Reconstructing signaling and dynamic regulatory networks
Input – Output Hidden Markov Model

Bengio and Frasconi, *NIPS* 1995

Input (Static transcription factor-gene interactions)

Hidden States Variables
(We constrain transitions between states to form a tree structure)

Output State Variables
(Gaussian distribution for expression values)

Log Likelihood

\[
r(G|M) = \sum_{g \in G} \sum_{q \in Q} \prod_{t=1}^{n-1} f_{q(t)}(o_g(t)) \prod_{t=1}^{n-1} P(H_t = q(t) | H_{t-1} = q(t-1), I_g)
\]

Product over all transition probabilities on path
Product over all Gaussian emission density values on path
Sum over all genes
Sum over all paths Q
Stress and hormone response


Stem cells differentiation

MSB 2011

Immune response

Fly development

Science 2010

DREM is useful, but several questions remain …

Who controls the master regulators?
Linking the signaling and regulatory networks

What’s going on here?!?
SDREM: Extending DREM to model signaling networks

Inputs:

• Condition specific inputs:
  - Time series expression data following treatment
  - (A few) receptors interacting with invader or activated by condition of interest

• General interaction data (not necessarily from the same condition):
  - Protein-DNA interactions
  - Motif information
  - Protein interaction networks
Iterative method for reconstructing dynamic signaling and regulatory networks

Identify TFs actively regulating gene expression

Determine which active TFs are well-connected in the PPI network?
How do we orient the network?

• We are given the receptors / sensing proteins and the TFs
• We need to determine an appropriate path from receptors to TFs
• The orientation should
  – Use short source-target pathways
  – Prefer high-confidence interactions
  – Encourage parallel pathways
Source target pairs:
\{A,E\}, \{F,B\}
Maximum Edge Orientation (MEO)

- Mixed graph $G = (V, E)$
- Known set of sources $S$ and targets $T$ (from DREM)
- Maximum path length $k$
- Consider all simple paths $P$ from $s \in S$ to $t \in T$
- Path weights given by

$$\sum_{p \in P} I_S(p) \cdot w(p)$$

where $I_S(p)$ indicates if $p$ is satisfied

Can be converted into a satisfiability problem and approximated using SAT solvers
Identify TFs actively regulating gene expression

Determine which active TFs are well-connected in the PPI network?

Yeast response to osmotic stress
HOG pathways analysis

- The high osmolarity glycerol (HOG) pathway is activated by increased environmental osmolarity and results in a rise of the cellular glycerol concentration.

- **Condition specific input:**
  - Time series expression data after treatment with sorbitol
  - 5 known proteins that sense this condition

[Diagram showing the HOG pathway with proteins and annotations]

Capaldi et al *Nature Genetics* 2008
Yeast response to osmotic stress

Identify TFs actively regulating gene expression

Determine which active TFs are well-connected in the PPI network?
Reconstructed HOG pathway: Short time series

Expression data from Romero-Santacreu et al RNA 2009

Gitter et al Genome Research 2013
Reconstructed HOG pathway:
Long time series

Even though *less than half* of the DE genes were shared between the two experiments, our reconstructed networks had a very large overlap: Of the 19 TFs in the short network, 16 were also identified in the long model.
HOG predictions

- We used two different HOG time series expression datasets

**Short time series**
- 11 of 19 TFs in gold standard (59%)
- 27 of 39 internal proteins in gold standard (69%)

**Long time series**
- 13 of 28 TFs in gold standard (46%)
- 16 of 23 internal proteins in gold standard (70%)
Validating computational predictions

• Predictions for active TFs

Before treatment  3 min after treatment

• Knockouts validate internal pathway members

Gitter et al Genome Research 2013
Applications of SDREM to immune response
Applications to viral infection

- Treatments directly targeting virus can fail
  - Viral mutation
- Instead interested in host response

- Reconstruct infection model
  - Start with host-virus PPI
  - Find pathways in human PPI network
  - Connect the TFs driving observed transcription
Disease-specific pathways

- Human networks contain millions of paths
  - Want pathways specific to the disease

- RNAi screening provides functional relevance

- Not reproducible across labs

<table>
<thead>
<tr>
<th>Virus</th>
<th>Independent screens</th>
<th>Genes common to 3+ screens</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1N1 influenza</td>
<td>5</td>
<td>0.7%</td>
</tr>
<tr>
<td>HIV</td>
<td>3</td>
<td>0.4%</td>
</tr>
</tbody>
</table>
Incorporating RNAi screens

Screens suggest relevant nodes in network

\[ w(p_1) = w(t) \prod_{v \in p_1} w(v) \prod_{e \in p_1} w(e) \]
H1N1: temporal TF activity

Recover major immune response TFs

12 groups of similarly responding genes

log$_2$ fold change

Time

HIF1A
ELK1
AHR
AIRE
MYC
TP53

RELANFKB2

TFAP2CTFAP2AEP300

IRF6
IRF4
NFKB2
IRF2
NFATC1

STAT1

XBP1

IRF3
IRF5
IRF7
IRF8

ESR1
MYOD1
FOXO1

E2F1
TFAP2A
TFAP2C
DSP
RB1
SOX9
HIF1A
TFDP1

PPARA
NR2F1
IRF2
HSF1
BRCA1
H1N1: signaling pathways

Source proteins

Internal proteins

Active TFs

Pathways contain many RNAi hits

Enriched for immune and viral GO terms

Active TFs

RNAi hits

Enriched for immune and viral GO terms
H1N1: signaling pathways

RNAi effect predictions can suggest which proteins to examine

RNAi hit

Source proteins

Internal proteins

Active TFs

UBE2I  TRIM28  HSP90AA1

AKT1  MDM2

RUNX1

EP300

IRF2  MYC
Predicting RNAi screen hits

\[ \sum_{T} \frac{\text{remaining path weight to target}}{\text{original path weight to target}} \leq |T| \]
Predicting RNAi screen hits

- Hold out all RNAi data and rerun SDREM
- Predict H1N1 knockdown effects

<table>
<thead>
<tr>
<th></th>
<th>Top 10</th>
<th>Top 20</th>
<th>Top 50</th>
<th>Top 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct predictions</td>
<td>6</td>
<td>8</td>
<td>18</td>
<td>42</td>
</tr>
<tr>
<td>Significance</td>
<td>1.97 E-5</td>
<td>3.44 E-5</td>
<td>3.24E-9</td>
<td>9.42 E-23</td>
</tr>
</tbody>
</table>

- Can also be used to make predictions about double KO which are infeasible to test genome-wide
- SDREM produces ranked pairs to test
  - Disease-specific predictions
Predicting genetic interactions

Redundant paths

Downstream TF only affected by double knockdown
Predicting genetic interactions

• Genetic interaction definition from pioneering yeast studies

\[ g = ob_{AB} - ex_{AB} = ob_{AB} - ob_A ob_B \]

• *In silico* observed phenotype is calculated using the directed pathways

\[ ob_A = \sum_{T} \frac{\text{remaining path weight to target}}{\text{original path weight to target}} \]

• Calculate same *in silico* phenotype for double knockdowns