

Peptide detection with Scribe

Tutorial based on Scribe version 2.11.2, last update on November 14, 2022

Written by Brian Searle (brian.searle@osumc.edu)

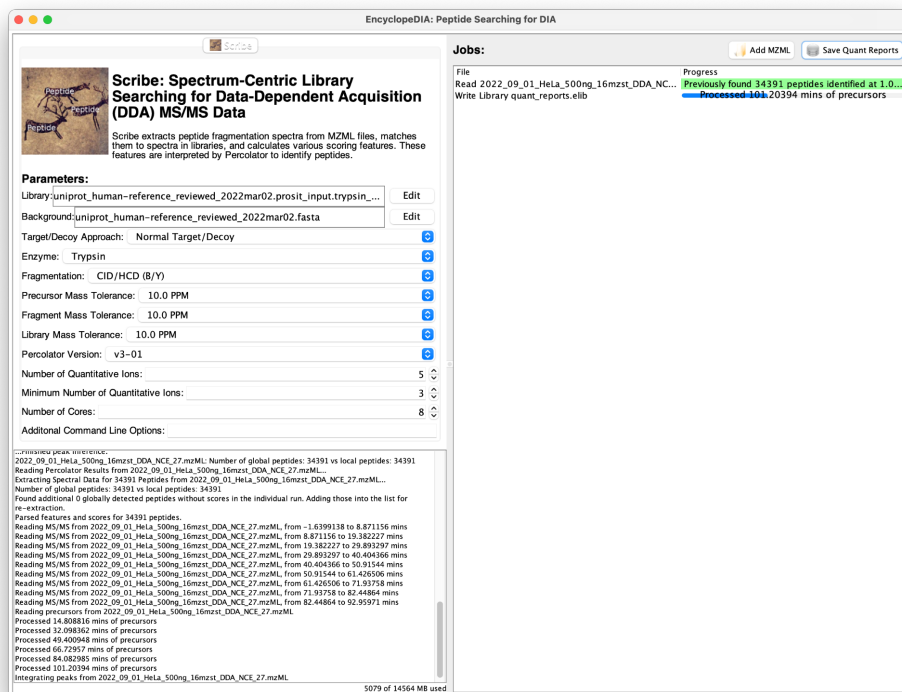
Spectrum library searching is a powerful alternative to database searching for data dependent acquisition experiments. Scribe is a new library search engine designed to leverage deep learning fragmentation prediction software such as Prosit. Scribe embeds Percolator for FDR correction and an interference tolerant label-free quantification integrator to enable an end-to-end proteomics workflow.

PREREQUISITES

Scribe is a cross-platform Java application that has been tested for Windows, Macintosh, and Linux. Scribe requires 64-bit Java 1.8. If you don't already have it, if you are using Windows, you can download the Windows "x64 Installer" from:

<https://www.oracle.com/java/technologies/downloads/#java8>

Other operating system options are available as well. After you have 64-bit Java 1.8, double click on the Scribe .JAR file to launch the GUI interface. If you are using a Macintosh, you may need to right click on the Scribe .JAR and select "Open" to execute it for the first time with the proper permissions.



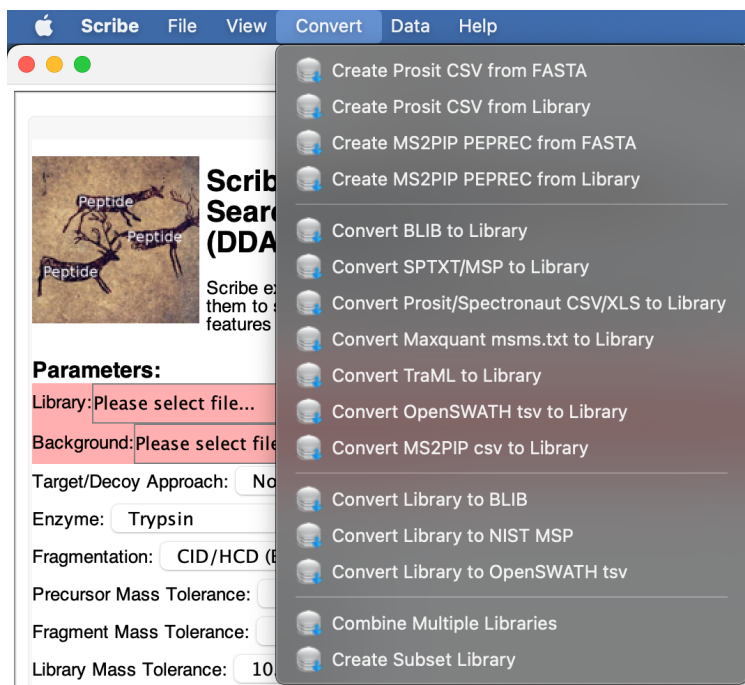
OVERVIEW

Scribe comes with a user-friendly GUI interface. The upper left pane contains search options, while the right pane contains a process queue. The bottom left contains a console that provides specific information about what EncyclopeDIA is doing at any given moment.

GENERATING MZMLS

We recommend using Proteowizard to create mzML files from your RAW files. You can freely download Proteowizard from here: <https://proteowizard.sourceforge.io/download.html>

GENERATING LIBRARIES



The primary library format Scribe uses is .DLIB. Spectrum libraries from Skyline, NIST, TraML and other formats can be converted to .DLIB using the “Convert” menu. In particular, Scribe can search fully predicted libraries generated by either Prosit or MS2PIP, and can produce input files for both of those tools. If you are using Prosit, we recommend saving Prosit results as NIST MSP files and converting the MSPs to DLIBs using Scribe.

SCRIBE SEARCH OPTIONS

Here is a screenshot of the Scribe default search options:

| Parameters: | |
|--|------|
| Library: Please select file... | Edit |
| Background: Please select file... | Edit |
| Target/Decoy Approach: Normal Target/Decoy | ⌵ |
| Enzyme: Trypsin | ⌵ |
| Fragmentation: CID/HCD (B/Y) | ⌵ |
| Precursor Mass Tolerance: 10.0 PPM | ⌵ |
| Fragment Mass Tolerance: 10.0 PPM | ⌵ |
| Library Mass Tolerance: 10.0 PPM | ⌵ |
| Percolator Version: v3-01 | ⌵ |
| Number of Quantitative Ions: 5 | ⌵ |
| Minimum Number of Quantitative Ions: 3 | ⌵ |
| Number of Cores: 8 | ⌵ |
| Additional Command Line Options: | |

Scribe has several options for searching files. Before you can start loading data, you need to specify both a DLIB or ELIB library to search as well as a background FASTA. These will be shaded red until they are properly specified.

Libraries can be either in the .ELIB (chromatogram library) or .DLIB (spectrum library) format.

Scribe has several other search settings. As a general rule, we recommend using the default search parameters at first. Other settings are defined below:

Target/Decoy Approach: In some circumstances it may be necessary to add additional decoys to improve statistical analysis. However, as a general rule this should be left at “Normal Target/Decoy”.

Enzyme: Several common digestion enzymes are supported.

Fragmentation: In general, we recommend using CID/HCD (B/Y) fragmentation for most CID or HCD experiments. However, if your library is particularly large or messy you may get improved results with “HCD (y- only)”.

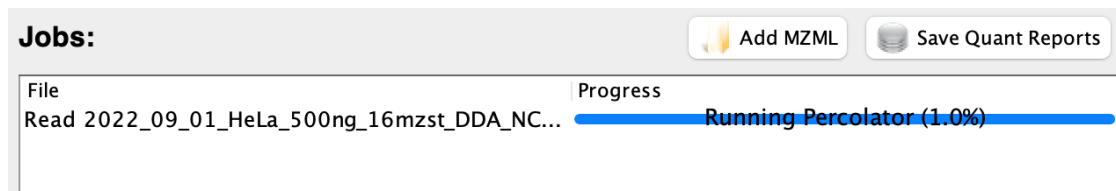
Precursor/Fragment/Library Mass Tolerance: Tolerances can be specified in PPM, AMU, or resolution. **Percolator Version:** Percolator 3.1 is recommended for most experiments.

Number of Cores: This is the number of CPU cores you allow EncyclopeDIA to use.

Additional command line options can be specified in the command line options box.

SCRIBE QUEUE

Here is a screenshot of the Scribe search queue:



Use the “Add MZML” button to add raw files for library searching. These will be automatically placed on the queue and executed in order using the current settings. If an MZML has been previously analyzed, Scribe will remember where it left off and try to not process it a second time.

Search results can also be saved for quantitative assessment using the “Save Quant Reports” button. Raw files between these experiments are expected to contain shared peptides so retention-time alignment is performed and match-between-runs quantification is calculated.

WALKTHROUGH WITH DEMO DATA

You can download data used in the Scribe manuscript from:

<https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=99b781c9c0b84ac9bfc3f93734e9ddab>

To start, we recommend downloading the following files:

[2022_09_01_HeLa_500ng_16mzst_DDA_NCE_27.mzML](#)

[uniprot_human-reference_reviewed_2022mar02.fasta](#)

[uniprot_human-reference_reviewed_2022mar02.prosit_input.trypsin_nce29_hcd2020.dlib](#)

Specifying the [[uniprot_human-reference_reviewed_2022mar02.prosit_input.trypsin_nce29_hcd2020.dlib](#)] as the library and [[uniprot_human-reference_reviewed_2022mar02.fasta](#)] as the background FASTA, use these settings:

Parameters:

| | | |
|---------------------------|---|------|
| Library: | uniprot_human-reference_reviewed_2022mar02.prosit_input.trypsin_... | Edit |
| Background: | uniprot_human-reference_reviewed_2022mar02.fasta | Edit |
| Target/Decoy Approach: | Normal Target/Decoy | ⬇ |
| Enzyme: | Trypsin | ⬇ |
| Fragmentation: | CID/HCD (B/Y) | ⬇ |
| Precursor Mass Tolerance: | 10.0 PPM | ⬇ |
| Fragment Mass Tolerance: | 10.0 PPM | ⬇ |
| Library Mass Tolerance: | 10.0 PPM | ⬇ |
| Percolator Version: | v3-01 | ⬇ |

Then queue up the [2022_09_01_HeLa_500ng_16mzst_DDA_NCE_27.mzML] mzML file using the “Add MZML” button. Processing should begin immediately. Scribe will complete several steps in order: conversion of the mzML to DIA format, library search scoring, Percolator filtering, retention time alignment, Refined Percolator filtering, and then quantification. The resulting search should detect and quantify 34,391 unique peptides, matching the manuscript result from Figure 3.

DLIB Libraries are paired to HeLa mzMLs based on experimental NCE and Prosit NCE, where NCEs are tuned to the specific instrument. Here are the pairings for this experiment in Figure 3 and Figure 4:

| Experimental NCE | Experimental File (mzML) | Library File (DLIB) |
|------------------|--|--|
| 22 | 2022_09_01_HeLa_500ng_16mzst_DDA_NCE_22.mzML | uniprot_human-reference_reviewed_2022mar02.prosit_input.trypsin_nce16_hcd2020.dlib |
| 27 | 2022_09_01_HeLa_500ng_16mzst_DDA_NCE_27.mzML | uniprot_human-reference_reviewed_2022mar02.prosit_input.trypsin_nce29_hcd2020.dlib |
| 32 | 2022_09_01_HeLa_500ng_16mzst_DDA_NCE_32.mzML | uniprot_human-reference_reviewed_2022mar02.prosit_input.trypsin_nce34_hcd2020.dlib |
| 37 | 2022_09_01_HeLa_500ng_16mzst_DDA_NCE_37.mzML | uniprot_human-reference_reviewed_2022mar02.prosit_input.trypsin_nce41_hcd2020.dlib |
| 42 | 2022_09_01_HeLa_500ng_16mzst_DDA_NCE_42.mzML | uniprot_human-reference_reviewed_2022mar02.prosit_input.trypsin_nce50_hcd2020.dlib |

If you run into error messages or Scribe shuts down, please check your version of Java. Ideally you will be using Java version 1.8. We know of errors that were introduced with Java versions 17 and 18 and we recommend using Java 16 or earlier to ensure stability.