

DIA-Pipe GUI tutorial

Release version 1.0

Jesse G. Meyer, Ph.D

February 9th, 2017

- 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](#) 'raw' folder (**ID:** MSV000080189 and **password:** \$diapip3)
- Make sure there are no spaces in path directories

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

- [AB SCIEX MS Converter.exe](#)
- [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, executable added to the system path variable or located at C:\Python27\python.exe](#)
- Java executable added to the system path variable
- [*Javac, java Compiler \(included in JDK\)](#), added to the system path variable
- [*R](#), added to the system path variable

Download and install [TPP 5.0 programs](#):

- xinteract.exe
- InterProphetParser.exe
- PTMProphetParser.exe [development version required](#)
- tandem.exe
- Tandem2XML.exe
- [comet.exe](#) – you must download the latest version of comet binaries to use the 64-bit version, NOT the 32-bit executable included with the TPP installation
- indexmzXML.exe

Suggested hardware:

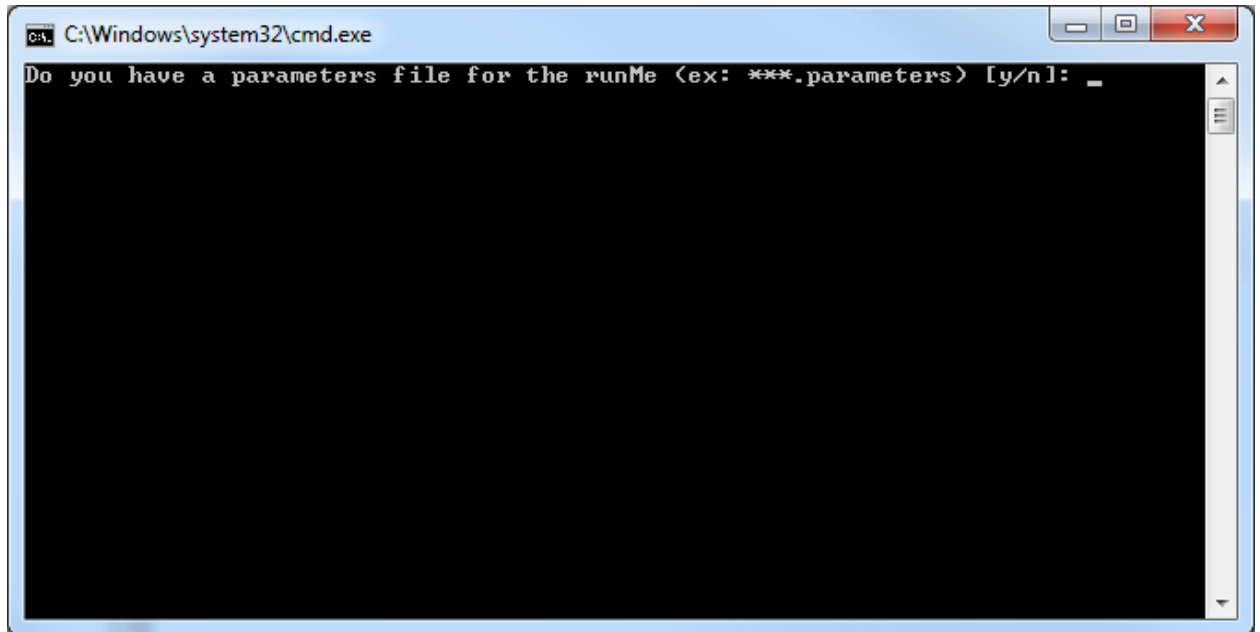
- Intel i5 quad-core minimum, Xeon >16 core recommended
- 4GB RAM minimum, >32 GB DDR4 RAM recommended
- 1TB 7200 rpm hard drive minimum, 1TB Solid-state hard drive recommended

*****IMPORTANT NOTES before you begin:**

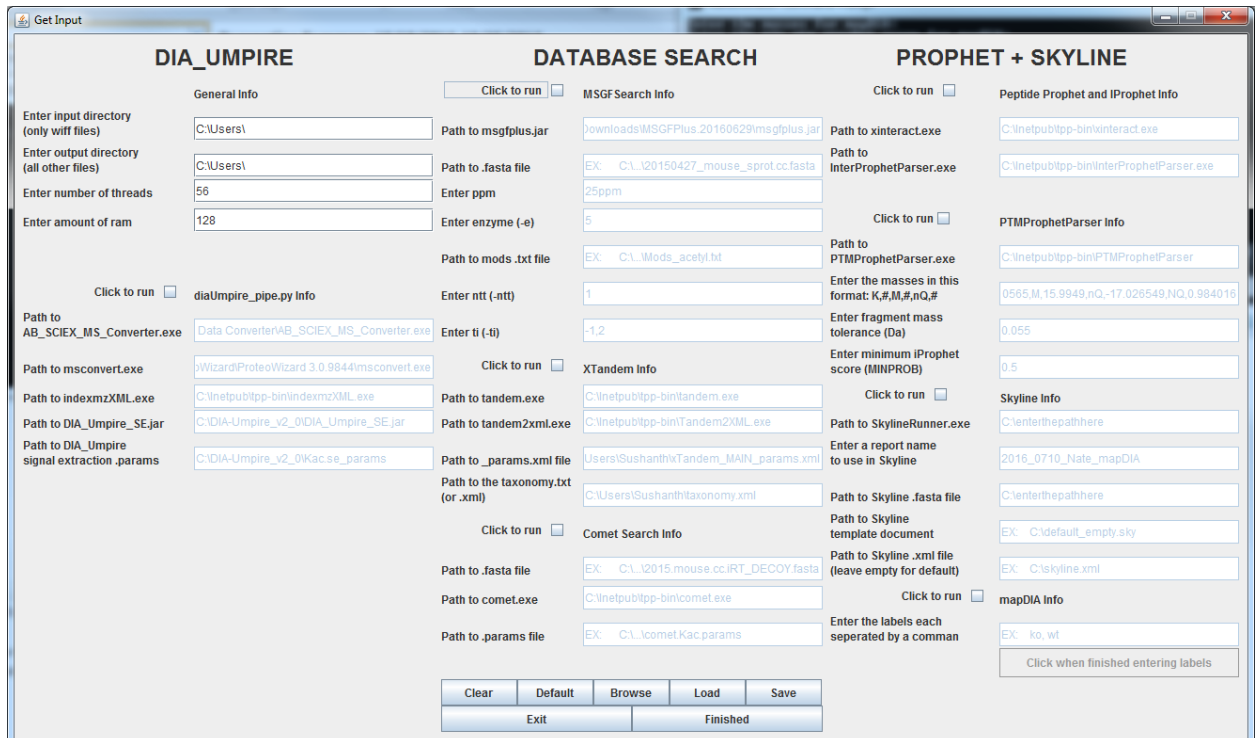
- **Avoid long path names** for all file locations you enter into the GUI because some steps are processed in parallel and command length becomes an issue
- **Prepare your parameter files** before starting the pipeline, for example:
 - DIA-Umpire Signal Extraction parameters file
 - Database search parameter files, e.g. FASTA, modifications, taxonomy.xml, etc.
 - Skyline template document
- **To complete the entire pipeline** 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”.
- **To enable protein-level correction**, include a file in the output directory named “proteinlevel.txt”, which will contain uniprot 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).\
- **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.
- **Use two different fasta files for the searches.** 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).

Starting DIA-Pipe and setting up files:

1. Right click on “runME.bat” and choose “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).
 6. Enter the amount of ram to use (equal to or less than the amount of RAM installed).
- Check the box, “Click to run” under each section to activate the fields.

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

***The pipeline detects the presence of either .wiff files or .raw files and performs the appropriate file conversions for either input file type. ***

1. In the bottom-left section, you need to specify the locations for all the file conversion executables.
2. In the bottom box, enter the path to **your specific** DIA-Umpire signal extraction parameter file containing your desired settings and possibly your variable window definitions. An example ‘DIA_umpire.se.params’ is available in the repository on github.

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.
 - MS-GF+ and X! Tandem searches are set to automatically convert the results to .pep.xml after completion, whereas the comet.params file should specify the output as .pep.xml

MSGF+ search: If using the example SWATH files from MASSIVE, under MSGF+ enter ntt = 1, enzyme= 1, ti= 0,1, ppm = 25ppm, fasta file = “20150810.mouse.cc.iRT.fasta”, and mods.txt file = “tutorial_MSGFmods.txt” from the github repository.

X! tandem search: DIA-Pipe looks for specific lines of text in the X! Tandem Parameters file. **Do not edit the input and output file locations within the tandem.params file.** You must edit two lines minimum: (1) the path to your scoring parameter file (e.g. kscore downloaded with TPP), and (2) the full path to your taxonomy.xml file containing the database location. If using the example SWATH files from MASSIVE, use the params.xml file “xTandem_Kac_params.xml” and the taxonomy.xml file from github. 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”.

COMET search: If using the example SWATH files from MASSIVE, use the “comet.Kac.DIA.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. If using the tutorial data, the defaults under PTMProphetParser, except for the path to PTMProphetParser, are appropriate settings. Otherwise, enter the mass corresponding to your modifications of interest, fragment mass tolerance appropriate for your instrument, and the minimum iprobability used to consider peptides for site-localization scoring (0.9 should be safe).

Skyline Signal Extraction and report generation

1. Skyline Template document, ****NOTE, template will be overwritten with completed import****:
 - a. Tutorial 5600 or Orbitrap data: Open Skyline and create a template Copy the appropriate skyline template document to your output directory, “..\DIA-Pipe\skyline\default_empty.sky” for data from ABSciex 5600, or “..\DIA-Pipe\skyline\orbi_empty.sky”
 - b. User datasets: Open Skyline and create a new blank document containing the transition settings you want to use for signal extraction. Add the report “2016_0826_mapDIA.skyr” to your document report list. Save the empty document to the output directory containing your .mzXML files.
3. Copy the skyline report file from “..\DIA-Pipe\skyline\2016_0826_mapDIA.skyr” to your output directory
4. Under the Skyline section, if using the tutorial files from MASSIVE, use the skyline report name “2016_0826_mapDIA,” and the skyline template document titled “default_empty.sky” from the github page.
5. If using the tutorial data accessible from Massive, use the .fasta file ‘20150810.mouse.cc.iRT.fasta’ for the skyline .fasta file.

mapDIA

notes: your directory must contain a tab-delimited text file named “filemapping.txt,” that includes at minimum one rows where the first row is 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. This is because of the mapDIA requirement that the input parameters file and the input data file have the same order.

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:

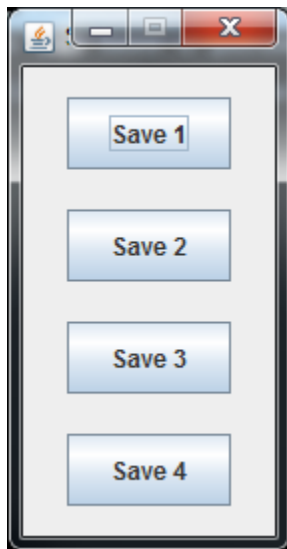
The screenshot shows a window titled "CheckBox Matrix" with the following elements:

- mapDIA Input = [text box]
- Skyline Report = [text box]
- Enter folder with .wiff files [text box]
- NORMALIZATION = ☐ None ☐ RT ☐ TIC
- MASSES = [text box]
- Min PTM Prophet Score = [text box]
- PTM Prophet Report = [text box]
- Below the PTM Prophet Report field, there is a 2x2 grid of checkboxes:
 - ko wt
 - ko ☐ ☐
 - wt ☐ ☐
- A "Finished" button at the bottom.

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.

6. If only running mapDIA (not the previous steps in DIA-Pipe), enter the mapDIA input file path and leave the Skyline Report box blank.
7. If running mapDIA using results from previous modules, 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.

8. 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)
9. Chose the type of normalization (see mapDIA documentation). If using the example SWATH files from MASSIVE, choose “None.”
10. Enter the integer mass of the modification you want to quantify (enter 42 for acetylation if using the example SWATH files from MASSIVE).
11. Enter the minimum localization score you are willing to accept for your quantified results (0.75 suggested).
12. 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”.
13. 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.
14. Once everything is correctly filled out, click “finished” to return to the main window.
15. Save the parameters you entered (in case of an error) by clicking the save button and choosing one of the save slots:



16. 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.
17. 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.