Bioinformatics

1016-Pos
Viral Disease Networks
Natali Gulbahce[1], Han Yan[2], Marc Vidal[2], Albert-Laszlo Barabasi[1].
[1]Northeastern University and Dana Farber Cancer Institute, Boston, MA, USA, [2]Dana Farber Cancer Institute and Harvard Medical School, Boston, MA, USA.
Viral infections induce multiple perturbations that spread along the links of the biological networks of the host cells. Understanding the impact of these cascading perturbations requires an exhaustive knowledge of the cellular machinery as well as a systems biology approach that reveals how individual components of the cellular system function together. Here we describe an integrative method that provides a new approach to studying virus-human interactions and its correlations with diseases. Our method involves the combined utilization of protein - protein interactions, protein - DNA interactions, metabolomics and gene - disease associations to build a ''viral diseasome''. By solely using high-throughput data, we map well-known viral associated diseases and predict new candidate viral diseases. We use microarray data of virus-infected tissues and patient medical history data to further test the implications of the viral diseasome. We apply this method to Epstein-Barr virus and Human Papillomavirus and shed light into molecular development of viral diseases and disease pathways.

1017-Pos
Mapping the Structural Locations of Disease-Associated SNPS
Michael Montesano.
University of Illinois Chicago, Chicago, IL, USA.
Non-synonymous single-nucleotide polymorphisms (nsSNPs) are the greatest source of genetic diversity within humans. The random mutation approach evolution has taken to instill this diversity leaves its footprint throughout the human genome in the form of SNPs. Unfortunately, many nsSNPs manifest themselves phenotypically as genetic diseases harmful to their host. Nathan Stitziel demonstrated that many of these disease-associated SNPs map to voids or pockets, rarely observed to be within the interior of the protein, and furthermore, there is no tendency for disease SNPs to be located in conserved regions. Both conclusions ring counter to to a significant portion of disease SNP research as well as to intuitive logic, since one would expect a mutation on either a buried or highly conserved residue to have a destabilizing effect. In the time since N. Stitziel produced his conclusions, the two major sources of SNPs he used, dbSNP and OMIM, have seen explosive growth in size and scope. The current build of dbSNP contains 92.3 million SNPs, up from 4.8 million SNPS at the time of Nathan's research. I propose to revisit his work equipped with updated repositories of nsSNPs and see if his his conclusions hold given the new data available.

1018-Pos
System Biology Pathway Exchange - Bridging Pathway Data And Quantitative Models
Oliver Ruebenacker, Michael L. Blinov, **Ion I. Moraru.**
University of Connecticut Health Center, Farmington, CT, USA.
Online databases store thousands of molecular interactions and pathways, and numerous modeling software tools provide users with an interface to create and simulate mathematical models of such interactions. However, modeling tools often have to deal with formats that are structurally and semantically different. Conversion between formats (making data present in one format available in another format) based on simple one-to-one mappings may lead to loss or distortion of data, is difficult to automate, and often impractical and/or erroneous. This seriously limits the integration of knowledge data and models. We introduce an approach for such integration based on a bridging format that we named *Systems Biology Pathway Exchange* (SBPAX) alluding to community standards for exchange of mathematical models (SBML) and storing pathway data (BioPAX). It facilitates conversion between data in different formats by a combination of one-to-one mappings to and from SBPAX and operations within the SBPAX data. The concept of SBPAX is to provide a flexible description expanding around essential pathway data - basically the common subset of all formats describing processes, the substances participating in these processes and their locations. SBPAX can act as a platform for converting between formats and documenting assumptions used during conversion, gluing (identifying related elements across different formats) and merging (creating a coherent set of data from multiple sources) data. This works was supported by NIH/NCRR grants RR022232 and RR13186.

1019-Pos
Docking by Structural Similarity at Protein-Protein Interfaces
Petras Kundrotas, Rohita Sinha, Ilya A. Vakser.
University of Kansas, LAWRENCE, KS, USA.
Rapid accumulation of the experimental data on protein-protein complexes drives the paradigm shift in protein docking from 'traditional,' template free approaches to template based techniques. Homology docking algorithms are based on sequence similarity between target and template complexes and can account for up to 20% of known protein-protein interactions. When the highly homologous templates for the target complex are not available, but the structure of the target monomers is known, docking by local structural alignment may provide an adequate solution, as a complement to the traditional template-free docking, which is notoriously sensitive to structural inaccuracies. Such an algorithm was developed based on structural comparison of monomers to known co-crystallized interfaces. A library of templates was compiled, consisting of 11,932 interfacial fragments, extracted from the asymmetric and biological units in PDB. The structural alignment was performed by TM-align program. The results were optimal with the interfaces defined by 12A distance cutoff. The benchmarking of the procedure was performed on the DOCKGROUND docking benchmark sets: 99 unbound complexes and the extended set of 372 bound complexes. The models were ranked by TMscore. Higher-accuracy models (i-RMSD < 5A) were found in top 10 predictions for 25 % (unbound set) and 33 % (extended bound set) of targets. Importantly, most of the successfully predicted complexes were in addition to those predicted by template-free docking (by GRAMM-X). Compared with the full structure alignment, the partial structure alignment succeeded in a significant number of targets on which the full alignment failed. The method was also tested on previous CAPRI targets. The results indicate that the partial structure alignment provides a much needed addition to the docking arsenal, with the combined structural alignment and template free docking success rate significantly surpassing that of the template free docking alone.

1020-Pos
Selection of Near-Native Protein Structures by Means of Molecular Dynamics Simulations
Bogdan Barz, Qingguo Wang, Jingfen Zhang, Zhiquan He, Dong Xu, Yi Shang, Ioan Kosztin.
University of Missouri - Columbia, Columbia, MO, USA.
In spite of recent advances, the problem of protein structure prediction from the amino acid sequence remains a challenging one. In general, once a large set of model protein structures is predicted one needs to define selection criteria for identifying the structure that is the closest to the native one. The most common way of discriminating between predicted structures of a given protein is to employ either knowledge or physics based energy functions. Here we present an alternative ranking method of the predicted structures of a protein by testing their stability during gradual heating achieved by all atom molecular dynamics (MD) simulations. In general, the smaller the RMSD of the structure of a protein is (with respect to its native one) the more stable this structure is. Thus, one can rank the quality of these structures by comparing the relative stability of the predicted structures against gradual heating. We refer to this approach as the MD-Ranking (MDR) method. We have successfully tested the MDR method on several sets of proteins. We have also tested the MDR method in the 2008 Critical Assessment of Techniques for Protein Structure Prediction (CASP8) competition as part of our MUFOLD-MD server, which worked as follows: i) it generated 10,000 structures using the ab initio method of the Rosetta software, ii) from these, 64 structures with the lowest Rosetta energy were selected, and iii) re-ranked with the MDR method. The top 5 models, with the best MDR score, were submitted to the CASP8 organizers. Based on the official CAP8 results, MUFOLD-MD was ranked as number one server in the Free Modeling category. Work supported by a grant from NIH [R21/R33-GM078601]. Major computer time was provided by the University of Missouri Bioinformatics Consortium.

1021-Pos
The Protein Circular Dichroism Data Bank (PCDDB) - First Release of a New Resource for Spectroscopic Data Sharing
Lee Whitmore[1], Benjamin Woollett[1], Andrew J. Miles[1], Robert William Janes[2], B.A. Wallace[1].
[1]Birkbeck College, University of London, London, United Kingdom, [2]Queen Mary, University of London, London, United Kingdom.
The Protein Circular Dichroism Data Bank (PCDDB) has been designed as a resource for the deposition of circular dichroism (CD) spectroscopic data and accompanying meta-data, with links to sequence and structure data bases and citation references. A key aim of the PCDDB is to provide a repository for spectroscopic data in a comparable manner to that of the long-established Protein Data Bank (PDB), which contains three-dimensional structures of proteins and their associated crystallographic, NMR or cryo-EM data. The PCDDB will be a searchable data bank of CD spectra, with associated tools and protocols for spectral matching, analysis and back-calculations available as part of the overall resource. In addition, validation software will be available for