LipidMatch Processing Manual

Jeremy P. Koelmel, Nicholas M. Kroeger, Mike Napolitano, and Candice Z. Ulmer

**For video tutorials go to:**

<https://www.youtube.com/playlist?list=PLZtU6nmcTb5mQWKYLJmULsfqNy9eCwy7K>

Step 1: Install Necessary Programs

* R, version 2.0.3 to 3.3.3 (**3.3.3 is suggested**, do not download the latest version of R as suggested in the tutorial video; this will cause errors) < <http://cran.rstudio.com/>>
* MSConvert (ProteoWizard) <<http://proteowizard.sourceforge.net/downloads.shtml>>
* MZmine, XCMS, MS-DIAL or other peak picking software
* Unzip LipidMatch folder (right click "extract all" in windows 10) and delete zipped file

Step 1b: Replace R packages with ones which come with LipidMatch:

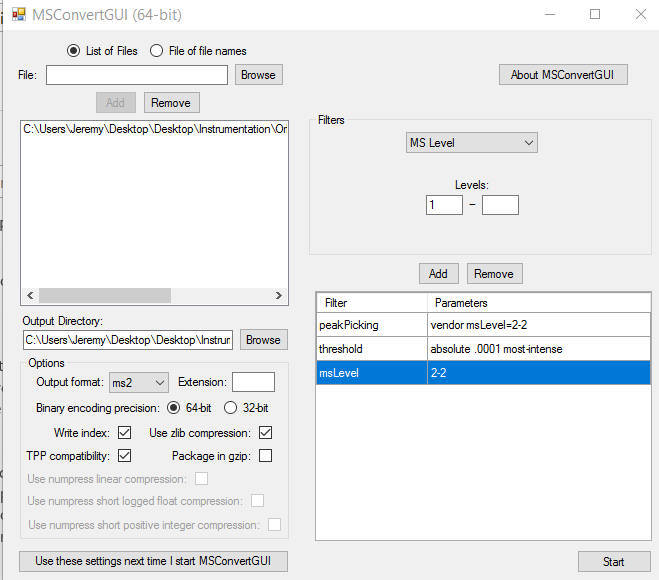
* Either manually navigate to the folder that contains the packages for R 3.3.X (**C:\Users\Your.Name.Here\Documents\R\win-library\3.3**) or run the "Remove All Installed Rackages.R" script (in athe additional\_tools folder).  If you have multiple versions of R installed, then the script will delete the packages for only the version from which the script was executed.
* Navigate to the folder in the step above (bold text) and into it copy the 19 folders (packages) (60 MB), which are included in the LipidMatch distribution as: "R\_Packages"
* Restart R 3.3.3
* Never update either R or the packages of the R installation reserved for LipidMatch.

Step 2: File Conversion using MSconvert

*Note: Depending on the vendor format, file conversion parameters and methods differ. Note that LipidMatch has only been tested using Thermo .raw files and Agilent .d files. Currently LipidMatch does not work for Waters .raw files. An alternative method for file conversion can be found in the troubleshooting document. This alternative is necessary for certain Agilent .d files to prevent an error.*

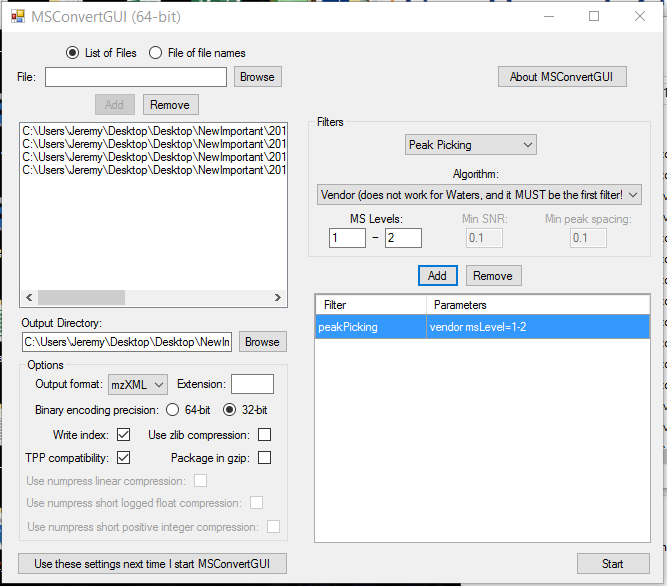
File Conversion in MSConvert (Thermo Parameters)

* MS2 for R script
  + Upload .raw
  + Peak Picking: Level: 2-2
  + threshold: absolute .0001 most intense
  + msLevel: 2-2
  + Output: MS2
  + *Put all MS2 files in one folder along with the inclusion lists*
* mzXML for mzMINE or XCMS *[or you can use any other alignment software]*
  + Upload .raw
  + Peak Picking: Level 1-1
  + msLevel: 1-1
  + threshold: absolute .0001 most intense
  + Output: mzXML

Thermo Parameters for MSconvert:

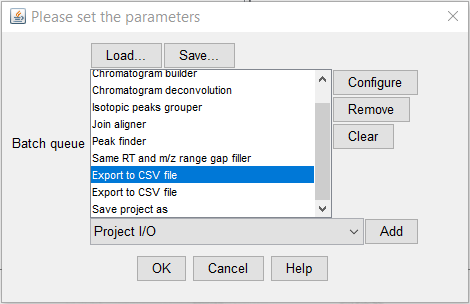
File Conversion in MSConvert (Agilent Parameters)

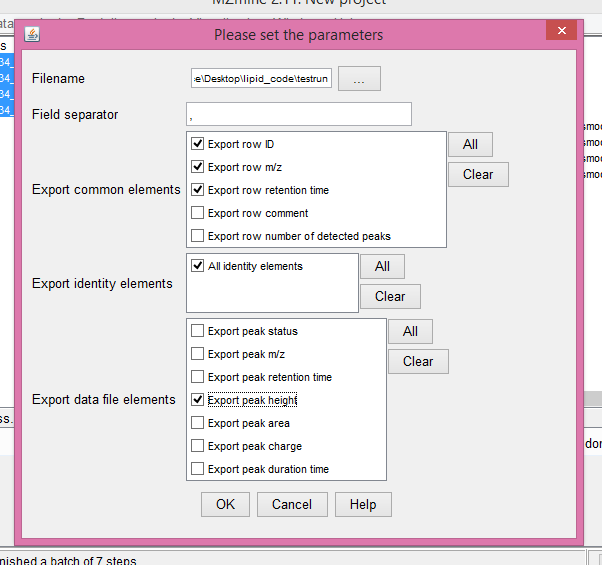
* MS2 for R script
  + Upload .d
  + Select 32 bit conversion, and deselect zlib compression
  + Peak Picking: Level: 2-2
  + Output: MS2
  + *Put all MS2 files in one folder along with the inclusion lists*
* mzXML for MZmine or XCMS *[or you can use any other alignment software]*
  + Upload .d
  + Select 32 bit conversion, and deselect zlib compression
  + Peak Picking: Level 1-1
  + Output: mzXML

Agilent Parameters for MSConvert:

Step 3: Create Feature Table (see example in "ExampleData" subfolder)

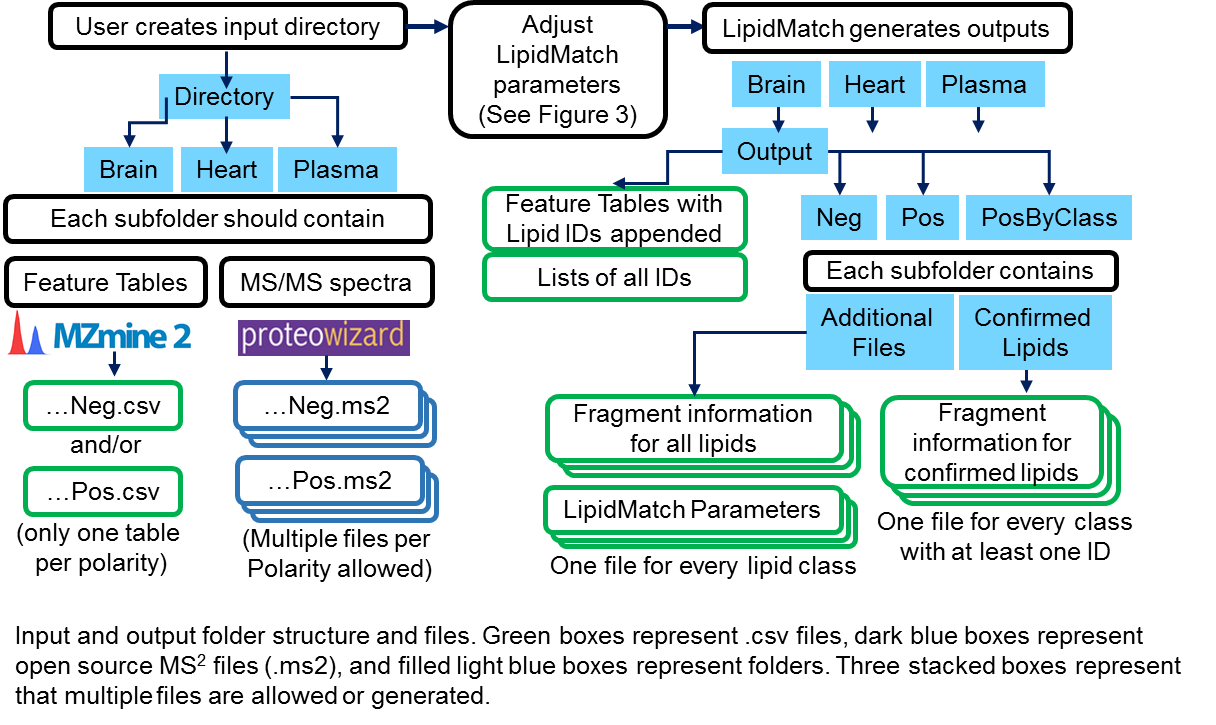
* MZmine 2.23 (alternative feature processing software can be used)
  + import mzXML files: raw data methods -> raw data import
  + process files: highlight files, select project -> batch mode -> load batch file (w/ processing parameters). Note the batch file is the .xml file in the LipidMatch folder.
  + A popup box may appear, select "replace"
  + In the "please set the parameters" window, two options for export data as csv exist. Select each one and click Configure on the right side of the popup box.
    - select the following below (you can choose to set one as peak height and the other as peak area)





Step 4: Organize data

* Place the .ms2 files containing the MS/MS data in an input folder (only a few representative MS/MS files are needed for each polarity) and the table of features, peak areas, etc. (you may have positive or negative mode data, or both in this folder).
  + One feature table should be included for each polarity run (and no more than one). All files (.ms2 and .csv) should end in N (for negative mode) or P (for positive mode).
  + If the files are acquired using AIF, have AIF in the file name.
  + If data was acquired using data-dependent analysis or using a targeted list, put dd somewhere in the filename. (e.g. for one folder: featureTableP.csv, featureTableN.csv, 1\_ddN.ms2, 2\_ddN.ms2, 1\_ddP.ms2, 2\_ddP.ms2...).
  + **Note that .ms2 file names should not be over 23 characters long**.
* Note that you can include multiple folders in the input directory (see figure below), each with their own feature table (Negative and/or Positive) and MS/MS files. All folders will automatically be included for analysis by LipidMatch.



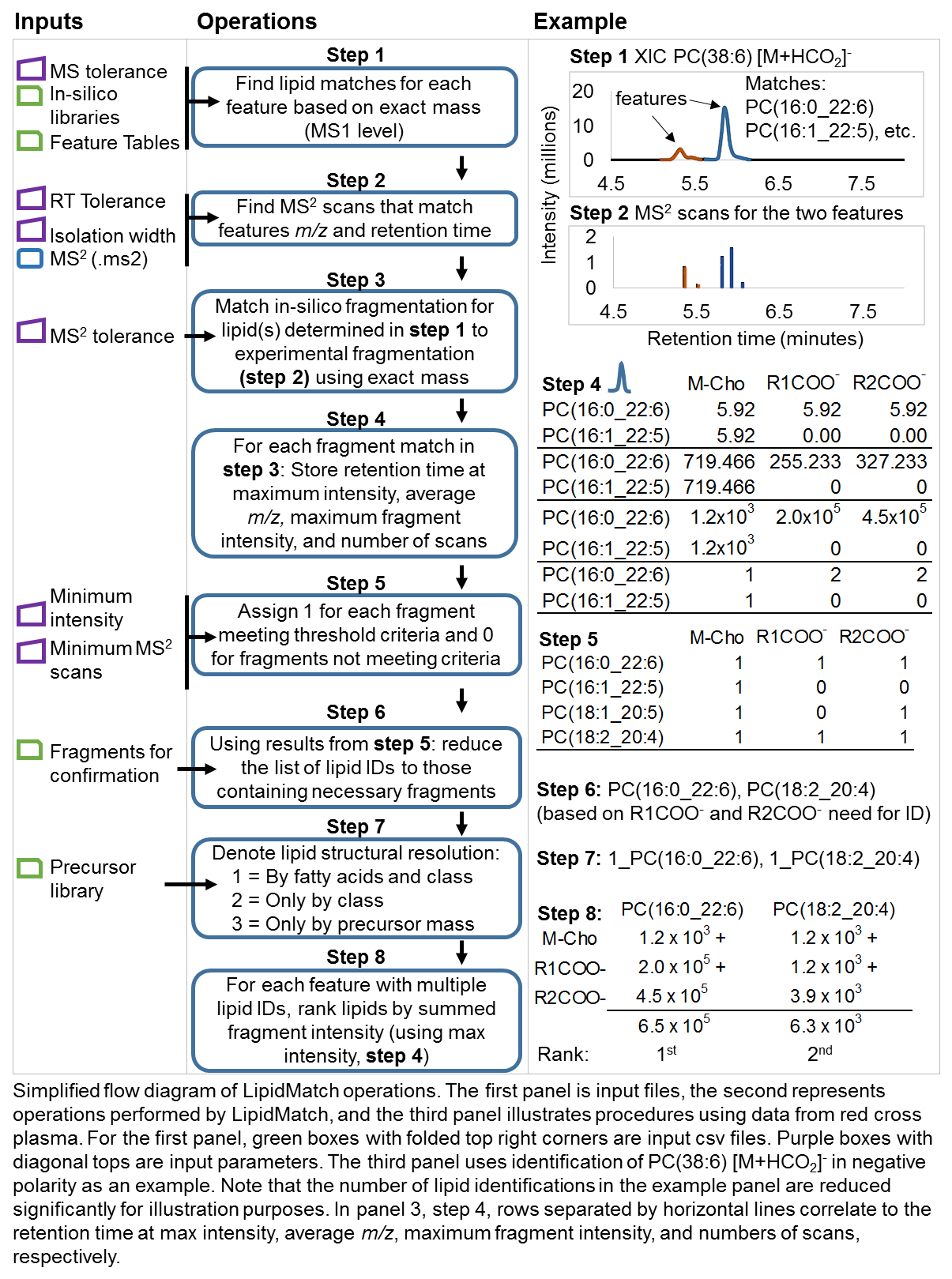
Step 5 (option 1): Run LipidMatch Using Popup Interface

* Open the code ([date]\_LipidMatchEnhanced.R), read the notes at the top of the script, select: Ctrl+Shift+S (or Ctrl+A and run)
* A number of pop-up boxes will appear, read the dialogue box and enter the correct information. You may select ‘d’ for most parameters which will use defaults based on an orbitrap experiment with ddMS2 at 35,000 resolution and UHPLC reverse phase chromatography. After the final pop-up box, the code will begin to output results in a folder named “output” which automatically populates in the input folder.

*Note: The pop-up boxes may appear behind all open windows, so minimize all open windows if pop-up boxes are not visible. The image below shows input parameters and how they relate to LipidMatch processes. Note for Agilent files to set minimum intensity to "1".*

Step 5 (option 2): Run LipidMatch Using .csv parameters file

* Open the code ([date]\_LipidMatch.R), (note LipidMatch.mc.r in the "additional\_tools" is the multicore version for Windows and Linux, and while it works it has not been as thoroughly tested)
* modify the line: csvInput <- FALSE to csvInput <- TRUE
* modify the lines below with the correct directories and file name, make sure to use "/" not "\":
  + parametersDir <- "C:/Users/Jeremy/Desktop/Desktop/Instrumentation/Software/MSms/LipidMatch\_Workflow/LipidMatch\_Flow\_Versions/LipidMatch\_Flow\_Portable/LipidMatch\_Flow/LipidMatch\_Distribution/"
  + parametersFile <- paste(parametersDir, "LIPIDMATCH\_PARAMETERS\_Agilent\_QTOF\_6530.csv", sep="")



Step 5: Interpreting the results

* Results after Step 4 will be saved in a .csv format which can be opened with excel. The final data is in a filename called either NegIDed or PosIDed. Appended in the next column after the original data from the feature table is a column labeled ID. IDs are contained in this column, and encase of multiple IDs are separated by “I”. In addition “NegIDed”, and “PosIDed”, contain all identification without alignment to the original data.
* There are currently 4 identity markers:
  + 1\_[ID] represents confirmation by ddMS2
  + 2\_[ID] represents confirmation by AIF
  + 3\_[ID] represent ddMS2 by clas
  + 4\_[ID] represents confirmation by exact mass.
* The order for multiple confirmations is based on summed fragment intensity.
* The data files in the folder “…Neg/Confirmed/”, “…Pos/Confirmed/”, and “…PosByClass/Confirmed/” can be used to better determine confidence and the most abundant lipids defining a specific feature (these tables contain fragment intensities, retention times at max intensity, experimental and actual mass of fragments, number of scans containing fragments, etc.).
* Definitions of acronyms can be found in the LIPID\_ID\_CRITERIA.csv file found in the LipidMatch libraries folder. For oxidized lipids: CHO is a short chain oxidation product ending in aldehyde and COOH is a short chain oxidation product ending in carboxylic acid. Long chain oxidation products include ketones (Ke), hydroxyl additions (OH), hydrogen peroxide additions (OOH), and mixtures of addition types (eg. a ketone and hydroxyl addition would be (Ke,OH)). Note that epoxy and ketone additions to fatty acids have the exact same mass and fragmentation (using CID or HCD MS level 2). There Ke could be an epoxy or ketone addition. Also fragments with just O, for example (OOOO), mean the addition of 4 oxygens without any subtraction of hydrogen mass (eg. OOOO could be 4 hydroxyl additions, 2 hydrogen peroxide additions, or a mixture of 2 hydroxyl and 1 hydrogen peroxide addition).

Developing your own in-silico MS/MS libraries or choosing which adducts/ions to search against

* Copy the "LipidMatch\_Libraries" subfolder in LipidMatch, and rename this folder (so as not to write over the original). In addition to these instructions video tutorial 6 covers adding new libraries in more detail.
* To change which ions are queried: Find and open the .csv file titled: LIPID\_ID\_CRITERIA.csv in your new subfolder with all the LipidMatch libraries. Under column D titled "Run DDA" change values to TRUE for lipids you want to search against or FALSE for lipids you want to remove from being searched against. The same applies to column E labeled "run AIF".
* To change which fragments are necessary for confirmation, add fragments by the columns associated with them in their respective .csv file. For example if you double click "LPC\_H.csv", column 2 refers to the precursor [M+H]+ ion, column 3 refers to MH-PC, column 4 refers to MH-Tail1+H2O, etc. In the LIPID\_ID\_CRITERIA.csv column 2 is used to list all fragments (by column number as noted above) which are necessary for confirmation, and column 3 is used to list fragments, with at least one fragmented needed to be observed for confirmation.
* For developing your own MS/MS libraries to search against: Please see the format of the MS/MS libraries by selecting any .csv in the "libraries". You can generate your library accordingly using the video tutorial for more in-depth instructions. The first column should be the name of the compound with the header the class acronym, the second column the precursor mass, and any number of columns thereafter will be fragment ions (even if precursors do not show up in MS/MS they are needed for exact mass matching and determining MS/MS scans in .ms2 files). Put all lipid libraries into one folder and add a row to LIPID\_ID\_CRITERIA.csv with information for any new libraries.