Supplementary Appendix

Supplement to: Liu Y-C, Kwon J, Fabiani E, et al. Demethylation and up-regulation of an oncogene after hypomethylating therapy. N Engl J Med 2022;386:1998-2010. DOI: 10.1056/NEJMoa2119771

This appendix has been provided by the authors to give readers additional information about the work.

Table of Contents

Section 1: Methods.....4

Figure S1: SALL4 expression in studied cohorts…….....10

Figure S2: Survival based on the change of SALL4 expression in both cohorts.....11

Figure S3: Distribution, frequency, and variant allele fraction of somatic mutations in both cohorts 12

Figure S4: Demethylation of a critical CpG island and SALL4 expression in leukemic cells treated with HMA

……..13

Table S1: Prognostic factors of cohort 1 MDS patients in OS…….....14

Table S2: Prognostic factors of cohort 2 MDS patients in OS…….....15

Supplementary Appendix References……..…….....17

Section 1: Methods

Cell culture and treatment protocol

The K562 and HL-60 human leukemia cell lines were purchased from ATCC and cultured in RPMI 1640 supplemented with 10% fetal bovine serum and antibiotics (penicillin-streptomycin 100U/100 μg/mL) and maintained in an incubator with 5% CO2 atmosphere at 37℃. Based on previous publications[1,2],100nM, 250nM, and 500nM 5-aza-2'-deoxycytidine (decitabine, DAC) daily was used to treat K562 and HL-60 cells for five days. Before adding new DAC or DMSO treatment, cells were washed with ice-cold PBS at 24-hour intervals. Cells were collected on day 5 for measuring SALL4 levels using droplet digital polymerase chain reaction (ddPCR) for RNA and western blot analysis for protein, as well as methylation studies.

CRISPR-DNMT1-interacting RNA (CRISPR-DiR): In vitro generation of sgRNA transcripts

Approximately 1.4kb of the genomic fragment spanning the SALL4 5' UTR-exon 1-intron 1 regionwas PCR amplified (Zymo Research) and cloned into the pGEM-T Easy vector. The vector was linearized with BamH1 restriction enzyme (New England Biolabs). SALL4-targeting sgRNA candidates were transcribed with HiScribe™ Quick T7 High Yield RNA Synthesis Kit (New England Biolabs) following the manufacturer's instructions. The sgRNA target sequences within the SALL4 locus is 5'-CCGGGCGAGCAGCAGCCGCA-3', targeting the opposite strand and using CGG as the PAM sequence.

CRISPR-DiR: In vitro cleavage and selection of sgRNA transcripts

An in vitro cleavage assay was performed using purified Cas9 nuclease from *S. pyogenes* (New England Biolabs) in order to select SALL4-specific sgRNAs among a number of candidates. The experiment was

performed according to the manufacturer's protocol. The sgRNAs were denatured at 95^0C for 3 minutes, then Cas9 protein and sgRNAs were incubated for 10 minutes at 25^0C to form a complex. Lastly, a linearized SALL4 DNA fragment was added to the mixture and the entire reaction was incubated at 37℃ for 1 hour. The reaction mixture was composed of purified Cas9 protein, an individual sgRNA, and a linearized SALL4 genomic fragment in a ratio of 10: 10: 1. 1 ul of Proteinase K was added to each sample after the cleavage reaction, and it was then incubated at room temperature for 10 minutes. The result was analyzed with a 1% agarose gel.

CRISPR-DiR: Lentiviral transduction of DiR-SALL4 and dCas9

Lentiviruses expressing dCas9 or sgRNA were packaged in 293T cells with the plasmids psPAX2 and pMD2.G. TransIT-LT1 Transfection Reagent (Mirus) was used for transfection into 293T cells. Virus was collected at 48 hours and 72 hours post-transfection. The collected virus was filtered through 0.45 μm microfilters and stored at -80 ℃. Transduction of SNU-387 cells was performed by mixing virus and 4 μg/mL polybrene (Santa Cruz) together to add to the cells seeded in T75 flasks 24 hours prior to the transduction. 24 hours after the transduction, the medium was replenished with normal RPMI culture medium. Transduction efficiency was determined by GFP (for sgRNA) or mCherry (dCas9) expression by FACS analysis, and the positive cells were sorted by a FACS Aria machine (BD Biosciences).

Real-time quantitative reverse transcriptase polymerase chain reaction (qRT-PCR)

Total RNA was extracted with Trizol reagent (Invitrogen) according to the manufacturer's instructions and treated with DNase. The RNA concentration was measured with ultraviolet spectrophotometry. Reverse

transcription and PCR were performed using the iScript One-Step RT-PCR Kit with SYBR Green (Bio-Rad, Hercules, CA, USA; catalog no. 170-8893). Triplicate reactions were run for each gene. The expression level was normalized to glyceraldehyde 3-phosphate dehydrogenase (GAPDH). For each sample, an amplification plot and corresponding dissociation curves were examined. Relative quantification analysis was performed using the comparative CT method ($2^{-\Delta\Delta CT}$). The formula used to determine fold change is as follows: $2^{-\Delta\Delta CT} = 2^{-[\text{ t4 CT (SALL4 after HMA) – t4 CT (GAPDH)] –[t0 CT (SALL4 at diagnosis)–t0 CT (GAPDH)]}}$ were used to define $SALL4^{up}$ or $SALL4^{down}$, respectively. The sequences of primers for genes tested as follows: *SALL4* (*SALL4* exon3/4 span), forward primer 5′-AAGGCAACTTAAAGGTTCACTACA-3′, reverse primer 5′-GATGGCCAACTTCCTTCCA-3′; *GAPDH*, forward primer 5′-GAAGGTGAAGGTCGGAGTCAAC-3′, reverse primer 5′-TGGAAGATGGTGATGGGATTTC-3′.

Western blot

Western blot was performed according to standard protocols. The following antibodies were used for western blotting: SALL4 (ab29112, Abcam), and β-actin. The dilution ratio of SALL4 antibody was 1:1000.

Droplet digital polymerase chain reaction (ddPCR)

Reactions for the ddPCR were prepared by harvesting 100,000 cells on each day for RNA extraction and cDNA preparation. The reaction mixture was prepared with the 2x ddPCR supermix for probes (Biorad, Cat #186-3026), 10-fold diluted cDNA, nuclease-free water, and forward and reverse primers. Once the reaction mixture was ready, it was loaded onto the DG8 cartridge for the QX200 Automated Droplet Generator (Biorad, catalog no.186-4003). Thermal cycling was performed using the Biorad C1000 Touch Thermal Cycler with the

following cycle conditions: 95℃ for 10minutes, 94℃ for 30 seconds (40 cycles), 60℃ for 2 minutes (40 cycles), 98℃ for 10 minutes, and 4℃ hold. The reaction plate was loaded into the QX200 Droplet Reader (Biorad, Cat#186-4003) for gene expression analysis. To detect SALL4 and β-actin, the following primers were used: SALL4, forward primer 5'-AAGCTTACATCTCCGCGGTGGATGT-3', reverse primer 5'-GGATCCTGCTCCGACACTTGTGCTTG-3'; β-actin, forward primer 5'-GGGAGATACCATGATCACGAAGGT-3', reverse primer 5'-CCACAAATTATGCAGTCGAGTTTCCC-3'.

Next generation sequencing (NGS) pipeline and validation methods

For cohort 1 patients:

DNA samples were extracted using the QIAamp DNA Mini Kit (Qiagen AG, Milan, Italy), in accordance with the manufacturer's instructions. NGS data about patients used in this study were extrapolated from our main cohort of genetically screened MDS patients. DNA samples collected at the time of diagnosis were processed and analyzed as previously reported[3]. In brief, NGS screening for common somatic mutations in thirty genes known to be involved in MDS pathogenesis was performed according to the commercial Myeloid Solution by SOPHiA GENETICS (SOPHiA GENETICS, Saint-Sulpice, Switzerland) on a MiniSeq® sequencing platform (Illumina, San Diego, California). The NGS analysis was performed on generated FASTQ sequencing files using the SOPHiA DDM® platform that allows for detection, annotation, and pre-classification of genomic mutations (SNVs and Indels) through its SOPHiA™ artificial intelligence. Reads were aligned to the human reference genome (hg19 assembly). Only mutations with a VAF ≥ 1% (variant allele frequency),

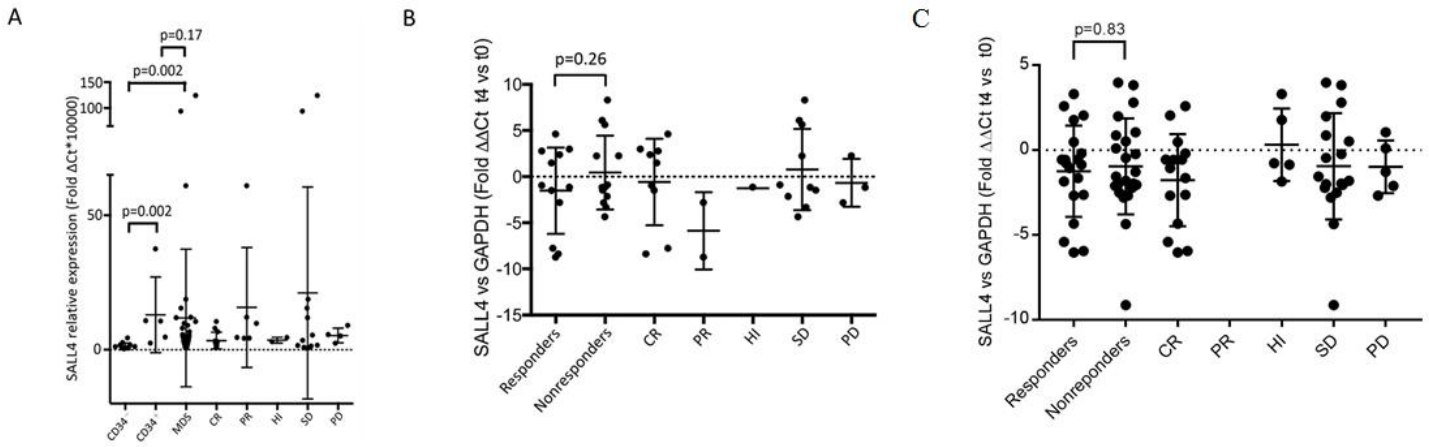
threshold coverage ≥ 1000x, and identified as highly or potentially pathogenic by the SOPHiA DDM® platform were considered for all subsequent steps of the analysis. Single nucleotide polymorphisms (SNP), variants localized in the intronic and UTR regions, as well as synonymous variants were also excluded from the analysis. Targeted-NGS sequencing data are stored at https://www.sophiagenetics.com (SOPHiA DDM platform) and can be extracted using the Sophia-DDM-v4 password-protected software. Raw data will be provided to researchers upon request. Validation of identified variants was performed using pyrosequencing technology (VAF ≥ 10%) and Sanger sequencing (VAF ≥ 20%). Pyrosequencing reagents (PyroMark Gold Q96, QiagenSrl, Milan, Italy), instrumentation, and software used for pyrosequencing analysis were as recommended by the manufacturers (PyroMark Q96 ID, DiatechPharmacogenetics,Jesi, Italy, PyroMark Assay Design and PyroMark Q24 version 2.0.6). Sanger sequencing reagents (BigDye Terminator v.3.1 cycle sequencing kit, Applied Biosystems/Life Technologies, Milan, Italy) and instrumentation were used for Sanger sequencing (ABI PRISM 3100; Applied Biosystems/Life Technologies, Milan, Italy). All primers were homemade designed, and all mutations were confirmed and quantified in independent experiments.

For cohort 2 patients:

DNA was extracted from bone marrow samples collected at the time of diagnosis and after treatment of hypomethylating drugs according to manufacturer's protocol. NGS screening for 114 myeloid neoplasm-related genes were performed on an Ion Torrent semiconductor platform and results were mapped to NCBI hg19 Ref Seq with a mean of >97% coverage of the targeted regions at an average depth of 1000X.All putative mutations were compared against multiple databases (e.g.1000genomes, COSMIC, PolyPhen, SIFT).Single nucleotide

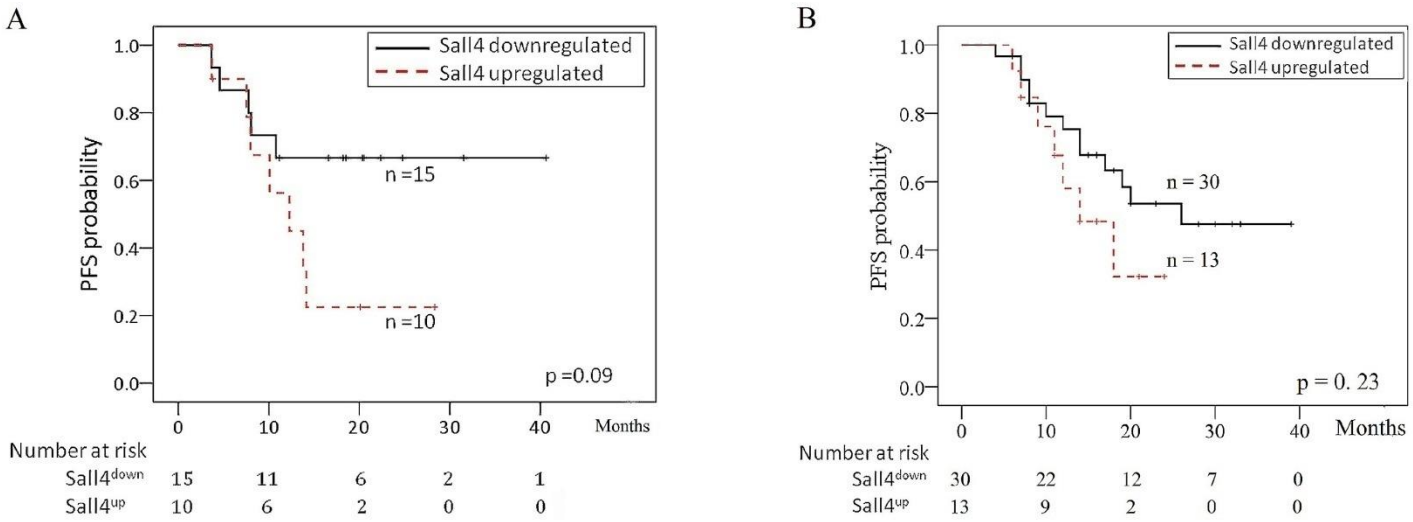
polymorphisms (SNP), variants localized in the intronic and UTR regions, as well as synonymous variants were excluded from the analysis.

Figure S1 (A-C): SALL4 expression in studied cohorts



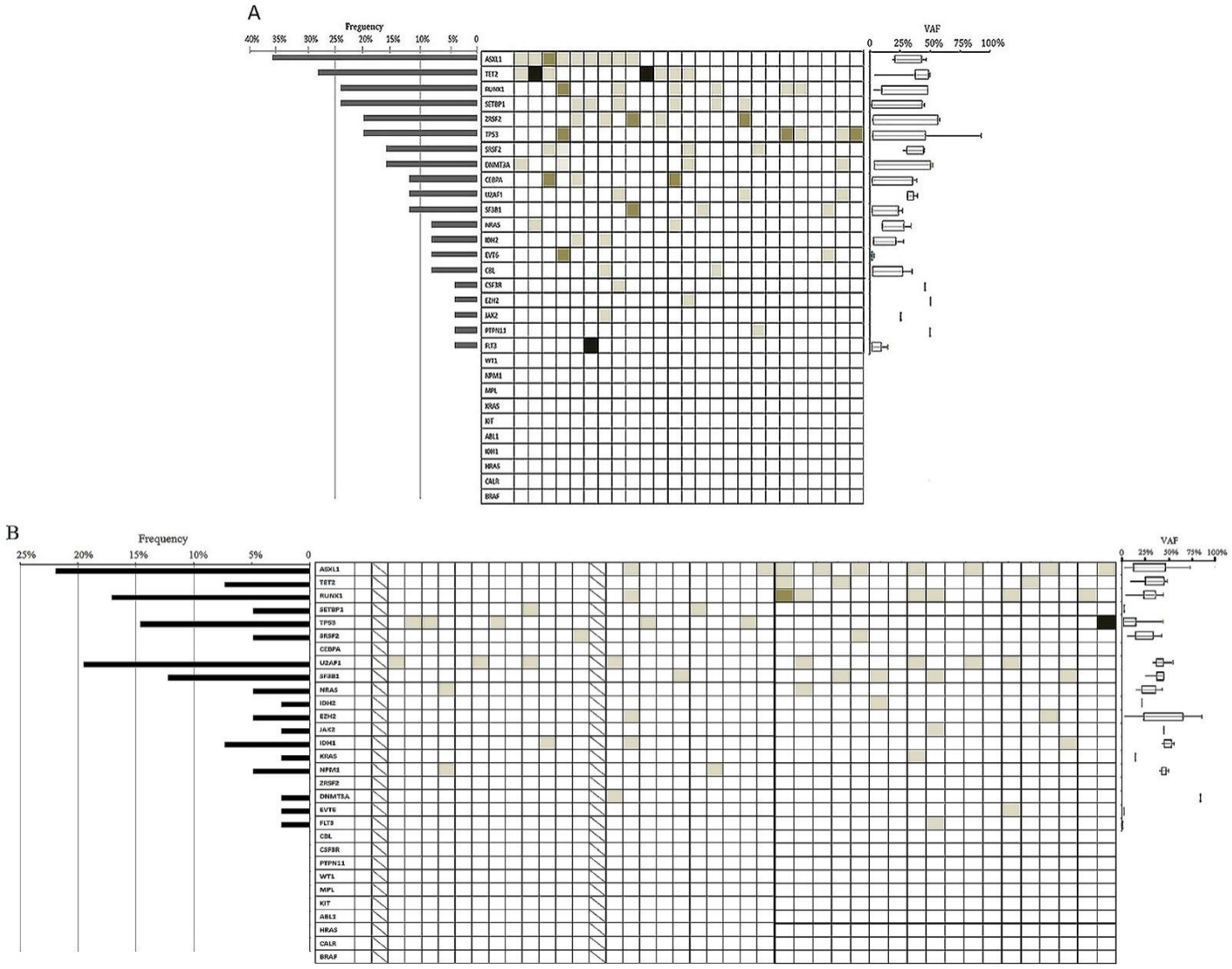
Footnote: (A) SALL4 expression in 37 MDS patients of cohort 1 at diagnosis before treatment in comparison to the controls (normal CD34(-) and CD34(+) cells). $CD34^-$ and $CD34^+$: from normal bone marrow; MDS: from bone marrow mononuclear cells of MDS patients; CR: complete remission; PR: partial remission; HI: hematologic improvement; SD: nonresponders with stable disease; PD: progressive disease. (B) Log_2 fold change of SALL4 based on responders and nonresponders in 25 patients in cohort 1. (C)Log_2 fold change of SALL4 based on responders and nonresponders in 43 patients in cohort 2.

Figure S2 (A-B): Survival based on the change of SALL4 expression in both cohorts.



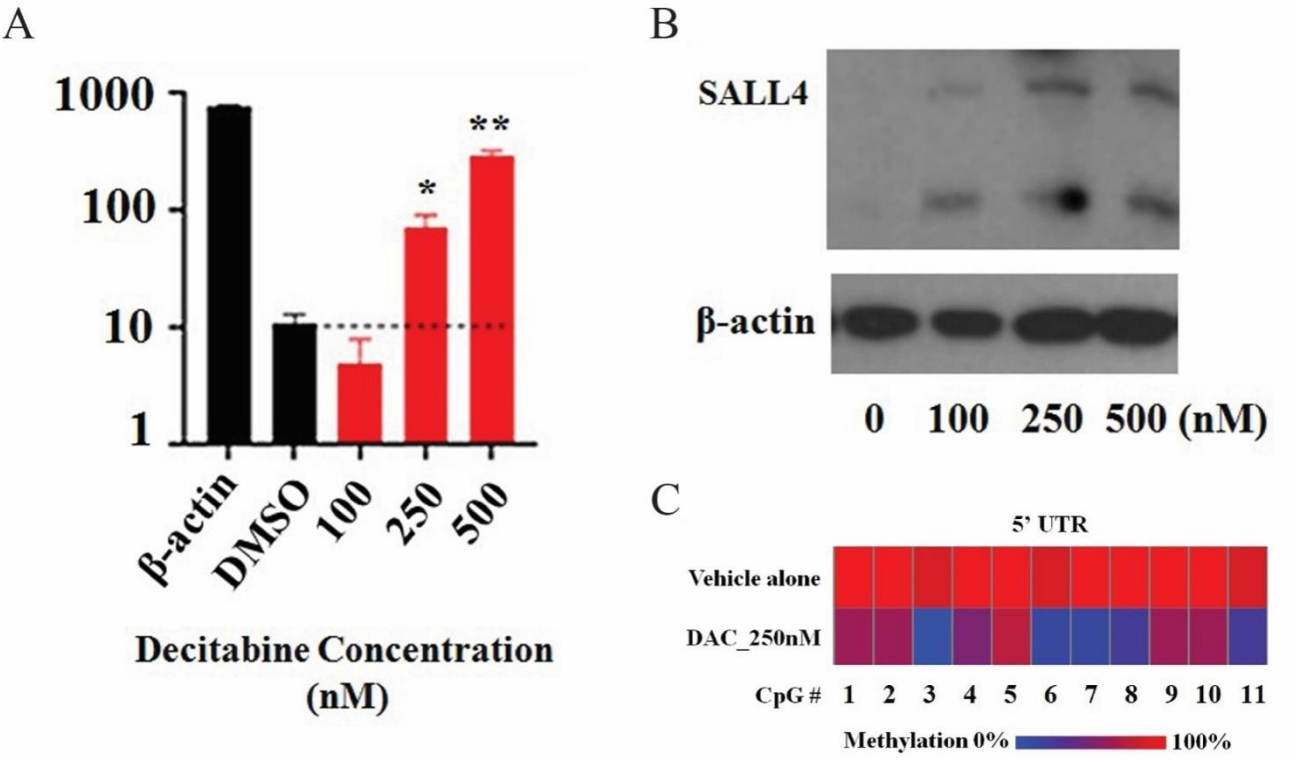
Footnote: (A) PFS between SALL4^{up} and SALL4^{down} in cohort 1. (B) PFS between SALL4^{up} and SALL4^{down} in cohort 2.

Figure S3 (A-B): Distribution, frequency, and variant allele fraction (VAF) of somatic mutations in both cohorts



Footnote: (A) Cohort 1 and (B) Cohort 2.Light-, intermediate-, and dark- black boxes indicate the presence of 1, 2, or ≥2 mutations in the same gene, whereas empty boxes indicate wild-type genes. Diagnoal boxes indicated no NGS in cohort 2.

Figure S4 (A-C): Demethylation of a critical CpG island and SALL4 expression in leukemic cells treated with HMA



Footnote: (A) SALL4 transcripts (copies per cell assessed by ddPCR) in K562 cells treated with DAC. (B) Western blot in K562 cells treated with DAC; (C) Methylation profiling in K562 cells treated with DAC. DAC treatment for all panels was for 5 days.

Table S1. Prognostic factors of cohort 1 MDS patients in OS (n = 25)

	Univariate analysis		Multivariate analysis	
Parameters	HR	95% CI	HR	95% CI
Age ≥ 60 years	1.84	0.49-6.92		
Male	4.06	0.50-32.56		
Nonresponders	1.86	0.46-7.47		
SALL4 upregulation after AZA.	4.24	1.05-17.22	6.48	1.06-39.67
IPSS intermediate-2 or high risk	24.71	0.00-x		
ANC < 1500/µl	0.42	0.10-1.72		
Hemoglobin < 10g/dL	2.04	0.50-8.25		
Platelet count < 100000/µl	0.93	0.23-3.74		
Poor karyotype	2.81	0.66-11.89		
Mutational profiles*				
ASXL1 mutation	1.97	0.52-7.40		
TET2 mutation	0.21	0.02-1.71		
RUNX1 mutation	3.91	1.00-15.19	10.66	1.25-90.72
SETBP1 mutation	4.45	1.16-17.04	1.85	0.37-9.12
TP53 mutation	2.80	0.68-11.46		
ZRSF2 mutation	3.64	0.85-15.52	2.94	0.30-28.19
DNMT3A mutation	0.70	0.08-5.65		
SRSF2 mutation	0.48	0.06-3.85		

HR, Hazard ratio; CI, Confidence Interval; SD, Stable disease; PD, Progression disease; AZA, Azacytidine; SALL4, Spalt like transcription factor 4; IPSS, International Prognostic Scoring System; ANC, Absolute neutrophil count
x:269250.47

*Mutations present in less than four patients were excluded from the analysis

Table S2. Prognostic factors of cohort 2 MDS patients in OS (n=43)

	Univariate analysis		Multivariate analysis	
Parameters	HR	95% CI	HR	95% CI
Age ≥ 60 years	2.60	0.59-11.52		
Male	7.71	1.02-58.41	3.59	0.46-28.20
Nonresponders	6.73	2.07-21.94	5.71	1.63-20.01
SALL4 upregulation	2.70	0.98-7.42	2.74	0.93-8.03
IPSS intermediate-2 or high risk	0.76	0.24-2.45		
ANC < 1500/µl	0.67	0.23-1.94		
Hemoglobin < 10g/dL	0.61	0.20-1.91		
Platelet count < 100000/µl	0.60	0.22-1.67		
Poor karyotype	2.88	0.62-13.33		
Mutational profiles*				
ASXL1 mutation	0.21	0.03-1.58		
RUNX1 mutation	1.36	0.30-6.10		
TP53 mutation	1.18	0.15-9.67		
SF3B1 mutation	2.98	0.82-10.89		
U2AF1 mutation	0.79	0.22-2.86		

HR, Hazard ratio; CI, Confidence Interval; SD, Stable disease; PD, Progression disease; AZA, Azacytidine; SALL4, Spalt like transcription factor 4; IPSS, International Prognostic Scoring System; ANC, Absolute neutrophil count

*Mutations present in less than four patients were excluded from the analysis

References:

1. Karahoca M, Momparler RL. Pharmacokinetic and pharmacodynamic analysis of 5-aza-2'-deoxycytidine (decitabine) in the design of its dose-schedule for cancer therapy. Clin Epigenetics 2013;5:3.
2. Qin T, Youssef EM, Jelinek J, et al. Effect of cytarabine and decitabine in combination in human leukemic cell lines. Clin Cancer Res 2007;13:4225-32.
3. Falconi G, Fabiani E, Piciocchi A, et al. Somatic mutations as markers of outcome after azacitidine and allogeneic stem cell transplantation in higher-risk myelodysplastic syndromes. Leukemia 2019;33:785-90.