AM5: ISMB2006 Tutorial

Title: Computing Biological Function: Bioinformatics approach to the analysis and prediction of protein function

Topic Area:
- Proteomics.

Main Presenter:
- Title: Dr
- Full name: Yanay Ofran
- Affiliation: Columbia University Bioinformatics Center (C2B2)
- Mailing Address: 1130 St. Nicholas Ave, 8th. floor, Room 802 New York, N.Y. 10032 USA
- Email address: ofran@maple.bioc.columbia.edu
- Telephone number – work and cell if available, with country and city codes: +1-212-851 4669
- Fax number: +1-212-305 7932
- Home page URL: http://cubic.bioc.columbia.edu/
- Teaching experience:

Developed and taught several undergraduate and graduate level courses. Including:
Intro to Bioinformatics (undergrad.), average student feedback: 4.6 (out of max of 5)
Intro Biomedical Informatics (grad.) average student feedback: 4.2 (out of max of 5)
Classes in "Computational Biology" (grad.)

- Earlier tutorial presentations:

Tutorial at ISMB 2004.
Title: "In-silico Analysis and Prediction of Protein Function"
with Marco Punta and Burkhard Rost
Feedback: voted "best quality tutorial" at ISMB 2004

Tutorial at Pacific Symposium on Biocomputing (PSB) 2005
Title: "Function Prediction: From High Throughput to Individual Proteins"
with Marco Punta and Burkhard Rost

Tutorial at Pacific Symposium on Biocomputing (PSB) 2006
Title: "Protein interaction and disease".

Second presenter:
- Title: Dr
- Full name: Marco Punta
- Affiliation: Columbia University Bioinformatics Center (C2B2)
- Mailing Address: 1130 St. Nicholas Ave, 8th. floor, Room 802 New York, N.Y. 10032 USA
- Email address: mp2215@columbia.edu
- Telephone number – work and cell if available, with country and city codes
Work: +1-212-851 4669   Cell: +1-917-498 5953

- Fax number: +1-212-305 7932
- Home page URL: http://cubic.bioc.columbia.edu/
- Teaching experience:

2005-2006 Lectures at the course "Computational Biology II" at Columbia University
2004-2005 Lectures at the course "Advances in Genome Technology and Bioinformatics" at the Marine Biological Laboratory (MBL) in Woods Hole, Massachusetts

- Earlier tutorial presentations and feedback if any – give tutorial title, conference name, location, year.

Tutorial at ISMB 2004.
Title: "In-silico Analysis and Prediction of Protein Function"
with Yanay Ofran and Burkhard Rost
Feedback: voted "best quality tutorial" at ISMB 2004

Tutorial at Pacific Symposium on Biocomputing (PSB) 2005
Title: "Function Prediction: From High Throughput to Individual Proteins"
with Yanay Ofran and Burkhard Rost

Other contributors to the tutorial presentation:

Prof. Burkhard Rost
Columbia University
Department of Biochemistry & Molecular Biophysics
1130 St. Nicholas Ave, 8th. floor, Room 802
New York, N.Y. 10032 USA
Tel +1-212-851 4669 / Fax +1-212-305 7932
e-mail: rost@columbia.edu

50-word abstract:

Currently, there are millions of sequences with little or no functional annotation. (1600 of them even have known 3D structure). In this tutorial we discuss the approaches and survey the tools available for studying function *in silico*, elaborating on open challenges. We use real-life examples from the literature to illustrate the strengths and weaknesses of current function prediction methods.

Tutorial level: Introductory - Intermediate.

Suitability of this tutorial for ISMB and prior knowledge required:

Protein function prediction is one of the merging fields in computational biology. This tutorial is geared toward biologists who want to familiarize themselves with the growing number of available tools for function prediction, and toward computational scientists who want an introduction to the problem of function prediction and the current attempts to address it.

A specific attention is given to open challenges and un-tackled questions. Thus, researchers who are interested in getting into the field will be exposed to suggestions and ideas for projects that could influence the community.

The tutorial does not require advanced knowledge in any field, but we assume basic understanding of biology (undergraduate level), and familiarity with the basic tools and algorithms of computational biology (e.g. sequence alignment).

Profile of Presenter 1

Yanay Ofran, PhD – Research Scientist at Columbia University Bioinformatics Center (C2B2). Current research focuses mainly on analysis of protein networks and on *ab-inito* prediction of protein function. Authored several papers, reviews and book chapters on function prediction. Most recently, he has coauthored an invited review on function prediction in the context of drug discovery (Ofran, Punta et al.: Beyond annotation transfer by homology: novel protein-function prediction methods to assist drug discovery. *Drug Discov Today.* 2005 Nov 1;10(21):1475-82). Teaching Experience: Developed and taught several undergraduate and graduate level courses; developed and taught tutorials and workshops on various topics (function prediction, protein-protein interaction, bioinformatics and disease) in different meetings (including ISMB2004, PSB05, PSB06).

Profile of Presenter 2

Marco Punta, PhD – Research Scientist at the Columbia University Bioinformatics Center (C2B2). Current research includes target selection and function annotation for Structural Genomics consortia NESG (globular proteins) and NYCOMPS (membrane proteins), as well as prediction of metal binding-sites through the use of machine learning approaches. Previously, he developed PROFcon, a method for predicting protein contact maps (http://www.predictprotein.org/submit_profcon.html) that was one of the best performing methods at CASP6 (Grana et al. *Proteins* 2005). He has recently co-authored an invited review on function prediction in the context of drug discovery (Ofran, Punta et al. *Drug Discov Today*. 2005, 10(21):1475-82) and co-organized a session on "Protein interactions and disease" at PSB 2006.

Teaching experience: lectures at 2004 and 2005 MBL course on "Advances in Genome Technology and Bioinformatics" in Woods Hole, lectures at 2005 and 2006 course "Computational Biology II" at Columbia University, tutorials at ISMB 2004 and PSB 2005. He co-organized the student meeting at CASP6 in 2004. Presentations at several meetings and seminars (most recently, BITS Italian bioinformatics society meeting 2005, P2P post-CASP symposium 2005).

Tutorial Outline:

Two years ago we presented a tutorial on function prediction at ISMB 2004 and it was well received (was voted "best quality tutorial"). During these two years, the field of function prediction has gone through some dramatic developments. Accordingly, this tutorial is based, to a large extent, on new findings, new approaches, new tools and new results.

Part 1 (90 min)

I. Introduction: – What is "function" (15 min)

1. Function means different things to different people. Protein function can be described in biophysical terms (e.g. kinase), biochemical terms (being a part of a certain pathway), biological terms (e.g. "development"), pathological terms (tumor suppressor), etc. Before predicting "function" we should clarify what is it that we attempt to predict.
2. Controlled vocabularies such as Gene Ontology (GO) can convey a rigorous framework for discussing function. We will discuss GO as well as other databases, annotations and theoretical metrics that could assist the computational analysis of protein function.

II. From individual molecules to networks (25 min)

3. Numerous studies and tools focus on the function of individual proteins and genes. However, while each individual macromolecule has a function, no biological process is carried out by a single macromolecule. We will look at protein-protein interaction and will see how interactions can help understand the function of proteins.
4. Interactions between genes and between proteins create networks that can teach us a lot about function. We will discuss approaches to define and study function in the context of biological networks.

III. High throughput methods (50 min)

5. Annotation transfer – using sequence or structure similarity as a means to infer function.
6. Methods and approaches for the prediction of sub-cellular localization.
7. Methods and approaches for the prediction of post-translational modifications.
8. Methods and approaches for the prediction of a functional class / type.

Part 2 (90 min) – Examples from the literature

IV. Function prediction through homology, motifs and identification of functional residues (25 min)

9. We will show how and to which extent sequence similarity can be used to predict sub-cellular localization and how correlated mutations have been used in combination with some prior experimental information to predict protein oligomerization interfaces.

V. Function prediction from structure (25 min)

10. In this section we will discuss function annotation in structural genomics along with a case in which knowledge of the protein structure is key to assigning its function.

VI. Function prediction using interaction maps (20 min)

11. We will present studies that use *in silico* analysis of experimental protein interaction maps to identify novel members of specific functional classes.

VII. Function prediction on a genomic scale (20 min)

12. We will illustrate how automatic function prediction methods can be used to annotate entire genomes.