Differential Flux Balance Analysis of Quantitative Proteomic Data on Protein Interaction Networks

Biaobin Jiang[1*] David F. Gleich[2] Michael Gribskov[1,2]

[1]Department of Biological Sciences,

[2]Department of Computer Science, Purdue University

[*]Presenter, bjiang@purdue.edu

December 16, 2015

Outline

Background: Protein Interactome

- Proteins are basic functional units in biological systems.
- Proteins carry out biological functions through protein-protein interactions.

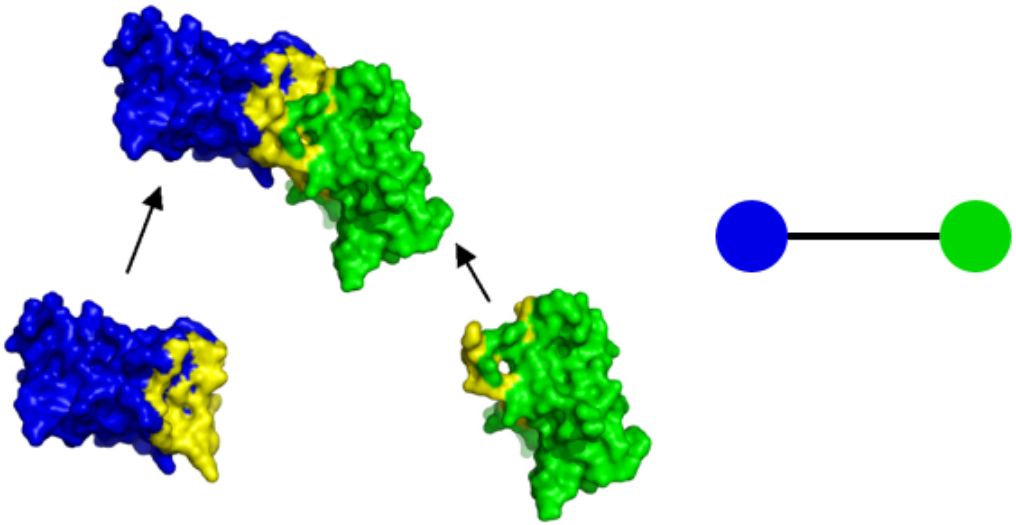


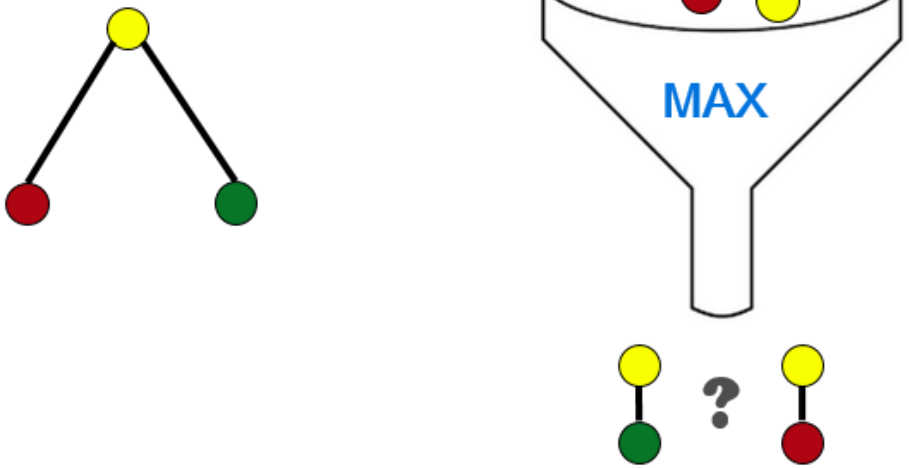
Image resource: lookfordiagnosis.com and Nature Review Genetics **5**, 101-113, Feb 2004.

Problem Statement

- **Ultimate Question:**
  How does a copy of proteins chooses its binding partners?
  - abundance;
  - binding affinity;
  - stoichiometry;
  - post-translational modification;
  - etc.

Problem Statement

- **Feasible Question:** How does a copy of proteins chooses its binding partners given proteome-wide abundance?



Motivation

1. Mutations on protein binding sites rewire cancer signaling;
   (Creixell *et al.*, Cell, 163(1), 2015)

2. Protein interaction dynamics can be experimentally quantified by AP-SWATH in a small scale;
   (Collins *et al.*, Nature Methods, 10(12), 2013)

3. Flux balance analysis is widely used to reconstruct genome-wide biochemical reaction network.
   (Thiele, *et al.*, Nature Biotechnology, 31(5), 2013)

Aim: To use flux balance analysis to estimate protein interaction dynamics caused by cancer-related mutations, which cannot be done experimentally in proteome-wide scale.

Flux Balance Analysis Model

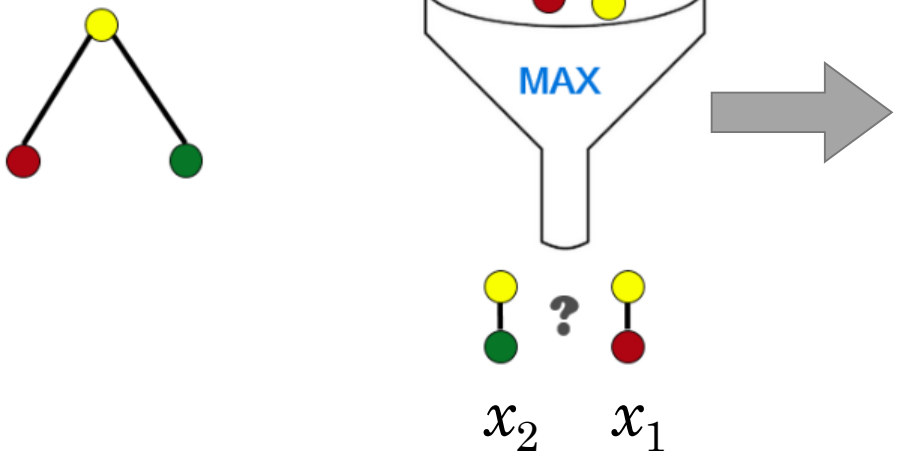
Notations:

- $\mathbf{x} \in \mathbb{R}^n$, protein flux for each interaction;
- $\mathbf{b} \in \mathbb{R}^m$, input protein absolute copies;
- $\mathbf{A} \in \mathbb{R}^{m \times n}$, the incidence matrix of protein-interaction graph, where $A(i,k) = A(j,k) = 1$ if protein i interacts with protein j;
- $\mathbf{c}$ is an all-one vector.

Goal: to maximize the total flux in the whole protei-protein interaction network. The flux balance analysis can be formulated as a linear programming model,

$$\begin{aligned} \underset{\mathbf{x}}{\text{maximize}} \quad & \mathbf{c}^T\mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

Model Illustration



$$\underset{x_1,x_2}{\text{maximize}} \quad x_1 + x_2$$

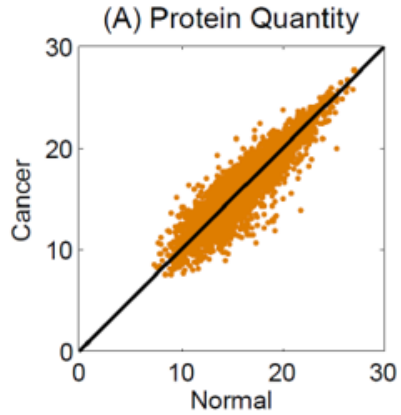
$$\text{subject to} \quad \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leq \begin{bmatrix} 8 \\ 4 \\ 3 \end{bmatrix}$$

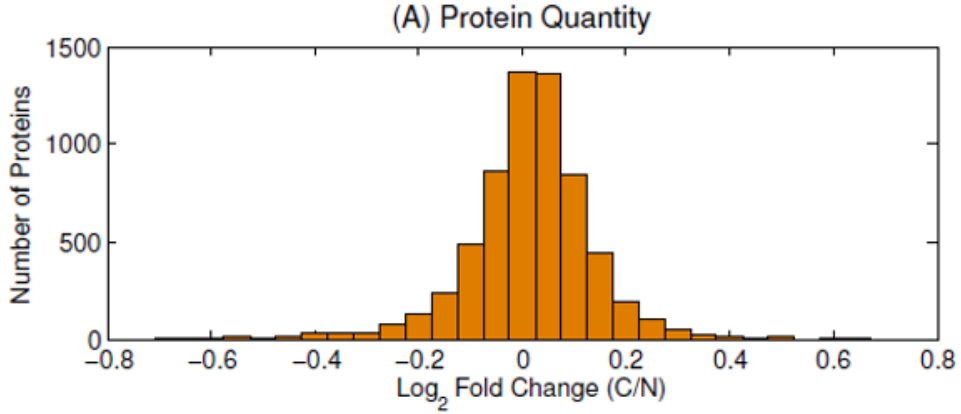
$$x_1, x_2 \geq 0.$$

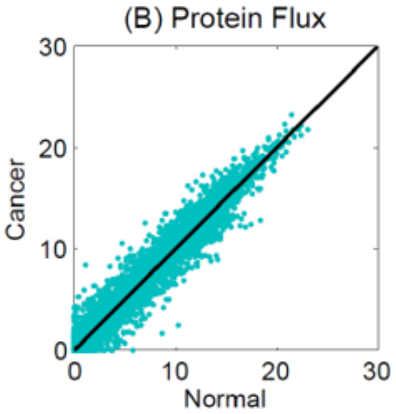
Note: Usually, the number of interactions is much larger than the number of proteins (that means **A** is a *fat* matrix).

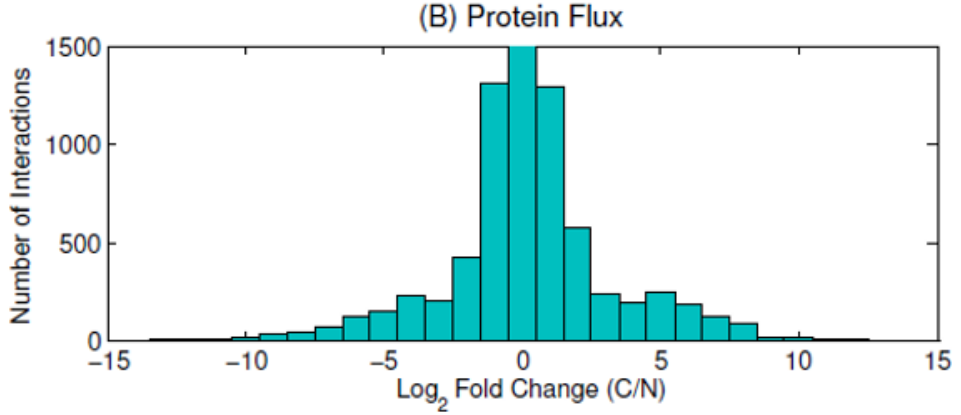
Data from Quantitative Proteomics in Colon Cancer

- **Estimates of Protein Absolute Copy Number**:
  Wiśniewski, Jacek R., *et al*. *Molecular systems biology* 8.611 (2012).
- **Network Data**: BioGRID
  Stark, Chris, *et al*. *Nucleic acids research* 34.suppl 1 (2006): D535-D539



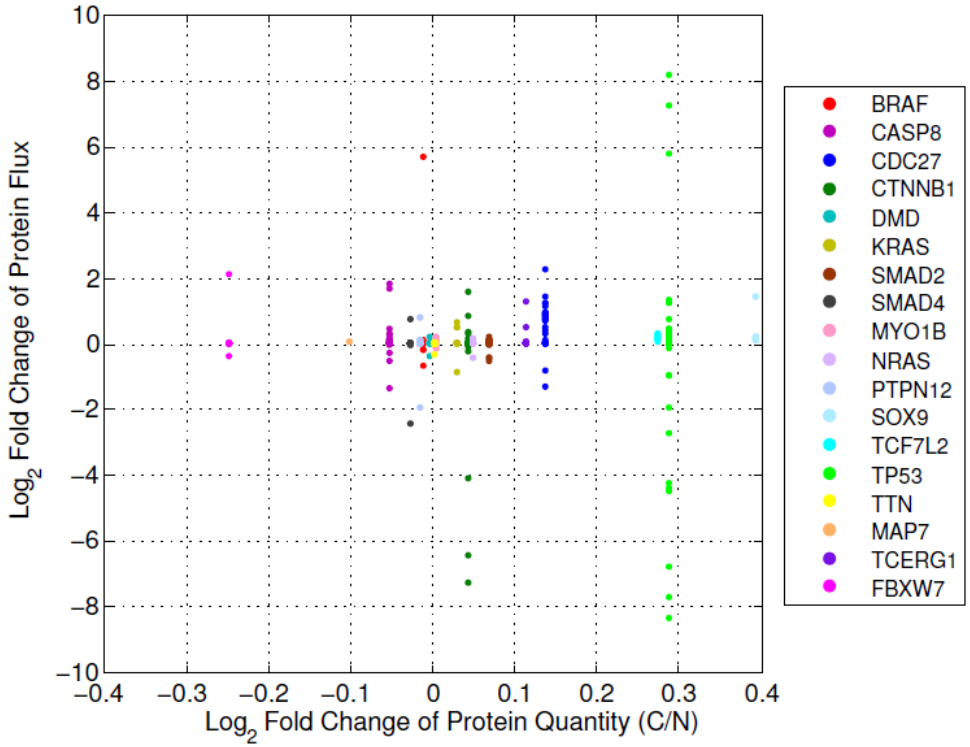






Results (1): Differential analysis, abundance vs. flux

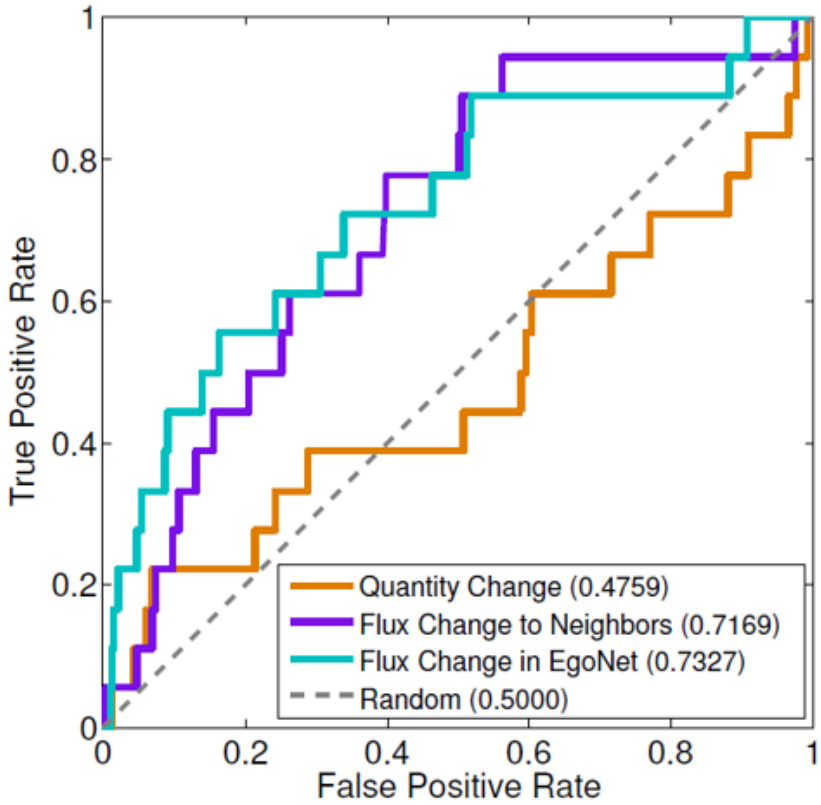
- Does an up-regulated (or down-regulated) protein definitely up-regulate (down-regulate) all of its interaction fluxes?
- Examine 18 genes (proteins) collected from TCGA colon cancer study. (Cancer Genome Atlas Network. Nature, 487(7407), pp.330-337)

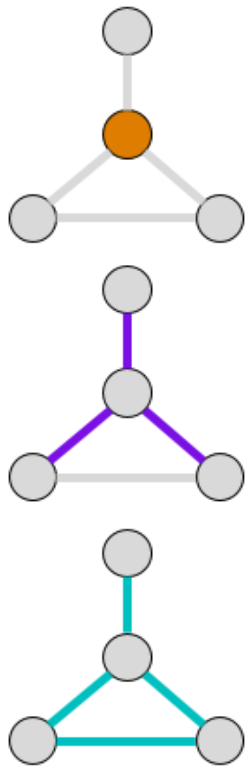


- **Conclusion**: Not necessarily true.

Results (2): Ego-centric network analysis

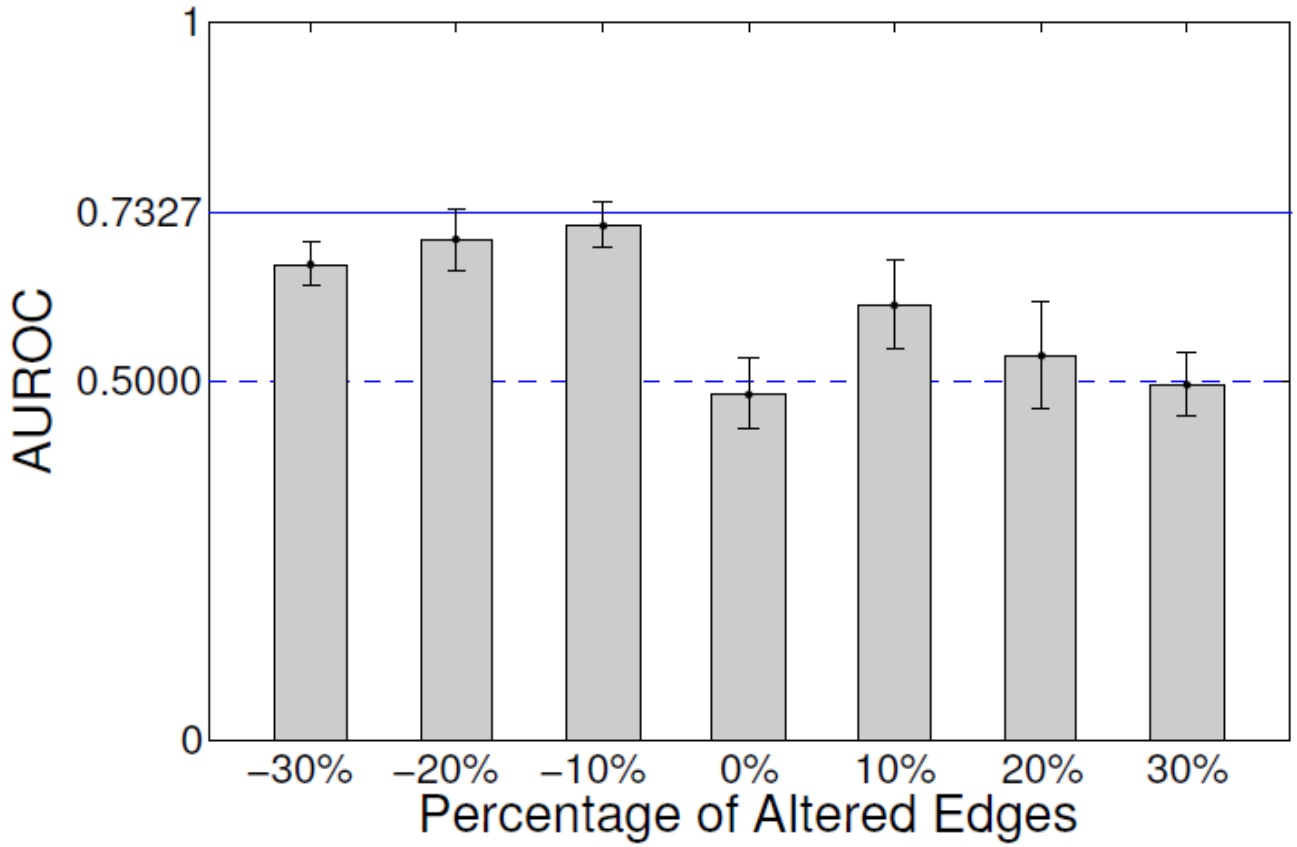
- Test whether cancer-related proteins have higher fluctuation in protein flux than cancer-unrelated ones.
- Examine the rankings of 3 metrics of the 18 cancer-related proteins versus other cancer-unrelated ones.





Results (3): Test of robustness against network noise

- Test whether the noise of network data affects the result (2) by:
  - randomly removing interactions (false positive);
  - randomly shuffling protein labels (random control); and
  - randomly adding interactions (false negative).



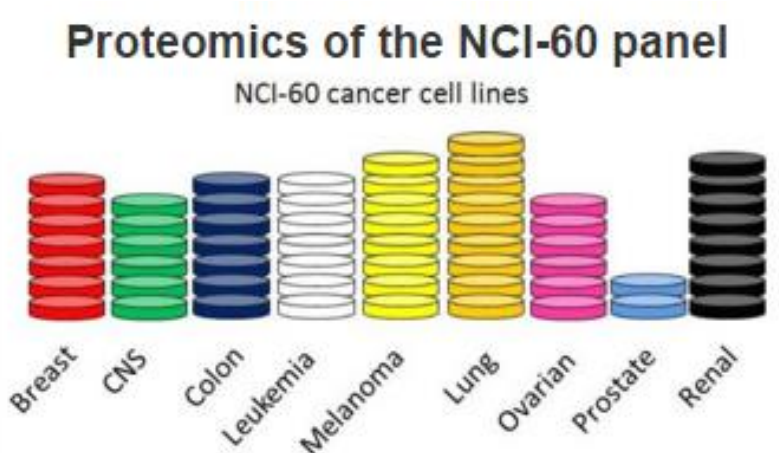
Conclusion

Take-home messages:

- Up/down-regulation of a protein abundance does not necessarily lead to up/down-regulation of its interaction fluxes.
- Cancer-related proteins, affected by genetic mutations, tend to have higher fluctuation in the fluxes of their ego-centric networks.
- Differential flux balance analysis are robust to noisy network data.

Future Plan

- Test the model using a large number of data sets (NCI-60 proteomes);



Statistics

Number of proteins: 10,350
Number of genes: 8,739
Number of peptides: 225,687
Number of ms/ms spectra: 6,894,688
Mean chromosome coverage: 44%

- Extend the model from interaction-wise to pathway-wise dynamics;
- Analyze flux fluctuation of crosstalk between pathways.

Image resource: wzw.tum.de/proteomics/nci60/.

Acknowledgement

I would like to thank,

- my advisor Dr. Gribskov, Purdue Biological Sciences Department;
- my mentor Dr. Gleich, Purdue Computer Science Department; and
- all my labmates.

The End