Networks and Modules: Lessons from Yeast to Humans
Aviv Regev

MIT Department of Biology
Broad Institute of MIT and Harvard

Reconstructing transcriptional circuits

From Parts to Function
Marcotte's Integrated functional network of yeast genes

Unifying concept: Modules
A set of biological entities that act collectively to perform an identifiable and distinct function

Unifying concept: Modules
Metabolism Molecular machines Signal transduction

Unifying concept: Modules
Operons, CRMs Complexes Regulons
Focus: Regulatory modules

A set of co-expressed and co-regulated genes

From modules to networks

Gene: Probabilistic function of regulators

Predict gene value based on its regulators

A Bayesian network view

- 300 deletion knockout in yeast [Hughes et al 2000]
- 600 genes

From Hubs to Modules

- Idea: enforce common regulatory program
- Robustness: Programs are estimated from m*k samples
- Biological interpretability: Organization of genes into regulatory modules
Module networks: Reconstructing modules

- A set of genes and their shared regulation program

Learning module networks

- Assumption: Predict target expression based on expression of its regulators

Learning regulation programs

**Goal:** Find regulator whose split best predicts behavior of module genes

Learning regulatory programs

Continue to split on each sub-tree

As long as the score improves
A module network for the yeast stress response

Learning Gene Assignments

- Each module has an expression model
- Assign each gene to module that best predicts it
- Reassignment guaranteed to improve overall predictiveness

EM iterations: Gene Assignment

Evaluating our convergence point

- EM sensitive to initialization
- Compare to
  - PCluster starting point
  - 100 runs initialized from randomized clusterings
  - 41%-52% are placed in the same module at convergence
  - 98/100 PCluster initialized runs score better (log likelihood of data and GO coherence)
  - Improvement in module coherence compared to PCluster starting point

Peaked co-expression (tight co-regulation) around specific sub-set of conditions
### Gene set coherence (GO, MIPS, KEGG)
- 41/50

### Match between regulator and targets
- 30/50

### Match between regulator and cis-reg motif
- 15/50

### Match between regulator and condition/logic
- 35/50

---

#### From modules to regulators

- **Regulators**

  - **Gene set coherence (GO, MIPS, KEGG)**
    - 41/50

  - **Match between regulator and targets**
    - 30/50

  - **Match between regulator and cis-reg motif**
    - 15/50

  - **Match between regulator and condition/logic**
    - 35/50

---

#### Condition specific regulation

- Significant enrichment of condition on one context (side of the “split”)
- **Best p-value on all splits**

---

#### Focused Prediction: Regulator ‘X’ activates process ‘Y’ under conditions ‘Z’
Focused predictions

Test:
An ‘X’ KO mutant under ‘Z’

Response under predicted conditions?

*Paired t-test to find differentially expressed genes*

Response in predicted modules?

*Rank modules by enrichment for diff. expressed genes*

Response of predicted processes?

*Test whether regulator-process annotations from matrices are enriched in diff. expressed genes*

---

**YPL230W**

Response under predicted conditions? Yes
Response in predicted modules? Yes
Response of predicted processes? Yes (partial)

---

**Kin82**

Response under predicted conditions? Yes
Response in predicted modules? Yes
Response of predicted processes? No

---

**PPT1**

Response under predicted conditions? Yes
Response in predicted modules? Yes (partial)
Response of predicted processes? Yes (partial)

---

**Why does it work (and where will it fail)?**

- Assumption: regulator’s expression is a proxy for activity
- But, many regulators are controlled post-transcription

---

**Why does it work (and where will it fail)?**

- Assumption: regulator’s expression is a proxy for activity
- But, many regulators are controlled post-transcription

---

**Microarray**