Lipid Data Analyzer
Study Data

1 Study data download instructions

The example data is available from our homepage:
http://genome.tugraz.at/lda2/lda_data.shtml
or from MetaboLights:
http://www.ebi.ac.uk/metabolights/
Data are available for three control experiments, one biological experiment, and the benchmarking of LDA against LipidBlast. Within the individual experiments, data is further subdivided to allow download of individual MS platforms, because we assume that users want to download only data relevant to their MS platforms. Details for downloading data from our homepage and from MetaboLights are given in the following two chapters.

1.1 Instructions for download from the lab's homepage

Study data is categorized in 'BENCHMARKING LIPIDBLAST', 'BIOLOGICAL EXPERIMENT', 'CONTROL EXPERIMENT 1', CONTROL EXPERIMENT 2', 'CONTROL EXPERIMENT 3', 'HCD CHARACTERIZATION' and 'IDENTIFICATION OF REGIO-ISOMERS'. Furthermore, the fragmentation rules and this description is provided in the sections 'FRAGMENTATION RULE SETS', and 'STUDY DATA DESCRIPTION', respectively.

For each platform in a category, a link is provided to download the corresponding data. In this figure, the link for downloading the biological results generated by Orbitrap Velos Pro in CID mode is highlighted.

Data from various experiments for various platforms contain typically the three folders: massLists (quantitation Excels used as input for LDA), and positive and negative for the polarities (except for control experiment 2, where there is no negative folder, since the experiments were performed in positive ion mode only). In most of the cases, the folders positive and negative contain MS runs, comments, and

evaluation results. Only for Control Experiment 1 for the platforms 'Orbitrap Velos Pro', 'Orbitrap Velos Pro HCD', 'SYNAPT G1 HDMS QTOF' and '4000 QTRAP', these folders are further subdivided in the collision energies used. Further details about the individual files are given in chapters 2 and 4, and their visualization in chapter 3.

1.2 Instructions for download from MetaboLights

Study data is organized in benchmarking (MTBLS397), biological experiment (MTBLS396), control experiment 1 (MTBLS394), control experiment 2 (MTBLS391), control experiment 3 (MTBLS398), and HCD characterization and *regio*-isomers (MTBLS462), and can be accessed by adding the identifier to the MetaboLights URL, e.g. http://www.ebi.ac.uk/metabolights/MTBLS394. In contrast to the homepage download, MetaboLights is organized much more granularly containing individual ZIP archives for each MS-run and data evaluation. For downloading data of a specific platform, click on the tab "Study Files", and type in the name identifying the platform, e.g. 'QTOF_G6550A', and press enter:

This will select all files containing the corresponding identifier in the name:

QTRAP6500_negative_Exp1_3.zip
QTRAP6500_negative_Exp1_Comments.zip
QTOF_G6550A_positive_Exp1_01.zip
QTOF_G6550A_positive_Exp1_02.zip
QTOF_G6550A_positive_Exp1_03.zip
QTOF_G6550A_positive_Exp1_Comments.zip
QTOF_G6550A_positive_Exp1_04.zip
QTOF_G6550A_positive_Exp1_05.zip
a_lda2_experiment1_metabolite_profiling_mass_spectrometry-38.txt
QTOF_G6550A_negative_Exp1_02.zip
QTOF_G6550A_negative_Exp1_03.zip
QTOF_G6550A_negative_Exp1_04.zip
QTOF_G6550A_negative_Exp1_05.zip
QTOF_G6550A_negative_Exp1_Comments.zip
QExactive_positive_Exp1_01.zip
QExactive_positive_Exp1_02.zip
QExactive_positive_Exp1_03.zip
QExactive_positive_Exp1_04.zip
QExactive_positive_Exp1_05.zip

The following study identifiers are used throughout the different categories:
- Orbitrap_velos_CID
- Orbitrap_velos_HCD
- QTOF_SYNAPTG1
- 4000QTRAP (for Control Experiment 2 use: ABSciex_QTRAP4000)
- Orbitrap_elite
- QTRAP6500
- QTOF_G6550A
- QExactive

Depending on the experiment and platform, data download can be further confined by polarity and collision energies.

For the LipidBlast benchmarking (MTBLS397), data of two platforms were used (Obitrap Velos Pro CID, 4000 QTRAP). The biological data was benchmarked on both platforms, and experiment 1 on the Orbitrap, only. Thus, data containing QTRAP in the name belongs to the 4000 QTRAP, the rest to the Orbitrap:
- Biological_positive
- Biological_negative
- Biological_QTRAP_positive
- Biological_QTRAP_negative
- Experiment1_positive
- Experiment1_negative

Finally, click on the "Download selected files" button which is located at the bottom of the page. Attention: some studies might contain several GB.

MetaboLights will create now one zipped archive containing the individual data files stored in zip files. Please note, depending on the size of data files, it might take a while to generate the zip file before the download starts. The zipped data files contain typically data in its vendor format, a 'chrom' directory (data in LDA format), and an Excel file containing the results produced by LDA. Data of the 4000QTRAP and the QTRAP6500 do not contain data in vendor format; the file can be found in the corresponding "Comments" or "Evaluation" file, since several MS runs from AB Sciex instruments are packed in one file. "Comments" and "Evaluation" files contain comments and overall analysis results.

In order to have a complete data set of an experiment (all MS runs of one quintuplicate including comments or evaluation results), all files located in the individual ZIP archives (downloaded by selecting one platform as described above) have to be put in one directory. Further details about the individual files are given in chapters 2 and 4, and their visualization in chapter 3.
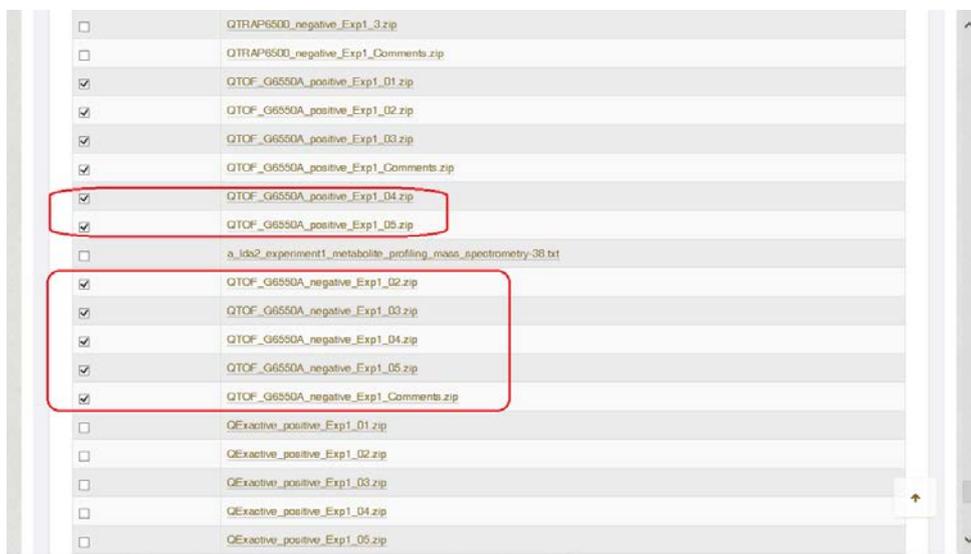
2 Data description of benchmarking (MTBLS397)

The final evaluation results are stored in the following Excel files:

    LDA-LipidBlast_positive.xlsx and LDA-LipidBlast_negative.xlsx (biological)

    LDA-LipidBlast_40000QTRAP_positive.xlsx and LDA-LipidBlast_40000QTRAP_negative.xlsx (biological 4000 QTRAP)

    LDA-LipidBlast_positive_standards.xlsx and LDA-LipidBlast_negative_standards.xlsx (control 1).

For MetaboLights, such a file can be found e.g. in the Biological_positive_Evaluation_Summary.zip.

These Excel files consist of a 'Summary' tab containing the final results, and five more tabs for the evaluation results of the individual MS runs. Detected lipid species and lipid molecular species are counted in two ways: adduct sensitive, where every species/adduct combination is one hit; adduct insensitive, where every detected species is one hit. E.g. PC 18:0/20:4 is found in protonated and sodiated form; in the first evaluation scheme, these are counted as two hits, in the second one, this corresponds to one hit only. Tables in the manuscript are based on the adduct insensitive numbers. For the biological experiment, data is evaluated for sensitivity (percent of correctly identified species relative to the total number of manually identified species) and positive predictive value (percent of correctly identified species relative to the total number of manually identified species) on the lipid species and lipid molecular species level. For control experiment 1, the sensitivity is evaluated in four aspects: 1) lipid species detection rate; 2) detection rate of correct chain combinations; 3) detection rate of correct chain combinations without reporting additional wrong chain combinations; 4) is the same as 3) including a correct position assignment. Data from control experiment 1 clearly shows that LipidBlast is not able to correctly assign *sn*-positions.

The folders containing the particular final result file contain 10 more Excel files (at MetaboLights, these files are zipped in one file which is e.g. called Biological_positive_Evaluation_Details.zip), where the names end by "LB10_comp.xlsx" or "LB450_comp.xlsx". These files contain details about the individual comparisons of LDA and LipidBlast, where 10 and 450 correspond to LipidBlast using a matching factor of 10 and 450, respectively. The "Comment" columns contain detailed reasons for counting a hit as false positive, or why LDA missed a hit.

Furthermore, the folders contain "mgf.tsv" (in the output folder of the home page download) and "mgf" files which are the original LipidBlast output files and the mgf files used for the searches, respectively. When downloaded from MetaboLights, these files can be found e.g. in Biological_positive_LipidBlast.zip.

For the sake of completeness, the folders contain the original Thermo RAW files. For MetaboLights download, the LDA's chrom and Excel results are provided too.

3 Data review and visualization

This chapter covers data description and visualization of all experiments, except for the benchmarking study which is described in chapter 2.

A complete data set (one polarity, one collision energy) consists of five MS runs (fifteen for biological data) and some txt or Excel files containing comments and/or evaluation results. One MS run comprises typically three files, i.e., data in its vendor format, a 'chrom' directory (data in LDA format), and an Excel file containing the results produced by LDA. This chapter contains two sections for the visualization of results, i.e. by LDA (recommended) and by MetaboLights.

3.1 Visualization by LDA (recommended)

In order to install the LDA, use the provided installation wizards. During the installation process, it does not matter which MS platform you choose, since this can be changed easily in the application itself. Usually the default settings in the wizard should do.

In order to visualize data of a single MS run, start the LDA, click on "Settings" tab, select the corresponding MS platform and the corresponding collision energies, and click on "Save as default":
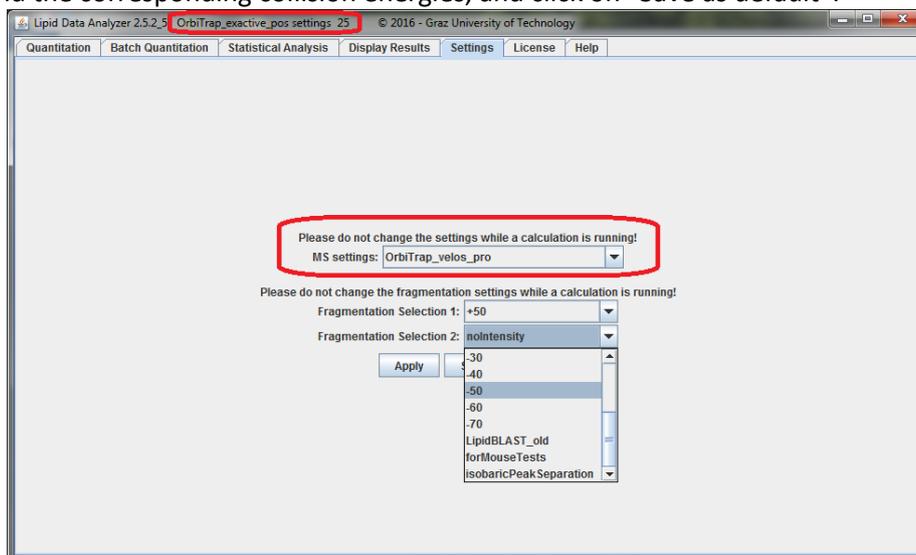
For study data, use the following settings:

| Platform	MS settings	Frag. Selection 1	Frag. Selection 2
Orbitrap_velos_CID	Orbitra_velos_pro	+50	-50
Orbitrap_velos_HCD	Orbitra_velos_pro_HCD	+25	-50
QTOF_SYNAPTG1	QTOF_centr	+30V	-30V
4000QTRAP	QTRAP_4000	+45V	-45V
Orbitrap_elite	OrbiTrap_Elite	+25	-50
QTRAP6500	QTRAP_6500	+45V	-45V
QTOF_G6550A	QTOF_Agilent_G6550	+30V	-30V
QExactive	OrbiTrap_exactive_pos/	25	
	OrbiTrap_exactive_neg	-30	

For control experiment 1, depending on the data set, different collision energy settings might apply for the platforms Orbitrap_velos_CID, Orbitrap_velos_HCD, QTOF_SYNAPTG1, and 4000QTRAP.

Then, click on the "Display Results" tab, select the corresponding chrom and LDA Excel result file, and click on "Start Display":
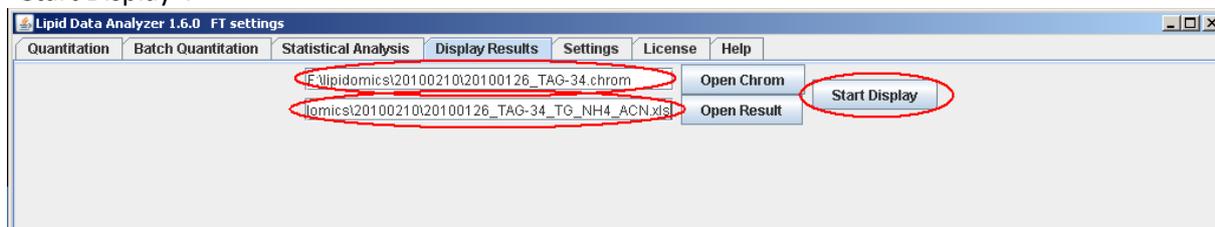
A menu on the left side will appear now containing a select box at the top, where you can switch between the different lipid subclasses. Then a table follows, whereupon the name of the analyte is in the first column, and the integrated mass-chromatographic area in the second. The display name consists primarily of "$name$:$double_bonds$_$retention_time$" (e.g. 38:5_24.98 lipid species name). If there are more than one adduct/modification, the modification name is added with _$modification_name$ (e.g. 38:5_24.98_H).
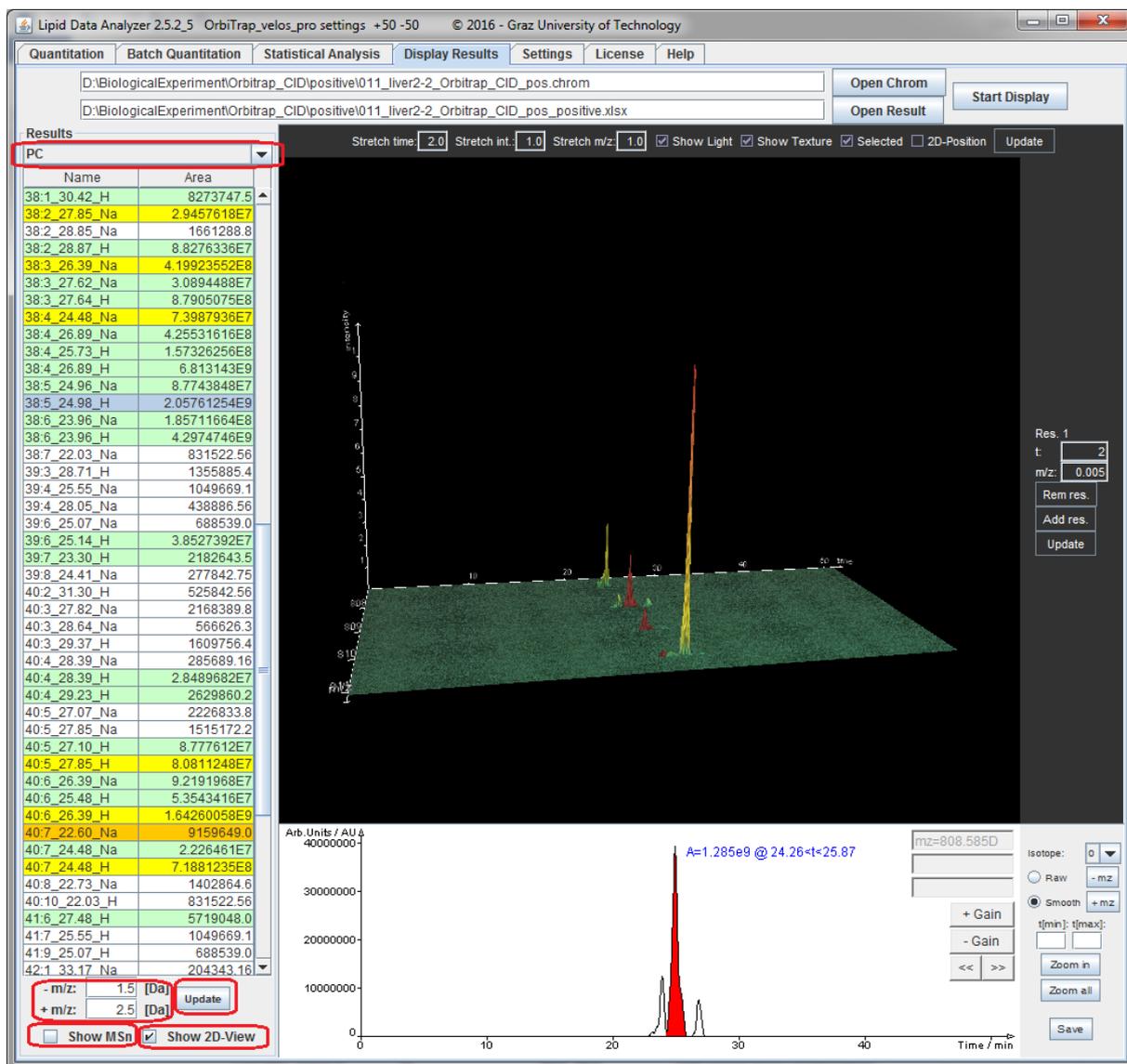
The color encoding of the table entries has the following meaning:

- White: MS^1 identification without any MS^n evidence (not relevant for the publication, and not exported to MetaboLights).
- Green: MS^1 identification verified by MS^n spectra.
- Yellow: The MS^n algorithm detected more than one lipid subclass in the MS^1 peak → the peak was split on the MS^1 level in a part for the one and a part for the other lipid subclass. The split is based on information derived from MS^n spectra.
- Orange: The MS^n algorithm detected more than one lipid subclass in this peak, but a split on the MS^1 level is not possible. The intensities of the intact MS^1 peak are split according to the distinct fragments of MS^n spectra. This way of splitting is rather unreliable, since usually fragment intensities do not correlate with MS^1 abundances.
- Blue: Current selection for display.

The "Show MSn" checkbox switches from the lipid species to the lipid molecular species view of the table where possible:

| PC	
Name	Area
20:1_18:0_30.42_H	8273747.5
38:2_27.85_Na	2.9457618E7
38:2_28.85_Na	1661288.8
18:1/20:1_28.87_H	8472244.0
18:0/20:2_28.87_H	7.9804088E7
18:0/20:3_26.39_Na	4.19923552E8
18:0/20:3_27.62_Na	3.0894488E7
21:1_17:2_27.64_H	9.5518848E7
18:0/20:3_27.64_H	7.835319E8
38:4_24.48_Na	7.3987936E7
18:0/20:4_26.89_Na	4.25531616E8
22:4_16:0_25.73_H	1.3572973E7
20:4_18:0_25.73_H	3.7298596E7
18:1/20:3_25.73_H	1.06454688E8
18:0/20:4_26.89_H	6.813143E9
18:1/20:4_24.96_Na	8.7743848E7
16:0/22:5_24.98_H	5.31407136E8
20:0/18:5_24.98_H	4.6825152E8
18:1/20:4_24.98_H	1.05795392E9
16:0/22:6_23.96_Na	1.85711664E8
16:0/22:6_23.96_H	4.2974746E9
38:7_22.03_Na	831522.56
39:3_28.71_H	1355885.4

The table displays now information about the obtained structure according to Liebisch, G. et al. Shorthand notation for lipid structures derived from mass spectrometry. J Lipid Res 54, 1523-1530 (2013). I.e., the sum of carbon atoms and double bonds in the chains if no structural information can be obtained (e.g. 38:5), the fatty acid chains separated by an underscore if fatty acids are detectable (e.g. 20:4_18:0), or the fatty acid chains separated by a slash if the position of the fatty acid chains is assignable (e.g. 18:1/20:4). Attention: For the manual inspection of MS^n, there is a difference in the displayed annotation of the spectra, if the lipid species consists of more than one lipid molecular species:

- The lipid species view is selected: The fragments of all detected lipid molecular species are annotated.
- The lipid molecular species view is selected: Only the fragments of the particularly selected lipid molecular species are annotated.

When a row of the table is activated by the left mouse button, the peaks are displayed in the 3D and the chromatogram viewer. When a row is activated by the right mouse button, a popup appears:

38:3_27.64_H	8.7905075E8
38:4_24.48_Na	7.3987936E7
38:4_26.89_Na	4.25531616E8
38:4_25.73_H	1.57326256E8
38:4_26.89_H	6.813143E9
38:5_24.96_Na	8.7743848E7
38:5_24.98_H	2.05761254E9
38:6_23.96_Na	Add analyte before
38:6_23.96_H	Add analyte after
38:7_22.03_Na	
39:3_28.71_H	Delete analyte
39:4_25.55_Na	Show MS/MS
39:4_28.05_Na	
39:6_25.07_Na	Sort by mass
39:6_25.14_H	Sort by intensity
39:7_23.30_H	Edit MSn rule
39:8_24.41_Na	277842.75
40:2_31.30_H	525842.56

By clicking "Show MS/MS", MS/MS spectra are shown as annotated by the used MS^n decision rules at the time of quantitation (see figure on next page). "Edit MSn rule" will start the interface for decision rule set definition/adaption/generation.
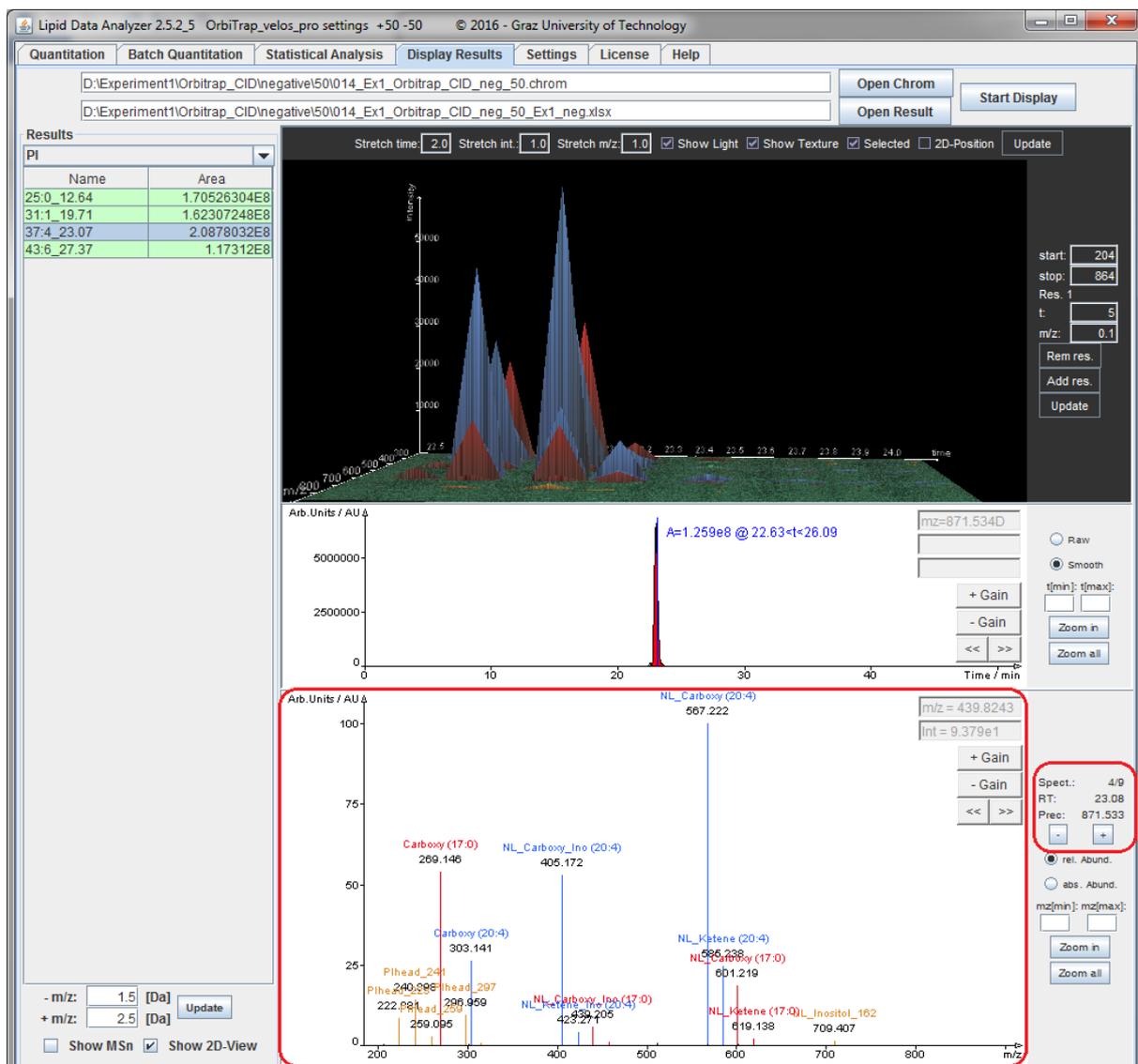
The big red box shows an annotated spectrum, and the smal red box shows the buttons +/- for scrolling through spectra. Further details regarding LDA visualization options are to be found in the user manual which is accessible in the "Help" tab of the LDA, or can be downloaded from the institute's homepage: http://genome.tugraz.at/lda2/lda_documentation.shtml.

3.2 Data reviewing in MetaboLights

MetaboLights does not allow visualization of spectra, but metabolite identifications are stored in tabular format in so-called MAF (metabolite annotation/assignment file) within assays. However, you can always download the MAF directly by selecting the m_…txt from "Study files" download, and view it using Excel. An assay consists of a name describing the respective MS platform, a table comprising information about the identified metabolites, and a table containing information about the data files and their generation.

Study Design Description   Protocols   Samples   ☑ Assay ▾   Study Files   Study Validation ✗   Pathways

Assay 1

Assay file name: 📄 a_lda2_experiment3_metabolite_profiling_mass_spectrometry-15.txt

Technology: **mass spectrometry**↓↓
Platform: **Orbitrap velos HCD positive 1to1 (Thermo Scientific)**   **Platform description**

Identfied metabolites

Metabolites

Show 10 ▾ entries   Search: [   ]

Metabolite identification	Chemical formula	Species	Smiles	InChl	Chemical Shift	Multiplicity	m/z	Fragmentation
PC 14:0/18:0)-|PG 14:0/18:0|PC 14:0/18:0|PC 14:0/18:0|PC 32:0,PC 14:0/18:0)	H80P1O8N1C40						734.5705	104.10,125.00,184.06,450.30,468.31,506.36,524.37,551.50,734.57|-[104.10
PC 14:0/18:0)|PC 14:0/18:0|PC 14:0/18:0|PC 14:0/18:0|PC 14:0/18:0|PC 32:0,PC 14:0/18:0)	H80P1O8N1C40						756.5502	146.97,413.20,450.30,469.26,506.36,528.34,551.50,573.49,697.48,756.54|
PC 16:0/16:0)|PC 16:0/16:0|PC 16:0/16:0|PC 16:0/16:0|PC 16:0/16:0|PC 32:0,PC 16:0/16:0)	H80P1O8N1C40						756.5502	146.97,441.24,478.32,500.31,551.50,573.49,697.48,756.54|146.97,441.24,
PC 16:0/16:0|PC	H80P1O8N1C40						734.5705	104.10,125.00,184.06,478.32,496.33,551.50,734.57|104.10,125.00,184.0

∎
∎
∎

18:0/18:2)|PC 18:0/18:2|PC 18:0/18:2|PC 18:0/18:2|PC 36:2,PG 18:0/18:2)								

Showing 1 to 10 of 26 entries   Previous **1** 2 3 Next

Data

Show 10 ▾ entries   Search: [   ]

Sample Name	Protocol REF	Post Extraction	Derivatization	Extract Name	Protocol REF	Chromatography Instrument	Column model	Column type	Labeled Extract Name	Label	Protocol REF
Standard mix3_1	Extraction				Chromatography	Dionex Ultimate XRS UHPLC	BEH C8 column (100 x 1 mm, 1.7 µm)	CHMO:reversed-phase liquid chromatography			Mass spectrometr
Standard mix3_1	Extraction				Chromatography	Dionex Ultimate XRS UHPLC	BEH C8 column (100 x 1 mm, 1.7 µm)	CHMO:reversed-phase liquid chromatography			Mass spectrometr
Standard mix3_1	Extraction				Chromatography	Dionex Ultimate XRS UHPLC	BEH C8 column (100 x 1 mm, 1.7 µm)	CHMO:reversed-phase liquid chromatography			Mass spectrometr
Standard mix3_1	Extraction				Chromatography	Dionex Ultimate XRS UHPLC	BEH C8 column (100 x 1 mm, 1.7 µm)	CHMO:reversed-phase liquid chromatography			Mass spectrometr
Standard mix3_1	Extraction				Chromatography	Dionex Ultimate XRS UHPLC	BEH C8 column (100 x 1 mm, 1.7 µm)	CHMO:reversed-phase liquid chromatography			Mass spectrometr
Standard mix3_1	Extraction				Chromatography	Dionex Ultimate XRS UHPLC	BEH C8 column (100 x 1 mm, 1.7 µm)	CHMO:reversed-phase liquid chromatography			Mass spectrometr

Showing 1 to 6 of 6 entries   Previous **1** Next

Information about the data files and their generation

The data columns contain a column called "Raw Spectral Data File" which indicates which files belong to one particular data set.

Scan m/z range	Instrument	Ion source	Mass analyzer	MS Assay Name	Raw Spectral Data File	Protocol REF	Normalization Name	Derived Spectral Data File	Protocol REF
400-1000	LTQ Orbitrap Velos	MS:electrospray ionization	MS:orbitrap	Orbitrap velos HCD positive 1to1 (Thermo Scientific)	Orbitrap_velos_HCD_pos_Exp3_1to1_034.zip	Data transformation			Metabolite identification
400-1000	LTQ Orbitrap Velos	MS:electrospray ionization	MS:orbitrap	Orbitrap velos HCD positive 1to1 (Thermo Scientific)	Orbitrap_velos_HCD_pos_Exp3_1to1_035.zip	Data transformation			Metabolite identification
400-1000	LTQ Orbitrap Velos	MS:electrospray ionization	MS:orbitrap	Orbitrap velos HCD positive 1to1 (Thermo Scientific)	Orbitrap_velos_HCD_pos_Exp3_1to1_036.zip	Data transformation			Metabolite identification
400-1000	LTQ Orbitrap Velos	MS:electrospray ionization	MS:orbitrap	Orbitrap velos HCD positive 1to1 (Thermo Scientific)	Orbitrap_velos_HCD_pos_Exp3_1to1_037.zip	Data transformation			Metabolite identification
400-1000	LTQ Orbitrap Velos	MS:electrospray ionization	MS:orbitrap	Orbitrap velos HCD positive 1to1 (Thermo Scientific)	Orbitrap_velos_HCD_pos_Exp3_1to1_038.zip	Data transformation			Metabolite identification
400-1000	LTQ Orbitrap Velos	MS:electrospray ionization	MS:orbitrap	Orbitrap velos HCD positive 1to1 (Thermo Scientific)	Orbitrap_velos_HCD_pos_Exp3_Evaluation.zip	Data transformation			Metabolite identification

Files belonging to one data set start typically by the same file name to ease selective download of the files (see chapter 1.2). In this case, "Orbitrap_velos_HCD_pos_Exp3_1to1" has to be entered in the text box as described in chapter 1.2 for selection of the five data files. Additionally, the file called Orbitrap_velos_HCD_pos_Exp3_Evaluation.zip has to be selected manually. This file is not called Orbitrap_velos_HCD_pos_Exp3_1to1_Evaluation.zip, since it is an evaluation file for all five ratios.

The "Metabolites" table shows the identified species in a collapsed form where each row corresponds to one lipid (molecular) species identified by one adduct in any of the associated 5 MS runs (or 15, in the case of the biological data).

The identifier for one species is shown in the "Metabolite identification" column in brackets at the end of the entry. The identifier starts with the lipid species (e.g. "PC 32:0"), and if present, followed by the lipid molecular species (e.g. "PC 14:0/18:0"), separated by a comma. Since data presentation is up to now not completely standardized in MetaboLights, and it was important for this study to reflect the actual identifications in each MS run, a pipe "|" was used to discriminate between the individual MS runs, while comma "," was used to account for an enumeration of identifications (e.g. several retention times) and

ambiguous identifications (e.g. DG 20:0/18:0/-,20:0/-/18:0 corresponds to a diacylglyceride where 20:0 is identified to be on the sn-1 position, while 18:0 might be at sn-2 or sn-3). For the first row in this example, PC 14:0/18:0 was found in all MS runs except for the second one where the missing detection is indicated by a hyphen "-".

For the column "Fragmentation", *m/z* values indicative for these identifications are presented as comma separated values.

The review of the control experiments by the MetaboLights interface might work quite well, however, for the biological data, we strongly recommend the LDA approach (see chapter 3.1). The reason is that the **"Metabolites" column does not allow for sorting by the species identifier**. Sorting by the column "Metabolite Identification" always uses the whole content of the cell. If the first MS run contains a missing detection (a hyphen), the hyphen is regarded as alphabetically first, and will be at the start of the table. Since **data is provided as derived by LDA including all false positives**, **by this mechanism, rare species and false positives are shown first**:

Nevertheless, class specific sorting is partially possible by entering the lipid subclass name in the "Search" text box which will filter the entries for the contained text. But e.g. entering of "PC" will always include LPC and P-PC too.

4 Available data

This chapter covers all datasets except for the benchmarking (MTBLS397) study which is explained in chapter 2. Data sets or assays consist of 5 or 15 MS runs (stored in ZIP files in MetaboLights), and 1 or more txt or Excel files containing results of detailed analysis, comments about information content that can be derived from the various adduct ions, and suggestions about optimal collision energies for subclasses/adducts (stored in one or more ZIP files in MetaboLights). Each MS run consists of a 'chrom' directory (raw data in LDA format), an Excel file containing the results produced by LDA, and the data in its vendor format, except for QTRAP data, where only one wiff is available for several files. The following chapters explain in detail the amount of additional information provided for each study.

4.1 Biological data (MTBLS396)

Each data set in biological data consists of 15 MS runs, and several Excel files containing manual evaluation results for both positive and negative ion mode. In the case of MetaboLights, these Excel files are packed in a file ending with "Evaluation.zip". These files contain at least one "SpeciesDetectable …" Excel file, reporting the identified species for mouse liver sample one, and one file containing "FP" in the name, reporting the identified false positives on the MS2 level (there is no evaluation of the MS1 level, since this study covers the fragmentation spectra only). For some platforms, an Excel document starting with "FNs …" is available, reporting missed identifications, and several Excel documents containing the lipid subclass and the "species" in the file name, reporting details which lipid species and lipid molecular species were found in the individual MS runs. For 4000 QTRAP data, the wiff file is provided in the

4000QTRAP_Liver_Evaluation.zip, and for QTRAP 6500 data, the wiff fie can be found in the Exp1_QTRAP6500.zip of control experiment 1 (MTBLS394 – contains raw data for both the control experiment 1 and the biological experiment).

4.2 Control experiment 1 (MTBLS394)

Each data set consists of 5 MS runs and several text files containing details about the detected species, missed species and false positives. In MetaboLights, these files are packed in one ZIP file whose name ends by Exp1_Comments.zip. For the platforms where the data was collected at more than one collision energy ('Orbitrap Velos Pro', 'Orbitrap Velos Pro HCD', 'SYNAPT G1 HDMS QTOF' and '4000 QTRAP'), additional comments are available providing general information about the individual subclasses and their optimal collision energies. In the download from the homepage, these files are located in the parent directory of the data set, while at MetaboLights, an additional Exp1_Comments.zip file is provided for these platforms. At MetaboLights, the used mass lists for the particular experiment are added to the "Exp1_Comments.zip" archives. For QTRAP data, the wiff files are provided in the Exp1_Comments.zip archives.

4.3 Control experiment 2 (MTBLS391)

This study was performed in positive ion mode only. Each data set consists of 5 MS runs and one text file containing information about unsuccessful splits. At MetaboLights, this file is provided in a file ending by "Exp2_Comments.zip" together with the actually used mass list, and the wiff file in the case of QTRAP data.

4.4 Control experiment 3 (MTBLS398)

Each data set (ratio) consists of 5 MS runs, one text file containing comments, and one Excel file containing evaluation of the performance of all five ratios for each platform and polarity. At MetaboLights, this file is provided in a file ending by "Exp3_Evaluation.zip" together with the actually used mass list, and the wiff file in the case of QTRAP data.

4.5 HCD characterization and regio-isomers (MTBLS462)

The HCD characterization data consists of 4 MS runs for the positive ion mode (longer and shorter inclusion list), and two for negative ion mode. Furthermore, two additional directories are provided containing the spectra of the novel species in SVG and MGF format. At MetaboLights, these files are provided in Novel_species_SVGs.zip and Novel_species_MGFs.zip.

The data of the regio-isomers consists of 3 MS runs for each sample, i.e. a mix of both regio-isomeric standards, and the individual standards. The results of the evaluation are stored in the Evaluation.xlsx. At MetaboLights, this file can be found in the PositionIsomers_Evaluation.zip.