Jun 18, 2025 Version 1

DIA data analysis in DIA-NN against a DDA library generated in FragPipe V.1

DOI
https://dx.doi.org/10.17504/protocols.io.eq2ly642pgx9/v1

Geremy Clair[1], Ernesto S Nakayasu[1], WEI-JUN QIAN[1]

[1]Pacific Northwest National Laboratory

Human BioMolecular Atl... Pacific Northwest Natio...

Geremy Clair
Pacific Northwest National Laboratory

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN ACCESS

DOI: https://dx.doi.org/10.17504/protocols.io.eq2ly642pgx9/v1

Protocol Citation: Geremy Clair, Ernesto S Nakayasu, WEI-JUN QIAN 2025. DIA data analysis in DIA-NN against a DDA library generated in FragPipe. **protocols.io** https://dx.doi.org/10.17504/protocols.io.eq2ly642pgx9/v1

License: This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working
We use this protocol and it's working

Created: February 13, 2025

Last Modified: June 18, 2025

Protocol Integer ID: 120147

Keywords: Data Independent Acquisition, DIA, Library, DDA library, Fragpipe, DIA-NN, dia data analysis in dia, dia data analysis, dda library, using dia, dia search, fragpipe this protocol, dia, fragpipe, protocol, data

Funders Acknowledgements:
LungMAP2 PNNL (NIH U01)
Grant ID: U01-HL148860
HubMAP Lung (NIH U54)
Grant ID: U54-HL165443
HubMAP Kidney (NIH U54)
Grant ID: U54-DK127823

Disclaimer

> While all softwares used for this protocol are free of use, copyrights and citations to the software developers are critical for their visibility.

Abstract

This protocols details how we typically perform DIA search using DIA-NN against a DDA library,

Guidelines

This is a step-by-step description of our procedure to perform "DIA data analysis in DIA-NN against a DDA library generated in FragPipe"

Materials

We performed all the steps of this protocol on a Windows 11 computer.

Before start

Before starting, ensure that all the software necessary are downloaded and installed on your computer.
links to download each software are provided as first step of each section.

The procedure detailled was performed successfully on Thermo Orbitrap and Astral instrument (*.raw files) and on Bruker TimsTOF (d folders).

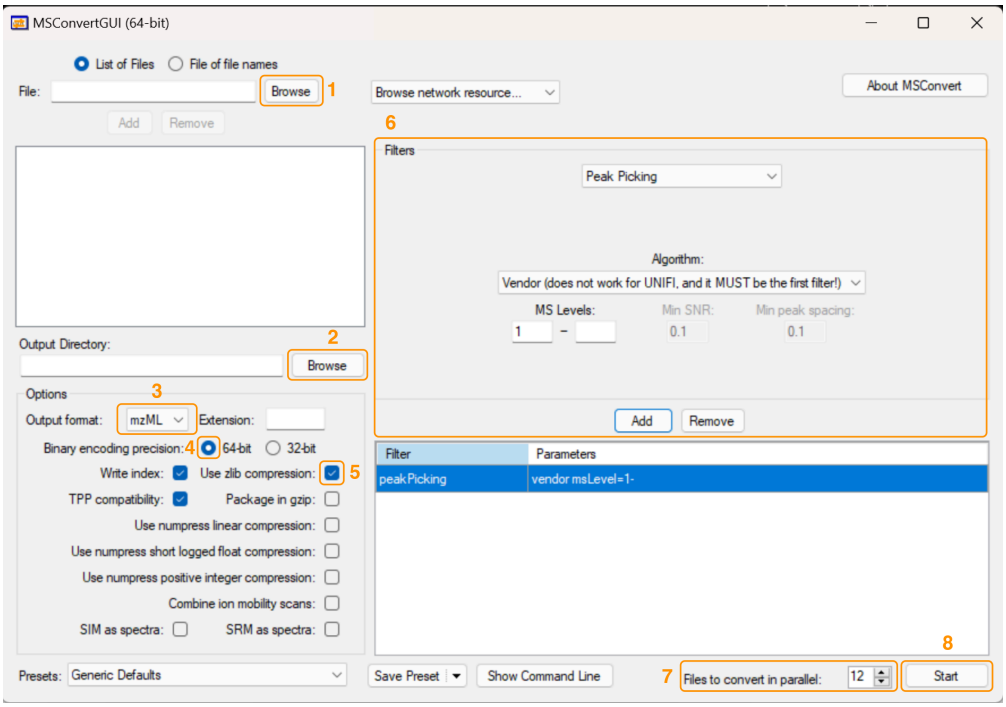
Note that if a FAIMS source is used with multiple a raw file for each CV for the DDA acquisition multiple MzML files will need to be used for each run. We have not tested the use of a FAIMS source for the generation of DIA runs.

Converting the raw files into MzML files

1 Download the raw files on your computer.
Both the DDA library files and DIA files from different instruments and instrument supplier can be converted using this method.

2 Download MS-Convert on your computer (https://proteowizard.sourceforge.io/download.html).

3 Convert the files to the MzML 64-bits zlib compression format with the peak picking filter activated using the following sub-steps (note that the substep numbers correspond to the numbers indicated on the image below):



MSConvert steps for the raw -> mzML conversion

3.1 Press the "File: Browse" button.
In the browser, select the files to be added and press "Open".

3.2 Press the "Output Directory: Browse button".
In the browser, select your directory and press "OK".

3.3 Ensure that "mzML" is selected as "Output format".

3.4 Ensure that "64-bit" is selected as "Binary encoding precision".

3.5 Ensure that "Use zlib compression" is selected.

3.6 In the "Filter" section select Peak Picking with the default parameters (see image) and press "Add".

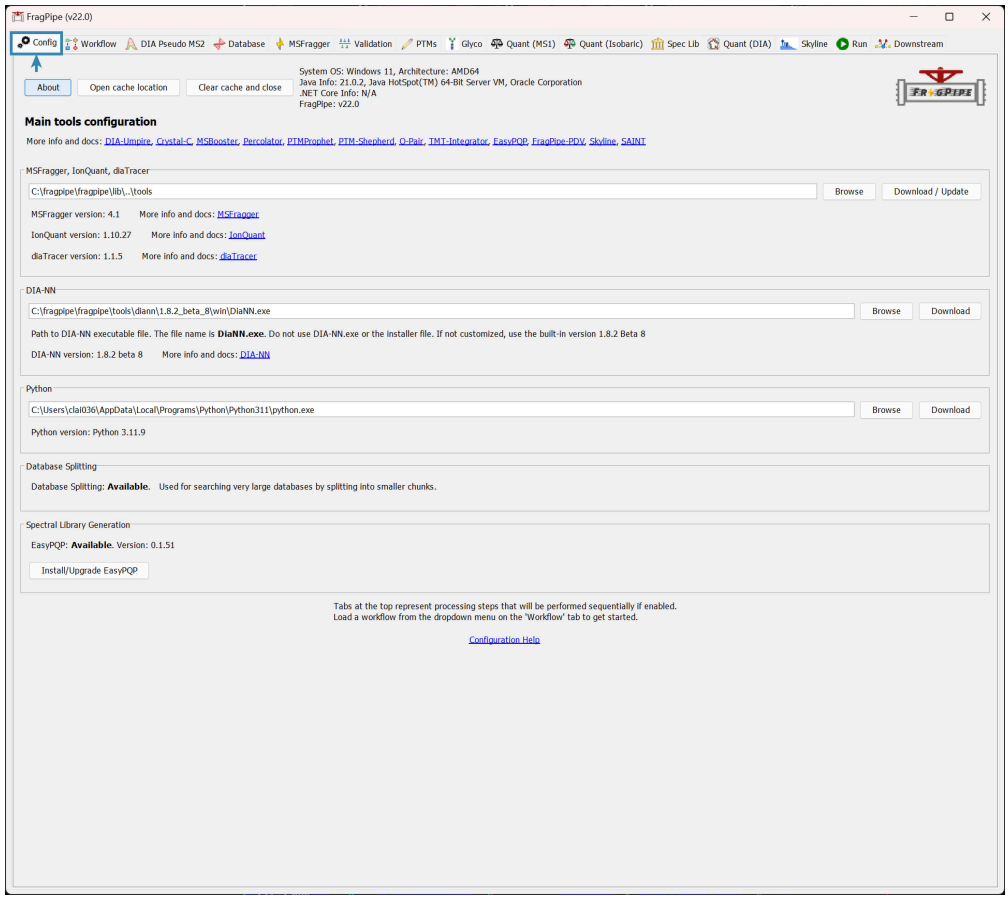
3.7 Select the number of "Files to convert in parallel", this number has to be lower than the number of "Logical processors" of your system.

3.8 Press "Start".

Generating a DDA library using FragPipe

4 Download fragpipe on your computer (**https://fragpipe.nesvilab.org/**)

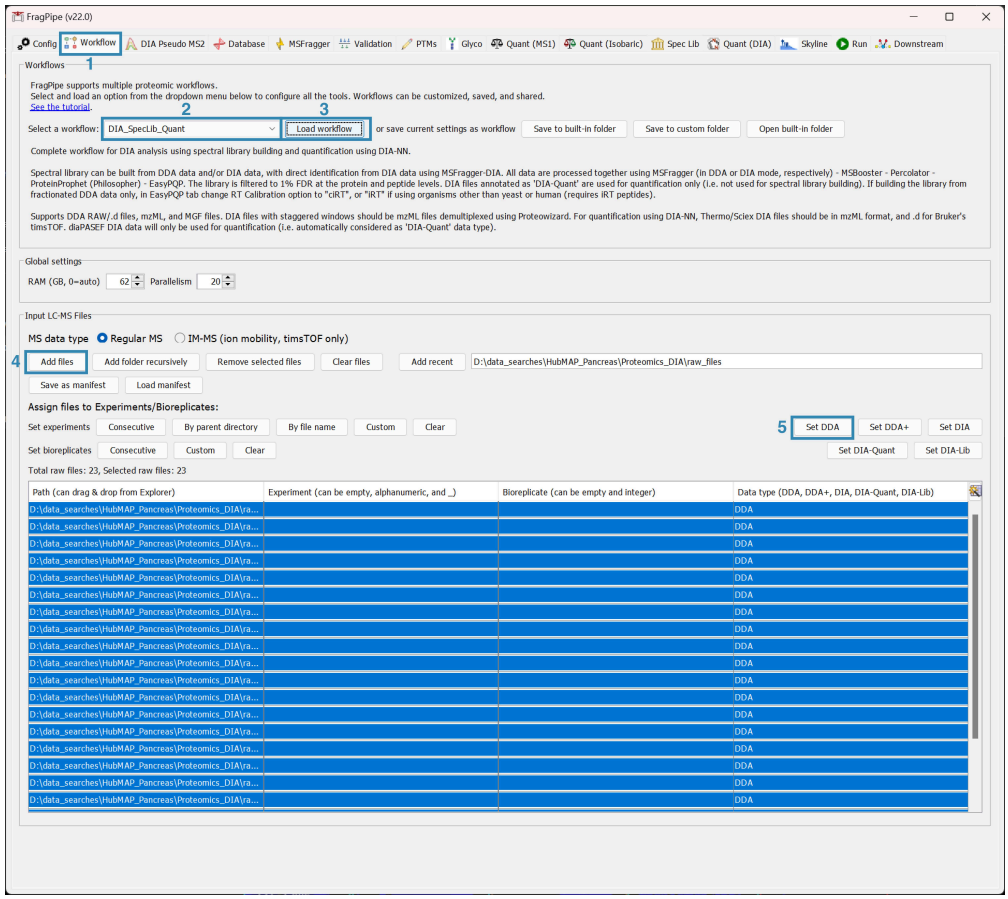
5 Open fragpipe using the fragpipe application located in the "fragpipe/bin" folder.
In the "Config" tab, ensure that all fragpipe components are up-to-date.



6 In this step, the objective is to create a "library" for the DIA-NN search using fragpipe, the reason for choosing this option is the enhanced level of certainty on the existence of the proteins in the samples over the library-free/fasta-library method in DIA-NN.

The steps below describe this process:

6.1 Go to the "Workflow" tab.



6.2 In the "Select a workflow" drop down menu, select DIA_SpectLib_Quant.

6.3 Press the button "Load workflow".

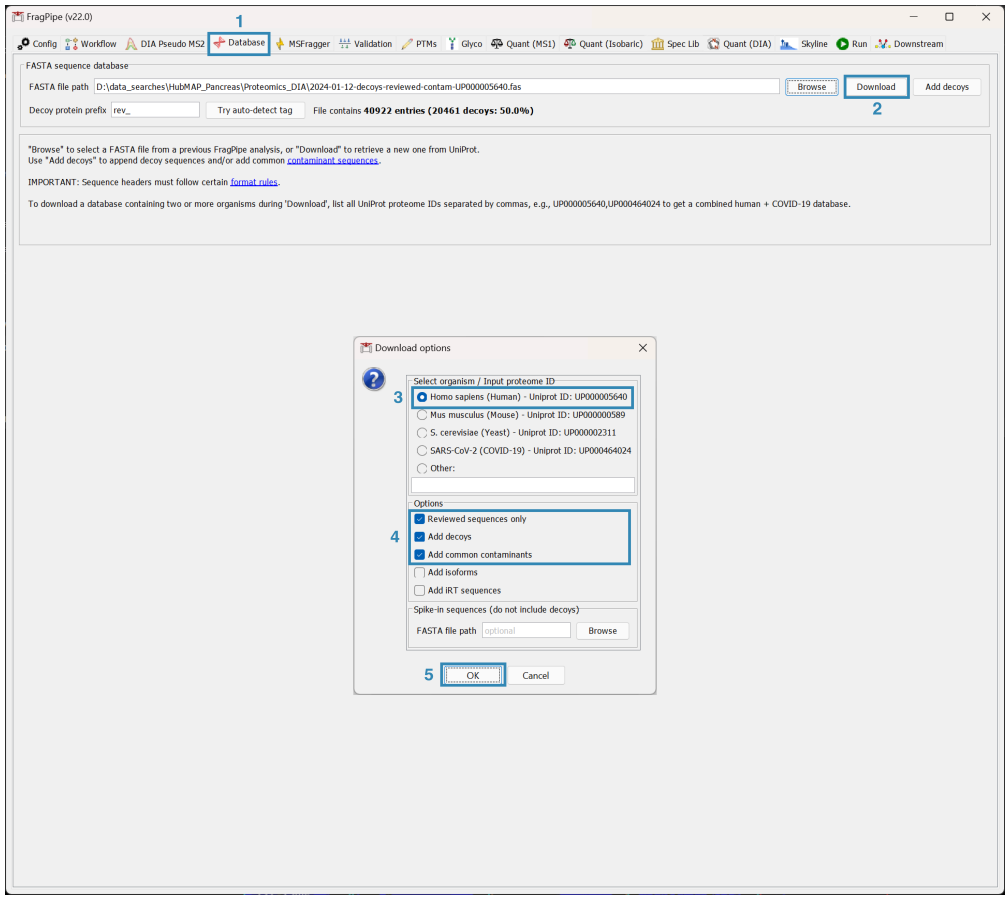
6.4 Press the "Add files" button, in the browser select the mzML files generated with MSConvert for the DDA library runs and press the "Select" button.

(note: we typically use a pooled sample with equivalent amount of peptides from each sample, then we fractionate this sample using LC-MS/MS into 96 fractions, that are then either recombined into 12 or 24 fractions. the recombination scheme is done to have representation of different retention times.)

6.5 The Data type should automatically being detected to DDA, if not, select all the files and press the button Set DDA.

7 Next we will use FragPipe to download a fasta file.
Note: a fasta file can also be prepared in advanced and loaded here instead of the steps described below, in this case the decoy and contaminants can be added using FragPipe.

7.1 Open the "Database" tab.



7.2 Press the "Download" button.

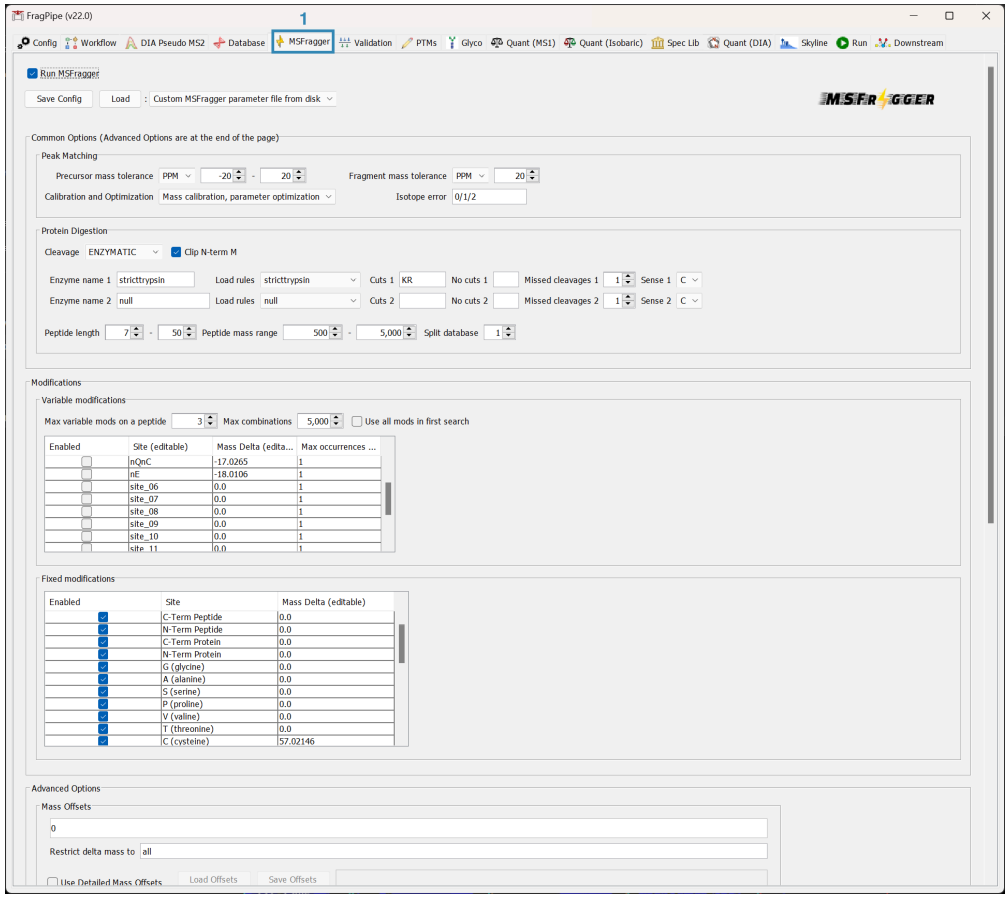
7.3 Select your organism or enter the Uniprot proteome number (typically starting with "UP")

7.4 Depending on your need tick or untick the boxes "Reviewed sequences only", "Add decoys", "Add common contaminants"

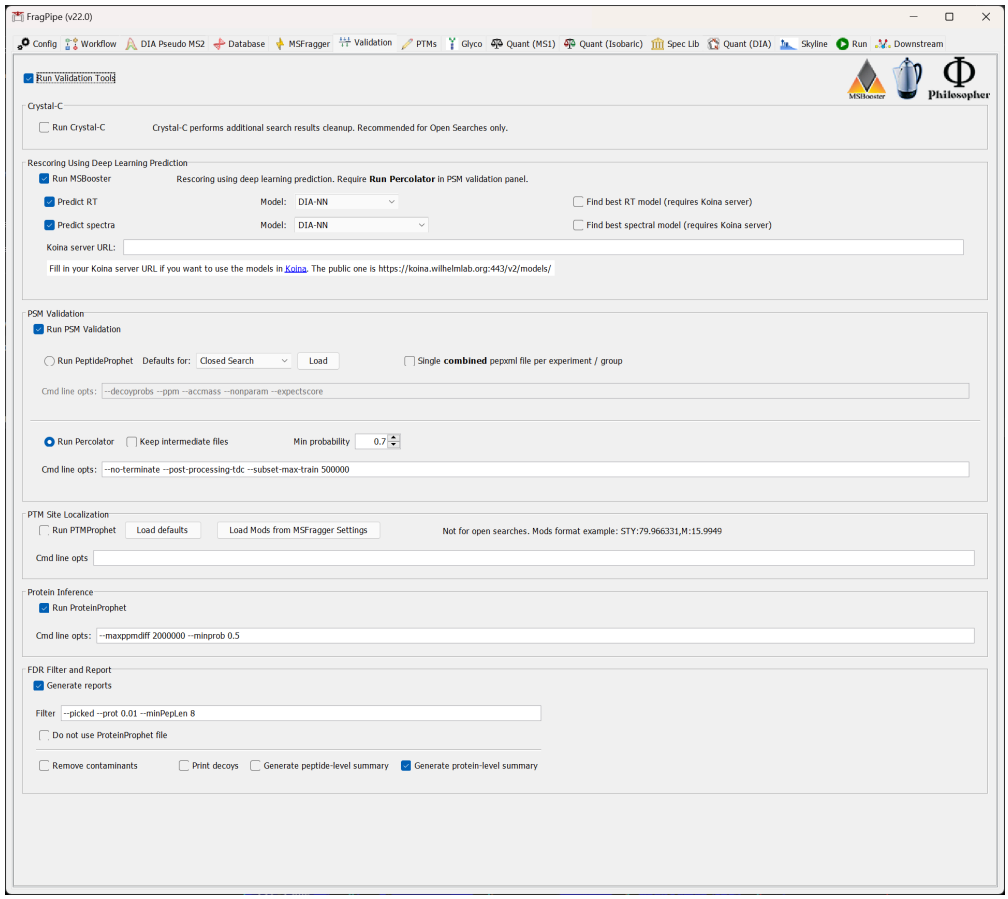
7.5 Press "OK" to start the download.

(note: the program might take a little time to download the sequences, please be patient!)

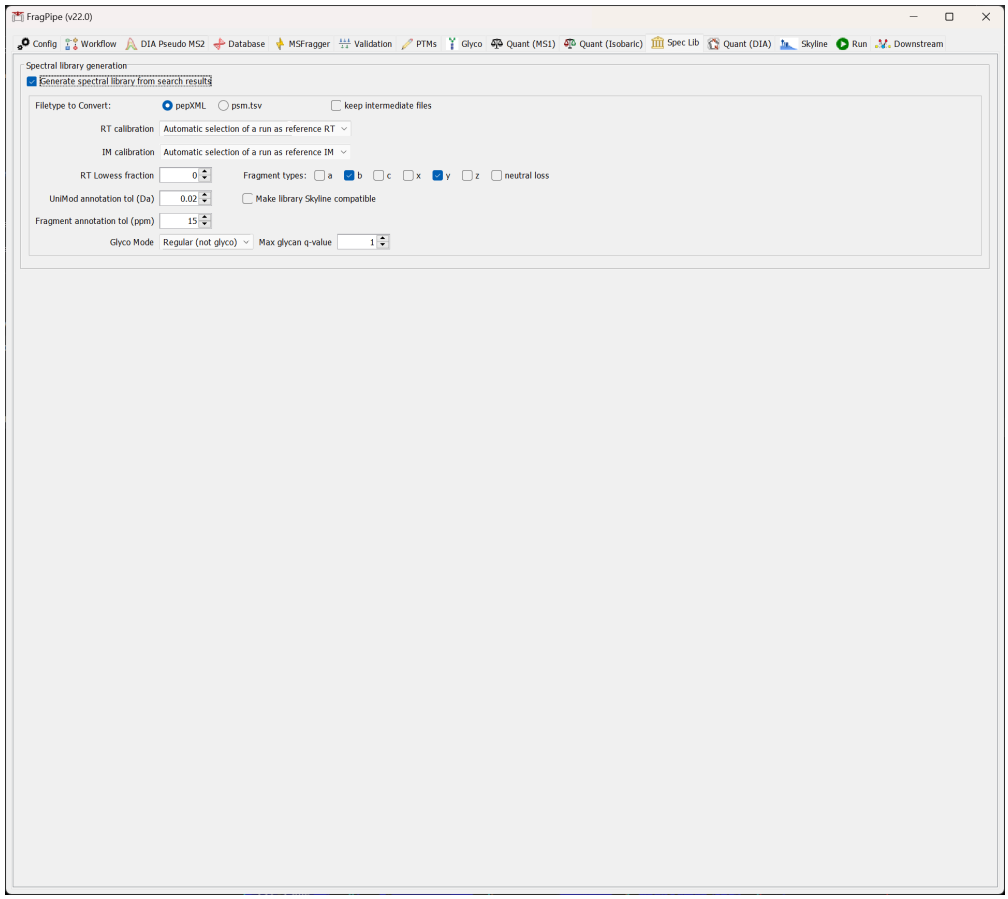
8 Go to the "MSFragger" tab
In this tab, you will be able to set up the search parameters, this will determine what search parameters will be used to create the peptide spectrum matches (PSM), peptides, and protein IDs that will be recorded in the library file generated.
(Note : We recommend to select parameters that are appropriate based on the way the samples were prepared)



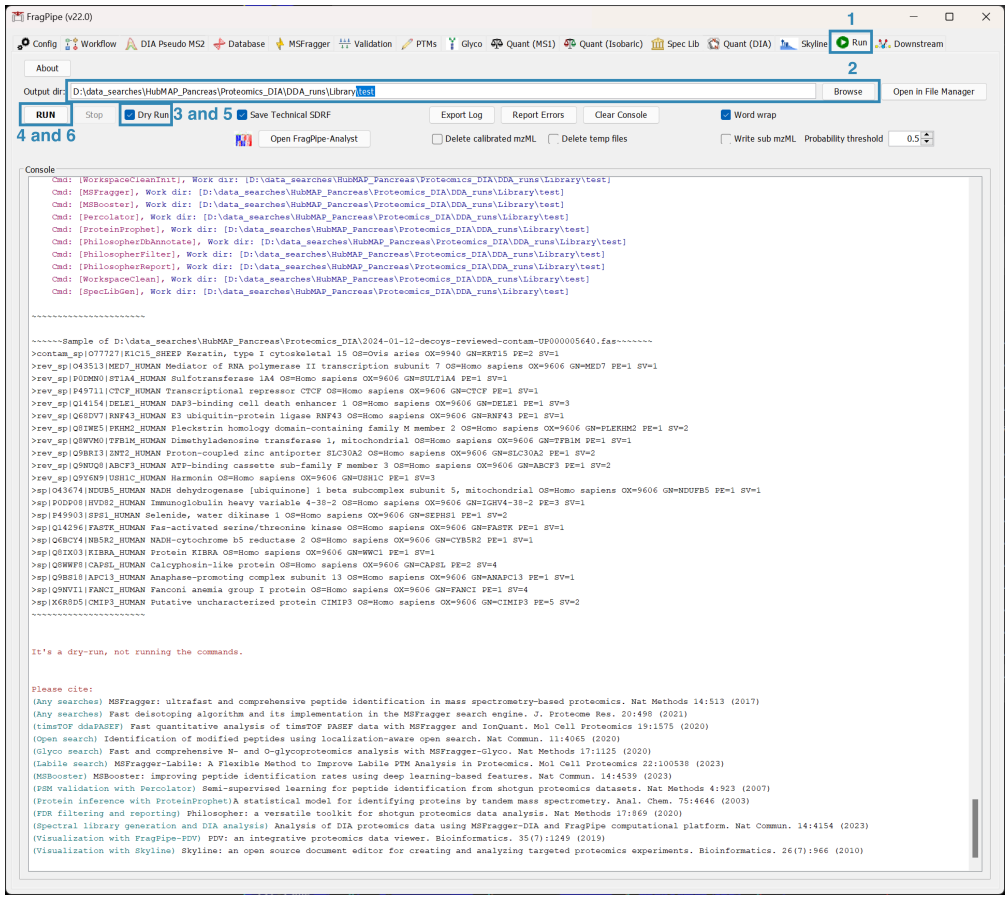
9 As needed, check that "Run the validation tools" is ticked the "Validation" tab.
for the Run MSBooster box, we recommend to tuse the default preset with "DIA-NN" selected as model for the predict RT and predict spectra.



10 Verify that the "Generate spectral library from search results" is ticked in the "Spec Lib" tab.



11 We will now run the search and generate the library.



11.1 Go to the "Run" tab.

11.2 Click on "Browse" to select where you'd like to save the library on your computer.

11.3 Tick the "Dry Run" box, this step is to ensure that there is no file conflicts in the destination folder.

11.4 Click on "Run" and verify that there is no error message

11.5 Untick the "Dry Run" box, now the search will be performed.

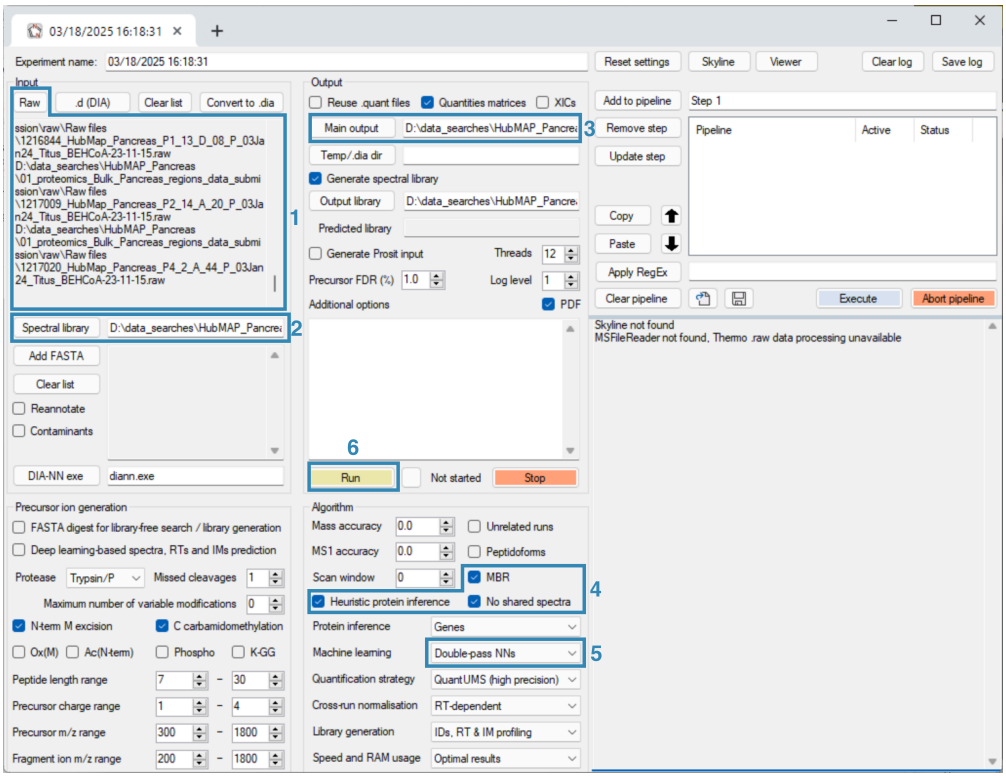
11.6 Click on "Run", when the search is finished, a tsv file named "library" will be generated in the destination folder. This file will be used as library in DIA-NN.

Performing the DIA analysis using DIA-NN

12 Download DIA-NN on your computer (follow the instructions on https://github.com/vdemichev/DiaNN).

13 Open DIA-NN.

14 Perform the DIA-NN analysis using the following substeps.



14.1 In the input section, click on the "Raw" button, then in the pop-up window, select the DIA raw or MzML files to be searched against the Fragpipe generated library.

14.2 In the input section, click on the "Spectral library" button, then in the pop-up window, select the "library.tsv" file generated by Fragpipe.

14.3 In the Output section, click on the "Main output" button, then in the pop-up window, select the target folder for the files generated.

14.4 In the Algorithm section tick the boxes "MBR" (Match Between Runs), "Heuristic protein inference", and "No shared spectra".

14.5 Select "Double-pass NNs for the "Machine learning" drop box.

14.6 Press the "Run" button.

Protocol references

Adusumilli R, Mallick P. **Data Conversion with ProteoWizard msConvert.** Methods Mol Biol. 2017;1550:339-368. doi: 10.1007/978-1-4939-6747-6_23. PMID: 28188540.

Yu F, Teo GC, Kong AT, Fröhlich K, Li GX, Demichev V, Nesvizhskii AI. **Analysis of DIA proteomics data using MSFragger-DIA and FragPipe computational platform.**Nat Commun. 2023 Jul 12;14(1):4154. doi: 10.1038/s41467-023-39869-5. PMID: 37438352; PMCID: PMC10338508.

Demichev V, Messner CB, Vernardis SI, Lilley KS, Ralser M. **DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput.** Nat Methods. 2020 Jan;17(1):41-44. Epub 2019 Nov 25. PMID: 31768060; PMCID: PMC6949130.