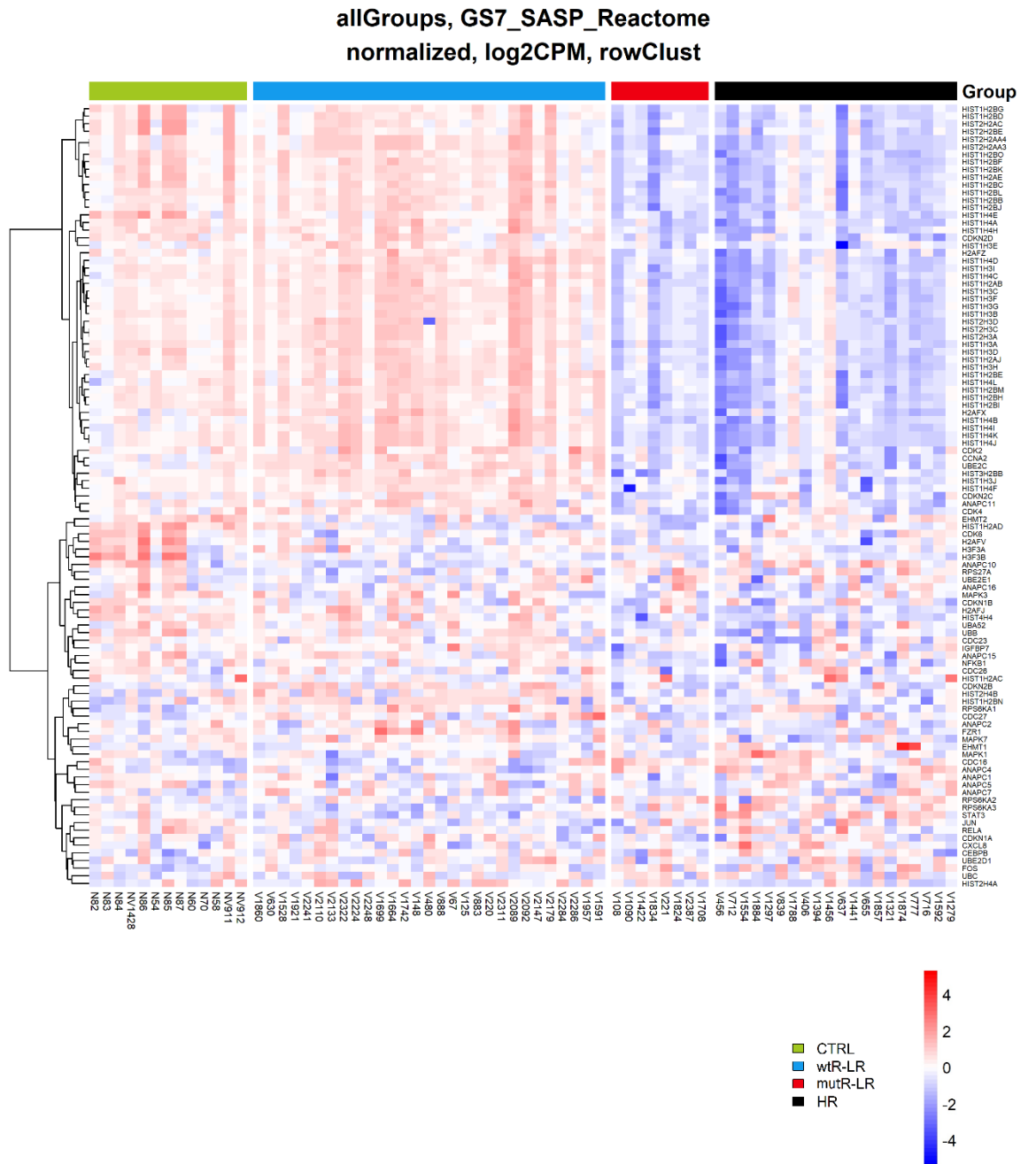
SI 20: Dysregulated expression profiles in selected GSEA pathways. Heatmaps show the expression profiles between CD34+ cells from healthy controls (CTRL), wtR-LR, mutR-LR and HR patients in A) Cellular senescence (KEGG), B) SASP (Reactome), C) DNA damage. The red color indicates upregulation, blue color downregulation of gene expression, and the color intensity indicates the level of differential expression.
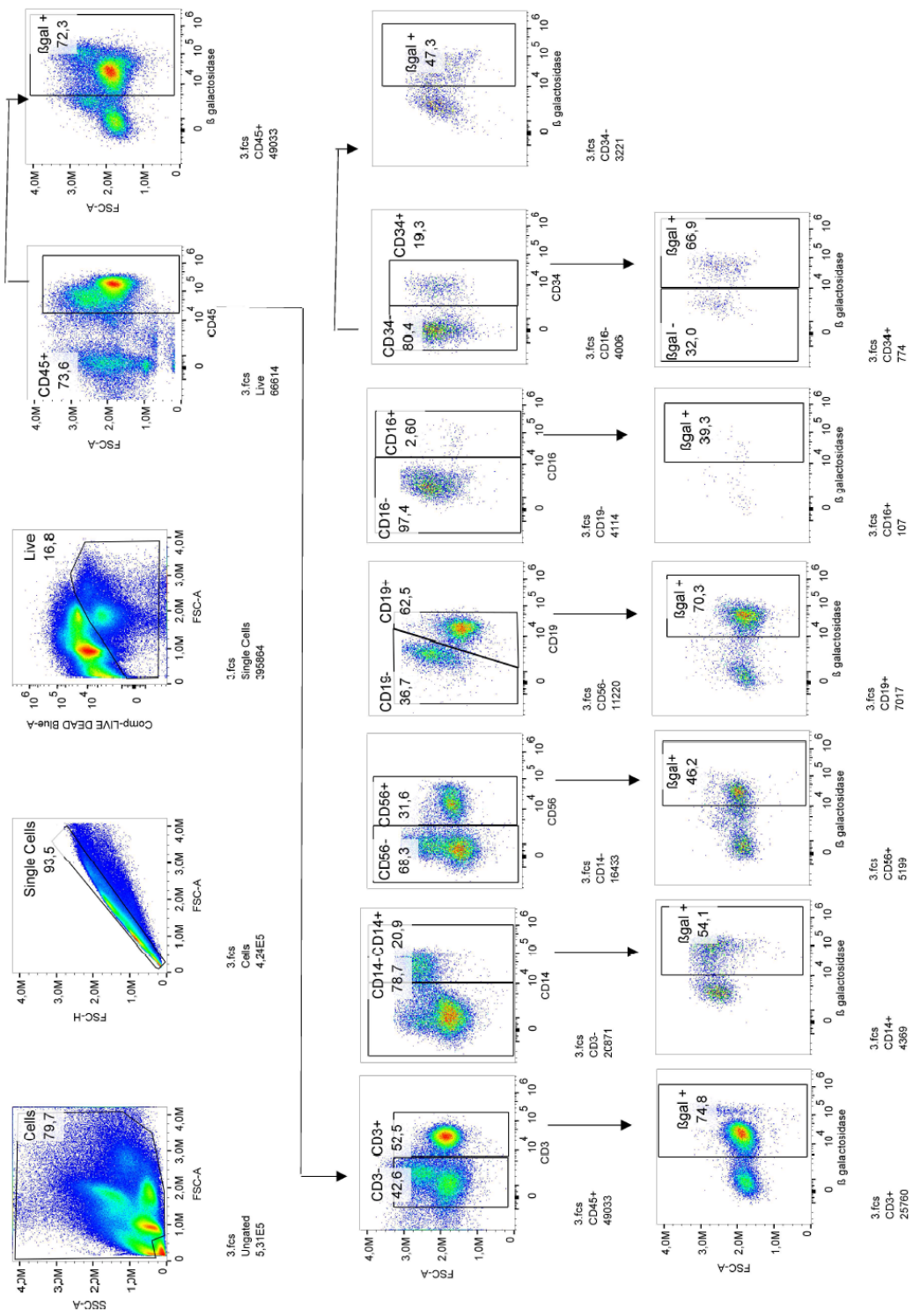
SI 21. Representative example of a gating strategy. The numbers in rectangles indicate the percentage of gated cells.

III.   The list of publications not included in this thesis

Hrustincova, A.; Krejcik, Z.; Kundrat, D.; Szikszai, K.; Belickova, M.; Pecherkova, P.; Klema, J.; Vesela, J.; **Hruba, M**.; Cermak, J.; Hrdinova, T.; Krijt, M.; Valka, J.; Jonasova, A.; Dostalova Merkerova, M. **Circulating Small Noncoding RNAs Have Specific Expression Patterns in Plasma and Extracellular Vesicles in Myelodysplastic Syndromes and Are Predictive of Patient Outcome.** Cells 2020, 9, 794. https://doi.org/10.3390/cells9040794

IF$_{2020}$ = 6.600

Koralkova, P.; Belickova, M.; Kundrat, D.; Dostalova Merkerova, M.; Krejcik, Z.; Szikszai, K.; **Kaisrlikova, M**.; Vesela, J.; Vyhlidalova, P.; Stetka, J.; Hlavackova, A.; Suttnar, J.; Flodr, P.; Stritesky, J.; Jonasova, A.; Cermak, J.; Divoky, V. **Low Plasma Citrate Levels and Specific Transcriptional Signatures Associated with Quiescence of CD34+ Progenitors Predict Azacitidine Therapy Failure in MDS/AML Patients.** Cancers 2021, 13, 2161. https://doi.org/10.3390/cancers13092161

IF$_{2021}$ = 6.575

Votavova, H.; Urbanova, Z.; Kundrat, D.; Dostalova Merkerova, M.; Vostry, M.; **Hruba, M.**; Cermak, J.; Belickova, M. **Modulation of the Immune Response by Deferasirox in Myelodysplastic Syndrome Patients.** Pharmaceuticals 2021, 14, 41. https://doi.org/10.3390/ph14010041

IF$_{2021}$ = 5.215

Hrubá M. **Význam sekvenování nové generace u MDS**. Myelodysplastic Syndrome News, 2020, vol. 8, s. 16-23.

A peer-reviewed journal without impact factor