Day 2: QualScore Tutorial

Introduction

The following tutorial / exercise is intended to familiarize the user with the operation of QualScore. First we will learn how to generate PeptideProphet results that include the low probability identifications that QualScore needs to train the classifier and then visually examine both high and low probability sequence to spectrum assignments to get a feel for what good and bad tandem mass spectra look like. Next, we will run the QualScore program to train the classifier and extract high quality spectra that were not correctly identified. Finally, we will examine the results of searching these spectra using a different algorithm and allowing for more potential peptide modifications.

1. Log on to PETUNIA and run XInteract with PeptideProphet to collate and validate first pass search results
   - 1.1. Open a web browser and go to http://localhost/tpp-bin/tpp_gui.pl.
   - 1.2. Log on using user name *guest* and password *guest*.
   - 1.3. Select Sequest from the pipeline selection drop down box. Select the Analysis Pipeline tab then select the Analyze Peptides tab.
   - 1.4. In the Select File(s) to Analyze pane, click on the Add Files button. Using the links under *DIRECTORIES:*, browse to `c:\Inetpub\wwwroot\ISB\data\class\QualScore\RaftFlow\`
   - 1.5. Select the check boxes next to the 9 raftflowXX xml files (`raftflow31.xml, raftflow32.xml, raftflow33.xml, raftflow34.xml, raftflow35.xml, raftflow36.xml, raftflow37.xml, raftflow38.xml, raftflow39.xml`) and click the Select button. The Select File(s) to Analyze pane should now list the 9 pepxml files.
   - 1.6. Enter `interact-qualscore.xml` in the Write output to file: field in the Output File and Filter Options pane.
   - 1.7. **!!! IMPORTANT !!!** Enter `0.0` in the Filter out results below this PeptideProphet probability: box. This is the most critical step. QualScore will not run correctly if this step is not done.
   - 1.8. Make sure that the RUN PeptideProphet check box is selected.
   - 1.9. Click Run XInteract button in the Run Analysis! pane. You can click on Show next to the Command Status pane to track the progress of the command.
2. Evaluate the results of the first pass search

2.1. When the commands have finished executing, click on the here link to view the output files (`Click here to view log file and output files`).

2.2. In the Output Files pane, click on View next to the interact file path (`c:\Inetpub\wwwroot\ISB\data\class\QualScore\RaftFlow\interact-qualscore.shtml [ View ]`). This will open the PeptideProphet result file using PepXMLViewer.

2.3. Click on any probability link to display the PeptideProphet analysis results. Answer the following questions:

2.3.1. Do the discriminate score distributions look reasonable? _____

2.3.2. How many estimated correct identifications are in the dataset? _____A

2.3.3. How many spectra were searched? _____

2.3.4. Approximately how many 'real' scans does the above number represent? _____B

[hint: average the number of 2+ and 3+ spectra and add the number of 1+ scans]

2.3.5. What is the percentage of scans positively identified? _____[A / B]

2.4. Close the PeptideProphet analysis result page.

3. Examine 'good' and 'bad' spectra, what do they look like?

3.1. Return to the PepXMLViewer and use the display options to sort the list of identifications by *probability descending*. Randomly visualize several identifications by clicking on the value in the ions column. Note that these spectra should exhibit similar features that lead to good identifications such as series of fragment ions well above the background noise.

3.2. Use the display options to sort the identifications by *probability ascending*. Randomly visualize several identifications by clicking on the value in the ions column. Note that these spectra should exhibit similar features that lead to poor identifications such as insufficient fragment ions or greater levels of noise (commonly called 'scratchy spectra').

4. Run QualScore

4.1. Specify search results file: Return to Petunia, navigate to the QualScore interface (on the Tools tab). Click the Add Files button in the Specify search results file pane. Use the controls to navigate to `c:\Inetpub\wwwroot\ISB\data\class\QualScore\RaftFlow\`, check the box for `interact-qualscore.xml` file and click the Select button.

4.2. Options: do not select any options. In default mode, QualScore creates a directory into which it will write any high quality unassigned spectra that it finds.

4.3. Look for Unassigned High Quality Spectra!: Click on the Run QualScore button.

4.4. Periodically click on the UPDATE THIS PAGE link to monitor QualScore's progress. When QualScore has finished, examine the log file in the Output so far pane and answer the following questions:

4.4.1. How many spectra were features calculated for? _____

4.4.2. How many spectra were used for the 'good' and 'bad' data sets? _____ _____

4.4.3. How many unassigned spectra are there? _____

4.4.4. How many high quality unassigned spectra did QualScore find? _____

[hint: this is half the number of dtas (spectrum files) written to the interact-qualscore.qdir directory]

5. Visually inspect the extracted high quality unassigned spectra

5.1. Open to the Browse files tab (under the Tools link) and use the links under *DIRECTORIES:* to browse to the c:\Inetpub\wwwroot\ISB\data\class\QualScore\RaftFlow\interact-qualscore.qdir directory.

5.2. Randomly select and visualize at least 10 spectra by clicking on their corresponding [spectrum] link. A separate browser window will open displaying the spectrum. Confirm that these spectra are good quality (they should resemble the high probability spectra observed in 3.1.

5.3. These spectra are now ready for further analysis. In theory, these spectra are enriched for PTMs or other modifications not considered in the original search or novel peptide sequences not represented in the protein database. A typical strategy for identifying these spectra might involve re-searching them using a wider selection of potential peptide modifications.

6. Re-searching the high quality unassigned spectra

6.1. The high quality unassigned spectra from the above dataset were searched using Mascot allowing for the following variable modifications: Carbamyl (N-term) +44,N-Acetyl (Protein) +43 ,Oxidation (M) +16,Oxidation (HW) +16 ,Pyro-glu (N-term E) -17,Pyro-glu (N-term Q) -17. Before we can view the results, we must first combine them using xinteract.

6.2. Switch to the Mascot pipeline by going to the Home tab and selecting Mascot from the drop-down box.

6.3. Select the Analysis Pipeline tab then select the Analyze Peptides tab. In the Select File(s) to Analyze pane, click on the Add Files button. Using the links under *DIRECTORIES:*, browse to `c:\Inetpub\wwwroot\ISB\data\class\QualScore\2nd_Search\`

6.4. Select the check boxes next to the 9 raftflowXX xml files (`raftflow31.xml, raftflow32.xml, raftflow33.xml, raftflow34.xml, raftflow35.xml, raftflow36.xml, raftflow37.xml, raftflow38.xml, raftflow39.xml`) and click the Select button. Enter `0.0` in the Filter out results below this PeptideProphet probability: box.

6.5. Make sure that the RUN PeptideProphet check box is selected and click Run XInteract button in the Run Analysis! pane.

6.6. Use the Command Status pane to track the progress of the command. When the commands have finished executing, click on the here link to view the output files (`Click here to view log file and output files`).

6.7. In the Output Files pane, click on View next to the interact file path (`c:\Inetpub\wwwroot\ISB\data\class\QualScore\2nd_Search\interact.shtml [ View ]`). This will open the PeptideProphet result file using PepXMLViewer.

6.8. Click on any probability link to display the PeptideProphet analysis results. Answer the following questions:

6.8.1. Do the discriminate score distributions look reasonable? _____

6.8.2. How many estimated correct identifications are in the dataset? _____C

6.8.3. How many spectra were searched? _____

6.9. Close the PeptideProphet analysis result page.

6.10. Sort the Search results by probability descending. Visualize at least 5 of the identifications by clicking on the value in the ions column. Do the identifications look good?

6.11. Use the PepXMLViewer's Display Options and Filter Options to answer the following questions:

6.11.1. How many spectra were identified as having Pyro-glu (N-term Q)? _____

[hint: enter ^Q in both the hilight peptide text (regex): and required peptide text (regex allowed): fields.]

6.11.2. How many spectra were identified as having Pyro-glu (N-term E)? _____

6.11.3. How many spectra were identified as having Carbamyl (N-term)? _____

6.11.4. How many spectra were identified as having N-Acetyl (Protein)? _____

6.11.5. What is the total number of modified spectra identified? _____

6.12. Combine the above counts with those made in section 2.3. What is the increase in the percentage of assigned spectra? _____