Association of Biomolecular Resource Facilities
Business Office:
2019 Galisteo Street, Bldg. I, Santa Fe, NM 87505
Tel: 505-983-8102 ♦ Fax: 505-989-1073 ♦ Email: abrf@abrf.org

Dec 7, 2007

Dear iPRG Study2008 Participant,

Thank you for participating in the first Proteome Informatics Research Group study. This letter describes the instructions to get the data, complete your analysis, and submit your results. Results returned by **January 4, 2008** will included in the iPRG presentation at the at the ABRF2008 meeting Feb 9-12, 2008 in Salt Lake City, UT.

Overview of the Analysis Task

We are providing a common data set (in several equivalent formats) and ask you to analyze the data, determine the set of proteins detectable in the sample, and report your findings as you would to a journal. We require that you provide your peptide identifications, although the focus of the study is on protein level identification. The peptide identifications will not be graded and are only collected to serve as evidence in the event that one or more participants assert protein IDs that are not considered part of the iPRG 'correct set' (recognizing that the iPRG members may also be wrong).

The data set for this study is an iTRAQ reagent experiment where the four mouse samples were digested with trypsin, reduced and alkylated with methyl methanethiosulfonate (MMTS), and labeled with the four different iTRAQ reagents. The samples were combined, resolved into 13 fractions by strong cation exchange chromatography, and then analyzed by LC-MS/MS on a 3200 QTRAP system. Most of the fractions were analyzed with successive rounds of exclusion in order to identify more peptides in each fraction. This resulted in a total of 29 acquisition sets/files. Although the sample happens to contain quantitative information, no quantitative results should be reported, as quantitation is not the focus of the study.

Study Materials

Data formats – The dataset is available in a number of formats to enable the use of many different software tools. You may use the format of your choosing. We have made a reasonable effort to generate the various formats with equivalent information content. You are welcome to try multiple input types, although we ask that you submit only one result set. The subject data set is available in (1) the instrument raw format – .wiff and .wiff.scan pairs of files, (2) .mgf, (3) .dta, (4) mzXML, and (5) mzData.

Materials checklist – From one of the download sources in the following section, you must download:

(1) **Study FASTA database** – You may choose to use either the provided forward+reversed decoy or forward-only variant of the study database. You are welcome to create your own derivative of this database to support other false discovery rate assessment techniques.
(2) **A data set** – The provided data set in the format of your choosing.
(3) **The Excel results template** – This is required to submit your results.

Download sources – The materials you will need to complete the analysis are available from two different sources. The files are available on the iPRG website near the bottom of the page under "2) iPRG 2008 Study". The password to all zip files is *iprgcode*.

http://www.abrf.org/iprg

All data files are also available on the Tranche system. The Excel template for results must be downloaded from the iPRG site. You can use the following Tranche links for all other files:

All Data (Peak Lists, Raw Data, and FASTA files):
http://www.proteomecommons.org/data-downloader.jsp?fileName=f6bbab28940e1387cd514bbfc97b4e5f2b76c8521394d04bfa10b48002cb91eb4e26ae428eb88db3fd94777ba28bfca57386a886068c69160d0df73d532d3d8b3c96fcbd000000000000563c

FASTA database files:
http://www.proteomecommons.org/data-downloader.jsp?fileName=371cfe453b2035f8878b90ea14d0e659660ca6622aba7fa339592e3182e0988cea9ae8055743967e545338d254cc89f0fbd54e719aef044e1660bba44543dd50e5a950d900000000000002c3

Raw data files (.wiff format):
http://www.proteomecommons.org/data-downloader.jsp?fileName=6b2d17af6ea5c4758732e596c969d6bf39ca9a99d4446bb3fbd0f2f1c55ff15d47df183acc2ee5e01c45d30ee43c86dfbc1da760c75e02db8fbf821c7ba60ca5069f8b320000000000000256b

MGF format peak list files:
http://www.proteomecommons.org/data-downloader.jsp?fileName=ca9cdf9bacfd95b733b367a574ae38432bbd69df0de119a13a85425589b22b31114cf9f272dfc3c22a624648d5845ad140e1be3e0cd150caa3b29775d4ba4d417bc87cc6000000000000108e

DTA format peak list files:
http://www.proteomecommons.org/data-downloader.jsp?fileName=516811853f07f837c4f17475075efeee28f499c6c9c6756865286d207b1d4d1bc7d49e4c1f1b37965d046646bf995cc6d053a18ec8a412c07f1548d9121b3c6dc2244afe0000000000000f9c

mzXML format peak list files:
http://www.proteomecommons.org/data-downloader.jsp?fileName=290f4afead27c4dd99fee529adb3e20c161da76992bb005eff386f49a6ecf2839e62a6d77408273cd5dfbf34353a7120e4edf695347046862f9d75a85a720bc3ca70845e0000000000001106

Component file names – Regardless of which data format you choose to work with, the component files of a set are named in a common way. By example,

03 CEX1 5uL DBS EMC 2ex.wiff

moving left to right through the file name:

- "03" is the LC acquisition run index
- "CEX1" indicates this is cation exchange fraction 1
- "5uL" indicates the relative injection amount
- "DBS" and "EMC" are acquisition settings, the former 'dynamic background subtraction and the latter 'enhanced multiple charge' scan type
- "2ex" indicates the second round of exclusion on this fraction. Nothing in this position indicates the first analysis of the fraction, and "1ex" would indicate the first round of exclusion

- ".wiff" is the raw data extension. The extension obviously varies with the data format you choose to work with.

Analysis Guidance

Searching guidance – As the focus of the study is protein identification, we provide the following guidance toward reasonable peptide identification:

FASTA database	You must use one of the required databases or a derivative of it with no unrelated proteins added (reversed or random decoys ok).
Modifications	The described workup produces iTRAQ on peptide N-termini and Lys in most cases, and occasionally on Tyr. The Cys will be almost completely labeled by MMTS. For reference, see: http://unimod.org/modifications_view.php?editid1=214 http://unimod.org/modifications_view.php?editid1=39
Instrument effects	The 3200 QTRAP system is like most ion traps in terms of the tolerances one would use for the precursor and fragments. However, the fragmentation tends to differ from an LTQ or LCQ such that one would normally set Mascot's *Instrument* setting to *ESI-Quad*, rather than *ESI-Trap*, for example.
Species filter	Do not set any species (taxonomic) filter. The database contains a small number of non-mouse contaminant proteins.

Use of the template – The provided template is required for submission of your results. You must include information for all the columns that are indicated as required in the INSTRUCTIONS tab and not delete any unused columns. You may also include additional columns to the right of these with additional information, although we ask you to avoid heading names or content that would obviously indicate which software tools were used. Each *Protein N* should correspond to one protein you would report to a journal. If you choose to indicate ambiguity among several accession numbers, this should be indicated as multiple rows with the same *Protein N* but different accession numbers in the *Accession* column. For example:

	A	B
1	N	Accession
2	1	Q3UBU0
3	1	Q91V38
4	2	Q9DC41
5	2	Q3U9G2

See the INSTRUCTIONS tab for additional clarification on the content for the protein and peptides levels.

Results Submission

There are two required steps to complete your contribution to the study that must be done on or before Friday January 4, 2008. First, please go to:

http://www.surveymonkey.com/s.aspx?sm=kANyjM9zAzfdEGqew0s_2bSg_3d_3d

and complete the study survey. This is required, but we estimate it should only take 10 minutes to complete. The survey tool will require you to create an identifier number. Be sure to keep this number.

Second, submit your results in the required Excel template (preferably zipped) to iPRG2008@gmail.com, naming the Excel file "**iPRG2008submission#####.xls**", using your identifier instead of #####. Please be advised that the survey will be available December 15, 2007.

The iPRG will enlist the services of an 'anonymizer', i.e., an individual not involved in the study or the iPRG to collect the submitted results from your emails in order to ensure the anonymity of participants prior to tabulating the aggregate results of the study. The anonymization process will include a check of the properties dialog in the Excel file, although you should clear this yourself before submission to be sure.

Any questions may be directed to the iPRG chair at seymousl@appliedbiosystems.com.

We look forward to your participation and thank you for your support of the ABRF.

Sincerely,

The ABRF Proteome Informatics Research Group

Sean L. Seymour (Chair) – Applied Biosystems|MDS Sciex
Jayson Falkner – University of Michigan
William S. Lane – Harvard University
Alexey I. Nesvizhskii – University of Michigan
Brian C. Searle – Proteome Software, Inc.
David L. Tabb – Vanderbilt University Medical Center
Jeffrey A. Kowalak (Executive Board Liaison) – National Institutes of Health

Vendors and Commercial Service Labs

PLEASE NOTE:
If you are a vendor or commercial service labs and wish to participate in this study, please read the following:

ABRF Research Group Studies are conducted for the benefit of our members and the field at large to help them evaluate their own technical level in comparison to their colleagues, to provide education in techniques, and strategies to which they normally might not be exposed, and to give an overview of the current capabilities of the 'average' lab in carrying out a challenging analysis.

The ABRF welcomes the participation of vendors and for-profit labs, provided that they abide by the ABRF guidelines for the use and distribution of data derived from these studies, as follows:

- An ABRF Research Study is not a competition and under no circumstances should it be referred to as such. Words and phrasing that imply a competition such as 'winner', 'best of', etc. are strictly forbidden.

- Representations and publications should not be deceptive and should fairly emphasize any differences between any data comparisons. For example, instrument reliability cannot be fairly concluded by comparing 5-year old instruments in the field used in the study with a vendor's new instrument.

- Any comparisons to or use of ABRF data should prominently indicate: the number of samples the vendor received, the number of runs performed by the vendor, and whether the actual characteristics of the sample were know by the vendor at the time the vendor's analysis was performed.
- Uses of or comparison to ABRF data should specifically emphasize that many factors will affect analytical results and that the data obtained in the company's R&D lab may exceed feasible expectations for an 'average' resource or research facility under routine conditions.

- Publications and presentations should contain a disclaimer stating that ABRF prepared and provided the sample to all members and vendors, but did not participate in the vendor's study and does not endorse any specific manufacturer, instrument, or strategy.

- Vendors are strongly encouraged to distribute potential publications to the ABRF Executive Board and Research Group Chairperson for comments regarding compliance with these guidelines.

Recipient: **We recommend that this document be distributed to the appropriate marketing and senior personnel in the company to ensure compliance. A copy of this document (Vendor ABRF Study Participation Guidelines.pdf) can be found at www.abrf.org under the Forms and Documents menu.**