Reconstruction of gene regulatory modules in cancers

Laura Poulton[1] with Duc Do[2] and Dr. Serdar Bozdag[2]

[1]Department of Electrical and Computer Engineering, Bucknell University

[2]Department of Math, Statistics, and Computer Science, Marquette University

Introduction

Background:

During protein synthesis, DNA are transcribed into mRNA, which are then translated into proteins. Transcription Factors (TFs) are a type of mRNA that bind to specific sections of a DNA sequence during the transcription stage[3]. microRNAs (miRNAs) are short, noncoding RNAs that can bind to the 3' UTR regions of mRNA in the translation stage of protein synthesis[4].

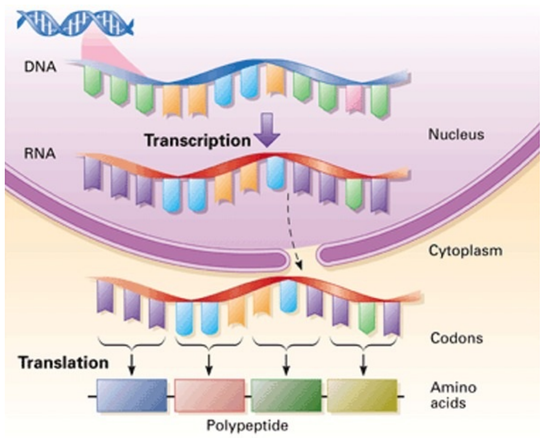


Figure 1. Protein synthesis within a human cell.

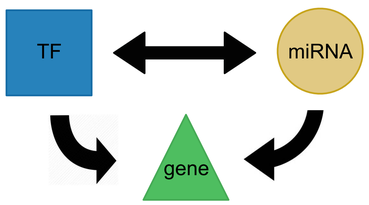
TFs and miRNAs can either enhance or repress the production of other mRNA. Both TFs and miRNAs can bind to multiple genes, the same genes, and even to each other[4].

Problem Statement:

Given the thousands of possible regulatory relationships between miRNA, TF, and mRNA, identify the most probable interactions using computational methods for later experimental validation.

MTTM Modules:

We propose integrating both expression and sequencing data in order to determine likely miRNA and TF co-regulatory relationships by constructing miRNA - TF - target gene modules.



Methods

- **Prepared expression data:**
  - 332 total KIRC samples from The Cancer Genome Atlas (TCGA)[7]
  - mRNA and miRNA expression counts
  - 23 normal, 309 tumor samples
- **Differential expression analysis:**
  - Identified mRNA and miRNA with significantly different expression counts in tumor samples when compared with normal samples
  - Filtered out genes with low expression in too many samples
  - Used Fisher's Exact Test
- **Prepared sequencing data:**
  - Downloaded miRNA - target gene and TF - target gene putative interactions from several databases[2,5,6,8]
  - Filtered interactions for differentially expressed miRNA, TFs, and target genes
- **Found expression correlations:**
  - Computed Pearson correlation of expression for each putative miRNA - mRNA and TF - mRNA pair across all of the tumor samples

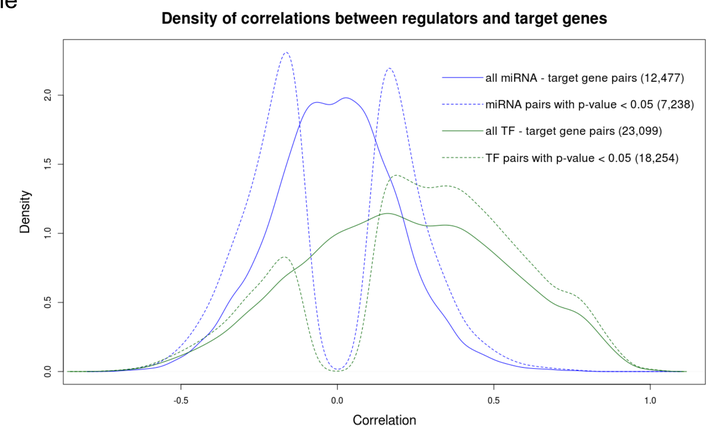


Figure 2. Correlation density plot. Dotted lines represent density of correlations with a significant p-value.

- **Filtered by correlation:**
  - Interested in miRNA tumor suppressors, want negative correlation
  - After analyzing density plots (see Figure 2), determined that miRNA - target gene correlations < -0.3 significant
  - 13% miRNA - target gene interactions had significant correlation and p-value
- **Filtered by correlation (cont.):**
  - Interested in both oncogene and tumor suppressor TFs, want absolute value of correlations
  - After analyzing density plots (see Figure 2), determined that TF - target gene correlations with absolute value > 0.5 significant
  - 27% TF - target gene interactions had significant correlation and p-value
- **Identified clusters:**
  - Used multilevel clustering algorithm on undirected graph object
  - 77 total clusters

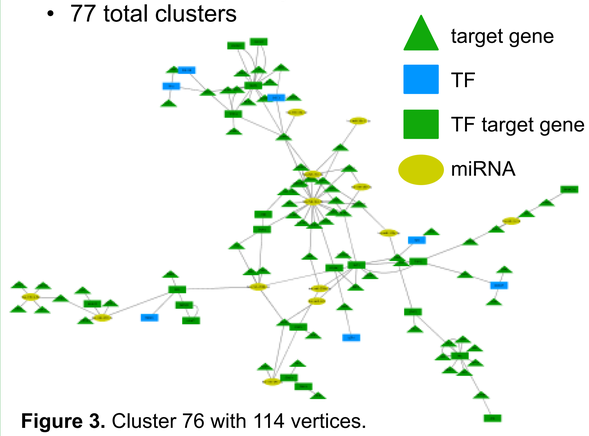


Figure 3. Cluster 76 with 114 vertices.

- **Performed enrichment analysis:**
  - Compared ontology of genes within cluster to all differentially expressed mRNA in 13 clusters large enough to analyze (> 20 vertices)
  - hypergeometric test was used to identify the biological processes and molecular functions active within each cluster

Results

ID	Description	GeneRatio	BgRatio	pvalue
GO:1903047	mitotic cell cycle process	115/199	461/6597	2.435671e-84
GO:0000280	nuclear division	82/199	279/6597	3.764313e-63
GO:0048285	organelle fission	84/199	306/6597	4.149000e-62
GO:0007059	chromosome segregation	64/199	152/6597	6.550957e-60
GO:0051301	cell division	76/199	281/6597	4.941299e-55
GO:0098813	nuclear chromosome segregation	57/199	129/6597	1.555684e-54
GO:0007067	mitotic nuclear division	70/199	239/6597	5.784039e-53

Figure 4. Enrichment Analysis results by p-value, cluster 17.

ID	Description	GeneRatio	BgRatio	pvalue
GO:0030155	regulation of cell adhesion	36/183	335/6597	9.024876e-13
GO:0048646	anatomical structure formation involved in morphoge...	20/64	452/6597	3.999370e-09
GO:0051338	regulation of transferase activity	34/199	456/6597	5.951423e-07
GO:0051347	positive regulation of transferase activity	26/199	316/6597	2.457841e-06
GO:2000145	regulation of cell motility	24/183	359/6597	4.789751e-05
GO:0030334	regulation of cell migration	23/183	336/6597	4.833387e-05
GO:0001501	skeletal system development	11/64	225/6597	8.590218e-06

Figure 5. Enrichment Analysis results by GO terms repeated in multiple clusters.

Module Construction:

- Extracted from clusters

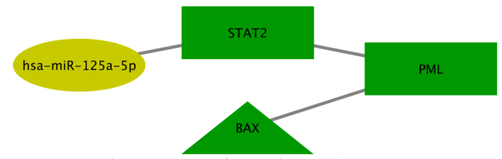


Figure 6. Module extracted from cluster 76.

Conclusions

- Many enriched GO terms are biological processes associated with cancer (see Figures 4 and 5)
- Known cancer genes BRCA1, BRCA2, MYBL2, E2F1 in multiple enriched GO terms
- Known tumor suppressor regulatory relationships intact in clusters (see Figure 6)

Future Work

- Complete survival analysis
- Complete module extraction and analysis
- Comparison to bi-clustering algorithm

References

1. Bioconductor - open source software for bioinformatics. Accessed June 28, 2017.
2. Chou, Chih-Hung, Nai-Wen Chang, Sirjana Shrestha, Sheng-Da Hsu, Yu-Ling Lin, Wei-Hsiang Lee, Chi-Dung Yang, Hsiao-Chin Hong, Ting-Yen Wei, Siang-Jyun Tu, Tzi-Ren Tsai, Shu-Yi Ho, Ting-Yan Jian, Hsin-Yi Wu, Pin-Rong Chen, Nai-Chieh Lin, Hsin-Tzu Huang, Tzu-Ling Yang, Chung-Yuan Pai, Chun-San Tai, Wen-Liang Chen, Chia-Yen Huang, Chun-Chi Liu, Shun-Long Weng, Kuang-Wen Liao, Wen-Lian Hsu, and Hsien-Da Huang. "MiRTarBase 2016: updates to the experimentally validated miRNA-target interactions database." *Nucleic Acids Research* 44, no. D1 (2015). doi: 10.1093/nar/gkv1258.
3. Latchman, David S. "Transcription factors: an overview." *The international journal of biochemistry & cell biology* 29.12 (1997): 1305-1312.
4. Le, T. D., L. Liu, J. Zhang, B. Liu, and J. Li. "From miRNA regulation to miRNA-TF co-regulation: computational approaches and challenges." *Briefings in Bioinformatics* 16, no. 3 (2014): 475-96. doi:10.1093/bib/bbu023.
5. "Prediction of microRNA targets." TargetScanHuman 7.1. 2016. Accessed July 19, 2017. http://www.targetscan.org/vert_71/.
6. Slowikowski, Kamil. "Human transcription factor target genes." Slowkow/tftargets. May 02, 2017. Accessed July 19, 2017. https://github.com/slowkow/tftargets.
7. "The Cancer Genome Atlas." National Institutes of Health. Accessed June 28, 2017. https://cancergenome.nih.gov/.
8. "The experimentally validated microRNA-target interactions database." MiRTarBase. Accessed July 19, 2017. http://mirtarbase.mbc.nctu.edu.tw/.

Acknowledgements

The author would like to thank Dr. Serdar Bozdag and Duc Do for their mentorship and guidance during the REU. The author would also like to thank Dr. Dennis Brylow, Dr. Kim Factor, and Dr. Petra Euccarius Brylow for their efforts in making this summer as successful as possible. This research was made possible through the funding of the National Science Foundation, NSF Award #ACI-1461264.