

Supplementary Note 1: DIA Data Analysis Without Spectral Libraries

This tutorial is a practical guide for how to use the Encyclopedia software suite to build chromatogram libraries for DIA-MS. We have a GUI-based workflow and also a command line workflow. We've included options for visualizing the results of the Encyclopedia analysis in Skyline or in Encyclopedia itself.

Citations

MSconvert (<https://www.nature.com/articles/nbt.2377>)

A cross-platform toolkit for mass spectrometry and proteomics. Chambers MC et al. *Nat Biotech* 30, 918-920 (2012). doi.org/10.1038/nbt.2377

Encyclopedia (<https://www.nature.com/articles/s41467-018-07454-w>)

Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry. Searle BC et al. *Nat Comm* 9, 5128 (2018). doi.org/10.1038/s41467-018-07454-w

You will need:

- MSConvert from Proteowizard: *Windows only!*
 - <http://proteowizard.sourceforge.net/download.html>
- EncyclopeDIA suite (*.jar file): *command line and cross-platform GUI*
 - <https://bitbucket.org/searlebc/encyclopedia/wiki/Home>

To exactly replicate the results here, you will also need:

- RAW data files from the tutorial HeLa dataset (MSV000084531)
 - 2018may16_hela_window_size_test_BCS_hela_narrow_1.raw
 - 2018may16_hela_window_size_test_BCS_hela_narrow_2.raw
 - 2018may16_hela_window_size_test_BCS_hela_narrow_3.raw
 - 2018may16_hela_window_size_test_BCS_hela_narrow_4.raw
 - 2018may16_hela_window_size_test_BCS_hela_narrow_5.raw
 - 2018may16_hela_window_size_test_BCS_hela_narrow_6.raw
 - 2018may16_hela_window_size_test_BCS_hela_narrow_7.raw
 - 2018may16_hela_window_size_test_BCS_hela_narrow_8.raw
 - 2018may14_hela_window_size_test_BCS_hela_wide_400_1000_1.raw
 - 2018may14_hela_window_size_test_BCS_hela_wide_400_1000_2.raw
 - 2018may14_hela_window_size_test_BCS_hela_wide_400_1000_3.raw
- FASTA of the Uniprot human reference proteome (reviewed; 20,350 entries)

SUMMARY: Three steps for DIA-MS analysis by chromatogram library

1. Convert .raw files to .mzML using MSConvert
2. Build library using Walnut or XCorDIA in EncyclopeDIA
3. Search wide-window data with library from step 2 using EncyclopeDIA

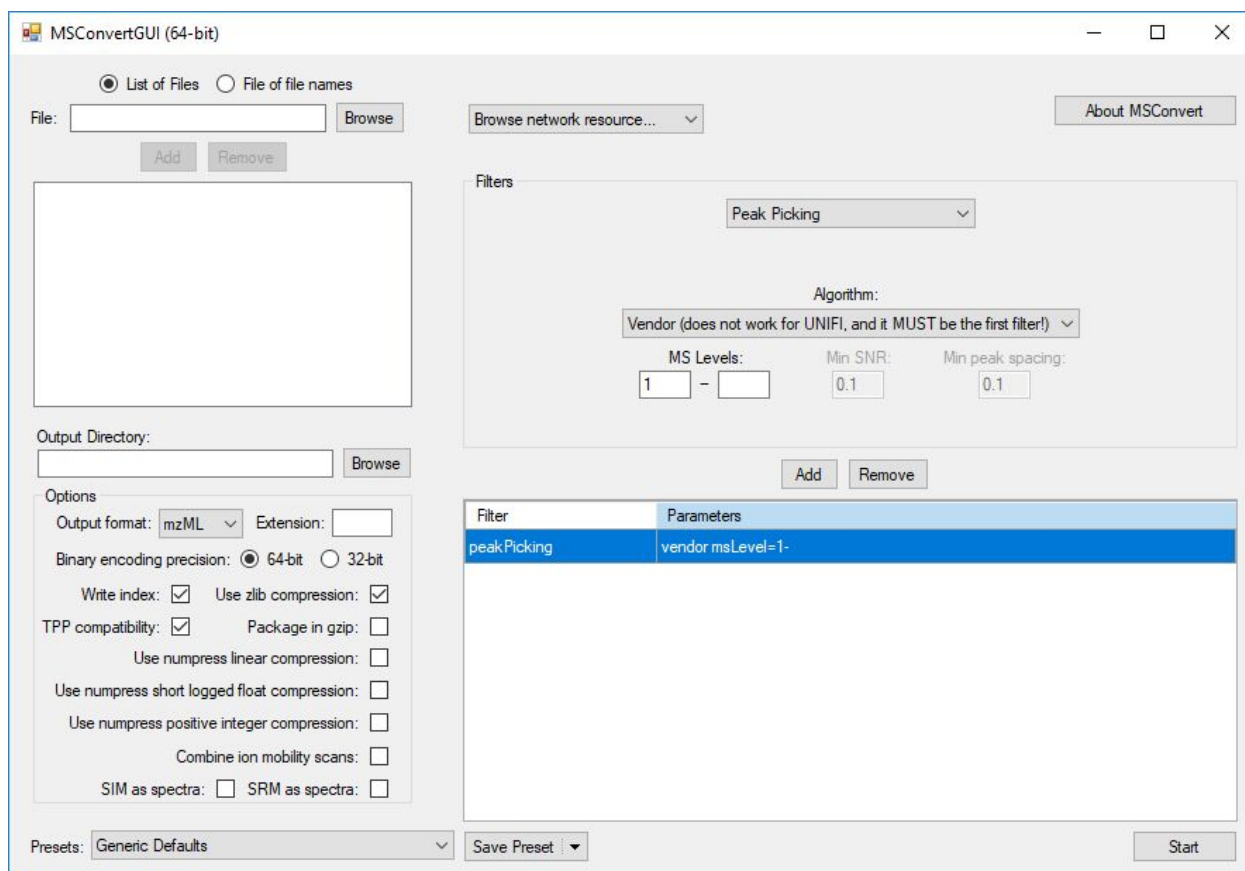
Appendix: Visualization options

- A. Skyline
- B. EncyclopeDIA viewer
- C. Viewing ELIB files with DB Browser for SQLite

GUI-BASED WORKFLOW

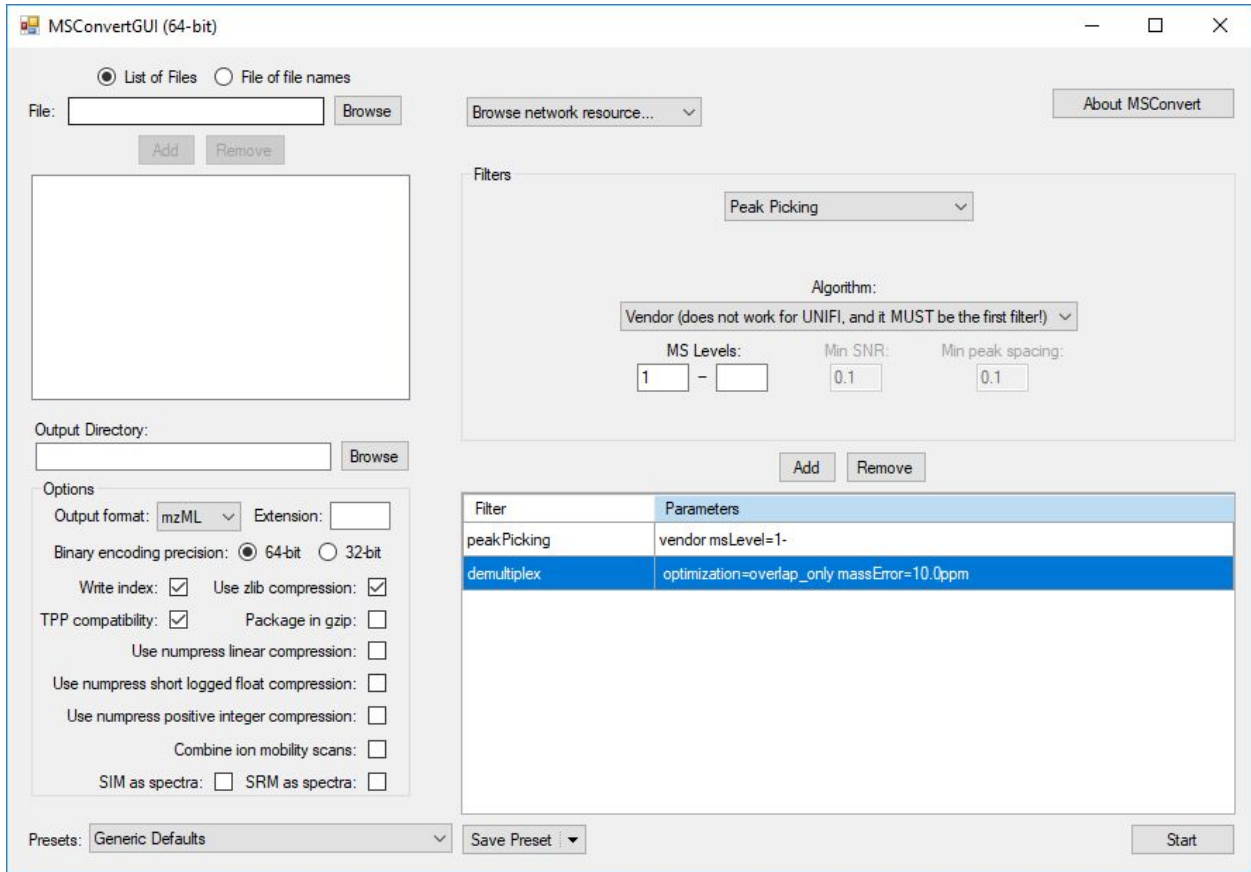
1. Convert .raw files to .mzML using MSConvert

For non-overlapping windows:

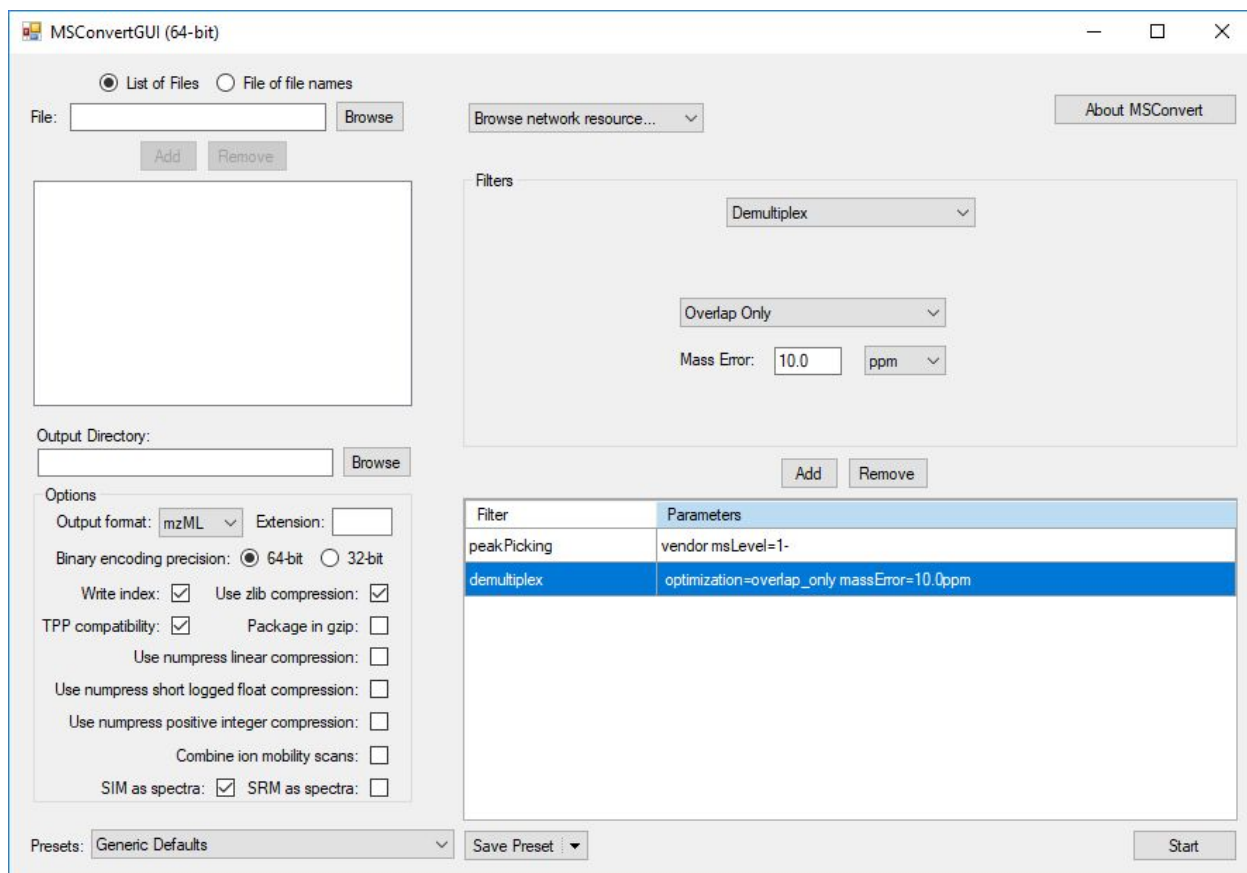


! NOTE: Make sure to have “peakPicking” as the first filter

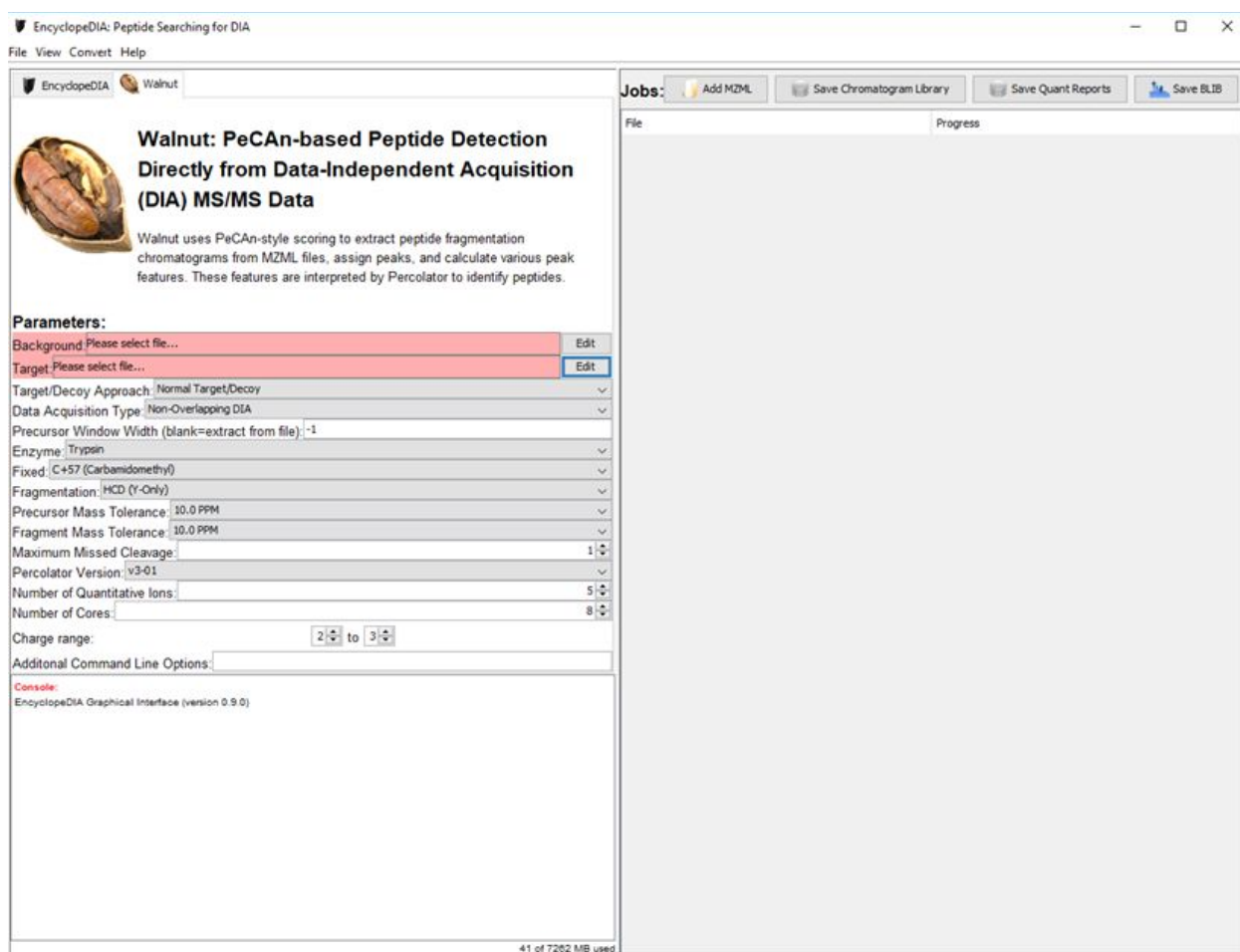
For overlapping windows:



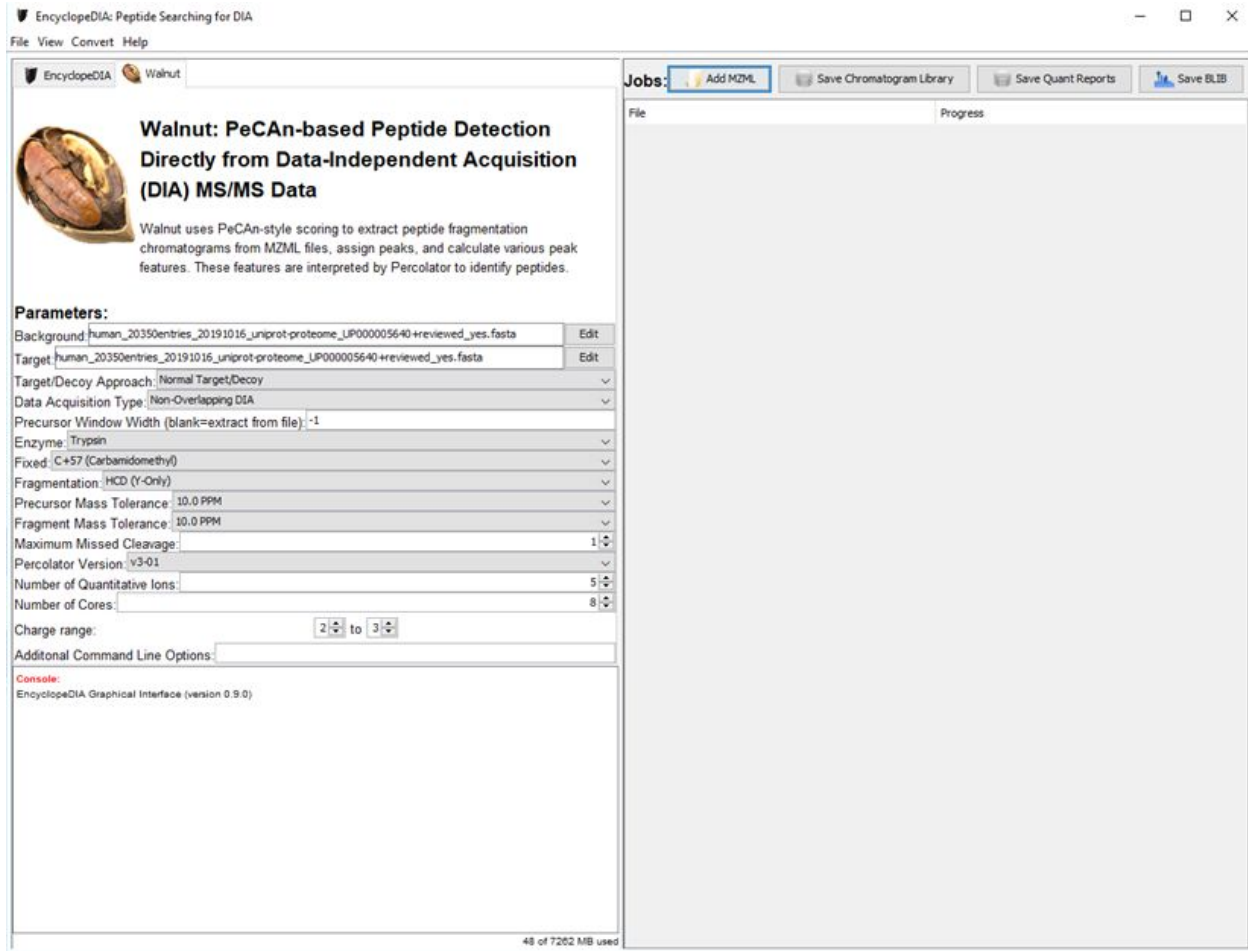
! NOTE: If the MS1 data is acquired as a SIM scan, an acquisition mode generally used for producing chromatograms, you must have another box checked (“SIM as spectra”, bottom left) so the SIM data are written to the mzML file as spectra and not as a wide extracted ion chromatogram. If so, your MSConvert should look like the screenshot below:



2. Build chromatogram library using Walnut in EncyclopeDIA



2.1 On the left hand side right underneath the Walnut icon, where it says **Parameters**, find the “**Background**” field and click the corresponding “**Edit**” button to select the background fasta file. The background fasta file should basically be the reference fasta for your model system (*E.coli*, yeast, human, etc). Here, we’ll use an example experiment in Hela, so we’ve downloaded the human reference proteome from Uniprot, and navigated the file explorer to that downloaded fasta file.



2.2 Again under “**Parameters:**”, just underneath the “**Background**” field, which should now contain the filename of that organism fasta, click the “**Target**” field corresponding “**Edit**” button. Navigate to a fasta of your target search proteome. Here, I’m interested in the whole proteome (no specific subcellular fraction like mitochondria) so I’ll select the same fasta that I used in the Background field.

! NOTE: *More about the Target/Background fasta: For experiments looking at a “whole proteome” (lysates, for example), both the Target and the Background fasta are the same file (the yeast reference fasta, human reference fasta, etc). For experiments where subcellular fractionation was performed or where you’re only interested in some subset of the proteome, use a “Background” fasta of the whole organism and a “Target” fasta just of the proteins you’re interested in (for example, a mitochondrial isolation might use the human proteome for a Background fasta, and a MitoCarta fasta that only includes mitochondrial proteins).*

! NOTE Both Background and Target files are .fasta format. PeCAAn users may recall processing a fasta to get a list of peptides for input, but Walnut includes the *in silico* digest step so you can just give Walnut the .fasta

2.3 Set the remaining parameters if you have experiment-specific details that deviate from the defaults (for example, a different digestion enzyme than the default trypsin, or a different fragmentation type than HCD, etc)

2.4 In the top right, where it says “Jobs:”, click “Add MZML”. Navigate the file explorer to your converted gas phase fractionated library files from step 1. Select all the gas phase fractionated library MZMLs and click “Open”. The MZML files you selected should now appear under the “Jobs” buttons. You can monitor progress using the GUI.

EncyclopeDIA: Peptide Searching for DIA

File View Convert Help

EncyclopeDIA Walnut

Walnut: PeCAAn-based Peptide Detection Directly from Data-Independent Acquisition (DIA) MS/MS Data

Walnut uses PeCAAn-style scoring to extract peptide fragmentation chromatograms from MZML files, assign peaks, and calculate various peak features. These features are interpreted by Percolator to identify peptides.

Parameters:

Background	human_20350entries_20191016_uniprot-proteome_LP000005640+reviewed_yes.fasta	Edit
Target	human_20350entries_20191016_uniprot-proteome_LP000005640+reviewed_yes.fasta	Edit
Target/Decoy Approach	Normal Target/Decoy	
Data Acquisition Type	Non-Overlapping DIA	
Precursor Window Width (blank=extract from file)	-1	
Enzyme	Trypsin	
Fixed	C+57 (Carbamidomethyl)	
Fragmentation	HCD (Y-Only)	
Precursor Mass Tolerance	10.0 PPM	
Fragment Mass Tolerance	10.0 PPM	
Maximum Missed Cleavage	1	
Percolator Version	V3-01	
Number of Quantitative Ions	5	
Number of Cores	8	
Charge range	2 to 3	

Additional Command Line Options:

```
python3 main.py --input: main_2018may16_hela_window_size_test_BCS_hela_narrow_2.mzML  
Adding mzML import to queue for  
[C:\Users\lindi\Desktop\tutorial\2018may16_hela_window_size_test_BCS_hela_narrow_6.mzML]  
Adding new job to queue: Read 2018may16_hela_window_size_test_BCS_hela_narrow_6.mzML  
Converting files...  
Adding mzML import to queue for  
[C:\Users\lindi\Desktop\tutorial\2018may16_hela_window_size_test_BCS_hela_narrow_7.mzML]  
Adding new job to queue: Read 2018may16_hela_window_size_test_BCS_hela_narrow_7.mzML  
Adding mzML import to queue for  
[C:\Users\lindi\Desktop\tutorial\2018may16_hela_window_size_test_BCS_hela_narrow_8.mzML]  
Adding new job to queue: Read 2018may16_hela_window_size_test_BCS_hela_narrow_8.mzML  
Indexing 2018may16_hela_window_size_test_BCS_hela_narrow_1.mzML ...  
Converting 2018may16_hela_window_size_test_BCS_hela_narrow_1.mzML ...  
Parsed 1%  
Parsed 2%  
Parsed 3%  
Parsed 4%  
Parsed 5%
```

117 of 7262 MB used

Jobs: Add MZML Save Chromatogram Library Save Quant Reports Save BLIB

File	Progress
Read 2018may16_hela_window_size_test_BCS_hela_narrow_...	Converting files...
Read 2018may16_hela_window_size_test_BCS_hela_narrow_...	
Read 2018may16_hela_window_size_test_BCS_hela_narrow_...	
Read 2018may16_hela_window_size_test_BCS_hela_narrow_...	
Read 2018may16_hela_window_size_test_BCS_hela_narrow_...	
Read 2018may16_hela_window_size_test_BCS_hela_narrow_...	
Read 2018may16_hela_window_size_test_BCS_hela_narrow_...	
Read 2018may16_hela_window_size_test_BCS_hela_narrow_...	
Read 2018may16_hela_window_size_test_BCS_hela_narrow_...	

2.5 When the six gas phase fractionated files have finished running, click “**SAVE CHROMATOGRAM LIBRARY**” and give your library some descriptive filename.

The screenshot shows the EncyclopeDIA software interface. The main window displays the 'Walnut: PeCAN-based Peptide Detection' section, which includes a description of the method and a list of parameters. The 'Jobs' panel on the right shows a progress table for the current run.

Walnut: PeCAN-based Peptide Detection
Directly from Data-Independent Acquisition (DIA) MS/MS Data

Walnut uses PeCAN-style scoring to extract peptide fragmentation chromatograms from MZML files, assign peaks, and calculate various peak features. These features are interpreted by Percolator to identify peptides.

Parameters:

Background	human_20350entries_20191016_uniprot-proteome_LP000005640+reviewed_yes.fasta	Edit
Target	human_20350entries_20191016_uniprot-proteome_LP000005640+reviewed_yes.fasta	Edit
Target/Decoy Approach	Normal Target/Decoy	
Data Acquisition Type	Non-Overlapping DIA	
Precursor Window Width (blank=extract from file)	-1	
Enzyme	Trypsin	
Fixed	C+57 (Carbamidomethyl)	
Fragmentation	HCD (Y-Only)	
Precursor Mass Tolerance	10.0 PPM	
Fragment Mass Tolerance	10.0 PPM	
Maximum Missed Cleavage	1	
Percolator Version	v3-01	
Number of Quantitative Ions	5	
Number of Cores	8	
Charge range	2 to 3	

Additional Command Line Options:

```

cmd=pepores c:\users\indv\desktop\tutorial\2018may16_hela_window_size_test_bcs_hela_narrow_1-
s_LIBRARY_concatenated_decoy.txt --no-terminate -N 200000 C:
/Users/indv/Desktop/tutorial/2018may16_hela_window_size_test_bcs_hela_narrow_1-
s_LIBRARY_concatenated_features.txt
Started Tue Nov 19 02:22:07 2019
Hyperparameters: selectionFdr=0.01, Cpos=0, Cneg=0, maxIter=10
Reading tab-delimited input from datafile C:
/Users/indv/Desktop/tutorial/2018may16_hela_window_size_test_bcs_hela_narrow_1-
s_LIBRARY_concatenated_features.txt
Features:
topN rank peakZScore peakCalibratedScore deltaSn avgldotp midldotp peakScore peakWeightedScore NCI
CIMassEnrMean CIMassEnrVar precursorMassEnrMean precursorMassEnrVar peakSimilarity sampledTimes midTime
spectraNorm pepLength charge2 charge3 precursorMz
Found 716572 PSMs
Separate target and decoy search inputs detected, using mix-max method.
Train/test set contains 104894 positives and 95116 negatives, size ratio=1.1027 and pi0=1
Warning: The mix-max procedure is not well behaved when # targets (104884) != # decoys (95116). Consider using target-
decoy competition (-Y flag).
  
```


Jobs: Add MZML Save Chromatogram Library Save Quant Reports Save BLIB

File	Progress
Read 2018may16_hela_window_size_test_bcs_hela_narrow...	Wrote 3721 peptides identified at 1.0% FDR
Read 2018may16_hela_window_size_test_bcs_hela_narrow...	Wrote 10100 peptides identified at 1.0% FDR
Read 2018may16_hela_window_size_test_bcs_hela_narrow...	Wrote 9227 peptides identified at 1.0% FDR
Read 2018may16_hela_window_size_test_bcs_hela_narrow...	Wrote 5808 peptides identified at 1.0% FDR
Read 2018may16_hela_window_size_test_bcs_hela_narrow...	Wrote 3640 peptides identified at 1.0% FDR
Read 2018may16_hela_window_size_test_bcs_hela_narrow...	Wrote 1506 peptides identified at 1.0% FDR
Read 2018may16_hela_window_size_test_bcs_hela_narrow...	Wrote 720 peptides identified at 1.0% FDR
Read 2018may16_hela_window_size_test_bcs_hela_narrow...	Wrote 249 peptides identified at 1.0% FDR
Write Library 2018may16_hela_window_size_test_bcs_hela...	

3288 of 7800 MB used

EncyclopeDIA: Peptide Searching for DIA

File View Convert Help



Walnut: PeCAN-based Peptide Detection Directly from Data-Independent Acquisition (DIA) MS/MS Data

Walnut uses PeCAN-style scoring to extract peptide fragmentation chromatograms from MZML files, assign peaks, and calculate various peak features. These features are interpreted by Percolator to identify peptides.

Parameters:

Background	human_20350entries_20191016_uniprot-proteome_LIP000005640+reviewed_yes.fasta	Edit
Target	human_20350entries_20191016_uniprot-proteome_LIP000005640+reviewed_yes.fasta	Edit
Target/Decoy Approach	Normal Target/Decoy	
Data Acquisition Type	Non-Overlapping DIA	
Precursor Window Width (blank=extract from file)	-1	
Enzyme	Trypsin	
Fixed	C+57 (Carbamidomethyl)	
Fragmentation	HCD (Y-Only)	
Precursor Mass Tolerance	10.0 PPM	
Fragment Mass Tolerance	10.0 PPM	
Maximum Missed Cleavage	1	
Percolator Version	v3-01	
Number of Quantitative Ions	5	
Number of Cores	8	

Charge range: 2 to 3

Additional Command Line Options:

```

walnut extracting 1176.8 to 1178.8 m/z (86.77353 to 83.971664 min)
Quant Extracting 1176.8 to 1178.8 m/z (86.77353 to 83.971664 min)
Quant Extracting 1178.8 to 1180.8 m/z (85.09182 to 74.9485 min)
Quant Extracting 1182.8 to 1184.8 m/z (42.08993 to 90.1653 min)
Quant Extracting 1184.8 to 1186.8 m/z (54.965083 to 106.29204 min)
Quant Extracting 1186.8 to 1188.8 m/z (90.23343 to 99.7341 min)
Quant Extracting 1188.8 to 1190.8 m/z (84.043785 to 81.60174 min)
Quant Extracting 1192.8 to 1194.8 m/z (74.8886 to 84.49803 min)
Quant Extracting 1194.8 to 1196.8 m/z (75.44697 to 77.06919 min)
Quant Extracting 1196.8 to 1198.8 m/z (86.37902 to 87.17902 min)
Quant Extracting 1198.8 to 1200.8 m/z (83.34446 to 84.134445 min)
Quant Extracting 1200.8 to 1202.8 m/z (59.71617 to 60.51617 min)
Writing EncyclopeDIA ELIB from 2018may16_hela_window_size_test_BCS_hela_narrow_8.mzML (226 entries)...
Writing 226 peptides to entries table...
Writing 226 peptides to peptidequants table...
Finished writing to EncyclopeDIA ELIB at Tue Nov 19 02:25:38 EST 2019
Writing global target/decoy peptides: 31642/316, pi0: 0.832606
Writing global target/decoy proteins: 3983/39

```

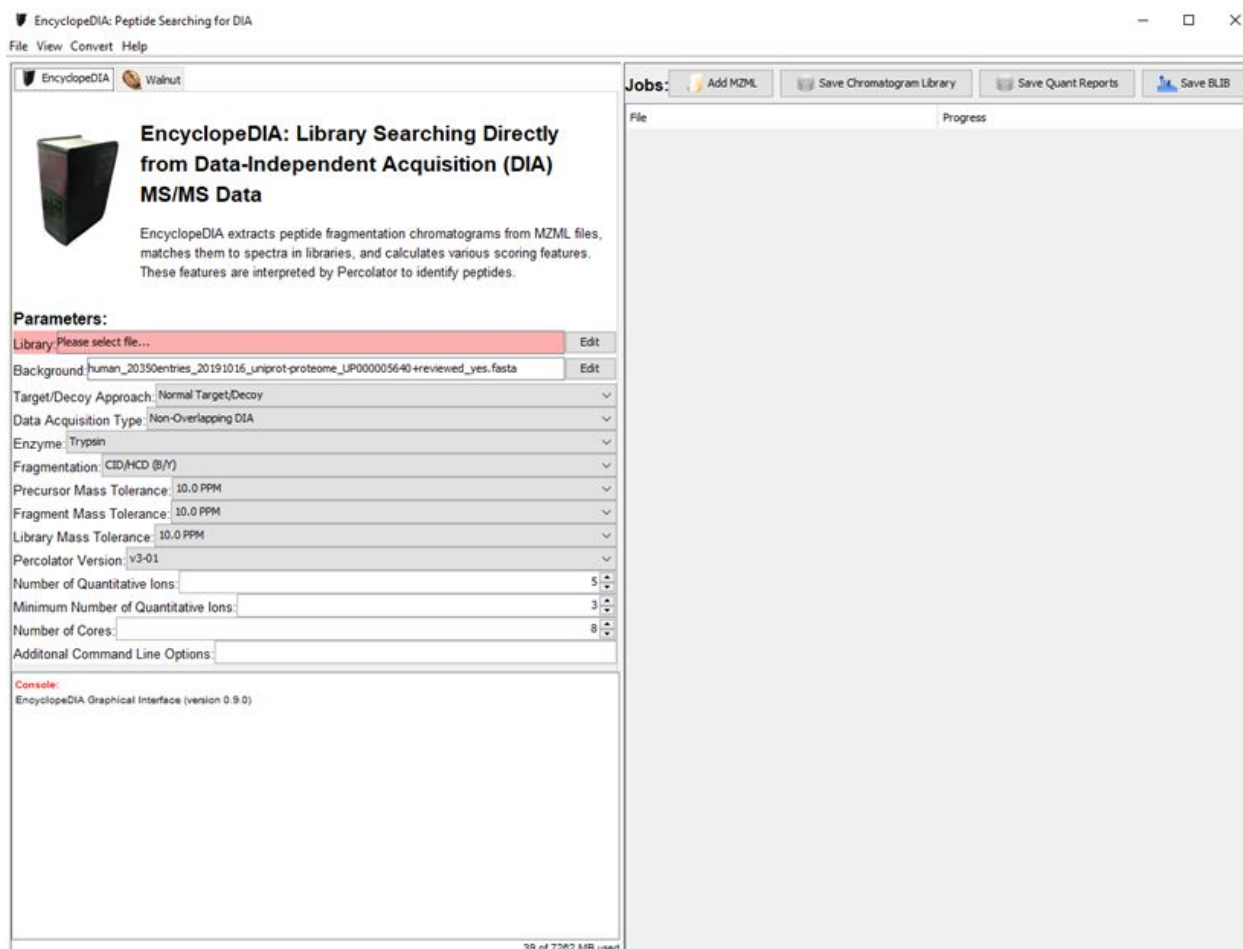
2283 of 7262 MB used

Jobs: Add MZML Save Chromatogram Library Save Quant Reports Save BLB

File	Progress
Read 2018may16_hela_window_size_test_BCS_hela_narrow...	Wrote 3721 peptides identified at 1.0% FDR
Read 2018may16_hela_window_size_test_BCS_hela_narrow...	Wrote 10100 peptides identified at 1.0% FDR
Read 2018may16_hela_window_size_test_BCS_hela_narrow...	Wrote 9227 peptides identified at 1.0% FDR
Read 2018may16_hela_window_size_test_BCS_hela_narrow...	Wrote 5808 peptides identified at 1.0% FDR
Read 2018may16_hela_window_size_test_BCS_hela_narrow...	Wrote 3640 peptides identified at 1.0% FDR
Read 2018may16_hela_window_size_test_BCS_hela_narrow...	Wrote 1506 peptides identified at 1.0% FDR
Read 2018may16_hela_window_size_test_BCS_hela_narrow...	Wrote 720 peptides identified at 1.0% FDR
Read 2018may16_hela_window_size_test_BCS_hela_narrow...	Wrote 249 peptides identified at 1.0% FDR
Write Library 2018may16_hela_window_size_test_BCS_hela...	31642 peptides identified at 1.0% FDR

3. Search wide-window data with chromatogram library from Step 2

3.1 Close and reopen the EncyclopeDIA GUI to clear EncyclopeDIA's cache/history



3.2 Within EncyclopeDIA GUI (not Walnut), on the left hand side under “Parameters:” across from the “Library” field, click the “Edit” button. Using the file explorer, select the .elib file you just saved in Step 2.5

EncyclopeDIA: Peptide Searching for DIA

File View Convert Help

EncyclopeDIA Walnut

EncyclopeDIA: Library Searching Directly from Data-Independent Acquisition (DIA) MS/MS Data

EncyclopeDIA extracts peptide fragmentation chromatograms from MZML files, matches them to spectra in libraries, and calculates various scoring features. These features are interpreted by Percolator to identify peptides.

Parameters:

Library: 2018may16_hela_window_size_test_BCS_hela_narrow_1-8_LIBRARY.elib

Background: human_20350entries_20191016_uniprot-proteome_UP000005640+reviewed_yes.fasta

Target/Decoy Approach: Normal Target/Decoy

Data Acquisition Type: Non-Overlapping DIA

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:

Console:
EncyclopeDIA Graphical Interface (version 0.9.0)

Jobs: Add MZML Save Chromatogram Library Save Quant Reports Save BLIB


File Progress

40 of 7262 MB used

3.3 Underneath the “Library” field, across from the “Background” field, click the corresponding “Edit” button and select the appropriate Background file (should be the same fasta you used in 2.1 above!)

EncyclopeDIA: Peptide Searching for DIA

File View Convert Help



EncyclopeDIA: Library Searching Directly from Data-Independent Acquisition (DIA) MS/MS Data

EncyclopeDIA extracts peptide fragmentation chromatograms from MZML files, matches them to spectra in libraries, and calculates various scoring features. These features are interpreted by Percolator to identify peptides.

Parameters:

Library: 2018may16_hela_window_size_test_BCS_hela_narrow_1-8_LIBRARY.elib Edit

Background: human_20350entries_20191016_uniprot-proteome_up000005640-reviewed_yes.fasta Edit

Target/Decoy Approach: Normal Target/Decoy

Data Acquisition Type: Non-Overlapping DIA

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:

```

writing global target/decoy peptides: 31642/316, pi0: 0.832606
writing global target/decoy proteins: 3983/39
adding mzML import to queue for [C]
/Users/ind/Desktop/tutorial/2018may14_hela_window_size_test_BCS_hela_wide_400_1000_1.mzML
opening library 2018may16_hela_window_size_test_BCS_hela_narrow_1-8_LIBRARY.elib (version: 0.1.14)
adding new job to queue: Read 2018may14_hela_window_size_test_BCS_hela_wide_400_1000_1.mzML
adding mzML import to queue for [C]
/Users/ind/Desktop/tutorial/2018may14_hela_window_size_test_BCS_hela_wide_400_1000_2.mzML
converting files...
adding new job to queue: Read 2018may14_hela_window_size_test_BCS_hela_wide_400_1000_2.mzML
adding mzML import to queue for [C]
/Users/ind/Desktop/tutorial/2018may14_hela_window_size_test_BCS_hela_wide_400_1000_3.mzML
adding new job to queue: Read 2018may14_hela_window_size_test_BCS_hela_wide_400_1000_3.mzML
indexing 2018may14_hela_window_size_test_BCS_hela_wide_400_1000_1.mzML ...
converting 2018may14_hela_window_size_test_BCS_hela_wide_400_1000_1.mzML ...
Parsed 1%
Parsed 2%

```

2529 of 7262 MB used

Jobs: Add MZML Save Chromatogram Library Save Quant Reports Save BLIB

File	Progress
Read 2018may14_hela_window_size_test_BCS_hela_wide_4...	Converting files...
Read 2018may14_hela_window_size_test_BCS_hela_wide_4...	
Read 2018may14_hela_window_size_test_BCS_hela_wide_4...	

3.4 In the top right, next to “Jobs”, click the “Add MZML” button and select all of the wide-window .mzML files that were acquired using this narrow-window library file.

The screenshot shows the EncyclopeDIA software interface. On the left, there is a header with the EncyclopeDIA logo and a brief description: "EncyclopeDIA: Library Searching Directly from Data-Independent Acquisition (DIA) MS/MS Data". Below this, the "Parameters:" section lists various settings such as Library, Background, Target/Decoy Approach, Data Acquisition Type, Enzyme, Fragmentation, and various mass tolerances. At the bottom of the parameters section, there is a log of command line options and their execution status.

On the right side, there is a "Jobs:" panel with buttons for "Add MZML", "Save Chromatogram Library", "Save Quant Reports", and "Save BLIB". Below these buttons is a table showing the progress of several jobs:

File	Progress
Read 2018may14_hela_window_size_test_BCS_hela_wide_4...	Wrote 26704 peptides identified at 1.0% FDR
Read 2018may14_hela_window_size_test_BCS_hela_wide_4...	Wrote 26768 peptides identified at 1.0% FDR
Read 2018may14_hela_window_size_test_BCS_hela_wide_4...	Wrote 28179 peptides identified at 1.0% FDR

3.5 Click **“Save Quant Reports”** to perform a final experiment-wide FDR correction and export peptide quant, and protein quant.

- Select **“Save Chromatogram Library”** to build a file with bonus information like integration boundaries.
 - This is only applicable if you do not want to retention time-align across the MZML files (for example, if your MZMLs are fractionated in a way such that you don't expect to sample the same peptides in each file)
- Select **“Save Quant Reports”** to get peptide/protein quantitation matrices in the form of a tsv.
 - Pick this if your MZMLs were the experimental samples you want to post-process.
 - If you are following this workflow as-is, this is what you should pick!
- Select **“Save BLIB”** to build a spectral library file that Skyline can use.
 - This option is effectively depreciated now that Skyline reads ELIB file formats

EncyclopeDIA: Peptide Searching for DIA

File View Convert Help

EncyclopeDIA Walnut

EncyclopeDIA: Library Searching Directly from Data-Independent Acquisition (DIA) MS/MS Data

EncyclopeDIA extracts peptide fragmentation chromatograms from MZML files, matches them to spectra in libraries, and calculates various scoring features. These features are interpreted by Percolator to identify peptides.

Parameters:

Library: 2018may16_hela_window_size_test_BCS_hela_narrow_1-8_LIBRARY.elib Edit

Background: human_20350entries_20191016_uniprot-proteome_LP000005640-reviewed_yes.fasta Edit

Target/Decoy Approach: Normal Target/Decoy

Data Acquisition Type: Non-Overlapping DIA

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:

```

Writing source peptides to pepinfo/quantar files...
Skipped 1879 integrated library entries because the RT inferrer could not find any fragment ions.
Finished writing to EncyclopeDIA ELIB at Tue Nov 19 04:41:20 EST 2019
Writing global target/decoy peptides: 28824/17204, p10: 0.0102301
Writing global target/decoy proteins: 3993/39
Outlining source files...
Found 3 data files...
10000 records processed...
20000 records processed...
30000 records processed...
40000 records processed...
50000 records processed...
60000 records processed...
Finished processing 65046 records, found 21682 quantitative unique peptides. Writing reports...
Inconsistent number of fragments in 5041 of 21682 peptides
Finished writing peptide report for 62154 unique peptides!
Finished writing protein report for 3503 protein groups!

```

773 of 7262 MB used

Jobs: Add MZML Save Chromatogram Library Save Quant Reports Save BLIB

File	Progress
Read 2018may14_hela_window_size_test_BCS_hela_wide_4...	Wrote 26704 peptides identified at 1.0% FDR
Read 2018may14_hela_window_size_test_BCS_hela_wide_4...	Wrote 26768 peptides identified at 1.0% FDR
Read 2018may14_hela_window_size_test_BCS_hela_wide_4...	Wrote 28179 peptides identified at 1.0% FDR
Write Library 2018may14_hela_window_size_test_BCS_hela...	28824 peptides identified at 1.0% FDR

! NOTE EncyclopeDIA is determining a lot of information about the DIA experiment. Important details include peptide detections, fragment refinement, and peak boundaries. We'll import that information into Skyline next so that we can visualize the results; however, there's also a visualizer built right into EncyclopeDIA. See Appendix for details.

! NOTE EncyclopeDIA's ELIB output can be used in Skyline for visualizing DIA-MS experiments. See Appendix for details.

COMMAND LINE WORKFLOW

1. Convert .raw files into .mzml using MSConvert

For non-overlapping windows:

```
msconvert.exe -v --zlib --64 --mzML --filter "peakPicking true 1-" *.raw
```

For overlapping windows:

```
msconvert.exe -v --zlib --64 --mzML --filter "peakPicking true 1-" --filter "demultiplex optimization=overlap_only" *.raw
```

NOTE: data acquired on Lumos instruments needs an extra flag (`--simAsSpectra`) to convert precursor scans correctly:

```
msconvert.exe --zlib --64 --mzML --filter "peakPicking true 1-" --filter "demultiplex optimization=overlap_only" --simAsSpectra *.raw
```

! NOTE If your file conversion is going slow, you probably aren't using a current version of MSconvert!

2. Build chromatogram library using Walnut (command line)

2.1. Searching DIA data against a FASTA with Walnut: for a full list of options and default values for your version of EncyclopeDIA:

```
java -jar encyclopedia.jar -walnut --help
```

The parameters you may want to change include the enzyme used to prepare the samples,

```
$ java -Xmx<GB_OF_MEM>G -jar encyclopedia.jar -walnut \  
-i <MZML_IN> \  
-f <BACKGROUND_FASTA> \  
-t <TARGET_FASTA> \  
-acquisition DIA \  

```



```
-enzyme <ENZYME> \  
-frag <FRAGMENTATION> \  
-ftol <FRAGMENT_TOLERANCE> \  
-ftolunits <FRAGMENT_TOLERANCE_UNITS> \  
-ptol <PRECURSOR_TOLERANCE> \  
-ptolunits <PRECURSOR_TOLERANCE_UNITS> \  
-minCharge <MIN_CHARGE> \  
-maxCharge <MAX_CHARGE>
```

With a typical DIA setup (trypsin digest, Orbitrap instrument, demultiplexing the RAW file overlapping windows with MSConvert), the command usually looks like this:

```
$ java -Xmx8G -jar encyclopedia.jar -walnut \  
-i DIA_narrow_run_400to500mz.mzML \  
-f human.fasta \  
-t human.fasta \  
-acquisition DIA \  
-enzyme trypsin \  
-frag YONLY \  
-ftol 10.0 \  
-ftolunits ppm \  
-ptol 10.0 \  
-ptolunits ppm \  
-minCharge 2 \  
-maxCharge 3
```

Running Walnut produces several results files, including:

```
<MZML_IN>.dia  
<MZML_IN>.mzML.pecan.txt.log  
<MZML_IN>.mzML.features.txt  
<MZML_IN>.mzML.pecan.txt  
<MZML_IN>.mzML.pecan.decoy.txt
```

2.1. Merge Walnut results into a chromatogram library file (.elib): this command must be run within the same directory as the search command from 2a. You have to be in the same directory because EncyclopeDIA will look for the result files in the current

directory in order to compile them into one ELIB, so if the result files aren't there, Walnut has nothing to compile.

For a full list of options and default values for your version of EncyclopeDIA:

```
java -jar encyclopedia.jar -libexport --help
```

OUTPUT_LIBRARY_NAME: (filename) Any name of your choice. Entering **narrow_merged** here would result in a final output of narrow_merged.elib (or narrow_merged.blib).

ALIGN_SPECTRA?: (**true** or **false**) You will likely want to set this to **false** for this step as you're probably doing a narrow isolation gas phase fractionation. In general, if each of your mzML acquisitions collect an identical precursor range, this should be **true** -- otherwise it should be **false**.

USE_BLIB_FLAG: To export an elib, leave this blank. To export a .blib, set to: **-blib**

Typical run:

```
$ java -Xmx<GB_OF_MEM>G -jar encyclopedia.jar -libexport \  
-i <INPUT_DIRECTORY> \  
-o <OUTPUT_LIBRARY_NAME> \  
-a <ALIGN_SPECTRA?> \  
<USE_BLIB_FLAG> \  
-f <BACKGROUND_FASTA> \  
-t <TARGET_FASTA> \  
-ftol <FRAGMENT_TOLERANCE> \  
-ftolunits <FRAGMENT_TOLERANCE_UNITS>
```

Example:

```
$ java -Xmx8G -jar encyclopedia.jar -libexport \  
-i ./ \  
-o narrow_merged \  
-a false \  
-f human.fasta \  
-t human.fasta \  
-ftol 10.0 \  
-ftolunits ppm
```

3. Search wide-window data with chromatogram library from step 2 using EncyclopeDIA (command line)

For a full list of options and default values for your version of EncyclopeDIA:

```
java -jar encyclopedia.jar --help
```

LIBRARY_ELIB_FILE: This should be the result from your narrow library search in step 2.b.ii, e.g. **path/to/your/narrow_merged.elib**

Typical run:

```
$ java -Xmx<GB_OF_MEM>G -jar encyclopedia.jar \  
-l <LIBRARY_ELIB_FILE> \  
-i <MZML_IN> \  
-f <BACKGROUND_FASTA> \  
-t <TARGET_FASTA> \  
-acquisition DIA \  
-enzyme <ENZYME> \  
-frag <FRAGMENTATION> \  
-ftol <FRAGMENT_TOLERANCE> \  
-ftolunits <FRAGMENT_TOLERANCE_UNITS> \  
-ptol <PRECURSOR_TOLERANCE> \  
-ptolunits <PRECURSOR_TOLERANCE_UNITS> \  
-minCharge <MIN_CHARGE> \  
-maxCharge <MAX_CHARGE>
```

Output files:

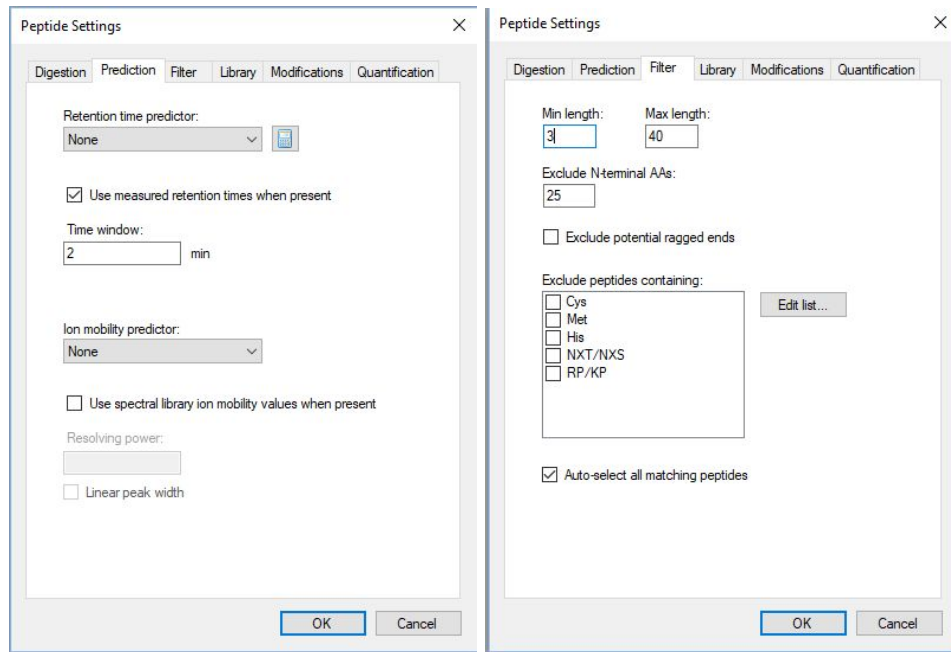
```
<MZML_IN>.dia  
<MZML_IN>.mzML.elib  
<MZML_IN>.mzML.encyclopedia.txt  
<MZML_IN>.mzML.encyclopedia.txt.delta_rt.pdf  
<MZML_IN>.mzML.encyclopedia.txt.log  
<MZML_IN>.mzML.encyclopedia.txt.rt_fit.pdf  
<MZML_IN>.mzML.encyclopedia.txt.rt_fit.txt
```

<MZML_IN>.mzML.features.txt
<MZML_IN>.mzML.first_round.txt

Appendix

Importing peptide detections into Skyline

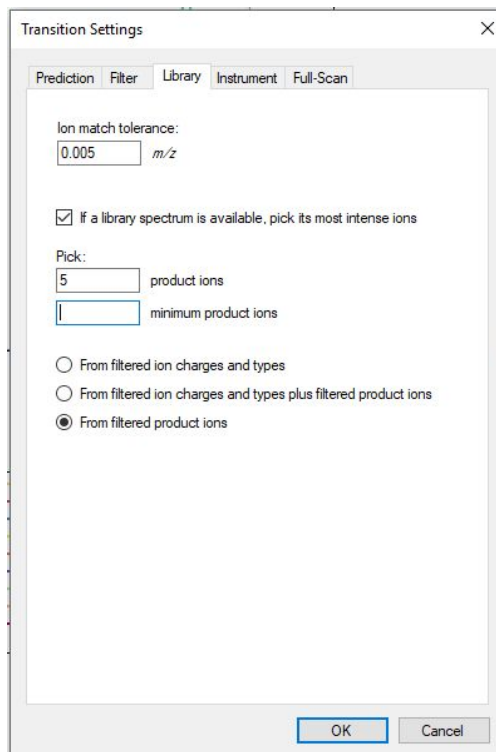
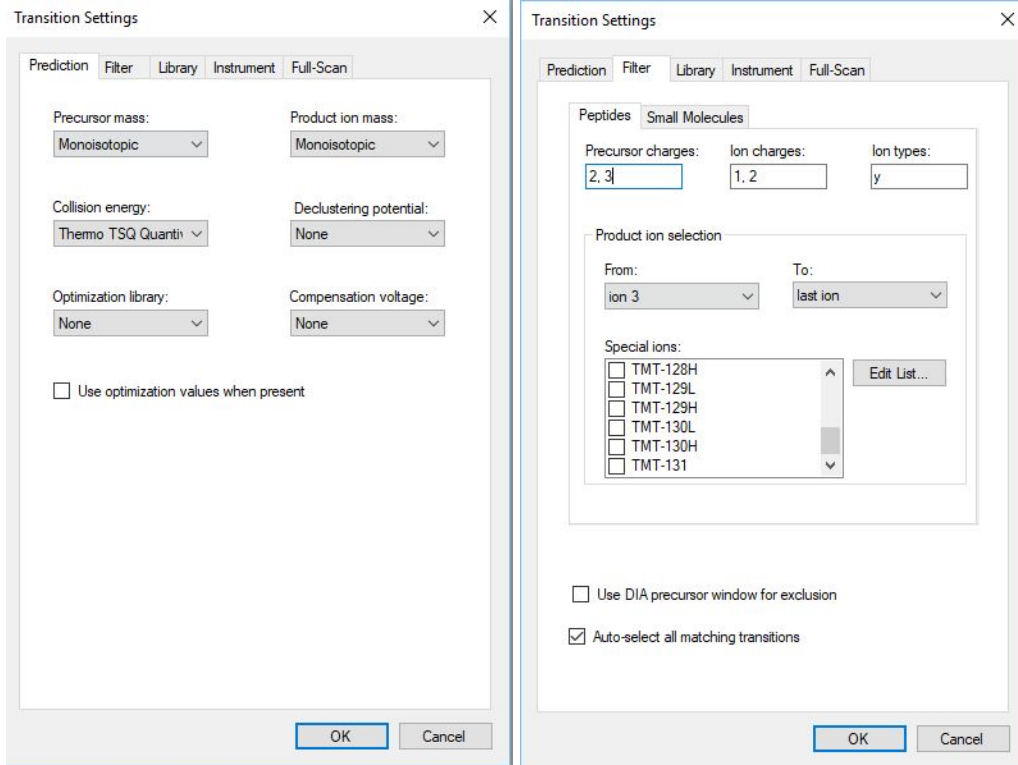
Settings > Peptide Settings.

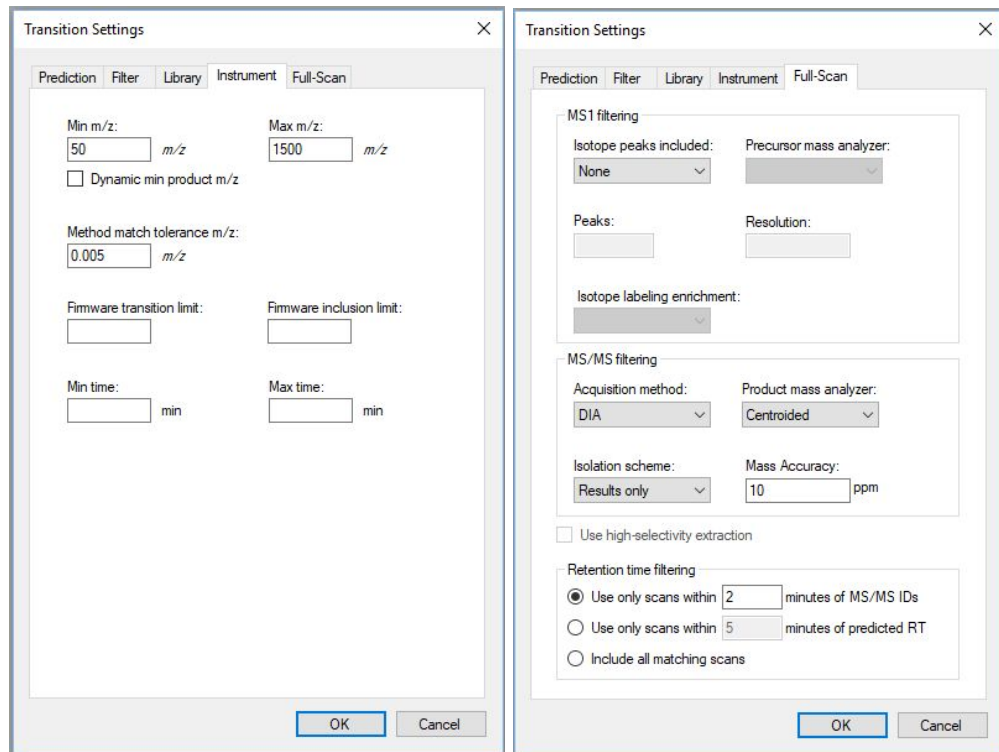


4a. Prediction: Check “Use measured retention times when present”, Time window=2

4b. Filter: Min length=3, Max length=40, no excluded amino acids checked

Settings > Transition Settings.





4c. Prediction: Precursor/Product ion mass="Monoisotopic"

4d. Filter: Ion charges=1,2 Ion types="y"*, From=Ion 3, To=last ion, no special ions
*Set "ion types" to reflect how you searched the data in Encyclopedia!

4e. Library: Ion match tolerance=0.005 m/z, check "If a library spectrum is available, pick its most intense ions", pick=5 product ions, select "From filtered product ions"

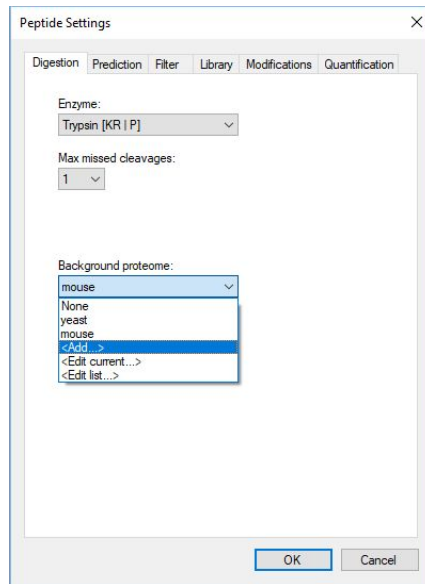
4f. Instrument: "Min m/z=50, Max m/z=1500, Method match tolerance m/z=0.005"

4g. Full-Scan (MS1): "Isotope peaks included=None"

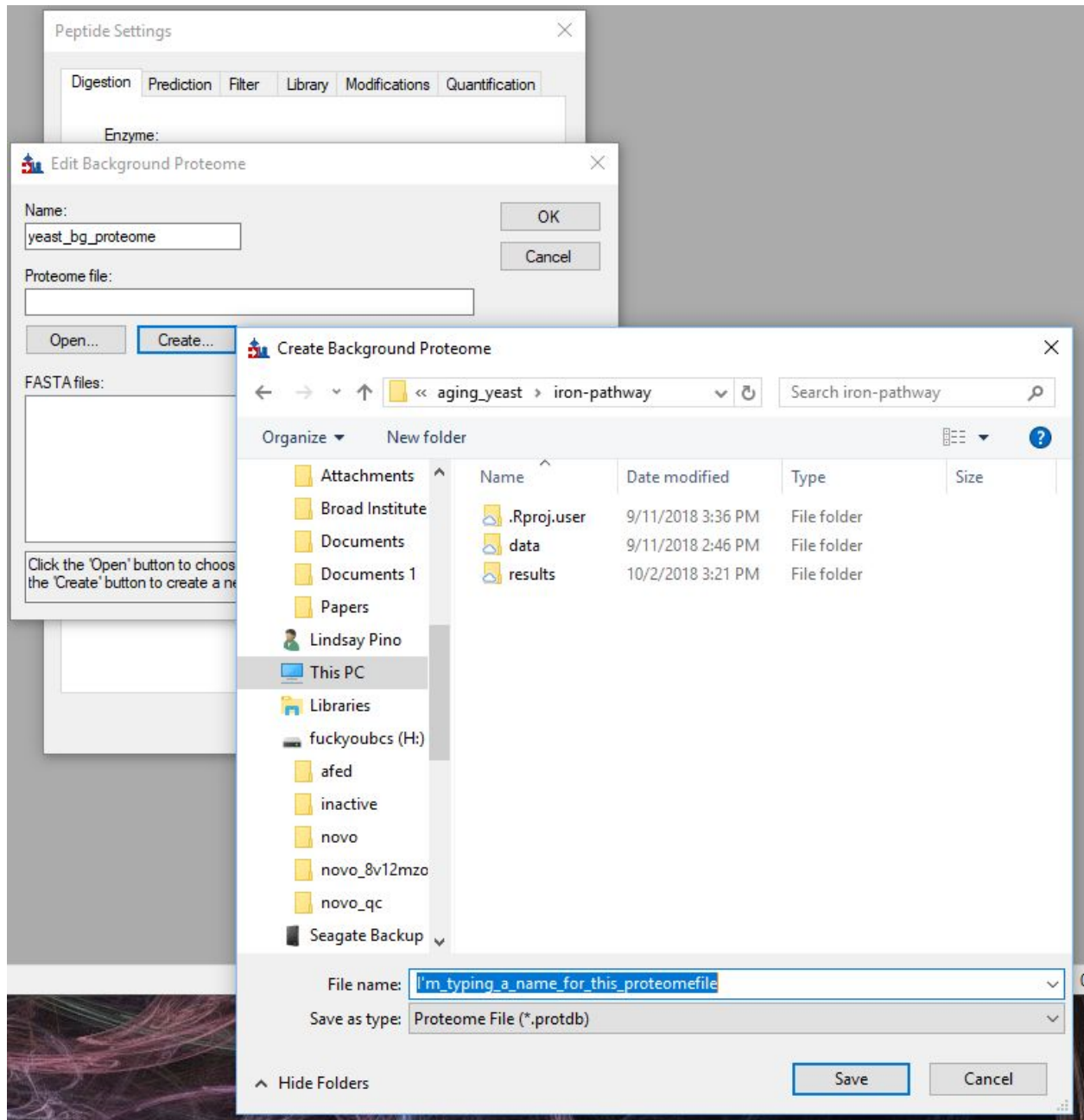
4h. Full-Scan (MS/MS): "Acquisition method=DIA, Product mass analyzer=Centroided, Isolation scheme=Results only, Mass Accuracy=10 ppm, check "Use only scans within "2" minutes of MS/MS IDs"

Load the FASTA database and the ELIB library from EncyclopeDIA into Skyline.

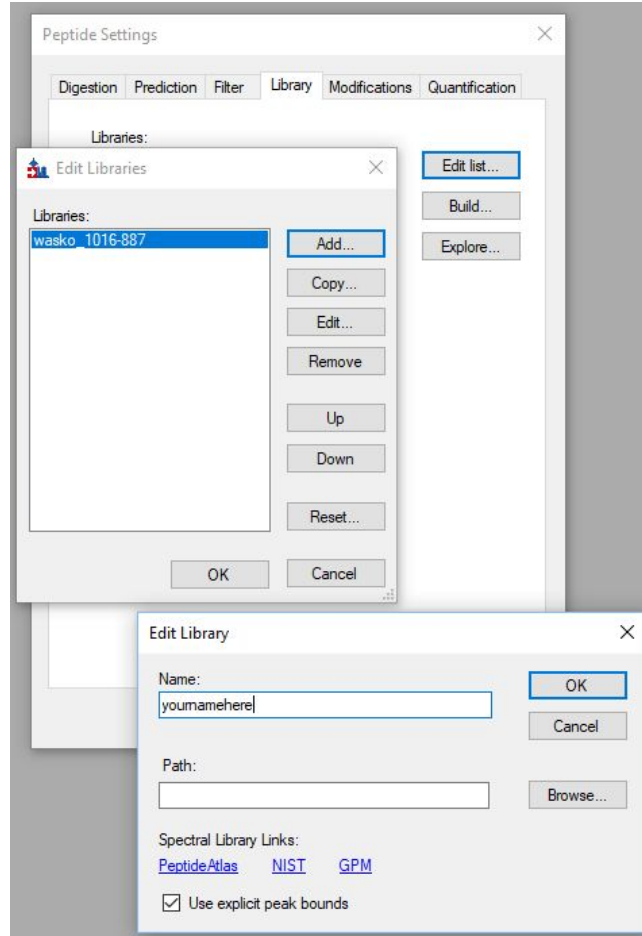
4i. Open Settings/Peptide Settings/Digestion. Select “<Add...>” for Background proteome



4j. Set Name="[yournamehere]", then click “Create...” and type in a descriptive name for the soon-to-be-created Proteome File and click “**Save**”. Select “Add File...” to add the background fasta you were just using in the Encyclopedia suite.



4k. Open Settings/Peptide Settings/Library. Select “Edit list...”, then “Add...”. Set the name to be “[yournamehere]” and select “Browse...” to select the .elib that you saved in EncyclopeDIA from the wide-window search (not the gas phase fractionated, narrow window library). “Ok” out to the Peptide Settings.



4l. Check the new “[yournamehere]” library and uncheck all other libraries. Click “Explore...” to view it

! If you get a pop up warning that ‘Peptide settings have been changed. Save changes?’ then click “Yes”

4m. Check “Associate proteins” and click “Add All...”. Select “Add to all matching proteins” and “Include all peptides” and hit “OK”.

4n. If you have more than one library (*.elib) for your experiment, you must select each of those libraries from the dropdown menu in "Spectral Library Explorer" and repeat step 4m for each.

4o. Save the Skyline file!

4p. Import .mzML data into Skyline

! Tip: Use the .mzml files, not the .raw files! If you prefer using the .raw files for some reason, you will need to set the windowing scheme parameters in the "MS/MS filtering" box under the "Full Scan" tab in Skyline's "Transition Settings".

Skyline: Export report for MSstats

To load the MSstats report format, install the MSstats external tool. Alternatively, you can quickly build a custom report using the little black binoculars to find whatever fields you're interested in exporting.

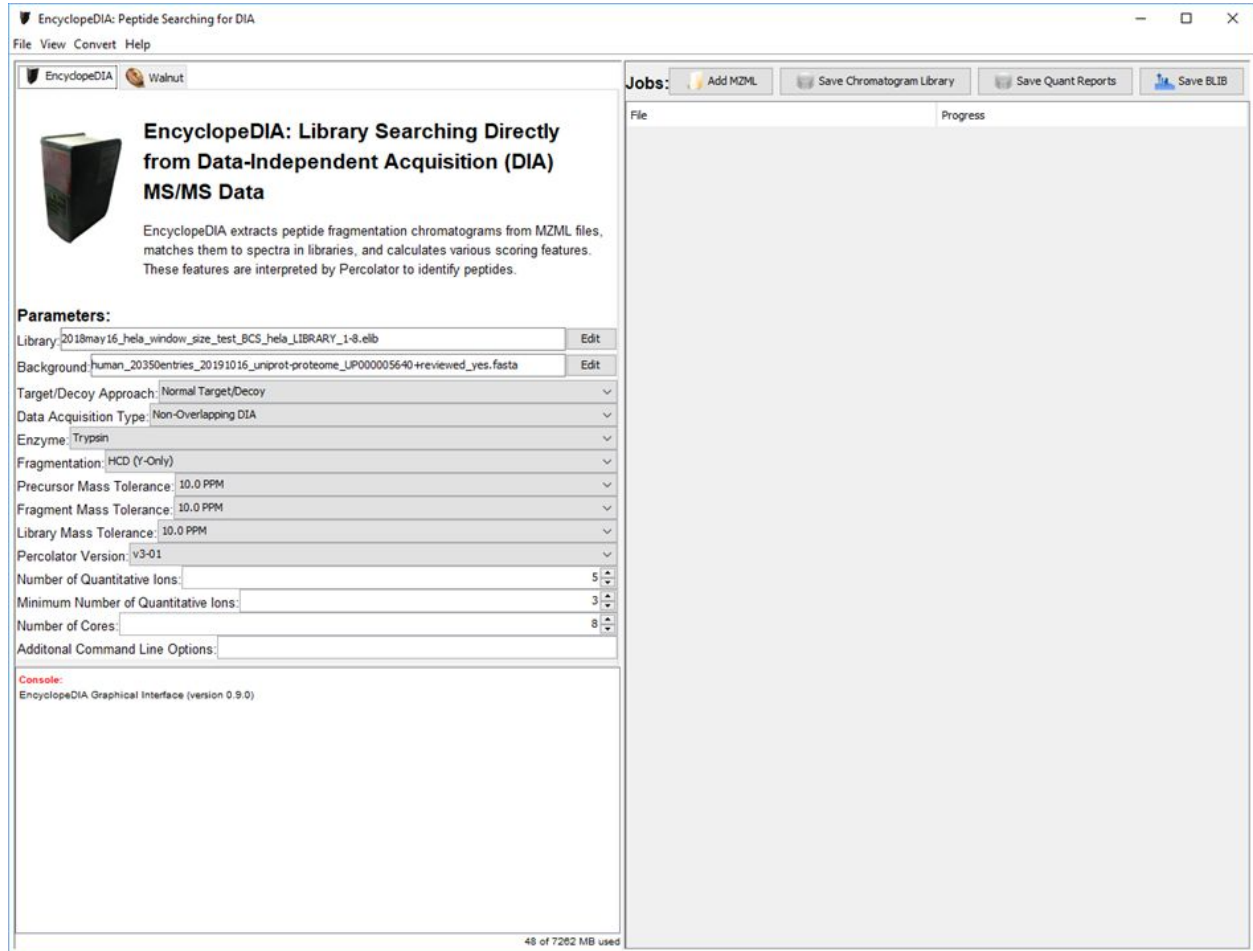
4q. File > Export > Report...

4r. Edit list... > Add...

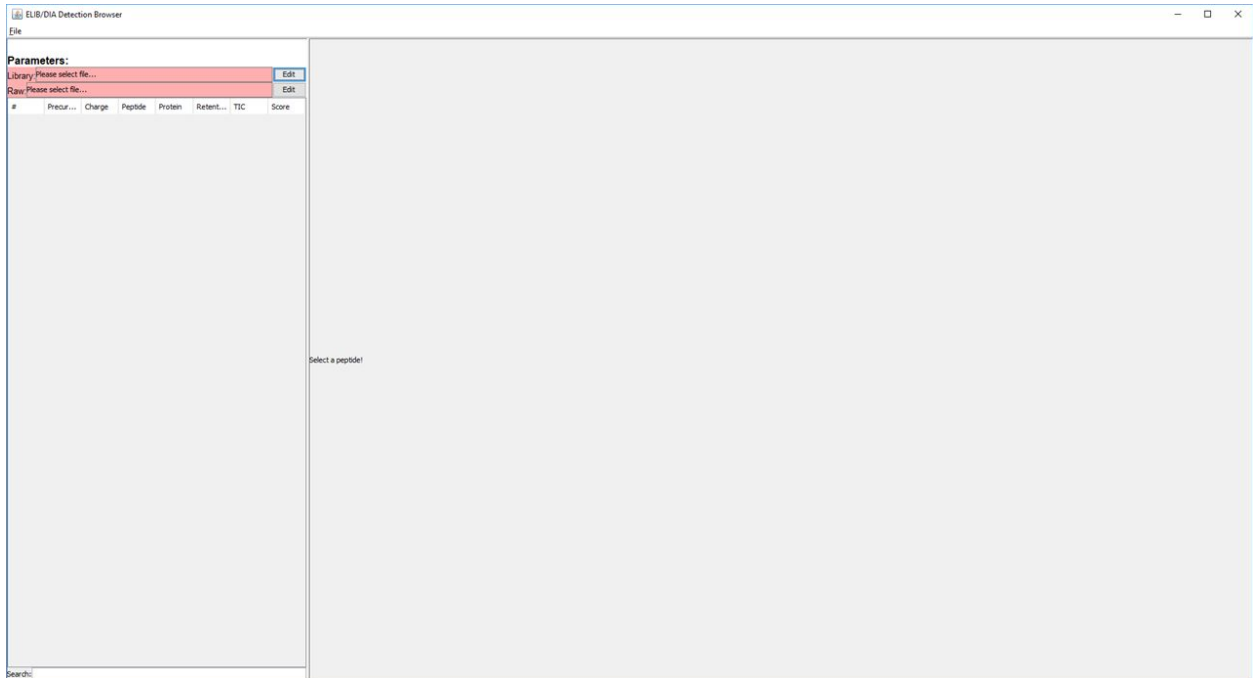
4s. Make sure to name your report ("View Name:"), and check the fields you want the report to include. Binoculars are at the top left, next to the Redo arrow, and will find fields by name.

EncyclopeDIA data visualizers

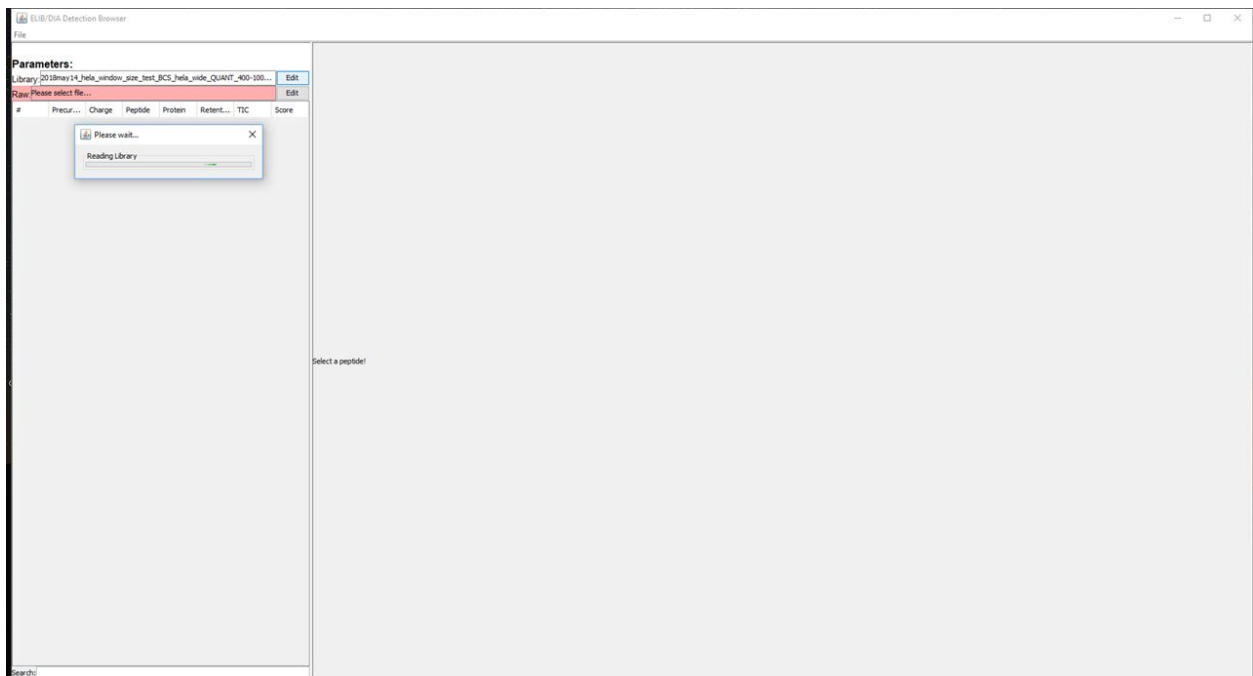
EncyclopeDIA ELIB Browser



The top left under "View", select the "Launch ELIB Browser" option

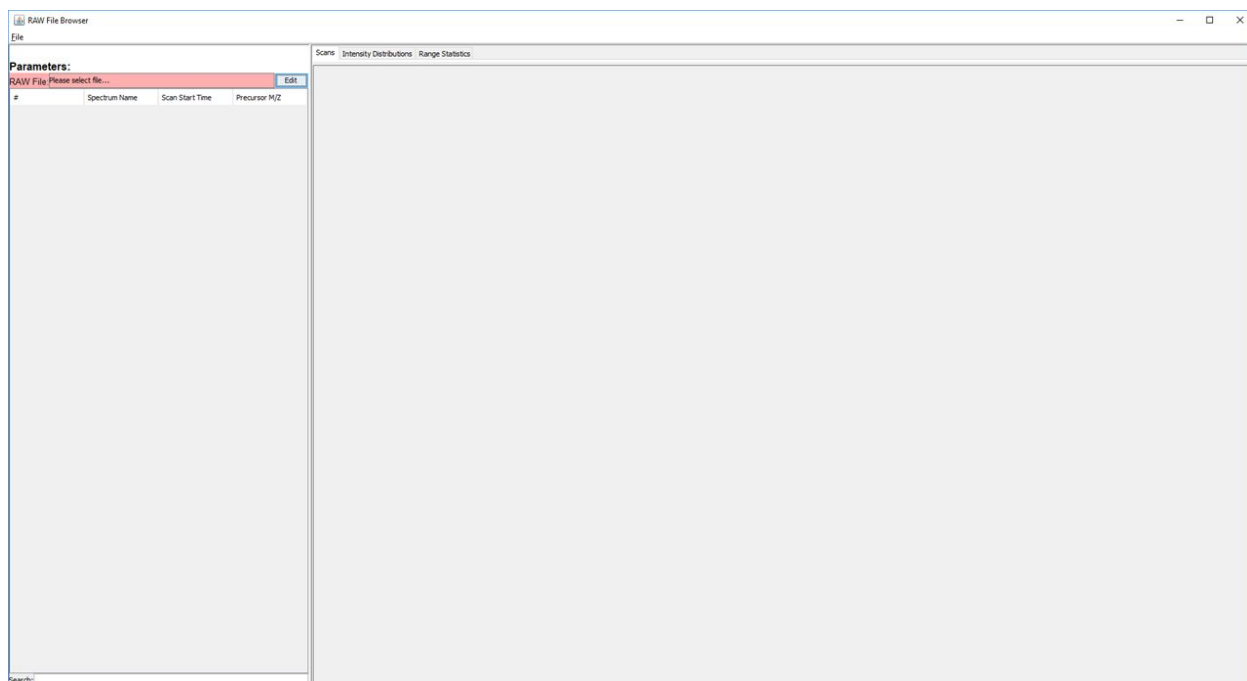


Next to the “Library” field, click the “Edit” button and navigate the explorer to the “Save Chromatogram Library” file you just saved. Once the library has loaded, you should have a populated target list like this:

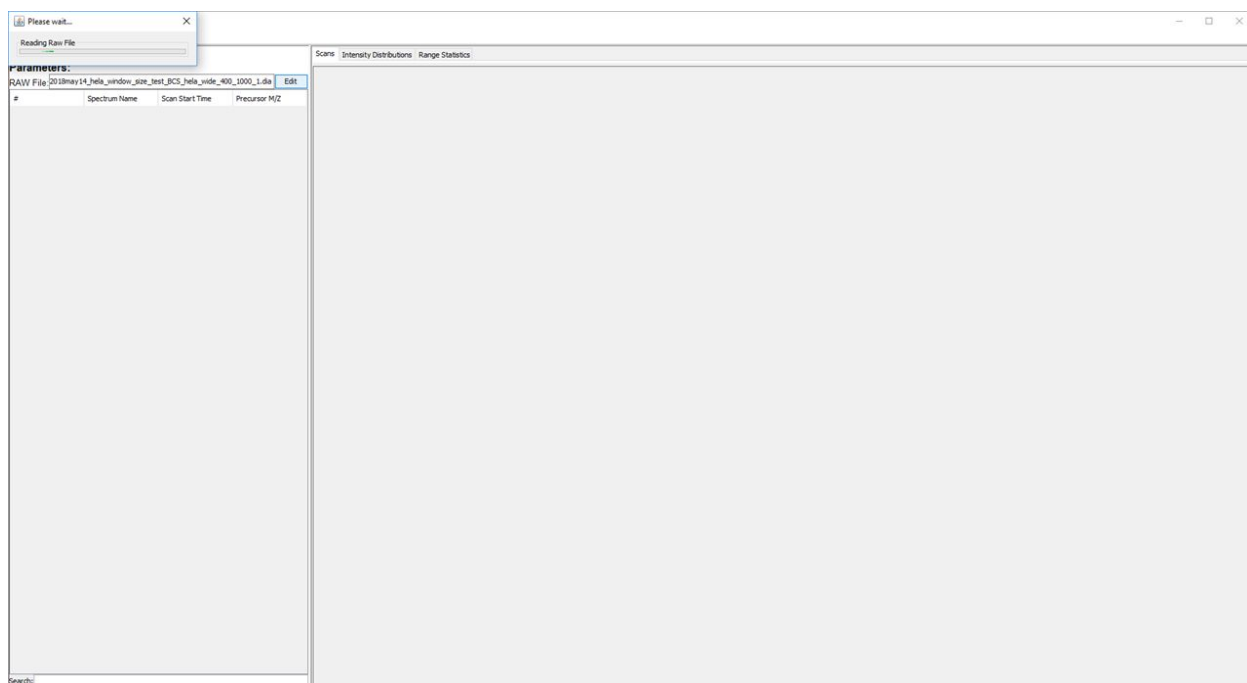


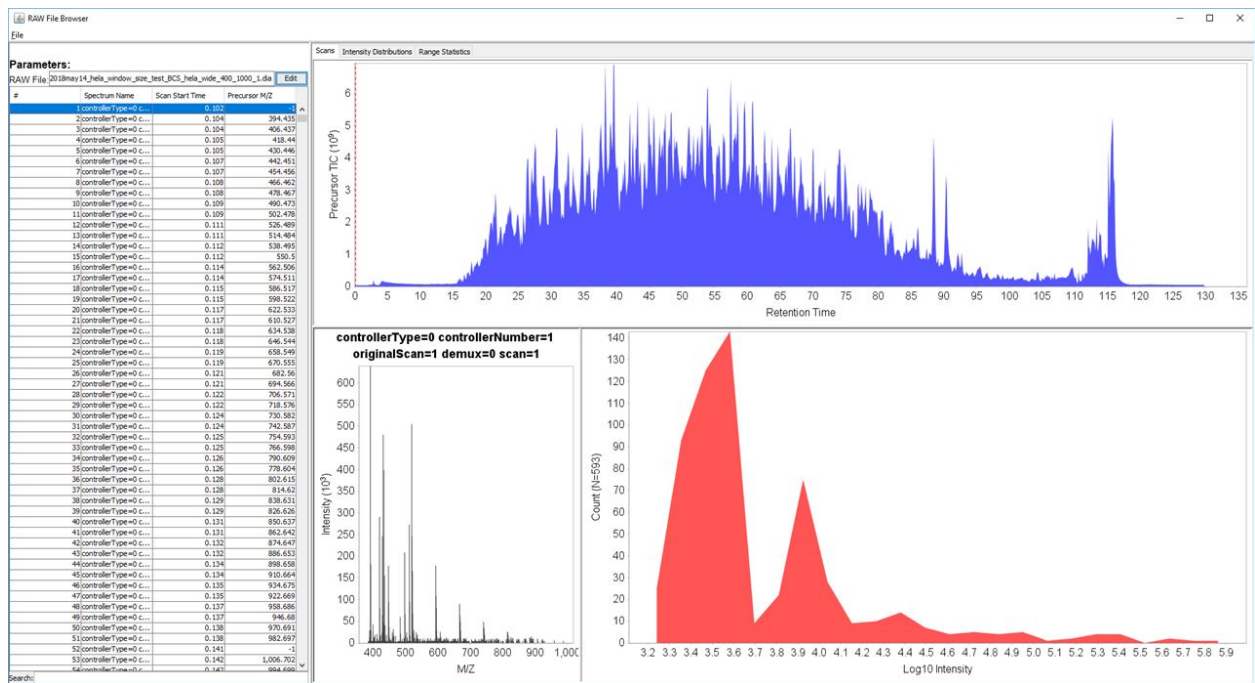
EncyclopeDIA RAW File Browser

Top left under “View”, select “Launch RAW File Browser”



Select a “RAW” file that has been previously analyzed with the Encyclopedia suite (Walnut or Encyclopedia)





The target list on the left lists the Scan Number (#), SpectrumName, Scan Start Time, and Precursor M/Z.

On the top above the graphics, there are three tabs. The “Range Statistics” gives some valuable information about the DIA method like the Ion Inject Time across each precursor window and across retention time bin:

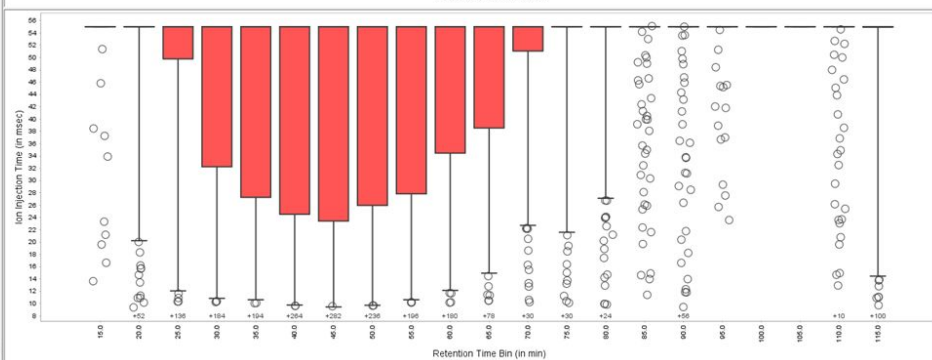
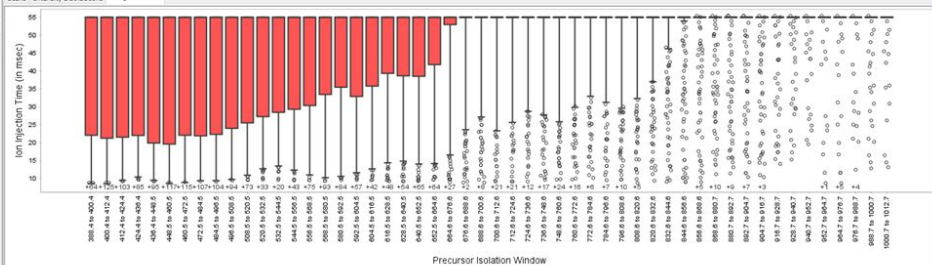
Parameters:

RAW File [2018may14_hela_widom_size_test_BCS_hela_wide_400_1000_1da]

#	Spectrum Name	Scan Start Time	Precursor M/Z
1	controller Type=0 c...	0.102	7.1
2	controller Type=0 c...	0.104	394.435
3	controller Type=0 c...	0.104	406.432
4	controller Type=0 c...	0.105	418.44
5	controller Type=0 c...	0.105	430.446
6	controller Type=0 c...	0.107	442.451
7	controller Type=0 c...	0.107	454.456
8	controller Type=0 c...	0.108	466.462
9	controller Type=0 c...	0.108	478.467
10	controller Type=0 c...	0.109	490.473
11	controller Type=0 c...	0.109	502.478
12	controller Type=0 c...	0.111	526.489
13	controller Type=0 c...	0.111	514.484
14	controller Type=0 c...	0.112	538.495
15	controller Type=0 c...	0.112	550.5
16	controller Type=0 c...	0.114	562.506
17	controller Type=0 c...	0.114	574.511
18	controller Type=0 c...	0.115	586.517
19	controller Type=0 c...	0.115	598.522
20	controller Type=0 c...	0.117	622.532
21	controller Type=0 c...	0.117	634.537
22	controller Type=0 c...	0.118	634.538
23	controller Type=0 c...	0.118	646.544
24	controller Type=0 c...	0.119	658.549
25	controller Type=0 c...	0.119	670.555
26	controller Type=0 c...	0.121	682.56
27	controller Type=0 c...	0.121	694.566
28	controller Type=0 c...	0.122	706.571
29	controller Type=0 c...	0.122	718.576
30	controller Type=0 c...	0.124	730.582
31	controller Type=0 c...	0.124	742.587
32	controller Type=0 c...	0.125	754.593
33	controller Type=0 c...	0.125	766.598
34	controller Type=0 c...	0.126	790.609
35	controller Type=0 c...	0.126	778.604
36	controller Type=0 c...	0.128	802.615
37	controller Type=0 c...	0.128	814.62
38	controller Type=0 c...	0.129	838.631
39	controller Type=0 c...	0.129	826.626
40	controller Type=0 c...	0.131	850.637
41	controller Type=0 c...	0.131	862.642
42	controller Type=0 c...	0.132	874.647
43	controller Type=0 c...	0.132	886.653
44	controller Type=0 c...	0.134	898.658
45	controller Type=0 c...	0.134	910.664
46	controller Type=0 c...	0.135	934.675
47	controller Type=0 c...	0.135	922.669
48	controller Type=0 c...	0.137	958.686
49	controller Type=0 c...	0.137	946.68
50	controller Type=0 c...	0.138	970.691
51	controller Type=0 c...	0.138	982.697
52	controller Type=0 c...	0.141	1006.707
53	controller Type=0 c...	0.142	1004.702
54	controller Type=0 c...	0.143	1004.696

Search:

Scans Intensity Distributions Range Statistics



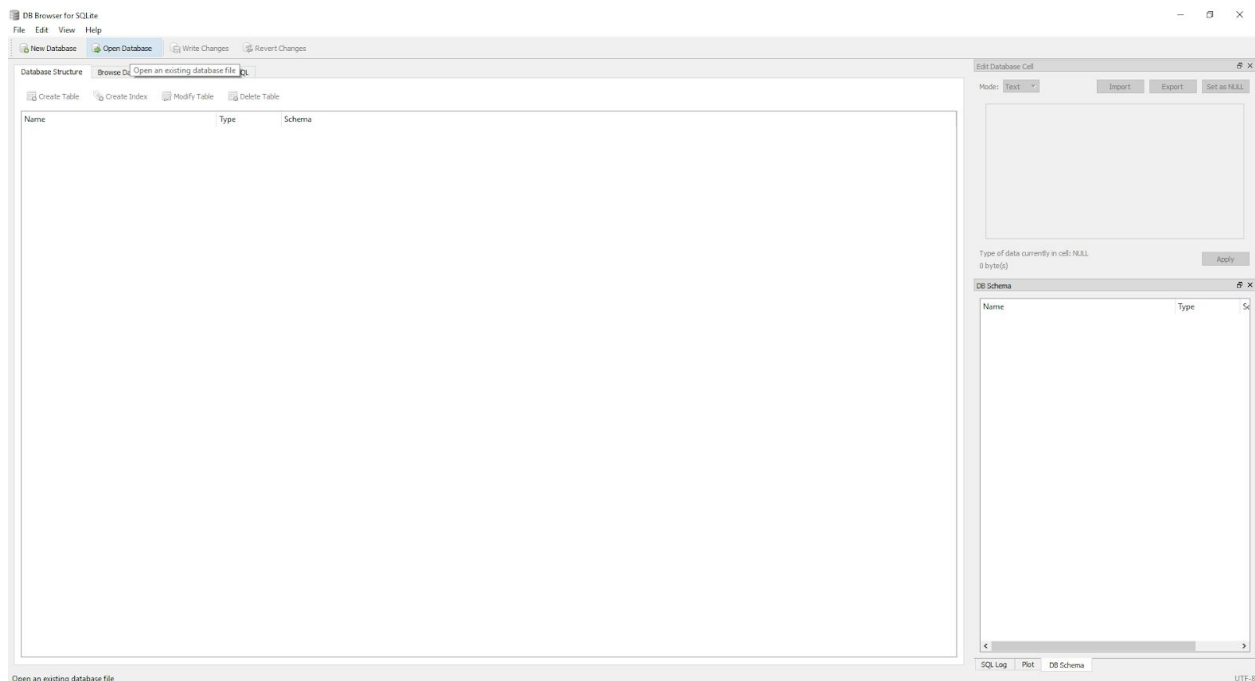
Viewing ELIB files with DB Browser for SQLite

The *.elib files that Encyclopedia builds are SQL databases, which are kind of like multi-tab Excel files but fancier. To view these files and browse the data stored in the elib, follow the steps below:

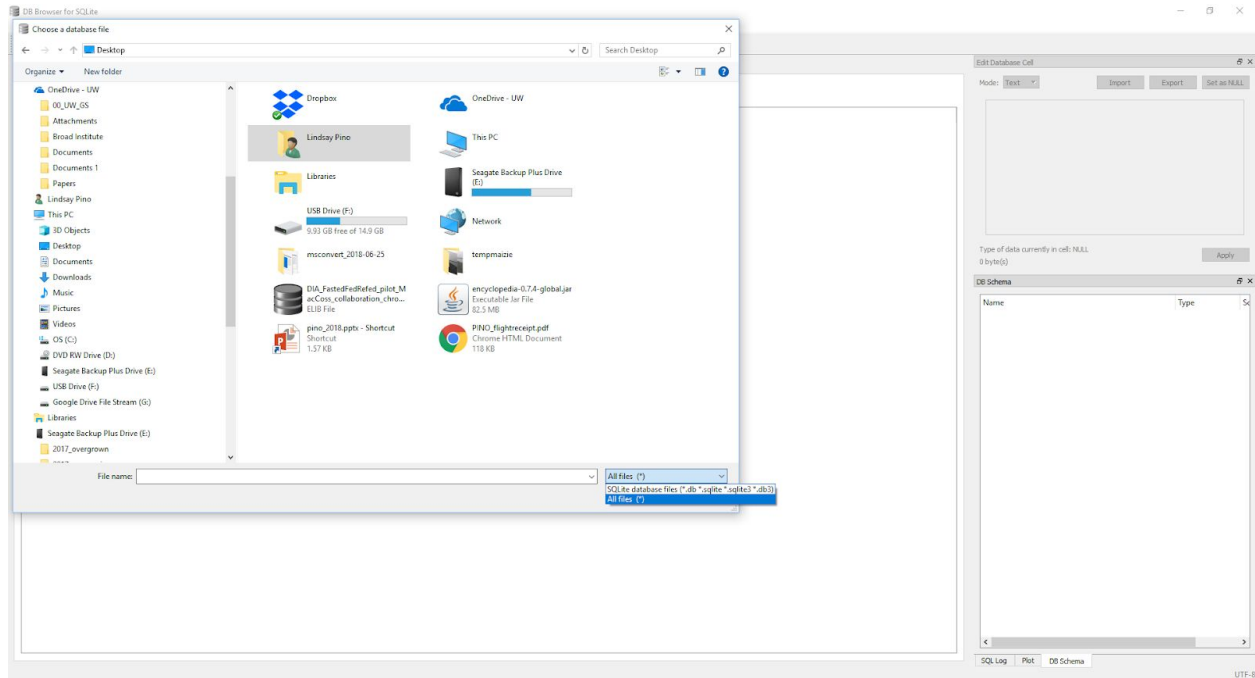
Download DB Browser for SQLite here: <https://sqlitebrowser.org/>

Open DB Browser for SQLite

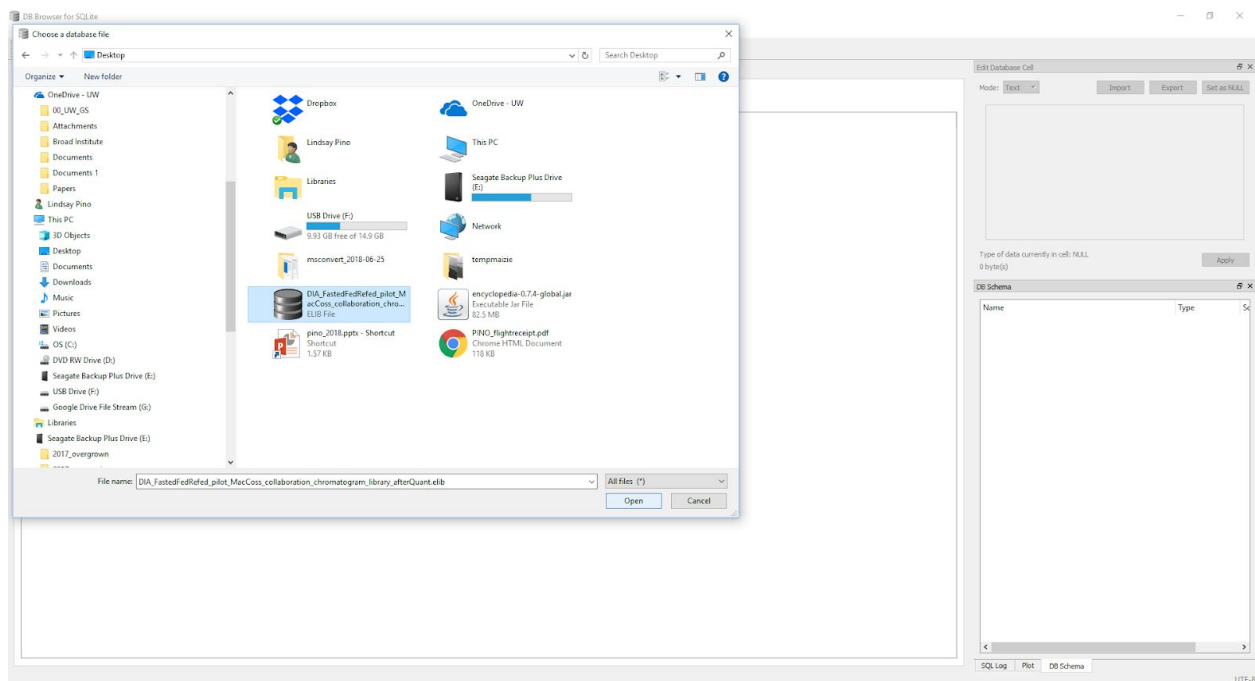
Navigate to “Open Database” button on top left



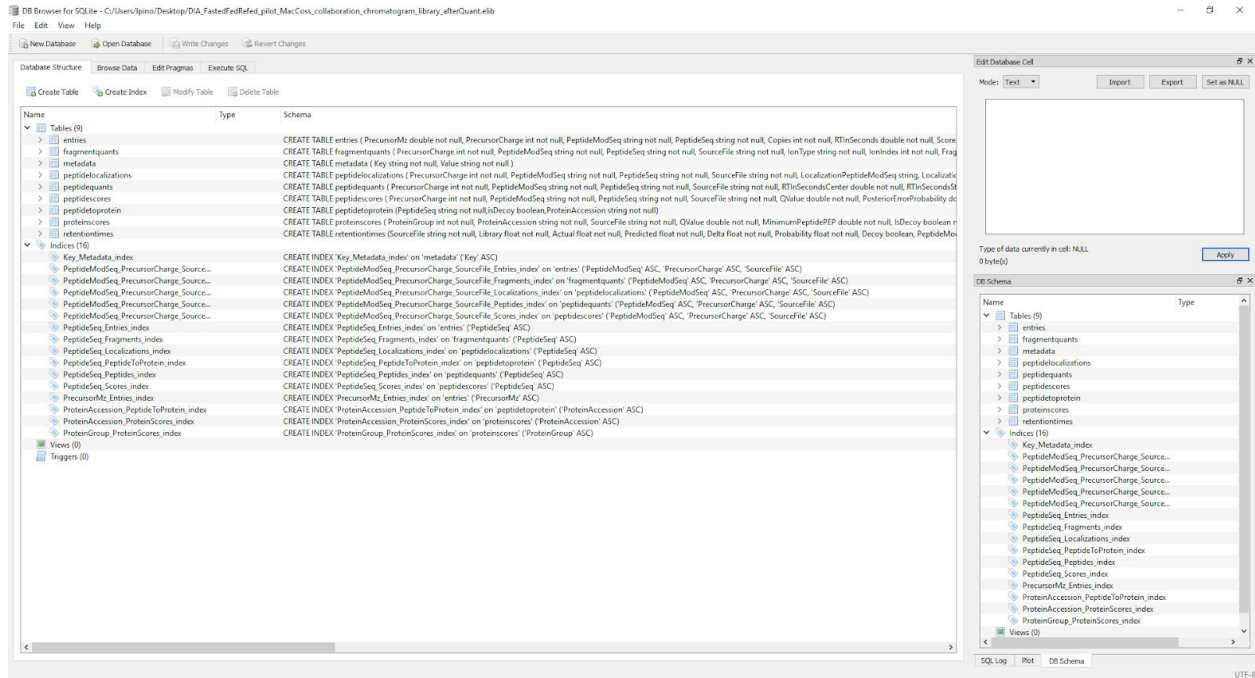
At the bottom of the “Choose a database file” pop-up next to the “File Name” field, select “All files” from the file type dropdown



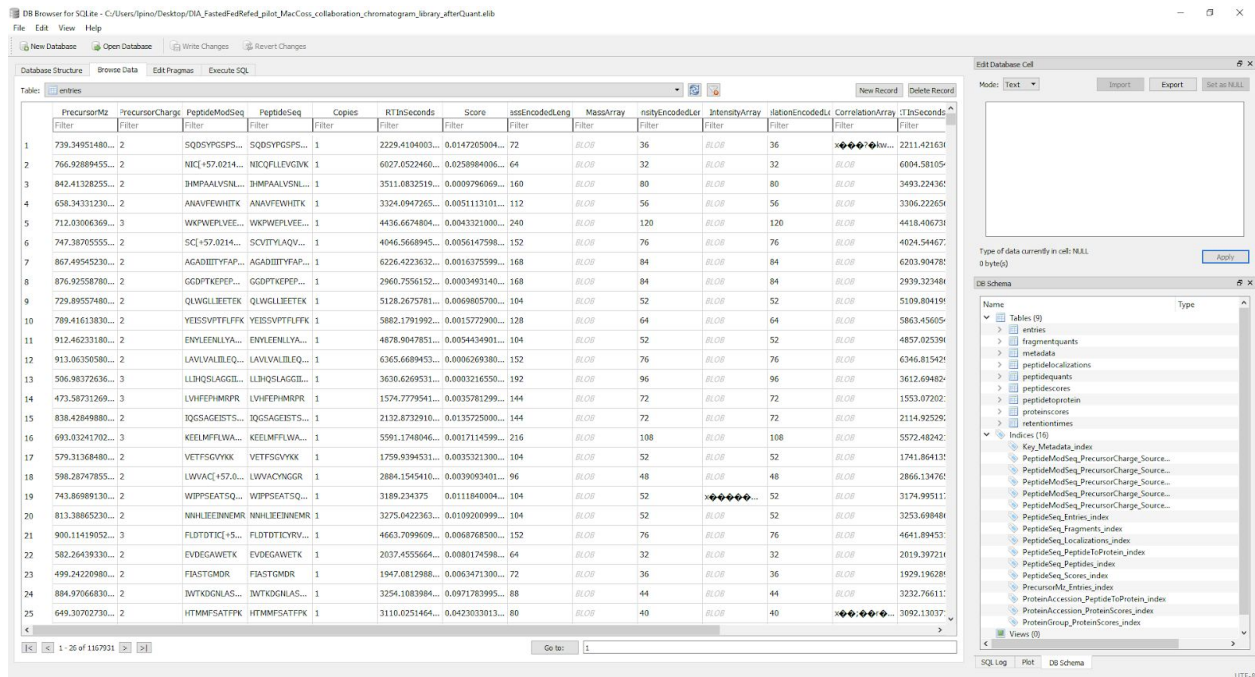
Navigate to the appropriate directory and select the elib file you want to view. Click .Open”



! It might take a minute to load. The window should look like this once it's loaded:



Click on “Browse Data” tab at top under the “Open Database” button.



Under the “Browse Data” tab, select the table you want to view from the “Table:” dropdown menu. For example, the meta data:

DB Browser for SQLite - C:\Users\lpsino\Desktop\DNA_FastedFedRefed_pilot_MacCoss_collaboration_chromatogram_library_afterQuant.elb

File Edit View Help

New Database Open Database Write Changes Revert Changes

Database Structure Browse Data Edit Pragma Execute SQL

Tables:

- entries
- fragmentquants
- metadata
- peptidecalibrations
- peptidequants
- peptidescores
- peptideprotein
- proteinscores
- retentionscores

Filter	IntensityArray	IsIonenEncoded	CorrelationArray	TimeSeconds
BL0B	36	BL0B	◆◆◆◆◆Inr...	2211.421631
BL0B	32	BL0B		4094.58105
BL0B	80	BL0B		3493.224361
BL0B	56	BL0B		3306.222654
BL0B	120	BL0B		4418.406731
BL0B	76	BL0B		4624.54467
BL0B	84	BL0B		6203.90478
BL0B	84	BL0B		2939.323484
BL0B	52	BL0B		5109.804191
BL0B	64	BL0B		5863.45605
BL0B	52	BL0B		4857.025391
BL0B	76	BL0B		6346.815421
BL0B	96	BL0B		3612.69482
BL0B	72	BL0B		1553.07202
BL0B	72	BL0B		2214.92529
BL0B	108	BL0B		5572.48242
BL0B	52	BL0B		1741.86413
BL0B	48	BL0B		2866.13476
BL0B	52	BL0B		3174.99511
BL0B	52	BL0B		3253.69948
BL0B	76	BL0B		4641.89453
BL0B	32	BL0B		2019.397214
BL0B	36	BL0B		1929.19628
BL0B	44	BL0B		3232.76611
BL0B	40	BL0B	◆◆◆◆◆	3092.13037

1 - 26 of 116791

Go to: 1

SQL Log Plot DB Schema UTF-8

Mode: Text Import Export Set as NULL

Type of data currently in cell: NULL
0 byte(s)

DB Schema

Name Type

- Tables (9)
 - entries
 - fragmentquants
 - metadata
 - peptidecalibrations
 - peptidequants
 - peptidescores
 - peptideprotein
 - proteinscores
 - retentionscores
- Indices (16)
 - Key_Metadata_index
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...

Choosing the "metadata" table will display information such as the parameter settings when Encyclopedia was run, the TIC for each raw file in the chromatogram library/quant report, etc

DB Browser for SQLite - C:\Users\lpsino\Desktop\DNA_FastedFedRefed_pilot_MacCoss_collaboration_chromatogram_library_afterQuant.elb

File Edit View Help

New Database Open Database Write Changes Revert Changes

Database Structure Browse Data Edit Pragma Execute SQL

Tables:

- entries
- fragmentquants
- metadata
- peptidecalibrations
- peptidequants
- peptidescores
- peptideprotein
- proteinscores
- retentionscores

Filter	Value
-filterPeaklists	false
-acquisition	DIA
-enzyme	Trypsin
-expectedPeakWidth	25
-fixed	
-offset	0
-frag	HCD
-ftol	10
-getNumberOfDraDecoyLibrariesSearched	0
-ftol	10
-localizationModification	none
-minIntensity	-1
-minimumOfQuantitativePeaks	3
-numberOfQuantitativePeaks	5
-numberOfThreadsUsed	12
-percolatorThreshold	0.01
-percolatorVersionNumber	3
-poffset	0
-precursorWindowSize	-1
-ptol	10
-quantifyAcrossSamples	true
-rWindowMin	-1
-scoringBreadthType	window
-targetWindowCenter	-1
-verifyModifications	true
20181120_LUMI_C8A_EASY04_053_30_SA_90minrad_808_DIA_400-1000_8mazl_15k_20HT_4e5acc_1601-01_01.ms.MS_library	DIA_FastedFe...

1 - 26 of 126

Go to: 1

SQL Log Plot DB Schema UTF-8

Mode: Text Import Export Set as NULL

Type of data currently in cell: NULL
0 byte(s)

DB Schema

Name Type

- Tables (9)
 - entries
 - fragmentquants
 - metadata
 - peptidecalibrations
 - peptidequants
 - peptidescores
 - peptideprotein
 - proteinscores
 - retentionscores
- Indices (16)
 - Key_Metadata_index
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...
 - PeptideModSeq_PrecursorCharge_Source...