Application of RASL-Seq in the analysis of immune response

Gabriel Hurtado, Yogendra Bordiya, Dr. Hong-Gu Kang, Dr. Habil Zare
Texas State University

Abstract

In the earliest stages of cancer, epigenetic changes, as opposed to genetic changes, have been shown to be more prominent.[1] ncRNA is the basis for the epigenetic mechanism, and its relationship with cancer has yet to be explained, as this region in the genome may contain biomarkers for early detection.[2] Current techniques that analyze the whole transcriptome are costly, thus by using eigengenes, we analyzed the robustness of a targeted approach called RASL-Seq on 226 genes in Arabidopsis and compared it to the gold standard, RNA-Seq. Our analysis showed a 0.92 spearman correlation, displaying the accuracy of our method. We then performed a hierarchical cluster analysis, which showed key differences in wildtype and mutant groups.

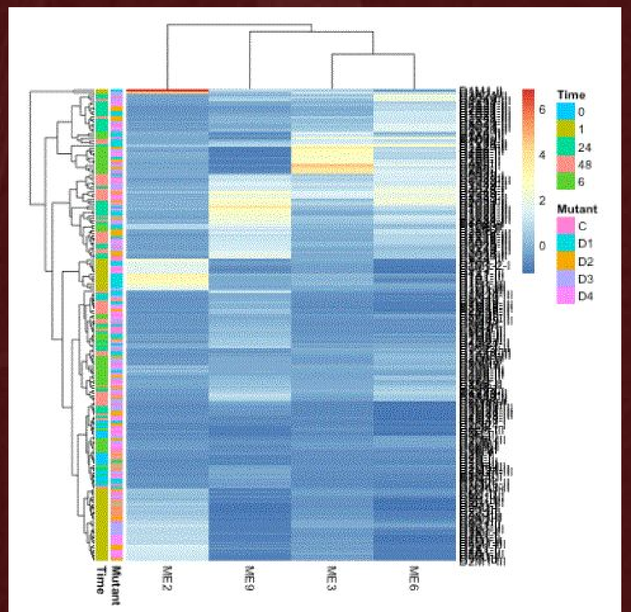


Fig. 1: A heatmap of eigengene expression in the four gene clusters (columns) that were created from the 226 genes in Arabidopsis. Samples are in rows. Time intervals and mutant groups in are color coded on left.

Citations

1. Hatziapostolou M, Iliopoulos D (2011) Epigenetic aberrations during oncogenesis. Cellular and molecular life sciences : CMLS 68: 1681-1702.
2. Kang HG, Zare H (2017 Development of massive parallel RNA sequencing for cancer diagnosis.
3. Consortium EP (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57-74
4. Hatziapostolou M, Iliopoulos D (2011) Epigenetic aberrations during oncogenesis. Cellular and molecular life sciences : CMLS 68: 1681-1702
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/.

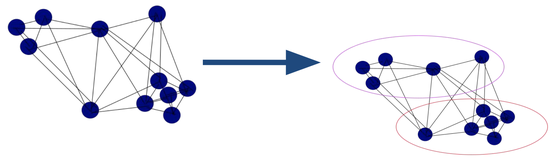
Methodology

Obtaining Data

Our gene expression data was generated by Dr. Hong-Gu Kang's laboratory from the Biology Department at Texas State. His lab performed RNA-Seq on thousands of genes from 39 Arabidopsis samples. 226 genes were selected for RASL-Seq analysis of 234 samples.

The Gene Network

The gene clusters were built using functions in the maSigPro package in R.



Each dot represents a gene. Each line is a connection (the correlation) between each gene.

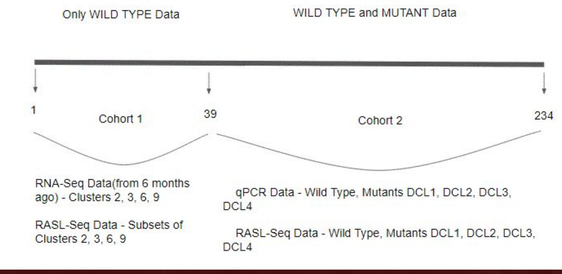
The network is then clustered into gene modules.

Clustering and Eigengenes

- An **eigengene** (first principle component of gene expression) was computed for each module.
- The R package Pigengene was used to perform compute an eigengene for each cluster.[5]

Treatment and Time Intervals

Treatment	Hours post infiltration (hpi)					Row total
	0 hpi	1hpi	6hpi	24hpi	48hpi	
Naïve(N) No treatment	1	N/A	N/A	N/A	N/A	1
Mock (M)	N/A	1	1	1	1	4
Virulent pathogen (V)	N/A	1	1	1	1	4
Avirulent pathogen (A)	N/A	1	1	1	1	4
Total number of samples from one biological replicate						13



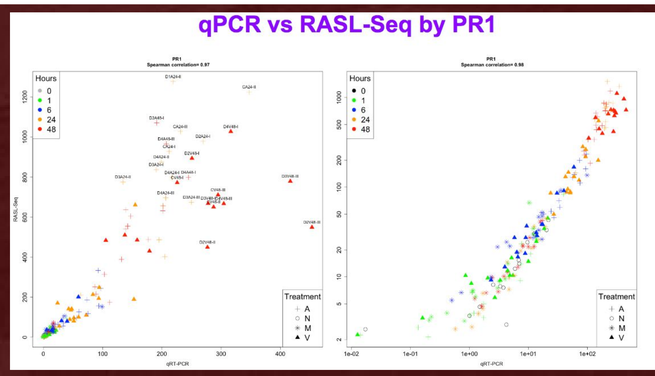
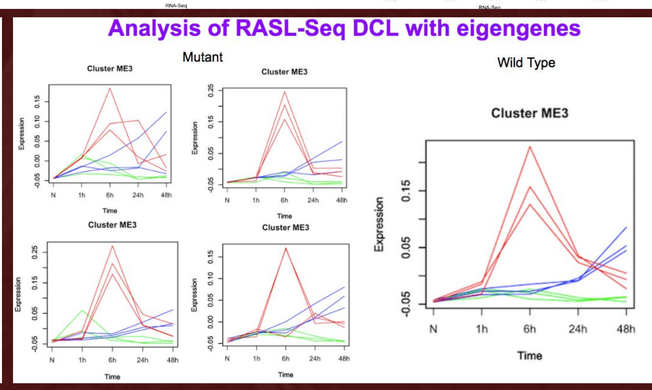


Fig. 2: Scatterplots and lineplots that compare the accuracy of RASL-Seq with two gold standards for gene expression, qPCR and RNA-Seq. From left to right, clockwise: (A) Plot showing an increase in correlation between RASL-Seq and qPCR for two rounds of sequencing. (B) Comparison of RASL-Seq and RNA-Seq using subset cluster 3. (C) Comparison of RASL-Seq mutant groups and wildtype using eigengene values. Treatments: A - Red, M - Green, V - Blue.



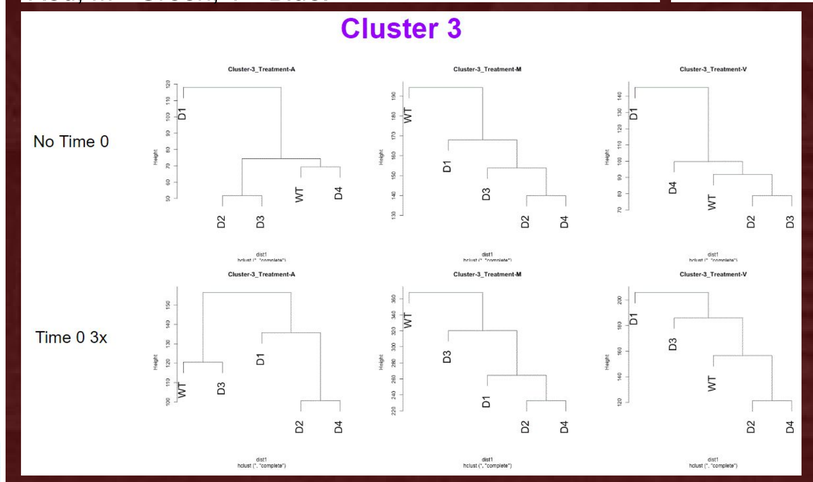


Fig. 3: Hierarchical cluster analysis for RASL-Seq mutant groups and WT. This method used an agglomerative approach and used complete linkage clustering as the method. This shows how mutants (WT, D1-D4) correlate to each other in terms of gene expression, specifically in cluster 3. The top three plots do not account for Time 0 (control), and the bottom three take into account Time 0 weighted 3x.

Results and Conclusion

In Fig. 2 (A) and (B), our two rounds of RASL-Seq differed in terms of the amount of sequence reads that Dr. Kang's lab obtained. With the second round having significantly more reads(37 mil vs 14 mil), as this was the metric we used to gauge the accuracy of RASL-Seq compared to the gold standard, the correlation increased for both comparisons. Thus this showed the accuracy of RASL-Seq when compared to gold standard analyses of RNA. In Fig. 2 (C) and Fig. 3, this analysis seemed to show that mutations in genes that control miRNA biogenesis affected the immune response in Arabidopsis and they alter gene expression on a marked level when compared to the wildtype.

Our future study include combining the mutant groups and then performing this analysis again and also performing RASL-Seq on cancer samples obtained from MD Anderson.

Acknowledgments

We thank the U.S Department of Education Grant #P031C160036 and the HSI STEM IMPACT Program for all their support. We also thank Rupesh Agrahari for his supply of the template for this poster.