Elucidation of MicroRNA-Gene Regulation in Human Cancer with Integrative Network Models

Haluk Dogan, Zeynep Hakguder, Stephen Scott, and Juan Cui*

Systems Biology and Biomedical Informatics Laboratory
https://sbbi.unl.edu/

Department of Computer Science and Engineering
University of Nebraska-Lincoln
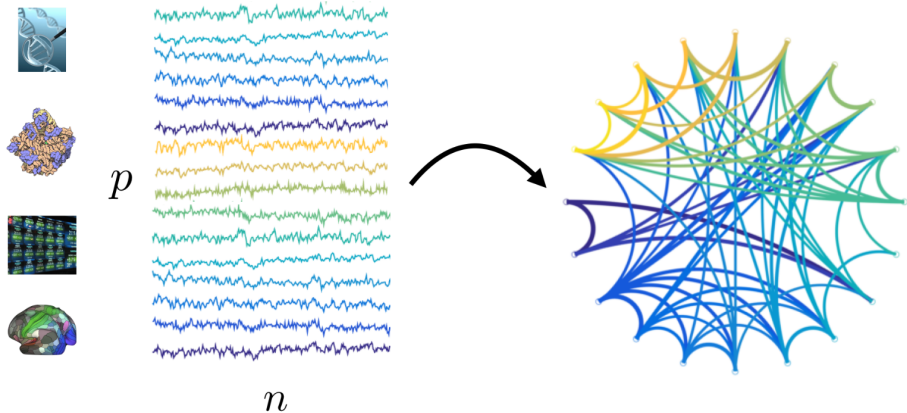
November 18, 2019

Introduction

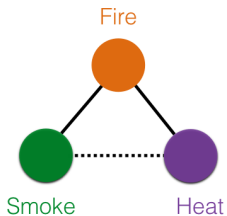Modeling interaction between cancer associated genes, miRNAs, and lncRNAs
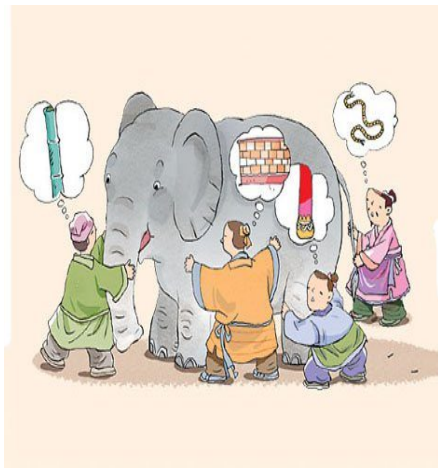
Network Analysis

Network Analysis (cont'd)

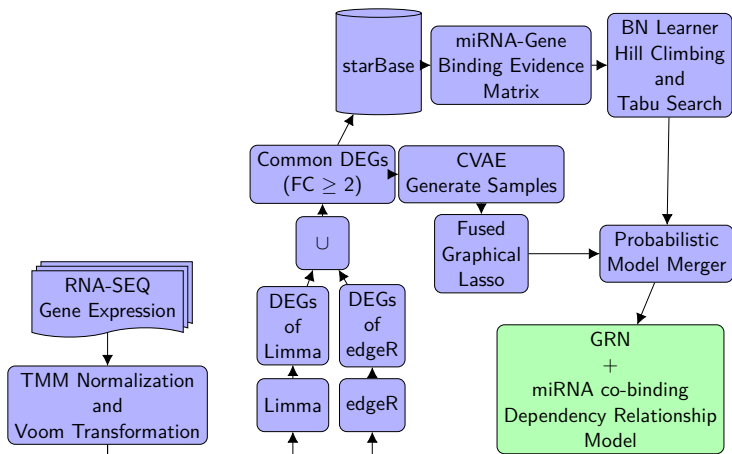Correlation network:

Occam's razor:

Our Approach

Gaussian Graphical Models

Gene regulatory network:

- $\boldsymbol{X} = [X_1, X_2, \ldots, X_p]$
- $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- $P(\boldsymbol{X}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})\right)$
- $\boldsymbol{\mu} \in \mathbb{R}^p, \boldsymbol{\Sigma} \in \mathbb{S}_{++}^p$
- Precision matrix: $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$
- Challenging problem when $n \ll p$
  Assumptions:
    - $\boldsymbol{\Theta}$ is sparse (ℓ_1)
    - Models for each group should not be too different (ℓ_2)
- Smoke $\perp\!\!\!\perp$ Heat | Fire

Gaussian Graphical Models (cont'd)

- $\mathcal{L}(\mathbf{\Omega}, \mathbf{x}) \equiv \log \det \mathbf{\Omega} - (\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{\Omega} (\mathbf{x} - \boldsymbol{\mu})$
- $\mathcal{L}(\mathbf{\Omega}, \widehat{\mathbf{\Sigma}}) \equiv \dfrac{1}{N} \sum_n \mathcal{L}\left(\mathbf{\Omega}, \mathrm{x}^{(n)}\right)$

  $= \log \det \mathbf{\Omega} - \mathsf{tr}(\widehat{\mathbf{\Sigma}} \mathbf{\Omega})$
- $\min_{\{\mathbf{\Omega} > 0\}} \mathcal{L}(\{\mathbf{\Omega}\}) := \sum_{k=1}^{K} \mathsf{tr}(\widehat{\mathbf{\Sigma}}^{(k)} \mathbf{\Omega}^{(k)}) - \log \det \mathbf{\Omega}^{(k)} + P(\{\mathbf{\Omega}\})$
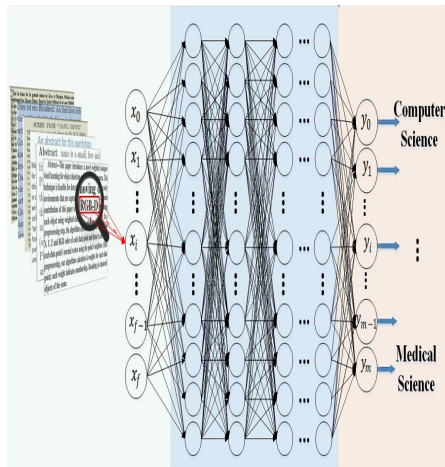
Data Augmentation

Conducting experiments:

- Unethical
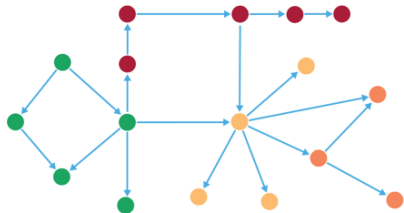- Expensive
- Difficult to repeat

We used deep learning to generate more in-vitro samples:

- number of samples in groups is imbalanced
- unbiased estimator for a learner

Constructing Bayesian Network

MiRNA Co-Binding Network:

- We constructed a Bayesian network to represent miRNA co-binding relationships
- Evidence matrix is from starBase

	Gene$_1$	Gene$_2$	...	Gene$_n$
miRNA$_1$				
miRNA$_2$		1/0		
$\vdots$				
miRNA$_p$				

- Intervention to a model is a lot easier

Connecting Models

- We test each miRNA and its dependencies to see if they make significant group difference
- We used a probabilistic distance metric to measure the data similarity with and without testing miRNAs and their dependencies
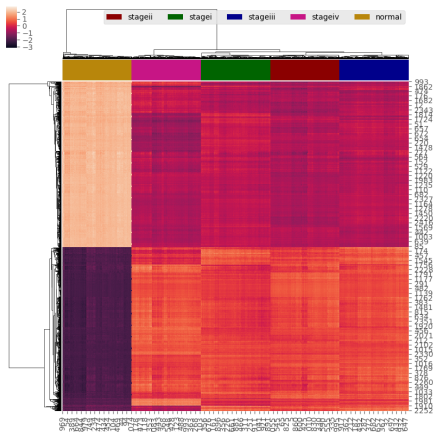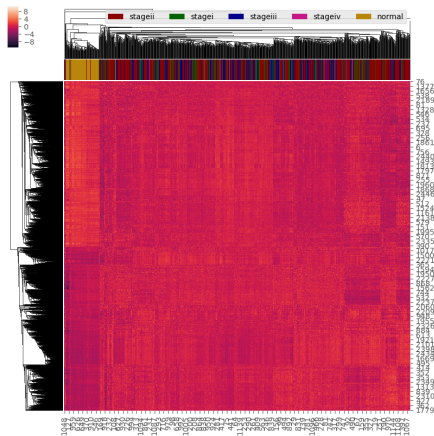 - Entropic Gromov Wasserstein
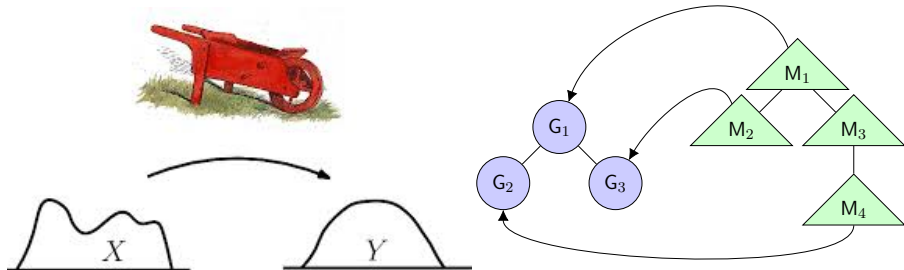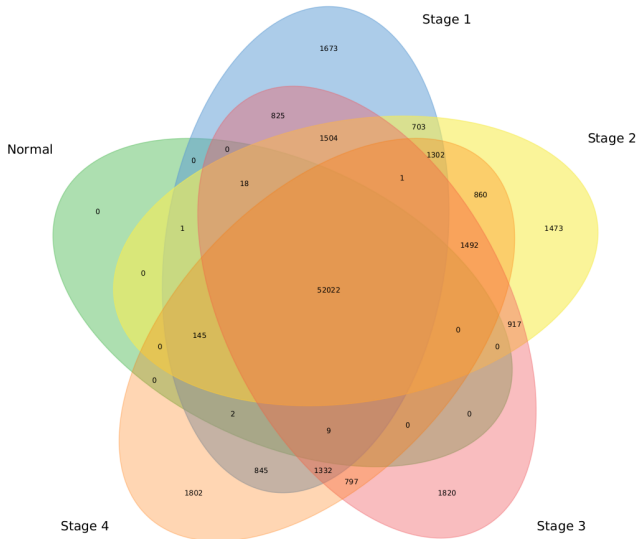
Data

- Data: TCGA-BRCA
- Samples:
  - Normal: 104
  - Stage 1: 179
  - Stage 2: 608
  - Stage 3: 242
  - Stage 4: 20
- DEGs (fold-change > 2):
  - Up: 1218
  - Down: 1236
- 617 miRNAs
  - 1,678 co-binding relationships
  - 20,441/166,669 bindings make significant group difference
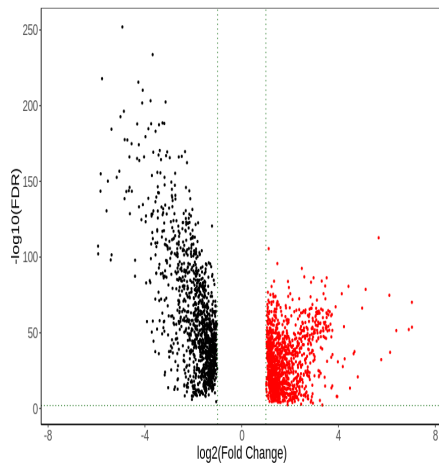
Results (cont'd)

Results (cont'd)

Results (cont'd)

Disease Enrichment on lncRNAs:

- DOID:2449 acromegaly and **GATA3-AS1**
- DOID:299 adenocarcinoma and **PVT1**
- DOID:3355 fibrosarcoma, DOID:8791 breast carcinoma in situ, DOID:8719 in situ carcinoma and **LINC00987**

Results (cont'd)

Low degree miRNAs that make group difference:

- hsa-miR-129-5p
- hsa-miR-140-3p
- hsa-miR-146b-5p
- hsa-miR-188-5p
- hsa-miR-193a-5p
- hsa-miR-28
- hsa-miR-346
- hsa-miR-3605-3p

- hsa-miR-361
- hsa-miR-455-5p
- hsa-miR-671-3p
- hsa-miR-320b
- hsa-miR-193a-3p
- hsa-miR-326
- hsa-miR-330
- hsa-miR-501-3p

Acknowledgement

Questions

Questions?