Protein Interaction Mapping:
Use of Osprey to map Survival of Motor Neuron Protein interactions

Presented by:
Meg Barnhart
Computational Biosciences
Arizona State University
The Spinal Muscular Atrophy project

• Support provided by Dr. Ron Nieman, director of ASU NMR facility, through Families of SMA and NIH research grants

• Osprey and GRID databases created in the Mike Tyers lab at Samuel Lunenfeld research institute at Mt. Sinai Hospital in Ontario, Canada
Why study the SMN protein?

- Most common genetic cause of infantile death
- SMA results in the specific loss of motor neurons and the atrophy of voluntary muscle groups (legs, arms)
- Broad clinical spectrum: Type I (severe), Type II (intermediate), and Type III (adult).

Low full-length levels
Alternatively spliced product
--80% Δ7 (lacks exon 7)
No protection from SMA

~100 % full-length SMN
Protects from SMA
SMA research efforts

- *SMN* gene discovered in 1995

- In recent years, at least 4 putative or proven functions for SMN protein

- >40 *primary* interactions of SMN with other proteins
SMN functions

- pre-mRNA splicing
- Transcriptional regulation
- Ribosome production
- Axon-specific RNP transport
Pre-mRNA splicing

- First established function of SMN in cells – “master assembler” (1997)

Adapted from Paushkin, Gubitz et al. 2002
Proposed axon-specific function

- SMN observed moving bi-directionally along axons (2003)

(Zhang, Pan et al. 2003)
Network visualization of protein interactions: Osprey

• Stand alone application (Red Hat, OS X, Windows) or online viewer for GRID databases
The GRID

- General Repository for interaction datasets
- Files may be interactively imported from any of the GRIDs while running Osprey
GRID datasets
Osprey accepts several types of tab-delimited text files:

### File variation #1

<table>
<thead>
<tr>
<th>GeneA</th>
<th>GeneB</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMN1</td>
<td>FBP</td>
</tr>
<tr>
<td>Htra 2B</td>
<td>hnRNPg</td>
</tr>
<tr>
<td>Htra 2B</td>
<td>RBM</td>
</tr>
<tr>
<td>hnRNPr</td>
<td>SMN1</td>
</tr>
<tr>
<td>hnRNPs</td>
<td>SMN1</td>
</tr>
<tr>
<td>Gemin5</td>
<td>SMN1</td>
</tr>
</tbody>
</table>

### File variation #2

<table>
<thead>
<tr>
<th>GeneA</th>
<th>GeneB</th>
<th>GeneA screen name</th>
<th>GeneB screen name</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMN1</td>
<td>FBP</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Htra 2B</td>
<td>hnRNPs</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>Htra 2B</td>
<td>RBM</td>
<td>C</td>
<td>E</td>
</tr>
<tr>
<td>hnRNPr</td>
<td>SMN1</td>
<td>F</td>
<td>A</td>
</tr>
<tr>
<td>hnRNPs</td>
<td>SMN1</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>Gemin5</td>
<td>SMN1</td>
<td>G</td>
<td>A</td>
</tr>
</tbody>
</table>
## Input Files

### File variation #3

<table>
<thead>
<tr>
<th>GeneA</th>
<th>GeneB</th>
<th>Experimental System</th>
<th>Source</th>
<th>PubMedID</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMN1</td>
<td>FBP</td>
<td>CoIP HEK293</td>
<td>Williams et al</td>
<td>10734235</td>
</tr>
<tr>
<td>Htra 2B</td>
<td>hnRNPG</td>
<td>CoIP HEK293</td>
<td>Hofmann et al</td>
<td>12165565</td>
</tr>
<tr>
<td>Htra 2B</td>
<td>RBM</td>
<td>CoIP HEK293</td>
<td>Hofmann et al</td>
<td>12165565</td>
</tr>
<tr>
<td>HnRNPR</td>
<td>SMN1</td>
<td>CoIP HEK293</td>
<td>Rossoll et al</td>
<td>11773003</td>
</tr>
<tr>
<td>HnRNPQ</td>
<td>SMN1</td>
<td>CoIP HEK293</td>
<td>Rossoll et al</td>
<td>11773003</td>
</tr>
<tr>
<td>Gemin5</td>
<td>SMN1</td>
<td>CoIP HEK293</td>
<td>Gubitz et al</td>
<td>11714716</td>
</tr>
</tbody>
</table>
## Input Files

### File variation #4

<table>
<thead>
<tr>
<th>GeneA</th>
<th>GeneB</th>
<th>GeneA screen name</th>
<th>GeneB screen name</th>
<th>Experimental System</th>
<th>Source</th>
<th>PubMedID</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMN1</td>
<td>FBP</td>
<td>A</td>
<td>B</td>
<td>CoIP HEK293</td>
<td>Williams et al</td>
<td>10734235</td>
</tr>
<tr>
<td>Htra 2B</td>
<td>hnRNPG</td>
<td>C</td>
<td>D</td>
<td>CoIP HEK293</td>
<td>Hofmann et al</td>
<td>12165565</td>
</tr>
<tr>
<td>Htra 2B</td>
<td>RBM</td>
<td>C</td>
<td>E</td>
<td>CoIP HEK293</td>
<td>Hofmann et al</td>
<td>12165565</td>
</tr>
<tr>
<td>hnRNPr</td>
<td>SMN1</td>
<td>F</td>
<td>A</td>
<td>CoIP HEK293</td>
<td>Rossoll et al</td>
<td>11773003</td>
</tr>
<tr>
<td>hnRNpq</td>
<td>SMN1</td>
<td>D</td>
<td>A</td>
<td>CoIP HEK293</td>
<td>Rossoll et al</td>
<td>11773003</td>
</tr>
<tr>
<td>Gemin5</td>
<td>SMN1</td>
<td>G</td>
<td>A</td>
<td>CoIP HEK293</td>
<td>Gubitz et al</td>
<td>11714716</td>
</tr>
</tbody>
</table>
SMN input files

- File variation #3 was most useful for SMN data
- ProCite database was used to download papers containing terms “SMN”, “SMA.”, “survival motor neuron protein” and “spinal muscular atrophy” from NCBI
- Abstracts were manually reviewed for interaction information
- Selected papers were read & interactions were recorded in text files

<table>
<thead>
<tr>
<th>GeneA</th>
<th>GeneB</th>
<th>Experimental System</th>
<th>Source</th>
<th>PubMedID</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMN1</td>
<td>FBP</td>
<td>CoIP HEK293</td>
<td>Williams et al</td>
<td>10734235</td>
</tr>
<tr>
<td>Htra 2B</td>
<td>hnRNPG</td>
<td>CoIP HEK293</td>
<td>Hofmann et al</td>
<td>12165565</td>
</tr>
<tr>
<td>Htra 2B</td>
<td>RBM</td>
<td>CoIP HEK293</td>
<td>Hofmann et al</td>
<td>12165565</td>
</tr>
</tbody>
</table>
SMN input files

- Disadvantage: manual data mining and entry is a huge bottleneck area!
- SMN-Osprey network was generated by creation of several files, each defined by a single type of experimental binding system
User interface

- Pubmed link
A functional interaction between the survival motor neuron complex and RNA polymerase II.

Pellizzoni L, Charroux B, Rappsilber J, Mann M, Dreyfuss G.

Howard Hughes Medical Institute, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA.

The survival motor neuron (SMN) protein, the protein product of the spinal muscular atrophy (SMA) disease gene, plays a role in the assembly and regeneration of small nuclear ribonucleoproteins (snRNPs) and spliceosomes. By nanoelectrospray mass spectrometry, we identified RNA helicase A (RHA) as an SMN complex-associated protein. RHA is a DEAH box RNA helicase which binds RNA polymerase II (pol II) and reportedly functions in transcription. SMN interacts with RHA in vitro, and this interaction is impaired in mutant SMNs found in SMA patients. Coimmunoprecipitation demonstrated that the SMN complex is associated with pol II, snRNPs, and RHA in vivo. In vitro experiments suggest that RHA mediates the association of SMN with the COOH-terminal domain of pol II. Moreover, transfection of cells with a dominant negative mutant of SMN, SMNDeltaN27, causes accumulation of pol II, snRNPs, and RHA in nuclear structures that contain the known markers of gems and coiled bodies, and inhibits RNA pol I and pol II transcription in vivo. These findings indicate a functional as well as physical association of the SMN complex with pol II and suggest a role for the SMN complex in the assembly of the pol II transcription/processing machinery.
Superimposition of files

- Osprey can superimpose two or more user files differentiated by experimental system
SMN:GRID superimposition

- User files and GRID datasets may also be superimposed
Color indices

- Color of nodes may be user defined or defined by GO components.
- GO consortium (http://www.geneontology.org/) categorizes gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner.
Color indices, continued

- Edge colors can be user defined, defined by experimental process, or defined by author

Edges colored by source (author)

Same edges colored by experimental system
Network and connectivity filters

- The network may be filtered by experimental system, source, or GO process.
- Applicable only if the edges are colored by experimental system or source and the nodes are colored by GO process.
Connectivity filters allow the user to define a minimum number of connections a node must have in order to be displayed. An iterative option is also available.
Network layouts: concentric circle
Network layouts: spoked dual ring
SMN: spoked dual ring layout
SMN interaction layouts
SMN interaction layouts - GO
Node report

- Table view

<table>
<thead>
<tr>
<th>YORF</th>
<th>Gene</th>
<th>Description</th>
<th>GO Component</th>
<th>GO Function</th>
<th>GO Process</th>
<th>GO Special</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOC10248</td>
<td>RPP20 POP7 RPP2</td>
<td>Homo sapiens, POP7 (processing of precursor, S. cerevisiae) homolog, clone MGC:1986 IMAGE:3138336, mRNA, complete cds.</td>
<td>- nucleolar ribonuclease P complex - nucleus</td>
<td>- ribonuclease P activity - hydrolase activity</td>
<td>- tRNA processing</td>
<td>- RNA processing - metabolism</td>
</tr>
<tr>
<td>LOC79760</td>
<td>GEMIN7 FLJ13956</td>
<td>Homo sapiens, hypothetical protein FLJ13956, clone MGC:14121 IMAGE:4053402, mRNA, complete cds.</td>
<td>- spliceosome complex</td>
<td>- NONE</td>
<td>- nuclear mRNA splicing, via spliceosome</td>
<td>- RNA processing - metabolism</td>
</tr>
</tbody>
</table>
Manual functional clustering

- Transcription
- snRNP assembly
- Actin dynamics
- Pre-mRNA splicing
- Ribosome assembly
- Nuclear localization(?)
- Cell cycle
Back to cartooning: too much information, too little information, or both?

- Cartoon images are widely used to visualize and document protein interactions
- Figures are often accepted as research results
- **Too little information:** interactions are only viewed in an isolated context, picture is limited to information deemed relevant by a particular research group
- **Too much information:** sequence of binding, oligomerization states, binding orientation, etc. are implied but not proven
Osprey advantages

- Protein-protein interactions are not shown with implications toward binding order or binding sites.
- Flexible interface allows rapid manipulation of the network, allowing immediate and creative formation of hypotheses regarding interactions.
Osprey disadvantages

- Manual extraction of data
- No automated deposition tools for interaction information
- Multiple interaction databases, main are Osprey and BIND
- Lack of standards for interaction quality
- Inability to record temporal or PTM dependence of interactions
Obstacles or opportunities?

- As shown with SMN, it is rapidly becoming necessary to be able to view protein interactions on the proteome scale, with a flexible and non-misleading graphic interface.

- In this new field, there are several exciting potential projects for students of the computational biosciences.

- A few ideas: GUI for normalized data entry into standardized protein interaction DBs, creation of decision-tree or NN data mining of PubMed for relevant interaction papers, n-dimensional interaction analysis.