### SUPPLEMENTARY APPENDIX

### IMPACT OF CONDITIONING INTENSITY AND GENOMICS ON RELAPSE AFTER ALLOGENEIC TRANSPLANTATION FOR PATIENTS WITH MYELODYSPLASTIC SYNDROME

Laura W. Dillon Ph.D., Gege Gui Sc.M., Brent R. Logan Ph.D., Mingwei Fei M.Sc., Jack Ghannam B.S., Yuesheng Li Ph.D., Abel Licon M.S., Edwin P. Alyea M.D., Asad Bashey M.D. Ph.D., Steven M. Devine M.D., Hugo F. Fernandez M.D., Sergio Giralt M.D., Mehdi Hamadani M.D., Alan Howard Ph.D., Richard T. Maziarz M.D., David L. Porter M.D., Erica D. Warlick M.D., Marcelo C. Pasquini M.D., Bart L. Scott M.D., Mitchell E. Horwitz M.D., H. Joachim Deeg M.D., Christopher S. Hourigan DM DPhil

#### **Table of Contents**

Supplementary Methods	Page 2
Supplementary Figure 1. Differences in overall survival, relapse, and transplant-related mortality between MAC and RIC censored at 18 months.	Page 10
Supplementary Figure 2. Clinical outcomes by mutation type, frequency, and co-occurrence	Page 11
Supplementary Figure 3. Clinical outcomes by disease classification, disease risk group, and cytogenetic risk group	Page 12
Supplementary Figure 4. Clinical outcomes by 10-gene mutational status and poor risk cytogenetics	Page 13
Supplementary Figure 5. Clinical outcomes by mutational classifications	Page 14
Supplementary Figure 6. Clinical outcomes by variant allele frequency	Page 16
Supplementary Table 1. Regions of interest in 29-gene targeted DNA sequencing panel	Page 17
Supplementary Table 2. Sample sequencing summary (n = 48)	Page 20
Supplementary Table 3. Variants detected by next-generation sequencing in the blood of MDS patients prior to conditioning	Page 21
Supplementary Table 4. Sensitivity and specificity of NGS mutational status prior to conditioning for predicting MDS patient relapse at 24 months	Page 23
Supplementary References	Page 24

#### SUPPLEMENTARY METHODS

#### SAMPLES AND DNA ISOLATION

Pre-conditioning baseline whole blood samples were collected a median of 12 days (100% within 30 days) prior to the date of transplantation. High quality genomic DNA (gDNA) was extracted from all 48 available frozen 1mL aliquots using the SPRI-TE gDNA Extraction kit (cat# A50087, Beckman Coulter, Inc., Brea, CA) on the SPRI-TE nucleic acid extractor (Beckman Coulter, Inc., Brea, CA). DNA concentration was measured using the Quant-iT PicoGreen dsDNA Assay kit (cat# P11496, Thermo Fisher Scientific, Waltham, MA) on the Synergy LX microplate reader (BioTek Instruments, Inc., Winooski, VT).

#### TARGETED NEXT-GENERATION SEQUENCING (NGS)

A custom anchored multiplex PCR based targeted DNA sequencing panel (VariantPlex, ArcherDx, Boulder, CO) was designed with coverage of 29 commonly mutated genes in Myelodysplastic Syndrome (MDS) (**supplementary Table 1**).

Libraries were prepared according manufacturer's instructions, with modifications, utilizing pre- and post-PCR separation on liquid handling robots. 200-ng of gDNA isolated from whole blood was brought to 50-µL using 10mM Tris-HCl pH 8.0 and was subjected to DNA fragmentation, end repair, A-tailing, purification using SPRIselect reagent (cat# B23318, Beckman Coulter, Inc., Brea, CA), and ligation with a universal ArcherDx molecular barcode (MBC) adapter, which tags each DNA molecule with a unique molecular index (UMI) and allows for unidirectional amplification of the sample using

gene-specific primers. Following molecular barcode ligation, the libraries were subjected to two rounds of nested PCR for target enrichment. DNA fragmentation, end repair, cleanup after end repair, ligation step 1, cleanup after ligation step 1, MBC adapter incorporation, ligation step 2, cleanup after ligation step 2, and setup of first PCR were all performed on a Sciclone G3 NGS Workstation (PerkinElmer Health Sciences, Inc., Shelton, CT). Cleanup after first PCR, second PCR, and cleanup after second PCR were all performed on an epMotion 5075 (Eppendorf AG, Hamburg, Germany). Incubations and PCR reactions were performed on a Mastercycler X50a (Eppendorf AG, Hamburg, Germany). For the first PCR, amplification was performed as follows: 95°C for 3 minutes; 16 cycles of 95°C for 30 seconds, 63°C for 5 minutes; 72°C for 3 minutes. For the second PCR, amplification was performed as follows: 95°C for 3 minutes; 30 seconds, 65°C for 5 minutes; 72°C for 3 minutes; 22 cycles of 95°C for 30 seconds, 63°C for 3 minutes.

The resulting libraries were subjected to paired-end 150-bp sequencing on a HiSeq 2500 in Rapid Run mode (Illumina, San Diego, CA), according to manufacturer's instructions. Libraries were sequenced in batches of 8 per lane, each with a unique dual index. A total of 108 million paired-end reads were acquired across the 48 MDS patient samples, with an average of 43 million paired-end reads per sample (**supplementary Table 2**). FASTQ files are available in the NCBI Small Reads Archive (SRA) (Accession: PRJNA644802).

#### NEXT GENERATION SEQUENCING DATA ANALYSIS

Raw sequencing FASTQ files were analyzed using the Archer Analysis software version 6.0.2.3, using default settings. The only exception was that the deep shallow threshold

3

was set to 3 (defining a deep error-correctable bin as  $\geq$ 3 reads per molecular barcode/UMI). DNA structural variation and SNP-InDel pipelines were utilized.

#### Preprocessing

Pre-processing steps include read cleaning, deduplication and error correction, read mapping, de novo and targeted variant calling, and variant annotation. During read cleaning, all reads were filtered based on Illumina's base quality calls and the presence of the appropriate Archer chemistry read signatures. Additionally, pairs of reads were aligned to determine overlap between R1 and R2. If overlap occurred, any discordant bases between R1 and R2 were converted to N's. Next, de-duplication and error correction of PCR duplicates was performed by extracting UMI read families and generating a single high-fidelity consensus read for use in downstream analysis. Any singleton reads were used as is, but not considered a deep bin (≥3 reads per molecular barcode). All reads (consensus and singleton) were then mapped to the human genome version hg19 (build GRCh37) using a combination of Bowtie2 and BWA-Mem. For the 48 MDS patient sample libraries, there was an average of 35 million mapped paired-end reads with an average on-target deduplication ratio of 6.9 (supplementary Table 2). Across all positions in the panel, there was a mean de-duplicated depth of 7,148 and a mean deep, error-correctable depth ( $\geq$ 3 reads per molecular barcode) of 3.429 (supplementary Table 2).

Variant calling was then performed on the aligned reads using *de novo* and targeted variant calling algorithms. *De novo* variant calling was performed using both Freebayes and LoFreq algorithms, with a minimum allele frequency for variant calling set to 0.001.

4

To improve coverage of well characterized insertion variants within the *NPM1* gene with established low background in our assay, targeted variant calling was performed by providing a VCF file with well-established *NPM1* variants to guarantee metrics were computed and supplied as an output in the bioinformatic pipeline. Annotation of variants was performed using the Variant Effect Predictor (VEP) algorithm.

*FLT3*-ITD mutations were detected using a structural variation algorithm. Beginning with deduplicated reads, clustering was performed for each gene specific primer, beginning with R2 reads and then assembled with their companion R1 reads using Velvet, creating a set of contigs for each R2/R1 cluster. Contigs were grouped together by target gene and merged based on pairwise local alignment, repeating until no further merges could take place. Contigs with low coverage, short length, or alignment to known transcripts or the reference genome were removed. The remaining sequences were annotated using NCBI blastn and those contigs sharing a breakpoint were merged. For each structural variant remaining, reads with coverage spanning the primary breakpoint by at least 20 bp were used to calculate the frequency of the mutation.

#### Error modeling

Following pre-processing, an outlier detection algorithm was performed to determine positions that have more alternate observations than would be expected due to error. In the outlier detection algorithm, two types of background error estimates are made. One based on all reads and another based only on consensus reads that originated from multiple PCR duplicates (i.e. deep bins with  $\geq$ 3 reads per molecular barcode/UMI). Using

5

the outlier detection algorithm, a position-specific background noise models based on a binomial distribution were generated for variants called in each sample compared to the Normal Data Set (ND). The ND background error model was generated using a set of 4 wild-type libraries made from gDNA isolated from the blood of de-identified normal donors, and corrected for the majority of errors introduced during library preparation, sequencing, and data analysis. For each variant called in a sample, a P value was calculated based on the ND background error models.

#### Variant filtering

#### De novo variants

Variants called by the *de novo* variant algorithms (LoFreq and Freebayes) underwent a first round of filtering in order to: 1) remove anomalies generated as the result of library preparation and sequencing; and 2) select for variants associated with leukemic progression. The following filter conditions were applied in step 1:

- 1. Deep Alternate Observations  $(DAO) \ge 5$ 
  - a. The number of consensus reads that come from deep molecular families must be greater than or equal to 5. A molecular family is considered deep if there are greater than or equal to 3 reads per UMI.
- 2. Unique Start for Alternate Observation (UAO)  $\geq$  5
  - a. The total number of unique starts across all reads that contained the alternate observations must be greater than or equal to 5.
- 3. gnomAD AF  $\leq 0.005$

- a. The population frequency of the variant in gnomAD database must not exceed 0.005.
- 4. VEP Consequence = coding\_sequence\_variant, feature\_elongation, feature\_truncation, frameshift\_variant, incomplete\_terminal\_codon\_variant, inframe\_deletion, inframe\_insertion, missense\_variant, protein\_altering\_variant, start\_lost, stop\_gained, stop\_lost, transcript\_ablation, transcript\_amplification, splice\_acceptor\_variant, or splice\_donor\_variant, splice\_region\_variant
  - a. The VEP consequence of any known isoform affected by the variant must include one of these consequences.
- 5. Has Sample Strand Bias = No
  - a. There must be no significant bias from primers that prime off of opposing strands.
- 6. Has Sequencing Direction Bias = No
  - a. There must be no significant bias from read pairs when using Illumina's paired-end chemistry
- 7. AF  $\geq$  0.001, excluding AF between 0.48 and 0.52
  - a. The AF of the variant must be greater than 0.001, but not be between 0.48 and 0.52 for exclusion of probable germline variants
- 8. HRUN < 9
  - a. The variant must not be flanked by a homopolymer run of greater than 8 bases.
- 9. ND background noise (BN)  $\leq 0.001$

- a. The ND error model position-specific background noise for all read families must be less than or equal to 0.001.
- 10. ND Deep Allele Frequency (DAF) outlier P value ≤ 0.0001
  - a. The ND background error model P value for deep consensus read families must be less than or equal to 0.0001.
- 11. Median distance from start site > 20
  - a. The median distance of the variant to the start site generated by random ligation of the molecular barcode adapter must be greater than 20.

All remaining *de novo* variants underwent manual curation to remove any remaining chemistry-related artifacts and to reannotate complex insertion-deletion variants present on the same allele.

#### **Targeted variants**

Targeted variants called within the NPM1 gene had the following filters applied:

- 1. Alternate Observations  $(AO) \ge 3$
- 2. VEP Consequence = frameshift\_variant

#### STATISTICAL ANALYSIS

Clinical characteristics of the 48 patients were compared with those of the 54 MDS patients in the original clinical trial using student's t-test or Mann-Whitney U test for continuous variables, and chi-squared test or Fisher's exact test for categorical variables. The same methods were applied for the baseline characteristics comparison of reduced

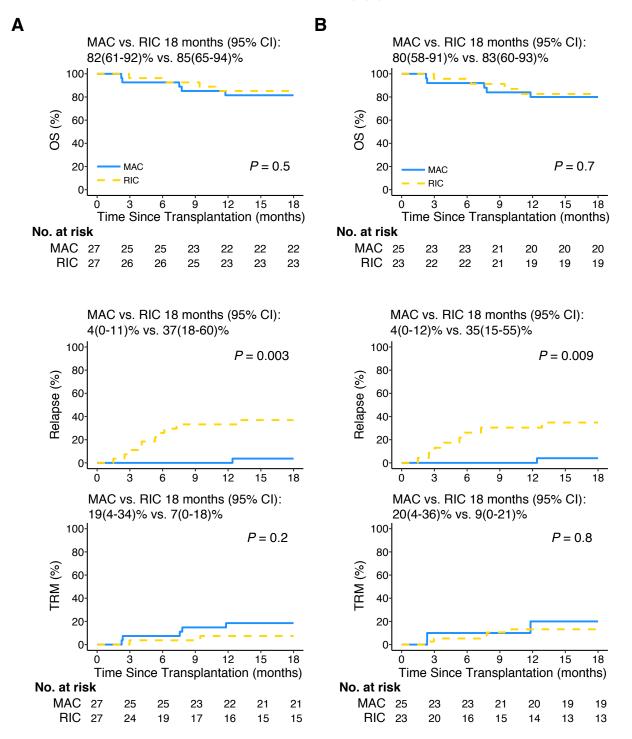
intensity conditioning (RIC) and myeloablative conditioning (MAC) patients, NGS positive and NGS negative patients.

The study of different clinical outcomes included the Kaplan-Meier estimation with logrank test for overall survival (OS) and relapse-free survival (RFS), the Gray's test for competing risks of treatment-related mortality (TRM) and relapse. The analysis was using the long-term Center for International Blood and Marrow Transplant Research (CIBMTR) data with P values reported, and the visualization was plotted at 3 years.

Two-sided P values <0.05 were considered significant. Statistical analysis and data visualization were performed with R software (version 3.6.0; R Foundation for Statistical Computing, Vienna, Austria) with its packages<sup>1-8</sup>.

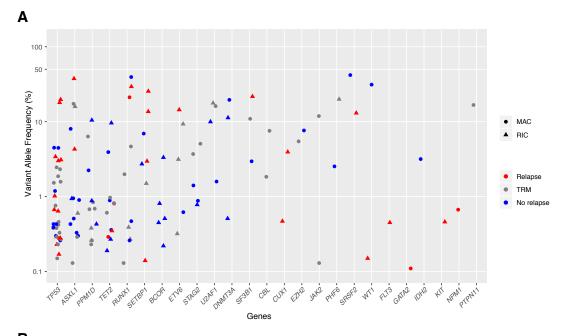
# Supplementary Figure 1. Differences in overall survival, relapse, and transplant-related mortality between MAC and RIC censored at 18 months.

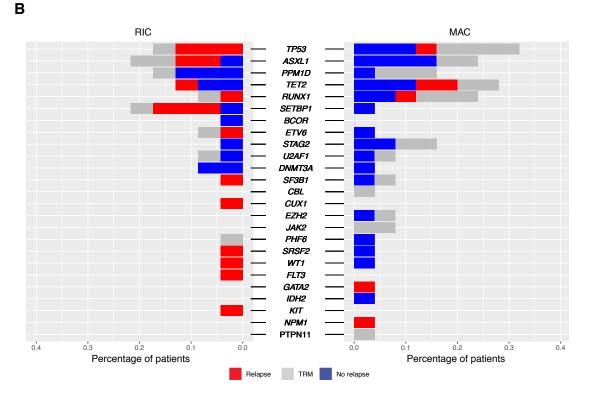
Rates of overall survival, relapse, and transplant related mortality for patients randomized to myeloablative conditioning (MAC; blue) and reduced intensity conditioning (RIC; yellow) are shown for the 54 myelodysplastic syndrome (MDS) patients in the original clinical trial (A) and the 48 MDS patients in the clinical cohort used in this study (B).



#### Supplementary Figure 2. Clinical outcomes by mutation type, frequency, and cooccurrence.

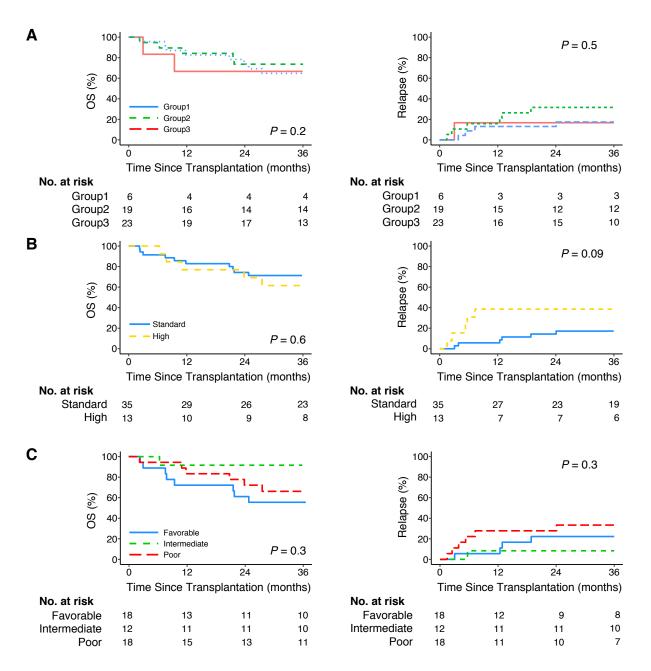
(A) The allele frequency of each mutation detected in each gene in the blood of myelodysplastic syndrome patients prior to transplantation is shown, with shapes indicating conditioning intensity and color indicating relapse outcome. (B) The percentage of patients with mutations in each gene is shown, separated by conditioning intensity. Red, relapse; Blue, no relapse; Grey, treatment related mortality (TRM).





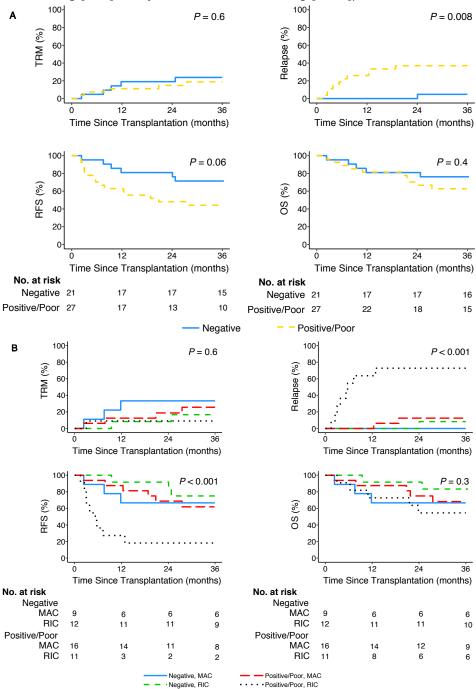
# Supplementary Figure 3. Clinical outcomes by disease classification, disease risk group, and cytogenetic risk group.

Models for overall survival (OS) and relapse were generating based on (A) disease classification (Group 1 [good risk] = Refractory Anemia with Ringed Sideroblasts and Refractory Cytopenia with Multilineage Dysplasia and Ringed Sideroblasts; Group 2 [evolving] = Refractory Anemia with Excess Blasts - 1 (5-9% blasts) and Refractory Anemia with Excess Blasts - 2 (10-19% blasts); and Group 3 [other] = Refractory Anemia, Refractory Cytopenia with Multilineage Dysplasia, MDS Unclassified, and MDS Associated with Isolated Del(5q)), (B) disease risk group, and (C) cytogenetic risk group.



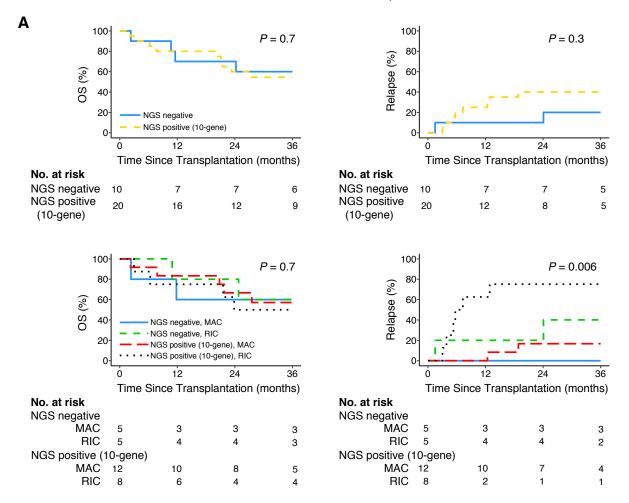
# Supplementary Figure 4. Clinical outcomes by 10-gene mutational status and poor risk cytogenetics.

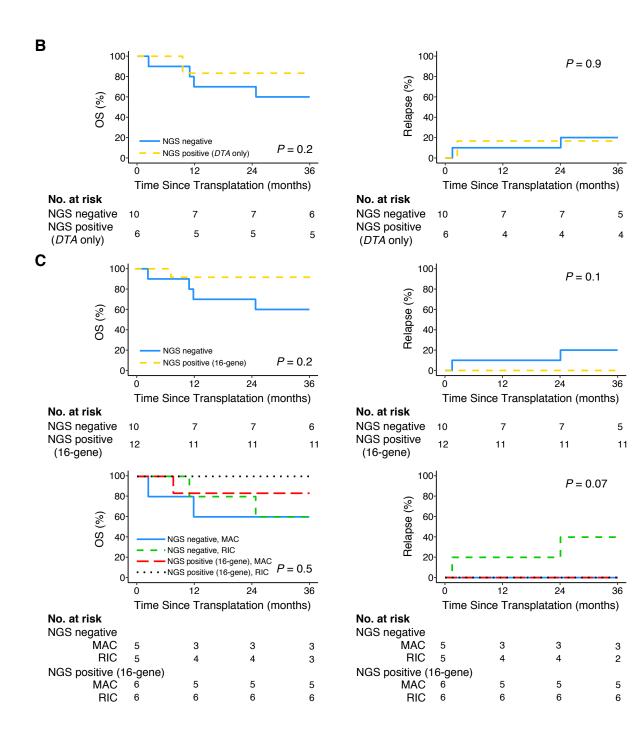
Patients were grouped into categories based on NGS mutational status and cytogenetic risk group. Patients were considered positive (Positive/Poor) if they had a mutation detectable in the blood by next-generation sequencing (NGS) within the 10-gene panel (*FLT3, IDH1, IDH2, JAK2, KIT, NPM1, NRAS, RUNX1, SF3B1,* and *TP53*) or had poor risk cytogenetics prior to conditioning and the remaining patients were considered negative. Models for rates of treatment related mortality (TRM), relapse, relapse-free survival (RFS), and overall survival (OS) were generated for (A) Positive/Poor versus negative status and (B) by conditioning intensity (reduced intensity conditioning [RIC] or myeloablative conditioning [MAC]).



#### Supplementary Figure 5. Clinical outcomes by mutational classifications.

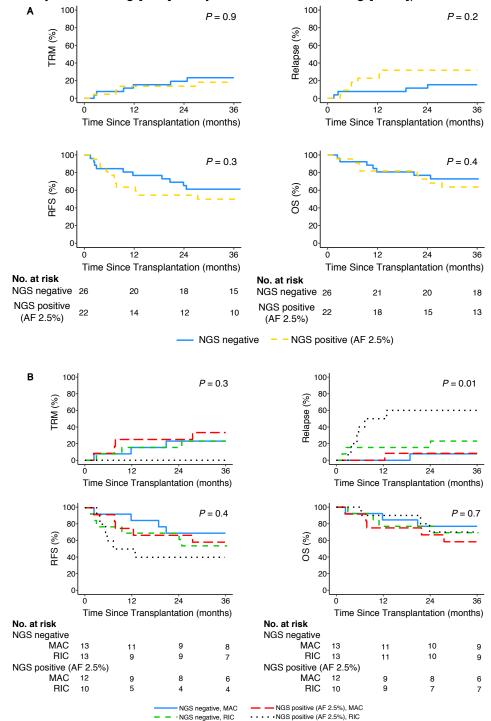
Patients were grouped into different categories based on the presence of mutations in groups of genes. Patients without mutations in any groups of genes were categorized as next-generation sequencing (NGS) negative. Models for overall survival (OS) and relapse were generated based on NGS positivity and conditioning intensity (reduced intensity conditioning [RIC] or myeloablative conditioning [MAC]) for patients with mutations in (A) 10-gene panel (*FLT3, IDH1, IDH2, JAK2, KIT, NPM1, NRAS, RUNX1, SF3B1,* and *TP53*), (B) *DTA* (DNMT3A, TET2, and ASXL1) only, and (C) 16-gene panel only (*BCOR, CBL, CUX1, ETV6, EZH2, GATA2, KRAS, PHF6, PPM1D, PTPN11, SETBP1, SRSF2, STAG2, U2AF1, WT1, ZRSR2*).





#### Supplementary Figure 6. Clinical outcomes by variant allele frequency.

Patients were grouped into categories based on the highest variant allele frequency (VAF) of a mutation detected within any of the 29 gene regions sequenced in the blood prior to conditioning. Patients with at least one mutation with a VAF  $\geq 2.5\%$  were classified as next-generation sequencing (NGS) positive and the remaining as NGS negative. Models for treatment related mortality (TRM), relapse, relapse-free survival (RFS), and overall survival (OS) were generated based on (A) NGS mutational status and (B) NGS mutational status and conditioning intensity (reduced intensity conditioning [RIC] or myeloablative conditioning [MAC]).



Gene ASXL1	Transcript NM_015338	Exons 11,12,13	ROI Genomic Coordinates (hg1 chr20:31020662-31020808 chr20:31021066-31021740
BCOR	NM_001123385	2,3,4,5,6,7,8,9,10,11,12,13,14,15	chr20:31022214-31025161 chrX:39937077-39937202
			chrX:39935686-39935805 chrX:39931581-39934453 chrX:39930869-39930963 chrX:39930205-39930432
			chrX:39923568-39923872 chrX:39922840-3992325 chrX:39921978-39922344
			chrX:39921371-39921666 chrX:39916387-39916594 chrX:39914600-39914786
			chrX:39913488-39913606 chrX:39913118-39913315 chrX:39911341-39911673
CBL	NM_005188	8,9	chr11:119148855-119149027 chr11:119149199-119149443
CUX1	NM_001913	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23	chr7:101459291-101459393 chr7:101559374-101559525 chr7:101671357-101671445 chr7:101713598-101713717 chr7:101740623-101740801 chr7:101747595-101747759
			chr7:101754957-101755074 chr7:10175466-101758573 chr7:101801819-101801908 chr7:101801819-101801908 chr7:101821728-101821957
			chr7:101833072-101833171 chr7:101837101-101837190 chr7:101837166-101838903 chr7:101916616-101916784 chr7:101917494-101917601
			chr7:101918497-101918650 chr7:101921199-101921356 chr7:101923308-101923432 chr7:101924075-101924172 chr7:101925111-10192523 chr7:101925983-101926088
			chr7:101926292-101926402
	NM_001202543	15,16,17,18,19,20,21,22,23,24	chr7:101839893-101840605 chr7:101842061-101842167 chr7:101842061-101843472 chr7:1018443330-101843472 chr7:101847650-101847856 chr7:1018478373-101847840 chr7:101877311-101877540 chr7:101882579-10188284 chr7:101891671-101882342
	NM_181552	1	chr7:1014609899-101460969
<u>NMT3A</u> ETV6	NM_022552 NM_001987	23 1,2,3,4,5,6,7,8	chr2:25457127-25457309 chr12:11803041-11803114 chr12:11905363-11905533 chr12:11992053-11992258 chr12:12026340-12006515 chr12:1202237-12022923 chr12:12023758-12037541 chr12:12038839-12038980
EZH2	NM_004456	2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20	chr12:12043854-12044000 chr7:148544253-148544410 chr7:148543541-148543710 chr7:148529705-148529862
			chr7:148526709-148525805 chr7:148526799-1485250960 chr7:148524235-14852437 chr7:148524525-148523744 chr7:148516667-148516799 chr7:148514948-148515290 chr7:148514293-148516593 chr7:148514293-1485145203 chr7:148514293-148513890
			chr7:14851257-148512658 chr7:148511985-148512151 chr7:148511985-148512151 chr7:14850896-1485082 chr7:1485087404-148507526 chr7:148506381-148506502
			chr7:148506142-148506267 chr7:148504717-148504818

## Supplementary Table 1. Regions of interest in 29-gene targeted DNA sequencing panel

FLT3	NM_004119	8,9,10,11,12,13,14,15,16,17,19,20,21	chr13:28623500-28623694
			chr13:28622391-28622600
			chr13:28611301-28611445
			chr13:28610051-28610200
			chr13:28609611-28609830
			chr13:28608417-28608564
			chr13:28608198-28608371
			chr13:28608003-28608148
			chr13:28602294-28602445
			chr13:28601204-28601398
			chr13:28597466-28597634
			chr13:28592583-28592746
			chr13:28589706-28589858
GATA2	NM 032638	2,3,4,5,6	chr3:128205625-128205894
		_,_,,_,_	chr3:128204549-128205231
			chr3:128202682-128202868
			chr3:128200641-128200807
			chr3:128199841-128200181
IDH1	NM 005896	3,4	chr2:209116133-209116295
	1100_003090	5,4	chr2:209113072-209113404
IDUIA	NIM 000160	4.0	
IDH2	NM_002168	4,6	chr15:90631798-90631999
			chr15:90630650-90630827
JAK2	NM_004972	12,13,14,15,18,19,20,21,22	chr9:5069904-5070072
1			chr9:5072471-5072646
1			chr9:5073677-5073805
1			chr9:5077432-5077600
1			chr9:5078285-5078464
			chr9:5081704-5081881
			chr9:5089653-5089883
1			chr9:5090425-5090590
			chr9:5090718-5090931
KIT	NM_000222	2,8,9,10,11,12,13,14,15,17,18	chr4:55561657-55561967
			chr4:55589729-55589884
			chr4:55592002-55592236
			chr4:55593363-55593510
			chr4:55593561-55593728
			chr4:55593968-55594113
			chr4:55594156-55594307
			chr4:55595480-55595671
			chr4:55597473-55597605
			chr4:55599215-55599378
			chr4:55602643-55602795
KRAS	NM_004985	2,3,4	chr12:25398187-25398338
			chr12:25380147-25380366
			chr12:25378527-25378727
NPM1	NM_002520	11	chr5:170837200-170837600
NRAS	NM_002524	2,3,4	chr1:115258650-115258801
			chr1:115256400-115256619
			chr1:115252169-115252369
PHF6	NM_032458	2,3,4,5,6,7,8,9,10	chrX:133511627-133511805
			chrX:133512014-133512156
			chrX:133512014-133512156
			chrX:133527510-133527684
			chrX:133527918-133528002
1			chrX:133547497-133547707
			chrX:133547832-133548016
			chrX:133549025-133549272
			chrX:133551178-133551352
1			chrX:133559210-133559380
PPM1D	NM_003620	6	chr17:58740335-58740933
PTPN11	NM_002834	3,4,7,8,12,13	chr12:112888101-112888336
1		-,.,.,,	chr12:112890978-112891211
			chr12:112910727-112910864
1			chr12:112915434-112915554
			chr12:112926226-112926334
1			chr12:112926807-112926999
1	NM 080601	11	chr12:112924258-112924457
DUNK		11	
RUNX1	NM_001754	2,3,5,6,7,8,9	chr21:36421118-36421216
1			chr21:36265201-36265280
1			chr21:36252833-36253030
			chr21:36231750-36231895
1			chr21:36206686-36206918
			chr21:36171577-36171779
			chr21:36164411-36164927
1	NM_001122607	1,5	chr21:36259119-36259429
			chr21:36193944-36194013
SETBP1	NM_015559	4	chr18:42531679-42532175
SF3B1	NM 012433	13,14,15,16	chr2:198267652-198267779
			chr2:198267259-198267570
			chr2:198266688-198266874
			chr2:198266445-198266632
	1 1		

SRSF2 STAG2	NM_003016	1,2	chr17:74732860-74733262
CTA CO			chr17:74732222-74732566
STAGZ	NM_006603	2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33	chrX:123156457-123156541
			chrX:123159669-123159788 chrX:123164790-123164995
			chrX:123171356-123171493
			chrX:123176398-123176515
			chrX:123178993-123179238
			chrX:123181183-123181375
			chrX:123182834-123182948 chrX:123184015-123184179
			chrX:123184950-123185089
			chrX:123185144-123185264
			chrX:123189957-123190105
			chrX:123191695-123191847 chrX:123195053-123195211
			chrX:123195600-123195744
			chrX:123196731-123196864
			chrX:123196945-123197075
			chrX:123197677-123197921
			chrX:123199705-123199816 chrX:123200004-123200132
			chrX:123200185-123200306
			chrX:123202393-123202526
			chrX:123204978-123205193
			chrX:123210161-123210341 chrX:123211786-123211928
			chrX:123215209-123215398
			chrX:123217250-123217419
			chrX:123220376-123220640
			chrX:123224404-123224634
			chrX:123227847-123228014 chrX:123229201-123229319
			chrX:123234403-123234467
	NM_001042749	32	chrX:123224683-123224834
TET2	NM_001127208	3,4,5,6,7,8,9,10,11	chr4:106155079-106158617
			chr4:106162475-106162606
			chr4:106163970-106164104 chr4:106164706-106164955
			chr4:106180755-106180946
			chr4:106182895-106183025
			chr4:106190746-106190924
			chr4:106193700-106194095 chr4:106196184-106197696
7050	NINA 000540		
TP53	NM_000546	1,2,3,4,5,6,7,8,9,10,11	chr17:7590674-7590888 chr17:7579679-7579960
			chr17:7579291-7579610
			chr17:7578350-7578574
			chr17:7578156-7578309
			chr17:7577478-7577628 chr17:7576998-7577175
			chr17:7576832-7576946
			chr17:7573906-7574053
			chr17:7572906-7573028
	NM_001276696	10	chr17:7576516-7576677
U2AF1	NM_006758	2,6	chr21:44524404-44524532 chr21:44514744-44514918
	NM 001025204	6	chr21:44515527-44515666
WT1	NM_000378	1,2,3,4,5,6,7,9	chr11:32456225-32456911
	_		chr11:32450022-32450185
			chr11:32449481-32449624
			chr11:32439102-32439220 chr11:32421473-32421610
			chr11:32417782-32421010
			chr11:32414191-32414321
			chr11:32410583-32410745
	NM_001198552	8	chr11:32413497-32413630
ZRSR2	NM_005089	1,2,3,4,5,6,7,8,9,10,11	chrX:15808598-15808679 chrX:15809036-15809156
			chrX:15817974-15818096
			chrX:15821790-15821939
			chrX:15822213-15822340
			chrX:15826335-15826414
			chrX:15827302-15827461 chrX:15833779-15834033
			chrX:15836689-15836785
	I		CIIIX. 13030009=13030703
			chrX:15838309-15838459

Total paired-end reads	2,058,856,848
Average paired-end reads per sample	$42,892,851 \pm 11,568,447$
Total mapped paired-end reads	1,678,734,150
Average mapped paired-end reads per sample	34,973,628 ± 9,472,727
Average on-target deduplication ratio	6.90 ± 1.78
Mean depth per sample (all bins)	7,148 ± 1,809
Mean error-correctable deep depth per sample	$3,\!429\pm830$
(≥3 reads/UMI)	
Unique molecular index, UMI	

# Supplementary Table 3 Variants detected by next-generation sequencing in the blood of MDS patients prior to conditioning.

				<u> </u>			Deep	Unique							
Patient	Gene	HGVSc	HGVSp	Allele Fraction	Deep Allele Fraction	Alternate Observations	Alternate Observations	Alternate Observations	Depth	Deep Depth	Variant Consequence	Chromosome	Position	Reference Sequence	Alternate Sequence
1	ASXL1 ETV6	NM_015338.5:c.2311_2315del NM_001987.4:c.1147C>G	NP_056153.2:p.Val771Ter NP_001978.1:p.His383Asp	0.1588	0.1581 0.0041	1059 17	566 8	241	6670 5276	3579 1941	frameshift missense	20	31022825 12037516	AGTAGC C	A G
1	ETV6 ETV6	NM_001987.4:c.1218_1220del NM_001987.4:c.1195C>G	NP_001978.1:p.lle407del NP_001978.1:p.Arg399Gly	0.0314 0.0931	0.0324 0.0861	180 521	41 100	107 197	5732 5597	1267 1161	inframe deletion missense	12 12	12038922 12038902	CATT C	C G
1	PHF6	NM_032458.2:c.346C>T	NP_115834.1:p.Arg116Ter	0.1992	0.2045	262	118	106	1315	577	nonsense	х	133527636	с	т
1	RUNX1 RUNX1	NM_001754.4:c.694C>T NM_001754.4:c.319C>T	NP_001745.2:p.Arg232Trp NP_001745.2:p.Arg107Cys	0.0039	0.0031	28 17	13	22	7230 6218	4201 3172	missense missense	21 21	36206818 36259172	G	A
1	SETBP1 U2AF1	NM_015559.2:c.2612T>C NM_006758.2:c.101C>T	NP_056374.2:p.lle871Thr NP_006749.1:p.Ser34Phe	0.015	0.0134 0.1672	196 712	113 163	114 239	13093 4015	8428 975	missense missense	18 21	42531917 44524456	T G	C A
2	ASXL1 ASXL1	NM_015338.5:c.2077C>T NM_015338.5:c.1804G>T	NP_056153.2:p.Arg693Ter NP_056153.2:p.Glu602Ter	0.0051 0.009	0.0056	64 74	28 41	45 54	12428 8220	5002 4580	nonsense	20 20	31022592 31022319	C G	T
2		<u> </u>												TCACCACTGCCATA	· · · · · · · · · · · · · · · · · · ·
2	ASXL1 ETV6	NM_015338.5:c.1900_1922del NM_001987.4:c.856_863del	NP_056153.2:p.Glu635ArgfsTer15 NP_001978.1:p.Asp286ThrfsTer11	0.0802	0.0801 0.0088	1102 19	577 14	310 13	13749 3048	7202 1594	frameshift frameshift	20	31022402 12022749	GAGAGGCGGC GGATTTCAA	G
2	EZH2 SETBP1	NM_004456.4:c.1021del NM_015559.2:c.2602G>A	NP_004447.2:p.Ala341336LeufsTer8 NP_056374.2:p.Asp868Asn	0.0763 0.0694	0.0716	509 1127	139 669	179 363	6667 16243	1941 9496	frameshift missense	7 18	148515187 42531907	GC	G
														TCACCACTGCCATA	-
3	ASXL1 ETV6	NM_015338.5:c.1900_1922del NM_001987.4:c.536_537insA	NP_056153.2:p.Glu635ArgfsTer15 NP_001978.1:p.His180AlafsTer16	0.006	0.0062 0.1229	77 344	49 143	58 140	12918 2387	7905 1164	frameshift frameshift	20 12	31022402 12022430	GAGAGGCGGC T	T TA
		NM_004119.2:c.1837_1837+1insTTAA TGAGTACTTCTACGTTGATTTCAGAGAA													CCAAACTCTAAATTTTCTCTT GGAAACTCCCATTTGAGATC
4	FLT3		GluTyrAspLeuLysTrpGluPheProArgGlu AsnLeuGluPhe	0.0045	0.0056	25	16	7	5577	2866	FLT3 ITD	13	28608218	c	ATATTCATATTCTCTGAAATC AACGTAGAAGTACTCATTAA
4	KIT	NM_000222.2:c.2484+1G>A		0.0046	0.0046	29	15	27	6246	3284	splicing	4	55599359	G	A
4	SETBP1 SRSF2	NM_015559.2:c.2608G>A NM_003016.4:c.284C>G	NP_056374.2:p.Gly870Ser NP_003007.2:p.Pro95Arg	0.1364 0.1304	0.1366 0.1307	965 173	635 91	345 78	7076 1327	4650 696	missense missense	18 17	42531913 74732959	G	A C
5	PPM1D PPM1D	NM_003620.3:c.1384C>T NM_003620.3:c.1412del	NP_003611.1:p.Gln462Ter NP_003611.1:p.Pro471GlnfsTer12	0.0068	0.0057	36 38	7	31 30	5273 5536	1231 1389	nonsense frameshift	17	58740479 58740505	C TC	T T
5	TET2	NM_001127208.2:c.1118_1127del	NP_001120680.1:p.Gln373ArgfsTer51	0.0061	0.0054	46	15	37	7583	2797	frameshift	4	106156216	CAAAATGAAAT	с
5	TP53 ASXL1	NM_000546.5:c.105dup NM_015338.5:c.2269C>T	NP_000537.3:p.Pro36AlafsTer7 NP_056153.2:p.Gln757Ter	0.0046	0.0051 0.1731	22 1003	11 523	16 245	4814 5789	2139 3022	frameshift nonsense	17 20	7579581 31022784	G	GC T
6	U2AF1 ASXL1	NM_006758.2:c.101C>T NM_015338.5:c.2762_2766del	NP_006749.1:p.Ser34Phe NP_056153.2:p.Ser921Ter	0.1613	0.1533	676 26	165 8	224 20	4191 7996	1076 3374	missense frameshift	21 20	44524456 31023275	G ATCTGT	A
7	ASXL1	NM_015338.5:c.2223del	NP_056153.2:p.Asp741GlufsTer3	0.003	0.0026	18	6	15	5930	2282	frameshift	20	31022737	AT	A
7	STAG2 BCOR	NM_001042749.1:c.1810C>T NM_001123385.1:c.4428+1G>A	NP_001036214.1:p.Arg604Ter	0.0141 0.0022	0.0141 0.0025	72	36 7	45 15	5103 7639	2547 2826	splicing	X	123197044 39921391	C C	T
8	BCOR BCOR	NM_001123385.1:c.464_467del NM_001123385.1:c.2474del	NP_001116857.1:p.lle155LysfsTer5 NP_001116857.1:p.Ser825ThrfsTer31	0.0045	0.0036	33 38	13 20	25 27	7406 7490	3576 3523	frameshift frameshift	x	39934131 39932124	TTGTA GC	T G
8	BCOR	NM_001123385.1:c.2373G>A	NP_001116857.1:p.Trp791Ter	0.0081	0.0096	52	28	42	6456	2902	nonsense	х	39932226	с	т
8	BCOR TET2	NM_001123385.1:c.2003dup NM_001127208.2:c.2797C>T	NP_001116857.1:p.Ser669LysfsTer71 NP_001120680.1:p.Gln933Ter	0.0333 0.0027	0.0374 0.0022	246 17	110 7	109 12	7378 6203	2944 3125	frameshift nonsense	X 4	39932595 106157896	T C	TA T
10	EZH2	NM_004456.4:c.401T>C	NP_004447.2:p.Met134Thr	0.0546	0.046	277	29	146	5072	631	missense	7	148526903	A TCACCACTGCCATA	G
11	ASXL1 STAG2	NM_015338.5:c.1900_1922del	NP_056153.2:p.Glu635ArgfsTer15	0.0094	0.0102	114	65 36	62 51	12092	6403	frameshift	20 X	31022402	GAGAGGCGGC	T
12	TP53	NM_001042749.1:c.3034C>T NM_000546.5:c.375G>A	NP_001036214.1:p.Arg1012Ter NP_000537.3:p.Thr125=	0.0078	0.0088	78 18	10	16	9987 7673	4084 3869	splicing	17	7579312	c	T
12	TP53 TP53	NM_000546.5:c29+1G>T NM_000546.5:c.536A>G	NP_000537.3:p.His179Arg	0.0064	0.0064	58 35	21	42	9038 5209	3270 1408	splicing missense	17	7590694 7578394	C T	A
12	TP53	NM_000546.5:c.673-1G>A		0.0102	0.0111	102	55	60	10044	4972	splicing	17	7577609	с	T
12	TP53 TP53	NM_000546.5:c.332T>G NM_000546.5:c.742C>T	NP_000537.3:p.Leu111Arg NP_000537.3:p.Arg248Trp	0.0301 0.031	0.0288	185 252	82 113	93 108	6143 8126	2851 3467	missense missense	17	7579355 7577539	A G	A
12 13	TP53 DNMT3A	NM_000546.5:c.607G>T NM_022552.4:c.2645G>A	NP_000537.3:p.Val203Leu NP_072046.2:p.Arg882His	0.0341 0.1954	0.0361 0.2011	200 1153	58 324	98 300	5859 5902	1606 1611	missense missense	17	7578242 25457242	C C	A
13	TET2	NM_001127208.2:c.4044+1G>A		0.0089	0.0109	50	21	42	5611	1922	splicing	4	106183006	G	A
14 14	TP53 TP53	NM_000546.5:c.683_686del NM_000546.5:c.636del	NP_000537.3:p.Asp228ValfsTer18 NP_000537.3:p.Arg213AspfsTer34	0.0017 0.18	0.0012 0.1795	13 923	5 363	13 215	7685 5128	4250 2022	frameshift frameshift	17	7577594 7578212	ACAGT GA	A G
14 15	TP53 PPM1D	NM_000546.5:c.376-1G>A NM_003620.3:c.1446_1450del	NP 003611.1:p.Thr483LysfsTer4	0.1974 0.0043	0.1906	1319 25	641 6	291 22	6681 5788	3363 2305	splicing frameshift	17	7578555 58740539	C CTGACT	T C
16	ASXL1	NM_015338.5:c.2555C>A	NP_056153.2:p.Ser852Ter	0.0095	0.0096	85	31	55	8910	3234	nonsense	20	31023070	с	А
16 17	TP53 PPM1D	NM_000546.5:c.375+1G>A NM_003620.3:c.1434C>A	NP_003611.1:p.Cys478Ter	0.0119 0.0038	0.0137	77 31	38 14	49 24	6453 8121	2767 3791	splicing nonsense	17	7579311 58740529	C C	T A
17 18	PPM1D RUNX1	NM_003620.3:c.1412dup NM_001754.4:c.562A>G	NP_003611.1:p.Glu472ArgfsTer4 NP_001745.2:p.Thr188Ala	0.0026	0.0049	10 3319	7 1843	9 531	3866 15691	1437 8878	frameshift missense	17 21	58740505 36231822	T	тс
18	TET2	NM_001127208.2:c.3660del	NP_001120680.1:p.Cys1221ValfsTer5	0.0081	0.0074	48	20	34	5902	2690	frameshift	4	106164790	AC	A
19 20	TP53 PTPN11	NM_000546.5:c.743G>A NM_002834.4:c.1403C>T	NP_000537.3:p.Arg248GIn NP_002825.3:p.Thr468Met	0.0026	0.002	23 1567	9 398	18 384	8893 9380	4597 2468	missense missense	17 12	7577538 112926270	C C	T
20 20	TP53 TP53	NM_000546.5:c.422G>A NM_000546.5:c.742C>T	NP_000537.3:p.Cys141Tyr NP_000537.3:p.Arg248Trp	0.0023	0.0028	13 29	6 10	10	5536 6920	2129 2831	missense missense	17	7578508 7577539	C G	T A
20	TP53	NM_000546.5:c.731G>A	NP_000537.3:p.Gly244Asp	0.0038	0.0036	27	11	21	7090	3030	missense	17	7577550	с	T
20 20	TP53 TP53	NM_000546.5:c.722C>A NM_000546.5:c.380C>T	NP_000537.3:p.Ser241Tyr NP_000537.3:p.Ser127Phe	0.0027	0.0037	20 27	12	18 20	7442 7117	3228 3061	missense missense	17	7577559 7578550	G	A
20 20	TP53 TP53	NM_000546.5:c.818G>T NM_000546.5:c.821T>A	NP_000537.3:p.Arg273Leu NP_000537.3:p.Val274Asp	0.0033	0.0042	17 40	7 14	12	5145 5261	1684 1626	missense missense	17	7577120 7577117	C A	A
20	TP53	NM_000546.5:c.637C>T	NP_000537.3:p.Arg213Ter	0.0153	0.0116	89	19	45	5800	1633	nonsense	17	7578212	G	A
20 20	TP53 TP53	NM_000546.5:c.533A>C NM_000546.5:c.783-1G>T	NP_000537.3:p.His178Pro	0.0158	0.0137	63 116	14 43	41 56	3991 6211	1022 2438	splicing	17 17	7578397 7577156	c	A
20 21	TP53 JAK2	NM_000546.5:c.527G>A NM_004972.3:c.1993-6T>G	NP_000537.3:p.Cys176Tyr	0.0246	0.0254 0.0008	98 17	26	47	3979 13590	1023 6001	missense splicing	17	7578403 5078300	C T	T G
21	PPM1D	NM_003620.3:c.1270G>T	NP_003611.1:p.Glu424Ter	0.0026	0.0027	22	5	19	8449	1882	nonsense	17	58740365	G	T
21 21	PPM1D PPM1D	NM_003620.3:c.1434C>A NM_003620.3:c.1615del	NP_003611.1:p.Cys478Ter NP_003611.1:p.Glu539LysfsTer8	0.0084 0.0634	0.0097	143 454	55 90	79 142	17061 7163	5666 1510	nonsense frameshift	17 17	58740529 58740709	C AG	A
21 21	RUNX1 TP53	NM_001754.4:c.532A>G NM_000546.5:c.711G>A	NP_001745.2:p.Thr178Ala NP_000537.3:p.Met237lle	0.0013	0.0005	21 227	5 95	16 94	16507 9764	9104 3987	missense missense	21 17	36231852 7577570	T C	C T
22 22	CBL	NM_005188.3:c.1211G>A NM_005188.3:c.1150T>G	NP_005179.2:p.Cys404Tyr	0.0184	0.0184	142 441	21 43	81 126	7702 5842	1144 532	missense missense	11 11	119148991 119148930	G	A
22	JAK2	NM_004972.3:c.1849G>T	NP_005179.2:p.Cys384Gly NP_004963.1:p.Val617Phe	0.1189	0.1149	1568	403	341	13186	3508	missense	9	5073770	G	T
22 22	RUNX1 SF3B1	NM_001754.4:c.430_431insATTGC NM_012433.2:c.1998G>C	NP_001745.2:p.Leu144HisfsTer3 NP_036565.2:p.Lys666Asn	0.0467 0.1093	0.0489 0.1038	697 809	308 142	240 222	14928 7404	6303 1368	frameshift missense	21	36252931 198267359	A C	AGCAAT G
22 22	STAG2 TET2	NM_001042749.1:c.2857C>T NM_001127208.2:c.3820C>T	NP_001036214.1:p.Arg953Ter NP_001120680.1:p.Gln1274Ter	0.0508	0.0469 0.0071	538 52	126	160 37	10599 6370	2689 991	nonsense	X 4	123215311 106180792	c	T
22	TET2	NM_001127208.2:c.3965T>A	NP_001120680.1:p.Leu1322Gln	0.0097	0.0084	110	32	68	11336	3816	missense	4	106182926	T	A
23 24	TET2 RUNX1	NM_001127208.2:c.286C>T NM_001754.4:c.292del	NP_001120680.1:p.Arg96Cys NP_001745.2:p.Leu98SerfsTer24	0.0036 0.3947	0.0037	22 3055	10 1345	18 471	6066 7741	2707 3443	missense frameshift	4 21	106155385 36259198	C AG	T A
24	WT1	NM_000378.4:c.1249C>T	NP_000369.3:p.Arg417Cys	0.312	0.3169	3472	1531	583	11129	4831	missense	11	32414251	G TCACCACTGCCATA	A
25	ASXL1	NM_015338.5:c.1900_1922del	NP_056153.2:p.Glu635ArgfsTer15	0.0013	0.0011	18	8	14	13435	7324	frameshift	20	31022402	GAGAGGCGGC	т
25 25	ASXL1 PPM1D	NM_015338.5:c.1312G>A NM_003620.3:c.1709C>G	NP_056153.2:p.Asp438Asn NP_003611.1:p.Ser570Ter	0.0029	0.0022	18 10	6 5	15 9	6280 4299	2774 1405	nonsense	20 17	31021313 58740804	G	A G
25 25	RUNX1 STAG2	NM_001754.4:c.601C>T NM_001042749.1:c.2049del	NP_001745.2:p.Arg201Ter NP_001036214.1:p.Asp683GlufsTer9	0.0199 0.037	0.0184 0.0305	243 155	127 41	129 100	12233 4194	6892 1346	nonsense frameshift	21 X	36231783 123199748	G AT	A
25	TP53 TET2	NM_000546.5:c.733G>T	NP_000537.3:p.Gly245Cys NP_001120680.1:p.Thr229AspfsTer25	0.0015	0.0018	10	5	10	6472	2850 3610	missense	17	7577548	C	A GA
27	DNMT3A	NM_022552.4:c.2645G>A	NP_072046.2:p.Arg882His	0.0051	0.0058	26	8	35 25	13364 5066	1372	frameshift missense	4	25457242	с	т
28 28	RUNX1 SETBP1	NM_001754.4:c.320G>A NM_015559.2:c.2608G>A	NP_001745.2:p.Arg107His NP_056374.2:p.Gly870Ser	0.2937 0.0297	0.2934 0.0285	3183 639	1375 336	405 238	10839 21482	4686 11770	missense missense	21 18	36259171 42531913	G	T A
28	SETBP1 WT1	NM_015559.2:c.2612T>C NM_000378.4:c.1250G>A	NP_056374.2:p.lle871Thr NP_000369.3:p.Arg417His	0.2548	0.2511 0.0011	5421 23	2906	671 20	21278 15103	11572 6106	missense missense	18 11	42531917 32414250	T	C T
29	ASXL1	NM_015338.5:c.1762C>T	NP_056153.2:p.Gln588Ter	0.0043	0.0037	64	29	49	14956	7805	nonsense	20	31022277	с	т
29 30	STAG2 IDH2	NM_001042749.1:c.1929_1933del NM_002168.3:c.418C>T	NP_001036214.1:p.Glu643AspfsTer16 NP_002159.2:p.Arg140Trp	0.0088	0.0054 0.0333	68 289	20 188	43 114	7747 9104	3675 5643	frameshift missense	X 15	123197803 90631935	GAGTTC	G
30 30	PHF6 RUNX1	NM_032458.2:c.827A>T NM 001754.4:c.58+5G>A	NP_115834.1:p.Lys276Ile	0.0253 0.0026	0.024 0.0026	129 26	61 11	79 22	5090 10190	2544 4212	missense splicing	X 21	133549143 36421134	A	T
30	RUNX1 RUNX1	NM_001754.4:c.58+5G>A NM_001754.4:c.611G>A	NP_001745.2:p.Arg204Gln	0.0026	0.0026	26 72	36	50	10190		splicing missense and splicing	21	36231773	c	T
														CGGCGGCTGTGGT	
30 31	SRSF2 PPM1D	NM_003016.4:c.284_307del NM_003620.3:c.1459dup	NP_003007.2:p.Pro95_Arg102del NP_003611.1:p.His487ProfsTer2	0.4188	0.3807	1596 173	718	310 83	3811 7715	1886 2549	inframe deletion frameshift	17	74732935 58740553	GTGAGTCCGGGG A	C AC
31	SF3B1	NM_012433.2:c.2294A>G	NP_036565.2:p.Tyr765Cys	0.0296	0.0332	170	48	75	5737	1445	missense	2	198266542	т	с
31	TET2	NM_001127208.2:c.1948del	NP_001120680.1:p.His650llefsTer50	0.0393	0.0424	184	51	89	4678	1202	frameshift	4	106157046	AC	A

31	TP53	NM 000546.5:c.783-1G>A		0.003	0.0032	19	7	18	6379	2203	splicing	17	7577156	С	т
31	TP53	NM 000546.5:c.725G>C	NP 000537.3:p.Cys242Ser	0.0043	0.0047	32	14	26	7422	2953	missense	17	7577556	С	G
31	TP53	NM_000546.5:c.832C>T	NP_000537.3:p.Pro278Ser	0.0043	0.0051	21	6	17	4905	1176	missense	17	7577106	G	A
31	TP53	NM 000546.5:c.584T>C	NP 000537.3:p.Ile195Thr	0.0039	0.0081	16	7	13	4127	868	missense	17	7578265	А	G
31	TP53	NM_000546.5:c.524G>A	NP_000537.3:p.Arg175His	0.0447	0.0376	181	30	105	4049	798	missense	17	7578406	С	т
31	TP53	NM_000546.5:c.640C>G	NP_000537.3:p.His214Asp	0.045	0.0495	263	71	114	5847	1434	missense	17	7578209	G	с
32	GATA2	NM_032638.4:c.917G>A	NP_116027.2:p.Trp306Ter	0.0011	0.0008	13	5	10	11544	6412	nonsense	3	128202803	с	т
32	NPM1	NM_002520.6:c.860_863dup	NP_002511.1:p.Trp288CysfsTer12	0.0067	0.0049	69	11	57	10256	2239	frameshift	5	170837543	с	CTCTG
32	TET2	NM_001127208.2:c.3009G>A	NP_001120680.1:p.Trp1003Ter	0.0029	0.0032	25	10	19	8533	3135	nonsense	4	106158108	G	A
32	TP53	NM_000546.5:c.407A>T	NP_000537.3:p.Gln136Leu	0.0028	0.0026	20	7	18	7207	2655	missense	17	7578523	т	A
33	ASXL1	NM_015338.5:c.2785G>T	NP_056153.2:p.Glu929Ter	0.3758	0.3748	2733	1453	404	7272	3877	nonsense	20	31023300	G	Т
33	TP53	NM_001126113.2:c.1037C>G	NP_001119585.1:p.Ser346Trp	0.0028	0.0034	11	5	10	3955	1451	missense	17	7576541	G	С
34	DNMT3A	NM_022552.4:c.26454C>T	NP_001307822.1:p.Arg882Cys	0.1128	0.1027	1199	249	259	10627	2424	missense	2	25457243	G	A
34	PPM1D	NM_003620.3:c.1469_1470dup	NP_003611.1:p.Asn491Ter	0.0088	0.0065	108	21	57	12241	3210	frameshift	17	58740563	т	TTG
34	U2AF1	NM_006758.2:c.101C>T	NP_006749.1:p.Ser34Phe	0.0997	0.0957	799	54	188	8013	564	missense	21	44524456	G	A
35	ASXL1	NM_015338.5:c.2773C>T	NP_056153.2:p.Gln925Ter	0.043	0.0434	592	233	186	13770	5364	nonsense	20	31023288	С	Т
35	CUX1	NM_181552.3:c.1756G>T	NP_853530.2:p.Val586Leu	0.0047	0.0048	49	12	39	10400	2500	missense	7	101840447	G	т
35	CUX1	NM_181552.3:c.3446T>C	NP_853530.2:p.Phe1149Ser	0.0394	0.0417	465	140	151	11817	3355	missense	7	101877344	т	С
35	SETBP1	NM_015559.2:c.2620G>C	NP_056374.2:p.Asp874His	0.0014	0.0013	40	21	31	28537	15799	missense	18	42531925	G	С
35	SF3B1	NM_012433.2:c.1998G>T	NP_036565.2:p.Lys666Asn	0.2167	0.2109	2028	503	371	9359	2385	missense	2	198267359	с	A
36	TP53	NM_000546.5:c.733G>A	NP_000537.3:p.Gly245Ser	0.0029	0.0019	22	6	15	7718	3195	missense	17	7577548	С	Т
37	U2AF1	NM_006758.2:c.101C>T	NP_006749.1:p.Ser34Phe	0.0159	0.0097	83	5	53	5212	516	missense	21	44524456	G	A
38	PPM1D	NM_003620.3:c.1563del	NP_003611.1:p.Met521llefsTer18	0.1048	0.1017	559	47	149	5334	462	frameshift	17	58740657	TG	Т
38	SETBP1	NM_015559.2:c.2608G>A	NP_056374.2:p.Gly870Ser	0.0273	0.0247	636	265	250	23270	10738	missense	18	42531913	G	A
38	TET2	NM_001127208.2:c.2746C>T	NP_001120680.1:p.Gln916Ter	0.0019	0.0025	17	6	12	8936	2411	nonsense	4	106157845	С	т
38	TET2	NM 001127208.2:c.5448 5449insAA	NP 001120680.1:p.His1817AsnfsTer4	0.0961	0.0868	625	81	163	6505	933	frameshift	4	106197114	т	TAA

#### Supplementary Table 4. Sensitivity and specificity of NGS mutational status prior to conditioning for predicting MDS patient relapse at 24 months.

Conditioning treatment	10-gene panel					26-gen	e panel	10 gene l	high VAF	10-gene panel + poor cytogenetics		
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity		
RIC	0.67	1.00	0.78	0.44	0.67	0.44	0.44	1.00	0.89	0.89		
MAC	1.00	0.64	1.00	0.29	1.00	0.36	0.50	0.79	1.00	0.36		
RIC + MAC	0.73	0.78	0.82	0.35	0.73	0.39	0.45	0.87	0.91	0.57		

 Image: Intermediate
 Image: Intermediate

10-gene high VAF (variant allele frequency): at least one variant detectable at VAF  $\geq$  2.5% in the 10-gene panel 10-gene panel + poor cytogenetics: 10-gene panel plus patients with poor risk cytogenetics present prior to hematopoietic stem cell transplantation

### SUPPLEMENTARY REFERENCES

1. Alboukadel Kassambara MK: survminer: Drawing Survival Curves using 'ggplot2'. R package version 0.4.3., 2018

2. Gray B: cmprsk: Subdistribution Analysis of Competing Risks. R package version 2.2-7, 2014

3. Kolde R: pheatmap: Pretty Heatmaps. R package version 1.0.10, 2018

4. Wickham H: Reshaping data with the reshape package. Journal of Statistical Software 21, 2007

5. Wickham H: Ggplot2 : elegant graphics for data analysis. New York, Springer, 2009

6. Wickham H: stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.0, 2019

7. Xie Y: Dynamic documents with R and Knitr (ed Second edition.). Boca Raton, CRC Press/Taylor & Francis, 2015

8. Zhang Z: Reshaping and aggregating data: an introduction to reshape package. Ann Transl Med 4:78, 2016