

Biological Statistics and
Computational Biology

Weill Institute for Cell
and Molecular Biology

Candidate for Cellular Systems Biology Position

Haiyuan Yu

Dana-Farber Cancer Institute
Harvard Medical School

Wednesday, March 18, 2009
10:00 – 11:00 am
226 Weill Hall, above Synapsis Cafe

Understanding Large-scale Interactome Networks

Proteins function through interactions with bio-molecules, DNA and other proteins especially. The set of all molecular interactions in an organism is its “interactome”. More specifically, interactome is the sum of all protein-protein interactions. Current yeast interactome network maps contain several hundred molecular complexes, with limited and at times controversial representation of direct binary interactions. We carried out a comparative quality assessment of current yeast interactome datasets, finding that high- throughput yeast two-hybrid (Y2H) provides high-quality binary interaction information. As most of the yeast binary interactome remains unmapped, we exploited an empirically- controlled mapping framework to produce a “second-generation” high-quality high- throughput Y2H dataset covering ~20% of all yeast binary interactions. Both Y2H and affinity purification followed by mass spectrometry (AP/MS) data are of equally high quality but are fundamentally different and complementary, producing networks with dissimilar topological and biological properties. Compared to co-complex interactome models, the binary Y2H map is enriched for transient signaling interactions and inter- complex connections. Protein connectivity correlates with genetic pleiotropy, not with essentiality. Furthermore, topological analysis of interactome networks can help identify key proteins involved in various human diseases, especially because many diseases result from disrupted interactions among proteins. In sum, mapping and analyzing interactome networks will lead to better understanding of human disease, especially cancer.

Short Bio

Dr. Yu received his B.S. from Peking University in China. He then went to Yale University for his PhD, in the research group of Dr. Mark Gerstein. He has extensive experiences in Computer Science and Statistics and is particularly specialized in machine-learning and protein network analysis. He also worked very closely with Dr. Michael Snyder's lab and designed many large-scale functional genomics experiments for them. He is now a post-doctoral fellow in Marc Vidal's lab in the Dana-Farber Cancer Institute at Harvard Medical School. Since joining Marc's lab three years ago, he has dedicated himself to quickly becoming an experimental expert, specializing in development and application of high-throughput proteomics technologies. Combining his computational and experimental experience, he has refuted and clarified many mis-conceptions and confusions in protein-protein interaction community, which was published in a recent issue of Science (Yu et al, 322: 104-10). In total, he has published 11 first-author papers and 38 papers overall.

Hosts: Andy Clark <ac347@cornell.edu>, Scott Emr <sde26@cornell.edu>, and Jim Booth <jb383@cornell.edu>