Reconstruction of gene regulatory modules in cancers

Laura Poulton¹ with Duc Do² and Dr. Serdar Bozdag²

¹Department of Electrical and Computer Engineering, Bucknell University
²Department 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. microRNAs (miRNAs) are short, noncoding RNAs that can bind to the 3' UTR regions of mRNA in the translation stage of protein synthesis.

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.

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.

MTM 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)
  • mRNA and miRNA expression counts
  • 23 normal, 309 tumor samples

• Differential expression analysis:
  • Identified miRNA 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
  • 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

• 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

• 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

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

Figure 3. Cluster 76 with 114 vertices.

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

Future Work

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

Conclusions

• Many enriched GO terms are biological processes associated with cancer (see Figures 4 and 5)
• Known cancer genes BRCA1, BRCA2, MYBL2, EZF1 in multiple enriched GO terms
• Known tumor suppressor regulatory relationships intact in clusters (see Figure 6)

References


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 Blyeow, Dr. Kim Factor, and Dr. Petra Eucalchus Blyeow 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 IIAI-140194.

Figure 1. Protein synthesis within a human cell.

Figure 6. Module extracted from cluster 17.