RESEARCH ARTICLE

eJHaem

British Society for Haematology

DNA methylation analysis improves the prognostication of acute myeloid leukemia

Hanie Samimi[1] | Isha Mehta PhD[2] | Thomas Roderick Docking PhD[3] |
Aamir Zainulabadeen[4] | Aly Karsan MD[3,*] (iD) | Habil Zare PhD[2,5,*] (iD)

[1] Department of Computer Science, Texas State University, San Marcos, Texas, USA

[2] Department of Cell Systems & Anatomy, The University of Texas Health Science Center, San Antonio, Texas, USA

[3] Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Research Centre, Vancouver, British Columbia, Canada

[4] Department of Computer Science, Princeton University, Princeton, New Jersey, USA

[5] Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, University of Texas Health Sciences Center, San Antonio, Texas, USA

Correspondence
Habil Zare, Department of Cell Systems & Anatomy, 7703 Floyd Curl Drive, San Antonio, TX 78229, USA.
Email: zare@uthscsa.edu

*Both the senior authors contributed equally to this work.

Abstract

Integration of orthogonal data could provide new opportunities to pinpoint the underlying molecular mechanisms of hematologic disorders. Using a novel gene network approach, we integrated DNA methylation data from The Cancer Genome Atlas ($n = 194$ cases) with the corresponding gene expression profile. Our integrated gene network analysis classified AML patients into low-, intermediate-, and high-risk groups. The identified high-risk group had significantly shorter overall survival compared to the low-risk group (p-value $\leq 10^{-11}$). Specifically, our approach identified a particular subgroup of nine high-risk AML cases that died within 2 years after diagnosis. These high-risk cases otherwise would be incorrectly classified as intermediate-risk solely based on cytogenetics, mutation profiles, and common molecular characteristics of AML. We confirmed the prognostic value of our integrative gene network approach using two independent datasets, as well as through comparison with European LeukemiaNet and LSC17 criteria. Our approach could be useful in the prognostication of a subset of borderline AML cases. These cases would not be classified into appropriate risk groups by other approaches that use gene expression, but not DNA methylation data. Our findings highlight the significance of epigenomic data, and they indicate integrating DNA methylation data with gene coexpression networks can have a synergistic effect.

1 | INTRODUCTION

Acute myeloid leukemia (AML), which is the most common acute leukemia among adults [1], has several subtypes that have various prognoses [2] depending on age, cytogenetic abnormalities [3], specific mutations [4], and other unknown risk factors (Note S1). Timely diagnostics of AML is critical for determining the best approach to clinical management [5].

We previously showed that coexpression gene network analysis is useful to identify subtle but consistent signatures of AML that otherwise would be difficult to pinpoint using conventional differential expression analysis [6]. In the current study, we tested the hypothesis

that integrating DNA methylation data into the model can be useful in detecting prognostic signatures, because more robust modules would be identified (Figure 1) [7,8]. To the best of our knowledge, this is the first study on systematic integrating of DNA methylation and transcriptome data in a single gene network. Using this approach, we developed a prognostic test based on data from The Cancer Genome Atlas (TCGA) consortium [9] ($n = 194$ cases), and we validated our findings using two independent AML cohorts.

Highlights

1. Incorporating DNA methylation into gene network analyses results in more robust and informative gene modules.
2. Using an integrative gene network approach, we identify high-risk AML cases that otherwise would be classified as intermediate-risk.

2 | METHODS

2.1 | Description of training and validation datasets

We used TCGA data to train our model (i.e., build our gene network and select the prognostic modules). We downloaded gene expression ($n = 179$) and DNA methylation ($n = 194$) data from the TCGA AML publication web page [9]. To validate our findings, we used two independent datasets: (1) the German Acute Myeloid Leukemia Cooperative Group (AMLCG) 1999 Cohort ($n = 562$, microarray) [10⁻13] and (2) the Beat AML RNA-Seq dataset ($n = 405$) [14].

The AMLCG cohort: We downloaded the gene expression profiles of 562 AML samples from Gene Expression Omnibus (series number GSE37642) [15]. We obtained the corresponding clinical data, including overall survival, life status, and risk category based on the European LeukemiaNet (ELN-2010) [16] from the authors.

The Beat cohort: We obtained the reads per kilobase of transcript per million mapped [17] (RPKM) values (i.e., the gene expression profile) and the survival data of the Beat AML cohort from the supplementary tables of the corresponding publication [14]. We transformed the RPKM values using natural logarithm [18]. In our survival analysis, we included only those 405 cases whose reported time from diagnosis to death is non-zero.

2.2 | Preprocessing data

Our approach can benefit from all DNA methylation data in constructing the network even if gene expression data are missing for some of the corresponding cases. We excluded genes and loci (i.e., DNA methylation probes) that did not pass the quality control criteria, showed too little variation, or had too weak correlation with overall patient survival. We used only the 24,649 loci that showed an absolute Pearson correlation of 0.2 or higher with overall survival. These loci correspond to 9377 genes, and the majority of these genes (95%) were associated with four loci at most (Figure S1). We then computed the effective DNA methylation level for each gene (Note S2).

2.3 | Gene network analysis

After data preprocessing, 12,535 genes were determined to have some relevance with overall survival based on either expression, or DNA methylation levels. To identify groups of genes that are associated with each other (i.e., gene modules), we built a network in which each node corresponds to one of these 12,535 genes (Figure 1). We computed the similarity between each gene pair based on both expression and DNA methylation levels. Specifically, for each gene pair, we computed (1) the
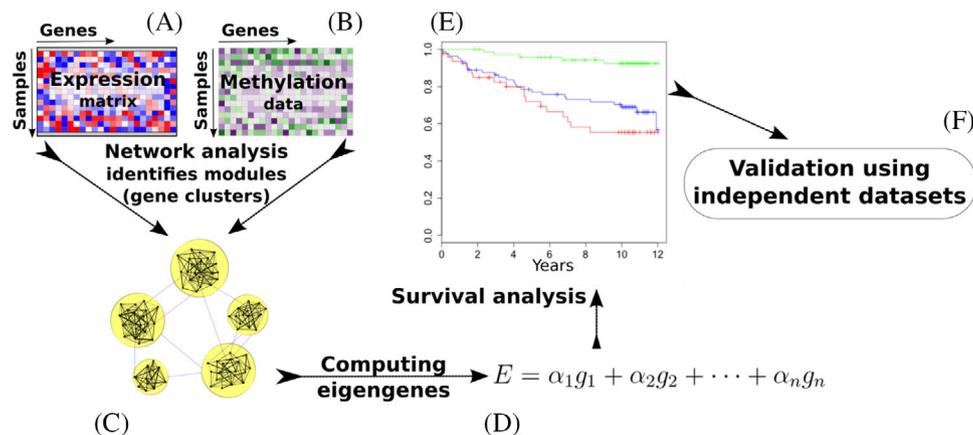
FIGURE 1 Schematic view of the methodology. The inputs include (a) a gene expression profile and (b) a DNA methylation profile. (c) We build an integrative network that models the association between individual gene pairs based on both expression and methylation data using Equation 1 (Supplementary Note 3). (d) For each module, we compute an eigengene as a weighted average of the expression of all genes in that module. (e) We use the eigengenes as robust biological signatures (biomarkers) to perform survival analysis. (f) The results are validated using independent cohorts of AML

Pearson correlation between their expression levels and (2) the Pearson correlation between their DNA methylation levels. We combined the resulting correlation values to derive the similarity (i.e., association) between each pair of genes (Note S3).

In our integrative gene network, each node represents a gene, and the edge between two genes (i.e., their connection) is weighted based on their similarity. We identified groups of similar genes (i.e., modules) using a hierarchical clustering approach [19] (Note S4). For each gene module, we then computed an eigengene based on a principal component analysis [6,20] (Note S5) and used these eigengenes to delineate the modules that are associated with overall survival of AML (Note S6).

2.4 | Comparison with other prognostication approaches

Gerstung et al approach provides continuous risk estimates [21]. To systematically compare this approach with our prognostic test, we discretized Gerstung et al risk estimates (i.e., we considered the first quartile low-risk, the second and third quartiles intermediate-risk, and the fourth quartile high-risk). We also compared our approach with European LeukemiaNet genetic risk stratification and the LSC17 score, which is based on a leukemic stem cell (LSC) signature [22] (Note S7).

2.5 | Data availability

All data analyzed during this study are included in the Supplementary Information files, which can be used to reproduce our results.

3 | RESULTS

3.1 | Integrative network analysis

Using our novel integrated network analysis (Figure 1, Methods), we identified 78 modules (clusters) of genes (Figure S4). The genes in each module are highly associated with each other with respect to their DNA methylation and gene expression patterns (Table S1). For each module, we computed an eigengene, which is a weighted average of the expression levels of the genes in that module (Note S5 and Table S2). Module 56 had the eigengene that showed the strongest correlation with overall survival (Pearson correlation $r = 0.3$), while Module 51 had the most anticorrelated eigengene ($r = 0.4$). In this correlation analysis, each module was considered individually. To determine whether a subset of modules together would be more useful for the prognostication of AML, we performed a survival analysis using Cox regression.

3.2 | Survival analysis

Among the participants in the TCGA cohort, the prevailing cytogenetic criteria classified 31 cases as low-risk, 92 as intermediate-risk, and 31

as high-risk cases (Figure 2a) [9]. To determine whether our integrative gene network analysis can improve prognostication, we used the 78 inferred eigengenes as covariates (i.e., potential prognostic features) in a penalized Cox regression analysis [23,24]. We found that the most associated subset of three modules with overall survival included Modules 46, 51, and 55 (Note S6). Some biological pathways are overrepresented in these three modules [26].

Module 46 has 19 genes (Table S1) with a significant overlap with the following pathways: (1) the immune system based on Reactome [27], (2) the signaling events mediated by class II histone deacetylases (HDAC) based on Pathway Interaction Database [28], and (3) the signaling by nerve growth factor [29] based on Reactome (hypergeometric test p-values = 0.001). The overlaps between these pathways and Module 46 include the following genes previously reported relevant to AML: *CAMK4* [30,31], *IL6ST* [32], *SOCS2* [33,34], *TUBB2A* [35⁻37], and *BCL2L11* [38]. While the expression of these genes anticorrelates with survival time, the overlaps also include *DUSP3* [39,40], which correlates with survival time.

Module 51 has 15 genes with a significant overlap with the post-translational protein modification (p-values = 0.001) and metabolism of proteins (p-values = 0.005) pathways. The overlaps include *PLAUR* (*CD87*) [41⁻44] with positive, and *PMM1* with negative, correlations with survival time, respectively.

Module 55 has 14 genes with a significant overlap with the following Reactome pathways: (1) apoptosis [45], (2) membrane trafficking [46], and (3) cellular responses to stress [47] (p-values = 0.01). The overlaps include *YWHAH* [48], *GPX1* [49], and *UBC*. The expression of these genes was negatively correlated with survival time.

To evaluate the significance of these selected modules in predicting overall survival, we fitted an *accelerated failure time* model to the selected eigengenes [50,51]. Based on this model, we predicted the expected survival time of each patient and then classified them into low-, intermediate-, and high-risk groups (Table S3). There was a significant difference between the survival times of the low-and high-risk groups that we identified (p-value $\leq 10^{-11}$, Figure 2c). When we stratified the patients into two age groups, the difference between the identified low- and high-risk groups was still significant in the group of patients who were diagnosed with AML after age 60 ($n = 74$, p-value ≤ 0.004), and also in the rest of relatively younger patients ($n = 92$, p-value $\leq 10^{-6}$, Figure S5). As expected, we identified more high-risk ($n = 19$) and fewer low-risk ($n = 5$) patients in the older group compared to the relatively younger group ($n = 7$ and $n = 22$, respectively). There were 15 males and 11 females in the high-risk group, which did not indicate a significant difference between the two genders (hypergeometric test p-value ≥ 0.2).

3.3 | Enhancement of current prognostication approaches

Because it is challenging to decide on the best treatment for the intermediate-risk group, we investigated whether combining our analysis with other classification approaches for AML would lead to
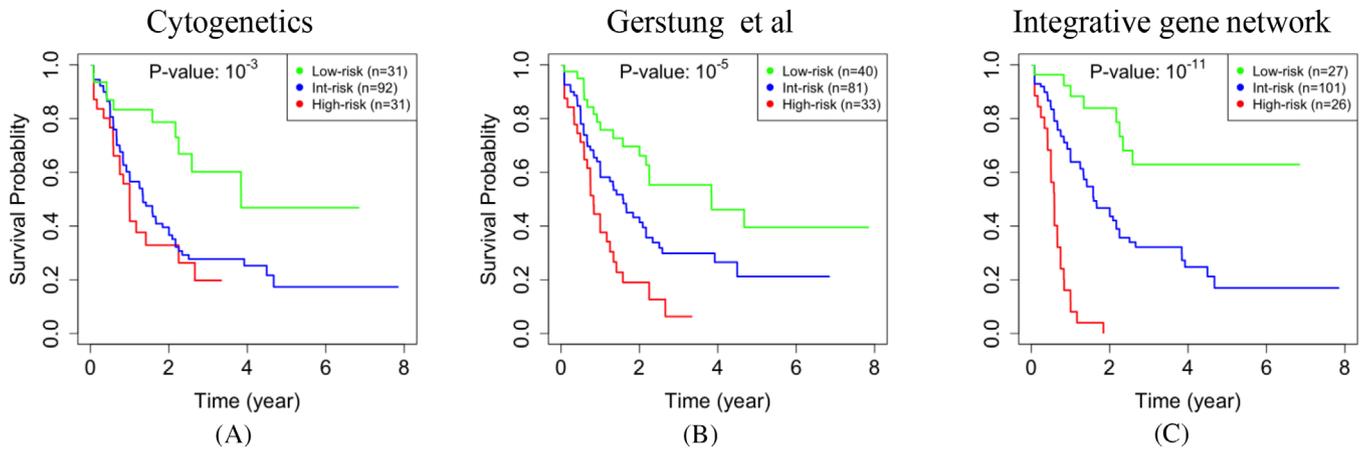
FIGURE 2 The Kaplan-Meier [25] (KM) curves for all cases in the TCGA-AML dataset. The log-rank *p*-values indicate that differences between the low-risk group (green) and the high-risk group (red) are statistically significant. Compared to cytogenetics criteria (**a**), the approach developed by Gerstung et al (**b**) led to a lower *p*-value (i.e., $10^{-5} < 10^{-3}$). Our gene network analysis, which includes gene expression and epigenome data but not mutation data, resulted in a *p*-value of 10^{-11} (**c**). In particular, all 26 cases that we identified as high-risk (i.e., 17% of the entire cohort) died within 2 years after diagnosis. Validation on two independent cohorts confirmed that the prognostic value of our analysis is not specific to the TCGA-AML dataset, which we used to train our model
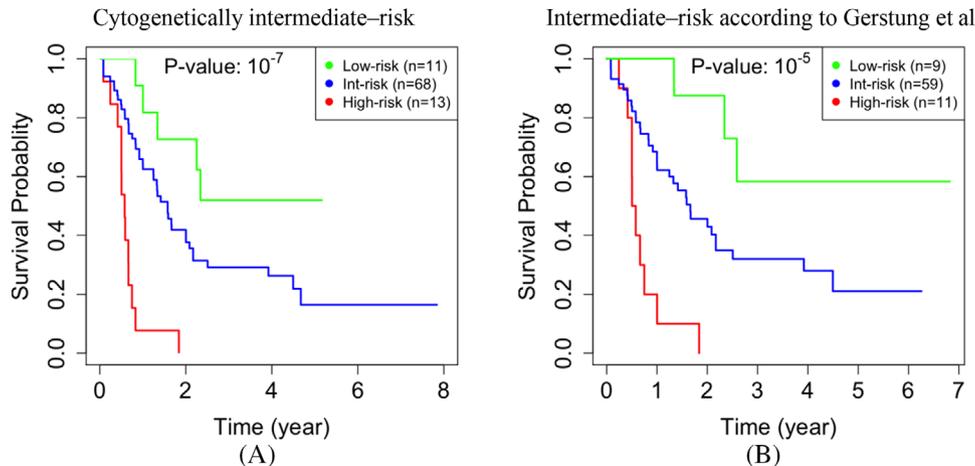
FIGURE 3 Restratification of purportedly intermediate-risk AML cases in the TCGA dataset. There are 92 cytogenetically intermediate-risk cases (the blue KM curve in Figure 2a). Using our integrative gene network approach, we further classified these 92 cases into 11 low-, 66 intermediate-, and 13 high-risk cases (**a**). The approach by Gerstung et al would classify 81 cases as intermediate-risk, with survival probabilities higher than the first quartile but lower than the fourth quartile (the blue KM curve in Figure 2b). After excluding two APL cases, we further classified the remaining 79 non-APL cases into nine low-risk, 59 intermediate-risk, and 11 high-risk cases (**b**). In each plot, the *p*-value corresponds to log-rank tests with a null hypothesis that the predicted high-risk cases have the same expected survival time as other cases in that plot

narrowing down the prognostication of purportedly intermediate-risk patients. Acute promyelocytic leukemia (APL) is a subtype of AML with a distinct cytogenetic signature, which allows clinicians to quickly and accurately diagnose it using fluorescent in situ hybridization and polymerase chain reaction [52]. Therefore, we excluded the 21 APL patients in the TCGA datasets in the following assessment. The group of 79 non-APL patients that Gerstung et al classified as intermediate-risk (i.e., the second and third quartiles based on the predicted survival probability) included a mix of actual high- and low-risk cases. In this group, 47 cases (60%) died of AML, while 14 cases (18%) were relatively low-risk cases. These 14 low-risk cases were followed for at least 2 years after

diagnosis, and they were found to be alive at the last time of contact. It is thus critical to further assess the clinical risk for this subset of cases.

We investigated the utility of our approach in the prognostication of the 79 non-APL cases that would be classified as intermediate-risk by Gerstung et al (Figure 3b). Our approach identified a subset of 11 high-risk cases (14%) who survived significantly shorter than other cases in this group (*p*-value $\leq 10^{-5}$). All of these 11 cases died of AML within 2 years after diagnosis, although the majority of them (i.e., six cases, 55%) had a normal karyotype, two cases had the del(7q) abnormality, and one case had the t(9;11)(p22;q23) translocation. Two cases
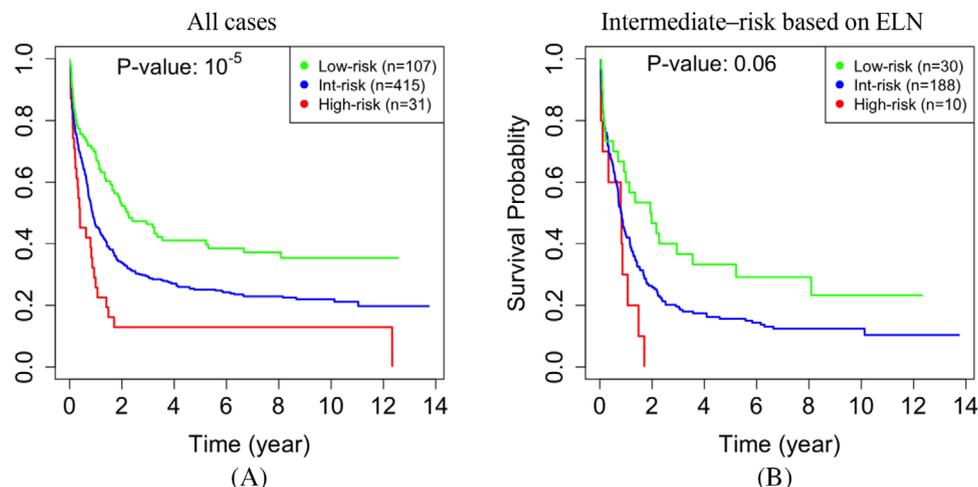
FIGURE 4 The KM curves for the AMLCG 1999 cohort. Cases that we identified as low- (green) and high-risk (red) have significantly different survival times (**a**). While 228 cases had an intermediate-I or intermediate-II ELN risk score, our approach determined that 10 of these cases were high-risk, all of whom died within 2 years after diagnosis (**b**)

had complex cytogenetics with three or more distinct abnormalities (Table S3).

3.4 | Validation using independent datasets

To show that the prognostication performance of the three identified modules is not specific to the TCGA dataset, we validated our findings using the German AMLCG 1999 and Beat cohorts.

The German AMLCG 1999 cohort. We identified 107 (19%) low-, 415 (75%) intermediate-, and 31 (6%) high-risk cases in the AMLCG cohort (Table S3). There was a significant difference between the survival time of the low- and high-risk groups that we identified (*p*-value $\leq 10^{-5}$, Figure 4a). More interestingly, all 10 cases with an intermediate ELN risk score that we identified as high-risk died within 2 years after diagnosis. This suggests that our approach can identify a subgroup of truly high-risk cases that would not be considered as such using the current clinical criteria. To further quantify and support this assertion, we compared this group of 10 cases with the other 218 patients in the AMLCG 1999 cohort with an intermediate-I or intermediate-II ELN-2010 risk score. After adjustment for multiple hypothesis testing, there was no significant association among the variables included in the ELN-2017 genetic risk classifier [53], nor other common clinical variables including white blood cell counts, hemoglobin, platelets, lactate dehydrogenase, and Eastern Cooperative Oncology Group score [54] (*p*-value ≥ 0.05).

The Beat cohort. Of the 405 cases with survival data in the BEAT cohort, 156 (39%) were not classified as favorable or adverse by ELN2017. Using the eigengenes that were identified based on TCGA data, we classified these 156 cases into 62 (40%) low-, 86 (55%) intermediate-, and eight (5%) high-risk cases (Table S3). Interestingly, none of the eight cases that we identified as high-risk were known to be alive long after diagnosis. That is, five cases died within 1 year after diagnosis, while the other three cases left the study within 16 months

(i.e., 458, 302, and 16 days, respectively) after diagnosis. A hypergeometric test showed that if five cases were randomly selected from the pool of 156 cases, it would have been unlikely that all of selected cases die within 1 year after diagnosis (*p*-value ≤ 0.002).

3.5 | The significance of DNA methylation in gene network analysis

To assess the significance of epigenomics in our study, we repeated our analysis using gene expression but without DNA methylation data. Without using DNA methylation data, our analysis identified fewer modules, which were generally larger. Specifically, we identified 36 modules, and their sizes had a mean, median, and standard deviation of 339, 38, and 917, respectively. The *p*-value for separation between the low- and high-risk cases was less significant (10^{-6} in Figure 5a vs. 10^{-11} in Figure 2c). We also compared the performance of prognostication for the subset of 92 cytogenetically intermediate-risk cases. The performance decreased when epigenomic data were excluded from the analysis. Specifically, the *p*-value of 10^{-2} in Figure 5b, which is equivalent to a chi-square statistic of 28, is less significant than the *p*-value of 10^{-7} in Figure 3a, which is equivalent to a chi-square statistic of 7. In particular, without DNA methylation data, we could identify only eight high-risk cases, three of which were cytogenetically classified as intermediate-risk. Collectively, the above results indicate that integrating DNA methylation data into the network is essential for identifying gene modules that are relevant to prognosis.

4 | DISCUSSION

Almost half of AML cases cannot be confidently classified into low- or high-risk groups using current prognostic criteria. Transcriptome and epigenome data have been used to prognosticate AML (Note S1). Our
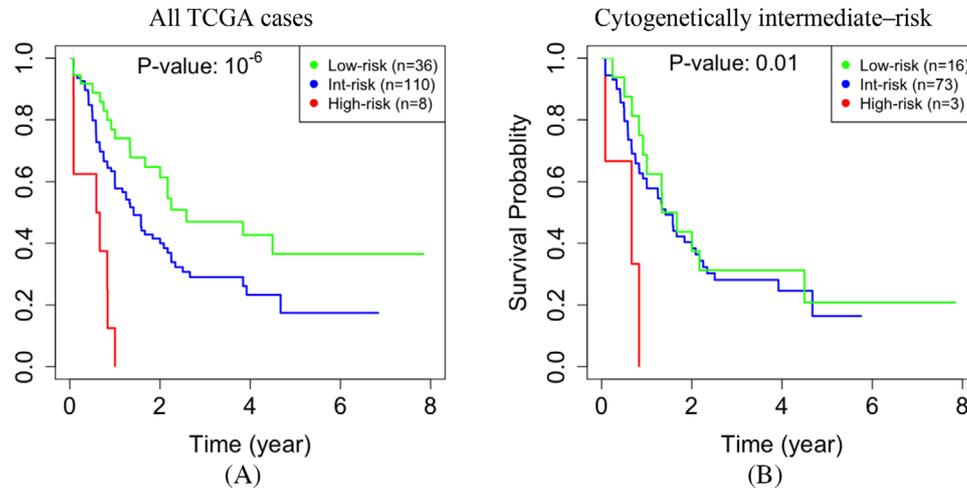
FIGURE 5 Network analysis without epigenomic data. When DNA methylation data were not used in constructing the gene network, the ability to predict survival decreases in the entire TCGA dataset (p-value = 10^{-6} vs. 10^{-11} in Figure 2c) **(a)**, and in the subset of 92 cytogenetically intermediate-risk cases (10^{-2} vs. 10^{-7} in Figure 3a) **(b)**

study showed that integrating these data in a single model can have a synergistic effect and improve prognostication (Figure 5). While both gene expression and DNA methylation data from TCGA cohort were used to identify gene modules, we used only gene expression data from two independent cohorts to validate our findings because DNA methylation data were not available in the later (Note S5). Interestingly, our approach is useful in prognostication of AML in the validation cohorts, although DNA methylation data were not available for these cases. Outperforming LSC17 score, which was developed based on transcriptome data, suggests that integration of epigenome data in the training phase results in more informative gene modules. This supremacy is expected because DNA methylation provides us with additional robust biological information to train the model.

Validation using the AMLCG and Beat AML cohorts showed that the prognostication performance of the three identified modules is not specific to the TCGA dataset. That is, for all studied cohorts, our approach allows for the identification of a particular group of high-risk cases that otherwise would be inaccurately considered intermediate-risk based on the current prognostic factors. In particular, when our risk assessment is compared to the approach presented by Ng et al, which is based on LSC [22], our assessment reveals that these high-risk cases are almost evenly distributed between the low-and high-risk groups based on their LSC17 score. This suggests that the modules we identified represent a signature that is distinct from the LSC17 score (Note S7).

Clinical utility. When combined with other integrative schemes, our approach can further improve risk assessment especially to identify some of the high-risk cases that otherwise would be classified as intermediate-risk. For instance, if a case is categorized in the gray intermediate-risk group by other approaches but our assessment suggests that case is high-risk, then the actual risk for the case is most likely high. Specifically, when combined with the Gerstung et al approach, we could identify 11 (6%) more high-risk cases in the TCGA dataset (Figure 3b). When combined with ELN-2010, we could identify 10 (2%)

more high-risk cases in the AMLCG dataset (Figure 4b). Interestingly, these cases exhibited a poor prognosis, with a survival rate of less than 2 years. Our results show that using DNA methylation data in building the gene network has synergistic benefits. In follow-up studies, integrating other multiomics data into the network could perhaps improve prognostication further.

In conclusion, using an integrative network approach to group thousands of genes into modules, we showed that incorporating DNA methylation into gene network analyses results in more robust and informative gene modules because each resulting module is a group of coexpressed and co-methylated genes. Some of these modules define survival signatures, which can be combined to develop a prognostic test (Figure 2c).

CONFLICT OF INTEREST

The authors have declared that no competing interests exist.

AUTHOR CONTRIBUTIONS

ORCID

Aly Karsan MD https://orcid.org/0000-0001-6753-892X
Habil Zare PhD https://orcid.org/0000-0001-5902-6238

REFERENCES

1. Jemal A, Thomas A, Murray T, Thun M. Cancer statistics, 2002. CA Cancer J Clin. 2002;52(1):23–47.
2. Borate U. Applicability of the AML knowledge bank approach. Blood. 2018;132(8):775–6.
3. Grimwade D, Hills RK, Moorman AV, Walker H, Chatters S, Goldstone AH, et al. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. Blood. 2010;116(3):354–65.
4. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al. Ge-nomic classification and prognosis in acute myeloid leukemia. N Engl J Med. 2016;374(23):2209–21.
5. Walter RB, Othus M, Paietta EM, Racevskis J, Fernandez HF, Lee JW, et al. Effect of genetic profiling on prediction of therapeutic resistance and survival in adult acute myeloid leukemia. Leukemia. 2015;29(10):2104.
6. Foroushani A, Agrahari R, Docking R, Chang L, Duns G, Hudoba M, et al. Large-scale gene network analysis reveals the significance of extracellular matrix pathway and homeobox genes in acute myeloid leukemia: an introduction to the Pigengene package and its applications. BMC Med Genet. 2017;10(1):16.
7. Li Y, Wu FX, Ngom A. A review on machine learning principles for multiview biological data integration. Brief Bioinform. 2016;19(2):325–40.
8. Baltrusˇaitis T, Ahuja C, Morency LP. Multimodal machine learning: A survey and taxonomy. IEEE Trans Pattern Anal Mach Intell. 2019;41(2):423–43.
9. Network CGAR, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013;368(22):2059.
10. Herold T, Metzeler KH, Vosberg S, Hartmann L, Roĺllig C, Stoĺlzel F, et al. Isolated trisomy 13 defines a homogeneous AML subgroup with high frequency of mutations in spliceosome genes and poor prognosis. Blood. 2014;124(8):1304–11.
11. Li Z, Herold T, He C, Valk PJ, Chen P, Jurinovic V, et al. Identification of a 24-gene prognostic signature that improves the European LeukemiaNet risk classification of acute myeloid leukemia: an international collaborative study. J Clin Oncol. 2013;31(9):1172.
12. Janke H, Pastore F, Schumacher D, Herold T, Hopfner KP, Schneider S, et al. Activating FLT3 mutants show distinct gain-of-function phenotypes in vitro and a characteristic signaling pathway profile associated with prognosis in acute myeloid leukemia. PLoS One. 2014;9(3):e89560.
13. Jiang X, Bugno J, Hu C, Yang Y, Herold T, Qi J, et al. Eradication of acute myeloid leukemia with FLT3 ligand–targeted miR-150 nanoparticles. Cancer Res. 2016;76(15):4470–80.
14. Tyner JW, Tognon CE, Bottomly D, Wilmot B, Kurtz SE, Savage SL, et al. Functional genomic landscape of acute myeloid leukaemia. Nature. 2018;562(7728):526.
15. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res. 2013;41(D1):D991–5.
16. Doĺhner H, Estey EH, Amadori S, Appelbaum FR, Buĺchner T, Burnett AK, et al. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. Blood. 2010;115(3):453–74.
17. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008;5(7):621–8.
18. Lee S, Seo CH, Lim B, Yang JO, Oh J, Kim M, et al. Accurate quantification of transcriptome from RNA-Seq data by effective length normalization. Nucleic Acids Res. 2010;39(2):e9–e9.
19. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9(1):559.
20. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of gene coexpression net-works in human and chimpanzee brains. Proc Natl Acad Sci. 2006;103(47):17973–8.
21. Gerstung M, Papaemmanuil E, Martincorena I, Bullinger L, Gaidzik VI, Paschka P, et al. Pre-cision oncology for acute myeloid leukemia using a knowledge bank approach. Nat Genet. 2017;49(3):332–40.
22. Ng SW, Mitchell A, Kennedy JA, Chen WC, McLeod J, Ibrahimova N, et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. Nature. 2016;540(7633):433.
23. Cox DR. Regression models and life-tables. Breakthroughs in statistics. New York: Springer; 1992. p. 527–41.
24. Gui J, Li H. Penalized Cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene expression data. Bioinformatics. 2005;21(13):3001–8.
25. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Statist Assoc. 1958;53(282):457–81.
26. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, et al. InnateDB: systems biology of innate immunity and beyond–recent updates and continuing curation. Nucleic Acids Res. 2013;41:D1228–33.
27. Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, et al. The Reactome pathway knowledgebase. Nucleic Acids Res. 2014;42(D1):D472–7.
28. Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, et al. PID: the pathway interaction database. Nucleic Acids Res. 2009;37(suppl 1):D674–9.
29. Chao MV. Neurotrophins and their receptors: a convergence point for many signalling pathways. Nat Rev Neurosci. 2003;4(4):299.
30. Monaco S, Rusciano MR, Maione AS, Soprano M, Gomathinayagam R, Todd LR, et al. A novel crosstalk between calcium/calmodulin kinases II and IV regulates cell proliferation in myeloid leukemia cells. Cell Signal. 2015;27(2):204–14.
31. Kang X, Cui C, Wang C, Wu G, Chen H, Lu Z, et al. CAMKs support development of acute myeloid leukemia. J Hematol Oncol. 2018;11(1):30.
32. Al-Asadi MG, Brindle G, Castellanos M, May ST, Mills KI, Russell NH, et al. A molecular signature of dormancy in CD34+ CD38-acute myeloid leukaemia cells. Oncotarget. 2017;8(67):111405.
33. Kazi JU, Rönnstrand L. Suppressor of cytokine signaling 2 (SOCS2) associates with FLT3 and negatively regulates downstream signaling. Mol Oncol. 2013;7(3):693–703.
34. Vitali C, Bassani C, Chiodoni C, Fellini E, Guarnotta C, Miotti S, et al. SOCS2 controls proliferation and stemness of hematopoietic cells under stress conditions and its deregulation marks unfavourable acute leukemias. Cancer Res. 2015;75(11):2387–99.
35. Abraham JE, Guo Q, Dorling L, Tyrer J, Ingle S, Hardy R, et al. Replication of genetic polymorphisms reported to be associated with taxane-related sensory neuropathy in early breast cancer patients treated with paclitaxel. Clin Cancer Res. 2014;20(9):2466–75.

36. Geyh S, Rodrʼıguez-Paredes M, Jáger P, Koch A, Bormann F, Gutekunst J, et al. TGFβ 1-mediated functional inhibition of mesenchymal stromal cells in myelodysplastic syndromes and acute myeloid leukemia. Haematologica. 2018;103(9):1462–71.

37. Zeddies S, Jansen SB, di Summa F, Geerts D, Zwaginga JJ, van der Schoot CE, et al. MEIS1 regulates early erythroid and megakaryocytic cell fate. Haematologica. 2014;99(10):1555–64

38. Hu T, Chong Y, Lu S, Wang R, Qin H, Silva J, et al. MicroRNA 339 promotes development of Stem Cell Leukemia/Lymphoma syndrome via downregulation of the BCL2L11 and BAX pro-apoptotic genes. Cancer Res. 2018;78(13):3522–31.

39. Kabir NN, Rónnstrand L, Kazi JU. Deregulation of protein phosphatase expression in acute myeloid leukemia. Med Oncol. 2013;30(2):517.

40. Arora D, Kólthe S, van den Eijnden M, van Huijsduijnen RH, Heidel F, Fischer T, et al. Expression of protein-tyrosine phosphatases in Acute Myeloid Leukemia cells: FLT3 ITD sustains high levels of DUSP6 expression. Cell Commun Signal. 2012;10(1):19.

41. Mustjoki S, Sidenius N, Sier CF, Blasi F, Elonen E, Alitalo R, et al. Soluble urokinase receptor levels correlate with number of circulating tumor cells in acute myeloid leukemia and decrease rapidly during chemotherapy. Cancer Res. 2000;60(24):7126–32.

42. Mustjoki S, Alitalo R, Stephens RW, Vaheri A. Blast cell-surface and plasma soluble urokinase receptor in acute leukemia patients: relationship to classification and response to therapy. Thromb Haemost. 1999;81(05):705–10.

43. Nadir Y, Katz T, Sarig G, Hoffman R, Oliven A, Rowe JM, et al. Hemostatic balance on the surface of leukemic cells: the role of tissue factor and urokinase plasminogen activator receptor. Haematologica. 2005;90(11):1549–56.

44. Graf M, Reif S, Hecht K, Pelka-Fleischer R, Pfister K, Schmetzer H. High expression of urokinase plasminogen activator receptor (UPA-R) in acute myeloid leukemia (AML) is associated with worse prognosis. Am J Hematol. 2005;79(1):26–35.

45. Kerr JF. History of the events leading to the formulation of the apoptosis concept. Toxicology. 2002;181:471–4.

46. Lippincott-Schwartz J, Roberts TH, Hirschberg K. Secretory protein trafficking and organelle dynamics in living cells. Annu Rev Cell Dev Biol. 2000;16(1):557–89.

47. Kültz D. Molecular and evolutionary basis of the cellular stress response. Annu Rev Physiol. 2005;67:225–57.

48. Neben K, Tews B, Wrobel G, Hahn M, Kokocinski F, Giesecke C, et al. Gene expression patterns in acute myeloid leukemia correlate with centrosome aberrations and numerical chromosome changes. Oncogene. 2004;23(13):2379–84.

49. Pei S, Minhajuddin M, Callahan KP, Balys M, Ashton JM, Neering SJ, et al. Targeting aberrant glutathione metabolism to eradicate human acute myelogenous leukemia cells. J Biol Chem. 2013;288(47):33542–58.

50. Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. Hoboken, NJ: John Wiley & Sons; 2011.

51. Zainulabadeen A, Yao P, Zare H. Underexpression of specific interferon genes is associated with poor prognosis of melanoma. PLoS One. 2017;12(1):e0170025.

52. Grignani F, Fagioli M, Alcalay M, Longo L, Pandolfi PP, Donti E, et al. Acute promyelocytic leukemia: from genetics to treatment. Blood. 1994;83(1):10–25.

53. Dó hner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Bú chner T, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood. 2017;129(4):424–47.

54. Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, Mcfadden ET, et al. Toxicity and re-sponse criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol. 1982;5(6):649–56.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.