Multiple Perturbation Mapping of Biological Systems

Gary Bader
Donnelly Centre for Cellular and Biomolecular Research
University of Toronto
Centennial Bioinformatics, Mar. 2 2009

http://baderlab.org
The Cell

How does it work?

How does it fail in disease?
Computational Cell Map

Map the cell
• Predict map from genome
• Active cell map
• Multiple perturbation mapping
• Map visualization and analysis software

Read map to understand
• Cell processes
• Gene function
• Disease effects
• Map evolution


Computational Cell Map

Map the cell
• Predict map from genome
• Active cell map
• Multiple perturbation mapping
• Map visualization and analysis software

Read map to understand
• Cell processes
• Gene function
• Disease effects
• Map evolution


Multiple Perturbation Mapping

Input

Black Box

Output
Direct Effect

Input

Black Box

Output

A Off → Light Off

A Off

B

C

D
Direct Effect

A Off → Light Off
Parallel Paths

Input

Black Box

Output

A

B

C

D
Parallel Paths

Input

Black Box

Output

B&C Off → Light Off
Parallel Paths

A & B

Input

Black Box

Output

B & C Off → Light Off
Upstream Effect

Input: A, B, C, D

Black Box

Output: Light bulb
Upstream Effect

A masks A effect
Upstream Effect

Input
A
B
C
D

Black Box

Output

D masks A effect
Our Black Box: Budding Yeast

- a.k.a. Baker’s Yeast, Brewer’s Yeast
- *Saccharomyces cerevisiae* (Fungi)
  - Greek “sugar mold” + Latin “of beer”
- >6000 genes (parts)
Why Yeast?

• Simple system (single cell)
• Easy to work with in a lab
• >20% yeast genes in human genome
  – Similar core processes
  – Nobel prize 2001: cell growth and cancer
• Pathogenic yeast, anti-fungal drugs
• We farm it: bread, beer, wine
  – Fermentation: sugar $\rightarrow$ ethanol, CO2
Mapping Yeast

6000 Genes  
Yeast Cell  
Growth  

Single Gene Deletions: 1000 Genes Essential

Double Gene Deletions

A
\[ \text{Wild-type} \]
B
aΔ
\[ \times \]
X
B
bΔ
\[ \times \]
Viable
Viable
Lethal

Genetic Interactions

Pathway A

A1
\[ \text{Essential biological function} \]
A2
A3
\[ \text{Cell proliferation} \]

Pathway B

B1
B2
B3

SL interaction

Brenda Andrews and Charlie Boone, UofT
Three Basic Types of Genetic Interactions

Fitness (colony size)

wt: 1
a: 0.5
b: 0.5
ab: 0.25 (ab = a x b)

“Neutral”
Expected Result
Multiplicative Model
Three Basic Types of Genetic Interactions

- **wt**: Fitness (colony size) = 1
- **a**: Fitness = 0.5
- **b**: Fitness = 0.5
- **ab**: Fitness = 0.25

- *Negative* Synthetic Lethal
- *Neutral* Expected Result

Parallel systems

Pathway A
- A1
- A2
- A3

Pathway B
- B1
- B2
- B3
Three Basic Types of Genetic Interactions

**Fitness (colony size)**

- **wt**: 1
- **a**: 0.5
- **b**: 0.5
- **ab**: 0.5

**“Neutral” Expected Result**: e.g., two genes whose products are in the same nonessential system

**“Positive”**
Three Basic Types of Genetic Interactions

- **wt**
- **a**
- **b**
- **ab**

**Fitness (colony size)**

- "Neutral" Expected Result
- "Positive" Expected Result
- "Negative" Synthetic Lethal

- 1
- 0.5
Large-scale Mapping of Genetic Interactions in Yeast

6000 Yeast Genes:

1000 Essential Genes
5000 Nonessential Genes

- Genetic Array ~ 5000 Viable Yeast Deletion Mutants
- ~ 1000 conditional alleles of essential genes
- Automated Genetics
- Examine 36 Million Double Mutants & Map GIs

Brenda Andrews and Charlie Boone, UofT
Synthetic Genetic Array (SGA)

Tong et al, 2001
First Genetic Interaction Network

- 8 screens
- 291 interactions
- 204 genes

Tong, 2001
Large-Scale Genetic Interaction Network

132 Screens
4000 Interactions
1000 Genes

~100,000 Interactions/genome

Tong, 2004
Automated Yeast Genetics

BioRad Colony Arrayer

SGA ROBOTICS @ CCBR

S&P ROBOTICS
BIOMATRIX

200 genome-wide screens/month
Quantitative Genetic Interactions from Double Mutant Colony Growth Modeling

Model: double mutant colony size is multiplicative in biological and experimental factors

\[ C_{ij} = \alpha f_i f_j \varepsilon_{ij} t \text{pos}(C_{ij}) \text{comp}(C_{ij}) \text{batch}(C_{ij}) \ldots e \]

Experimental Factors
- E.g. Plate Position
- Nutrient Competition
- Screen Batch

Log-normal

Challenge: Most of the variation in colony sizes is due to systematic experimental factors, which we can estimate and normalize out, not biological effects
Fitting the Model: Deriving Precise Measures of Double and Single Mutant Fitness

Experimental factor normalization:
- Camera adjustment
- Plate normalization
- Competition normalization
- Spatial gradient normalization
- Plate-specific row/col. normalization
- Batch normalization
- Cell # adjustment

\[ C_{ij} = \alpha f_i f_j \varepsilon_{ij} \, t \, \text{pos}(C_{ij}) \, \text{comp}(C_{ij}) \, \text{batch}(C_{ij}) \ldots e \]

Single mutant fitness estimation

Genetic interaction/double mutant fitness estimation

(with error estimates)

\[ f_i, f_j, \varepsilon_{ij} \]
Normalizing Spatial and Batch Effects

Unnormalized interactions

- Array Plates (1-14)

Screens

Normalized interactions

Array Plates (1-14)

Screens

Correction

- Negative interaction (SL/SS)
- Positive interaction
SGA Interaction Score Distribution

~$3 \times 10^6$ gene pairs tested

At 95% confidence:
~45,000 negative interactions
~20,000 positive interactions

More negatives than positives
Full genome SGA matrix
994 x ~5000
Quantitative Genetic Interaction Data Display Wall, Princeton

Chad Myers
Olga Troyanskaya
Anastasia Baryshnikova
Full genome SGA matrix
994 x \~5000
Full genome SGA matrix
994 x ~5000

DNA replication & repair
Vacuolar ATPase
Retrograde transport
Retrograde-transport complex
(endosome-Golgi transport)
Vacuolar ATPase
(organelle acidification)
SGA genetic interaction matrix

- Red: aggravating interaction
- Green: alleviating interaction
Vacuolar H\(^+\) ATPase
Retrograde complex

System Reconstruction: Monochromatic effects

Vacuolar H\(^+\) ATPase

Retrograde complex

ER assembly
The cell map

4000 nodes
13,754 edges
The cell map

2068 nodes
7400 edges

- Ribosomes
- Mitochondria
- DNA repair
- Cell cycle, chrom. segregation
- Nucleus localization
- Microtubules/mitosis
- Cell polarity
- Protein modification
- Secretion
- Signaling
- Peroxisomes
- RNA processing
- Transcription, Chromatin modification
- Mitochondria
- Chromatin remodelling
- DNA repair
System Level Map of Yeast

- RNA processing
- metabolism
- secretion/trafficking
- cell cycle/tubulin cytoskeleton
- polarity/actin cytoskeleton/cell wall
- DNA replication and repair
- transcription
- general transport: small molecule, ion, drug
- protein biosynthesis/modification
- mitochondria/energy/peroxisome
Systems Level Map

Edge Significance
Z-score > 10.0
Systems Level Map of Yeast
Genetic interactions with morphological profiling

single mutants, bim1 & bni1 double mutants

40 parameters per cell
4,800-16,000 data points per mutant
~1,000 statistical parameters per mutant

- % unbudded, small, medium and large budded cells
- cell length, breadth, elliptical factor, area
- number of spindle pole bodies
- length, position, orientation of the spindle
- breadth of the budneck
- distance of the spindle from the budneck
- ...
Cell map exploration and analysis

Can we accurately predict genetic interactions?

Pathway Information

Pathway Analysis (Cytoscape)

Databases

Literature

Expert knowledge

Experimental Data
Complete Listing of All Pathguide Resources

Pathguide contains information about 222 biological pathway resources. Click on a link to go to the resource home page or 'Details' for a description page. Databases that are free and those supporting BioPAX, CellML, PSI-MI or SBML standards are respectively indicated.

If you know of a pathway resource that is not listed here, or have other questions or comments, please send us an e-mail.

### Protein-Protein Interactions

<table>
<thead>
<tr>
<th>Database Name</th>
<th>Full Record</th>
<th>Availability</th>
<th>Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>3DID - 3D interacting domains</td>
<td>Details</td>
<td>Free</td>
<td></td>
</tr>
<tr>
<td>ABCdb - Archaea and Bacteria ABC transporter database</td>
<td>Details</td>
<td>Free</td>
<td></td>
</tr>
<tr>
<td>AICS - Alliance for Cellular Signaling Molecule Pages Database</td>
<td>Details</td>
<td>Free</td>
<td></td>
</tr>
<tr>
<td>AllFuse - Functional Associations of Proteins in Complete Genomes</td>
<td>Details</td>
<td>Free</td>
<td></td>
</tr>
<tr>
<td>ASEdb - Alanine Scanning Energetics Database</td>
<td>Details</td>
<td>Free</td>
<td></td>
</tr>
<tr>
<td>ASPD - Artificial Selected Proteins/Pepptides Database</td>
<td>Details</td>
<td>Free</td>
<td></td>
</tr>
<tr>
<td>BID - Binding Interface Database</td>
<td>Details</td>
<td>Free</td>
<td></td>
</tr>
<tr>
<td>BIND - Biomolecular Interaction Network Database</td>
<td>Details</td>
<td>Free</td>
<td>PSI-MI</td>
</tr>
<tr>
<td>BindingDB - The Binding Database</td>
<td>Details</td>
<td>Free</td>
<td></td>
</tr>
<tr>
<td>BioGRID - General Repository for Interaction Datasets</td>
<td>Details</td>
<td>Free</td>
<td>PSI-MI</td>
</tr>
<tr>
<td>BRITE - Biomolecular Relations in Information Transmission and Expression</td>
<td>Details</td>
<td>Free</td>
<td></td>
</tr>
<tr>
<td>CA1Neuron - Pathways of the hippocampal CA1 neuron</td>
<td>Details</td>
<td>Free</td>
<td></td>
</tr>
<tr>
<td>Cancer Cell Map - The Cancer Cell Map</td>
<td>Details</td>
<td>Free</td>
<td>BioPAX</td>
</tr>
<tr>
<td>CSP - Cytokine Signaling Pathway Database</td>
<td>Details</td>
<td>Free</td>
<td></td>
</tr>
<tr>
<td>CTDB - Celmodulin Target Database</td>
<td>Details</td>
<td>Free</td>
<td></td>
</tr>
<tr>
<td>DDIB - Database of Domain Interactions and Bindings</td>
<td>Details</td>
<td>Free</td>
<td></td>
</tr>
<tr>
<td>DIP - Database of Interacting Proteins</td>
<td>Details</td>
<td>Free</td>
<td>PSI-MI</td>
</tr>
<tr>
<td>Doodle - Database of oligomerization domains from lambda experiments</td>
<td>Details</td>
<td>Free</td>
<td></td>
</tr>
<tr>
<td>DopaNet - DopaNet</td>
<td>Details</td>
<td>Free</td>
<td></td>
</tr>
<tr>
<td>DRC - Database of Ribosomal Crosslinks</td>
<td>Details</td>
<td>Free</td>
<td></td>
</tr>
<tr>
<td>DSM - Dynamic Signaling Maps</td>
<td>Details</td>
<td>Free</td>
<td></td>
</tr>
<tr>
<td>FIMM - Functional Molecular Immunology</td>
<td>Details</td>
<td>Free</td>
<td></td>
</tr>
<tr>
<td>FusionDR - Prokaryote Gene Fusion Events</td>
<td>Details</td>
<td>Free</td>
<td></td>
</tr>
</tbody>
</table>
BioPAX Pathway Language

• Represent:
  – Metabolic pathways
  – Signaling pathways
  – Protein-protein, molecular interactions
  – Gene regulatory pathways
  – Genetic interactions

• Community effort: pathway databases distribute pathway information in standard format
<bp:biochemicalReaction rdf:ID="biochemicalReaction37">
  <bp:DATA-SOURCE rdf:resource="#dataSource14"/>
  <bp:LEFT>
    <bp:physicalEntityParticipant rdf:ID="physicalEntityParticipant26">
      <bp:STOICHIOMETRIC-COEFFICIENT>1.0</bp:STOICHIOMETRIC-COEFFICIENT>
      <bp:PHYSICAL-ENTITY>
        <bp:smallMolecule rdf:ID="smallMolecule27">
          <bp:SHORT-NAME rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
            a-D-glucose-6-P
          </bp:SHORT-NAME>
          <bp:CHEMICAL-FORMULA rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
            C6H13O9P
          </bp:CHEMICAL-FORMULA>
          <bp:SYNONYMS rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
            &lt;FONT FACE="Symbol">&amp;lt;/FONT&gt; d-glucose-6-phosphate</bp:SYNONYMS>
        </bp:smallMolecule>
        <bp:xref rdf:ID="unification#xref30">
          <bp:ID rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
            C00668</bp:ID>
          <bp:DB rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
            KEGG/DB</bp:DB>
          <bp:xref/>
        </bp:xref>
      </bp:PHYSICAL-ENTITY>
    </bp:physicalEntityParticipant>
    <bp:CHEMICAL-STRUCTURE rdf:ID="chemicalStructure28">
      <bp:structure-format>SMILES</bp:structure-format>
      <bp:structure-data>C(OP(=O)(O)O)[CH]1([CH](O)[CH](O)[CH](O)[CH](O)[CH](O)O)1</bp:structure-data>
    </bp:CHEMICAL-STRUCTURE>
  </bp:LEFT>
  <bp:DELTA-G rdf:datatype="http://www.w3.org/2001/XMLSchema#double">
    0.4</bp:DELTA-G>
  <bp:SYNONYMS rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    alpha-D-glucose 6-phosphate &lt;=> beta-D-fructose 6-phosphate</bp:SYNONYMS>
  </bp:DELTA-G>
  <bp:physicalEntityParticipant rdf:ID="physicalEntityParticipant38">
    <bp:CELLULAR-LOCATION rdf:resource="#openControlledVocabulary15"/>
    <bp:physicalEntityParticipant>
      <bp:DELTA-G rdf:datatype="http://www.w3.org/2001/XMLSchema#double">
        0.4</bp:DELTA-G>
      <bp:SYNONYMS rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
        alpha-D-glucose 6-phosphate &lt;=> beta-D-fructose 6-phosphate</bp:SYNONYMS>
      </bp:DELTA-G>
    </bp:physicalEntityParticipant>
  </bp:physicalEntityParticipant>
</bp:biochemicalReaction>
cPath Pathway Database Software

Collect → Browse / Query → Analyze

Cluster

Query

Analyse

Web Site

Cytoscape

XML Web Services API
Pathway Commons: A Public Library

http://pathwaycommons.org

- Books: Pathways
- Lingua Franca: BioPAX
- Index: cPath pathway database software
- Translators: translators to BioPAX
Network visualization and analysis

Pathway comparison
Literature mining
Gene Ontology analysis
Active modules
Complex detection
Network motif search

UCSD, ISB, Agilent, MSKCC, Pasteur, UCSF
Challenges

• Data: Author entry systems
  – From individual publications
  – For pathways (review)
  – Curator tools (advanced)
• Semantic integration (Identifier resolution)
• Visualization
  – Pathway diagrams (SBGN)
  – Automated layout
• Algorithms for compound graphs
• Linking discrete and dynamic representations
  – Including use by modelers
Where we want to be with cellular visualization…

CELL!!!
Acknowledgements

Genetic Interactions
Charlie Boone (UofT)
Anastasia Baryshnikova
Michael Costanzo
Huiming Ding
Judice L. Y. Koh
Kiana Toufighi
Brenda Andrews (UofT)
Chad Myers (UMinn)
Olga Troyanskaya (Princeton)

Bader Lab
MP
Anastasija Baryshnikova
Ben Elliott
Iain Wallace
Laetitia Morrison
G2N
Chris Tan
David Gfeller
Moyez Dharsee
Shirley Hui
ACM
Daniele Merico
Ruth Isserlin
Vuk Pavlovic

Cytoscape
Trey Ideker (UCSD)
Ryan Kelley, Kei Ono, Mike Smoot, Peng Liang Wang (Nerius Landys, Chris Workman, Mark Anderson, Nada Amin, Owen Ozier, Jonathan Wang)

Lee Hood (ISB)
Sarah Killcoyne (Iliana Avila-Campillo, Rowan Christmas, Andrew Markiel, Larissa Kamenkovich, Paul Shannon)

Benno Schwikowski (Pasteur)
Melissa Cline, Tero Aittokallio

Chris Sander (MSKCC)
Ethan Cerami, Ben Gross (Robert Sheridan)

Pathway Commons
Chris Sander
Ethan Cerami
Ben Gross
Emek Demir
Robert Hoffmann

Funding
CIHR, NSERC, NIH
Genome Canada
Canada Foundation for Innovation/ORF

http://baderlab.org