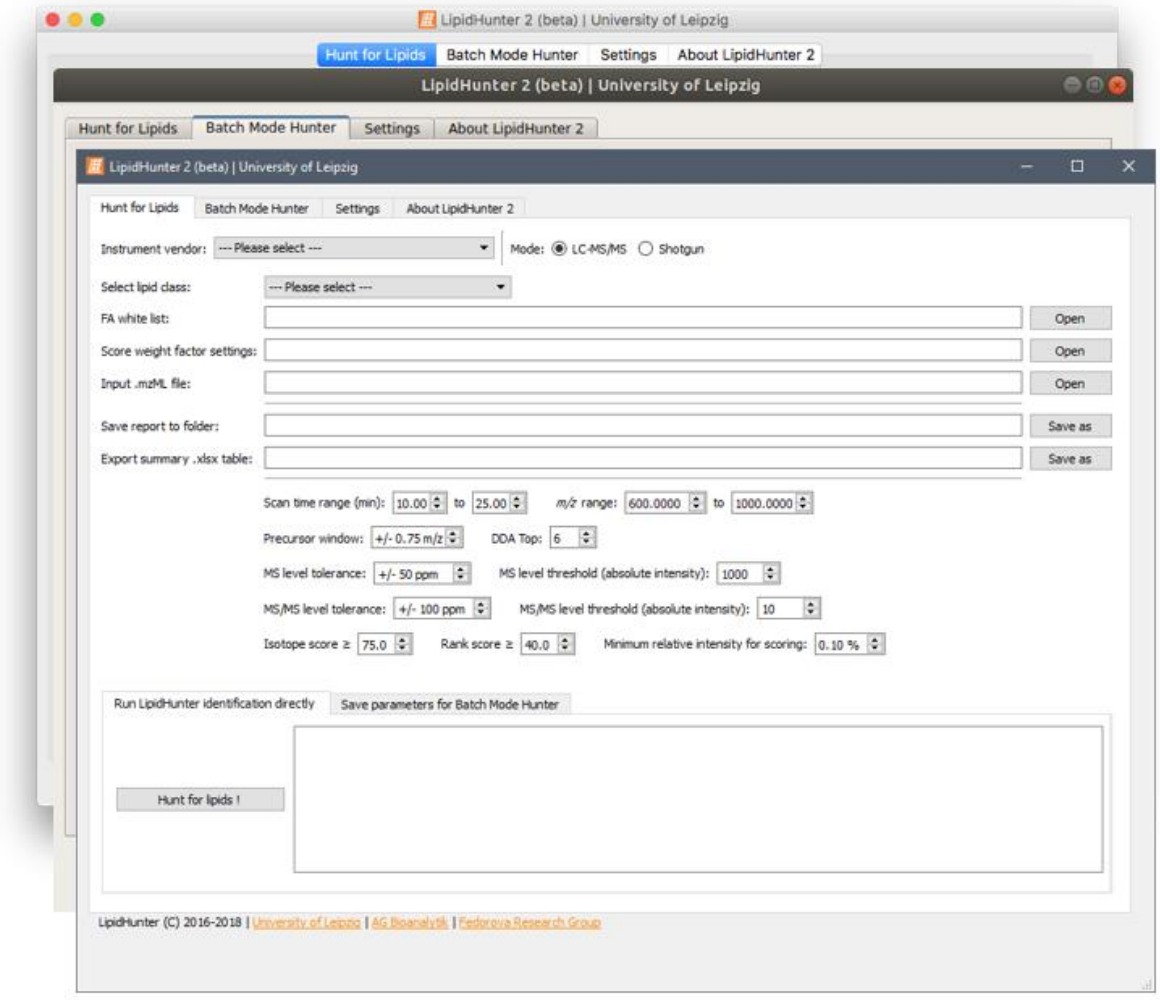
LipidHunter 2

User Guide



For LipidHunter 2 version 6 March 2019

User guide version 6 March 2019

UNIVERSITÄT LEIPZIG

Contents

1. Introduction

LipidHunter is a pure python open source program for Lipid identification from LC-MS or shotgun MS data based on specific RANK score system and isotope score of lipids.

LipidHunter is aimed to provide solid identification of lipids and simply the lipid identification workflow. We provide user friendly cross platform graphic interface, terminal commands, and KNIME integration.

2. License

LipidHunter is Dual-licensed

- For academic and non-commercial use: GPLv2 License
  - https://www.gnu.org/licenses/old-licenses/gpl-2.0.en.html
- For commercial use: please contact the develop team by email.

Please cite our publication in an appropriate form.

- Ni, Zhixu, Georgia Angelidou, Mike Lange, Ralf Hoffmann, and Maria Fedorova. "LipidHunter identifies phospholipids by high-throughput processing of LC-MS and shotgun lipidomics datasets." Analytical Chemistry (2017).
  DOI: 10.1021/acs.analchem.7b01126

3. Downloads and installation

There are two versions of LipidHunter. The Windows executable version and the source code version. General information and installation instructions can be found here:
https://github.com/SysMedOs/lipidhunter

- Windows executable version
  - Executable version is provided for Windows 7, 8 and 10 64bit system only.
  - Please read the instructions of LipidHunter windows version:
  - https://github.com/SysMedOs/lipidhunter/releases
  - The download link:
  - https://github.com/SysMedOs/lipidhunter/releases
- Source code version
  - For developers or other platform users (Linux, macOS), LipidHunter source code is available.
  - Please read the instructions of LipidHunter source code version:
  - https://github.com/SysMedOs/lipidhunter
- Sample spectra
  - The sample spectra in .mzML format can be downloaded from:
  - https://bitbucket.org/SysMedOs/lipidhunter_exe/downloads/
- Updates of LipidHunter user guide
  - Please check following link for the latest version of LipidHunter user guide
  - https://github.com/SysMedOs/lipidhunter/releases

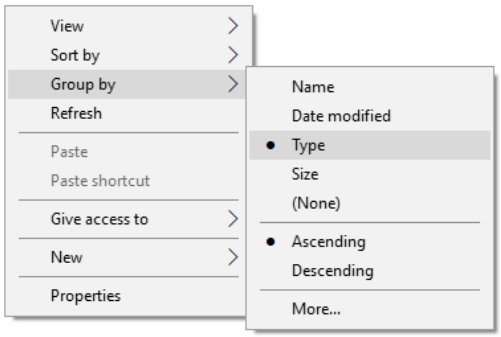
4. Quick start guide

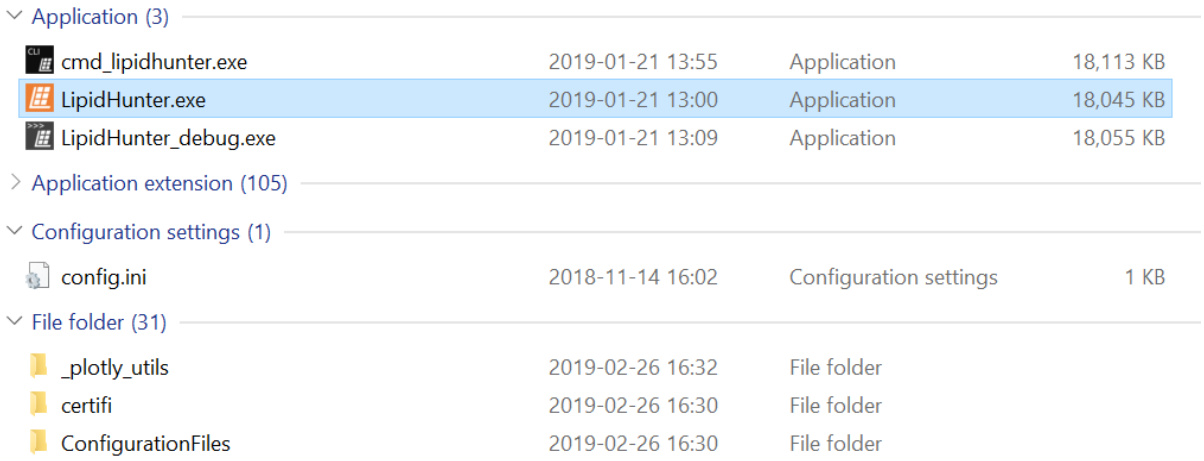
This is a quick start guide for windows executable version users to run a test analysis with the sample dataset included in the tutorial package.

We recommend you to go through this tutorial before you start with your own dataset.

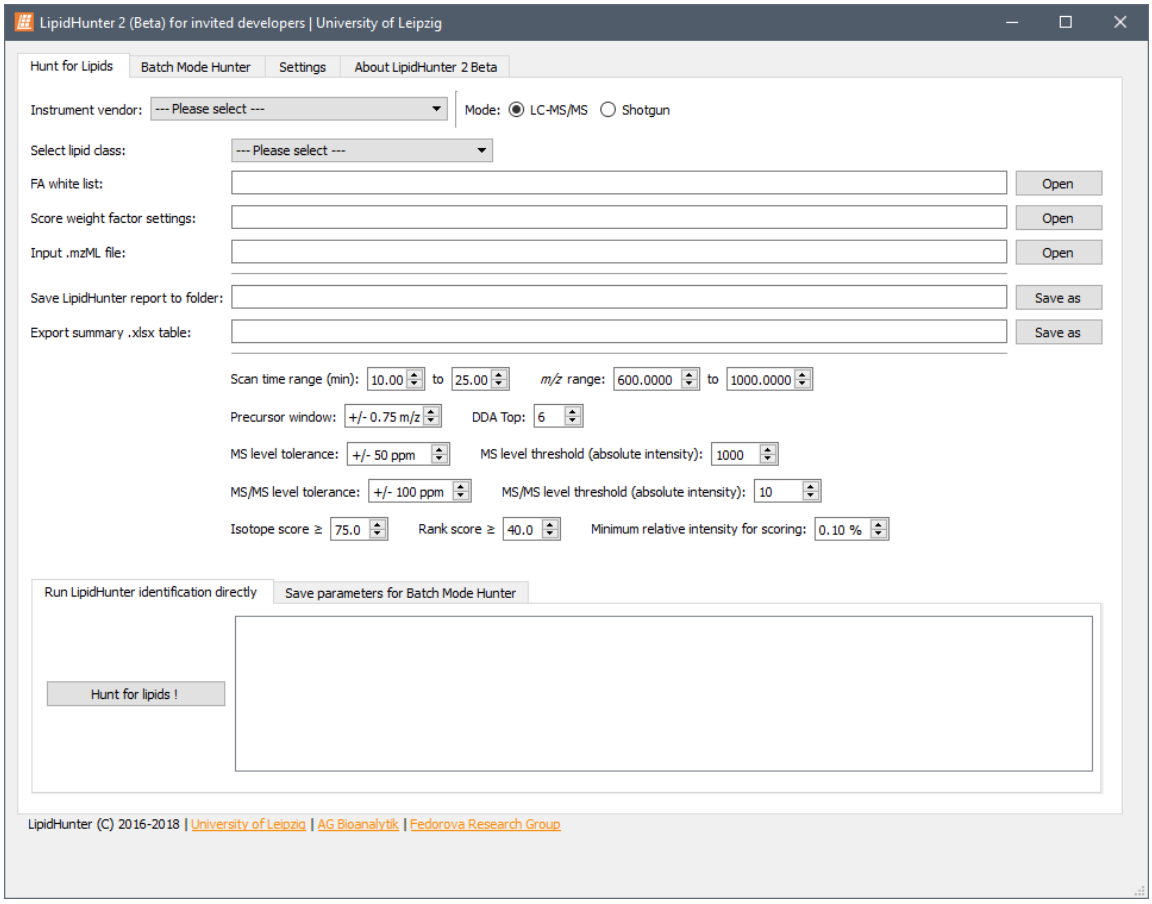
4.1. Launch LipidHunter

Unzip the LipidHunter2.zip to a folder and find the **"LipidHunter.exe"** file. Double click it to start Lipidhunter. You can sort the folder contents by right click on empty space and select "Group by → Type". Then you can find **"LipidHunter.exe"** listed on the top.



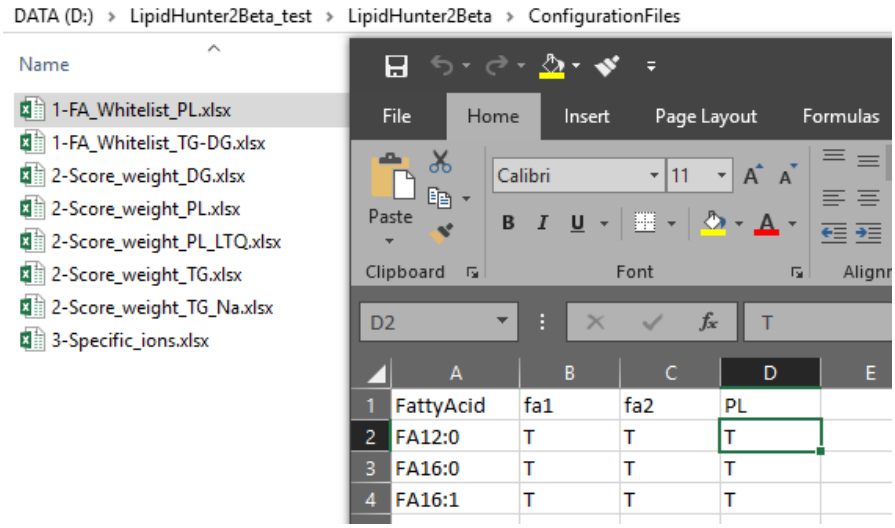


You will see the LipidHunter 2 interface after the checked by your antivirus software.



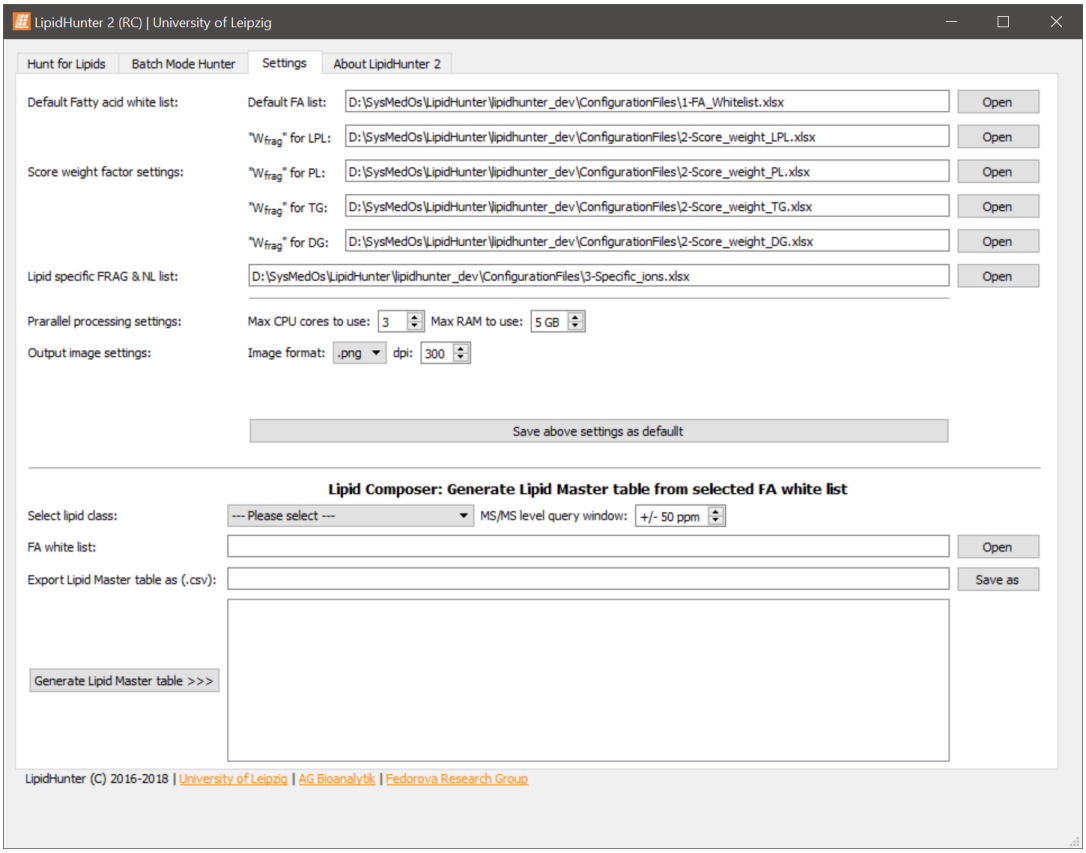
4.2. Basic configuration before your analysis

LipidHunter automatically detected the default configuration tables in the **"ConfigurationFiles"** folder. Open that folder, check if all files are inside.



After checking default files, go back to LipidHunter 2 Settings Tab and set CPU and RAM settings.

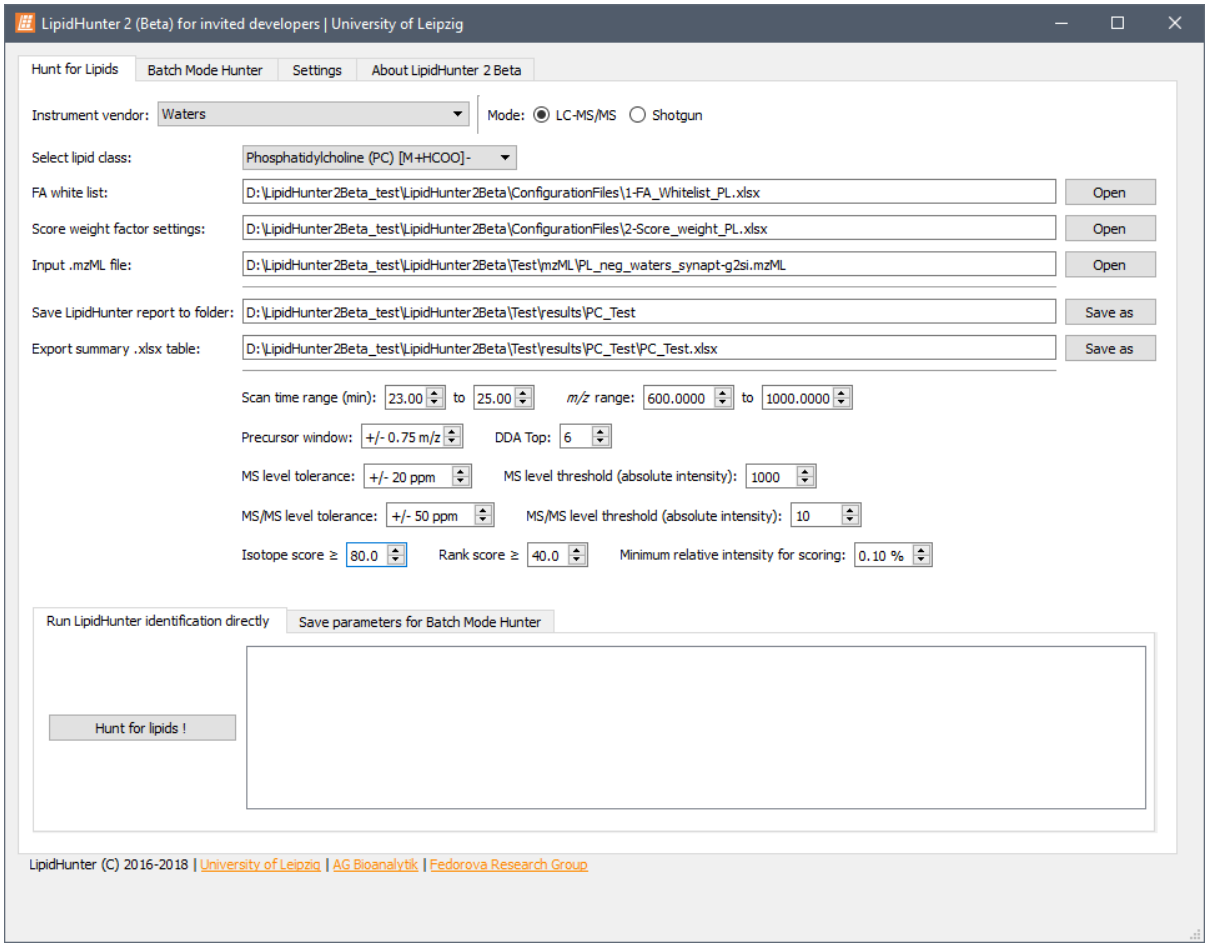
We recommend at least 2 CPU cores 3 GB RAM to run LipidHunter and max 8 CPU cores with 12 GB RAM. The default settings can be used in most computers.



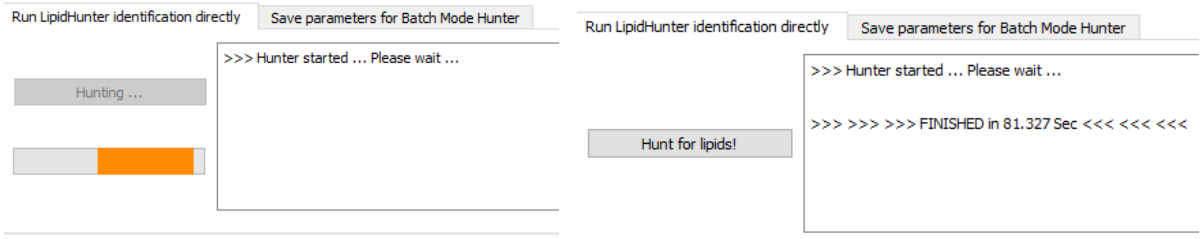
Now press **“Save above settings as default”**, quit LipidHunter by click the red **“X”** on top right, and restart LipidHunter. We recommend you restart LipidHunter every time after you changed default settings in the **“Settings”** tab.

4.3. Start your first hunt with LipidHunter

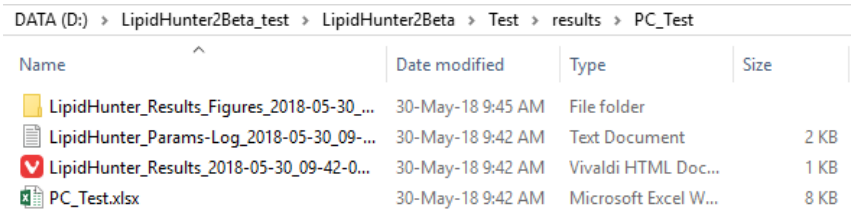
Now you can start to run your first analysis. Please use the mzML from the **“Test/mzML”** folder or download test spectra from LipidHunter repository to try.



Set parameters as above and click on **"Hunt for Lipids!"**. You will see the progress bar and terminal information.

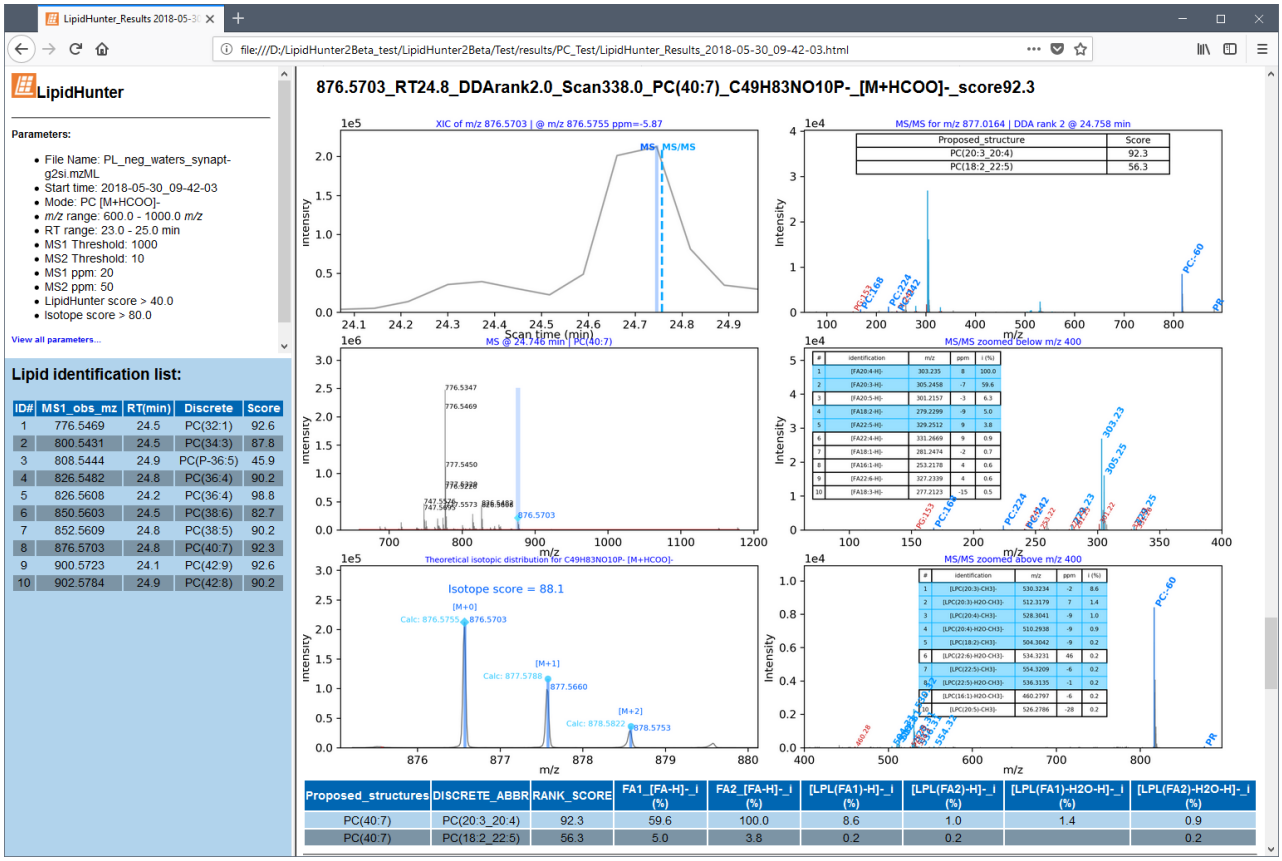


LipidHunter will display the run time when the analysis finish. And you will get the results in the folder selected above.



4.4. View report

Click on the HTML report file, open it by any browser e.g. Firefox, Chrome, you will see the report with images. You can use the table on the left to navigate and click on **"View all parameters"** to check the parameters of this run.



ID#	MS1_obs_mz	RT(min)	Discrete	Score
1	776.5469	24.5	PC(32:1)	92.6
2	800.5431	24.5	PC(34:3)	87.8
3	808.5444	24.9	PC(P-36:5)	45.9
4	826.5482	24.8	PC(36:4)	90.2
5	826.5608	24.2	PC(36:4)	98.8
6	850.5603	24.5	PC(38:6)	82.7
7	852.5609	24.8	PC(38:5)	90.2
8	876.5703	24.8	PC(40:7)	92.3
9	900.5723	24.1	PC(42:9)	92.6
10	902.5784	24.9	PC(42:8)	90.2

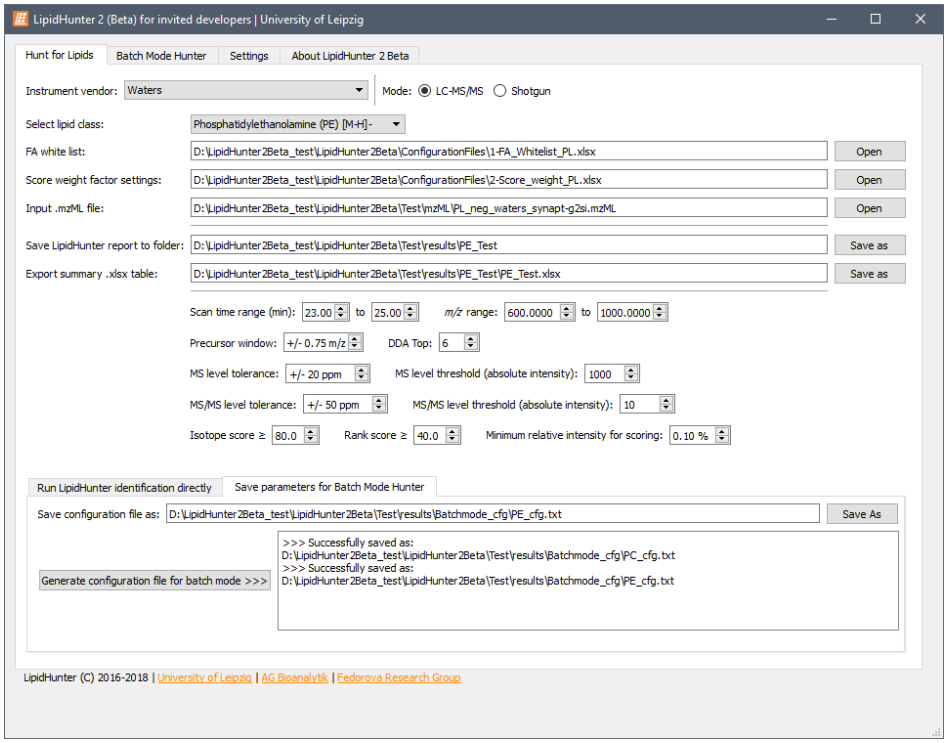
Proposed_structures	DISCRETE_ABBR	RANK_SCORE	FA1_[FA-H]-_I (%)	FA2_[FA-H]-_I (%)	[LPL(FA1)-H]-_I (%)	[LPL(FA2)-H]-_I (%)	[LPL(FA1)-H2O-H]-_I (%)	[LPL(FA2)-H2O-H]-_I (%)
PC(40:7)	PC(20:3_20:4)	92.3	59.6	100.0	8.6	1.0	1.4	0.9
PC(40:7)	PC(18:2_22:5)	56.3	5.0	3.8	0.2	0.2		0.2

You can also open the excel file to sort your results.

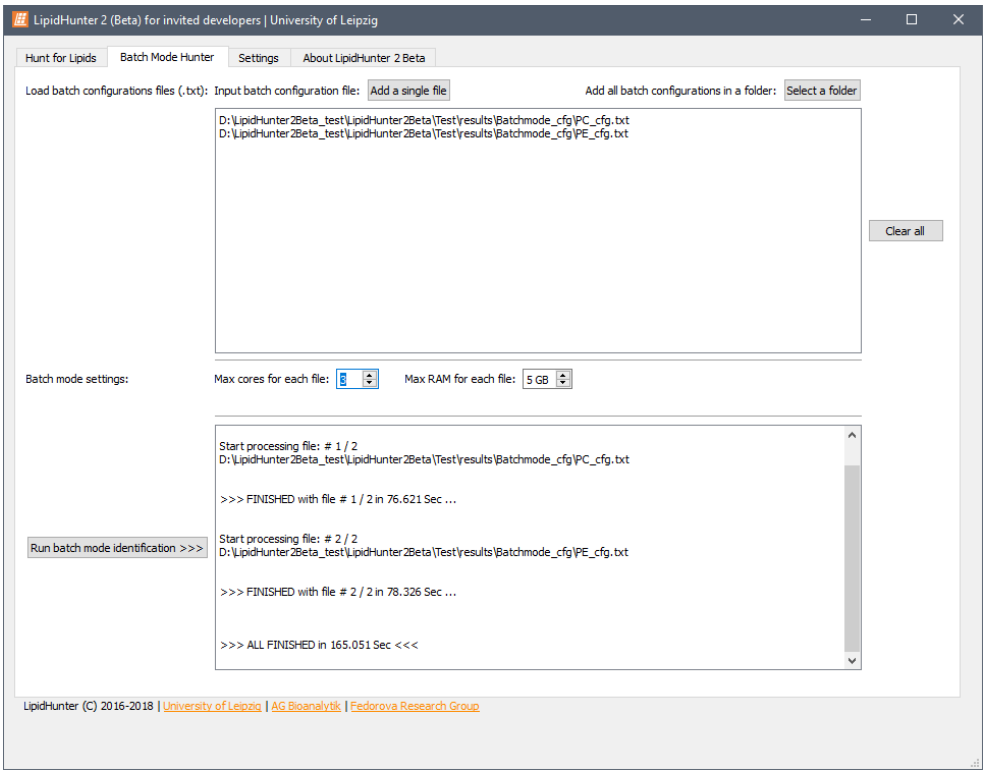
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	osed_struc	DISCRETE_ABBR	Formula_neutral	Formula_ion	Charge	Lib_mz	ppm	ISOTOPE_SCORE	RANK_SCORE	MS1_obs_mz	MS1_obs_i	MS2_PR_mz	MS2_scan_time	DDA#
2	PC(32:1)	PC(16:0_16:1)	C40H78NO8P	C41H79NO10P-	[M+HCOO]-	776.5442	3.58	94.89423775	92.6	776.5469	2.59e+04	776.937805	24.544	5
3	PC(34:3)	PC(16:0_18:3)	C42H78NO8P	C43H79NO10P-	[M+HCOO]-	800.5442	-1.33	92.47161077	87.8	800.5431	1.52e+05	800.947205	24.524	1
4	PC(34:3)	PC(16:1_18:2)	C42H78NO8P	C43H79NO10P-	[M+HCOO]-	800.5442	-1.33	92.47161077	67.5	800.5431	1.52e+05	800.947205	24.524	1
5	PC(P-36:5	PC(P-16:0_20:5)	C44H78NO7P	C45H79NO9P-	[M+HCOO]-	808.5492	-6.03	99.40637225	45.9	808.5444	2.79e+04	808.952881	24.897	1
6	PC(36:4)	PC(16:0_20:4)	C44H80NO8P	C45H81NO10P-	[M+HCOO]-	826.5598	-14.02	92.27187277	90.2	826.5482	4.35e+05	826.966675	24.753	1
7	PC(36:4)	PC(18:1_18:3)	C44H80NO8P	C45H81NO10P-	[M+HCOO]-	826.5598	-14.02	92.27187277	56.3	826.5482	4.35e+05	826.966675	24.753	1
8	PC(36:4)	PC(18:2_18:2)	C44H80NO8P	C45H81NO10P-	[M+HCOO]-	826.5598	-14.02	92.27187277	47.4	826.5482	4.35e+05	826.966675	24.753	1
9	PC(36:4)	PC(18:2_18:2)	C44H80NO8P	C45H81NO10P-	[M+HCOO]-	826.5598	1.19	89.43099368	98.8	826.5608	7.61e+05	826.979248	24.222	1
10	PC(36:4)	PC(16:1_20:3)	C44H80NO8P	C45H81NO10P-	[M+HCOO]-	826.5598	1.19	89.43099368	67.7	826.5608	7.61e+05	826.979248	24.222	1
11	PC(36:4)	PC(18:1_18:3)	C44H80NO8P	C45H81NO10P-	[M+HCOO]-	826.5598	1.19	89.43099368	43.9	826.5608	7.61e+05	826.979248	24.222	1
12	PC(38:6)	PC(18:1_20:5)	C46H80NO8P	C47H81NO10P-	[M+HCOO]-	850.5598	0.58	91.6371397	82.7	850.5603	1.12e+05	850.992004	24.529	2
13	PC(38:6)	PC(18:2_20:4)	C46H80NO8P	C47H81NO10P-	[M+HCOO]-	850.5598	0.58	91.6371397	68.1	850.5603	1.12e+05	850.992004	24.529	2
14	PC(38:6)	PC(16:1_22:5)	C46H80NO8P	C47H81NO10P-	[M+HCOO]-	850.5598	0.58	91.6371397	46	850.5603	1.12e+05	850.992004	24.529	2
15	PC(38:5)	PC(18:2_20:3)	C46H82NO8P	C47H83NO10P-	[M+HCOO]-	852.5755	-17.06	96.55177518	90.2	852.5609	1.88e+05	852.993713	24.825	1
16	PC(38:5)	PC(16:0_22:5)	C46H82NO8P	C47H83NO10P-	[M+HCOO]-	852.5755	-17.06	96.55177518	51	852.5609	1.88e+05	852.993713	24.825	1
17	PC(38:5)	PC(18:1_20:4)	C46H82NO8P	C47H83NO10P-	[M+HCOO]-	852.5755	-17.06	96.55177518	41.8	852.5609	1.88e+05	852.993713	24.825	1
18	PC(40:7)	PC(20:3_20:4)	C48H82NO8P	C49H83NO10P-	[M+HCOO]-	876.5755	-5.87	88.14505798	92.3	876.5703	2.12e+05	877.016418	24.758	2
19	PC(40:7)	PC(18:2_22:5)	C48H82NO8P	C49H83NO10P-	[M+HCOO]-	876.5755	-5.87	88.14505798	56.3	876.5703	2.12e+05	877.016418	24.758	2
20	PC(42:9)	PC(20:4_22:5)	C50H82NO8P	C51H83NO10P-	[M+HCOO]-	900.5755	-3.48	95.64931074	92.6	900.5723	4.61e+04	901.031799	24.093	4
21	PC(42:9)	PC(20:5_22:4)	C50H82NO8P	C51H83NO10P-	[M+HCOO]-	900.5755	-3.48	95.64931074	46.9	900.5723	4.61e+04	901.031799	24.093	4
22	PC(42:8)	PC(20:3_22:5)	C50H84NO8P	C51H85NO10P-	[M+HCOO]-	902.5911	-14.11	84.20768633	90.2	902.5784	9.89e+03	903.03894	24.917	5
23	PC(42:8)	PC(20:4_22:4)	C50H84NO8P	C51H85NO10P-	[M+HCOO]-	902.5911	-14.11	84.20768633	52.7	902.5784	9.89e+03	903.03894	24.917	5

4.5. Batch mode

LipidHunter 2 provide a batch mode to enhance the efficiency when analyzing large amount of data files. You can click on **"Save parameters for Batch Mode Hunter"** and save your previous run for PC as "PC_cfg.txt", and change your settings for PE and save as "PE_cfg.txt".



Then navigate to the **"Batch Mode Hunter"**, you can load all batch mode settings file here, and run all together.



5. Step by step guide to LipidHunter

Step by step guide to LipidHunter to run your own data in the different platforms (Windows, Linux, Mac). All the parameters that you need to consider and how to change the configure files according to your data. The following information is important to avoid any errors and get accurate results.

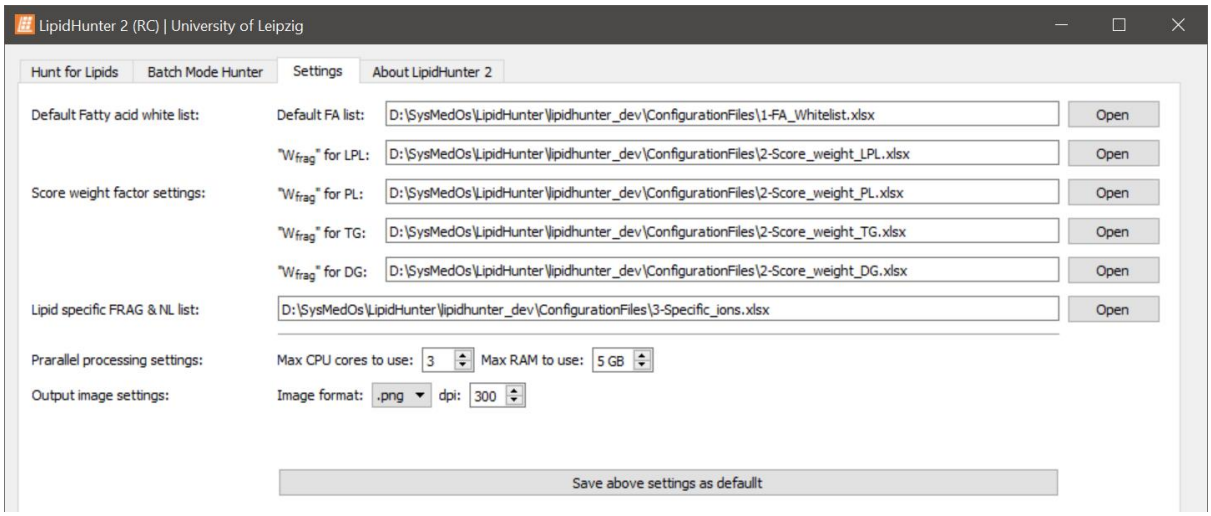
5.1. Convert spectra to mzML

The first step of the analysis is to convert the raw data to mzML file by using the ProteomeWizard MSconvert tool. The program is available in http://proteowizard.sourceforge.net/downloads.shtml and download the version suitable for your system.

Open the **"MSConvertGUI"** from the downloaded folder. When the window open upload the file/files in the program. Click on the **"Browse"** button find the location of the raw data and then click the **"Add"**. The names of the files that you choose should be shown in the text box on the left. The next step is to choose the **"Output Directory"** with the **"Browse"** button to define a new location or use the default one that the program will give the minute you give the raw files. For the conversion choose the **"32-bit as binary"** encoding precision. The final step you need to use **"Filters"** which are located on the right. From the options choose **"Threshold Peak Filter",** then threshold type **"Absolute"** and define the Value base on the minimum intensity that you want the spectra to have and click on **"Add"** button. Keep in mind you define the minimum intensity for the MS/MS spectra. Next you can extract only a small subset or get the full data from all the different measurements. In case you want to get the information only from a specific time range from the **"Filters"** choose **"Subset"** and put the time range expressed in seconds and click **"Add"**. You should see the option in the text box below. Click **"Start"** to convert the raw data to the .mzML format.

5.2. Prepare configuration files

The program needs user define configuration files. The files are separated in three main categories: the **"FA white list"**, the **"Score Weight"** and the **"Specific ions"**. As you can see for the different lipid types and ionization modes there the necessary files with predefine values. All of the needed configuration files are located inside the folder **"Configuration folder"** inside the LipidHunter download folder. All the configure files where optimize according to the raw data that we manually analyze.



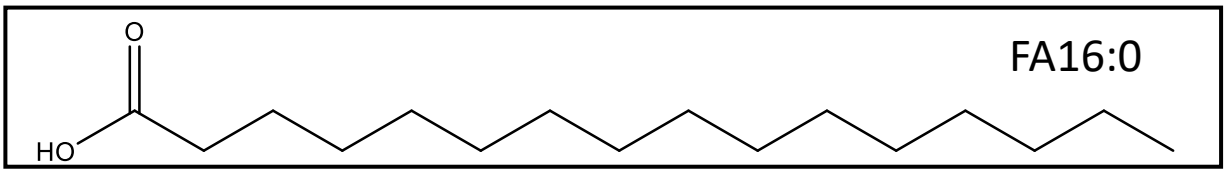
5.2.1. FA Whitelist

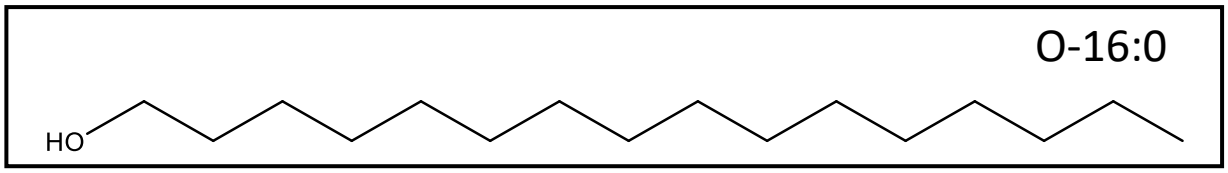
Here you can define all the different fatty acids that you want to include inside your analyses. Below you will find some information about how you can modify this file without influence the program. Also, you can have only one file but still we provide the possibility for each lipid class to support different files.

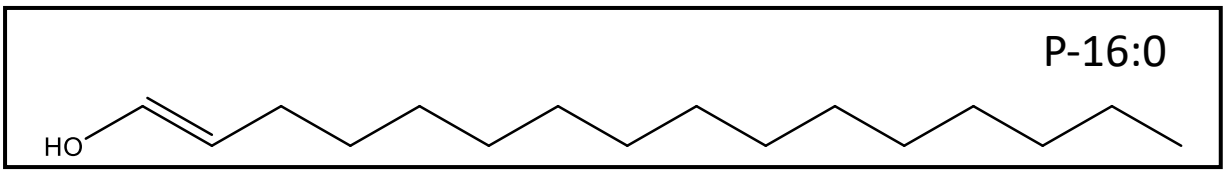
A B C

	A	B	C	D	E	F	G	H
1	FattyAcid	fa1	fa2	fa3	PL	TG	DG	LPL
2	FA14:0	T	T	T	T	T	T	T
3	FA14:1	T	T	T	T	T	T	T
4	FA16:0	T	T	T	T	T	T	T
5	O-16:0	T			T	T		
6	P-16:0	T			T	T		
7	FA18:0	T	T	T	T	T	T	T

As you can see from the picture above, the file is separated in three main sections (A, B and C). The A section is only one column which provides information about the different fatty acids the program will use for the identification. As you can see from the first line the lipids are abbreviated with the following format FA25:0. The number 25 indicates the total number of C in the lipid chain and 0 indicates the total number of double bonds in the chain. The letters indicate the type of fatty acids like the example below (FA16:0, O-16:0, P-16:0).







You can define who many times the fatty acids should be consider for each discrete structure prediction according to the total number of C and db. There three columns because of triacylglycerols (TG) which have three fatty acid chains. Remember the columns **do not** indicate the position in the glycerol where the fatty acid chain is esterified.

Final, on the C section you will find all the possible lipid classes that currently the program can support. You can define for each lipid class which group of lipids you want to use. You just type a capital **"T"** in

the corresponding cells of the lipid that you want to add or remove from the cells you want to remove the lipid.
Since the default fatty acids is different for the different lipid classes, there more than one configure file, one for each lipid class. It is possible to define and use only one according what you like to use for your data.

5.2.2. Weight Factor List

Since in the algorithm it does not use (directly) the intensity of the peaks the weight score was built to show the importance of the different fragments. Each different class it has different fragmentation pattern. Some fragments have higher intensity and some other lower. According to what observe from our data and some online publicly available data we predefined how much this score should be. Higher the score indicates that fragment have higher intensity compare to the others.

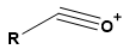
The Weight factor list contains information about how much impact the different fragments will have in the final rank score. The sum score of all the fragments should be 100. The different fragments have different intensities with the weight factor we define which fragments have higher intensity in the spectra compare to the others and they have higher influence in the identification.

The file contains three main columns. The first columns show the abbreviation name of the different fragments. The fragment names should not be change. In case of the addition of more fragmentation pattern you should contact with us.

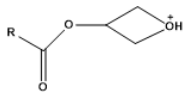
Type	Weight	Group
FA1_[FA-H2O+H]+	1.6	1
FA2_[FA-H2O+H]+	1.6	1
FA3_[FA-H2O+H]+	1.6	1
[MG(FA1)-H2O+H]+	1.6	2
[MG(FA2)-H2O+H]+	1.6	2
[MG(FA3)-H2O+H]+	1.6	2
[M-(FA1)+H]+	30	2
[M-(FA2)+H]+	30	2
[M-(FA3)+H]+	30	2

Fragments abbreviation for TG:

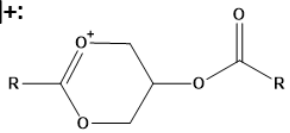
FA1_[FA-H2O+H]+:



[MG(FA1)-H2O+H]+:



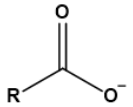
[M-(FA1)+H]+:



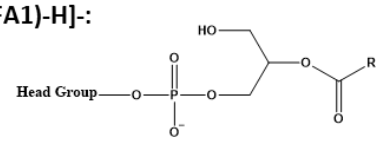
Type	Weight	Group
FA1_[FA-H]-	35	1
FA2_[FA-H]-	35	1
[LPL(FA1)-H]-	9	2
[LPL(FA2)-H]-	9	2
[LPL(FA1)-H2O-H]-	6	2
[LPL(FA2)-H2O-H]-	6	2

Fragments abbreviation for PL:

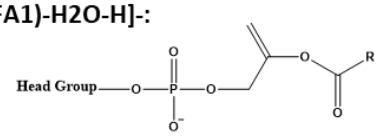
FA1_[FA-H]-:



[LPL(FA1)-H]-:



[LPL(FA1)-H2O-H]-:



5.2.3. Specific FRAG & NL list

This file contains information about specific fragments that are present and they are independent from the fatty acid composition. These fragments can be unique ion fragments or neutral losses correspond to the different head group of the phospholipids or water loss in the diacylglycerols.

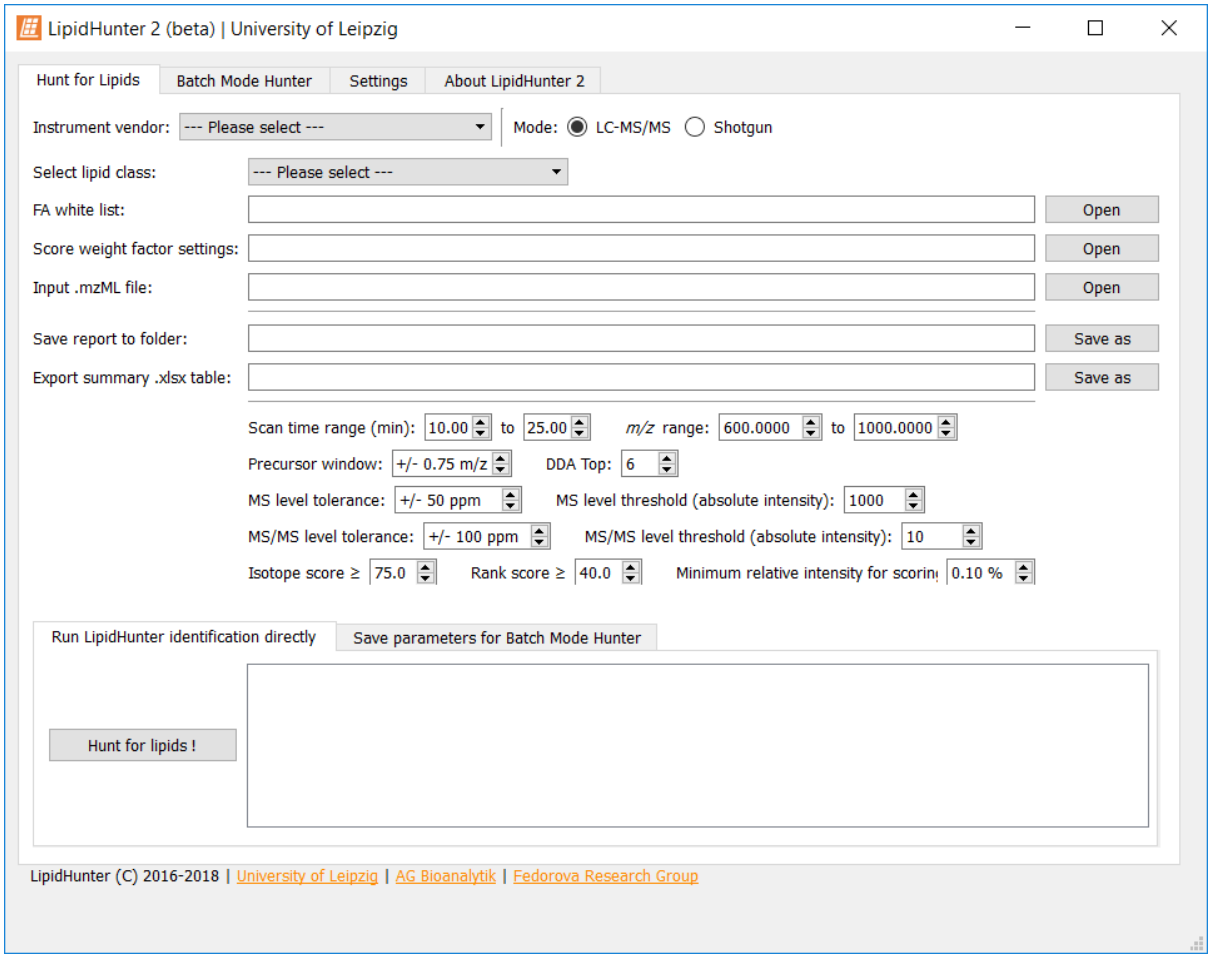
The list contains eight different columns. The first columns contain information about the class which the fragment belongs. Then the second column indicates if it is fragment as FRAG or neutral loss with NL. The third column should contain the exact mass of the neutral loss/fragment and should follow the elemental composition of the FRAG/NL. One important column is also to define in which charge mode we say this fragment by using NEG as negative mode and POS for positive mode. The sixth column should have more detail information about the ionization mode. The last two column should have the labeling of the two FRAG/NL which would be show in the final output and some details about the FRAG/NL.

Keep in mind the LABEL should be clear and short so it will give the mean information if it corresponds to a specific phospholipid head group and if it is a fragment or neutral loss.

CLASS	TYPE	EXACTMASS	FORMULA	CHARGE_MODE	PR_CHARGE	LABEL	REMARKS
PA	NL	97.9769	H3O4P	NEG	[M-H]-	PA:-98	-PA Head Group
PC	FRAG	168.0458	C4H11O4NP-	NEG	[M+HCOO]-	PC:168	demethylated PC [M-H]-
PC	FRAG	224.0688	C7H15O5NP-	NEG	[M+HCOO]-	PC:224	demethylated PC dehydrated glycerol ester [M-H]-
PC	FRAG	242.0794	C7H17O6NP-	NEG	[M+HCOO]-	PC:242	demethylated PC glycerol ester [M-H]-
PC	NL	60.0211	C2H4O2	NEG	[M+HCOO]-	PC:-60	-methyl formate (-CH3COOH)
PC	NL	183.0660	C5H14NO4P	NEG	[M+HCOO]-	PC:-183	-PC Head Group
PC	FRAG	168.0458	C4H11O4NP-	NEG	[M+OAc]-	PC:168	demethylated PC [M-H]-
PC	FRAG	224.0688	C7H15O5NP-	NEG	[M+OAc]-	PC:224	demethylated PC dehydrated glycerol ester [M-H]-
PC	FRAG	242.0794	C7H17O6NP-	NEG	[M+OAc]-	PC:242	demethylated PC glycerol ester [M-H]-
PC	NL	74.0368	C3H6O2	NEG	[M+OAc]-	PC:-74	-CH3COOCH3
PC	NL	183.0660	C5H14NO4P	NEG	[M+OAc]-	PC:-183	-PC Head Group
PC	FRAG	184.0739	C5H15NO4P+	POS	[M+H]+	PC:184	PC Head Group ion in positive mode [M+H]+
PE	FRAG	140.0113	C2H7O4NP-	NEG	[M-H]-	PE:140	PE Head Group [M-H]-
PE	FRAG	196.0375	C5H11O5NP-	NEG	[M-H]-	PE:196	Deprotonated dehydrated glycerol phosphoethanolamine [M-H]-
PE	NL	141.0191	C2H8NO4P	NEG	[M-H]-	PE:-141	-PE Head Group [M-H]-
PE	NL	43.0422	C2H5N	NEG	[M-H]-	PE:-43	-PE Head Group part [M-H]-
PE	FRAG	142.0269	C2H9O4NP+	POS	[M+H]+	PE:142	PE Head Group in positive mode
PG	FRAG	171.0059	C3H8O6P-	NEG	[M-H]-	PG:171	PG Head Group [M-H]-
PG	FRAG	152.9953	C3H6O5P-	NEG	[M-H]-	PG:153	PG Head Group [M-H2O-H]-
PG	NL	172.0137	C3H9O6P	NEG	[M-H]-	PG:-172	-PG Head Group
PI	FRAG	241.0113	C6H10O8P -	NEG	[M-H]-	PI:241	PI Head Group [M-H]-
PI	NL	162.0528	C6H10O5	NEG	[M-H]-	PI:-162	-inositol
PS	FRAG	184.0011	C3H7NO6P-	NEG	[M-H]-	PS:184	PS Head Group [M-H]-
PS	NL	87.0320	C3H5NO2	NEG	[M-H]-	PS:-87	-(serine-H2O)
PS	FRAG	186.0168	C3H9NO6P+	POS	[M+H]+	PS:186	PS Head Group [M+H]+
TG	NL	17.026549	NH3	POS	[M+NH4]+	[M+H]+	TG ammonium loss
DG	NL	35.0371	NH5O	POS	[M+NH4]+	[M+H-H2O]+	TG ammonium loss + water
DG	NL	18.0106	H2O	POS	[M+H]+	[M+H-H2O]+	TG water loss

5.3. Run single analysis

After define the configuration files according to your experiments and your data you can continue with the analysis. You can use two options for the analyses: 1) run a single file and 2) create and save all the parameters for each different run that you want and use the batch mode. The first step for both options you need to go to the first tab call **“Hunt for lipids”** where you need to give more specific parameters according to your .mzml files.



The first think that you need to defined the **"Instrument vendor"**. There is the option to choose from the following Vendors:

- Agilent
- SCIEX
- Thermo
- Waters

Then you should choose if there are **LC-MS/MS or Shotgun Mode**.

The next part is to define the lipid class you want to analyses and in which ionization mode.
In the current version are available the following lipid ionizations modes:

Lysophospholipids	Phospholipids	Glycerolipids
Lyso PA (LPA) [M-H]-	Phosphatidic acid (PA) [M-H]-	Triacylglycerol (TG) [M+NH4]+, [M+H]+, [M+Na]+
Lyso PC (LPC) [M+HCOO]-, [M+CH3COO]-	Phosphatidylcholine (PC) [M+HCOO]-, [M+CH3COO]-	Diacylglycerol (DG) [M+NH4]+
Lyso PE (LPE) [M-H]-	Phosphatidylethanolamine (PE) [M-H]-	
Lyso PG (LPG) [M-H]-	Phosphatidylglycerol (PG) [M-H]-	
Lyso PI (LPI) [M-H]-	Phosphatidylinositol (PI) [M-H]-	
Lyso PS (LPS) [M-H]-	Phosphatidylserine (PS) [M-H]-	

According to the lipid class and mode it in the next 2 boxes (**"FA white list"** and **"Score weight factor settings"**) it will give the default files (defined in the settings tab) that the program is going to use. In case you want to change these files and use different ones you can do it here otherwise you can continue with the default's ones.

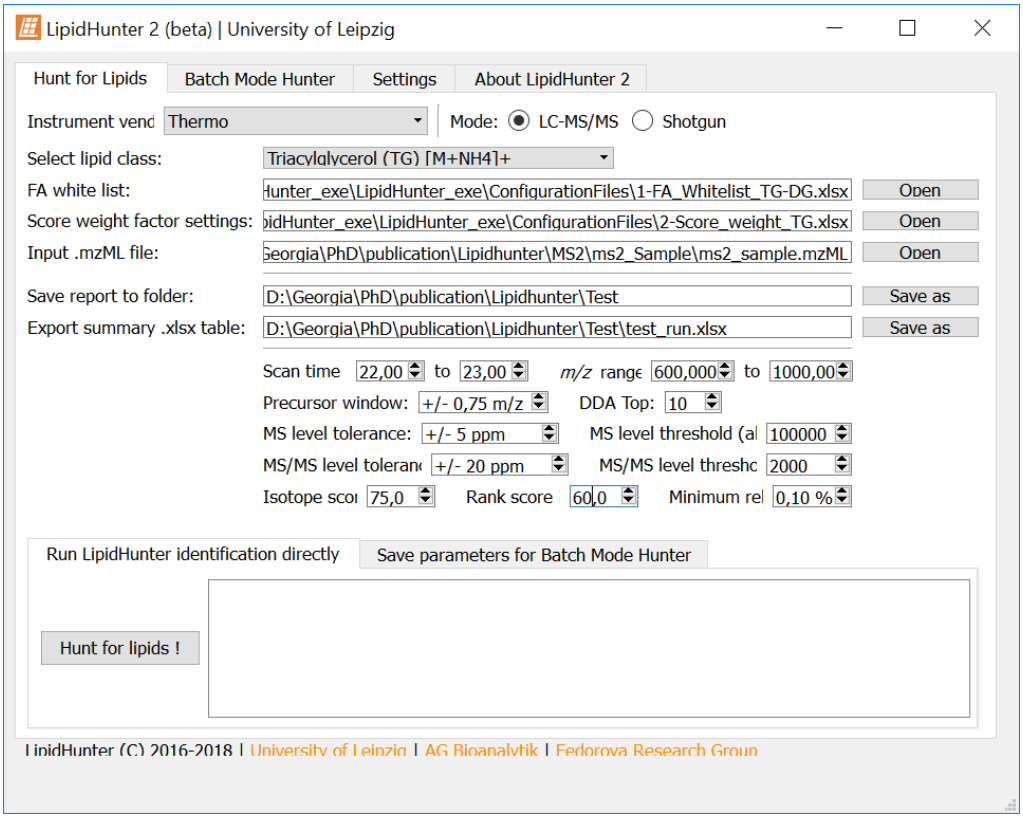
The third box with the name **"Input .mzML file"** you need to upload your .mzML file. Click in the button **"Open"** and find the location of the file that you want to analyze. Then, the next box **"Save LipidHunter report to folder"** is to find the location of a folder where the data will be saved. Click the button **"Save as"** and find the location of the folder. It should be an existing folder and in there the program will save your results. Finally, for the .xlsx file you need to define the location and the name in the sixth box called **"Export summary .xlsx table"** by clicking the button **"Save as"**.

The next parameters are depending from the instruments, lipid types and also your experiments.

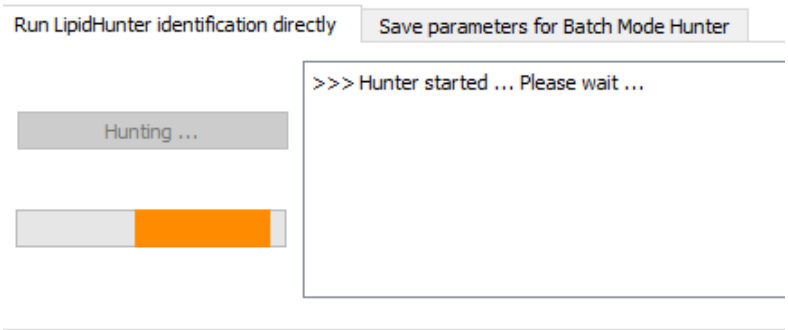
- **"Scan time":** is important to define the time range where the lipid class is eluting according to the separation technique that you use (LC). It is possible to give a bigger rt range but consider there is possibility it will give some false positive identifications.
- **"m/z range":** gives the opportunity to be more specific about the m/z values according the minimum and maximum mass range of the lipid class that you want to analyze. This option it is useful for a faster performance of the program.
- **"Precursor window"**: defines the mass range used to fit MS1 and MS2 reported precursors.
- **"DDA top":** you need to define the method that you run the analyses
- **"MS level tolerance":** The maximum ppm difference between the theoretical to the experimental ionize lipids for the MS level. Strict values indicate more accurate results.
- **"MS level threshold":** The minimum peak intensity for the MS level
- **"MS/MS level tolerance":** The maximum ppm difference between the theoretical to the experimental fragments for the MS/MS spectra.
- **"MS/MS level threshold":** The minimum peak intensity of the fragments to be consider for the identification in the MS/MS spectra
- **"Isotope score":** The program calculates an isotope score for the MS level. This way gives the possibility for a better filtering of the spectra. The higher the score then better identification.
- **"Rank score":** The rank score is use to define the probability of the different predicted structures base of the fragments. Description of who it is calculate is in the publication which you can find in publication Ni, Zhixu, et al. "LipidHunter identifies phospholipids by high-throughput processing of LC-MS and shotgun lipidomics datasets." *Analytical chemistry* 89.17

(2017): 8800-8807. The weight factor configuration files will influence how strict the value should be here.

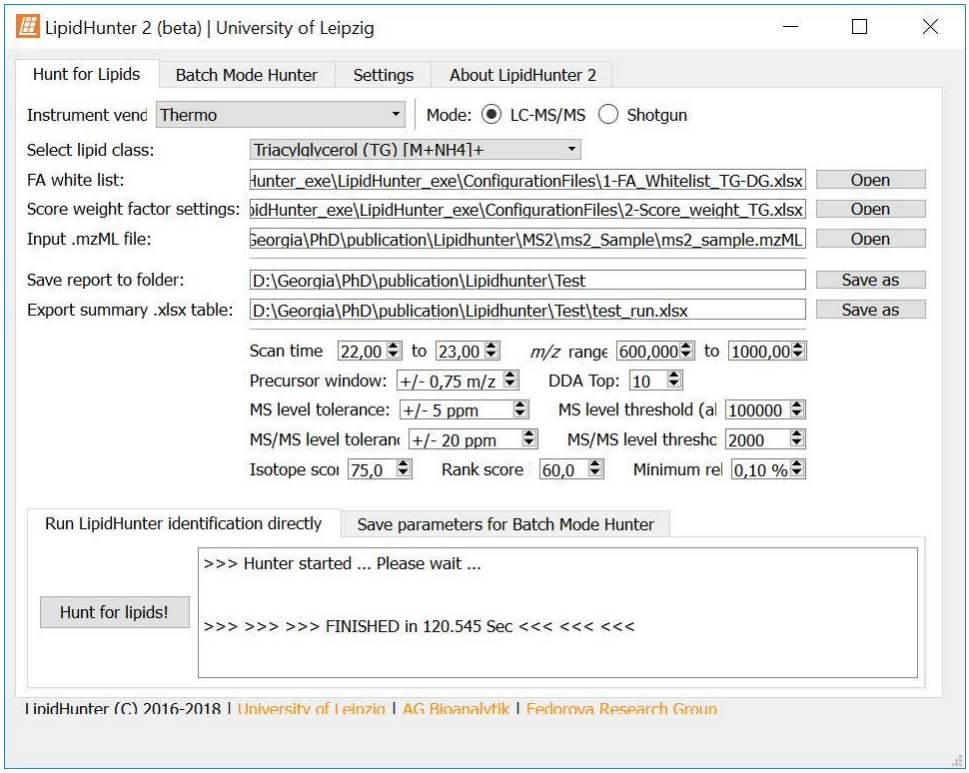
- **"Minimum relative intensity for scoring":** peaks that are below the threshold compare to the higher one from the MS/MS spectra would not consider for the scoring. This is a way to avoid peaks that are at the noisy level but still can be identify from the program.



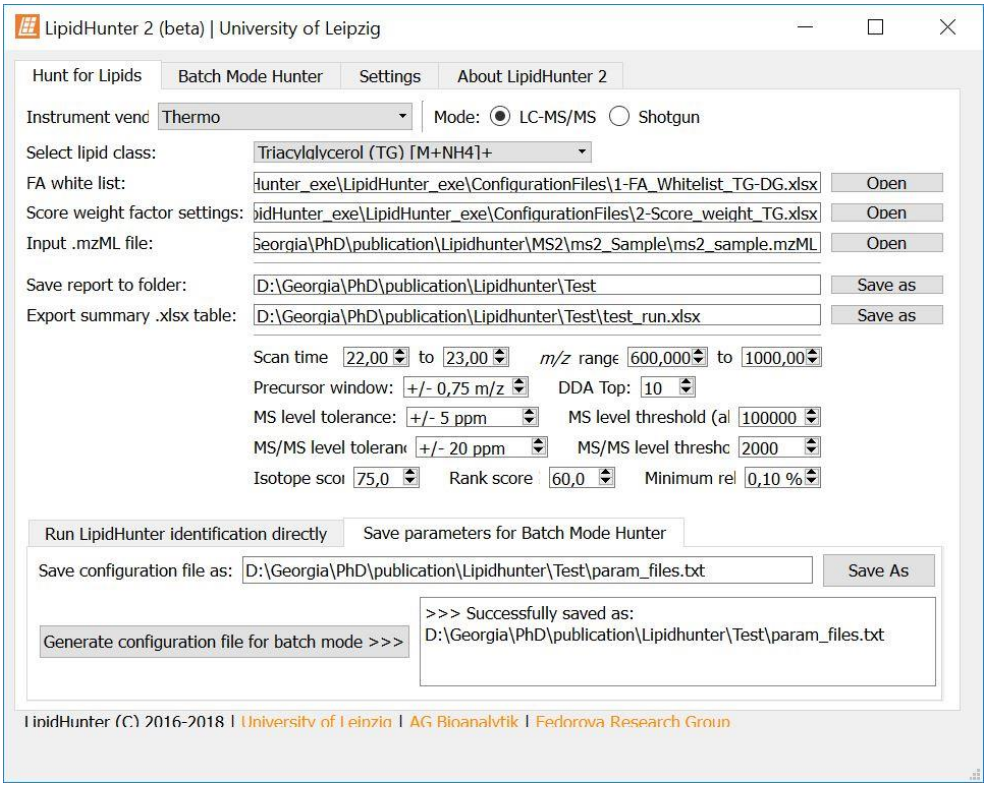
After setting the necessary parameters to run a single file you need to click in the button **"Hunt for lipids!"** from the **"Run LipidHunter Identification directly"** tab.



During the run, a progress bar will be displayed and the **"Hunt for lipids!"** button will be temporally disabled. When the program finish it will give you a message as it shows in the figure below, and the **"Hunt for lipids!"** button will be enabled for the next analysis:



Otherwise to save the file and run the program later with other files, then you need to go to the tab **"Save parameter for Batch Mode Hunter"**. And then in the **"Save configuration file as"** Click on the **"Save as"** button and choose the directory and the name of the file to be saved. Then Click on the **"Generate configuration file for batch mode >>>"** and a message will show up next to it as the figure below.



5.4. Run batch mode

The batch mode is a new figure of the program which it gives you the opportunity to run multiple files at the same time.

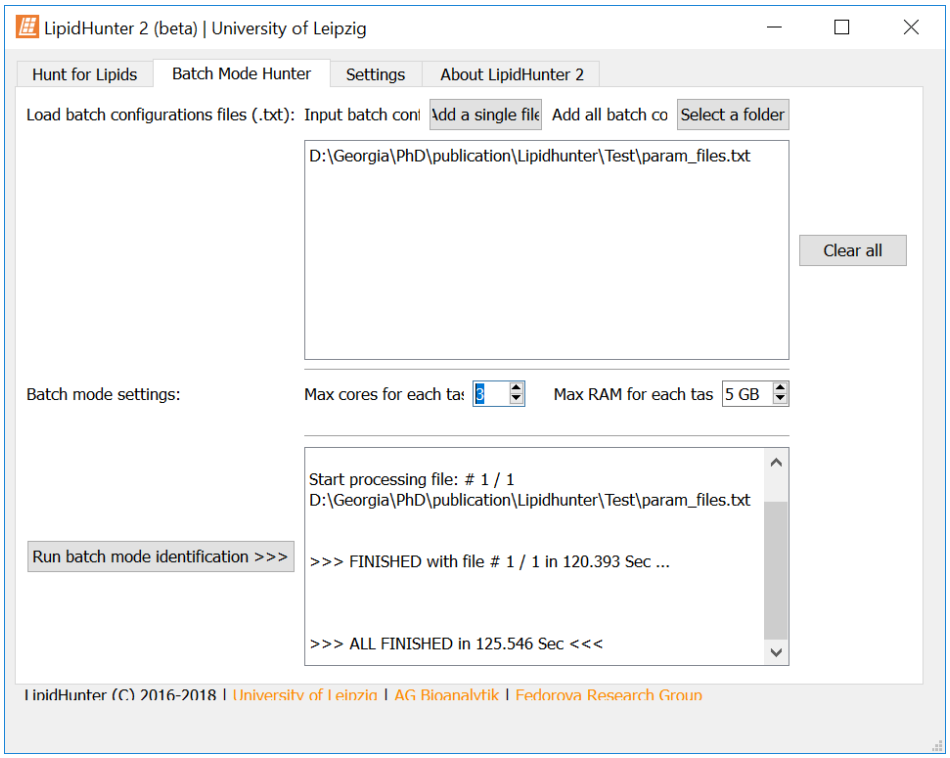
5.4.1. Geminate batch mode configuration file

Before you run the batch mode you need to create the necessary files. Follow the instructions at the 5.3 Section and at the end go to the go to the tab **"Save parameter for Batch Mode Hunter"**. You need to do this process for all the files that you want to run with the different parameters.

5.4.2. Run batch mode

You have two option to upload your data in the program. You can select a folder where you had save all the .txt files that you want to run. You just Click on the second button **"Select a folder"** and find the location of the folder. Alternative you can select one by one the files. Click on **"Add a single file"** and each time you can upload one file at the time.

When you select all your file if you want you can change the number of cores and RAM that the program will use. The identification will start after you click on the **"Run batch mode identification >>>"**. In the window next you will get a short report about which file it analyze and it is finish as it shows in the figure below.

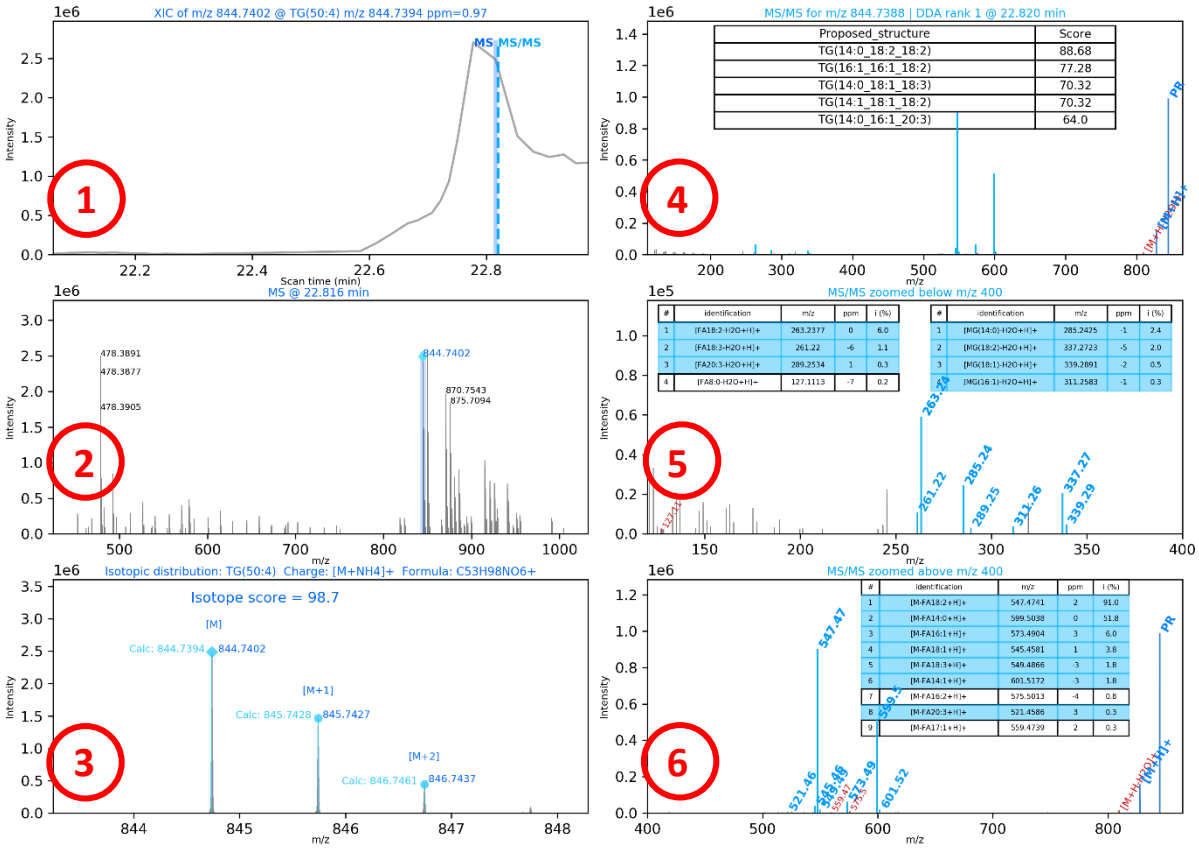


5.5. Review output files

The program will give the results of the identification in different formats. First of all, it will summarize the information in the excel file, then for each MS/MS spectra it will create a figure, it will create a .txt file with the parameters that you use to run the program and finally an html page for a better visualization of the information.

LipidHunter_Results_Figures_2019-01-16_20-50...	16/01/2019 20:52	File folder
LipidHunter_Params-Log_2019-01-16_20-50-21	16/01/2019 20:50	Text Document
LipidHunter_Results_2019-01-16_20-50-21	16/01/2019 20:52	Chrome HTML Docu...
test_run	16/01/2019 20:52	Microsoft Excel Work...

LipidHunter gives you the opportunity to manually check all the spectra that it identifies throw the six panel figures. This way you can see which fragment-peaks the program uses for the identification throw the different panels which provides different type of information.

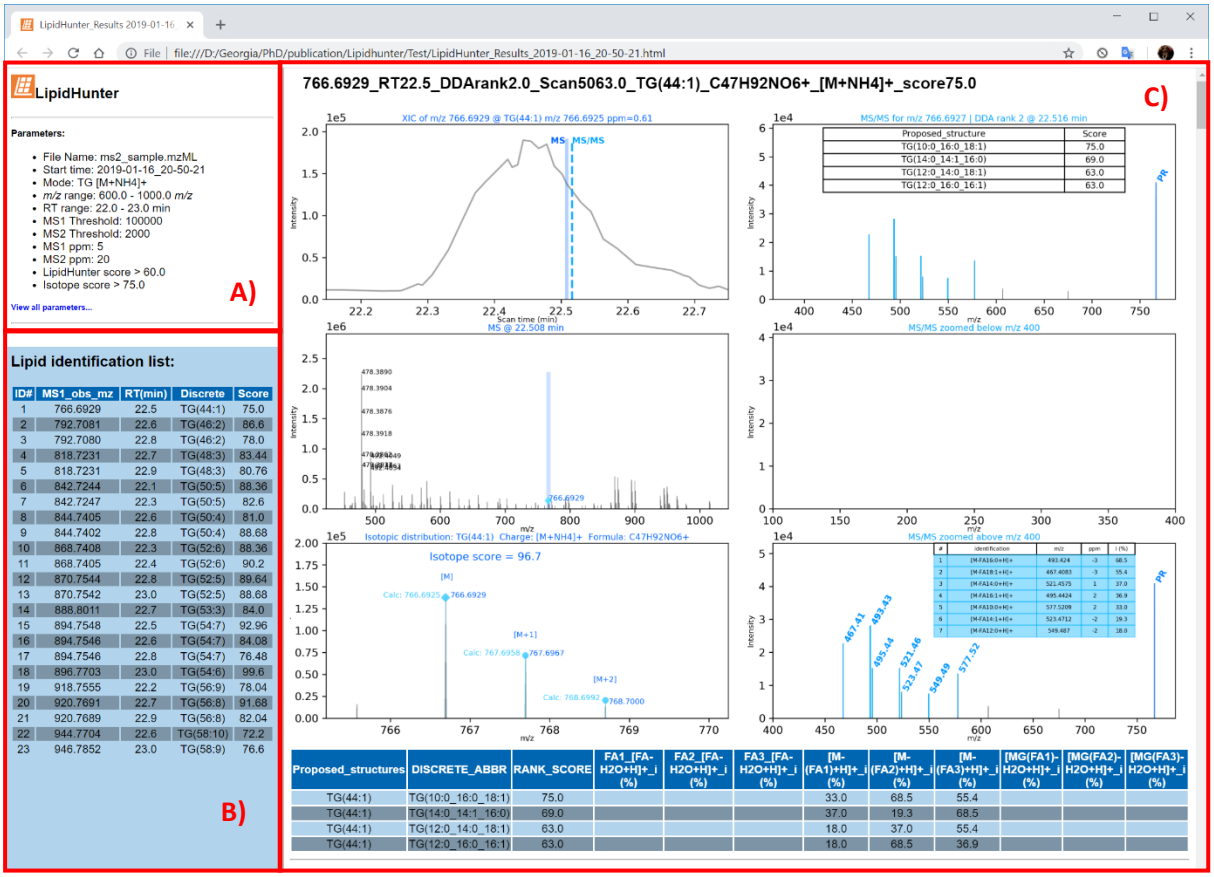


Panels information:

1) Show the extracted ion chromatogram and the position where the MS and MS/MS where taken.
2) Full MS scan and in blue you see where the m/z is located.
3) Shows the isotopic pattern of the m/z and the isotopic score
4) Full MS/MS spectra with the identify peaks. Blue colors shows the peaks used for the proposed structures and red the ones that couldn't feet in the proposed structure and the m/z.
5) Zoom area of the MS/MS between 0 – 400 showing the free fatty acids fragments and the monoglycerol fragments (in the case of TG)
6) Zoom area from 400-precursor m/z showing the neutral loss of one fatty acid.

5.5.1. Interactive HTML report

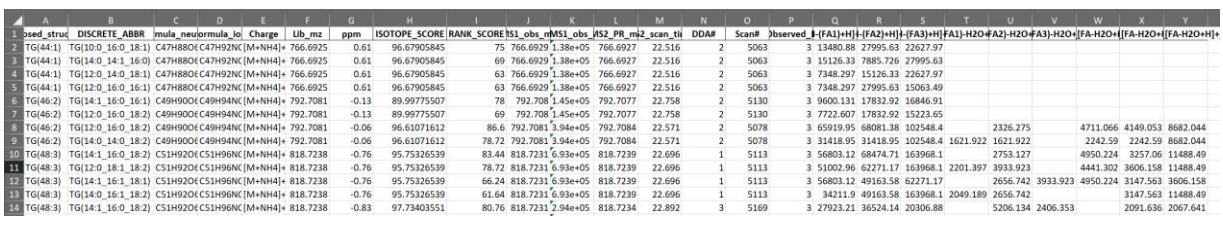
The Html report it gives you the opportunity to navigate throw the data and visualize the figures that are produced for the different files. Below you can see who the Html report looks like.



The window can be split into three sections. The A section it shows the information about the parameters which were used for the lipid identification. In the B section it contains a table having information about the different figures containing only the basic information such as: the observed MS1 m/z, the retention time (RT), the bulk identification (discrete) and final the score of the highest identification. You can scroll up and down and save a general view about the data and also you have the possibility to click to the different m/z values and go directly to the correspond figure without spending your time search in the C section by scrolling down. You can see each figure from the C section and a simple table showing more details about the discrete structure and the relative intensity of the fragments that it identifies.

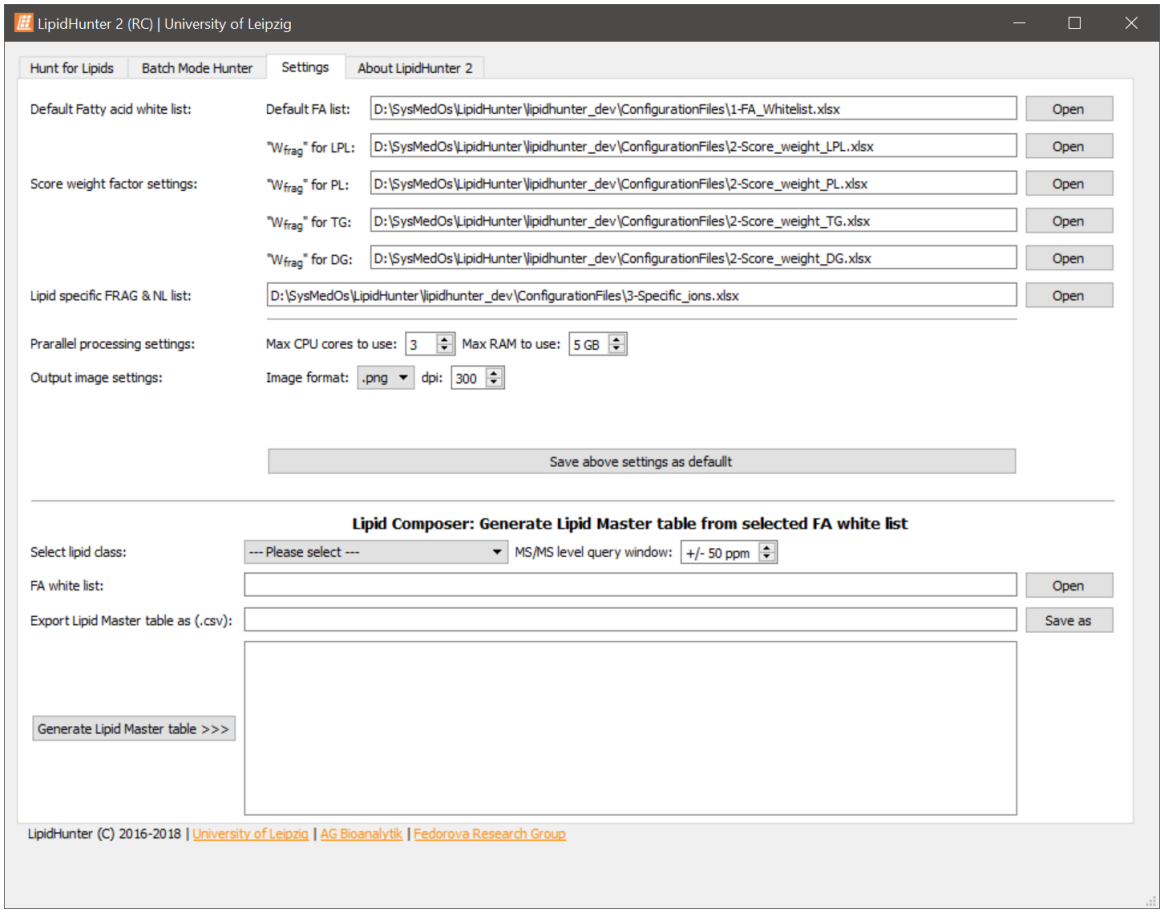
5.5.2. Summary excel table

All the information in the figures and the html report were saved inside the excel table. The excel table it gives you the opportunity to filter the data and also further used for illustrated them in graphs or used them for statistics.



5.6. Generate LipidMaster table using LipidComposer

Another new feature of lipid hunter is the generation of all the possible structures according to the fatty acid list and lipid class. It is available on the second half from the **“Settings”** tab, as you can see below.



The only thing that you need to define is the lipid class and the ionization mode, the MS/MS level query window, the Fatty acid list that you want to use and the location you want to save your list.

The parameter **“MS/MS level query window”** is need to define the range that m/z can have according to the value that you use. The program used this range to define which spectra should take for identification and which not.

In the future it will be possible to upload the LipidMaster list instead of the Fatty acid list.

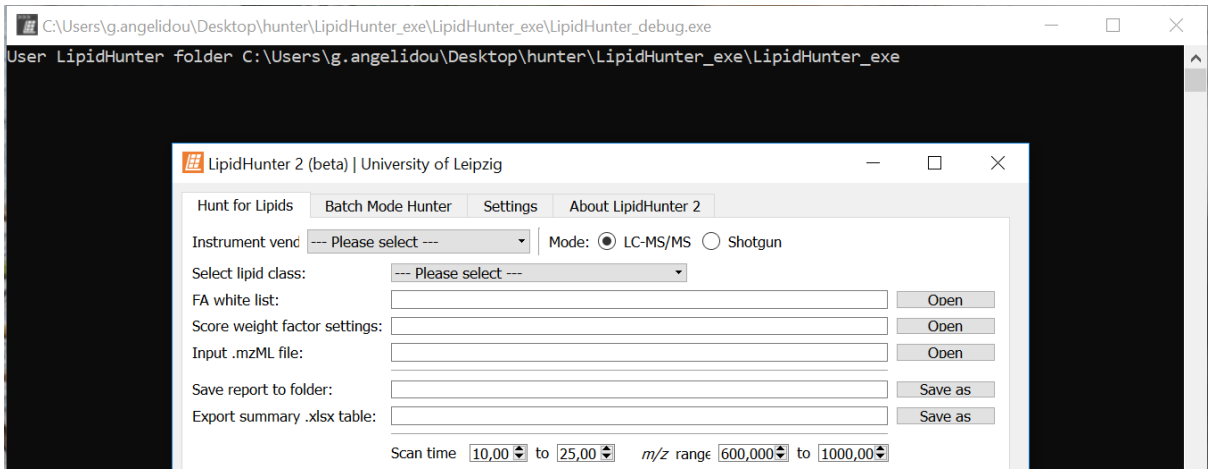
5.7. Debug mode in windows version

The debug mode of the program it gives you the opportunity to follow the process of the program. Now you can see in which step is currently the program and which precursor it tries to identify. In the case of any errors or bugs, the debug mode it gives you more detail information about the error/bug and then it will be easier to identify it and solve it.

Open the folder you downloaded for the program. In the main folder where LipidHunter.exe is located, you should be able to find the following icon with name **“LipidHunter_debug.exe”**.

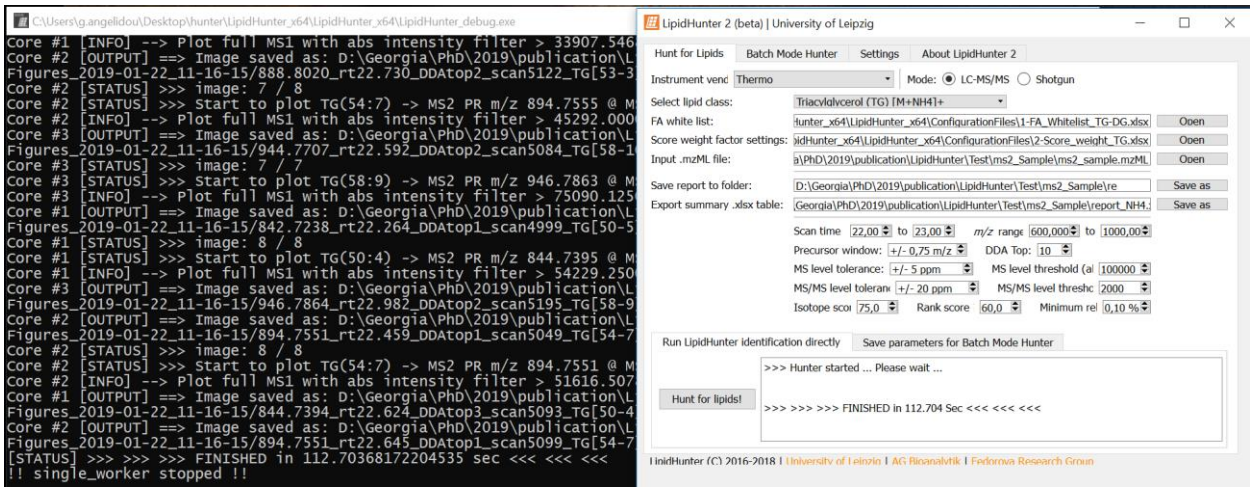
LipidHunter_debug

When you double click LipidHunter window should popup and at the same time a terminal window where will return the report of the program.



In debug mode you will see all essential information, status, and warnings during the run of LipidHunter like the source code version.

Click and stroll in the terminal may paus the analysis in some cases. Please press **"Enter"** to continue the analysis.



6. Notes for source code version users

LipidHunter 2 source code version is developed using Python 3.6 64bit. This release is still compatible with Python 2.7 64bit, but further release is planned to switch to Python 3 completely.

Please note that LipidHunter is using several multiprocessing features and only compatible with 64bit Python environment.

We recommend latest Anaconda release for Python 3.6 64bit as the environment for LipidHunter. The following guide is based on Anaconda environment. For users with low disk space or advanced users, the Miniconda version for Python 3.6 64bit from the Anaconda project can be used as well.

Please read more about Anaconda and Miniconda on https://www.anaconda.com/download/.

6.1. Using LipidHunter in Conda environment

After successfully installation of Anaconda, you can easily create a virtual environment for LipidHunter using the built-in graphic interface "Anaconda-Navigator".

After creating the virtual environment, go to channels menu and add following channel:

- `conda-forge`

Refresh the index list after updating the channel list, then install following python libraries:

```
cython matplotlib mkl natsort numba numpy numexpr
opencv openpyxl pandas plotly pyside scipy xlrd xlwt
```

The pyside package may result in a downgrade of Python from Python 3.6.5 to 3.6.2 or 3.4.x, and in rare case may downgrade to Python 2.7 in some Linux version, please accept the changes and continue.

After successful installation of above mentioned packages, start a new terminal, activate the virtual environment, and install pymzml version 0.7.8 using following command:

```
pip install pymzml==0.7.8
```

Now your environment is ready.

Navigate to LipidHunter source code folder and use following command to start LipidHunter GUI:

```
Python LipidHunter.py
```

If any import error occurs, try to install the missing package by conda or pip command.

6.2. Notes for Linux users

For Linux users, additional packages may be required. Please try to install following packages before starting LipidHunter (example under Ubuntu Linux, package name for Fedora or other Linux distribution may be different, please search for corresponding package names):

```
sudo apt-get install libsm6 libxrender1 xfonts-base
```

Linux version may have GUI text display issue due to customized font settings. You can change font size settings or use screen shots from Windows version to read corresponding text.

6.3. Notes for macOS users

macOS users may have GUI text display issue due to display settings. You can change display settings or use screen shots from Windows version to read corresponding text.

Pyside was observed to have compatibility issue in old macOS versions. LipidHunter is fully tested under macOS Sierra (version 10.12). For users failed to install pyside, LipidHunter terminal version or KNIME integrated LipidHunter can be used alternatively.

6.4. Notes for developers

The LipidHunter project is open for developers, if you made changes of LipidHunter source code, please find us on GitHub project repository (https://github.com/SysMedOs/lipidhunter) and feel free to setup a pull request or report your issue.

7. Use LipidHunter in terminal

LipidHunter terminal version can be used from source code or by compiled windows executable.

A configuration file containing all parameters is required for the terminal version of LipidHunter.

7.1. Prepare configuration files

The configuration file can be generated by the GUI or by manual construction of a plain text file in txt format.

- The configuration files generated for batch mode by LipidHunter GUI can be used directly.
- User can also edit the log file of LipidHunter runs or create new files using the same template.

7.2. Run commands

For windows version:

- Start PowerShell and navigate to LipidHunter folder:
- Run command like: `cmd_lipidhunter.exe -i configuration_file.txt`
    - Example: `cmd_lipidhunter.exe -i PC_cfg.txt`
- To get help, run command: `cmd_lipidhunter.exe -h`

For source code version:

- Start the terminal
- Activate the python environment for LipidHunter
- Run command like:
    - `python cmd_lipidhunter.py -i configuration_file.txt`
    - Example: `python cmd_lipidhunter.py -i PC_cfg.txt`
- To get help, run command: `python cmd_lipidhunter.py -h`

8. Use LipidHunter in KNIME

KNIME is a free and open platform for data analysis. LipidHunter can be integrated into KNIME workflow through Python Nodes, which allows users to construct customized pipeline for complex data analysis projects.

Currently, only LipidHunter source code version can be executed through Python extension in KNIME. The Windows compiled executables of LipidHunter cannot be used in KNIME.

Please download and install the latest 64bit version of KNIME from: https://www.knime.com

8.1. Prepare Python environment in KNIME

Before launch KNIME, please make sure that you have created a Python environment for LipidHunter and successfully pass the test of LipidHunter source code terminal version.

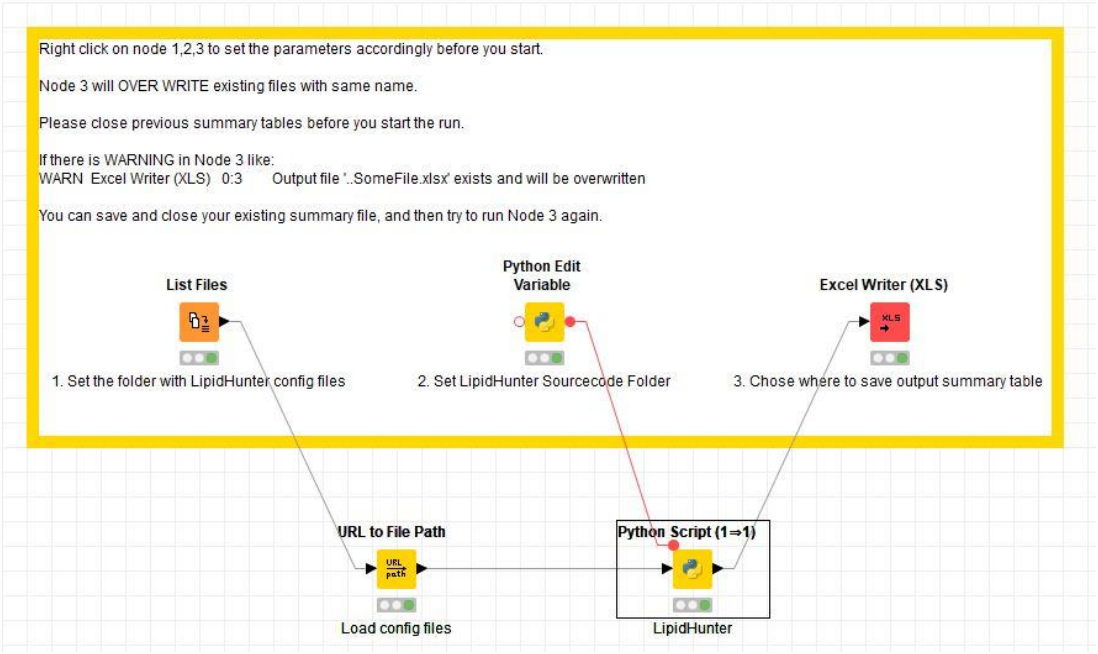
Follow the KNIME official guide (https://www.knime.com/blog/how-to-setup-the-python-extension) to install Python extension in KNIME and set Python path to the previous created LipidHunter environment.

8.2. Import sample workflow

Follow the guide to import the "LipidHunter_KNIME.knar" workflow file into KNIME:

https://www.knime.com/knime-introductory-course/chapter1/import-export-workflows

Following sample workflow will be displayed.



Edit node 1, 2, and 3 according to instructions, and then start the analysis.

The LipidHunter Python Script node is the core of the workflow, which can be modified to build customized workflow.