Signatures of Poor Prognosis in Leukemia Identified Using Gene Network Analysis

Aamir Zainulabadeen[1,2], Dr. Habil Zare[1], Rod Docking[3], Dr. Aly Karsan[3]
1 Department of Computer Science, Texas State University, Texas, USA
2 Department of Computer Science, Princeton University, New Jersey, USA
3 Karsan Lab, British Columbia Cancer Agency, Vancouver, CA

TEXAS STATE UNIVERSITY
The rising STAR of Texas

Background

The normal human karyotype (NK) consists of 46 chromosomes and all loci being coherent. In some subtypes of leukemia, there can be major aberrations to the karyotype, i.e. more or less than 46 chromosomes through duplications as well as loci transposing between chromosomes.[8] Classifying risk amongst patients with a normal karyotype in Acute Meyeloid Leukemia (AML) has been important and challenging for clinicians.[9]
Our analysis uses in vivo data from three different studies with AML-NK patients to determine an expression signature that stratifies normal karyotype patients into high and low risk groups.

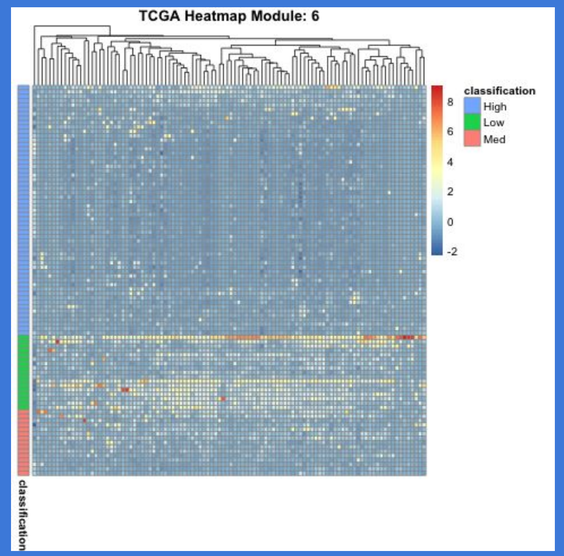


Fig. 1: A heatmap of gene expression (columns) in each patient (row) for the Focal Adhesion Module in the TCGA dataset. Patients are grouped by status.

Citations

1. Cancer Genome Atlas Network. (2015). Genomic classification of cutaneous melanoma. *Cell*, *161*(7), 1681-1696.
2. Cancer Genome Atlas Network. Genomic classification of cutaneous melanoma. Cell. 2015 Jun 18;161(7):1681-96.
3. Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*, *9*(1), 1.
4. Grimm EA, Sikora AG, Ekmekcioglu S. Molecular pathways: inflammation-associated nitric-oxide production as a cancer-supporting redox mechanism and a potential therapeutic target. Clinical Cancer Research. 2013 Oct 15;19(20):5557-63.
5. Amir Foroushani et al.(2016) Large-scale gene network analysis reveals the significance of extracellular matrix pathway and homeobox genes in acute myeloid leukemia, Foroushani et al., In preparation. URL: http://oncinfo.org/.
6. Cox DR. Regression models and life-tables. In: Breakthroughs in statistics. Springer; 1992. p. 527–541.
7. 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):3001–3008. pmid:15814556
8. Nicholson, J.M. and Cimini, D., 2013. Cancer karyotypes: survival of the fittest. Frontiers in oncology, 3.
9. Schlenk, R.F., Dohner, K., Krauter, J., Cornelissen, J.J., van Putten, W.L.J. and Verdonck, L.F., 2008. Cytogenetically normal acute myeloid leukemia. N Engl J Med, 2008(359), pp.651-653.

Methodology

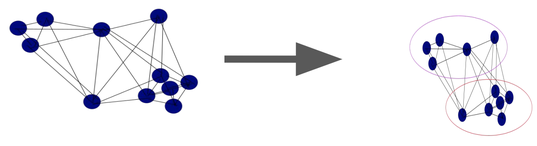
Obtaining Data

Our analysis used three datasets:

- We downloaded gene expression data from the TCGA study on Leukemia.[1]
- We wrote a script to download and bind the data from the Leucegene dataset.
- Expression data from AML-NK patients were provided to us from the Karsan Lab at the British Columbia Cancer Agency.

The Gene Network

We build a weighted gene correlation network using functions in the WGCNA R package.[3]



Each dot represents a gene and each connection is based on the similarity (computed from correlation) between each gene. RIght: groups of highly correlated genes are identified using hierarchical clustering.

Clustering and Eigengenes

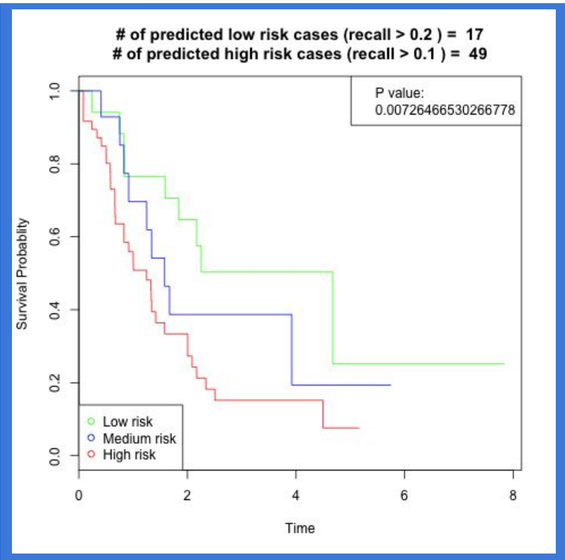
- The R package Pigengene was *modified* to perform consensus clustering and eigengene computation over all three datasets.[5]
- The clustering yielded 14 clusters with 3808 outliers. An **eigengene** (first principle component of gene expression) was computed for each module.

Survival Analysis

A Penalized Cox Regression was performed to obtain the modules that are associated with poor prognosis from **eigengenes** in the TCGA dataset.[6, 7]

The best three modules across many runs with different parameters were:

- A module associated with Focal Adhesion (102 genes).
- A module associated with Metabolism (32 genes).
- A module not significantly annotated to any pathway (182 genes).



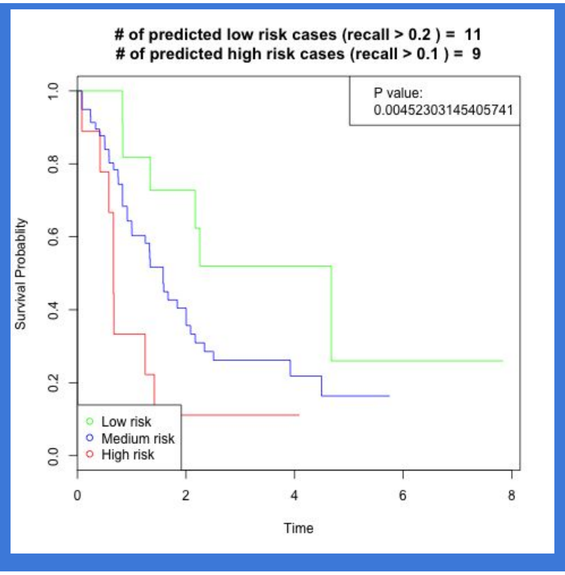
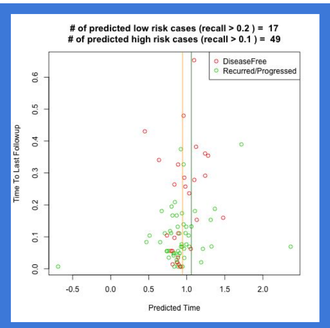


Fig. 2: Kaplan-Meier plots of survival. The probability of survival is on the y-axis and time is mapped on the x-axis. Left (A): the risk classification from Focal Adhesion and Metabolism modules. Right (B): the risk classification from all the three modules. Because using only the two modules leads to classification of more patients with similar significance thresholds, they are of greater interest than all three modules together.



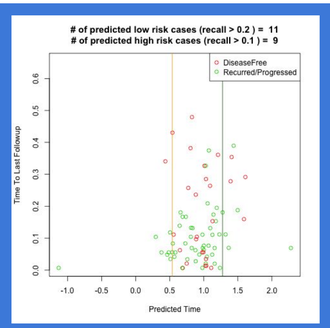


Fig. 3: Left to right:

(A) Plot of survival predictions (x-axis) for each patient using the Focal Adhesion and Metabolism pathways.

(B) Plot of predictions for each patient using all three selected modules.

Results and Conclusion

We found that in the module associated with Focal Adhesion, 100 genes correlate to survival, while 2 genes anti correlate to survival. In the module associated with Metabolism, 23 genes correlate to survival, while 9 anti correlate to survival. In the intersection of the Focal Adhesion Pathway and associated module as well as the Metabolism Pathway and corresponding module, all genes correlate to survival.

Other researchers should also use different methodologies to confirm whether the Metabolic and Focal Adhesion pathways are differentially expressed in AML-NK patients.

Acknowledgments

We thank the National Science Foundation for funding the research under the Research Experiences for Undergraduates site programs (CNS-1358939, CNS-1659807) at Texas State University to perform this piece of work and the infrastructure provided by a NSF-CRI 1305302 award.