Association of Biomolecular Resource Facilities
Proteome Informatics Research Group (iPRG)

9650 Rockville Pike, Bethesda, MD 20814
Tel: 301-634-7306 ♦ Fax: 301-634-7455 ♦ Email: abrf@abrf.org

Re: iPRG-2013: Proteome Informatics Research Group Study:

Using RNA-Seq Data to Refine Proteomic Data Analysis

Dear Potential Study Participant,

The cornerstone of proteomic data analysis is identification of proteins and peptides from fragmentation data using a search engine. This type of software requires a reference database to define which peptide sequences should be considered as possibly being present in the sample. This database normally consists of all protein sequences known to ever be produced by the species under analysis. However, in any given cell type or tissue, only a subset of these proteins will ever be expressed. In addition, the particular cells in question may synthesize sequence and/or splice variants that differ from the canonical sequence that a database has selected to represent a given protein.

With the recent increases in speed and decreases in cost of RNA-Seq analysis, it is now practical to produce a catalog of all gene sequences that are expressed in a given sample. If these data are paired to proteomic analysis of the same sample, it is possible to create refined, more specific protein databases for querying by proteomic search engines.

The Proteome Informatics Research Group (iPRG) of the Association of Biomolecular Resource Facilities (ABRF) invites you to participate in a collaborative data analysis study to evaluate the benefits of using databases derived from RNA-Seq data for peptide identification. Participants will be asked to compare peptides that are identified when searching against a standard reference protein database to those that can be assigned using databases that make use of complementary RNA-Seq data. The proteomic dataset provided consists of high mass accuracy tandem mass spectra acquired from a large number of precursors. A variety of different types of sequence databases will be supplied. These include a standard protein sequence database; a database containing only sequences of proteins expressed in the sample based on RNA-Seq data; a database that includes sequence and splice variants; a database of sequences that cannot be reconciled to known expressed gene sequences. The raw RNA-Seq data are also provided.

Participants will be supplied with proteomic data in a range of formats and a template for reporting the results. Participants will be asked to report spectral identifications, highlighting those identifications that were only identified using one of the RNA-Seq derived specialized sequence databases. Upon completion, participants will be asked to complete a web-based questionnaire summarizing the methods they used. Results submission will be anonymous, so you are encouraged to participate, regardless of your experience level in database searching.

The datasets and databases are now available for downloading and results must be returned by February 8 2013 in order to enable sufficient time to prepare the presentation for the 2013 ABRF Meeting (March 2 – 5, 2013 in Palm Springs, CA). Submissions received after the 8th, but before the ABRF conference starting on 2 March, will be included in further analysis and presentations. If you would like to take part or get more information, please download the study participation instructions from the ABRF iPRG web site:
http://www.abrf.org/index.cfm/group.show/ProteomicsInformaticsResearchGroup.53.htm

This study is open to both ABRF members and non-members. However, we do strongly encourage non-members to join, and thus help support the ABRF. (For more information, visit http://www.abrf.org). A summary of the results of this study will be presented orally and as a poster at the ABRF 2013 conference. The results will be subsequently posted on the ABRF website, and published in a peer reviewed journal.

We thank you for your support of the ABRF and look forward to your participation in this year's study.

Sincerely,

The ABRF Proteome Informatics Research Group (iPRG)

Robert Chalkley – UCSF (Chair)
Nuno Bandeira – UCSD
Matt Chambers – Vanderbilt University
John Cottrell – Matrix Science Ltd
Eric Deutsch – Institute for Systems Biology
Eugene A. Kapp – WEHI
Henry Lam – Hong Kong University of Science and Technology
Ruixang Sun – Chinese Academy of Sciences
Olga Vitek – Purdue University
Sue Weintraub – Univ. of Texas Health Science Center at San Antonio
Thomas Neubert (EB Liaison) – New York University