Relating Protein Abundance & mRNA Expression

Mark B Gerstein
Yale (Comp. Bio. & Bioinformatics)

NIDA site visit at Yale
2008.12.03
Why relate amounts of protein & mRNA

Gene expression - major place for regulation (easy to measure)

vs.

Concentration of protein - major determinant of activity

Expectations from simple kinetic models:

\[
\frac{dP_i}{dt} = k_{s,i}[\text{mRNA}_i] - k_{d,i} P_i
\]

At steady state: \( P_i = \frac{k_{s,i}[\text{mRNA}_i]}{k_{d,i}} \)

where \( k_{s,i} \) and \( k_{d,i} \) are the protein synthesis and degradation rate constants

Outliers from trend interesting

[Greenbaum et al. *Bioinformatics* 2002, 18, 587.]
Relationship of Protein Abundance to Complexes and Pathways

In protein complexes, one expects stoichiometric abundance of component proteins and that mRNA expression levels should be correlated with protein abundance.

...Among pathways, this is expected to a lesser degree between interacting proteins.


Protein complexes
[Graphic: http://proton.chem.yale.edu]

Protein interaction networks
### Sources of experimental data

#### mRNA expression levels
- Microarrays
  - Affymetrix
  - PCR
- SAGE

#### Protein abundance
- **2D Gel Electrophoresis**
  - Multiple staining options
  - Small dynamic range
- **DIGE**
  - Cy3 vs. Cy5 labeling
  - Large dynamic range
- **ICAT, iTRAQ**
  - MS-based
  - Relative abundance (ratio of isotopically labeled species)
  - Large dynamic range
- **MudPIT**
- **SILAC**
  - Stable isotope labeling with amino acids in cell culture - for MS analysis
- **TAP-Tag**
  - Weissman and O’Shea (Oct. 2003)

PARE: proteomics.gersteinlab.org

Choose datasets

1. Select organism: Yeast

[required if selecting datasets from menus in (2) and (3) below]

2. mRNA expression

Select complete citations | or upload example file
REL Ideker et al. (2001)

3. Protein abundance

Select complete citations | or upload example file
REL Ideker et al. (2001)

Please see external mRNA expression and protein abundance databases to retrieve additional datasets for analysis.

Analyze

Correlate everything

Perform correlation for selected categories (subsets for selection appear on next page)
- pick a MIPS complex of proteins for the analysis
- pick a GO biological process subset
- pick a GO molecular function subset
- pick a GO cellular component subset
- upload your own subset example file

Submit | Reset

Other tools

Sequence variation (SNP) substitution generator

Upload or use pre-loaded mRNA, protein datasets

Open-source code

Downloadable

Analyze all or analyze MIPS or GO subset

[Yu et al., BMC Bioinfo. '07]
PARE: a web-based tool for correlating mRNA expression and protein abundance

(1) Select mRNA, protein datasets:
   - use pre-loaded datasets
   - upload datasets

(2) Choose categorization method:
   - correlate all
   - MIPS complexes
   - GO biological processes
   - GO molecular function
   - GO cellular component

(3) Display
   - Linear or log-log correlation for selected subset(s)
   - Tabulate data, correlation values for selected subset(s)
   - Label (on plot) and tabulate outlying datapoints

[Yu et al., BMC Bioinfo. '07]
PARE output

Correlated data

Log-log plot of correlation
- linear fit
- outliers labeled

Calculation of mutual information

[Yu et al., BMC Bioinfo. '07]
Correlation of subsets (GO, MIPS)

Yeast ref. datasets:
“Correlate all”

vs.

GO cellular component subsets
for particular cellular locations
# PARE: pre-loaded datasets

**mRNA expression datasets**

<table>
<thead>
<tr>
<th>Data type</th>
<th>Organism</th>
<th>Number of ORFs</th>
<th>Method</th>
<th>Rel./Abs.</th>
<th>Citation</th>
</tr>
</thead>
</table>

[Yu et al., BMC Bioinfo. '07]
Connecting PARE with datasets from NIDA investigators

**Rat** (Nairn lab) - samples from 3 brain regions:
- cortex, striatum, hippocampus

*Protein abundance* - MudPIT datasets
(protein profiling core)
samples for each brain region pooled from 4 rats

*RNA expression* - Affymetrix microarray
(microarray core)
microarrays for brain regions for each of
4 rat individuals + replicate measurements (13 datasets)

**Mouse** (Nairn lab) - 3 brain regions
only iTRAQ data available; no matched mRNA datasets
Rat brain: mRNA expression

mRNA samples
4 individual rats
CTX: indiv. 2, 3, 4 (+2 replicates each for 3, 4)
HIPP: indiv. 1-4
STR: indiv. 1-4
28,894 probes

mRNA expression clustering for rat brain tissue samples
Rat brain: protein abundance

Rat brain
3 tissues: Cortex (CTX), striatum (STR), hippocampus (HIPP)

Number of matched mRNA/proteins pairs between Affymetrix and MudPIT data:
  CTX: 830 proteins
  STR: 926 proteins
  HIPP: 701 proteins

Number of proteins observed in 2 or more tissues (for MudPIT data):
  CTX and STR: 517 proteins
  CTX and HIPP: 538 proteins
  STR and HIPP: 485 proteins
  CTX, STR, and HIPP: 421 proteins
Rat brain: correlation of mRNA expression and protein abundance

Rat brain: \text{emPAI(STR)} v. \text{average mRNA expression (4 rats: 1STR to 4STR)}

926 proteins matched with corresponding mRNA
-Ribosomal proteins observed for STR appear to have low emPAI and a wide range of mRNA expression.

-Proteins that comprise tubulin have a wide range of emPAI (including 2 of the 4 highest values), but a narrow range of mRNA expression.
Rat brain: mRNA expression and protein abundance for functional categories
MudPIT datasets:
Angus Nairn, Erika Andrade
Kathy Stone, Chris Colangelo, Mark Shifman (YPED)

mRNA datasets:
Shrikant Mane, Aiping (Amy) Lin

PARE:
Anne Burba, Eric Yu