DISCO – Tutorial version 0.2

For this tutorial we will be using some peak picked Orbitrap Elite DIA data

The data were already converted to mzML using msconvert. The search results were processed individually by PeptideProphet and combined using iProphet. This tutorial starts by running PTMProphet on the iProphet combined pepXML results.

1. Ensure that you are connected and have the latest Trans-Proteomic Pipeline (TPP)
    - This tutorial is written with details for TPP on a Microsoft Windows operating system. If you have a different kind of operating system, this tutorial should still work, but the details of the installation and some file path locations will be somewhat different
    - Ensure that you are Internet connected for this tutorial because you will download files
    - This tutorial requires that you have TPP 6.3.2 or later installed on your system.
    - If you have a pre-6.3.2 version of TPP already installed on your system, uninstall the older version first by clicking **[Start]:[Trans-Proteomics Pipeline]:[Uninstall TPP]**
    - If TPP 6.3.2 or greater is not installed on your system, install that first by downloading and following the instructions at http://tools.proteomecenter.org/wiki/index.php?title=TPP:5.2_Installation. Note that you will need to restart your computer or manually start the Apache web server service

2. Launch the TPP web browser Graphical User Interface
    - **[Start]:[Trans-Proteomic Pipeline]:[TPP Web Interface]** or open a web browser to: ***http://localhost:10401/tpp/cgi-bin/tpp_gui.pl***
    - Login with the username ***'guest'*** with password ***'guest'*** (or alternate account if you have one)

3. In the web browser GUI, download the tutorial data:
    - Click the **[Files]** tab at the top
    - In the bottom right, create new directory: tutorials
    - Click **[TPP Tools]:[Fetch Datasets]**
    - Click on **[Show version information and available features]**
    - If there is a newer version, click on **[Update to the latest version of fetchDataset]**. Version ***0.8.1*** or greater is required
    - Click **[TPP Tools]:[fetch datasets]** again
    - Click **(Add Files)**
    - Checkmark the ***"tutorials"*** directory and click **(Select)**
    - Paste the following URL into the [Dataset Identifier or URL] box:
      http://www.tppms.org/tools/disco/DISCO.zip
    - Click **(Fetch Dataset)**
    - Monitor the job by clicking **[Refresh]** until download and unzip is complete

4. Run DISCO on the example dataset
    - Click **[TPP Tools]:[DIA:Extract MS2 Fragments]**
    - Click on **(Add Files)**, and navigate to ***tutorials/DISCO***
    - This directory contains the mzML data: ***011618-DIA-02.mzML***.
    - Checkmark ***011618-DIA-02.mzML*** and click **(Select)**

- Under **[Choose Hardklor params]** select: ***c:/TPP/data/tutorials/DISCO/Hardklor.conf***
- Under **[Disco Options]**
  a. **[Scan Range]:[Start]: *50000***
  b. **[Scan Range]:[End]: *55000***
  c. If you know you have multiple core, you can specify number of threads to number of cores
  d. **[Additional (expert) options]: *MAXPPM=8***
- Click on **(Run Disco)**
- Click on **[Refresh]** to monitor progress. (About 2-4 minutes on a modern computer)
- Ensure that the job did not end with an error

5. Run Comet on the DISCO generated data file
   - Open **[TPP Tools]:[Comet Search]**
   - Remove ***011618-DIA-02.mzML*** as a selected file (checkmark file on right and click **[Remove]**)
   - In **"Choose from list"**, select the ***011618-DIA-02_ds.mzML*** file (note the _**ds** suffix)
   - Remove the ***Hardklor.conf*** from the Comet parameters file (checkmark and click **[Remove]**)
   - Click **[Add Files]**, navigate into ***"com"*** subfolder and then select the ***comet.params*** file and click **[Select]**
   - Next Choose a sequence database. Go up to ***DISCO*** folder. Then into dbase folder. Checkmark ***PA_THISP_Level1_2018-05-01_targetdecoy_iRT.fasta*** and click **[Select]**

- Click on **(Run Comet)**
- Click on **[Refresh]** to monitor progress. (About 2-4 minutes on a modern computer)
- Ensure that the job did not end with an error

6. Run the Prophets on the Comet results
    - Open **[TPP Tools]:[Analyze Peptides]**
    - Remove any selected input files if present by checkmarking them and clicking **[Remove]**
    - Navigate to the ***DISCO/com*** folder
    - Checkmark the ***011618-DIA-02_ds.pep.xml*** file and click **[Select]**
    - Checkmark **"Use accurate mass binning, using PPM"**
    - Checkmark **"Only use Expect Score as the discriminant"**
    - **Enter additional options to pass directly to the command-line** (expert use only!): ***-PREC***
    - Under iProphet options, select **"RUN iProphet"**
    - Click on **(Run XInteract)**
    - Click on **[Refresh]** to monitor progress. (About 1-2 minutes on a modern computer)
    - Ensure that the job did not end with an error
7. Explore the results
    - Click on the **[PepXML]** link next to ***interact.pep.xml*** (first output file)
    - Click on a probability and examine the models. Do they seem reasonable?
    - What is the probability threshold, number correct and incorrect at 1% FDR?

- Close the model viewer and PepXML Viewer tabs
- Click on the **[PepXML]** link next to ***interact.ipro.pep.xml*** (second output file)
- Click on a probability and examine the models. Do they seem reasonable?
- What is the iProphet probability threshold, number correct and incorrect at 1% model-estimated error-rate?
- Close models and **filter** the PepXML Viewer with an **iProphet probability** of ***0.87*** and **sort by iProphet probability** in ***descending order***
- Explore the Comet matches to DISCo generated spectra
    i. How many PSMs and Peptides pass the threshold?
- Filter for Proteins containing the string "DECOY", check the box next to: **"all mapping proteins and alternatives must match"**
    i. How many of the results passing the 1% model-based error-rate threshold match to DECOY peptides?
    ii. What is the DECOY-estimated error-rate (count 2 wrong results for every passing DECOY)

8. Run reSpect to find more peptides from chimeric specta
    a. Open **[TPP Tools]:[Reanalyze Spectra]**
    b. Add the following file to
    "Choose PepXML File With Probabilities:"
    c:/TPP/data/class/disco_tutorial/DISCO/com/interact.ipro.pep.xml
    c. Set **"m/z tolerance (ions):" 0.1**
    d. Click **(Run reSpect)**
9. Run Comet on the reSpect generated data file ***011618-DIA-02_ds_rs.mzML***

- Open **[TPP Tools]:[Comet Search]**
- Remove ***011618-DIA-02_ds.mzML*** as a selected file (checkmark file on right and click **[Remove]**)
- In **"Choose from list"**, select the ***011618-DIA-02_ds_rs.mzML*** file (note the **_ds_rs** suffix)
- Remove the ***.../com/comet.params*** from the Comet parameters file (checkmark and click **[Remove]**)
- Click **[Add Files]**, navigate into ***"com_rs"*** subfolder and then select the ***comet.params*** file and click **[Select]**
- Next Choose a sequence database. Go up to ***DISCO*** folder. Then into dbase folder. Checkmark ***PA_THISP_Level1_2018-05-01_targetdecoy_iRT.fasta*** and click **[Select]**

10. Run the Prophets on the reSpect+Comet results
    - Open **[TPP Tools]:[Analyze Peptides]**
    - Remove any selected input files if present by checkmarking them and clicking **[Remove]**
    - Navigate to the ***DISCO/com_rs*** folder
    - Checkmark the ***011618-DIA-02_ds_rs.pep.xml*** file and click **[Select]**
    - **Uncheck** checkmark **"Use accurate mass binning, using PPM"**
    - Checkmark **"Only use Expect Score as the discriminant"**
    - **Enter additional options to pass directly to the command-line** (expert use only!): ***-MASSWIDTH=20***
    - Click on **(Run XInteract)**
11. Explore the reSpect results
    a. Click on the **[PepXML]** link next to ***interact.pep.xml*** (first output file)

b. Click on a probability and examine the models. Do they seem reasonable?
c. Close models and **filter** the PepXML Viewer with an **PeptideProphet probability** of ***0.9385*** and **sort by PeptideProphet probability** in ***descending order***
d. Explore the Comet matches to DISCo generated spectra, processed with reSpect
    i. How many PSMs and Peptides pass the threshold?
e. Filter for Proteins containing the string "DECOY", check the box next to: **"all mapping proteins and alternatives must match"**
    i. How many of the results passing the 1% model-based error-rate threshold match to DECOY peptides?
    ii. What is the DECOY-estimated error-rate (count 2 wrong results for every passing DECOY)

12. Run iProphet to combine results of the first search with the results of reSpect search
    a. Using Petunia's File Browser create a new directory
        c:/TPP/data/class/disco_tutorial/DISCO/ipro
    b. Open **[TPP Tools]:[Refine/Combine Analyses]**
    c. Add the following *two files* in
        "FILE(S) TO ANALYZE:"
        c:/TPP/data/class/disco_tutorial/DISCO/com/interact.pep.xml
        c:/TPP/data/class/disco_tutorial/DISCO/com_rs/interact.pep.xml
    d. Set the directory you just created in
        "OUTPUT FILE AND LOCATION"
        Output Directory:
        c:/TPP/data/class/disco_tutorial/DISCO/ipro

e. Checkmark **"Run ProteinProphet on these results"**
f. Click **(Run iProphet)**

13. Explore the iProphet results that include both the DISCo analysis and Comet search and the reSpect results
    a. Click on the **[PepXML]** link next to ***interact.ipro.pep.xml***
    b. Click on a probability and examine the models. Do they seem reasonable?
    c. Close models and **filter** the PepXML Viewer with an **iProphet probability** of ***0.8974*** and **sort by iProphet probability** in ***descending order***
        i. How many PSMs and Peptides pass the threshold?
    d. Filter for Proteins containing the string "DECOY", check the box next to: **"all mapping proteins and alternatives must match"**
        i. How many of the results passing the 1% model-based error-rate threshold match to DECOY peptides?
        ii. What is the DECOY-estimated error-rate (count 2 wrong results for every passing DECOY)
14. Group the PSMs by protein, click **[Display Options]**
    a. In **"Sorting:"** dropdown menus select to sort on ***"protein"*** and by: ***"peptide"*** and then by ***"spectrum"***
    b. In **"Highlight spectrum text"** type ***_rs***
    c. In **"Rows per page"** select ***1000***
    d. Click **(Update Page)**
    e. Are there any proteins identified only by _rs spectra?