LipidMatch Flow Instructions

Jeremy Koelmel 03/13/2021

jeremkoelmel@gmail.com

YouTube tutorials and walk throughs are also available at:

**Innovative-omics.com**

1) Dependencies: (most windows computers will already have the two dependencies below installed, but if you do not, you will need to install them)

Windows (built for Windows 10)

Microsoft .NET (Developed using Ver 4.7.2)

Java 64-bit version: if you use a 32 bit version LipidMatch Flow will give you a warning.

Uninstall java: <https://java.com/en/download/uninstalltool.jsp>

Install java 64 bit: <https://www.java.com/en/download/manual.jsp>, [Windows Offline (64-bit)](https://javadl.oracle.com/webapps/download/AutoDL?BundleId=241536_1f5b5a70bf22433b84d0e960903adac8)

2) Acquire data for the following data types:

You will need 3-4 data types: Blanks, Target, Samples, and MS/MS

Blanks (at least 3, recommended 3-4, blanks are not mandatory, but strongly recommended): Data with full scan (can also have MS/MS) of blank injections with the same polarities and chromatography used for samples. Ideally, extraction blanks would be used, since features found in the blanks at similar intensities to those in samples are removed. Therefore, extraction blanks would remove ions/molecules contributed from the extraction procedure or other processes. Solvent blanks will work as well, but won't remove any features introduced during extraction.

Target (at least 1, recommended 3-5): these can be randomly picked samples, pooled samples, or any samples which are representative of all other samples. We recommend using 3-5 target samples, the less you use the faster the program will go (they are used for doing untargeted feature finding, which is used to generate a targeted list of peaks to search across all samples). The more targets you use the more features will be detected, and hence more lipid identifications will be obtained. They are also treated as normal samples and will be aligned with all you other samples in the final dataset. So for instance if you have 3 diseased and 3 healthy samples, you could run 1 diseased and 1 healthy as targeted samples, and the other 2 as normal samples (for each polarity). Or you could pool all 3 diseased and all 3 healthy and run the pools as targeted samples. **Note: The current blank filtration methods uses the following formula:**

Quartile 1 (pools or targets) > 5 x (Average (Blanks) + 3 x Standard deviation (Blanks))

Therefore, **the pools/targets should have the highest levels of lipids of species of concern or these lipid will be filtered out if they are not greater than the blanks.** Blank filtering parameters can be changed accordingly.

Samples (0 needed, rule of thumb is you will need to GB of RAM per sample, so if you have 250 samples you will need a 512 GB RAM computer): These are data of samples for final feature/lipid molecule intensities to be determined for statistics or determining the relative amount of lipids in a certain sample type. Running both targeted and sample data in full scan mode only maximizes the number of scans across a peaks, but as long as you have 11+ full-scans across peaks than other acquisition modes (e.g. data-dependent analysis) can be used for samples. More scans across a peak means more accurate calculation of peak area, etc. More samples generally don't linearly increase data-processing time, since a rapid peak integration method using a targeted list from the "target" file type is employed.

MS/MS (ddMS2): this can be auto-MS/MS (otherwise known as data-dependent MS/MS) or data of MS/MS data acquired using a targeted inclusion list. We recommend using between 1-12 MS/MS files for each polarity. The more MS/MS files used, generally the more identifications will be obtained, especially if intelligent data-acquisition methods such as iterative exclusion are used ([IE-Omics – Innovative Omics](http://innovativeomics.com/software/ie-omics/)). But, the more data-files are used, the longer data-processing will take. All-ion-fragmentation data using thermo instruments is also accepted, and takes at least twice as long to process as data-dependent files.

2) Change vendor file names (.d, .raw, or other vendor format):

You can change your vendor formatted files to the right naming convention manually (or when you setup your sequence for data-acquisition) or using the tool that comes with LipidMatch Flow (LipidMatch\_RenamingTool).

All file names should end in \_Neg.raw or \_Pos.raw, depending on the polarity.

All ion fragmentation data should have AIF somewhere in the name

Data-dependent are targeted MS/MS files should have "ddMS2" somewhere in the name

Blanks, for blank filtration (for example extraction blanks) should have "blank" somewhere in the file name

Files which are representative of all samples (eg. pooled samples) and can be processed to determine which features to target in all samples, should have "target" somewhere in the file name.

Length of the file names of MS/MS (ddMS2 or AIF) should not be longer than 23 characters. It is OK if full scan data used for feature detection is over 23 characters.

For example: an MS/MS file in positive mode may be name: 2021\_ddMS2\_blood1\_Pos.d (.raw, etc.), or even more simply: 1\_ddMS2\_Pos.d.

3) Input Files:

Double click LipidMatch\_Flow.bat to open the GUI interface.

You should have at least one MS/MS file per polarity (AIF or ddMS2), atleast 3 blanks (optional but strongly recommended), and 1 file with "target" in their name for each polarity

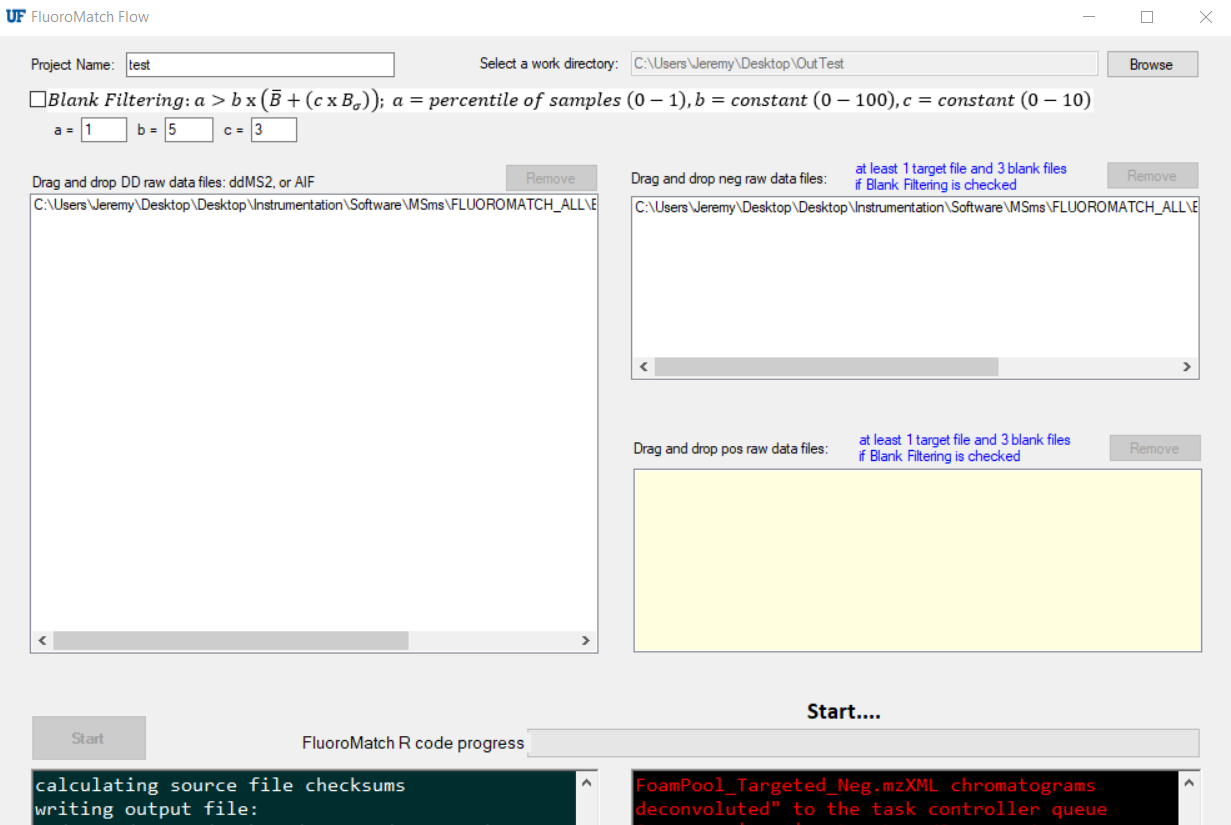
You can run only a single polarity if desired, or both simultaneously (Current LipidMatch and LipidMatch Flow only have negative ion mode libraries)

After uploading files (just drag them onto the GUI in the appropriate place) and choosing project name and working directory the start button should no longer be grayed out.

You can uncheck blank filtering or change the parameters if you would like before running LipidMatch. A is the percentile of samples, for example 1 would mean the maximum of samples, and 0 would be the minimum. This is the peak area which must be higher than the blank threshold. The blank threshold is set by the blank average (B̂) multiplied by 3 x the standard deviation (this is termed the limit of detection (LOD), 3 can be changed by changing the parameter c). Furthermore, the LOD is multiplied by 5, 5-10 are used for limit of quantitation (LOQ), this parameter can also be changed, for example 2 is a more wide net if you want to make sure to see compounds even if they are close to the blank signal.

Select start.

**NOTE ON LOADING BAR:** The loading bar only starts when the annotation step begins, this will take quite sometime for large samples sets, so even if the loading bar has not started the code is likely still running. When it says task complete there is actually one more step. Do not close the application until you see the following final output file: NegIDed\_FIN\_KMD\_scored.csv (in the work directory chosen in the application, under LipidMatch\_Run\Output.



4) Interpreting results:

You will get multiple files outputted in the working directory you selected in the previous step; these include the converted .raw files to .mzXML and the MZmine files in the Temp\_Work folder which is generated, and the LipidMatch output files in the LipidMatch\_Run directory (directory name from historical use of the LipidMatch framework for LipidMatch). Your final data with each feature's *m/z*, retention time, intensities across samples, and lipid annotation(s) will be in the LipidMatch\_Run directory saved as NegIDed\_FIN\_KMD\_scored.csv. This file contains all annotations for negative mode data, this is the file you will most likely want for further statistics, etc.

Please see the youtube tutorials available at **innovative-omics.com** for in-depth walk through on understanding outputs.

MZmine project files can be found in the Temp\_Work folder (use the most recently saved files for negative and positive mode). These files can be open in MZmine and the gap-filled table can be double clicked to manually investigate the quality of peak picking.