SBEAMS – Proteomics Tutorial
Eric Deutsch

Please follow along in this tutorial as we go through it in the class, or go your own pace if you choose. Feel free to add your own notes. If you make notes or have suggestions or bug reports that might be useful for others, please email them to edeutsch@systemsbiology.org.

1. Open a web browser and login (recommend Firefox to use FireGoose plugin):
    - http://www.sbeams.org/sbeams/
    - Click "Login to SBEAMS"
    - Login with the account information provided (classNN)
2. Go to Proteomics Module Home Page (link on left nav bar) and test controls
    - Current Project
    - My Projects
    - Accessible Projects
    - Click on project names under "Accessible Projects" to switch current projects
    - Explore several different projects
    - Click on [View/Edit Full Project Information] hyperlink to view project attributes
    - Click BACK and then [View/Edit Experiment Description] under Experiments.
3. Explore the mmarelli – pxproteome project/experiment:
    - Set the Current Project to mmarelli – pxproteome
    - Click on [Proteomics Home] or [My Home] to get back to the top page
    - View project and experiment attributes. READ the Experiment Description for "pxproteome" experiment! Some questions below assume you understand some of the ideas behind the experiment!
    - Click on "Number of MS Runs: 14" (scroll down when page appears!!)
    - Click on some TIC Plots in the table
    - Click on some other hyperlinks in the table to look around
4. Browse this dataset using hyperlinks on left navigation bar
    - Choose Browse Search Hits to get an Interact / pepXML Viewer style view of individual spectrum identifications
        - Select mmarelli – pxproteome – YeastORF experiment
        - P > 0.9
        - QUERY
        - Click on all hyper links across the table
        - Re-sort by Xcorr descending. Re-sort by precursor m/z, % ACN, etc.
        - Page through resultset
        - View in Excel, TSV
        - Make a continuous-value histogram of Precursor m/z, mass diff, % ACN
        - Make a scatter plot of mass diff vs EstRT
        - Click on the PA link for a protein and a peptide
    - Choose Summarize over Proteins/Peptides to get a higher level summary of proteins and/or peptides on pre-ProteinProphet data
        - Select mmarelli – pxproteome – YeastORF experiment
        - P > 0.9
        - QUERY

    - In Display options, CTRL click to multi-select: [Group By Reference], [show GO Columns] & [Show SQL] and QUERY
    - Click on all hyperlinks across the table
    - At top, select [Show All Query Constraints], then near bottom set Gene Annototation Level to 2 and then [QUERY]
    - At bottom choose [Discrete-value Histogram] and [Molecular Function] and do [VIEWRESULTSET] below the list boxes. Then scroll to bottom. Do another discrete-value histogram for [Cellular Component]
    - Click on Display option [Group By Peptide] and QUERY
    - Click on the 29 next to FOX2 LCTPTMPSNGTLK and examine
- Choose Compare Experiments to compare results from different experiments
    - Select jranish – gricat experiment
    - Also CTRL Select mmarelli – pxproteome experiment
    - QUERY
    - Examine differences between experiments, and see summary table
    - Find overlaps by selecting [Show All Query Constraints] and entering ">0" for both "# in Experiment 1" and "# in Experiment 2". QUERY
    - Select [Group by Peptides] and QUERY and examine
- Choose Compare Experiments to compare two search batches of the same experiment
    - Select edeutsch – YeastSILACProfile – SEQ_ScSGD2006_TargDecoy
    - Also CTRL Select edeutsch – YeastSILACProfile – XTK_ScSGD2006_TargDecoy
    - Set Probability >= 0.9
    - Display Option: Group by peptide
    - QUERY
    - Examine differences between the two searches, and see summary table
    - Click on PA link next to YDR502C to see what's in PeptideAtlas
        - What is the protein coverage? (Answer A1 at bottom)
        - Find 2 peptides deemed highly proteotypic both empirically and predicted (A2)
        - How many other proteins and genome locations do the constituent peptides map to? (A3)
        - Click on Cytoscape link to visualize
        - Click on PAp00018124 (SLVAAGLCK) for the Peptide View
        - Click on [Compare Proteins] in Genome Mappings section
        - How do the proteins YLR180W and gi|172534 differ? (A4)
- Choose Compare By Spec to compare two search batches by spectrum to examine the differences between search results using different engines or parameters
    - Select edeutsch – YeastSILACProfile – SEQ_ScSGD2006_TargDecoy
    - Also CTRL Select edeutsch – YeastSILACProfile – XTK_ScSGD2006_TargDecoy
    - Probability in Experiment 1: <.5
    - Probability in Experiment 2: >.95
    - QUERY
    - Examine resulting list

    - Click on OR20080317_S_SILAC-LH_1-1_01.00626.00626.3 and examine spectra (click on <Ions> column entries)
      - Is the one with high probability likely to be correct? What's good or bad about the spectrum IDs? (A5)
      - Has either peptide been seen other experiments in PA? (A6)
  - Choose Compare By Spec to compare two search batches by spectrum to examine the quality of PeptideProphet validation
    - Select edeutsch – YeastSILACProfile – SEQ_ScSGD2006_TargDecoy
    - Also CTRL Select edeutsch – YeastSILACProfile – XTK_ScSGD2006_TargDecoy
    - Probability in Experiment 1: >0.9
    - Probability in Experiment 2: >0.9
    - Select Display Option: "Show different-sequence matches"
    - QUERY
    - Examine resulting list. What is the most common reason for these discepancies? (A7)
    - What about OR20080320_S_SILAC-LH_1-1_11.03716.03716.2?
  - Choose Browse Biosequences to view the contents of the reference FASTA files
    - Select Yeast ORFs Database 2004-04-22
    - Type %perox% in Molecular Function Constraint field
    - QUERY
  - Choose Protein Summary to explore Protein Prophet output for one or more experiments
    - Select mmarelli – pxproteome experiment
    - Protein Group Probability >= 0.9 and Protein Probability >= 0.9
    - QUERY
    - (if you will be doing the Gaggle tutorial later, click URL to "recall this resultset" and then bookmark that page.)
    - Enter %perox% in cellular component (don't forget [Show All Query Constraints] and QUERY
    - Remove %perox% and reQUERY
    - Re-sort by descending Protein Probabilty
    - Download ResultSet in Format [Cytoscape]
    - Within Cytoscape:
      - Click {i} balloon
      - Expand GO, Molecular Function
      - Click on 3
      - Click "Apply Annotation to all Nodes"
      - Click on Go Molecular Function (level 3) on right
      - Click Layout
      - Expand Go Molecular Function (level 3) on right
      - Click on individual GO categories and see them highlighted
      - Click on "hydrolase" and then right click on graph and see attributes
      - Exit Cytoscape. There will be more Cytoscape this afternoon!

5. Try browsing directly in the public yeast PeptideAtlas:
    - Go to http://www.peptideatlas.org/ and click [Search Database] on left
    - Select Build Type: Yeast
    - Type in search box: %peroxisom%   (leave off the e) and click GO
    - How many proteins match? And how many with at least 2 hits? (A8)

6. Now use the interface to answer these questions:
    - How many (GO-annotated) yeast peroxisomal proteins were identified in the pxproteome Protein Summary with P>=0.9 (i.e., by ProteinProphet)? (A9)

    - Of all the (GO-annotated) yeast peroxisomal proteins, how many were found in (use Compare Exps and P>=0.9): (A10)
        - Pxproteomics
        - jranish-gricat
        - both
        - how many not seen at all? (hint: remove default "# of matches constraint")

    - Why so few peroxisomal proteins in gricat? (A11)

    - Create a list of a few proteins not currently annotated as peroxisomal but that might be peroxisomal based on a comparison of identifications in pxproteome and gricat. (A12) (note that a "!" can be used to negate, e.g.: "!%perox%").

Answers:

- A1: 96.6% (although 99.7% if one excludes regions unlikely to be observed)
- A2: TCNVLVAIEQQSPDIAQGLHYEK & ICDQVSDAILDACLEQDPFSK are ranked 1 & 3 observed and 8 & 7 predicted
- A3: Up to 5 proteins and up to 2 different genome locations
- A4: They differ by just two residues and both the YLR180W variants are seen (Hint: realign just these two proteins)
- A5: Both spectrum matches certainly look plausible. However, the Tandem-K hit is doubly tryptic, while the SEQUEST one is just singly tryptic. A peek at the SEQUEST .out file shows that the correct hit was #2.
- A6: The SEQUEST hit peptide has not been seen at all by PeptideAtlas, while the Tandem-K hit peptides has been seen several times in the Yeast PeptideAtlas in multiple charge states (although not the exact modification seen here). This lends considerable credibility to the assignment.
- A7: Most of the discrepancies are really just I/L substitutions, which cannot be differentiated in a mass spectrometer since they have the same mass.
- A8: 57 matches, 51 of which have > 1 hits (sort increasing N Peptides and count)
- A9: With P>=0.9 for both proteins and protein groups and cellular component constraint "%perox%", there are 29 proteins.
- A10: For the first 3, query using Compare Experiments with cellular component constraint "%perox%" and use the summary table at the bottom to get:
  - Pxproteomics - 29
  - jranish-gricat - 6
  - both – 5 of total 30
  - not seen at all? (hint was: remove default "# of matches constraint"). By removing this constraint, the query will display all proteins that have the %perox% annotation, not just the ones seen somewhere in the selected experiments. Now there should be a total of 50 proteins. So 50 – 30 = 20 were not seen at all.
- A11: As described in the experimental description, yeast does not form peroxisomes under most normal growth conditions, but rather they must be specifically induced.
- A12: Start a fresh Compare Experiments query, choose the two yeast experiments we've been working with, set probability to ">.9", set "# in Experiment 1 constraint" to "<3" and to ">3" for Experiment 2. Using the number 3 is subject to personal taste, but in effect this displays proteins observed 4 or more times in pxproteomics and 2 or fewer times in gricat. Explicitly remove annotated peroxisomal proteins with "!%perox%" in the cellular component constraint. A set of 17 proteins matching these criteria should be displayed. Some of these may be suitable for follow-up experiments.