Manual for QuantPipe

1. The requirements for QuantPipe running on Windows.

1.1 Download and install Microsoft .NET Framework 4.5 or newer on Microsoft website; if you have installed .NET framework, go to the next step.

1.2 Download and install Java 8.0 or newer at www.java.com; if you have installed java 8.0+, go to the next step.

1.3 Download and install Python 3.6 or newer.

1.3.1 From the Anaconda download site (https://www.anaconda.com/), click 'Download' and select the latest Python version (3.6 or higher) and launch the installer.

1.3.2 Follow the prompts in the graphical installer. Note the install location that you choose and complete the installation. We do not recommend adding Anaconda to your PATH environment variable, but you can choose to register Anaconda as your default Python.

1.3.3 From the start menu, search for "Anaconda Prompt" and launch it.

1.3.4 In the Anaconda Prompt window that opens, type "pip install msproteomicstools" and hit enter to install the package.

1.3.5 In the Anaconda Prompt window that opens, type "pip install pyprophet" and hit enter to install the package.

1.4 Download and install OpenMS 2.5 at https://github.com/OpenMS/OpenMS/releases/tag/Release2.5.0.

1.5 Download the latest QuantPipe at https://github.com/tachengxmu/QuantPipe/releases.

1.6 Unzip the QuantPipe package into OpenMS installation folder "bin", and four additional files, namely "QuantPipe", "Lib2tsv", "TricParser.class" and "UpdateReport.class" should appear in the "bin" folder.

1.7 Double click "QuantPipe" icon to run QuantPipe software.

2. Files required for DIA data analysis using QuantPipe.

2.1 QuantPipe needs iRT file, UniMod file, the spectral library, database, Python and swathwindow files and profile mzML/mzXML files to run. Please get these files ready before running QuantPipe.

2.2 iRT and UniMod files have been provided in the example folder. iRT files contain all endogenous ciRT peptides and 11 iRT peptides (Biognosys), and you can custom the iRT files by adding the peptides specific for your experiments. Unimod.xml file for non-PTM DIA data and Phos.Unimod.xml for phospho-DIA data have been provided. You can also custom the PTM.Unimod.xml for specific-PTM DIA data analysis.

2.3 The spectral library can be generated using several tools. FragPipe and Skyline can be used for building of the spectral library (https://github.com/Nesvilab/FragPipe and https://skyline.ms/project/home/software/Skyline/begin.view).

2.4 The database file is the fasta file that was used in the database searches without any decoy or reverse sequences.

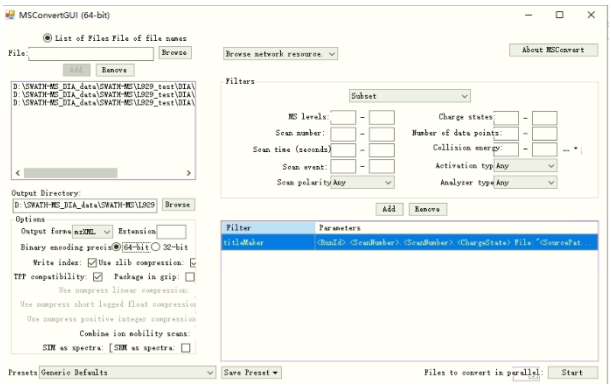
2.5 The swathwindow file should be a no overlap window file for OpenSWATH extraction. A DIA window example file is provided in the example folder.

3. The tutorial of running QuantPipe.

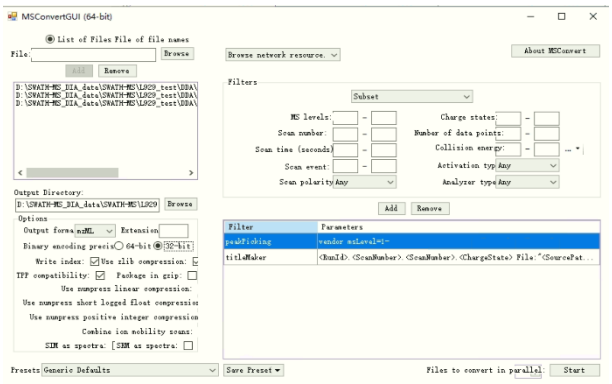
3.1 We analyze SWATH-MS to illustrate the usage of QuantPipe. Download the tutorial data at http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD021390, and raw data included three technical SWATH-MS runs and three technical DDA runs from murine L929 cell lysate sample.

3.2 The spectral library was generated using FragPipe (Note: DDA files were used for generation of the external library. SWATH-MS files were first analyzed with DIA-Umpire to generate pseudo-DDA files, which can be used to build the internal spectral library. DDA files and pseudo-DDA files were altogether searched with FragPipe for the generation of the combined spectral library, which can provide the most in-depth coverage of DIA analysis. However, you can build the internal library or the external library alone. The steps of the generation of the combined spectral library are shown below).

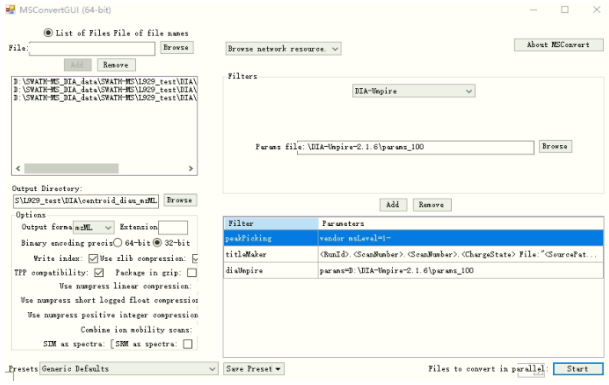
3.2.1 Convert SWATH-MS wiff files to profile mzXML using Protowizard MSConvert (V.3.0.20149) (http://proteowizard.sourceforge.net/download.html) for OpenSWATH analysis. The settings of MSConvert are shown below.



3.2.2 Convert DDA wiff files to centroid mzML.



3.2.3 Convert SWATH-MS wiff files to pseudo-DDA mzML files, and the DIA-Umpire params file can be downloaded in the tutorial folder.



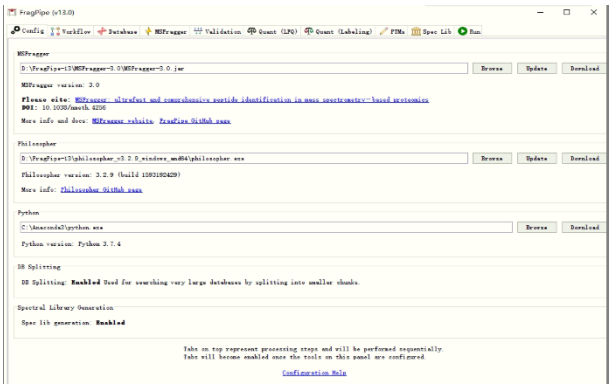
3.2.4 Because there are no 11 iRT peptides (Biognosys) in our SWATH-MS samples, we must modify the Python script to use endogenous ciRT peptides. Once you have set the ciRT peptides in the Python script, then you do not need to spike commercial iRT peptides into your samples from human, murine or yeast species.

3.2.4.1 Go to C:\Anaconda3\Scripts, and backup the original spectrast2spectast_irt.py file. Copy spectrast2spectrast_irt.py file provided in the tutorial folder into this folder.

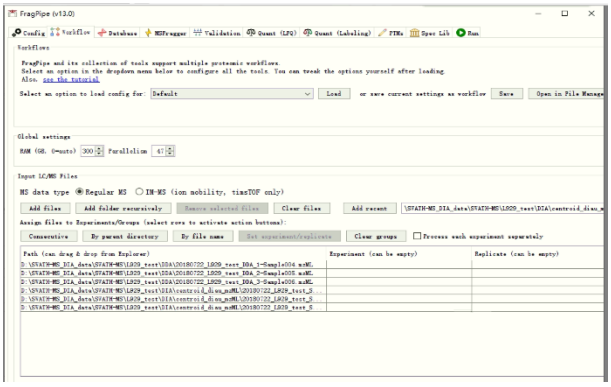
3.2.4.2 Open spectrast2spectrast_irt.py with notepad++. Search “rtkit={“, and the screenshot is shown below.

3.2.5 Perform the database searches and the spectral library generation using FragPipe (V.13).

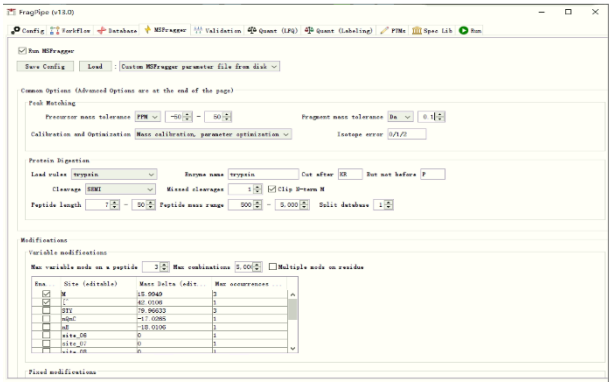
3.2.5.1 The Configs tab of FragPipe is shown below.



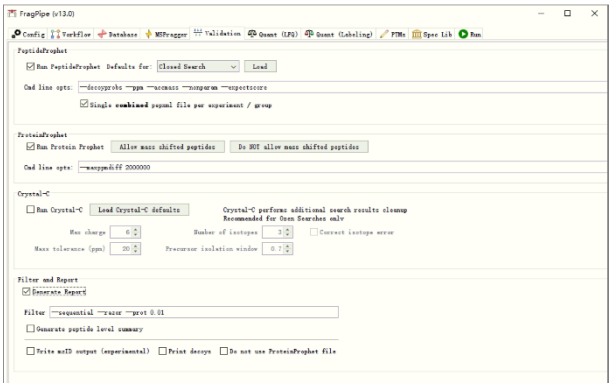
3.2.5.2 DDA and pseudo-DDA mzML files are input into the FragPipe. The Workflow tab of FragPipe is shown below.



3.2.5.3 The MSFragger tab is shown below.

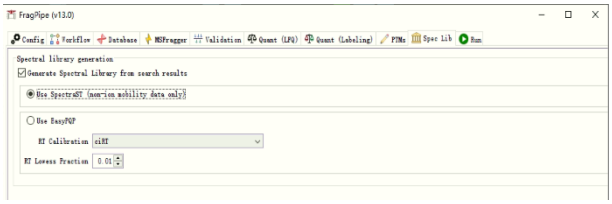


3.2.5.4 The Validation tab is shown below.

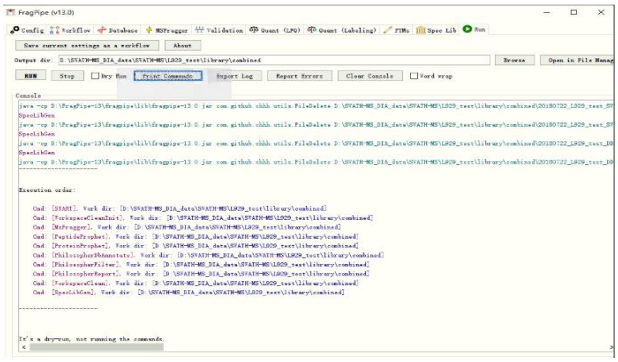


3.2.5.5 The Quant (LFQ), Quant (Labeling) and PTMs are disabled.

3.2.5.6 The Spec Lib tab is shown below.



3.2.5.7 Specify the output path, and Run.



3.2.5.8 The spectral library (output_file_irt_con.splib) is in the output folder. Rename *.splib file to *.sptxt file.

3.3 Perform a targeted analysis of SWATH-MS using QuantPipe.

3.3.1 Double click “iRT” textbox to select “iRT” file. iRT file should contain the iRT peptides that were spiked into DIA samples. An iRT example file is provided along with the installation files, which includes all 11 iRT peptides and ciRT peptides. This iRT example file should be applicable for most DIA data from human, murine or yeast samples.

3.3.2 Double click “unimod” textbox to select “unimod” file, an unimod example file is provided. This unimod is used for “lib2tsv” to build the assay library from the spectral library. The checkbox beside “IPF” is used for IPF function.

3.3.3 Double click “Library” textbox to select “library” file, suffix name of which can be “sptxt” or “blib” or “tsv”. If you got “splib” files from FragPipe, you can rename the suffix name of “splib” to “sptxt”. “blib” files are generated by Skyline, and “tsv” files are the assay libraries.

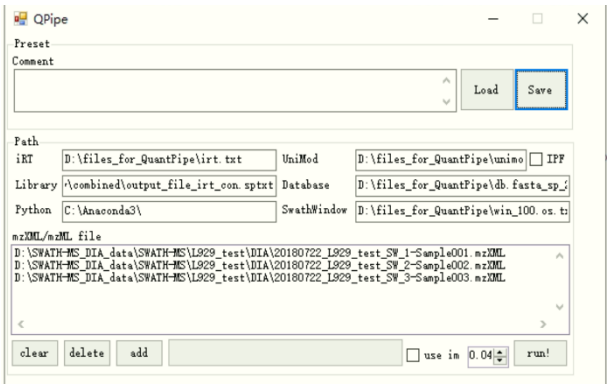
3.3.4 Double click “database” textbox to select “database” file, which should be the fasta file that was used in the database searches without any decoy or reverse sequences.

3.3.5 Double click “Python” textbox to select the path of Python program located at.

3.3.6 Double click “Swathwindow” textbox to select “DIA window” file used in DIA MS data acquisition. A DIA window example file is provided, which should be a no overlap window file for OpenSWATH extraction. The Swathwindow file for these tutorial data is provided in the tutorial folder.

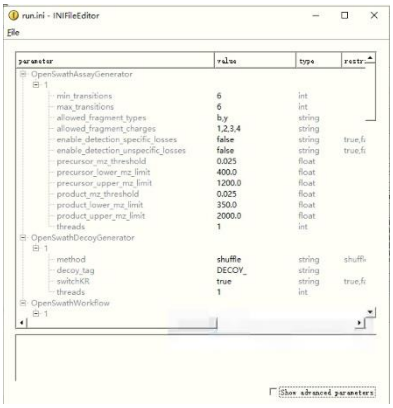
3.3.7 Click “add” to add profile “mzXML” files.

3.3.8 The screenshot of QuantPipe is shown below.



3.3.9 Click “Save” to save all settings to a cfg file. Note: the folder the cgf file is located will be the working directory, and all intermediate files will be saved in this folder.

3.3.10 Click “run” to open the parameter editing window, which allows user to edit the workflow parameters including “rt_extraction_window”, “mz_extraction_window” and “the number of threads allowed to be used by the TOPP tool” etc. In most cases, you can run the workflow with the default parameters except for “rt_extraction_window”. After all parameters were set, close the window by clicking “close” icon on the top right corner. Then click “OK” to save the modified parameters. The screenshot is shown below.



3.3.11 The “OpenSWATH-PyProphet-TRIC” workflow start running. The first step is converting the spectral library (splib and blib) to the assay library. The folder named “log” is used for storage of log files, which can be examined for which step is being executed and what errors happened.

3.3.12 When all steps are accomplished, the file “aligned.tsv_quant.pep.csv” contains all quantified peptides with their intensities in each run, and the file “aligned.tsv_quant.prot.csv” contains all quantified proteins with their intensities in each run. The peptides and proteins are filtered at 1% global protein FDR by default. The protein intensities are calculated by summing the top three peptide intensities.