Research • Technology
Communication • Education

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-2012: Proteome Informatics Research Group Study: *Detecting modified peptides in a complex mixture*

Dear iPRG 2012 Study Participant,

Thank you for participating in this year's Proteome Informatics Research Group study. This letter provides the instructions needed to get the data, complete your analysis, and submit your results. ***Results returned by Friday, February 3rd, 2012 will be included in the iPRG presentation at the 2012 ABRF meeting (March 17-20, 2012 in Orlando, FL).***

Overview of the Analysis Task

This collaborative LC-MS/MS data analysis study focuses on the evaluation of proteomics laboratories in identifying modified peptides present in a complex mixture.

This study requires you to perform the following bioinformatic analysis:

- Identify the CID spectra present in the sample with < 1% false discovery rate (FDR) for matches to the target database.
- For modified peptides, report which modification site assignments can be reliably localized.
- Complete a brief survey and attach a 1-2 page description of your methodology.

We are providing a common data set (in several equivalent file formats) and ask you to analyze the data and identify the CID spectra present in the sample. The data was acquired on a 5600 TripleTOF (AB SCIEX) mass spectrometer, so has high mass accuracy and resolution on both precursor and fragment ions. The fragmentation is quadrupole CID type fragmentation (similar to HCD fragmentation). We ask that you provide long lists that include identifications that are both above and below your 1% FDR threshold at the spectrum level and indicate those matches that were to the decoy database. This will help us provide feedback for participants on the number of false negative answers reported; i.e., correct answers were found, but were not of high enough confidence to be above the acceptance threshold. There are a wide variety of modifications present in this sample, both biological and chemical in nature. Naturally occurring modifications include, but are not limited to, acetylation, methylation, dimethylation, trimethylation, phosphorylation and sulfation. In the subsequent iPRG analysis of submitted results, special emphasis will be placed on the characterization of modifications not introduced by sample handling.

Description of Sample Preparation and LC-MS/MS Data Acquisition

The sample consists of a yeast lysate that has some additional non-yeast proteins spiked in. Undigested Saccharomyces cerevisiae lysate, Reference Material (RM) 8323, was obtained from the National Institute of Standards and Technology (NIST) and is described at https://www-s.nist.gov/srmors/view_report.cfm?srm=8323.
The S. cerevisiae, strain BY4741, was grown at Boston Biochem Inc. (Cambridge, MA) in rich (yeast peptone dextrose) medium and harvested by continuous-flow centrifugation. The cell pellet was then washed twice with ice-cold water, and lysed by incubation with ice-cold trichloroacetic acid (10 mL/L) in

water for 1 h at 4 °C. The precipitate was collected by centrifugation, washed twice with 100 mL/L water in acetone, and pelleted again. The lyophilized yeast lysate was homogenized at NIST through manual grinding. The ground yeast lysate powder was suspended in 50 mmol/L ammonium bicarbonate containing 6 mol/L urea in water, pH 7.85. After gently stirring at 5 ºC overnight, the yeast lysate solution was filtered through a 0.22 µm cellulose acetate filter. To remove urea from the yeast lysate solution, the solution was thoroughly dialyzed (6,000 Da to 8,000 Da cutoff) at 5 °C using 50 mmol/L ammonium bicarbonate in water as the dialysis buffer. 40 µg of lysate was vacuum centrifuged to almost dryness, then resuspended in 6M guanidine HCl / 25mM ammonium bicarbonate (AmBic), reduced with 2mM TCEP, then alkylated with 5mM iodoacetamide. After diluting to 1.5M guanidine HCl / 25mM AmBic, 1ug Trypsin (Promega) was added for overnight digestion. Digestion was stopped by addition of 5% formic acid, then desalted using a C18 sep-pak (Waters) cartridge. Tryptic peptides from additional non-yeast proteins were spiked into this tryptic yeast lysate.
LC-MS/MS: The iPRG sample was dissolved in 0.1% Formic Acid and 5ul was analyzed by an AB Sciex 5600TripleTOF mass spectrometer interfaced with a Waters nanoAcquity UPLC system. For peptide separation a Waters Symmetry C18 180 µm x 20 mm trap column and a 1.7 µm, 75 µm x 150 mm nanoAcquity UPLC column (45ºC) was used. Trapping was performed at 5µl / min, 99% Buffer A (100% water, 0.1% formic acid) for 1 min. Peptide separation was performed at 500 nl / min with Buffer A: 100% water, 0.1% formic acid and Buffer B: 100% CH3CN, 0.075% formic acid. A linear gradient was run with 5% buffer B at initial conditions, 30% B at 70 minutes, and 85% B at 70.33 minutes. Data acquisition was performed with a TripleTOF 5600 System (AB SCIEX, Concord, ON) fitted with a Nanospray III source (AB SCIEX, Concord, ON) and a pulled quartz tip as the emitter (New Objectives, Woburn, MA). Data was acquired using an ion spray voltage of 2.2 kV, curtain gas of 20 PSI. For data-dependent acquisition, survey scans were acquired in 250 ms and as many as 20 product ion scans were collected if exceeding a threshold of 150 counts per second and with a 2+ to 5+ charge-state. Total cycle time was fixed to 1.3 sec. Four time bins were summed for each scan at a pulser frequency value of 15.420 kHz through monitoring of the 40 GHz multichannel TDC detector with four-anode/channel detection. A sweeping collision energy setting of 35 ± 15 eV was applied to all precursor ions for collision-induced dissociation.

Study Materials

Data formats: The dataset is available in a number of formats, enabling the use of many different software tools. You may use the format(s) of your choosing. You are welcome to try multiple input types, although we ask that you submit only one result set. We also supply the databases to be used for analysis.

Materials checklist: The materials you will need to complete the analysis can be downloaded from:
http://ftp3.worlddesign.com/2012
user: iprg
passwd: edWE4252er

1. Study FASTA databases – The database provided consists of the yeast, human and bovine portions of the most recent SwissProt database, with a few additional potential contaminating proteins appended. The database is provided with and without scrambled sequence decoy proteins appended, and the scrambled database is also available separately. Target and decoy sequences are distinguishable by header, as shown below:
Target:
> sp|A2P2R3|YM084_YEAST Putative glucosamine--fructose-6-phosphate aminotransferase [isomerizing] OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=YMR084W PE=3 SV=1
Decoy:
>DECOY_sp|A2P2R3|YM084_YEAST Decoy sequence

2. Data – The provided dataset in the file format of your choosing.
http://ftp3.worlddesign.com/2012/iPRG_2012_wiff.zip
Raw data file created by instrument acquisition software.
http://ftp3.worlddesign.com/2012/iPRG2012_mzML.zip

mzML 1.1.0 file created by AB Sciex MS Data Convertor 1.2 with the -proteinpilot switch. Fragments are de-isotoped m/z values. If precursor charge could not be determined, none has been specified.

http://ftp3.worlddesign.com/2012/iPRG2012.mgf

MGF file created from iPRG2012_mzML.mzML using ProteoWizard 2.2.

http://ftp3.worlddesign.com/2012/iPRG2012_nd.mgf

MGF peak list created by Mascot Distiller 2.4 with de-isotoping disabled
If precursor charge could not be determined, none has been specified.

http://ftp3.worlddesign.com/2012/iPRG2012_DTA.zip

Unpacks to a collection of DTA files. The content is identical to iPRG2012.mgf except that, where precursor charge could not be determined, DTA files were output for both 2+ and 3+ charge states.

3. The Excel results template – This is required to submit your results.
http://ftp3.worlddesign.com/2012/iPRG2012_reporting_template.xlsx

Results Submission

There are two required steps to complete your contribution to the study that must be done on or before Friday, February 3rd, 2012.
Submit your results in the required Excel template (preferably zipped) to anonymous.iPRG2012@gmail.com, naming the Excel file "iPRG2012submission#####.xls", using your identifier to replace the ##### section.
Please go to: http://www.surveymonkey.com/s/X6L9FHM
and complete the study survey. This is required, but we estimate it should only take 15 minutes to complete. The survey tool will require you to create an identifier number. Be sure to keep this number.
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.

Where to Send Questions

You can send questions directly to the iPRG at anonymous.iPRG2012@gmail.com. Of course, all identifying information will be removed prior to forwarding the question to the iPRG group members.

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
Karl Clauser - Broad Institute of MIT and Harvard
John Cottrell – Matrix Science Ltd
Eric Deutsch - Institute for Systems Biology
Eugene A. Kapp - WEHI
Henry Lam - Hong Kong University of Science and Technology
W. Hayes McDonald - Vanderbilt University
Ruixang Sun – Chinese Academy of Sciences
Thomas Neubert (EB Liaison) - New York University