4599

Methylation Biology of a Blood-Based MDS Risk Stratification Test

ASH 2023
December 9-12, 2023
San Diego, CA, USA

Alvin Shi, PhD,[1,*] Abdullah M. Ali, PhD,[2,*] Gordon Cann, PhD,[1] Rita Shaknovich, MD, PhD,[1] Oliver Venn, PhD,[1] Azra Raza, MD,[2,†] Qinwen Liu, PhD[1,†]
*Joint lead authors; †Joint senior authors; [1]GRAIL, LLC, Menlo Park, CA, USA; [2]Edward P. Evans Myelodysplastic Syndrome Center and Department of Medicine, Columbia University Medical Center, New York, NY, USA.

INTRODUCTION

- Myelodysplastic syndromes (MDS) consist of a heterogeneous group of myeloid malignancies with 10,000 new cases diagnosed annually in the US and a 5-year survival of 36.9%, according to the Surveillance, Epidemiology, and End Results database[1,2]
- During MDS treatment, risk stratification is critical for optimal decision making. The current gold-standard for risk stratification is the 2012 revised International Prognostic Scoring System (IPSS-R). However, IPSS-R is limited by the requirement of bone marrow (BM) aspirate and lack of accurate survival prediction for all patients[3,4]
- A promising alternative for MDS risk stratification is methylation-based profiling. MDS-related somatic mutations often affect key epigenetic regulators (e.g., *TET2*, *DNMT3A*, *IDH1*, *IDH2*, and *WT1*), which leads to downstream epigenetic changes[5]
- GRAIL's blood-based, cell-free DNA (cfDNA) targeted methylation (TM) assay has previously demonstrated promise as a non-invasive approach for multi-cancer early detection.[6] The information from the TM assay may be used in other applications, such as predicting risk stratification with customized machine learning-based classifiers
- We previously demonstrated that an MDS prognostic classifier trained with methylation-based features from either BM whole-genome bisulfite sequencing (WGBS) or serum TM sequencing could stratify risk with comparable performance relative to IPSS-R.[7] Here, we provide additional information regarding the methylation signatures associated with differential survival outcomes, as well as further characterize survival stratification performance of the classifiers relative to IPSS-R

OBJECTIVE

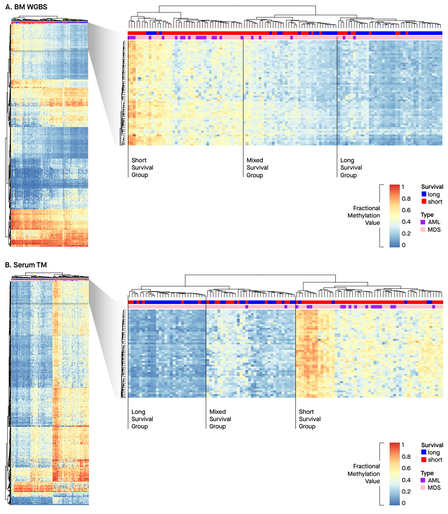
- Elucidate the biological underpinnings of the methylation-based MDS prognostic classifier and further evaluate its survival stratification performance

KEY RESULTS: BLOOD-BASED METHYLATION PROFILING RECAPITULATES KNOWN ASPECTS OF MDS BIOLOGY AND OFFERS A POTENTIAL ALTERNATIVE TO IPSS-R

Analysis of Differentially Methylated Regions (DMRs) Reveals Aspects of MDS Biology

- Significantly different DMRs between long (> 3 years) and short (< 3 years) survivor groups were identified after multiple hypothesis correction[8,9]
  - BM WGBS: 7742 / 2,812,497 (0.27%)
  - Serum TM: 14,093
- Hierarchical clustering of the top 500 DMRs (as ranked by adjusted p-value) demonstrated clustering into 3 distinct groups for both BM WGBS and serum TM: a predominantly short survivor group, a mixed group, and a long survivor group. The DMRs associated with short survivors were predominantly hypermethylated (**Figure 1A and B**)

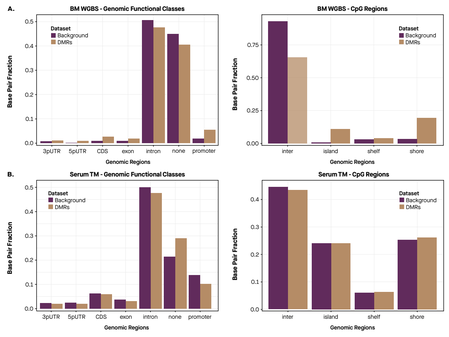
Figure 1. Top DMRs From BM WGBS and Serum TM Distinguish Survivorship Groups.



Clustered heatmaps of fractional methylation values (beta-values) in the top 500 survival-associated DMRs for (**A**) BM WGBS and (**B**) serum TM.
AML, acute myeloid leukemia; BM, bone marrow; DMR, differentially methylated region; MDS, myelodysplastic syndromes; TM, targeted methylation; WGBS, whole-genome bisulfite sequencing.

- The survival-associated DMRs in BM WGBS were associated with promoter, CpG island, and CpG shore regions; whereas, survival-associated DMRs in serum TM did not show any significant enrichment relative to existing panel regions, which are already enriched in cancer-associated functional regions (**Figure 2A and B**)

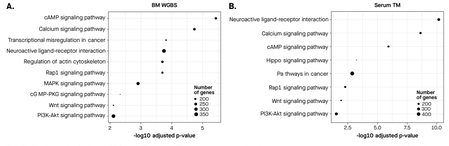
Figure 2. BM WGBS and Serum TM DMRs Were Differentially Enriched in Specific Genomic Regions.



DMR enrichment, in terms of total number of base pairs covered by DMRs, across genomic functional classes and CpG island/shore/shelf in (**A**) BM WGBS and (**B**) serum TM relative to background. BM background refers to the whole genome, and serum background refers to TM panel regions.
BM, bone marrow; CDS, coding sequence; DMR, differentially methylated region; TM, targeted methylation; UTR, untranslated region; WGBS, whole-genome bisulfite sequencing.

- For both BM WGBS and serum TM DMR data, pathway analysis showed significant enrichment of several pathways containing genes linked to top DMRs (**Figure 3A and B**)
- Enriched pathways included cAMP, calcium, Rap1, and Wnt signaling, which have been previously implicated in MDS progression[11-14] (**Figure 3A and B**)

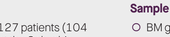
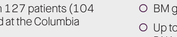
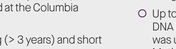
Figure 3. Both BM WGBS and Serum TM DMR Data Had Enrichment of MDS-Related Pathways.



Top DMRs for both (**A**) BM WGBS and (**B**) serum TM were annotated by the closest gene and used to perform KEGG pathway enrichment analysis. The FDR-corrected q-value is reported for selected pathways.
BM, bone marrow; DMR, differentially methylated region; FDR, false discovery rate; KEGG, Kyoto Encyclopedia of Genes and Genomes; MDS, myelodysplastic syndromes; TM, targeted methylation; WGBS, whole-genome bisulfite sequencing.

- The survival-associated DMRs in BM WGBS were used to identify driver transcription factors and genes with HOMER.[15] We identified several driver genes that have been previously associated with MDS to AML progression, such as EWS/FLI-1,[16] ERG,[17] Pu.1,[18] and RUNX1[19] (**Figure 4**)

Figure 4. HOMER Enrichment Analysis of Driver Transcription Factors in BM WGBS DMRs Revealed Potential Contributors to Differential Survival Outcomes of the MDS/AML Cohort.

Rank	Motif	Name	P-value	q-value (Benjamini)
1		PU.1	1e-24	<0.00001
2		REST-NRSF	1e-24	<0.00001
3		ERG	1e-24	<0.00001
4		Fli1	1e-19	<0.00001
5		ETS1	1e-19	<0.00001
7		EWS: ERG-fusion	1e-15	<0.00001
9		EWS: FLI1-fusion	1e-14	<0.00001
15		RUNX1	1e-13	<0.00001

Driver enrichment as reported by HOMER in BM WGBS DMRs associated with survival. Motif represents the binding motif of the transcription factor. p-value represents the significance of enrichment relative to background.
AML, acute myeloid leukemia; BM, bone marrow; DMR, differentially methylated region; MDS, myelodysplastic syndromes; WGBS, whole-genome bisulfite sequencing.

BM WGBS and Serum TM Methylation Data Can Be Utilized to Create Methylation-Based Classifiers

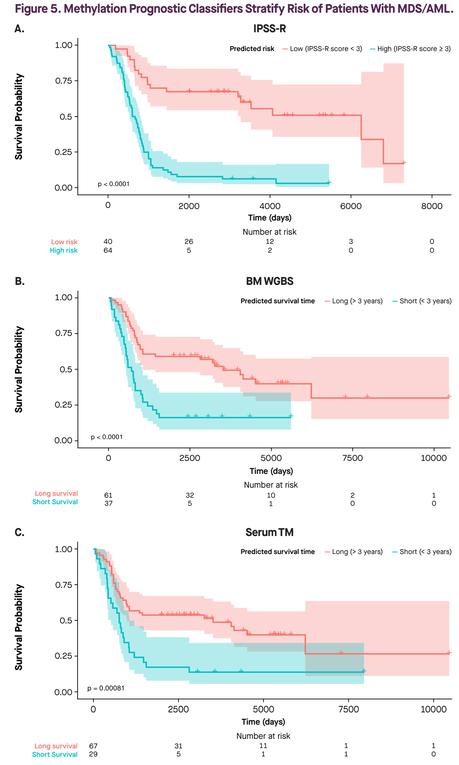
- The survival-associated DMRs identified from BM WGBS and serum TM sequencing were used to create machine learning-based prognostic classifiers (i.e., the methylation prognostic classifiers) that performed on par with IPSS-R when evaluated using concordance indices (**Table 3**)
- The methylation prognostic classifiers stratified predicted survival of patients with MDS/AML into short (< 3 years) or long (> 3 years) survival groups (**Figure 5**)

Table 3. Concordance Indices for IPSS-R and Methylation Prognostic Classifiers Evaluated With BM WGBS and Serum TM MDS/AML Patient Samples

	BM WGBS		Serum TM	
Survival Model	IPSS-R	Methylation Prognostic Classifier	IPSS-R	Methylation Prognostic Classifier
Concordance Index (95% CI)[a]	0.71 (0.64-0.77)	0.69 (0.61-0.77)	0.70 (0.63-0.78)	0.68 (0.60-0.75)

[a]95% CIs generated by bootstrapping.
AML, acute myeloid leukemia; BM, bone marrow; CI, confidence interval; IPSS-R, revised International Prognostic Scoring System; MDS, myelodysplastic syndromes; TM, targeted methylation; WGBS, whole-genome bisulfite sequencing.

Figure 5. Methylation Prognostic Classifiers Stratify Risk of Patients With MDS/AML.



Kaplan-Meier curves and log-rank test results for the (**A**) IPSS-R high-risk vs low-risk stratification, and the (**B**) BM WGBS and (**C**) serum TM methylation prognostic classifiers' survival predictions at 80% specificity cutoff.
AML, acute myeloid leukemia; BM, bone marrow; IPSS-R, revised International Prognostic Scoring System; MDS, myelodysplastic syndromes; TM, targeted methylation; WGBS, whole-genome bisulfite sequencing.

CONCLUSIONS

- Analysis of BM WGBS and serum TM sequencing yields insights into aspects of MDS biology, including key biological pathways and drivers previously implicated in progression to AML
- Methylation-based prognostic classifiers were trained and evaluated on either BM WGBS or serum TM features. These classifiers were able to demonstrate comparable performance to IPSS-R in predicting binarized short (< 3 years) and long (> 3 years) overall survival as measured in terms of concordance index
- Liquid biopsy utilizing targeted methylation of cfDNA offers an alternative to currently available methods (e.g., bone marrow biopsy) for MDS risk stratification
- Future studies with an independent hold-out cohort are planned to improve and validate performance of the presented liquid biopsy approach using TM sequencing

METHODS

Patient Cohort

- Biobanked BM aspirates and matched serum samples from 127 patients (104 MDS, 23 secondary acute myeloid leukemia [AML]) treated at the Columbia University Medical Center MDS center were used
- **Table 1** displays the cohort characteristics stratified by long (> 3 years) and short (< 3 years) overall survival times

Table 1: Patient Characteristics of Short vs Long Survivor Groups.

	Short Survivor Group (< 3 years)	Long Survivor Group (> 3 years)
N (% of total)	76 (59.8%)	51 (40.2%)
Age (years)	75.2 ± 8.58	67.5 ± 14.4
Gender	59.2% male	58.8% male
Survival (years)	1.7 ± 0.76	10.3 ± 5.16
Time between diagnosis and collection (years)	0.94 ± 0.65	4.26 ± 4.07
Blast cell %	16.1 ± 19.0	1.45 ± 1.59
Platelets (10^9/L)	105 ± 145	201 ± 237
Hemoglobin (g/dL)	9.47 ± 1.48	10.4 ± 2.36

Values are reported as mean ± standard deviation unless otherwise noted.

Sample Processing

- BM genomic DNA samples (gDNA) were sheared using a Covaris E220
- Up to 75 ng of sheared gDNA was bisulfite converted using a Zymo EZ DNA Methylation-Lightning™ MagPrep Kit. The bisulfite-converted gDNA was used to prepare sequencing libraries using a custom-fill IDT xGen™ Methylation-Sequencing DNA Library Preparation Kit on a Hamilton Star liquid handler
- WGBS was done on a NovaSeq 6000 at ~30x mean coverage per sample
- cfDNA from serum samples was extracted and subjected to bisulfite conversion, enrichment, and TM sequencing through a previously described GRAIL protocol (Gallerì[6], GRAIL, LLC).[7] The TM panel design covers 17.2 Mb and 11.6 million cancer-associated CpGs in the human genome

Data Processing

- Sample demultiplexing and sequencing data metrics were generated using GRAIL internal pipelines
- Clinical and sequencing data of evaluable samples were incorporated into TidyData tables for use in secondary analyses

Quality Control (QC)

- Evaluable samples subjected to WGBS or TM sequencing had ≥ 99% conversion of cytosine residues
- Multiple QC checks were used to eliminate potential sources of confounders
- GRAIL in-house contamination and concordance pipelines were used to filter out potentially problematic samples prior to final analysis (**Table 2**)

Table 2: QC Filtering of BM WGBS and Serum TM Samples.

	BM WGBS Samples (N = 127)	Serum TM Samples (N = 127)
Library prep dropout	1 (0.7%)	0 (0%)
Genotype-based contamination	3 (2.4%)	8 (6.3%)
Genotype concordance swaps	3 (2.4%)	3 (2.4%)
WBC serum contamination	0 (0%)	3 (2.4%)
Total removed	7 (5.5%)	14 (11%)
Final study population	120 (94.5%)	113 (89%)
Paired samples	109 (86%)	109 (86%)

BM, bone marrow; TM, targeted methylation; WBC, white blood cell; WGBS, whole-genome bisulfite sequencing.

Differentially Methylated Region (DMR) Calling, CpG Island Hypermethylation Signature Identification, and KEGG Enrichment Analysis

- DMR calling was performed using a beta-binomial model with arcsine link function for calling DMRs from a predefined list of regions.[8] Parameter estimation is based on transformed data with a generalized least square approach. False discovery rate was controlled using the Benjamini-Hochberg procedure[9]
- The top 500 DMRs associated with short and long survival were extracted based on adjusted p-values. Hierarchical clustering was performed using complete linkage with euclidean distance with the 'NMF' R package
- CpG island and closest gene annotations for DMRs were generated using a custom R script
- The presence of a CpG island hypermethylation signature in our survival-associated DMRs was investigated by plotting the difference in average fractional methylation values (beta-values) between short and long survivor groups as a histogram
- Biological enrichment analysis via the KEGG database of biological pathways was performed by annotating each DMR with the closest gene and conducting a hypergeometric test for enrichment against background[10]

HOMER Driver Calling

- A bed file consisting of the BM WGBS DMRs was used as input to the Hypergeometric Optimization of Motif EnRichment (HOMER)[15] 'findMotifs.pl' function with default settings and hg19 background as reference to identify driver transcription factors/genes

Methylation Classifier Training and Evaluation

- The ability of methylation-based features to predict binary short (< 3 years) vs long (> 3 years) overall survival in a subset of patients with MDS (using 98 BM and 96 serum MDS patient samples) was compared to IPSS-R as a benchmark
- Two methylation-based classifiers were created; one was trained with BM WGBS methylation data and the other with serum TM data. The entire beta-value matrix (2,812,497 1 kb windows for BM WGBS) was used as input for a Principal Component (PC) Random Forest classifier to predict binarized < 3 / > 3 year overall survival
- Classification performance evaluation and hyperparameter (# of PCs, # of Trees) optimization were conducted using nested 6-fold cross-validation with either BM WGBS samples or serum TM samples for the respective classifier. Cross-validated results were used to calculate concordance indices using the R glmnet package
- To quantify classification performance of IPSS-R, a logistic regression model was used to predict survival on IPSS-R raw scores, and this score was then used to calculate concordance indices using the R glmnet package. 95% confidence intervals for the concordance indices were generated by 1000 bootstrap permutations
- Kaplan-Meier curves detailing classifier performance were produced by setting a cutoff at the 80% specificity threshold for both the IPSS-R and methylation-based classifiers. The R survival package was then used to plot the Kaplan-Meier curves along with the log-rank test statistic

References

1. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov). SEER*Stat Database: Incidence - SEER Research Data. 2021.
2. Adès L, et al. *The Lancet*. 2014;383(9936):2239– 2252.
3. Greenberg PL, et al. *Blood*. 2012;120(12):2454–2465.
4. Ball S, et al. *Leukemia & Lymphoma*. 2022;63(6):1281–91.
5. Heuser, et al. *Seminars in Cancer Biology*. 2018;51(August): 170–79.
6. Klein EA, et al. *Ann Oncol*. 2021;32(9):1167-1177.
7. Shi A, et al. *Journal of Clinical Oncology* 2023 41:16_suppl. 7058.
8. Park Y, et al. *Bioinformatics*. 2015;32(10):1446–1453.
9. Benjamini Y and Hochberg Y. *Journal of the Royal Statistical Society*. Series B. 1995;289–300.
10. Kanehisa M, et al. *Nucleic Acids Research*. 2016;44(D1):D457–462.
11. Guillemin MC, et al. *JEM*. 2002; 196(10):1373–80.
12. Ishida D, et al. *Cancer Cell*. 2003;4(1):55–65.
13. Gruszka AM, et al. *Cells*. 2019; 8(11).
14. Lieu YK, et al. *PNAS*. 2022;119(1):e2111703119.
15. Heinz, S. et al. Mol. *Cell*. 2010; 38:576–589.
16. Torchia, EC, et al. *Mol. Cell. Biol*. 2007; 27:7918–7934.
17. Lee, WY, et al. *Genes Chromosomes Cancer*. 2022; 61: 399–411.
18. Will, B. et al. *Nat. Med*. 2015; 21:1172–1181.
19. Gaidzik, VI. et al. *Leukemia*. 2016; 30:2160–2168.

Disclosures

Study funded by GRAIL, LLC. RS is an employee of GRAIL, LLC. AS, GC, OV, and QL are employees of GRAIL with equity in the company and stock options in Illumina, Inc. OV has received research funding from GRAIL. AA has equity in TFC Therapeutics, consulted with VOR Biopharma and TF Therapeutics, has patents with VOR Biopharma and Brahma Therapeutics, and received research funding from VOR Biopharma and Actinium Pharmaceuticals. AR has equity in TFC Therapeutics, consulted and/or served an advisory role with TFC Therapeutics, GRAIL, Genzyme, Takeda, and Janssen, has patents with TFC Therapeutics, received research funding from Regeneron and Immuneel Therapeutics, received honoraria and/or other expenses paid from GRAIL, Epizyme and Taiho Oncology, and Rarecells.

Acknowledgements

This study was funded by GRAIL, LLC. We would like to acknowledge Tracy Nance, PhD and Yuefan Huang, PhD (GRAIL, LLC) for code review and Prescott Medical Communications Group (Chicago, IL) for editorial and graphic assistance.