MRMkit

August 11, 2020

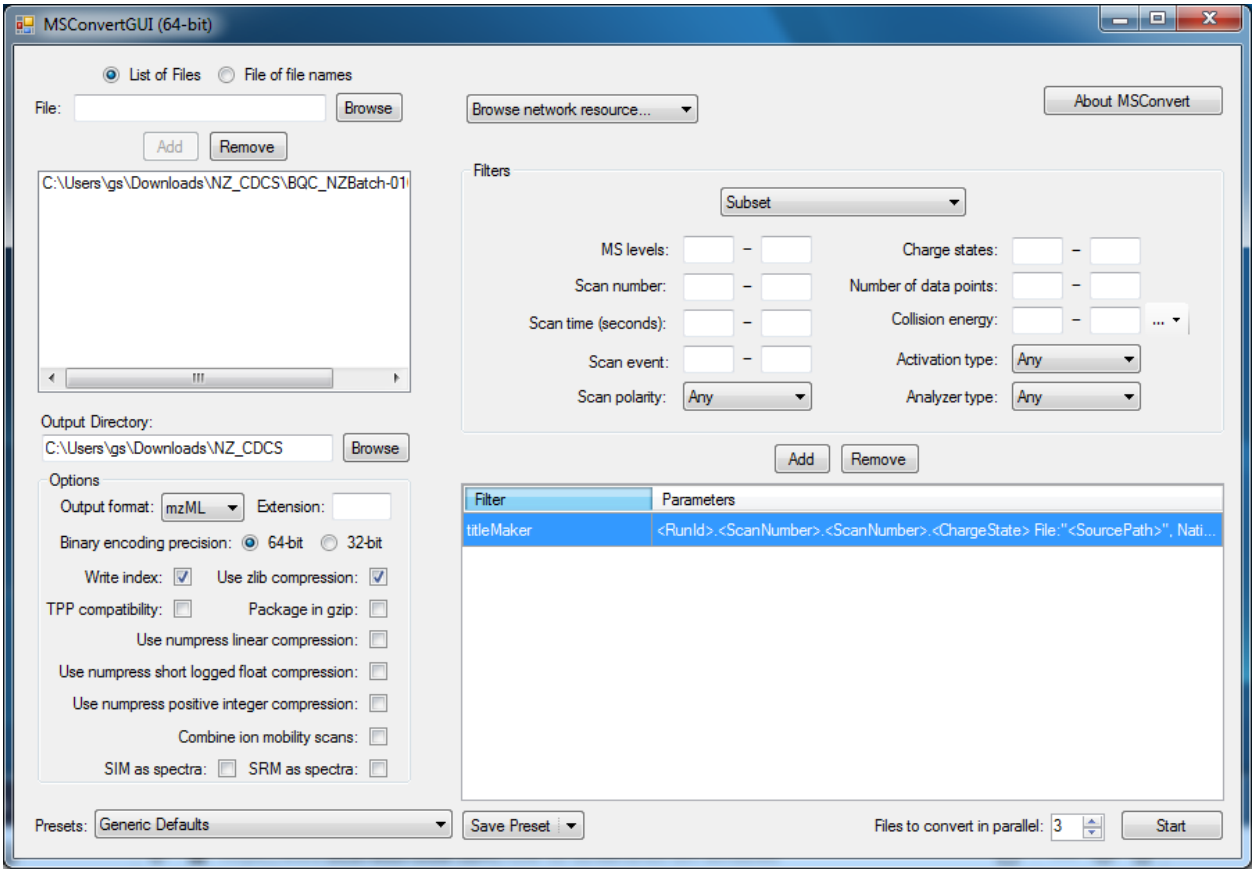
1 Raw Data Conversion into mzML

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

1. Download and install a recent version ProteoWizard on a Windows computer (the versions for other OS platforms do not work).

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 (see figure below).



2 Input Parameters

Input parameters must be specified according to the user's preference in order to control various aspects of the data processing. Input parameters are to be specified in the "param.txt" file. MRMkit requires that the user provide the following options for data processing:

- **mzML_files**: Names of the mzML files. Use "*.mzML" to select all mzML files in the current working directory.
- **batch_info**: analytical sequence of samples with indication of metabolite extraction / sample preparation batches. An example shown in Figure 1.
- **length_of_ion_chromatogram**: range of ion chromatogram to be detected in each MRM window (in seconds)
- **transition_list**: assay information with compound name, Q1, Q3, retention time and the name of the designated internal standards (ISTD). An example shown in Figure 2.
- **ISTD_trace_all**: Set "1" to use the peak shape of the corresponding ISTD for peak integration for all transitions, "0" to directly take the area under peak for transitions

with unimodal chromatograms. Peak areas will still be extracted based on the peak shape of ISTD for multi-modal ion chromatograms

- **batch_correction**: Set "1" to enable time trend and batch effect correction, "0" to disable.
- **smoothing_sigma**: Optional parameter. A smaller number will result in more smoothing correction. Default is set to 10.
- **RT_difference**: Optional parameter. This will restrict the search space for peaks to within ± the given RT value in **transition_list**.

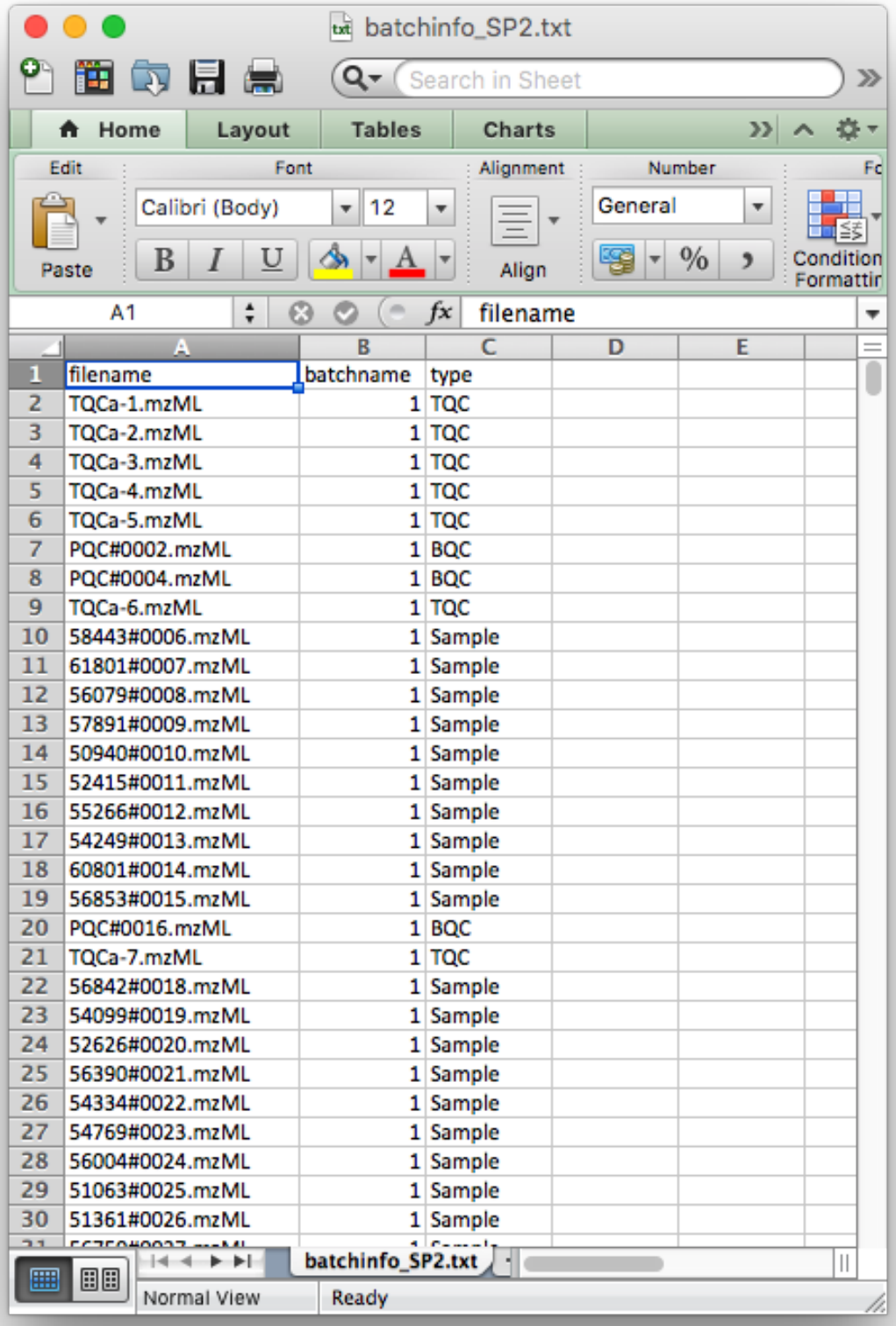
3 Analysis Output Table

- **batch_adjusted.txt**: Quantitative data (compound peak area / ISTD) after correction of time trends and batch effects
- **quant_auc_rt.txt**: Integrated peak areas for target analytes and ISTDs
- **BQCtable.txt**: Batch specific and overall coefficient of variation using batch pooled QC samples (a pooled sample with metabolite extraction performed along with the samples from a particular batch)
- **TQCtable.txt**: Batch specific and overall coefficient of variation using technical QC samples (a pooled sample formed at the beginning of the entire analytical sequence, intended to capture instrument variation only)
- **run_seq.pdf**: Visualization of concentration data before and after time trend and batch effect correction, per compound basis (produced from R script)
- **IonChromatogram**: A folder containing ion chromatogram images with integrated peaks across all samples, per transition
- **peak_picking.txt**: The peak picking table, refer to next section.
- **quant_table.txt**: The final quantification table after the user indicates the desired peak in "peak_picking.txt".

4 User's peak picking step

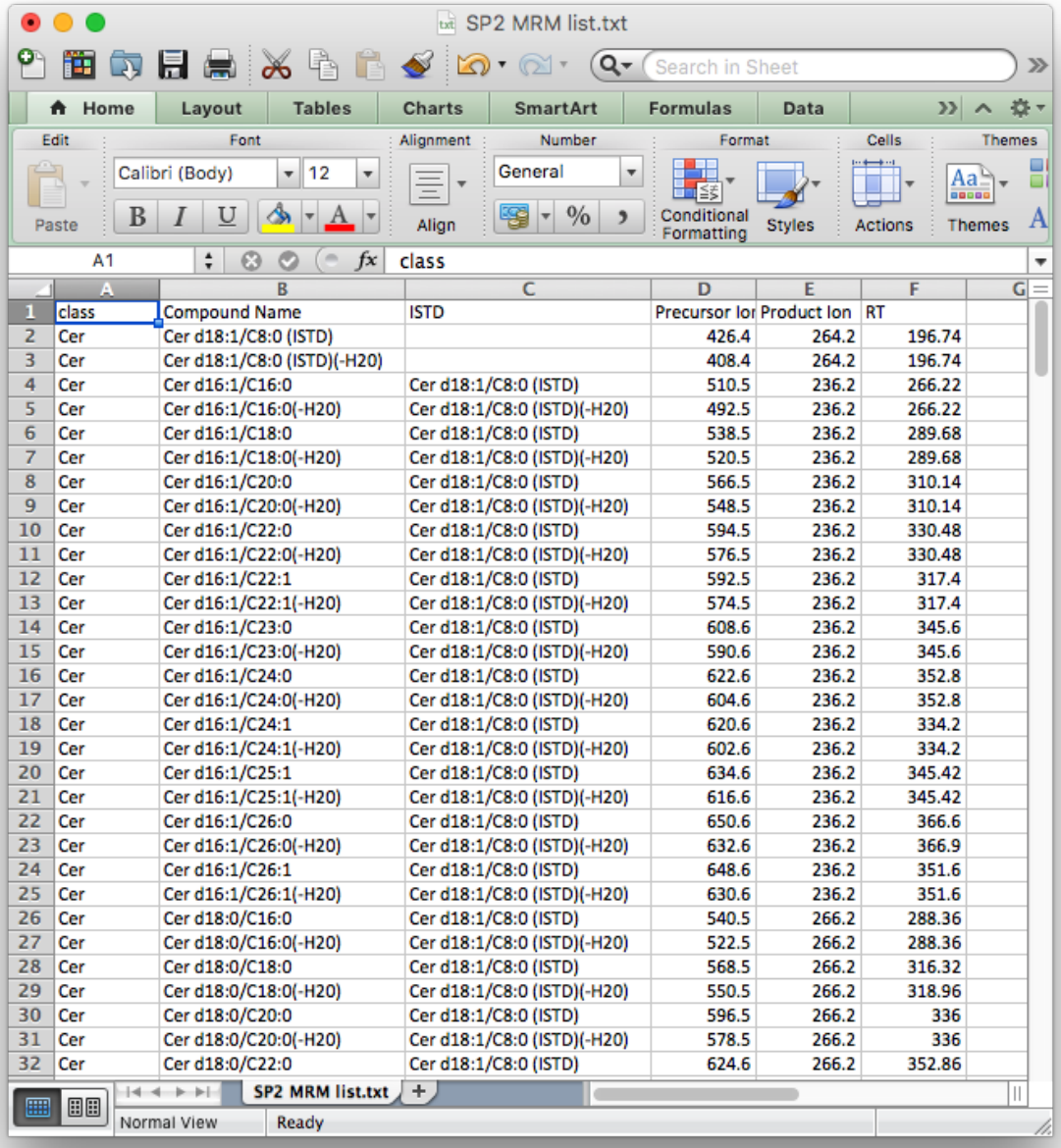
Upon running the analysis, MRMkit produces the aforementioned output files, but it also reports a peak picking table. This step is necessary: for transitions with a multi-modal ion chromatogram, the user must specify which peak he/she wishes to use to quantify the target compound. To facilitate the process, the template file shows the TQC/BQC-based

Figure 1: **batch_info** file



	A	B	C	D	E
1	filename	batchname	type		
2	TQCa-1.mzML	1	TQC		
3	TQCa-2.mzML	1	TQC		
4	TQCa-3.mzML	1	TQC		
5	TQCa-4.mzML	1	TQC		
6	TQCa-5.mzML	1	TQC		
7	PQC#0002.mzML	1	BQC		
8	PQC#0004.mzML	1	BQC		
9	TQCa-6.mzML	1	TQC		
10	58443#0006.mzML	1	Sample		
11	61801#0007.mzML	1	Sample		
12	56079#0008.mzML	1	Sample		
13	57891#0009.mzML	1	Sample		
14	50940#0010.mzML	1	Sample		
15	52415#0011.mzML	1	Sample		
16	55266#0012.mzML	1	Sample		
17	54249#0013.mzML	1	Sample		
18	60801#0014.mzML	1	Sample		
19	56853#0015.mzML	1	Sample		
20	PQC#0016.mzML	1	BQC		
21	TQCa-7.mzML	1	TQC		
22	56842#0018.mzML	1	Sample		
23	54099#0019.mzML	1	Sample		
24	52626#0020.mzML	1	Sample		
25	56390#0021.mzML	1	Sample		
26	54334#0022.mzML	1	Sample		
27	54769#0023.mzML	1	Sample		
28	56004#0024.mzML	1	Sample		
29	51063#0025.mzML	1	Sample		
30	51361#0026.mzML	1	Sample		

Figure 2: **transition_list** file


SP2 MRM list.txt

Search in Sheet

Home | Layout | Tables | Charts | SmartArt | Formulas | Data

Edit | Font | Alignment | Number | Format | Cells | Themes

Paste | Calibri (Body) | 12 | Align | General | Conditional Formatting | Styles | Actions | Themes

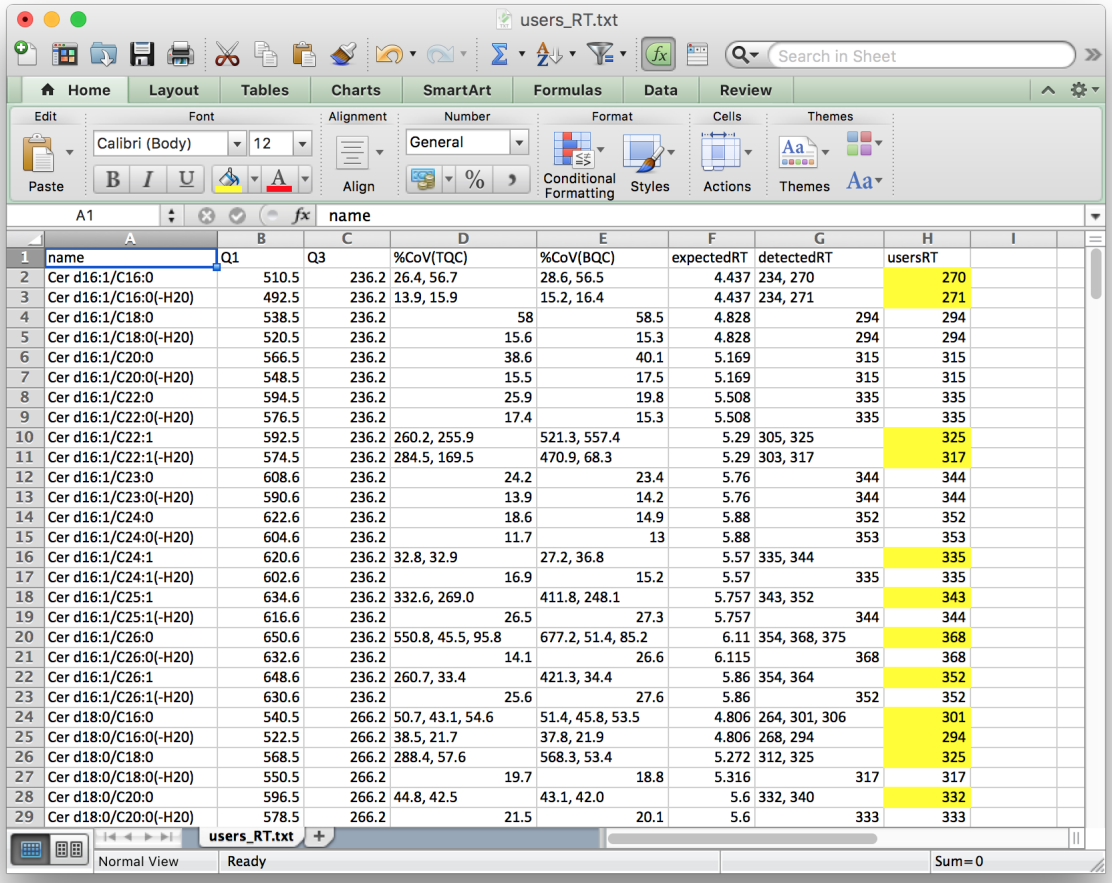
A1 | class

	A	B	C	D	E	F
1	class	Compound Name	ISTD	Precursor Ion	Product Ion	RT
2	Cer	Cer d18:1/C8:0 (ISTD)		426.4	264.2	196.74
3	Cer	Cer d18:1/C8:0 (ISTD)(-H20)		408.4	264.2	196.74
4	Cer	Cer d16:1/C16:0	Cer d18:1/C8:0 (ISTD)	510.5	236.2	266.22
5	Cer	Cer d16:1/C16:0(-H20)	Cer d18:1/C8:0 (ISTD)(-H20)	492.5	236.2	266.22
6	Cer	Cer d16:1/C18:0	Cer d18:1/C8:0 (ISTD)	538.5	236.2	289.68
7	Cer	Cer d16:1/C18:0(-H20)	Cer d18:1/C8:0 (ISTD)(-H20)	520.5	236.2	289.68
8	Cer	Cer d16:1/C20:0	Cer d18:1/C8:0 (ISTD)	566.5	236.2	310.14
9	Cer	Cer d16:1/C20:0(-H20)	Cer d18:1/C8:0 (ISTD)(-H20)	548.5	236.2	310.14
10	Cer	Cer d16:1/C22:0	Cer d18:1/C8:0 (ISTD)	594.5	236.2	330.48
11	Cer	Cer d16:1/C22:0(-H20)	Cer d18:1/C8:0 (ISTD)(-H20)	576.5	236.2	330.48
12	Cer	Cer d16:1/C22:1	Cer d18:1/C8:0 (ISTD)	592.5	236.2	317.4
13	Cer	Cer d16:1/C22:1(-H20)	Cer d18:1/C8:0 (ISTD)(-H20)	574.5	236.2	317.4
14	Cer	Cer d16:1/C23:0	Cer d18:1/C8:0 (ISTD)	608.6	236.2	345.6
15	Cer	Cer d16:1/C23:0(-H20)	Cer d18:1/C8:0 (ISTD)(-H20)	590.6	236.2	345.6
16	Cer	Cer d16:1/C24:0	Cer d18:1/C8:0 (ISTD)	622.6	236.2	352.8
17	Cer	Cer d16:1/C24:0(-H20)	Cer d18:1/C8:0 (ISTD)(-H20)	604.6	236.2	352.8
18	Cer	Cer d16:1/C24:1	Cer d18:1/C8:0 (ISTD)	620.6	236.2	334.2
19	Cer	Cer d16:1/C24:1(-H20)	Cer d18:1/C8:0 (ISTD)(-H20)	602.6	236.2	334.2
20	Cer	Cer d16:1/C25:1	Cer d18:1/C8:0 (ISTD)	634.6	236.2	345.42
21	Cer	Cer d16:1/C25:1(-H20)	Cer d18:1/C8:0 (ISTD)(-H20)	616.6	236.2	345.42
22	Cer	Cer d16:1/C26:0	Cer d18:1/C8:0 (ISTD)	650.6	236.2	366.6
23	Cer	Cer d16:1/C26:0(-H20)	Cer d18:1/C8:0 (ISTD)(-H20)	632.6	236.2	366.9
24	Cer	Cer d16:1/C26:1	Cer d18:1/C8:0 (ISTD)	648.6	236.2	351.6
25	Cer	Cer d16:1/C26:1(-H20)	Cer d18:1/C8:0 (ISTD)(-H20)	630.6	236.2	351.6
26	Cer	Cer d18:0/C16:0	Cer d18:1/C8:0 (ISTD)	540.5	266.2	288.36
27	Cer	Cer d18:0/C16:0(-H20)	Cer d18:1/C8:0 (ISTD)(-H20)	522.5	266.2	288.36
28	Cer	Cer d18:0/C18:0	Cer d18:1/C8:0 (ISTD)	568.5	266.2	316.32
29	Cer	Cer d18:0/C18:0(-H20)	Cer d18:1/C8:0 (ISTD)(-H20)	550.5	266.2	318.96
30	Cer	Cer d18:0/C20:0	Cer d18:1/C8:0 (ISTD)	596.5	266.2	336
31	Cer	Cer d18:0/C20:0(-H20)	Cer d18:1/C8:0 (ISTD)(-H20)	578.5	266.2	336
32	Cer	Cer d18:0/C22:0	Cer d18:1/C8:0 (ISTD)	624.6	266.2	352.86

SP2 MRM list.txt

Normal View | Ready

Figure 3: **peak_picking.txt** file, the user will fill in RT of true peak where multiple peaks are detected (highlighted in yellow)



users_RT.txt

	A	B	C	D	E	F	G	H	I
1	name	Q1	Q3	%CoV(TQC)	%CoV(BQC)	expectedRT	detectedRT	usersRT	
2	Cer d16:1/C16:0	510.5	236.2	26.4, 56.7	28.6, 56.5	4.437	234, 270	270	
3	Cer d16:1/C16:0(-H20)	492.5	236.2	13.9, 15.9	15.2, 16.4	4.437	234, 271	271	
4	Cer d16:1/C18:0	538.5	236.2	58	58.5	4.828	294	294	
5	Cer d16:1/C18:0(-H20)	520.5	236.2	15.6	15.3	4.828	294	294	
6	Cer d16:1/C20:0	566.5	236.2	38.6	40.1	5.169	315	315	
7	Cer d16:1/C20:0(-H20)	548.5	236.2	15.5	17.5	5.169	315	315	
8	Cer d16:1/C22:0	594.5	236.2	25.9	19.8	5.508	335	335	
9	Cer d16:1/C22:0(-H20)	576.5	236.2	17.4	15.3	5.508	335	335	
10	Cer d16:1/C22:1	592.5	236.2	260.2, 255.9	521.3, 557.4	5.29	305, 325	325	
11	Cer d16:1/C22:1(-H20)	574.5	236.2	284.5, 169.5	470.9, 68.3	5.29	303, 317	317	
12	Cer d16:1/C23:0	608.6	236.2	24.2	23.4	5.76	344	344	
13	Cer d16:1/C23:0(-H20)	590.6	236.2	13.9	14.2	5.76	344	344	
14	Cer d16:1/C24:0	622.6	236.2	18.6	14.9	5.88	352	352	
15	Cer d16:1/C24:0(-H20)	604.6	236.2	11.7	13	5.88	353	353	
16	Cer d16:1/C24:1	620.6	236.2	32.8, 32.9	27.2, 36.8	5.57	335, 344	335	
17	Cer d16:1/C24:1(-H20)	602.6	236.2	16.9	15.2	5.57	335	335	
18	Cer d16:1/C25:1	634.6	236.2	332.6, 269.0	411.8, 248.1	5.757	343, 352	343	
19	Cer d16:1/C25:1(-H20)	616.6	236.2	26.5	27.3	5.757	344	344	
20	Cer d16:1/C26:0	650.6	236.2	550.8, 45.5, 95.8	677.2, 51.4, 85.2	6.11	354, 368, 375	368	
21	Cer d16:1/C26:0(-H20)	632.6	236.2	14.1	26.6	6.115	368	368	
22	Cer d16:1/C26:1	648.6	236.2	260.7, 33.4	421.3, 34.4	5.86	354, 364	352	
23	Cer d16:1/C26:1(-H20)	630.6	236.2	25.6	27.6	5.86	352	352	
24	Cer d18:0/C16:0	540.5	266.2	50.7, 43.1, 54.6	51.4, 45.8, 53.5	4.806	264, 301, 306	301	
25	Cer d18:0/C16:0(-H20)	522.5	266.2	38.5, 21.7	37.8, 21.9	4.806	268, 294	294	
26	Cer d18:0/C18:0	568.5	266.2	288.4, 57.6	568.3, 53.4	5.272	312, 325	325	
27	Cer d18:0/C18:0(-H20)	550.5	266.2	19.7	18.8	5.316	317	317	
28	Cer d18:0/C20:0	596.5	266.2	44.8, 42.5	43.1, 42.0	5.6	332, 340	332	
29	Cer d18:0/C20:0(-H20)	578.5	266.2	21.5	20.1	5.6	333	333	

coefficients of variation and their respective RTs. In addition, the user can refer to the ion chromatogram image files to see which peak corresponds to the "real" peak, based on the knowledge of the assay. See the template in Figure 3:

Once the user fills up this form, run the last command of the software **MRM_RT.py**.

The user does not have to do this for every assay – as long as the MRM method stays the same and the LC-MS instrumentation stays the same, this template file can be applied to all future experiments.

5 Example "param.txt"

```
mzML_files
*.mzML

batch_info
batchinfo_SP2.txt

length_of_ion_chromatogram
4 40

transition_list
SP2 MRM list.txt

ISTD_trace_all
1

batch_correction
1
```

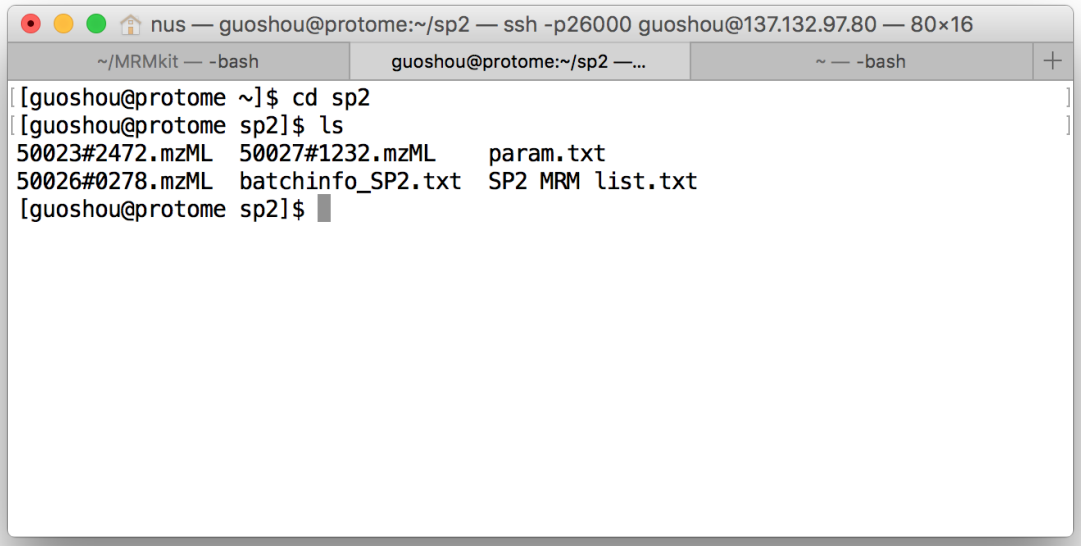
6 Command line for Unix/macOS users

Installation of python3 (>= 3.7), numpy, matplotlib and R (>= 4.0) is required.

```
cd {path_to_folder_with_mzML_files,_param.txt,_assay_info_and_batch_info}

# internal standards (ISTDs) peak extraction
# Run time for SP2 data: 2 hours
python3 {path_to_MRMkit}/MRMistd.py

# peak detection
# Run time for SP2 data: 1 hour
python3 {path_to_MRMkit}/MRMgetpeak.py

# draw ion chromatograms
# Run time for SP2 data: 1 hour
Rscript {path_to_MRMkit}/MRMionc.r

# batch correction
# Run time for SP2 data: a few minutes
python3 {path_to_MRMkit}/MRMcorrect.py

# peak picking (optional)
# Run time for SP2 data: a few seconds
```

```
# users to input RT of true peaks in user_RT.txt before running the command below
python3 {path_to_MRMkit}/MRM_RT.py
```

1. Change directory to the folder (e.g. sp2) containing mzML files (a truncated list of mzML files shown here), param.txt, assay info and batch info files.



2. internal standards (ISTDs) peak extraction, MRMkit's path in this example was "~/02srm".

3. peak detection

4. draw ion chromatograms

A series of IonC_{transition_name}.pdf will appear upon completion.

5. batch correction

"batch_adjusted.txt", "peak_picking.txt" will appear upon completion.

7 Windows users

Installation of R (>= 4.0) is required.
Users can copy and paste the provided executables (.exe) and batch files (.bat) to the folder containing the mzML files, param.txt, assay info and batch info files.
Then, double-click on the batch files in the sequence of "MRMistd.bat", "MRMgetpeak.bat", "MRMcorrect.bat" and "MRM_RT.bat".

1. internal standards (ISTDs) peak extraction
   MRMistd.bat

2. peak detection
   MRMgetpeak.bat

3. draw ion chromatograms
   MRMionc.r. Run this R script using R
   A series of IonC_{transition_name}.pdf will appear upon completion.

4. batch correction
   MRMcorrect.bat
   "batch_adjusted.txt", "peak_picking.txt" will appear upon completion.

8 Copyrights

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.