CS 5984: Topics and Schedule

T. M. Murali

January 19, 2006
Continuum of Models in Systems Biology

We will cover “high-level” models.  
Emphasise a data-driven approach to systems biology.  
Focus on large-scale properties of biological systems.  
Integrate massive quantities of different types of data  
Learn techniques from clustering, data mining, and graph theory and apply them to solve specific biological questions.
Sources of Data

- Gene expression data
- Gene knockouts and external perturbations such as drugs.
- Samples belonging to various classes
- Time-series data.
- GEO, SGD, the Whitehead institute.

- Protein-protein interaction data
- Large-scale Yeast 2-hybrid assays (yeast, worm, fruitfly).
- Affinity precipitation + mass spectometry (yeast).
- Literature (HPRD).

- Transcriptional regulation
- Protein-DNA binding data (yeast, human liver TFs).
- Binding profiles for known TFs (SCPD, TRANSFAC).
Sources of Data

- Gene expression data
  - Gene knockouts and external perturbations such as drugs.
  - Samples belonging to various classes
  - Time-series data.
  - GEO, SGD, the Whitehead institute.

- Protein-protein interaction data
  - Large-scale Yeast 2-hybrid assays (yeast, worm, fruitfly).
  - Affinity precipitation + mass spectometry (yeast).
  - Literature (HPRD).

- Transcriptional regulation
  - Protein-DNA binding data (yeast, human liver TFs).
  - Binding profiles for known TFs (SCPD, TRANSFAC).
Sources of Data

✿ Gene expression data
  ▶ Gene knockouts and external perturbations such as drugs.
  ▶ Samples belonging to various classes
  ▶ Time-series data.
  ▶ GEO, SGD, the Whitehead institute.

✿ Protein-protein interaction data
  ▶ Large-scale Yeast 2-hybrid assays (yeast, worm, fruitfly).
  ▶ Affinity precipitation + mass spectometry (yeast).
  ▶ Literature (HPRD).
Sources of Data

- Gene expression data
  - Gene knockouts and external perturbations such as drugs.
  - Samples belonging to various classes
  - Time-series data.
  - GEO, SGD, the Whitehead institute.

- Protein-protein interaction data
  - Large-scale Yeast 2-hybrid assays (yeast, worm, fruitfly).
  - Affinity precipitation + mass spectometry (yeast).
  - Literature (HPRD).

- Transcriptional regulation
  - Protein-DNA binding data (yeast, human liver TFs).
  - Binding profiles for known TFs (SCPD, TRANSFAC).
Sources of Data

- Literature, Computation, Databases
  - Transcriptional regulators (TRANSFAC)
  - Protein-protein interactions (DIP, GRID, Predictome, MIPS)
  - Metabolic networks (KEGG, EcoCyC, BioCarta, GenMAPP)
  - Functional annotations (GO, MIPS, species-specific databases)
  - Genetic Associations with Disease (GAD, MEDGENE, i-HOP).
Specific Topics

1. Clustering of molecular profiles
   ▶ Basic clustering, application to finding cancer gene modules.
   ▶ Biclustering, application to yeast cellular networks, interpretable disease classification.

2. Functional annotation of genes

3. Comparative systems biology
   ▶ Finding conserved protein interaction modules.
   ▶ Cross-species (bi)clustering of gene expression data.
Specific Topics

1. Clustering of molecular profiles
   ▶ Basic clustering, application to finding cancer gene modules.
   ▶ Biclustering, application to yeast cellular networks, interpretable disease classification.

2. Functional annotation of genes

3. Comparative systems biology
   ▶ Finding conserved protein interaction modules.
   ▶ Cross-species (bi)clustering of gene expression data.

4. Experimental detection, structure, and properties of biological networks.
   ▶ Transcriptional regulatory networks
   ▶ Protein interaction and signalling networks
   ▶ Metabolic networks

5. Prediction of disease outcome.

6. microRNAs.

Course Schedule

Week 1, Jan 17, 19  Topics and schedule.
Week 2, Jan 24, 26  Basic clustering, application to cancer gene modules.
Week 3, Jan 31, Feb 2  Biclustering algorithms.
Week 4, Feb 7, 9  Applications of biclustering.
Week 5, Feb 14, 16  Functional annotation.
Week 6, Feb 21, 23  Comparative systems biology.
Week 7, Feb 28, Mar 2  Comparative systems biology.
Week 8, Mar 14, 16  Data Integration
Week 9, Mar 21, 23  Student Group 1
Week 9, Mar 28, 30  Student Group 2
Week 10, Apr 4, 6  Student Group 3
Week 11, Apr 11, 13  Student Group 4
Week 12, Apr 18, 20  Student Group 5
Week 13, Apr 25, 27  Project presentations
Week 14, May 2  Wrap-up
**Gene Regulation**

Diagram showing the regulation of gene expression, focusing on activators (C/EBP, HNF1, HNF3, HNF4, AP1) and co-activators. The diagram also illustrates the involvement of general transcription factors (TFIA, TFIIB, TBP, TFIIF, TFIIE, TFIID, TFIH, Srb/Mediator) and the association with enhancers and promoter-proximal regions.
Regulatory Networks

Focus

Topics

T. M. Murali January 19, 2006 CS 5984: Topics and Schedule
Regulatory Networks

C Module A functions:

Vegetal plate expression in early development:

Synergism with modules B and G enhancing endoderm expression in later development:

Repression in ectoderm (modules E and F) and skeletogenic mesenchyme (module DC):

Modules E, F and DC with LiCl treatment:
Regulatory Networks

B
if \((F = 1 \text{ or } E = 1 \text{ or } CD = 1)\) and \((Z = 1)\)
\[\alpha = 1\]
else \[\alpha = 0\]

if \((P = 1 \text{ and } CG_r = 1)\)
\[\beta = 2\]
else \[\beta = 0\]
if \((CG_s = 1 \text{ and } CG_t = 1 \text{ and } CG_r = 1)\)
\[\gamma = 2\]
else \[\gamma = 1\]
\[\delta(t) = B(t) + G(t)\]
\[\epsilon(t) = \beta^*\delta(t)\]

if \((\epsilon(t) = 0)\)
\[\zeta(t) = Otx(t)\]
else \[\zeta(t) = \epsilon(t)\]
if \((\alpha = 1)\)
\[\eta(t) = 0\]
else \[\eta(t) = \zeta(t)\]
\[\Theta(t) = \gamma^*\eta(t)\]

Repression functions of modules F, E, and DC mediated by Z site
Both P and CG_r needed for synergistic link with module B
Final step up of system output
Positive input from modules B and G
Synergistic amplification of module B output by CG_s-P subsystem
Switch determining whether Otx site in module A, or upstream modules (i.e., mainly module B), will control level of activity
Repression function inoperative in endoderm but blocks activity elsewhere
Final output communicated to BTA
Signal Transduction Cascades

RTK → GNRP → RAS → RAF → MEK → ERK

Transcription factor → Transcription

Ligand → Receptor

RAS → GTP

GAP → GTP

Inactive transcription factor → Active transcription factor

Modulation of transcription
Protein-Protein Interaction Networks
Protein-Protein Interaction Networks
Metabolic Networks