Subnetworks of Disrupted Genes in Cancer: A Preliminary Investigation

Lucas Lochovsky

gNets (Gerstein Nets) meeting

2012-01-17

Overview

- Integrate several types of molecular interaction network data
  - Find connections and relationships between different types of molecules
  - My network spans PPI, TFBS, and miRNA networks
- Use whole genome mutation and expression data from cancer genomes to find subnetworks of molecules that exhibit “disruption”
  - Sequence mutations or differential expression

The Network

- **PPI Interactions:** HPRD Release 9 (latest release)
  - ~37,000 interactions between some ~9000 proteins
- **TFBS:** ~400,000 binding sites of 74 human TFs (same ones from GENCODE v7) based on Peakseq coordinates
  - Follow disrupted regulatory relationships
  - Assume binding site regulates gene within 1 kb of TSS
- **miRNA:** ~900 human miRNA and their targets (~20,000) from microRNA.org database

The Network

- Genes shared by PPI and TF networks:
  - 7452
- Genes shared by PPI and miRNA networks:
  - 9458
- Genes shared by TF and miRNA networks:
  - 12,993
- Given these nontrivial intersections, this demonstrates that there is considerable “cross-talk” possible between these networks if merged

Cancer Data

- Set of 7 prostate cancer genomes from the Rubin lab
  - All have mutation data
  - Two have expression data
- Nine colorectal cancer genomes from Peter Good (published in *Nature Genetics*)
  - Whole genome mutation data
- Whole genome mutation data for various cancers from TCGA
  - 44 colon adenocarcinoma genomes (BCM; SOLiD_DNASeq)
  - 126 glioblastoma multiforme genomes (BI (Broad Institute); ABI)
  - 59 acute myeloid leukemia genomes (WUSM (Washington University School of Medicine), IlluminaGA_DNASeq)
  - 142 ovarian serous cystadenocarcinoma genomes (BI; IlluminaGA_DNASeq)
  - 35 rectum adenocarcinoma genomes (BCM; SOLiD_DNASeq)

Subnetwork Creation

- Start with known cancer-associated genes
  - Online Mendelian Inheritance of Man (OMIM)
- Iteratively investigate neighbors of these genes in the surrounding network
  - Neighbors added to the subnetwork if node has at least one sequence mutation, and/or expression difference between normal and cancer tissue is significant at a *t*-test *p*-value of 0.05
  - Algorithm allows TFBS nodes to serve as intermediates, i.e. a TFBS node may not have any disruptions, but if it connects two nodes that have disruptions, then that node and its neighbors are added

Results

- For each cancer, exactly one subnetwork was produced
  - Each contains some fraction of the seed (OMIM) genes
  - Likely corresponds to the main disease module
  - Stricter parameters can be used to filter size of the network → pick out most important sub-subnetworks

Cancer	Subnetwork Size
Prostate	6868
Colorectal	10410
AML	423
COAD	733
GBM	0
OV	3192
READ	4566

Results

Pairwise intersection matrix

	Prostate	Colorectal	AML	COAD	OV	READ
Prostate		4680	153	268	1081	1541
Colorectal			214	392	1492	2245
AML				27	78	109
COAD					183	287
OV						862
READ						

- Each disease module occupies a unique space in the entire network
- However, a small fraction of each disease module is shared with at least one other disease module → potential for development of one therapy that affects multiple diseases?

Future Work

- Experiment with stricter parameters for subnetwork creation to filter subnetworks down to most essential components
- Biological characterization of network components
  - Especially those that intersect multiple disease subnetworks