PIQED Installation and Usage Tutorial

Jesse G. Meyer, Ph.D.
Release version 1.0
May 10, 2017
Please send problems or suggestions to jessegmeyer@gmail.com and/or bschilling@buckinstitute.org

Getting started:

- Download the github repository, un-zip, and move the folder to C drive.
- Download the example raw MS/MS data (.wiff and .wiff.scan files) from Massive using dataset ID: MSV000080189

Download and install all of the following programs (*denotes executables that must be added to the system path variable):

- msconvert.exe (Proteowizard)
- DIA Umpire, DIA_Umpire_SE.jar
- MS-GF+, msgfplus.jar
- mapDIA, copy mapDIA.exe into C:\...\DIA-pipe\bin\
- Skyline and SkylineRunner.exe
- *Python 2.7
- *Java AND *Javac, java Compiler (included in JDK), added to the system path variable
- *R version > 3.1.0, added to the system path variable
- comet binaries – must use 64-bit version, NOT the 32-bit included with the TPP

Download and install TPP 5.0 programs:

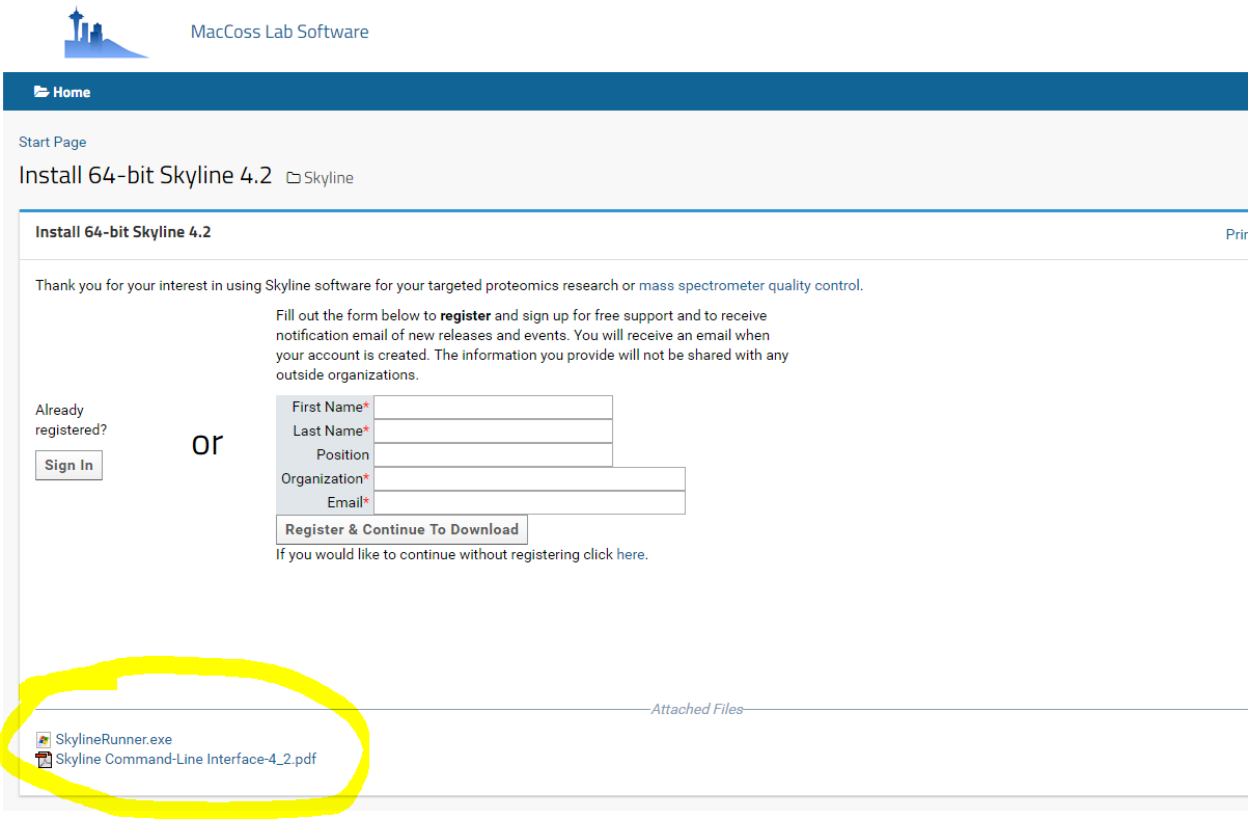
- xinteract.exe
- InterProphetParser.exe
- tandem.exe
- Tandem2XML.exe
- indexmzXML.exe
- PTMProphetParser.exe development version required

Suggested hardware:

- Windows 7
- Intel i5 quad-core minimum, Xeon >16 core recommended
- 4GB RAM minimum, >32 GB DDR4 RAM recommended
- 1TB 7200 rpm hard drive minimum, 1TB Solid-state hard drive recommended

Skylinerunner.exe download:

The automated workflow uses skylinerunner.exe, which is downloaded separately from Skyline's GUI version. Get it from the link at the bottom of the download page:

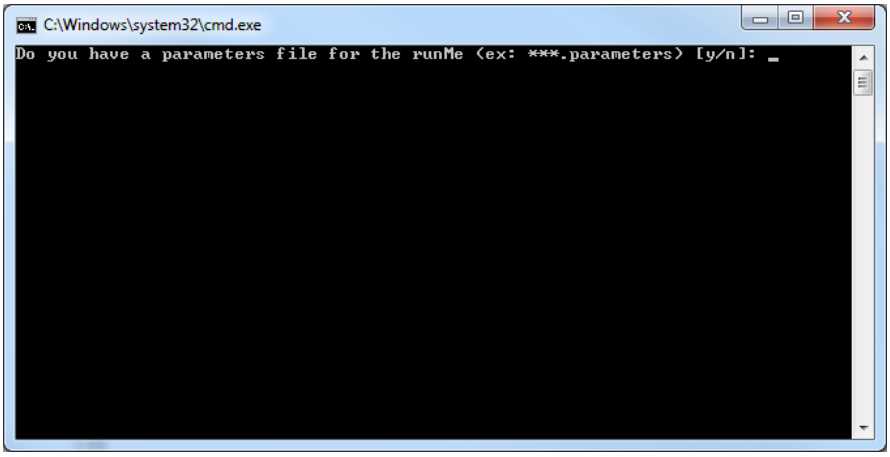


IMPORTANT NOTES before you begin:

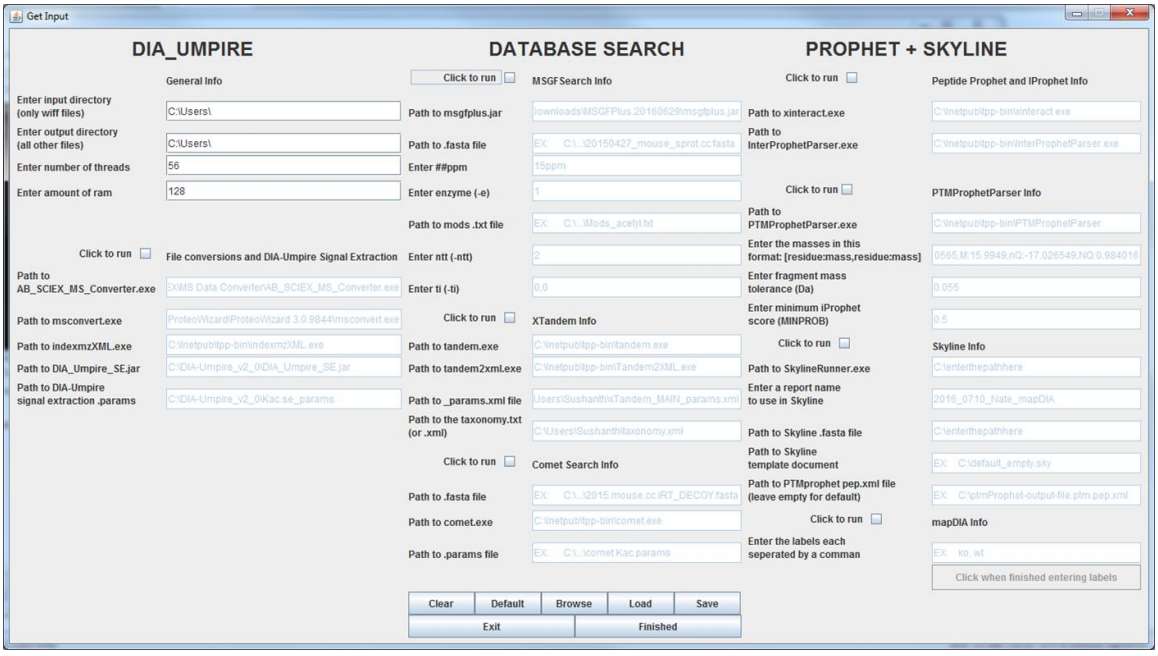
- **Each pipeline step associated with a "click to run" checkbox can be run independently**
- **Prepare your parameter files** before starting the pipeline, for example:
  - DIA-Umpire Signal Extraction parameters file
  - Database search parameter files, e.g. FASTA, modifications, taxonomy.xml, etc.
  - Skyline template document
- **Modify example parameter files whenever possible** instead of using your own because some steps in the program look for specific lines within these files.
- **Place all raw mass spectrometry files into a folder on your C:\ drive**. From SWATH that means your ".wiff" and ".wiff.scan" files, or from Thermo DIA that means all ".RAW" files. You don't need to explicitly specify whether the data is from Sciex or Thermo, the pipeline will automatically detect the file type and convert the files to .mzXML as appropriate. The parameters used for downstream steps for each type of data are quite different.
- **Use two different fasta files for the database searches (module 2)**. The MSGF+ search uses a database not containing decoys (e.g. 20150810.mouse.cc.iRT.fasta), as they will be created automatically by the search engine. The X! Tandem and COMET searches use a database containing decoys (e.g. 20150810.mouse.cc.iRT_DECOY.fasta).
- **For the mapDIA section to work properly** through mapDIA quantification, files from each condition must all contain the same unique text identifier, e.g. when comparing two replicates each of wild-type (WT) and knockout (KO) samples, files could be named: "DATE_WT1.wiff, DATE_WT2.wiff, DATE_KO1.wiff, DATE_KO2.wiff". **Alternatively,** all file names must either contain text with their group name, or the directory must contain a tab-delimited text file named "name_mapping.txt" where the first row contains names of the group names in the same order entered into the main PIQED GUI before generating the pop-up pictured below, and the rows below list the names of the files in each condition.
- **To enable protein-level correction,** include a file in the output directory named "proteinlevel.txt", which will contain uniport identifiers that match those in your FASTA databases in the first column, followed by the measured areas for each PTM analysis replicate in subsequent columns that are in the same order as the skyline document (alphabetical).
- **If it doesn't work:** Check that you entered file paths correctly. Look in your working directory for the last windows batch file that worked, which would be indicated by the presence of the appropriate files created after the batch file.

Starting DIA-Pipe and setting up files:

1. Within the unzipped “PIQEDia-master” folder you downloaded from github, open the folder “bin” and locate the file “runME.bat” right-click on the file and chose “Run as Administrator.”
2. Command prompt should appear with the question, ”Do you have a parameters file for the runME (ex: ***.parameters) [y/n]:” type “n” and press enter.

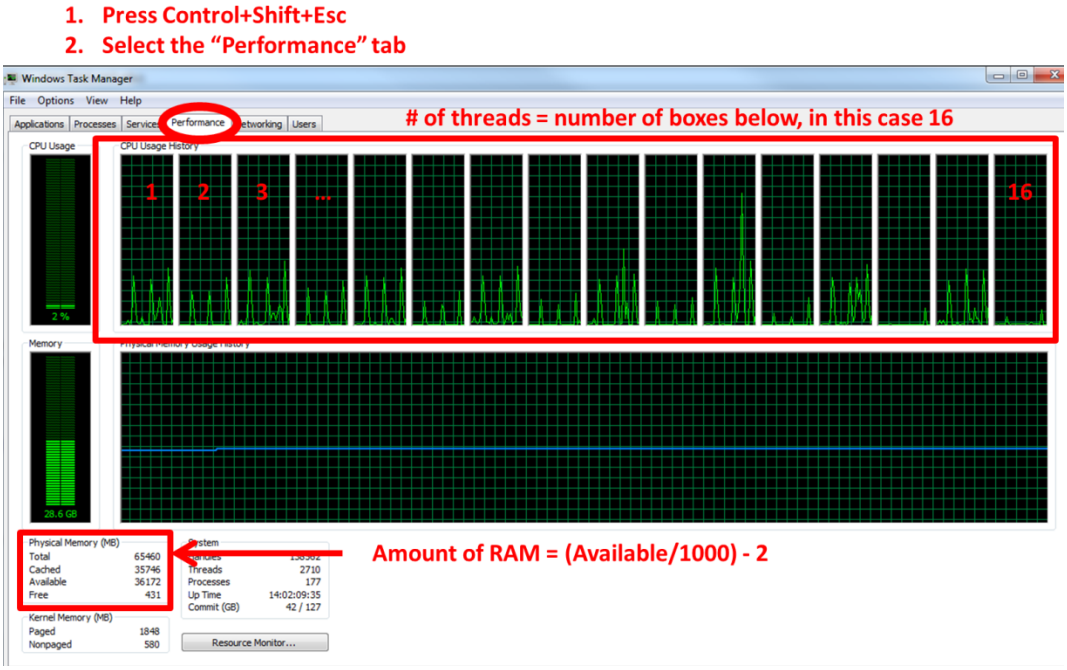


3. The GUI should appear:



4. At the top left under **General Info**, enter the full path to the directory where all .wiff + .wiff.scan files OR all .raw files are located as the input and output directories.

5. Enter the number of threads (equal to or less than the number of CPU cores), and enter the amount of ram (equal to or less than the amount of RAM installed). To determine the number of cores and the amount of RAM, press control+shift+Esc, which opens the task manager, then click on the performance tab, and count the boxes:



Module 1, File Conversions and DIA-Umpire signal extraction:

1. Check the box, "Click to run" under each section to activate the fields.
2. In the bottom-left section, specify the locations for all the file conversion executables. **NOTE: Use of Sciex data converter is currently depreciated! You do not need to specify the actual location, but this box must contain text, e.g. "none".**
3. In the bottom-left box, enter the path to **your specific** DIA-Umpire signal extraction parameter file

DIA-Umpire Signal extraction parameter notes:

Sciex and/or tutorial data example parameter file 'diaumpire_se.params' is available in the github repository under the "params" folder. This example can be used with the "halfDIA" and "fullDIA" files from massive. Open the params file with a text editor and edit the first variable "Threads=" to reflect the # of threads available on your machine.

Orbitrap data example parameter file is available from github under the parameter folder: (1) "diaumpire_se_orbi_strict.txt" was used for the non-enriched urine data in the manuscript, or (2) "diaumpire_se_orbi.txt" for enriched samples.

For more details on these parameters, please see the DIA-Umpire documentation: http://diaumpire.sourceforge.net/?page_id=19

Module 2, Database Searches:

1. Check the boxes next to each of the database searches you want to run.
2. Enter the locations of the files for all the search program executables and the parameters.

MSGF+ search: If using the example SWATH files from MASSIVE, under MSGF+ enter:

Path to .fasta file = C:\[your path here]\20150810.mouse.cc.iRT.fasta
ppm = 25ppm
enzyme= 1
mods.txt file = tutorial_MSGFmods.txt
ntt = 1
ti= 0,0

X! tandem search: If using the example SWATH files from MASSIVE, use the params.xml file "xTandem_Kac_params.xml" and the taxonomy.xml file from github. DIA-Pipe looks for specific lines of text in the X! Tandem Parameters file. **You must edit two lines inside the tandem.params.xml :** (1) the path to your scoring parameter file (e.g. kscore downloaded with TPP), and (2) within the taxonomy.xml file, update the path to the database to reflect the full path to the .fasta database on your computer, e.g. "20150810.mouse.cc.iRT_DECOY.fasta". **Do not edit the input and output file locations within the tandem.params file.**

COMET search: If using the example SWATH files from MASSIVE, use the "comet64.sciex.Kac.params" file from github, and open the file to update the first line "database_name =" to reflect the location of "20150810.mouse.cc.iRT_DECOY.fasta" on your computer.

Module 3, Prophet Refinement, Skyline Signal Extraction, and mapDIA:

PeptideProphet, iProphet, and PTMprophet

1. Check the boxes for all the desired steps and enter the requested information.

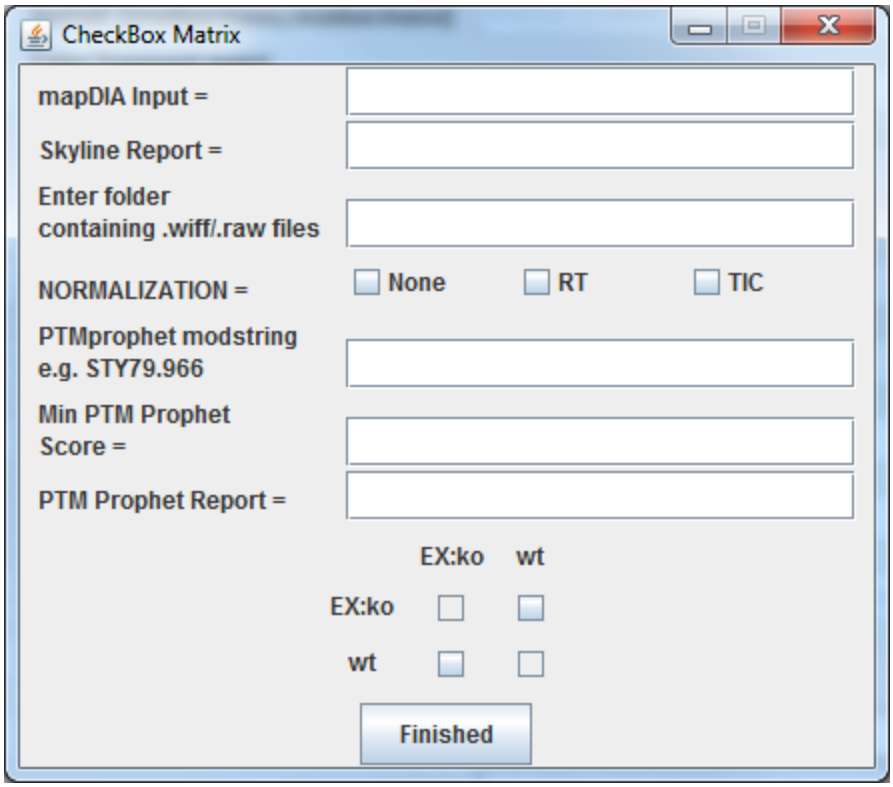
2. Under “Enter the masses in this format..”, enter the mass corresponding to your modifications of interest in the form [residue:modmass,residue:modmass]. For the tutorial enter:
   K:42.0105,M:15.9949,nQ:-17.026549
3. Enter the fragment mass tolerance appropriate for your instrument (tutorial enter 0.055)
4. Enter the minimum iprobability used to consider peptides for site-localization scoring (0.99 should be safe).

Skyline Signal Extraction and report generation

1. Enter the full path to SkylineRunner.exe
2. Enter the report name “PIQED_mapDIA”
3. Enter the full path to the .fasta file used for the MS-GF+ database search.
4. Skyline Template document:
   a. Tutorial data: Copy the skyline template document file “~\PIQEDia\skyline\5600_tutorial_template.sky” to your output directory as specified under the “General Info” section.
   b. User datasets: Start with the 5600 template “~\PIQEDia\skyline\default_empty.sky” or the orbitrap template document “~\PIQEDia\skyline\default_orbi.sky” or open the template document and edit the settings as appropriate for your data. Add the report “PIQED_mapDIA.skyr” to your document report list. Save the empty document to the output directory containing your .mzXML files.
5. Copy the skyline report file from “..\DIA-Pipe\skyline\PIQED_mapDIA.skyr” to your output directory as specified under the “General Info” section.
6. If using the tutorial files from MASSIVE, enter the full path to the skyline template document “default_empty.sky” that you copied to your output directory in step 4.
7. Erase the contents of the box to the right of “Path to PTMprophet pep.xml file”

mapDIA

1. To setup mapDIA, enter the conditions-specific labels contained in the raw file names separated by commas, and then click, “Click when finished entering labels.” If using the tutorial files, enter “fulldia” and “halfdia” as condition labels.
2. The following box should appear:



NOTE: The mapDIA module requires either the path to your mapDIA input file (if running this module independent of the other steps), OR the name of the skyline report created from the previous steps. Enter only one of these.

5. If running mapDIA using results from previous modules (e.g. with tutorial data), and using the skyline template file with template report, then enter "2016_0826_mapDIA.csv" as the **skyline report** and leave the **mapDIA Input** box blank.
6. Enter the starting directory containing your .wiff files (if completing the entire pipeline, this path will match the path under the first section "Enter Input directory)
7. Choose the type of normalization (see mapDIA documentation). If using the example SWATH files from MASSIVE, choose "None."
8. Enter the amino acid and modification mass **rounded to the 3rd decimal** (e.g. enter K:42.011 for acetylation if using the example SWATH files from MASSIVE).
9. Enter the minimum localization score you are willing to accept for your quantified results (0.99 suggested).t
10. Enter the name of the PTM Prophet report to use for finding the localization scores. If this was produced using the PTMProphet section, use the name: "ptmProphet-output-file.ptm.pep.xml".

11. Make checks in the boxes where you want comparisons to be made. For each comparison, the column title will be the denominator. For example, to compute KO/WT in the above example, check the top-right box. To compute, WT/KO, check the bottom-left box. **NOTE:** The same comparison cannot be made twice with alternate denominators, e.g. WT/KO and KO/WT will cause an error from mapDIA.
12. Once everything is correctly filled out, click "finished" to return to the main window.
13. Save the parameters you entered (in case of an error) by clicking the save button and choosing one of the save slots:

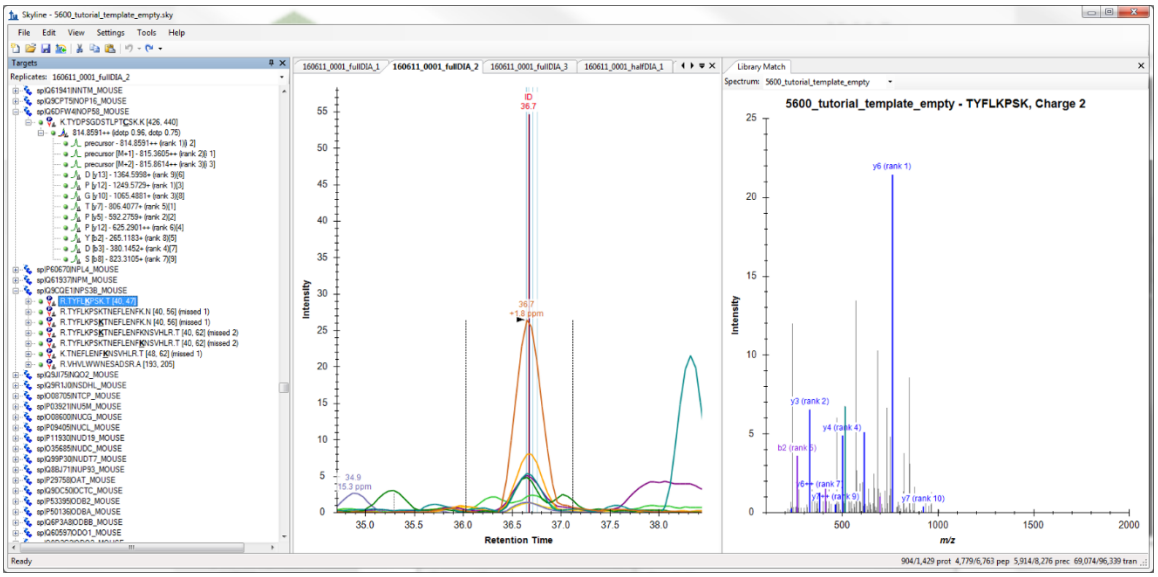
14. To run the pipeline, click the "finished" button at the bottom of the page. A command prompt will pop up and display the status of the current commands as they run.
15. After clicking the finished button, a window will open saying that 'errors may occur'. This does not mean that the system paths were entered incorrectly. It is just a warning. This warning is telling you that if the database searches do not find your specified modification (acetylation if using tutorial data), then the Prophet refinement will fail.

Expected results:

The pipeline will produce:

1. .mzXML files for each of the input .wiff/.raw files
2. Q1-3.mzXML files from the DIA-Umpire signal extraction containing searchable pseudo-MSMS spectra from your DIA runs
3. MS-GF+ outputs in .mzid and .pep.xml format
4. X!Tandem outputs in the format of .tandem and .pep.xml
5. Comet output in pep.xml format

6. PeptideProphet pep.xml outputs for each individual database search – a total of 9 for each initial .wiff/.raw file (1 file -> Q1, Q2,Q3 X 3 searches = 9 results)
7. One combined iProphet search result from the combination of all database searches
8. One PTM-localized pep.xml output from PTMprophet containing all the database search results
9. Skyline .sky and .skyd file containing all your identified peptides (including unmodified peptides):



10. Skyline .csv report containing all identified peptides.
11. Filtered and reformatted report for input to mapDIA
12. mapDIA output files including “analysis_output.txt” which contains the statistical results from all comparisons