CS 5984: Topics and Schedule

T. M. Murali

January 18, 2007
Continuum of Models in Systems Biology

Focus of the Course

- 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).

- Protein abundance and activity
- Metabonomics
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.
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

- 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).

- Protein abundance and activity
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).

- **Protein abundance and activity**

- **Metabonomics**
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. Predicting gene function

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

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

5. microRNAs.

Course Schedule

Weeks 1/2, Jan 16, 18, 23  Topics and schedule.
Week 2, Jan 25  Basic clustering
Weeks 3, Jan 30, Feb 1  Application to cancer gene modules.
Week 4, Feb 6 ,8  Biclustering algorithms.
Week 5, Feb 13, 15  Applications of biclustering.
Weeks 6, 7, Feb 20, 22, 27, Mar 1  Functional annotation.
Week 8, Mar 13, 15  Comparative systems biology.
Weeks 9, 10, Mar 20, 22, 27, 29  Structure of biological networks.
Week 10, Apr 4, 6  Predicting molecular interactions.
Week 11, Apr 11, 13  Network legos
Week 12, Apr 17, 19  Invited lectures
Week 13, Apr 24, 26  Project presentations
Week 14, May 1  Wrap-up
Gene Regulation
Regulatory Networks

Focus of the Course

T. M. Murali: 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 or E = 1 or CD = 1) and (Z = 1)
    α = 1
else
    α = 0
if (P = 1 and CG, = 1)
    β = 2
else
    β = 0
if (CG, = 1 and CG, = 1 and CG, = 1)
    γ = 2
else
    γ = 1
δ(t) = B(t) + G(t)
ε(t) = β*δ(t)

if (ε(t) = 0)
    ξ(t) = Otx(t)
else
    ξ(t) = ε(t)
if (α = 1)
    η(t) = 0
else
    η(t) = ξ(t)
Θ(t) = γ*η(t)

Repression functions of modules F, E, and DC mediated by Z site
Both P and CG, needed for synergistic link with module B
Final step up of system output
Positive input from modules B and G
Synergetic amplification of module B output by CG, -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

Diagram showing the pathway of signal transduction from Ligand to ERK, involving RTK, GNR, RAS, RAF, MEK, and GAP proteins.
Protein-Protein Interaction Networks
Protein-Protein Interaction Networks
Metabolic Networks