Inferring Regulatory Networks
Background

• Early studied covered relatively few regulatory factors

• Initial models often defined by Boolean expression assumption
  – Gene expression as \{0,1\} or ON/OFF
  – Many models were purely artificial, generated randomly

• Gene Regulatory Networks were sought after using mRNA profiling data alone (from microarrays)
  – Too many many GRN solutions can explain the expression output
Random Boolean Network (RBN) Representations

**Figure 2.** A three-node RBN with the logical updating functions of its nodes (left) and the corresponding state space (right). Nodes in the state space with self-loops are the network’s attractors; connected sets of nodes are basins of attraction.
Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data.

Method Overview

Segal et al. Summary

- Define “regulatory module” as TF and the genes they coordinately regulate
- Gene regulation network is hierarchical
- Regulated genes have promoter regions enriched for DNA motifs
- Regulation is combinatorial
- Generates testable hypothesis about the role of regulators
Regulatory Network for Cell Cycle Control

Functional Discovery via a Compendium of Expression Profiles

A.

selected transcript clusters

- PAU
- ergosterol
- amino acid biosynthesis
- PKC/calcineurin
- mating

Gcn4 down

Gcn4 up

ergosterol

cell wall

tup1, ssn6

ERG11
ERG3
YER044C
HIS3
ARG4
LEU4
CMK2
FKS2
AGA1,2
FUS3

ste7

gcn4

lovastatin

ery11

ery044c

tunicamycin

ery083c

mating

selected profile clusters

- mating
- Gcn4 down
- Gcn4 up
- ergosterol
- cell wall
- tup1, ssn6
Method

Validation and refinement of gene-regulatory pathways on a network of physical interactions
Chen-Hsiang Yeang*†, H Craig Mak*‡, Scott McCuine‡, Christopher Workman†, Tommi Jaakkola§ and Trey Ideker†

Addresses: *Center for Biomolecular Science and Engineering, Baskin School of Engineering, University of California at Santa Cruz, Santa Cruz, CA 95064, USA. †Department of Bioengineering, University of California at San Diego, La Jolla, CA 92033, USA. §Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

Infer inducer or repressor
Infer inducer or repressor

Diagram showing the interactions between DAL81, PDR1, and HOR7 genes.
Relevant Interactions

• Subset of Rosetta compendium used

• 28 deletions were TF (red circles)
  - 355 diff. exp. genes (white boxes)
  - 755 TF-deletion effects (grey squiggles)
Computational methods

- Problem Statement:
  - Find regulatory paths consisting of physical interactions that explain functional relationship

- Method:
  - A probabilistic inference approach
  - To assign annotations
  - Formalize problem using a factor graph
  - Solve using max product algorithm
    - Mathematically similar to Bayesian inference, Markov random fields, belief propagation
Test and refine approach

Alternate network models

Candidate single gene knock-outs

Knock-out priority scoring: (B, D, C, E, G, F)

Run top priority microarray experiments: (B, D)

Re-assembly

Remaining consistent models

Validation

Genes
- original knockout
- affected
- new knockout

Interactions
- Protein-DNA
  - inducer
  - activator
- Protein-protein
  - repressor
  - inhibitor
50/132 protein-DNA interactions had been confirmed in low-throughput assays (Proteome BioKnowledge Library).

Inferred regulatory roles (induction or repression) for 48 out of 50 of these interactions agreed with their experimentally determined roles. (96%, binomial $p$-value $< 1.22 \times 10^{-7}$)
Inferred Network Annotations

A network with ambiguous annotation
Target experiments to one network region

Expression for: SOK2Δ, HAP4 Δ, MSN4 Δ, YAP6 Δ
Expression of Msn4 targets

Msn4-regulated genes

Coherence

swi4Δ  sok2Δ  hap4Δ  msn4Δ  yap6Δ

Average Z-score

Unrelated control (Msn1-regulated genes)

swi4Δ  sok2Δ  hap4Δ  msn4Δ  yap6Δ

Negative control
Expression of Hap4 targets

Hap4-regulated genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Coherence</th>
</tr>
</thead>
<tbody>
<tr>
<td>swi4Δ</td>
<td>2.0</td>
</tr>
<tr>
<td>sok2Δ</td>
<td>1.0</td>
</tr>
<tr>
<td>hap4Δ</td>
<td></td>
</tr>
<tr>
<td>msn4Δ</td>
<td>-2.6</td>
</tr>
<tr>
<td>yap6Δ</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Diagram showing regulatory relationships between SWI4, SOK2, MSN4, and HAP4.
Refined Network Model

- Caveats
  - Assumes target genes are correct
  - Only models linear paths
  - Combinatorial effects missed
  - Measurements are for rich media growth
How are the maintenance and repair processes coordinated?

- Transcriptional regulation,
- Signaling cascades,
- Post-translational modifications
Figure 9. Summary of genomic responses to MMS and ionizing radiation. This diagram summarizes the functional features of the genomic expression responses observed in this study (purple), transcription factors (blue) and protein kinases (yellow) that have been implicated in those genomic responses, and the hypothetical cellular signals that trigger the responses (orange). Regulatory factors that were investigated in this study are indicated as colored ellipses.

Overview

- Experimental factors and selection
  - Multiple criteria used
- ChIP-on-chip
  - The technique
  - Differential binding analysis
- Gene expression analysis of TF-deletion mutants
  - Clustering analysis
  - Deletion-buffering analysis
- Data integration and pathway reconstruction
Transcription factors that regulate DNA damage responsive genes may be critical

Activated regulatory network

TF knockout

“Deletion-buffered”
Environmental “epistasis analysis”: (deletion-buffering)

<table>
<thead>
<tr>
<th>Genetic Factor</th>
<th>Environmental Factor</th>
<th>-E</th>
<th>+E</th>
</tr>
</thead>
<tbody>
<tr>
<td>+G</td>
<td>(WT, control)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-G</td>
<td>(KO, control)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mRNA Exp</td>
<td>mRNA Exp</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(WT, treated)
Specific deletion-buffering effects

A

MBP1  SWI6

DUN1  RFA1

RAD9  MEC1

RNR4  RFX1

RNR3  SRB4

Wild Type (+/- MMS)

<table>
<thead>
<tr>
<th>protein-DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>YPD</td>
</tr>
<tr>
<td>YPD or stress</td>
</tr>
<tr>
<td>stress only</td>
</tr>
<tr>
<td>MMS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>protein-protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>proteins</td>
</tr>
<tr>
<td>TF</td>
</tr>
<tr>
<td>TF in MMS study</td>
</tr>
<tr>
<td>enzyme</td>
</tr>
</tbody>
</table>

**log_{10}(ratios)**

-0.3  0  0.3
Introduction to the Cytoscape Exercises

Regulatory Networks
Aim of the exercise

• Gain experience integrating data on larger, mixed interaction networks
• Get experience with a realistic workflow:
  – Build interaction network
  – Annotate nodes and interactions
  – Filter based on attributes, network properties or both
  – Set visualization options
P-value and log-ratio

- P-value: The chance of rejecting the null hypothesis by coincidence

- For gene expression analysis we can say:
  the chance that a gene is categorized as differentially expressed by coincidence

- The lower the value, the more significant is the difference in gene expression → the gene is either up- or down-regulated.

- Log-ratio: \( \log_2(\frac{X_{\text{exp}}}{X_{\text{ctr}}}) \)
  - \( X_{\text{exp}} \) is the expression level in the experiment
  - \( X_{\text{ctr}} \) is the expression level in the control
Integrated Genomic and Proteomic Analyses of a Systematically Perturbed Metabolic Network

Trey Ideker, Vesteinn Thorsson, Jeffrey A. Ranish, Rowan Christmas, Jeremy Buhler, Jimmy K. Eng, Roger Bumgarner, David R. Goodlett, Ruedi Aebersold, Leroy Hood

www.sciencemag.org  SCIENCE  VOL 292  4 MAY 2001

Fig. 1. Model of galactose utilization. Yeast metabolize galactose through a series of steps involving the GAL2 transporter and enzymes produced by GAL1, GAL7, GAL10, and GAL5. These genes are transcriptionally regulated by a mechanism consisting primarily of GAL4, GAL80, and GAL3. GAL6 produces another regulatory factor thought to repress the GAL enzymes in a manner similar to GAL80. Dotted interactions denote model refinements supported by this study.
Integrated Genomic and Proteomic Analyses of a Systematically Perturbed Metabolic Network

Trey Ideker,1,2* Vesteinn Thorsson,1,2 Jeffrey A. Ranish,1,2 Rowan Christmas,1 Jeremy Buhler,3 Jimmy K. Eng,1 Roger Bumgarner,4 David R. Goodlett,1 Ruedi Aebersold,1,2 Leroy Hood1,2

www.sciencemag.org SCIENCE VOL 292 4 MAY 2001