Lipid Data Analyzer 1.5

Examples

0 Customization to FT parameters

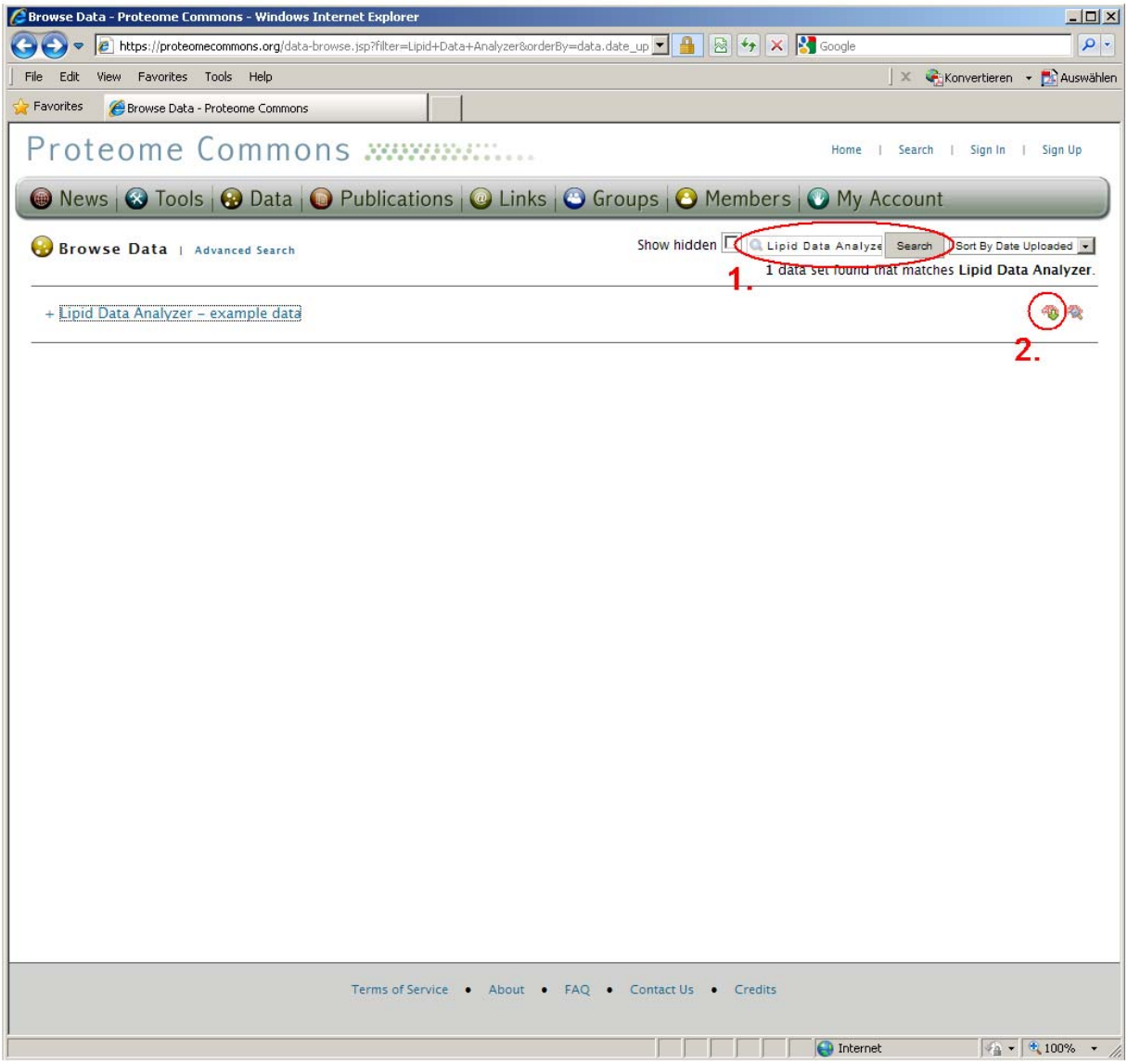
The provided example files are FT data. In order to show the data appropriately in the 3D viewer, to make quantifications in the viewer, or to re-quantify the data automatically the LDA has to be reconfigured to FT machines (just if in the installation not FT has been selected). For the visualization in heat maps and bar charts the adaptations described here are not necessary. Despite the installation to your corresponding MS machine, the installed version contains the parameters for other machines as well in the LDA/properties directory (LDA corresponds in this document to the installation directory of the LDA, e.g. C:\Program Files\Lipid Data Analyzer). In order to change to a different MS machine (e.g. FT) follow this instruction:

- Make a backup copy of your LDA/LipidDataAnalyzer.properties (for later usage)
- Copy the LDA/properties/LipidDataAnalyzer_$Machine_I_need$.properties file (e.g. LipidDataAnalyzer_FT.properties) to LDA.
- Rename the file to LipidDataAnalyzer.properties
- Restart the LDA and it will start with the changed settings

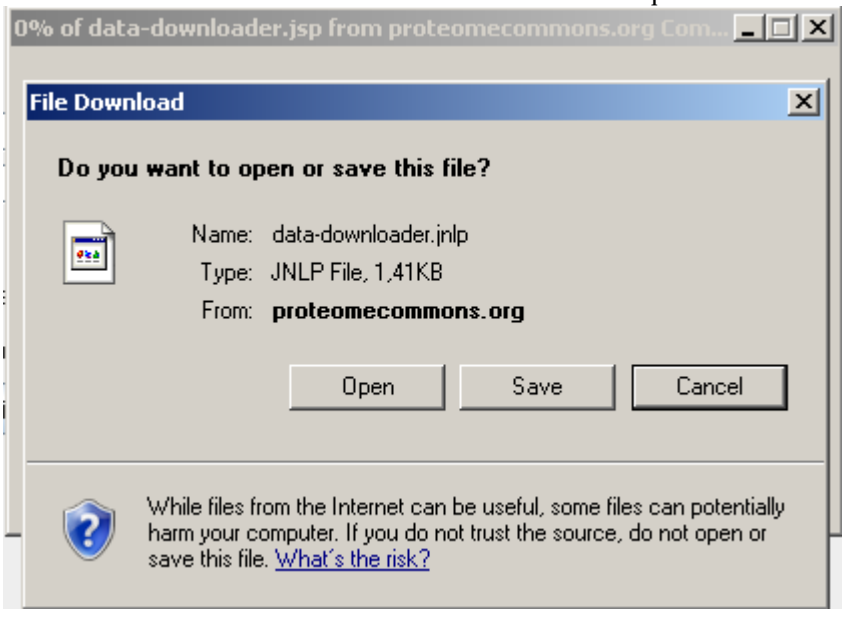
If you want to undo the changes (back to original installation), replace the LipidDataAnalyzer.properties file with your backup copy and restart the program.

1 Tranche example data download instructions

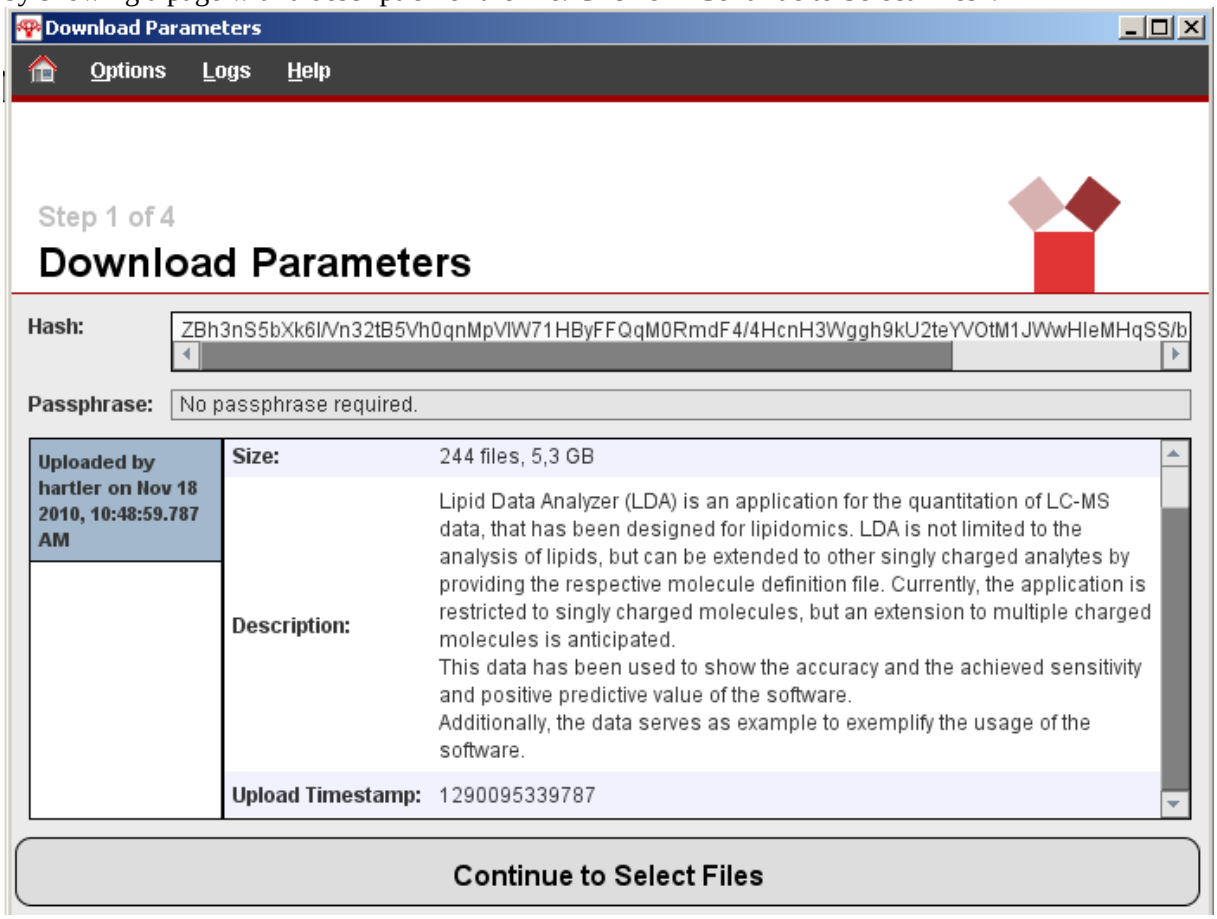
The example data is publicly available from the Tranche repository:
https://proteomecommons.org/tranche
There are two ways to access the data. The first one is to use the direct link:
http://genome.tugraz.at/lda/lda_download_example_data.php
The second one is to use the web interface. Go to the tranche home page, and search for “Lipid Data Analyzer” (1. in next figure). Then click on the download icon (2. in next figure).



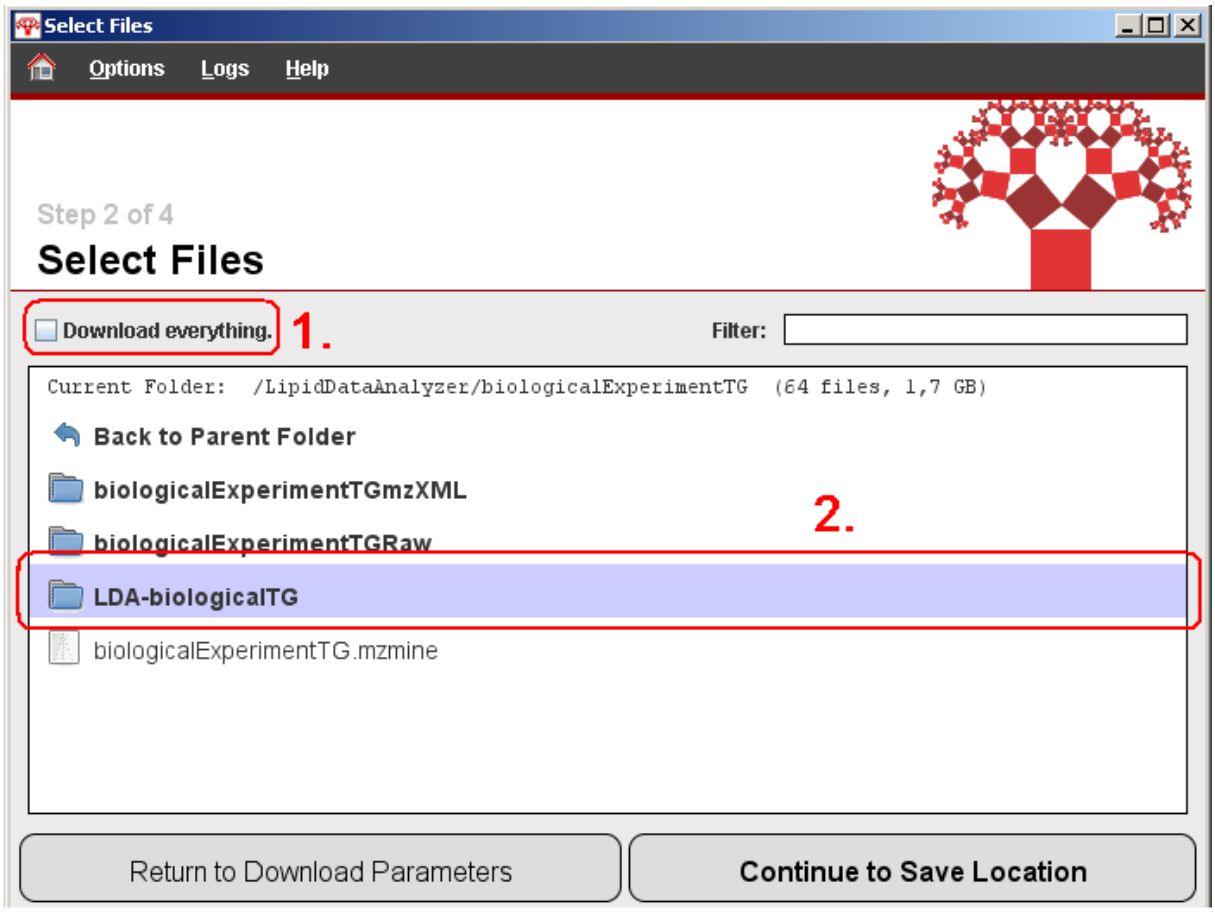
Then, in both cases, a popup will appear, that asks you if you want to download the applet for the downloader. For this downloader at least Java 1.5 is required.



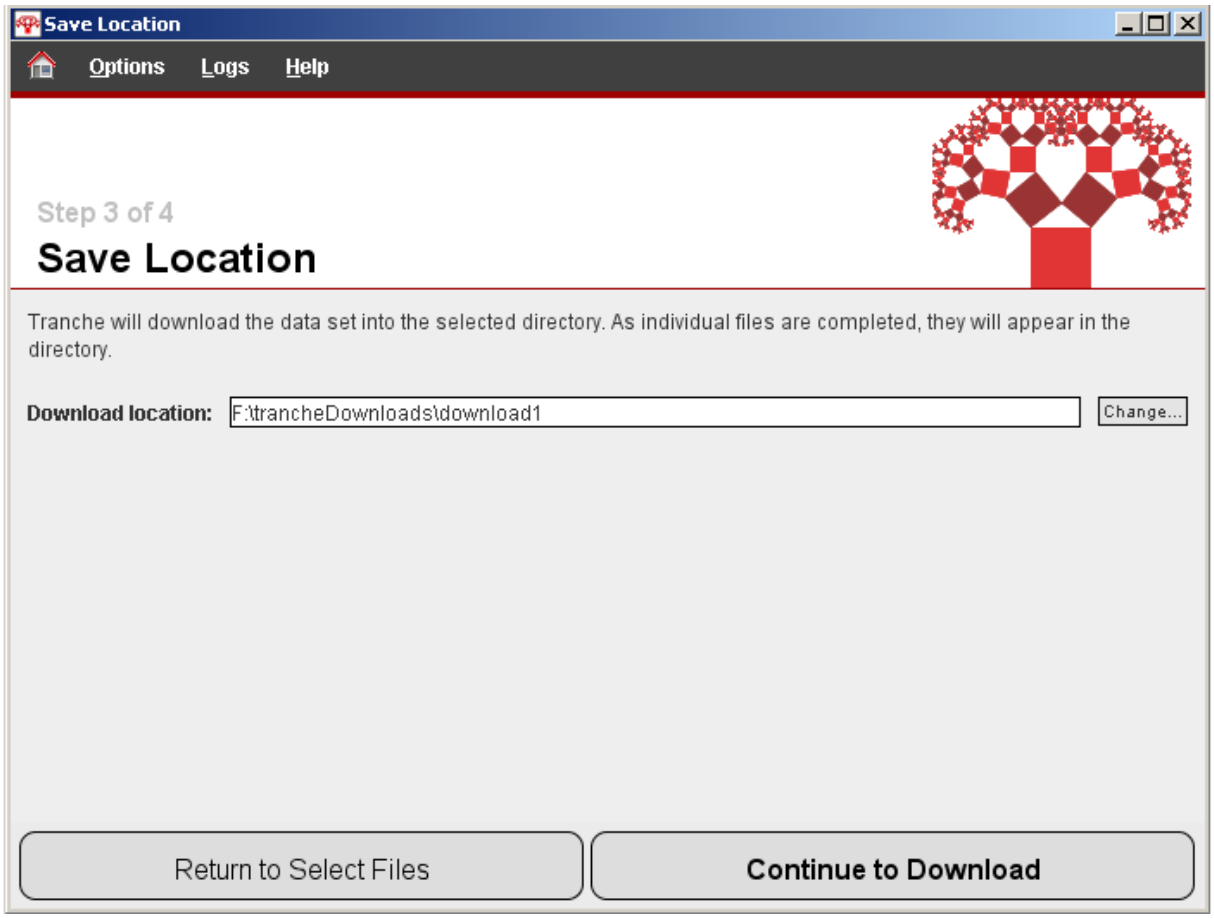
Click on “Open”, again a popup with a security warning might appear that asks if you want to execute the “ProteomeCommons.org Tranche Download Tool”, please accept. Next, the downloader will start by showing a page with a description of the file. Click on “Continue to Select Files”.



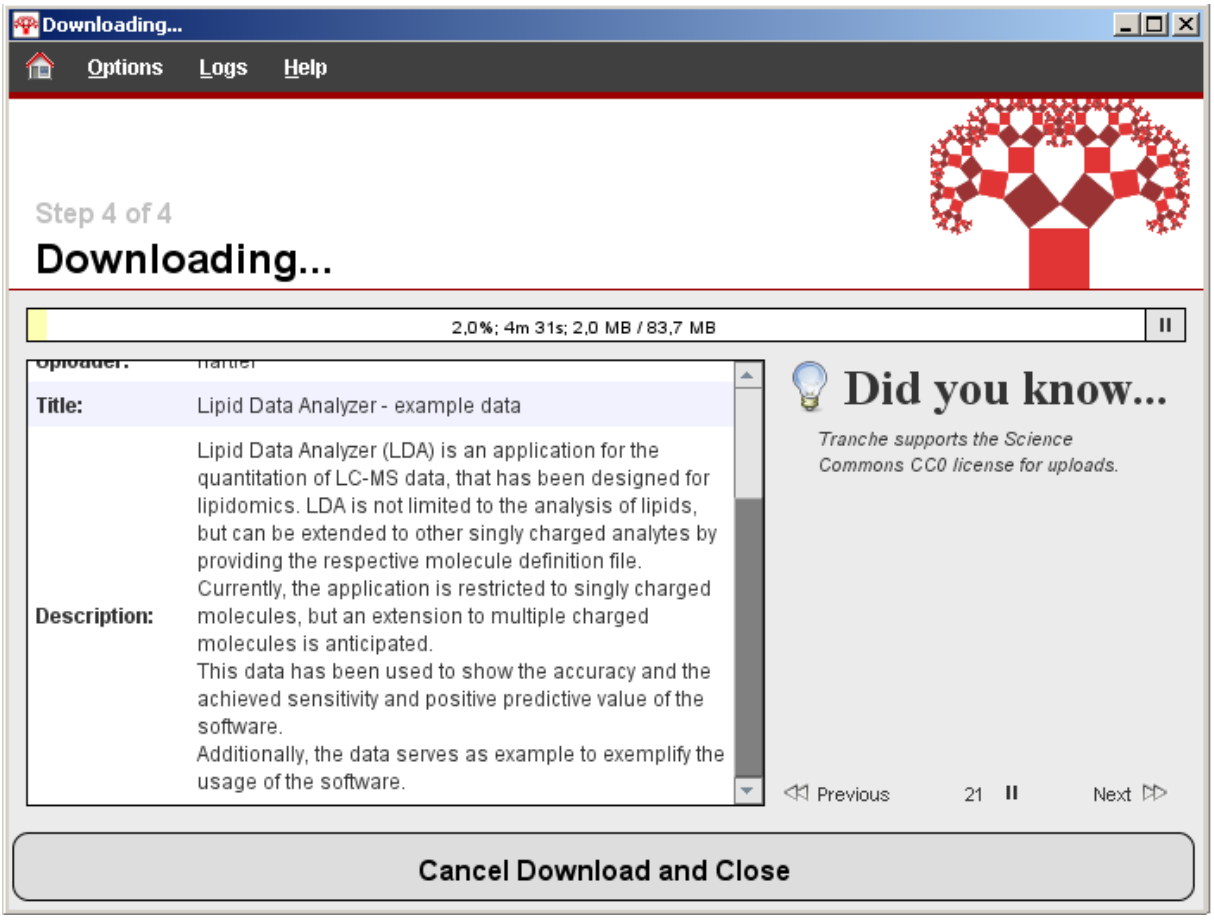
As you can see from the previous figure the data is 5.3GB which is quite huge. Of course it is possible to download everything however, for the examples not all of the data is required. The examples require the data in chrom format, whereas the data is additionally available in XCalibur RAW mzXML format, which is available rather for completeness. The most of the data volume is produced by the mzXML files. In order to select just the parts required unselect the check box “Download everything” (1. in next picture). Then a folder icon will appear called “LipidDataAnalyzer”. Now browse to the folder you need for your example and select it (don’t open it with double click; 2. in next figure). The paths to the required data for the specific examples are given in the examples. E.g. in the figure I chose /LipidDataAnalyzer/biologicalExperimentTG/LDA-biologicalTG which is required for the example in chapter 2. Then click on “Continue to Save Location”.



In the next page, you can select the location on your PC where you want to store the data. Comment: Tranche will keep the path structure in the repository. E.g. if you selected the same path like in the example above, and you chose e.g. C:\trancheDownloads as your destination directory, you will have the folder “LDA-biologicalTG” in the folder:
C:\trancheDownloads\LipidDataAnalyzer\biologicalExperimentTG



Then the download will start. Wait until it is finished. As you can see from the file size the data is much smaller than the 5.3GB total data volume.



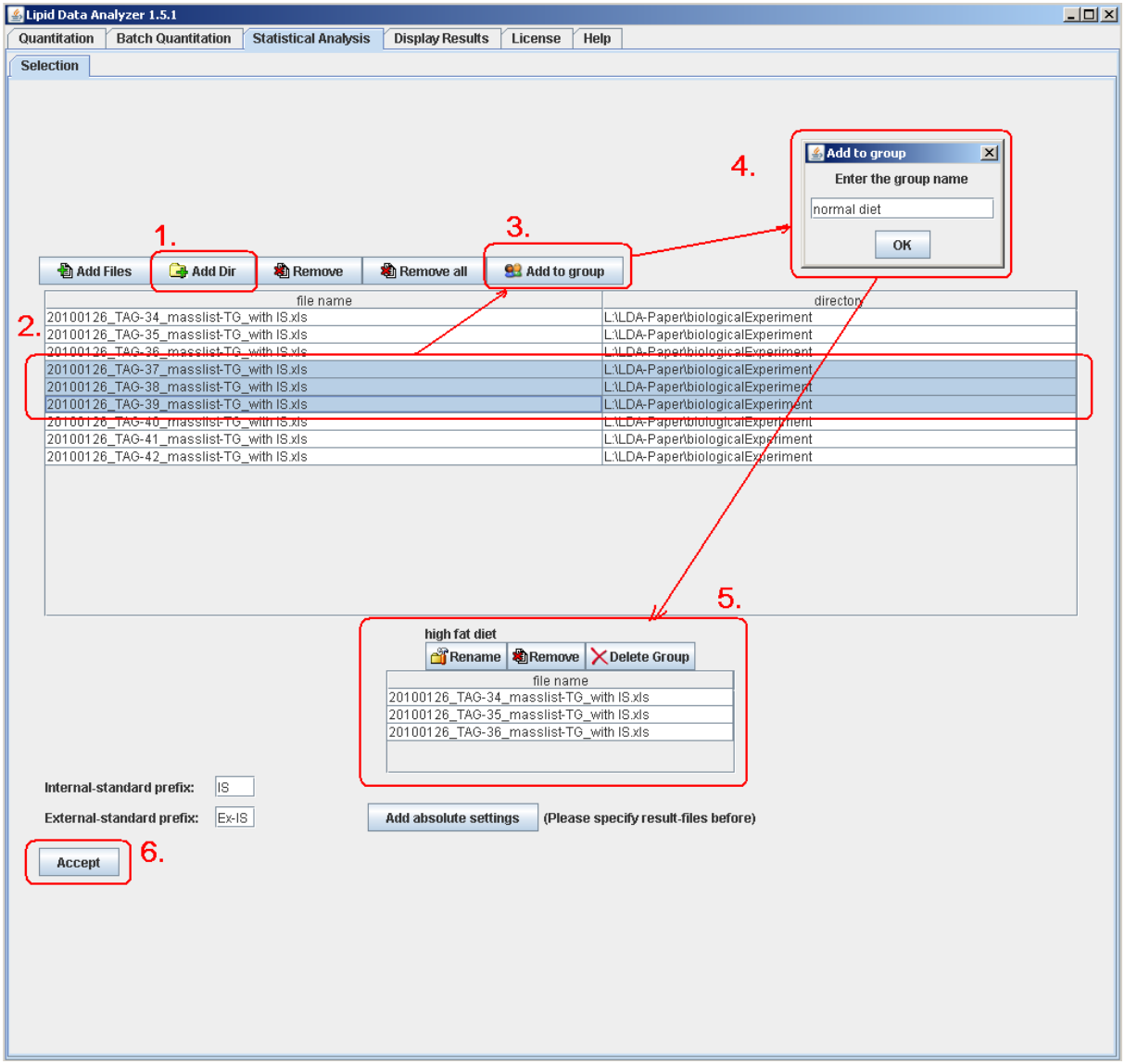
2 Visualization quantitation results with the aid of the published TG data

This chapter should show how the quantified data can be visualized with LDA and exported. Furthermore, it explains how to work with the 3D viewer for the manual correction. I started the examples section with this part, because it shows the features of the LDA, even though the quantitation (see chapter 3) would be the first part of the workflow. The test data is the one that has been used for the calculation of the sensitivity and positive predictive value (the example will reproduce Figure 3, Supplementary table TS2 and subsequently Table 2 of the publication). Since this data is FT-data, LDA should be set to use the FT parameters (see chapter 0) if necessary.

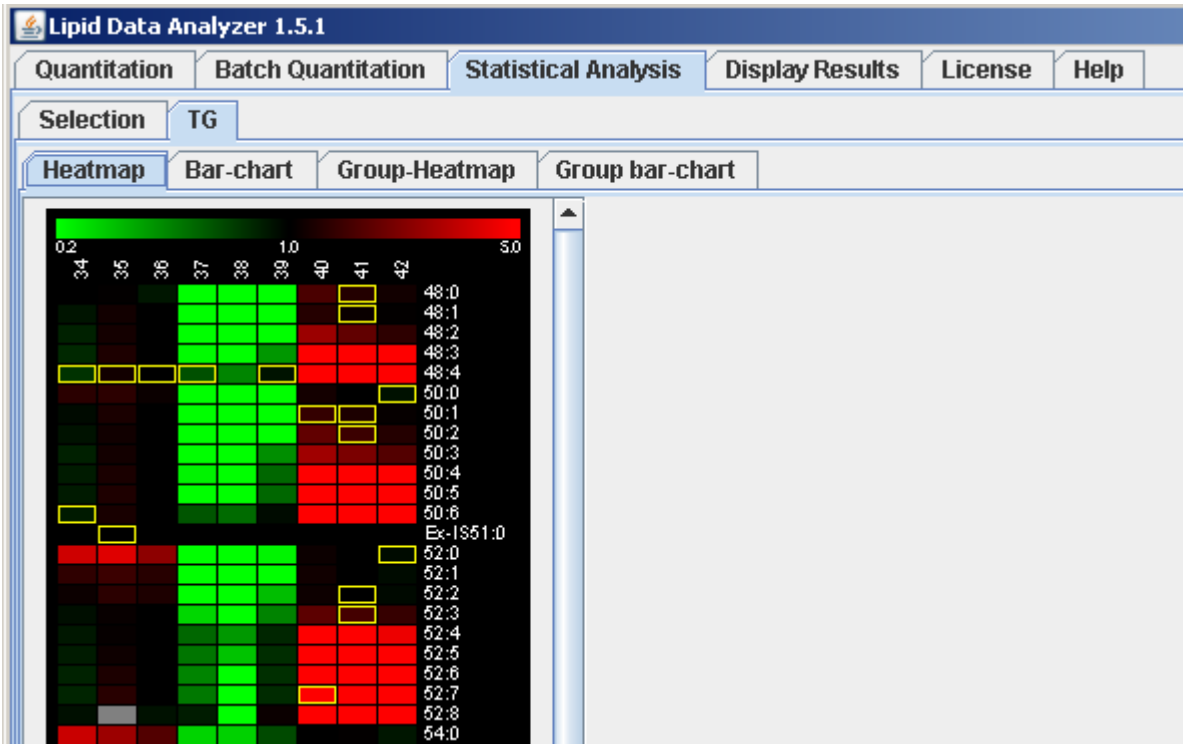
2.1 Selection of data

Please follow the following instructions to select the data:

- Download the data from Tranche repository as described in chapter 0. **Just the data from the folder /LipidDataAnalyzer/biologicalExperimentTG/LDA-biologicalTG is required.**
- Start the LDA and click on the "Statistical Analysis" tab.
- In this tab click on the "Add Dir" button (1. in next figure). Then select the folder that contains the downloaded files (the one that has the "xls" file in it; these contain the results from the quantitation, e.g. C:\trancheDownloads\LipidDataAnalyzer\biologicalExperimentTG\LDA-biologicalTG). If you then click OK in the directory chooser the same files (first column) should appear in the top table like in the figure below, with the proper paths in column 2.
- Next, the experiments should be added to groups (2. in next figure). For this step select the files from the table (same selection like in Windows; with "Shift" you can select items by clicking on the first and last entry of the list; with "Ctrl" you can specifically select single entries).
- Then click on "Add to group" (3. in next figure) and a popup will appear where you can enter the name of the group (4. in next figure).
- After you clicked "OK" the group will appear as a separate table at the bottom (5. in next figure).
- In order to reproduce the published data, select 34-36 and call the group "high fat diet", 37-39 "normal diet" and 40-42 "fasted".
- Then click "Accept" (6. in next figure) and the program starts to calculate statistics about the file (normalization etc.). After some time the heat map will appear. If you want to read more details about the results selection, please visit chapter 5 of the user manual.



The first heat map that will be displayed will look like this:

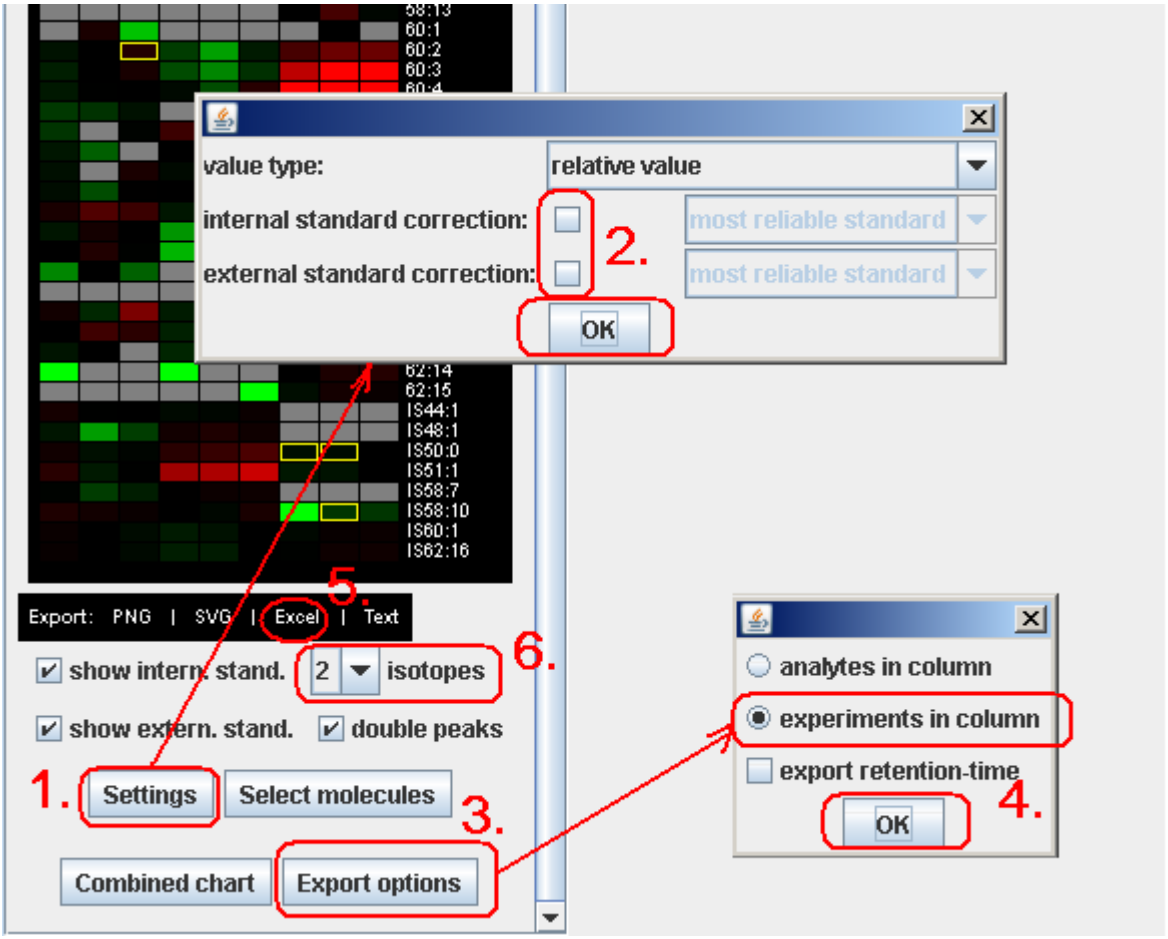


The output has now different tabs. Below the main tabs you can see “Selection” and “TG”. “Selection” is the page where you can select your results. Then tabs for the lipid groups will follow (here just TG because just TG is quantified). In the next row below you can see a tab for the “Heatmap” and one for the “Bar-chart” (the bar chart is currently empty), and the same for the group (high fat, normal fasted).

2.2 Data export and usage of the 3D viewer

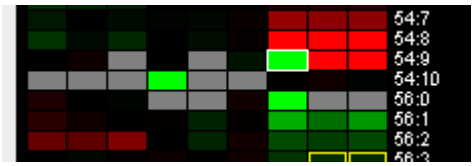
This example should show how to export data to other file formats and the possible settings. Furthermore, the access to the 3D viewer is description. These instructions will reproduce table TS2 of the publication, which has been used for Table 1 in the publication.
A detailed description about the features of the heat map can be found at chapter 5.1 of the user manual. The 3D viewer is described in detail at chapter 4 of the user manual.
To extract the data in TS2 please follow these instructions:

- Click on the settings button at the bottom of heat map (1. in next figure).
- A dialog window will appear where it is possible to select the normalization of the values (2. in next figure). For a detailed description about the visualization settings, please visit the user manual chapter 5.2. For the determination of the sensitivity and the positive predictive value a normalization on a standard is not necessary, thus deselect the checkboxes after “internal standard correction” and “external standard correction (2.) and click OK.
- Then click on the “Export options” button at the bottom (3. in the next figure).
- The data in TS2 has the names of the samples in the column and the analytes in the row. Thus choose the radio button “experiments in column”, so that the data will be exported in the same manner and click OK (4. in next figure).

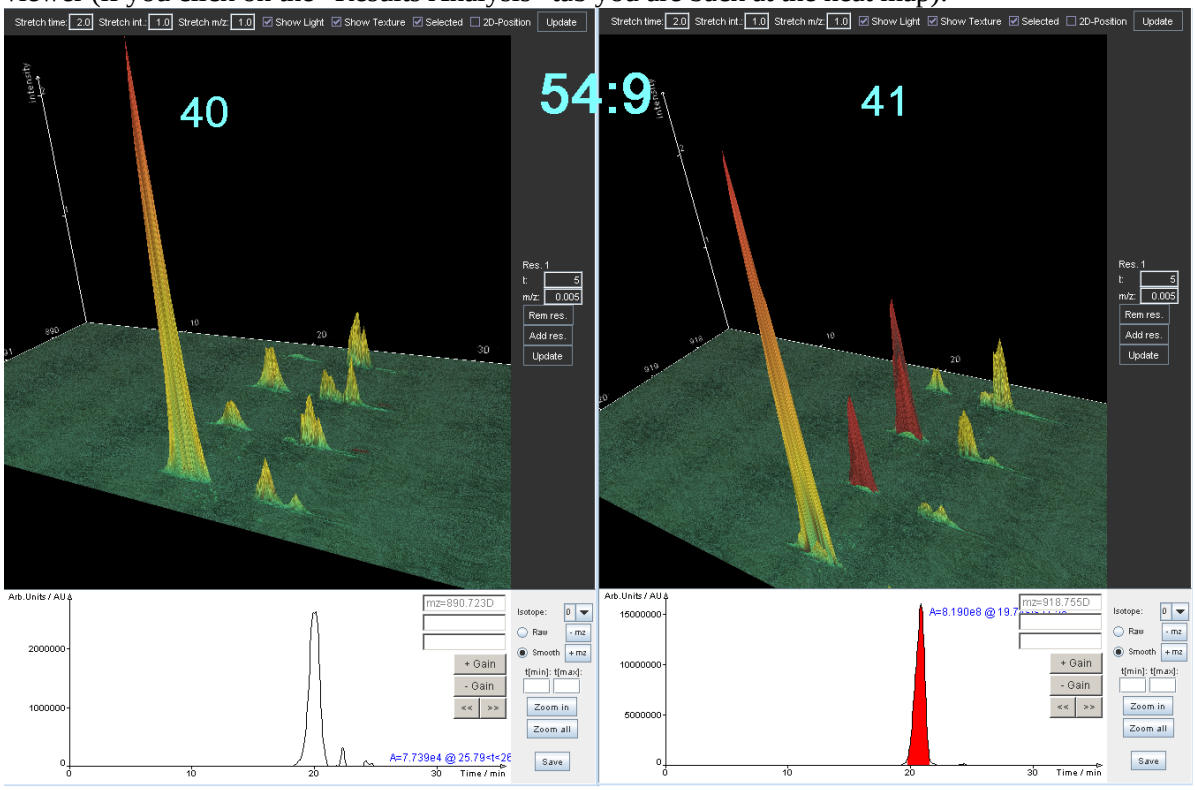


- Click on “Excel” to export the data in an Excel file (5. in next figure). The exported values will be the same like in TS2 sheet “assignments”. The only difference: 58:14, 60:0, 60:14 and 62:16 are missing, since they have not been identified (which is correct, since they have not been detected by MS). They have been added manually to TS2 since they have been searched for. For the sheet “assignments-relative”, you have to select value type “relative to measured class amount” in the settings screen before exporting to Excel (2. in next figure). The exported values will be in per mille and have to be divided by 10 (in Excel) to get the used percent values. The rest is the same like in the export of “assignment” sheet.

To find out now which are the ambiguous hits (yellow colored ones in TS2 sheet “assignment”) is easy. A yellow rectangle surrounding a heat map cell indicates more than one candidate peak for this identification (this is not necessarily wrong, because the peak could be part of the peak tail and OK, it should just warn the user to be a little bit cautious with this hit and check it manually). This rectangle is not only drawn if there is more than one candidate for the 0 isotope, but every isotope is considered. In order to consider just the 0 isotope, change the “isotopes” value from 2 to 0. The ones that are now yellow are the ones that are reckoned ambiguous in TS2. Just the 0 isotope is taken, because the mzMine2 returns just the 0 isotope, thus a comparison between the programs is just possible for this isotope. Furthermore, all of the yellow appearing hits of LDA are counted as ambiguous for fairness reasons, since it is most of the times not possible with mzMine2 to figure out if the ambiguous hit is just part of the tail or not, because the hits are normally reported in separate chromatograms.
To determine the wrong hits (red ones) requires manual work (looking at the chromatograms). This can be done with the aid of the retention time. Start with identifications with few double bonds (e.g. 54:0) and check the peak. Then check the peak of the hit with one more double bond (e.g. 54:1). If this hit is slightly before the retention time of 54:0 it should be correct. But sometimes the wrong identifications are easily found with the aid of the heat map. E.g. if you look at 54:9 at sample 40 of the group, you see that the value is green, while the members of the group (40-42) show a red value.



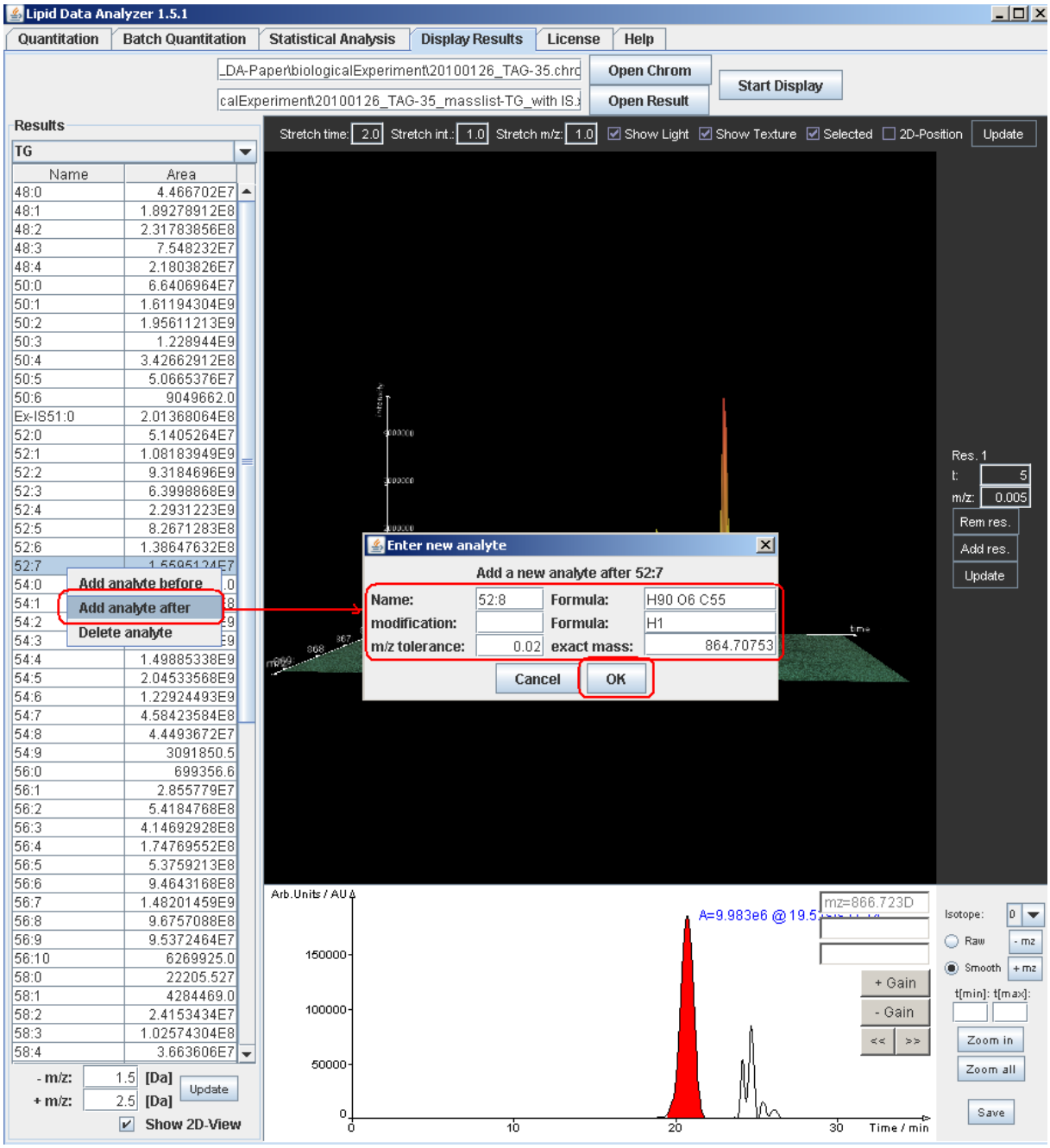
If you hover the mouse over the heat map cell of 54:9 a white rectangle will appear. If you click the left mouse button the program will jump to the “Display Results” tab and show the hit in the 3D viewer (if you click on the “Results Analysis” tab you are back at the heat map).



The last figure shows the comparison of 54:9 found in sample 40 and 41. In 40 definitely the wrong peak has been identified.

If you want to check if the algorithm missed a peak or if the peak is there at all, you have to enter the molecule manually. Go to the “Display Results” tab (or click directly on a cell in the heat map), click with the right mouse button on an analyte before or after the one you want to add. A popup will appear where you can select if you want to add the analyte before or after this one. Then, a dialog window will appear, where the properties of the peak can be added (the m/z value of the analyte can be found in the provided examples/masslist-TG_with IS.xls; the mass in the column mass(…) has to be taken).

A detailed description about the usage of the 3D viewer can be found in the user manual at chapter 4.

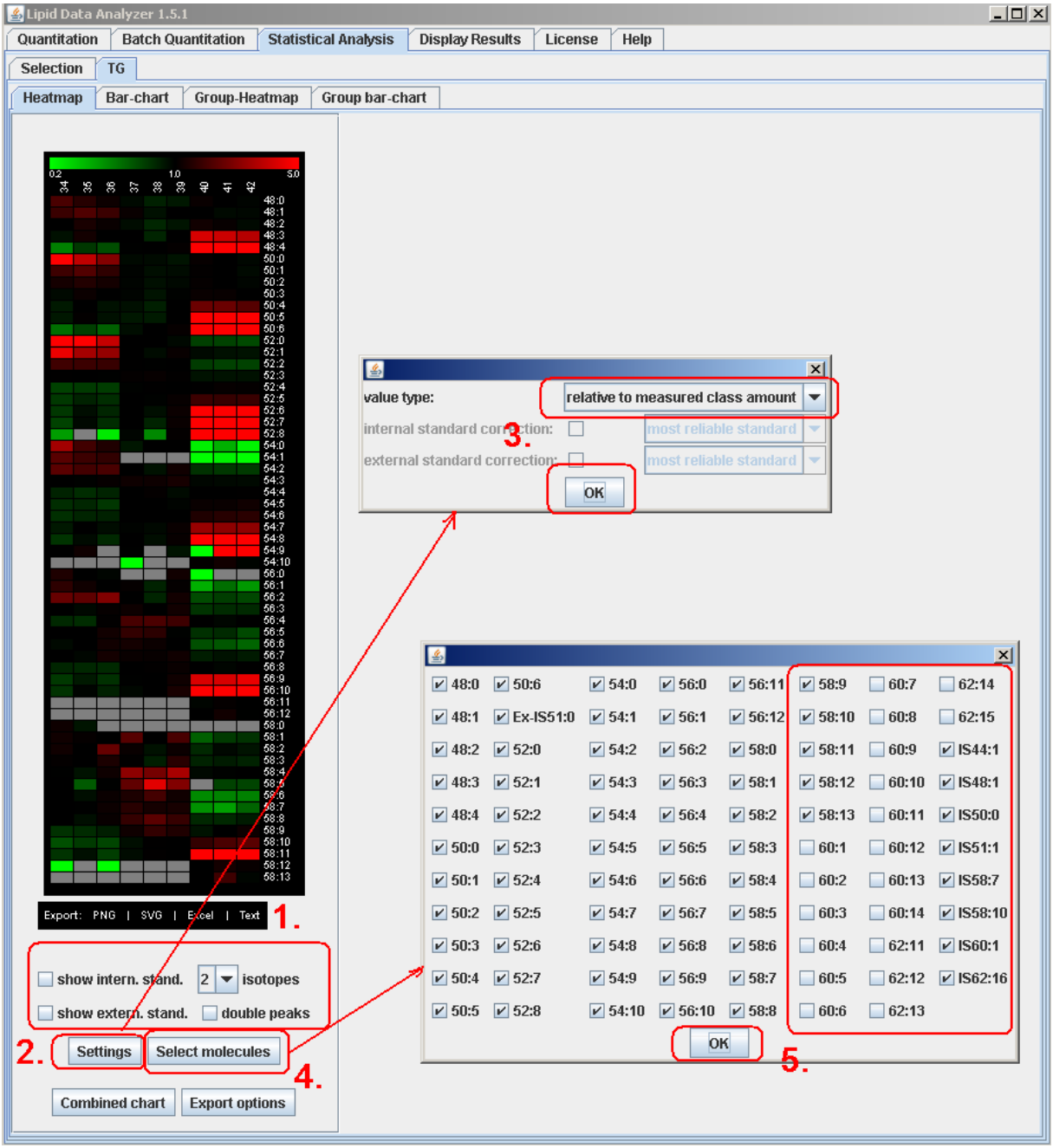


The manual validation was entered in TS2 sheet “assignments”, which served as source for the LDA values for the sheet “details” in TS2, and the resulting values have been used to calculate the sensitivity and the positive predictive value.

2.3 Export of figures of heat maps and bar charts

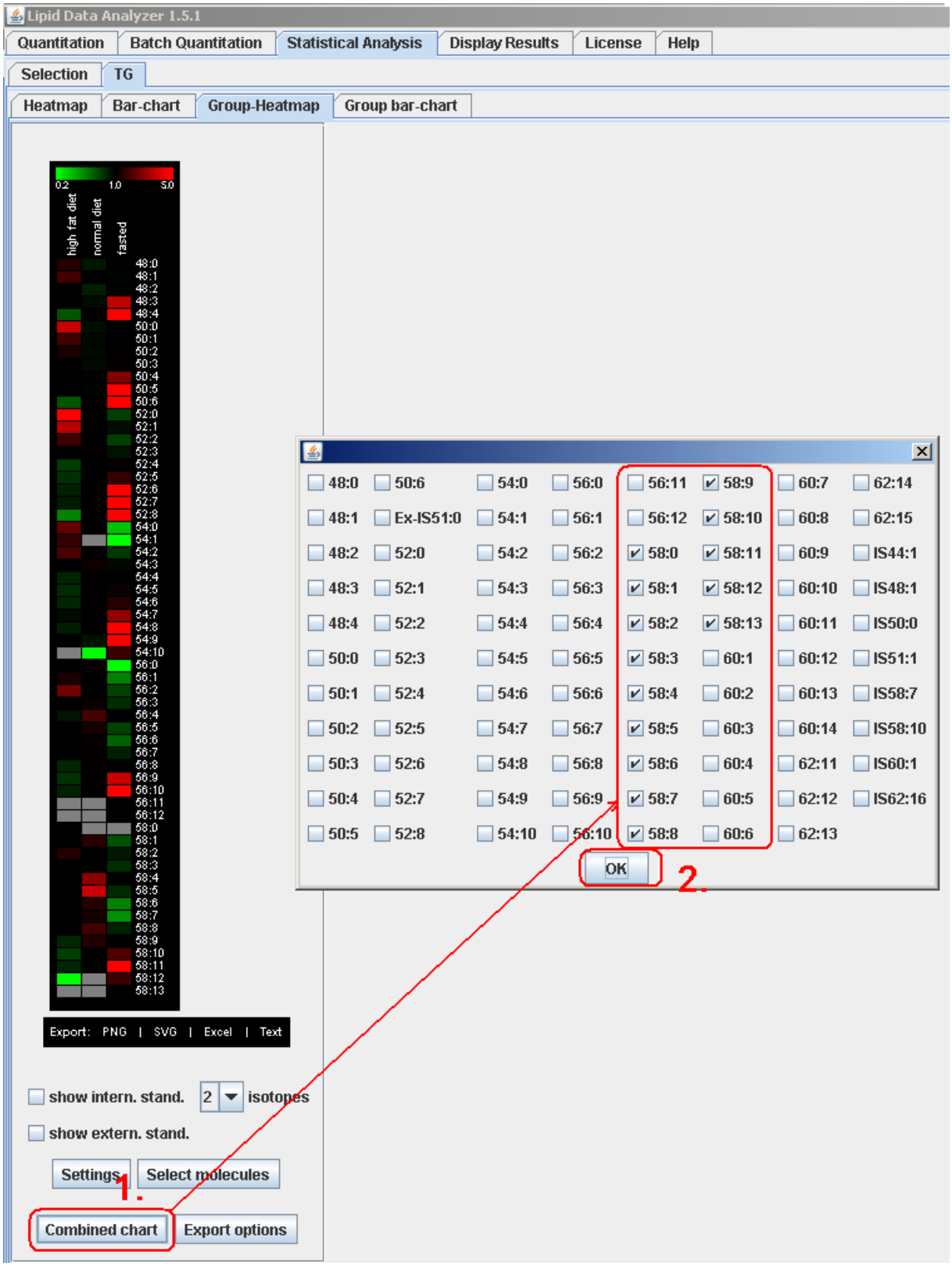
This chapter shows how export figures. As example the figure 3 of the paper is reproduced. First, the data has to be selected as describe in 2.1 and the heat map must be visible. Please follow the following instruction to visualize the heat map:

- First (in next figure), deselect “show intern. stand.”, “show extern. stand.”, and “double peaks.
- Second (in next picture), click on “Settings”
- Third (in next picture), select “relative to measured class amount” and click OK.
- Fourth (in next figure), click on “Selected molecules”.
- Fifth (in next figure), unselect the molecules of 60, and 62 and click OK
- The heat map can now be exported as “PNG” or “SVG”.

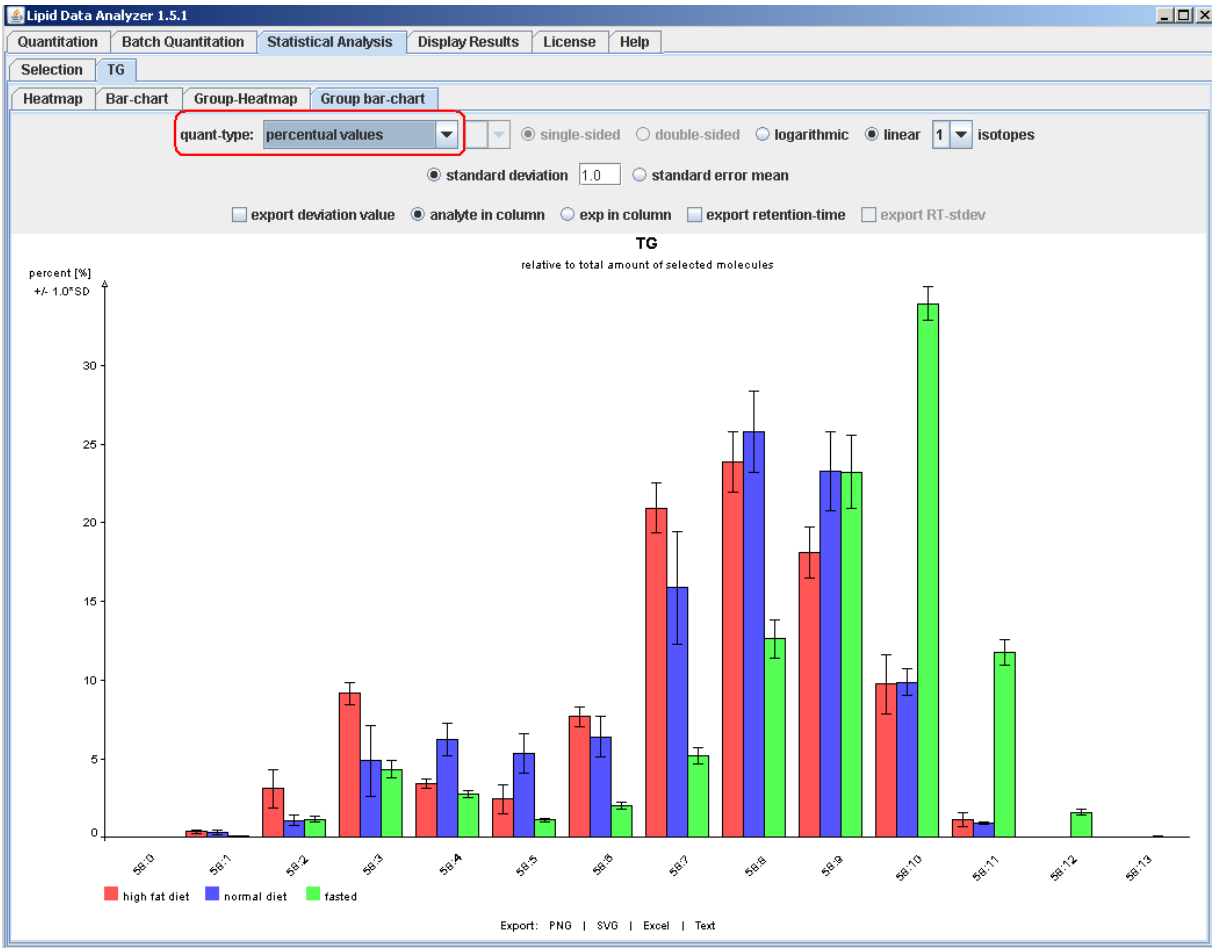


To show the group bar chart please follow the following instruction:

- First, click on the tab “Group-Heatmap”.
- Second, click on the button “Combined chart”. This is for the manual selection of analytes for the heat map. Bar charts for single experiments or analytes can be generated directly in the heat maps by clicks on the name of the analyte or experiment (details see user manual chapter 5.1)
- Third, select all of the 58 molecules and click OK. Then the bar chart will appear.



In the bar chart select as “quant-type” percentual values and the same figure like in figure 3 will appear. “percentual values” means that all of the selected molecules correspond to 100% and the contribution of a single analyte is reported in per cent. This “quant-type” can be used to detect changes in the distribution of lipids with the same amount of C atoms, but a different amount of double bonds. The picture can be exported as PNG or SVG. More details about the bar chart can be found in the user manual chapter 5.3.



3 Starting the quantitation

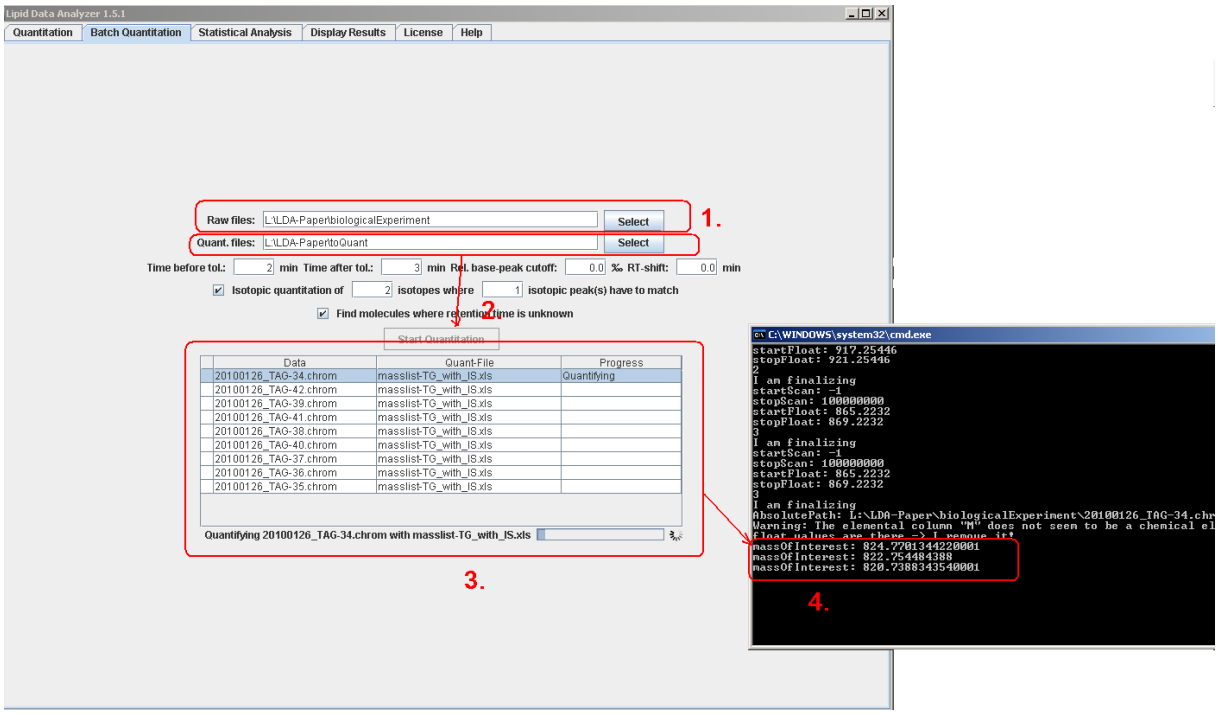
If you download the data as described in 2.1, the folder contains 4 different file types (chrom, head, idx, rtt, xls).

- chrom, head, idx, and rtt: these files belong to the chrom file format and the quantitation is executed on them. These files allow fast access to the data and settable resolution levels. In the chrom the data is stored in small chromatograms (standard setting of one chromatogram is 1mDa range). The head file stores general information about the chrom file. The idx is a binary file that contains indices to allow quick access to parts of the chrom file. The rtt file stores the retention time for the corresponding scan number.
- xls: these are the result files of the quantitation.

The quantitation works based on the chrom file (chrom, head, idx, and rtt). If you want to be sure if the Excel files are really quantified, delete the Excel files from the folder.

The first thing to start batch quantitation is to copy the file LDA/examples/masslist-TG-with_IS.xls to a directory that does not contain any other Excel files. This Excel file contains the analytes to look at, their chemical formula and their m/z values. A detailed description about how to prepare such an Excel file and the options of the quantitation page can be found at the user manual chapter 2 and 3. Next, start the LDA and click on “Batch Quantitation”. Press the “Select” button next to “Raw files” to open the directory that contains the data files, and the one next to “Quant files” for the directory where you put the masslist-TG-with_IS.xls (1. in next figure).
Then, set the “Rel. base-peak cutoff to 0.0‰. With this value, very small peaks can be neglected from analysis. In the original analysis no analytes have been neglected, so please set the value to 0.0 if you want to exactly reproduce the results. Then click on “Start Quantitation” (2. in next figure).
Note: The quantitation can take quite long. You can continue with the rest of the examples while it is running. If you want to stop it, close the program. The program scans now the “Raw” folder for RAW,

mzXML or chrom files (in our case just chorm files are there), and the “Quant” folder for Excel files. Then a table will appear that contains the data files (chrom, mzXML or RAW) and the Excel file that has been used for quantitation (3. in next figure). If you started the LDA with the console window, you can see additionally in the console which analyte is currently quantified by the program (4. in next figure). If you deleted the result Excel files, you can see them regenerated after the quantitation has finished.



It is possible to start the quantitation based on the XCalibur RAW format or on mzXML files. However, you can test the RAW files just on machines where XCalibur is installed, because the used program ReadW requires a dll of XCalibur!! The data can be download from the Tranche repository (these files are quite huge). Just download the files as described in chapter 0 with the following path settings

- /LipidDataAnalyzer/biologicalExperimentTG/biologicalExperimentTGmzXML for mzXML
- /LipidDataAnalyzer/biologicalExperimentTG/biologicalExperimentTGmzRaw for RAW

The only difference of this calculation compared to the ones before is that you have to select these folders for the “Raw files” below. If you choose RAW the data is first translated into mzXML. Then in both cases the files are translated into chrom. If your initial files are RAW files the mzXML files are deleted automatically since they are quite huge and are just needed for the translation to chrom.
For the other experiments, the data is available in mzXML and RAW as well.
For the phospholipid data:

- /LipidDataAnalyzer/biologicalExperimentPhospholipids/biologicalExperimentPhospholipidsMzXML
- /LipidDataAnalyzer/biologicalExperimentPhospholipids/biologicalExperimentPhospholipidsRaw

For the controlled experiment:

- /LipidDataAnalyzer/controlledExperiment/controlledExperimentMzXML
- /LipidDataAnalyzer/controlledExperiment/controlledExperimentRaw

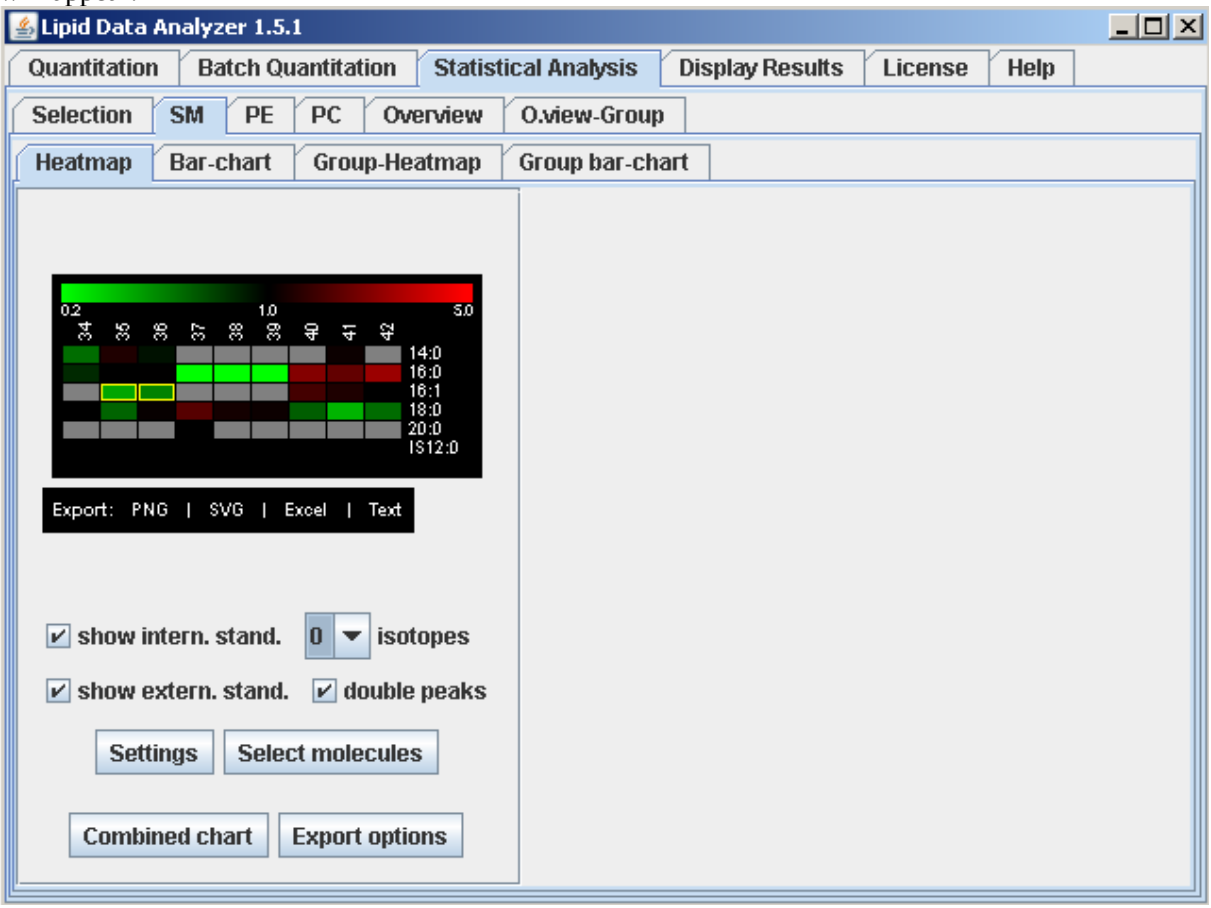
The mass list files are available in this distribution:

- LDA/examples/masslist-TG_with_IS.xls: for the biological TG data
- LDA/examples/masslist-positive_with IS_wo_TG.xls: for the biological phospholipid data
- LDA/examples/TG_with_IS.xls: for the controlled experiment

Moreover, further mass list files, containing different lipid classes, are available in the LDA/exampleMassList directory.

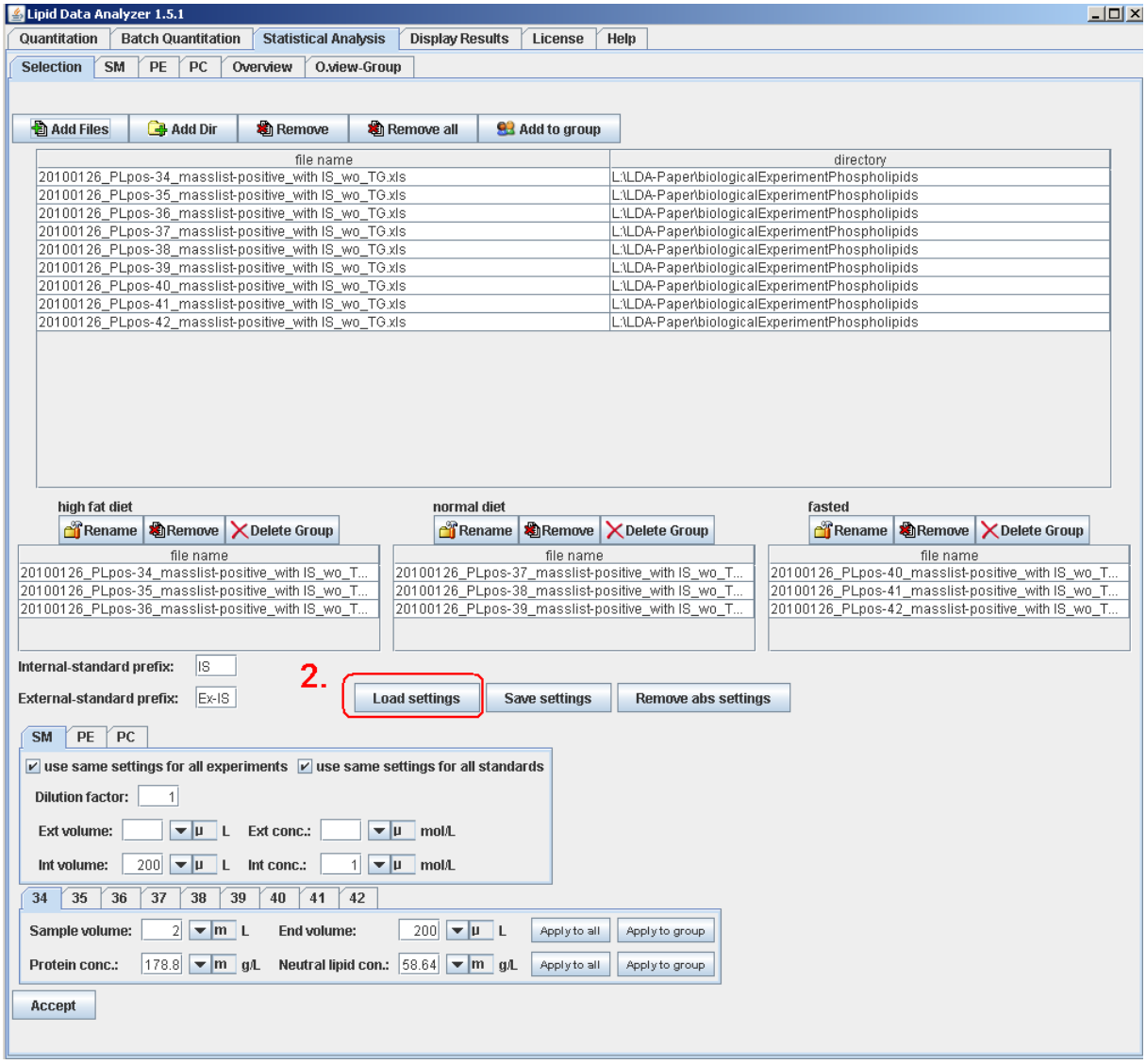
4 Visualization of several lipid classes the phospholipid and sphingomyelin data from the biological experiment

The selection of the data is similar like in chapter 2.1, the only difference is that the data from /LipidDataAnalyzer/biologicalExperimentPhospholipids/LDA-biologicalPhospho has to be used. The end of this process will be the data used in table Supplementary table TS4 of the publication. Please follow the instructions at 2.1. After the "Accept" button has been clicked, a slightly different output will appear:

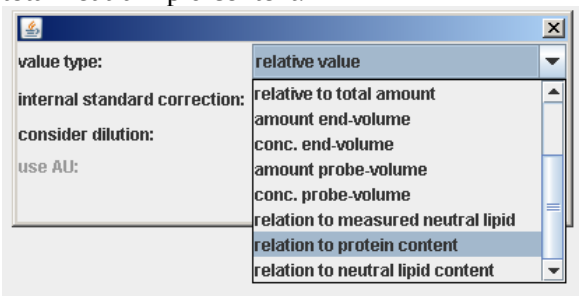


Instead of TG, the lipid classes SM, PE, and PC appear. Additionally the tabs "Overview" and "O.view-Group" appear that can serve as overview over the lipid classes. If you want to see an overview click on the tab "O.view-Group", "consider standard" and "consider dilution" should be selected, and click on "Accept". In the appearing dialog window, click on "OK".

In the appearing overview bar chart, the sum of the analytes of a group is taken as measurement for the total amount of a class. If the “consider standard” is clicked, the areas are standardized to correct ionization differences of the single lipid classes. In this version, this overview is just true if the same amounts of standards are added and the molecules are diluted equally. If you want to add specific content of standard you have to click on the “Selection” tab and then on the button “Add absolute settings”. Then click on “Load settings” and select the file LDA/examples/phospholipids.wqs.xml (LDA is the directory where LDA has been installed; see chapter 0), which contains the standard settings for this probe. The input fields are filled out then. Next, click on “Accept” that the new settings take effect.

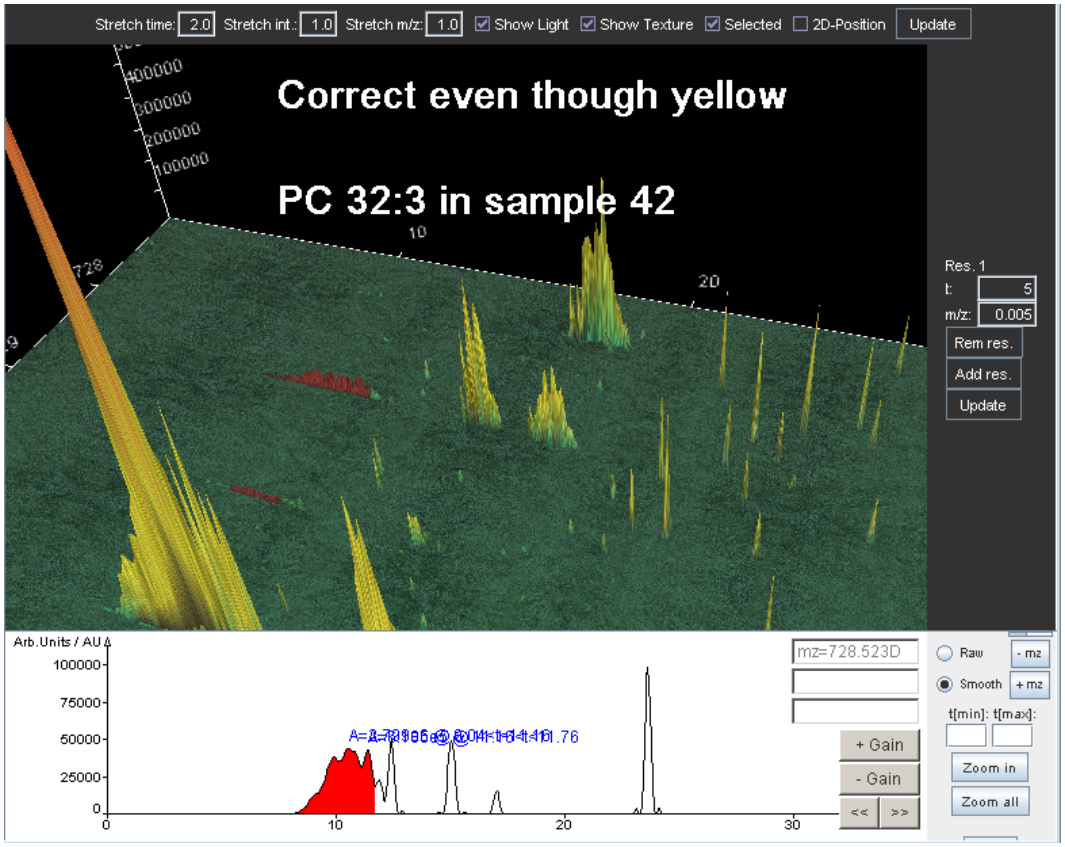


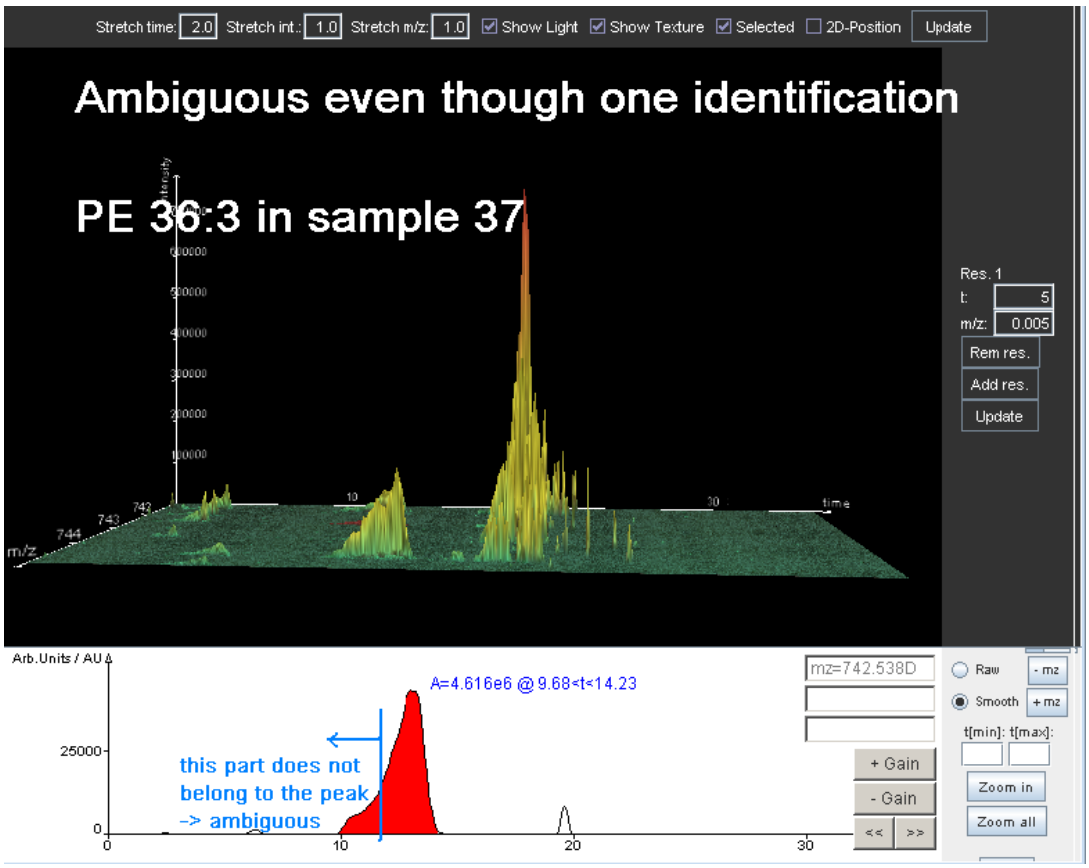
Nevertheless if you generate the overview panel, it will not look very different, since the standards are added at the same concentration. However, if you click on the “Settings” button (bottom of the heat map), there are many more value types available, like the normalization on the protein content or the total neutral lipid content.



If you want to have a more detailed description about the standard settings, please visit the user manual chapter 5.

The values for the published Excel sheet TS4 can be exported in the same way like in chapter 2.2, and for the sensitivity analysis the data has be checked manually as described in chapter 2.2. The only difference here, the analytes with a yellow rectangle around the heat map cell are not directly taken as ambiguous hit like in the comparison to mzMine2. We checked if the yellow rectangle is caused by separate identifications that belong together (no ambiguous hit) or if a correct hit is made by 2 peaks that do not belong together(is an ambiguous hit event though no yellow rectangle).



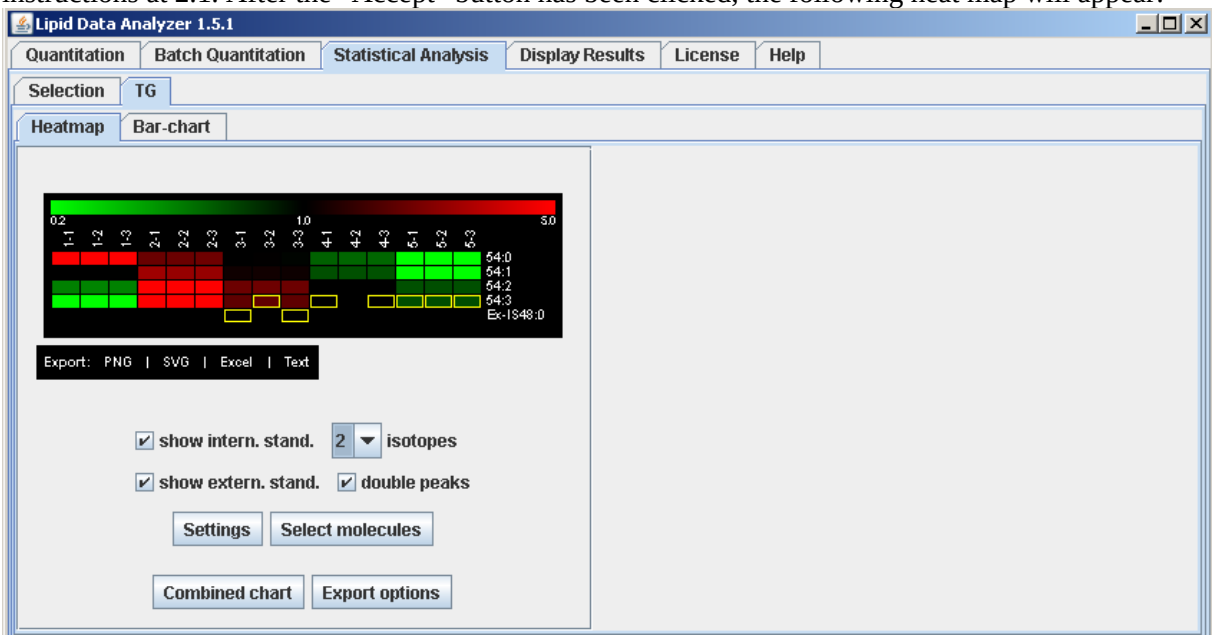


The quantitation works in the same manner like described in chapter 3, but instead of masslist-TG-with_IS.xls the file examples/masslist-positive_with IS_wo_TG.xls has to be taken. Furthermore, the "Rel. base-peak cutoff" was set to 0.05‰, and the "Time after/before tolerance" was set to 0.5min for

both input fields. This retention time setting has just been used for the quantitation of the internal standards.
In addition, for the same raw data the quantitation results including TG are available. They can be downloaded from Tranche using the directory:
/LipidDataAnalyzer/biologicalExperimentPhospholipids/biologicalExperimentPhosphoResultsCombinedTG
Here just the result files are present. To fully use them, backup and delete afterwards the Excel files in the LDA-biologicalPhospho folder und put the ones of this directory there. The rest works in the same manner.

5 Adducts/modifications

This chapter should exemplify how to work with analytes that appear with different adducts/modifications in the MS analysis. The used data is from the controlled experiment of the paper. The selection of the data is similar like in chapter 2.1, the only difference is that the data from /LipidDataAnalyzer/controlledExperiment/LDA-controlled has to be used. Please follow the instructions at 2.1. After the "Accept" button has been clicked, the following heat map will appear:



In order to reproduce the data used for publication (Supplementary table TS1 and subsequently Table 1), the data can be exported in the same manner described in chapter 2.2. The only difference: "isotopes" was set to 2 and not 0. The only novelty of this example compared to the previous ones is that the quantitation is based on peaks that are from analytes with different adducts/modifications. If you click now on one cell in the heat map, the application will jump to one of the identifications. You will see the other adducts/modifications next to the select hit. For the quantitation of this data, the file examples/TG_with_IS.xls has been used and can be seen as an example file for the quantitation of analytes with different adducts/modifications. The important thing is that the column header contains mass(name[…] form[…]) that the column is accepted for quantitation. A detailed description about how to prepare such Excel files can be found at the user manual Appendix A.