7058

Methylation-Based Prediction of Myelodysplastic Syndrome Survival Outcomes

ASCO 2023
June 2–6, 2023
Chicago, IL, USA

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

INTRODUCTION

- Myelodysplastic syndromes (MDS) consists of a heterogeneous group of myeloid malignancies with 10,000 new cases diagnosed annually in the US and a 5-year SEER survival of 36.9%[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 aspirate and lack of accurate survival prediction for all patients[3,4]
- A promising alternative for MDS risk stratification is methylation-based profiling, as 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 contained in the TM assay may be used in other applications, such as predicting risk stratification using a custom classifier
- To discover prognostic methylation patterns contained in the TM panel, a methylation-based classifier can be trained to predict outcomes on a cohort of patients with differential survival

OBJECTIVE

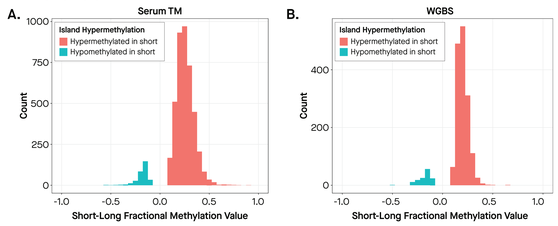
- Evaluate the potential of methylation-based features, as assessed by analysis of Serum TM and verified by Bone Marrow (BM) Whole-Genome Bisulfite Sequencing (WGBS), to stratify MDS survival relative to IPSS-R in a cohort of 127 patients with MDS (N=104) and secondary AML (N=23)

BLOOD-BASED METHYLATION PROFILING RECAPITULATES KNOWN ASPECTS OF MDS BIOLOGY AND OFFERS SIMILAR RISK-STRATIFICATION PERFORMANCE AS IPSS-R.

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

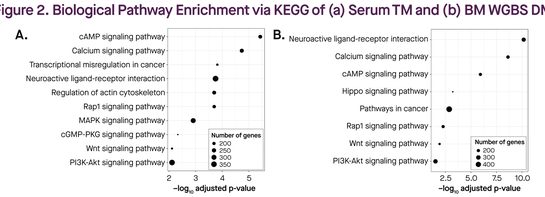
- Significantly different DMRs between long (>3 years) and short (<3 years) survivor groups were identified
  - Serum TM: 14,093 out of 103,142 (13.66%)
  - BM WGBS: 7,742 out of 2,812,497 (0.27%)
- CpG island hypermethylation was more prevalent in the short survivor group compared to the long group (**Figure 1**), across both Serum TM and BM WGBS DMRs, confirming the previous finding[11] that CpG island hypermethylation is associated with AML progression
  - The left-skew observed in the histogram suggests that CpG islands have higher methylation values in the short survivor group compared to the long survivor group
- In both Serum TM and BM WGBS data, significant enrichment was observed for several pathways, such as cAMP, calcium, Rap1, and Wnt signaling[12-14], all previously implicated in MDS progression (**Figure 2**)

Figure 1. Recapitulation of CpG Island Hypermethylation Signature in (a) Serum TM and (b) BM WGBS Data



Average fractional methylation values computed as the fraction of methylated reads over all reads for a particular DMR location. Short-Long denotes the subtraction of the long from the short fractional methylation values at each individual DMR. BM, bone marrow; DMR, differentially methylated region; TM, targeted methylation; WGBS, whole-genome bisulfite sequencing.

Figure 2. Biological Pathway Enrichment via KEGG of (a) Serum TM and (b) BM WGBS DMRs

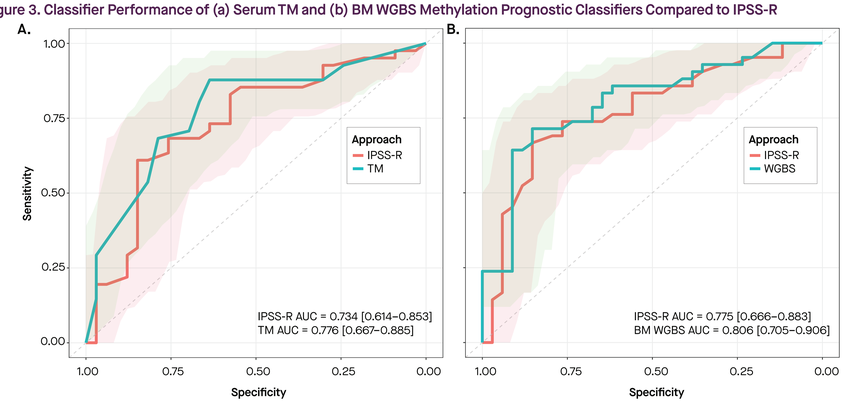


The FDR-corrected q-value of pathway enrichment is reported for selected pathways. BM, bone marrow; DMR, differentially methylated region; FDR, false discovery rate; KEGG, Kyoto Encyclopedia of Genes and Genomes; TM, targeted methylation; WGBS, whole-genome bisulfite sequencing.

Methylation Features Offer Similar Levels of Risk Prediction Performance as IPSS-R

- The ability of methylation-based features to predict binary short (<3 years) vs long (>3 years) overall survival using a methylation prognostic classifier was compared to the benchmark IPSS-R measure in a subset of 96 Serum and 98 BM MDS patients
- Both BM WGBS and Serum TM had higher AUC values than IPSS-R alone for predicting binarized <3/>3 year overall survival (**Figure 3**)

Figure 3. Classifier Performance of (a) Serum TM and (b) BM WGBS Methylation Prognostic Classifiers Compared to IPSS-R



ROC of methylation prognostic classifiers vs. IPSS-R logistic regression for (a) Serum TM and (b) BM WGBS. Classifier performance evaluation and hyperparameter optimization were conducted using 6-fold cross-validation. 95% Confidence Intervals for the ROC and AUC were derived from 2000 bootstrap replicates. AUC, area under the curve; BM, bone marrow; IPSS-R, 2012 revised International Prognostic Scoring System; ROC, receiver operator characteristic; TM, targeted methylation; WGBS, whole-genome bisulfite sequencing.

- A multivariable Cox regression indicated that methylation prognostic score is a significant predictor of survival even after accounting for IPSS-R, age, and sex (**Table 3**)

Table 3. Multivariable Cox Regression Shows That Both Methylation Prognostic Score and IPSS-R Are Significant Predictors of Survival

	Serum TM		BM WGBS	
	HR	p-value	HR	p-value
Methylation prognostic score	4.85	0.002	4.82	0.016
IPSS Score	1.30	0.003	1.29	0.003
Age	1.04	0.007	1.03	0.022
Sex	1.57	0.143	1.26	0.439

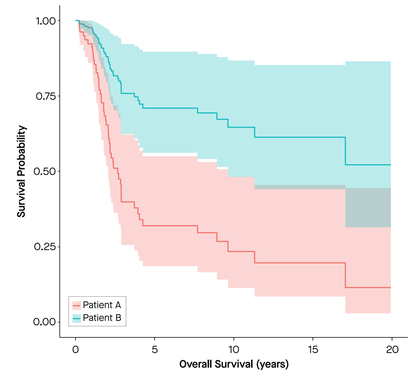
Methylation prognostic score predicted probability of the Serum TM or BM WGBS methylation prognostic classifier; IPSS score, raw IPSS-R score; sex, binarized coding for gender. BM, bone marrow; HR, hazard ratio; IPSS-R, 2012 revised International Prognostic Scoring System; TM, targeted methylation; WGBS, whole-genome bisulfite sequencing.

Serum TM Methylation Features Can Be Leveraged to Generate Personalized Survival Predictions

- The Serum TM-based methylation prognostic classifier was able to generate personalized survival[10] predictions using multivariable Cox regression. It differentiated between two held-out patients with similar IPSS-R scores but differing survival outcomes, and better predicted outcomes compared to IPSS-R (**Figure 4**)

Figure 4. Serum TM Methylation Features Can Be Used to Generate Personalized Predictions

	IPSS-R category	IPSS-R score	Gender	Age (years)	Overall survival (years)
Patient A	Low-risk	2	Male	82	2.64
Patient B	Low-risk	2	Male	73	11.08



Personalized survival curves predictions for 2 patients (A and B; held out from classifier training) generated from a multivariable Cox regression model with age, sex, and methylation prognostic score as covariates. IPSS-R, 2012 revised International Prognostic Scoring System.

METHODS

Patient Cohort

- Biobanked bone marrow aspirates and serum samples were used from 127 patients (104 MDS, 23 secondary AML) treated at the Columbia University Medical Center MDS center
- All samples were collected under the same specimen collection and storage protocol
- The cohort characteristics stratified by long (>3 years overall survival) and short (<3 years overall survival) are displayed in **Table 1**

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 and then 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
- Serum cfDNA samples were subjected to TM sequencing using an MCED assay (Galleri®, GRAIL, LLC)

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	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 percent %	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.

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 sequenced by TM and WGBS sequencing had ≥99% conversion of Cytosine residues
- Multiple QC checks were used to eliminate potential sources of confounders
- GRAIL in-house contamination, genotype concordance, and white-blood cell (WBC) contamination pipelines were used to filter out potentially problematic samples prior to final analysis (**Table 2**)

Table 2: QC Filtering for the Serum TM and BM WGBS Samples

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

Differential Methylation Region (DMR) Calling, CpG Island Hypermethylation Signature Identification, and KEGG Enrichment

- DMR was performed using a beta-binomial model with arcsine link function for calling DMRs from a predefined list of regions.[7] Parameter estimation is based on transformed data with a generalized least square approach. False discovery rate was controlled using the Benjamini-Hochberg procedure[8]
- CpG island and closest gene annotations for DMRs were generated using a custom R script
- The presence of 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[9]

Methylation Prognostic 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 96 Serum and 98 BM MDS patients was compared to IPSS-R as a benchmark
- To train a methylation-based classifier, the entire beta-value matrix (2,812,497 1kb windows for BM WGBS and 103,142 panel regions for Serum TM) was used as input to 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. Confidence intervals for AUROC and the ROC were generated using 2000 bootstrap samples
- To quantify binary classification performance of IPSS-R, a logistic regression model was used to predict survival on IPSS-R raw scores

Multivariable Cox Regression

- The output from the methylation prognostic classifier ("methylation prognostic score") was evaluated as a predictor of survival through a multivariable Cox regression with Serum TM and BM WGBS methylation prognostic score, IPSS-R score, age, and sex

Personalized Survival Curve Generation

- To investigate whether Serum TM methylation classifiers are better able to stratify risk in patients with similar IPSS-R scores but dramatically different outcomes, a personalized risk survival model[10] was built based on a multivariable Cox regression model with sex, age, and methylation prognostic score

CONCLUSIONS

- Methylation-based prognostic classifiers were trained and evaluated on either Serum TM or BM WGBS features. These classifiers were able to:
  - Recapitulate known aspects of MDS biology, including key biologically pathways and CpG island hypermethylation previously implicated in progression to AML
  - Demonstrate comparable performance to IPSS-R in predicting binarized short (<3 years) and long (>3 years) overall survival
- Both Serum TM- and BM WGBS-based methylation prognostic scores were significantly associated with survival, even after accounting for IPSS-R score, age, and gender
- Methylation-based features can be used to generate personalized risk predictions that capture additional prognostic information over IPSS-R and may be able to provide more efficient predictions of clinical outcomes for some individuals
- Serum TM offers a less invasive approach to MDS risk stratification than currently available methods
- Future studies with an independent hold-out cohort are planned to validate performance

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. Park Y, et al. *Bioinformatics*. 2015;32(10):1446-1453. 8. Benjamini Y and Hochberg Y. *Journal of the Royal Statistical Society. Series B*. 1995;289-300. 9. Kanehisa M, et al. *Nucleic Acids Research*. 2016;44(D1):D457-462. 10. Bersanelli M, et al. *JCO*. 2021; 39 (11): 1223-33. 11. Spencer DH, et al. *Cell*. 2017;168(5):801-16.e13. 12. Guillemin MC, et al. *JEM*. 2002; 196(10):1373-80. 13. Ishida D, et al. *Cancer Cell*. 2003;4(1):55-65. 14. Gruszka AM, et al. *Cells*. 2019; 8(11).

Disclosures

Study funded by GRAIL, LLC. Please see abstract online for author disclosure information.

Acknowledgements

This study was funded by GRAIL, LLC. We would like to acknowledge Tracy Nance, Ph.D. (GRAIL LLC) for code review and Kristi Whitfield from PosterDocs (Oakland, CA) for graphic assistance.