MetaboKit

October 29, 2020

1 Installation

Executables for all three major platforms (OS X,Windows,linux) are included.

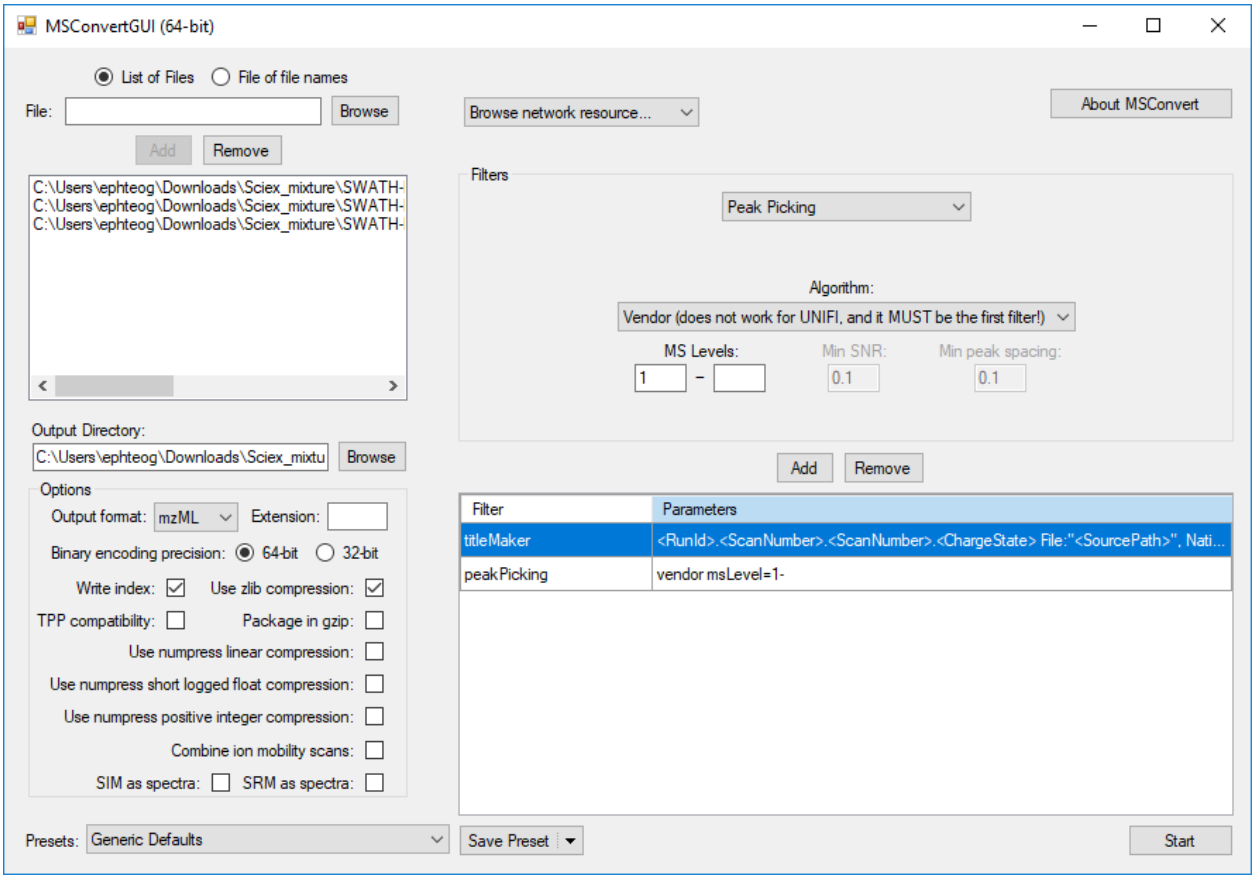
2 Raw Data Conversion into mzML

The input files need to be in mzML format for MetaboKit. MSConvert, provided through the ProteoWizard software suite, enables conversion of proprietary raw data files (.wiff for SCIEX, .raw files for Thermo Fisher instruments) into mzML.

1. Download and install a recent version ProteoWizard on a Windows computer.

2. Start the software by opening the Start Menu, type "MSConvert" in the search field and click on "MSConvert".

3. In the MSConvert window, use the "Browse" button to select the raw files to be converted (the .wiff and .wiff.scan files need to be located in the same folder). Select "mzML" for "Output format", "64-bit" for "Binary encoding precision", check "Write index" and "Use zlib compression" checkbox. "Peak Picking" filter should be added (see figure below).



3 Input Parameters

Input parameters must be specified according to the user's preference for various aspects of the data processing. Input parameters are to be specified in the "param.txt" file.

MetaboKit requires that the user provide the following options for data processing:

- **mzML_files**: Names of the centroided mzML files. Use "*.mzML" to select all mzML files in the folder.
- **num_threads**: Number of threads to use
- **length_of_ion_chromatogram**: range of ion chromatogram of feature in seconds
- **MS2_score**: Minimum dot product score for match between library spectra and experimental spectra
- **ms1_ppm**: MS1 ppm tolerance (e.g. 15)
- **ms2_ppm**: MS2 ppm tolerance (e.g. 30)
- **library**: External libraries to be used for analysis

- **ISF_rt_diff**: Retention time difference between parent and fragment
- **RT_shift**: Maximum allowable RT shift
- **min_peaks**(DDA only): Minimum number of matching peaks between library spectra and experimental spectra
- **ISF_score**(DDA only): Minimum dot product score for match between parent spectra and fragment spectra
- **adduct**(DDA only): Adducts to be considered for identification in the data
- **pfcor**(DIA only): precursor-fragment ion chromatogram correlation threshold
- **rt_diff**(DIA only): Retention time difference between library entry and feature in the DIA sample
- **topNfrag**(DIA only): Number of high intensity fragments in library entry to score against DIA sample
- **ms2_auc_w/o_feature**(DIA only): Fragment-level quantitation if MS1 feature not detected
- **window_setting**(DIA only): DIA MS/MS window setting

4 Analysis Output Table

- **quant_...All.txt** A table of MS1-based or MS2-based quantification.
- **ann_...All.txt**(DDA only) This file contains the information of compounds identified and corresponding spectra. This file will be used as rt-annotated library in SWATH DIA extraction
- **una_....mzML.txt**(DDA only) This gives information of non-matched spectra

5 Example

The example datasets are located in the "example" folder.

5.1 Library generation from DDA and DDA quantitation

The data folder with all DDA mzML files and param.txt file.

```
C:\Users\ephteog\Downloads\standards_example>dir
 Volume in drive C is WINDOWS
 Volume Serial Number is 805C-0F96

 Directory of C:\Users\ephteog\Downloads\standards_example

29/08/2019  12:51 PM    <DIR>          .
29/08/2019  12:51 PM    <DIR>          ..
31/07/2019  12:51 PM               886 param.txt
05/07/2019  03:12 PM       126,755,018 SWATH-POS-MODE-PathwayMetabolites_PosIDA-1.wiff.mzML
05/07/2019  03:12 PM       127,658,748 SWATH-POS-MODE-PathwayMetabolites_PosIDA-2.wiff.mzML
05/07/2019  03:13 PM       128,593,806 SWATH-POS-MODE-PathwayMetabolites_PosIDA-3.wiff.mzML
               4 File(s)    383,008,458 bytes
               2 Dir(s)  1,264,493,162,496 bytes free

C:\Users\ephteog\Downloads\standards_example>
```

First, run the feature detection executable.

```
C:\Users\ephteog\Downloads\standards_example>..\metabokit_dda\DDAfeature.exe
```

Next, MS/MS spectra scoring.

```
C:\Users\ephteog\Downloads\standards_example>..\metabokit_dda\DDAscore.exe
```

Next, features and annotations are aligned between samples.

```
C:\Users\ephteog\Downloads\standards_example>..\metabokit_dda\DDAalign.exe
```

5.2 DIA

The data folder should contain all DIA mzML files and param.txt file, in addition copy "ann_...All.txt" from DDA output to this folder.
First, run the feature detection executable.

```
C:\Users\ephteog\Downloads\calibration_example>..\metabokit_dia\DIAfeature.exe
```

Next, MS/MS spectra scoring.

```
C:\Users\ephteog\Downloads\calibration_example>..\metabokit_dia\DIAscore.exe
```

Next, features and annotations are aligned between samples.

```
C:\Users\ephteog\Downloads\calibration_example>..\metabokit_dia\DIAalign.exe
```

6 Example “param.txt”

6.1 DDA

```
mzML_files
*.mzML

num_threads
3

length_of_ion_chromatogram
3 50

ISF_rt_diff
2

ISF_score
.5

MS2_score
.5

min_peaks
2

RT_shift
10

library
msdial
LipidBlast
LipidBlast-fork

adduct
M     0 1+
M+H       1.007825 1+
M+2H      2.014102 2+
M+Na      22.98977 1+
M+NH4     18.03437 1+
M+K     38.96371 1+
M+Li      7.016004 1+
#M-H      -1.007825 1-
```

```
#M-2H     -2.014102 2-
#M+HCOO    44.99765 1-
#M+Na-2H  20.97412 1-
#M+Cl   34.96885 1-


ms1_ppm
15

ms2_ppm
30
```

6.2 DIA

```
mzML_files
*.mzML

num_threads
3

length_of_ion_chromatogram
3 50

library
metabokit

ms1_ppm
15

ms2_ppm
30

ISF_rt_diff
2

rt_diff
10

topNfrag
6
```

```
MS2_score
0.5

pfcor
0.5

ms2_auc_w/o_feature
1

RT_shift
10

#30
window_setting
40 70
69 100
99 130
129 160
159 190
189 220
219 250
249 280
279 310
309 340
339 370
369 400
399 430
429 460
459 490
489 520
519 550
549 580
579 610
609 640
639 670
669 700
699 730
729 760
759 790
789 820
819 850
849 880
```

```
879 910
909 940
939 970
969 1000
```

7 Copyrights

BSD 3-Clause License

Copyright (c) 2020, Guoshou Teo All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.