The Human Massome of Protein Interactions: A New Mass Spectrometry-Based Perspective

Gil Alterovitz[1], Dmitriy Patek[2], Marco F. Ramoni[2, 3], and Isaac S. Kohane[2,3]

Keywords: proteomics, mass spectrometry, protein-protein interactions, networks

1 Introduction

Two of the main approaches used in proteomics are involve studying protein-protein interactions and protein identification/quantification [1]. These have been facilitated by technologies in high-throughput two-hybrid screening and mass spectrometry (e.g. Surface Enhanced Laser Desorption and Ionization (SELDI); Electrospray Ionization). While protein-protein interactions are useful for pathway discovery and network analysis [3, 5], mass spec technology is better suited in certain ways for protein quantification and biomarker discovery. On the other hand, there are several issues with mass spec-based analysis including protein identification, especially for SELDI and Matrix-assisted Laser Desorption and Ionization (MALDI) mass spectrometry.

This research gives a new mass spectrometry-based perspective vis-à-vis a human massome of protein interactions. By integrating multiple existing sources in a non-redundant manner, a network of over 162,000 interactions was created (double the number previously published [4]). One of the benefits of the human massome approach is that the interactions are accessible and searchable by masses of interaction participants (including both cleavage products and mutant proteins). In addition, it can help facilitate protein identification for mass spectrometry from High-Throughput Mass Spectrometric Protein Complex Identification (HMS-PCI) to SELDI.

A public (searchable) version of the database is available at: http://www.chip.org/proteomics/massome.html

2 Methods and Results

An automated program was implemented in Matlab to integrate data from a variety of databases/literature sources [2]. Where possible, proteins identification numbers (which vary by database) were then converted to NCBI Entrez Protein GI numbers via SeqHound, AliasServer, IPI cross-reference indexes, Ensembl cross-reference indexes, and the Entrez Protein database.

The best annotated non-redundant version of each protein was then selected for inclusion in the database. Over 20,000 proteins were analyzed. Protein cleavage sites where extracted from Entrez protein feature information and the different corresponding masses were calculated (with consequent amino acid regions marked in the database). This was stored in a MySQL database. A web interface was developed and is searchable by mass ranges for potential interactors (see Figure 1).

[1] Health Science and Technology/Electrical Engineering & Computer Science, Massachusetts Institute of Technology, Cambridge, 02139, USA, Email: ga@alum.mit.edu
[2] Children's Hospital Informatics Program, Harvard Medical School, Boston, 02115, USA.
[3] Harvard Partners Center for Genetics and Genomics, Boston, 02115, USA.

The human massome of protein interactions was then analyzed for overall statistical properties. Preliminary results suggest that protein interactions are not uniformly distributed across masses. In particular, proteins with similar masses were found to be more likely interact.

3 Figures

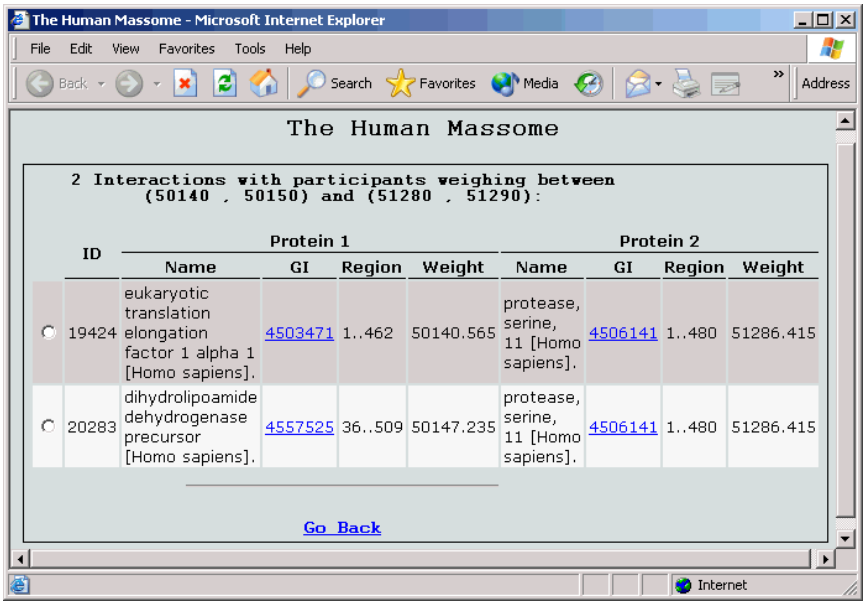

Figure 1: Web interface to the human massome of protein interactions database

4 References

[1] G. Alterovitz, E. Afkhami, and M. Ramoni, "Robotics, Automation, and Statistical Learning for Proteomics," in *Focus on Robotics and Intelligent Systems Research*, vol. 1, F. Columbus, Ed. New York: Nova Science Publishers, Inc., 2005 (In press).

[2] G. Alterovitz, D. Patek, I. S. Kohane, and M. Ramoni, "Human Protein Meta-Interaction Database (HPMD) Potentiates Integration for Meta-Analysis," presented at IEEE Genomic Signal Processing and Statistics (GENSIPS), 2005 (accepted).

[3] H. Jeong, S. P. Mason, A. L. Barabasi, and Z. N. Oltvai, "Lethality and centrality in protein networks," *Nature*, vol. 411, pp. 41-2, 2001.

[4] B. Lehner and A. G. Fraser, "A first-draft human protein-interaction map," *Genome Biol*, vol. 5, pp. R63, 2004.

[5] H. Yu, X. Zhu, D. Greenbaum, J. Karro, and M. Gerstein, "TopNet: a tool for comparing biological sub-networks, correlating protein properties with topological statistics," *Nucleic Acids Res*, vol. 32, pp. 328-37, 2004.