Experimental Techniques 2
• High-throughput interaction detection
• Yeast two-hybrid - pairwise
  • organisms as machines to learn about organisms
  • yeast, worm, fly, human,...
  • low intersection between repeated experiments
  • *in vivo*, but takes place inside the nucleus.
  • Estimated 50% FP rate
• TAP-MS (co-immunoprecipitation) - complexes
Tandem Affinity Purification (Puig et al, 2001)

Want to find interaction partners for protein encoded by this gene:

Add a tag to the end of its DNA sequence.

Calmodulin binding peptide

“Protein A” from Staphylococcus aureus
Binds to IgG protein

TOBACCO ETCH VIRUS PROTEASE

TEV protease cleavage (cutting) site

Glu-Asn-Leu-Tyr-Phe-Gln-Gly
Fishing for Proteins

- Tag may not be exposed
- Tag may change folding / binding properties
- Tag may change expression levels

Co-complexed proteins

Tagged protein

Contaminants

Grab with Immunoglobulin G protein

Wash contaminants cleave with TEV

Retrieve with calmodulin beads

IgG

Calmodulin
Sequencing Proteins (Tandem Mass Spectrometry)

Trypsin digestion

AAVEK → Tandem Mass Spec → b-ions:
A
AA
AAV
AAVE
AAVEK

y-ions:
K
EK
VEK
AVEK

Database of known or predicted spectra

Frequency of seeing given mass

Search

Database (KYNLSQMDVFVK)
Gavin et al, 2002 Results:

- 589 tagged proteins (78% of which returned some interaction partners)
- 232 complexes (grouping those with substantial overlap)
- Covering 1440 proteins
- Not binary interactions
- In this picture: edges mean complexes share a protein
Other ways to Convert to a Graph

Goll & Uetz, 2006
TAP-MS vs. Yeast 2 Hybrid

**Yeast 2-hybrid:**
- Pro: better at transient interactions (b/c they only have to happen long enough to “turn on” the reporter gene)
- Con: take place in nucleus (may be unnatural)
- Con: only binary interactions

**TAP-MS:**
- Pro: can find higher-order interactions (> binary)
- Con: requires more stable interactions
Ho et al, 2002 Results:

725 yeast proteins chosen to be “bait”:

<table>
<thead>
<tr>
<th>#</th>
<th>Protein Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>Kinases</td>
</tr>
<tr>
<td>36</td>
<td>Phosphatases</td>
</tr>
<tr>
<td>86</td>
<td>DNA damage response</td>
</tr>
<tr>
<td>503</td>
<td>Other proteins</td>
</tr>
</tbody>
</table>

600 baits worked (~10% of yeast proteins)
493 specific baits
1,578 proteins involved in ≥ 1 interaction
3,617 interactions
Kinases / Phosphatases

**kinase**: class of enzyme (protein) that adds a phosphate group to other molecules (usually a protein).

**phosphorylation**: the process of adding a phosphate group (PO₄) to a protein.

Phosphorylation often changes the shape (conformation) of a protein, thereby turning it “on” or “off”.

For example, phosphorylation can make a hydrophobic residue hydrophilic.

It is an important regulatory mechanism.

Estimate: >30% of proteins are phosphorylated in humans

- 518 known kinases in human
- 122 known kinases in yeast
Von Mering et al, 2002 Comparisons

10,907 “trusted” interactions from YPD & MIPS (T)

Coverage = % of T also in D

Accuracy = % of D also in T

Combining methods again helps significantly. (But of 80,000 predicted interactions, only 2,400 were seen in more than 1 method.)
Von Mering estimate for # of interactions in yeast

- $M =$ interactions seen more than once (2,400)
- 1/3 of them were previously known
- At the time: $\sim 10,000$ interactions known
- Therefore, expect 30,000 interactions total
- (Sprinzak et al estimate $\sim 16,000$)
Transcription network, aka regulatory network:

**Transcription Factors** = proteins that bind to DNA to activate or repress the nearby, downstream genes.

The regulated gene might also be a transcription factor.

leads to a directed graph.
ChIP-chip (ChIP-seq)

Chromatin immunoprecipitation - chip

TF Binds to DNA

TF Cross-linked to DNA (covalent bonds)

Cell is lysed, DNA fragmented

Antibodies used to pull out protein-DNA complexes

DNA is “read” using microarray or short-read sequencing
Synthetic Lethality

- Predicts a particular kind of functional interaction ("genetic interactions")
- "Synthetic" b/c manufactured mutations

\[
\begin{align*}
-A & = \text{survive} \\
-B & = \text{survive} \\
-A \& -B & = \text{die}
\end{align*}
\]

Proteins A and B are likely to be involved in similar functions.

A & B are "redundant" or complementary (parallel pathways)

Pretty course measurement
Explanations

- Two copies of the same protein.
- Complexes that can function without one of their constituent proteins.
- Two "redundant" pathways.
- 3 pathways, where any 2 are required.

Complex abcde can function when a single one of its proteins is removed, but not if 2 are removed.

A & B are “redundant” or complementary (parallel pathways)
SSL network from 2001

8 query genes
4500 “array” genes

(Tong et al., Science, 2001)
Databases & Software Resources
All data is freely available to both commercial and academic users for research purposes only and is provided WITHOUT ANY WARRANTY. Publications that make use of this data are requested to please cite Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: A General Repository for Interaction Datasets. Nucleic Acids Res. Jan1;34:D535-9.

Downloads of BIOGRID data is available via four types of data query: By Gene, By Publication, By Organism, and By Experimental System. Downloads by Organism and by System are available below, while downloads by Gene or by Publication are available by searching for your Gene of interest. In addition to these sets, you can also download the entire dataset in a single file by choosing ALL.

Each of these datasets is provided in three formats: Microsoft Excel Tab Delimited Text (TAB), PSI-MI XML Version 1 (PSI), and PSI-MI XML Version 2.5 (PSI25).

**The Following Downloads are Available Now**

Click the name of the file you wish to download...

**NOTICE:** Some curated datasets may not currently be available due to on-going updates/recompilation. All will be available shortly - if you would like to be notified upon availability please gridadmin@mskcc.org.

**Most Recent Release**

<table>
<thead>
<tr>
<th>File Name</th>
<th>File Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIOGRID-ALL-2.0.32.psi.zip</td>
<td>8.01 MB</td>
</tr>
<tr>
<td>BIOGRID-ALL-2.0.32.tab.zip</td>
<td>3.48 MB</td>
</tr>
<tr>
<td>BIOGRID-IDENTIFIERS-2.0.32.tab.zip</td>
<td>30.23 MB</td>
</tr>
<tr>
<td>BIOGRID-ORGANISM-2.0.32.psi.zip</td>
<td>8.11 MB</td>
</tr>
<tr>
<td>BIOGRID-ORGANISM-2.0.32.psi25.zip</td>
<td>11.33 MB</td>
</tr>
<tr>
<td>BIOGRID-ORGANISM-2.0.32.tab.zip</td>
<td>3.49 MB</td>
</tr>
<tr>
<td>BIOGRID-SYSTEM-2.0.32.psi.zip</td>
<td>14.4 MB</td>
</tr>
<tr>
<td>BIOGRID-SYSTEM-2.0.32.psi25.zip</td>
<td>16.72 MB</td>
</tr>
<tr>
<td>BIOGRID-SYSTEM-2.0.32.tab.zip</td>
<td>3.91 MB</td>
</tr>
</tbody>
</table>

**Interaction Statistics**

- Total Raw: 188109
- Total Raw Physical: 130309
- Total Raw Genetic: 57800
- Total Non-Redundant: 122582
- Non-Redundant Physical: 86199
- Non-Redundant Genetic: 36383

**Database Statistics**

- Proteins: 322372
- Publications: 20722
- Organisms: 13

BIND (part of BOND)

http://bond.unleashedinformatics.com/
Must register to use ("free")
Database of Interacting Proteins

Search by: [protein] [sequence] [motif] [article] [pathBLAST]

NODE SEARCH

Node Identifier

Node ID

Reset
Query DIP

or

Node Annotation

Name/Description

Organism

Keyword

Reset
Query DIP

http://dip.doe-mbi.ucla.edu/dip/Search.cgi?SM=3
Welcome to the MIPS *Saccharomyces cerevisiae* genome database

The MIPS Comprehensive Yeast Genome Database (CYGD) aims to present information on the molecular structure and functional network of the entirely sequenced, well-studied model eukaryote, the budding yeast *Saccharomyces cerevisiae*. In addition the data of various projects on related yeasts are used for comparative analysis.

News:

- 2006-12-04: Restart of the CYGD. Restructured entry views, GBrowse integration and new search engines for a more convenient use

http://mips.gsf.de/genre/proj/yeast/
STRING is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations; they are derived from four sources:

- Genomic Context
- High-throughput Experiments
- (Conserved) Coexpression
- Previous Knowledge

STRING quantitatively integrates interaction data from these sources for a large number of organisms, and transfers information between these organisms where applicable. The database currently covers 2,483,276 proteins from 630 organisms.

STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) is being developed at EMBL, SIB and UZH.


What's New? This is version 8.0 of STRING, including new datasources, an API, and better navigation. For more check our Blog.

http://string.embl.de/