



The 65th ASH Annual Meeting Abstracts

POSTER ABSTRACTS

617.ACUTE MYELOID LEUKEMIAS: BIOMARKERS, MOLECULAR MARKERS AND MINIMAL RESIDUAL DISEASE IN DIAGNOSIS AND PROGNOSIS**Identifying a Relapse Signature in CBF-B-MYH11 Acute Myeloid Leukemia**

Jack H Peplinski, BS¹, Logan K Wallace², Rhonda E Ries, MA¹, Benjamin J Huang, MD³, Amanda R Leonti, MS², Danielle C Kirkey, MD², Xiaotu Ma, PhD⁴, Scott N Furlan, MD^{2,5}, Soheil Meshinchi, MDPhD⁶

¹Translational Sciences and Therapeutics, Fred Hutchinson Cancer Center, Seattle, WA

²Translational Science and Therapeutics, Fred Hutchinson Cancer Center, Seattle, WA

³Department of Pediatrics, University of California, San Francisco, San Francisco, CA

⁴St. Jude Children's Research Hospital, Memphis, TN

⁵Department of Pediatrics, University of Washington, Seattle, WA

⁶Translational Sciences and Therapeutics, Fred Hutchinson Cancer Research Center, Seattle, WA

Acute myeloid leukemia (AML) is an aggressive cancer characterized by many mutational and cytogenetic subtypes with varied risk strata. The CBF-B-MYH11 fusion is a subtype associated with favorable disease outcomes in pediatric patients despite a relapse rate of over 30%. While overall survival remains relatively high, above 65%, more work can be done to improve outcomes. Considering the frequent rate of relapse, identifying a genetic signature that can predict relapse at diagnosis is vital to outcome improvement. Through k-means clustering, differential expression analyses, and recursive feature elimination, we were able to identify twenty genes that were strongly associated with relapse risk in the CBF-B-MYH11 fusion at the time of diagnosis.

All analysis was performed in R, version 4.3.0, including all listed packages. Ribodepleted RNA sequencing data was generated on 3023 patients and included a total of 119 diagnostic CBF-B-MYH11; 64 were considered non-relapse and 55 were considered relapse. A Poly(A)-seq validation cohort was generated on 495 patients included 60 diagnostic CBF-B-MYH11. Of these patients, 42 were considered non-relapse and 18 were considered relapse.

Unsupervised clustering was performed using a k-means algorithm in the *ComplexHeatmap* package which resulted in three "clusters" of 48, 54, and 17 samples with distinct variation in z-scores when the most-variable 500 genes were plotted. These clusters were labeled 1, 2, and 3 respectively.

Survival analysis was then performed using Kaplan-Meier method and curves were stratified by cluster. Clusters 1 and 2 had nearly identical event-free survival rates near 60%, yet cluster 3 had an event-free survival rate of less than 12% ($p < 0.001$). Clusters also differentiated in overall survival, with cluster 3's survival rate near 50% compared to 80% and 95% in cluster 1 and 2 respectively ($p < 0.001$).

To identify the underpinnings of aggressive relapse in CBF-B-MYH11, differential expression analysis was performed using *DESeq2*, contrasting cluster 3 against the other clusters, yielding a list of 6,204 differentially expressed genes (p -adjusted cutoff of 0.05, using a Benjamini-Hochberg procedure to reduce false discovery rate).

The list of differentially expressed genes were input into a feature elimination machine learning model using the *RandomForest* package. This model identified 20 genes that were most predictive of relapse risk with diminishing returns when the number of genes was increased. The list of 20 genes in order of predictive influence includes: CBWD3, DRAP1, GNL3LP1, MAF1, GRAMD1C, MAGED2, TICAM1, MPP1, NBPF11, RAB13, RN7SL329P, HAL, TLK2P2, CD151, XRCC6, HIST1H2AM, MSANTD3-TMEFF1, CDCA4, LRRC8A, and IPCEF1.

CBWD3 was the most predictive gene and is under-expressed in relapsing patients in CBF-B-MYH11. When survival analysis was performed on CBWD3, above-median expressors had an event-free survival probability of near 75% and below-median expressors had an event-free survival probability of just 30% with a combined p -value of less than 0.0001.

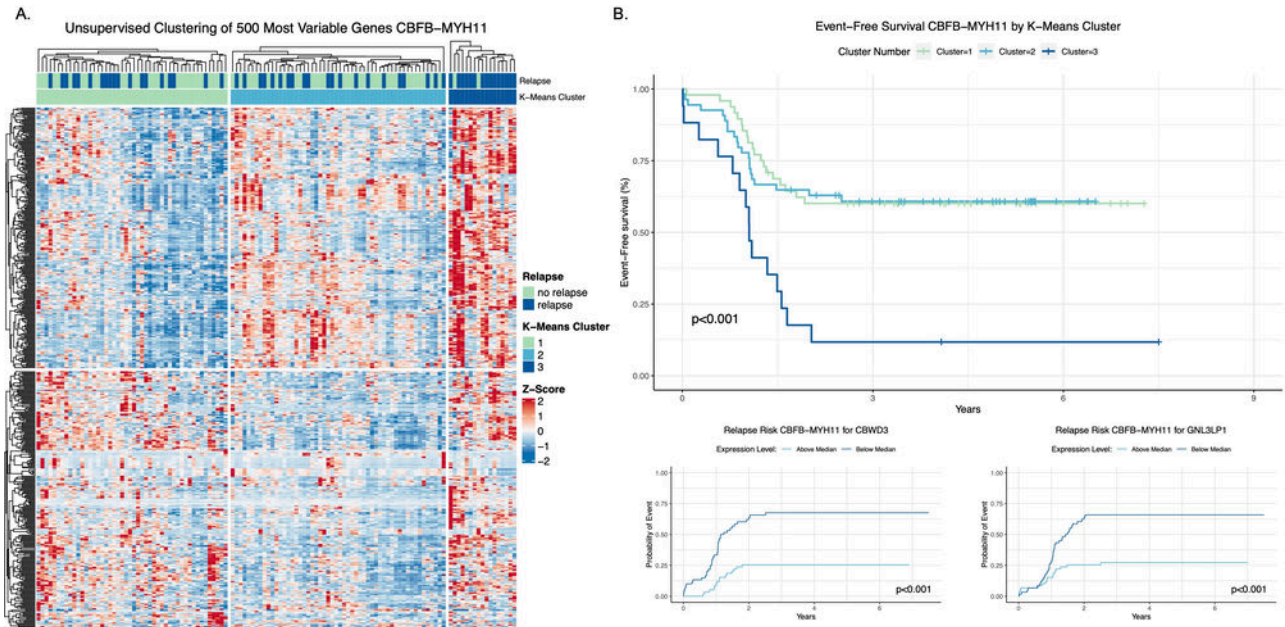
To validate our findings, we used a separate CBF-B-MYH11-positive cohort that had been sequenced using Poly(A)-seq, which is sequenced at a lower depth than ribodepleted RNA-seq data. We saw similar event-free survival estimates in both cohorts for some genes-including CBWD3 and GNL3LP1.

The 20 genes of interest were then input into a Random Forest model, using 70% of the CBF-B-MYH11 cohort as a training subset. When the training model was optimized, relapse was correctly predicted in 32 of 39 samples (82%) and non-relapse

was correctly predicted in 40 of 45 samples (89%). Future analyses will aim to test the efficacy of the model on the testing subset and the Poly(A) cohort for validation. A LASSO-Cox regression will also be used to calculate variable coefficients for risk determination in a clinical setting.

While the fusion group CBF-B-MYH11 is associated with improved survival outcomes in pediatric AML, relapse rate remains high. Through k-means clustering, differential expression analysis, and recursive feature elimination, we've identified a subset of genes that have predictive power in determining relapse risk. As our machine learning model approaches improve, a robust relapse signature may be identified and used to improve patient outcomes.

Disclosures No relevant conflicts of interest to declare.



A. Heatmap showing the 500 most-variable genes and subsequent k-means clustering of CBF-B-MYH11 patients divided into columns. B. Kaplan-Meier event-free survival curves for CBF-B-MYH11 patients by k-means cluster including relapse risk for strongly associating genes CBWD3 and GNL3LP1

Figure 1

<https://doi.org/10.1182/blood-2023-187637>

Downloaded from http://ashpublications.net/blood/article-pdf/142/Supplement_1/2955/2193638/blood-9557-main.pdf by guest on 24 May 2024