SBEAMS – Proteomics Tutorial
Proteomics Course 2006-10-20
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. Login:
    - 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 Summarize Fractions to view the msruns (fractions) available for experiments
        - Choose jranish – gricat experiment and QUERY (scroll down!)
    - 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 Input Form Format [Full Detail] 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 nbaliga – haloICAT – Mascot
    - Also CTRL Select nbaliga – haloICAT – Sequest
    - 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 Chr_ORF1215
        - What is the protein coverage? (Answer A1 at bottom)
        - How many experiments was this protein seen in? (A2)
    - Click on PA link next to Chr_ORF1950.
        - How many peptides are shared with another protein? (A3)
        - Click on Cytoscape link to visualize
- Choose Compare By Spec to compare two search batches by spectrum to find out examine the differences between search results using different engines or parameters
    - Select nbaliga – haloICAT – Mascot
    - Also CTRL Select nbaliga – haloICAT – Sequest
    - Probability in Experiment 1: >.95
    - Probability in Experiment 2: <.5
    - QUERY
    - Examine resulting list

        - Click on haloICAT2_33.1078.1078.3 and examine spectra
            - Which is likely to be correct? What's good or bad about the spectra? (A4)
            - Has either peptide been seen other experiments? (A5)
    - Choose Browse Biosequences to view the contents of the reference FASTA files
        - Select Yeast ORFs Database
        - 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
        - Enter %perox% in cellular component 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. This afternoon will be all Cytoscape!

5. 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)? (A6)

    - Of all the (GO-annotated) yeast peroxisomal proteins, how many were found in (use Compare Exps and P>=0.9): (A7)
        - Pxproteomics
        - jranish-gricat
        - both
        - how many not seen at all? (hint: remove default "# of matches constraint")

    - Why so few peroxisomal proteins in gricat? (A8)

    - 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. (A9)

Answers:

- A1: 50%
- A2: 21
- A3: 6
- A4: The Mascot hit is more likely to be correct. It hits a Halobacterium protein instead of a human one and is doubly tryptic instead of singly. However, the two highest peaks are not identified, which is disturbing. The SEQUEST spectrum has a lot of nice matched peaks, but the highest is a B2 series, which seems implausible.
- A5: The SEQUEST hit peptide has not been seen at all by PeptideAtlas (not even in the human atlases), while the Mascot hit peptides has been seen many times in the Halobacterium PeptideAtlas with multiple modifications and charge states. This lends considerable credibility to the assignment.
- A6: With P>=0.9 for both proteins and protein groups and cellular component constraint "%perox%", there are 29 proteins.
- A7: 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.
- A8: As described in the experimental description, yeast does not form peroxisomes under most normal growth conditions, but rather they must be specifically induced.
- A9: Start a fresh Compare Experiments query, choose the two yeast experiments we've been working with, 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. You can 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 to follow up experiments.