

Step-by-step LION/web enrichment analysis tutorial

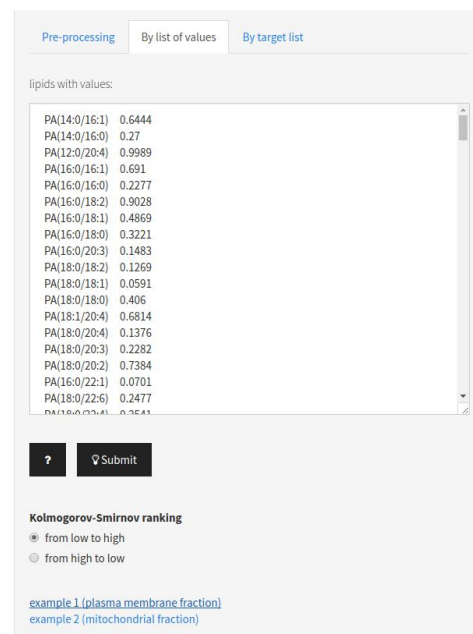
Suppose researchers measured the lipidome of cells in different conditions. They want to assess differences between condition A and B. LION/web provides an environment that helps researchers to do this in an unbiased and systematic way. In this tutorial, we use a published dataset that includes lipidomes of several organelles isolated from macrophages (Andreyev *et al.* 2010).

LION (Lipid ONtology) is an ontology database that contains information ('terms') about classification, lipid function, sub-cellular localization and biophysical properties of lipids. LION/web (performs LION enrichment analysis) calculates enrichment of LION-terms of provided datasets.

To perform LION enrichment analysis, follow these steps:

- 1) Calculate association-values for all lipids. This can be, for instance, fold-increase values of condition A vs. condition B, or one-tailed t-test p-values. Alternatively, LION/web has a 'pre-processing' option; here a csv-file is be uploaded and processed. In this example, this is already done for you. You can find how to enter this dataset at step 6.
- 2) Go to the website of LION/web; <http://www.lipidontology.com>.
- 3) Lipidomics data can be processes in two ways. There is an input-field on the left side of the page. On top of this, there are three tabs, supporting two different enrichment methods;
 - 'by list of values': lipids are ranked by the provided values and terms that are associated with highly ranked lipids are reported.
 - 'by target list': a hitlist of lipids that satisfy certain criteria (*i.e.* p-value threshold, part of a cluster of interest) is compared with the full lipidome; terms that are enriched in the hitlist are reported.
- 4) In this example, we choose 'by list of values', as this circumvents the arbitrary use of a threshold.
- 5) Now, lipids and values can be entered in the input field. Lipids and values have to be separated by a *commas* or *tabs*. Simply copy-paste data from a spreadsheet is usually the most convenient way to enter data. Lipids should be formatted as:

CLASS (#C:#DB/#C:#DB/...)



The screenshot shows the LION/web interface. At the top, there are three tabs: 'Pre-processing' (selected), 'By list of values', and 'By target list'. Below the tabs, there is a text input field labeled 'lipids with values:'. The input field contains a list of lipids and their corresponding values, separated by tabs. The list is as follows:

Lipid	Value
PA(14:0/16:1)	0.6444
PA(14:0/16:0)	0.27
PA(12:0/20:4)	0.9989
PA(16:0/16:1)	0.691
PA(16:0/16:0)	0.2277
PA(16:0/18:2)	0.9028
PA(16:0/18:1)	0.4869
PA(16:0/18:0)	0.3221
PA(16:0/20:3)	0.1483
PA(18:0/18:2)	0.1269
PA(18:0/18:1)	0.0591
PA(18:0/18:0)	0.406
PA(18:1/20:4)	0.6814
PA(18:0/20:4)	0.1376
PA(18:0/20:3)	0.2282
PA(18:0/20:2)	0.7384
PA(16:0/22:1)	0.0701
PA(18:0/22:6)	0.2477
PA(18:0/22:4)	0.3541

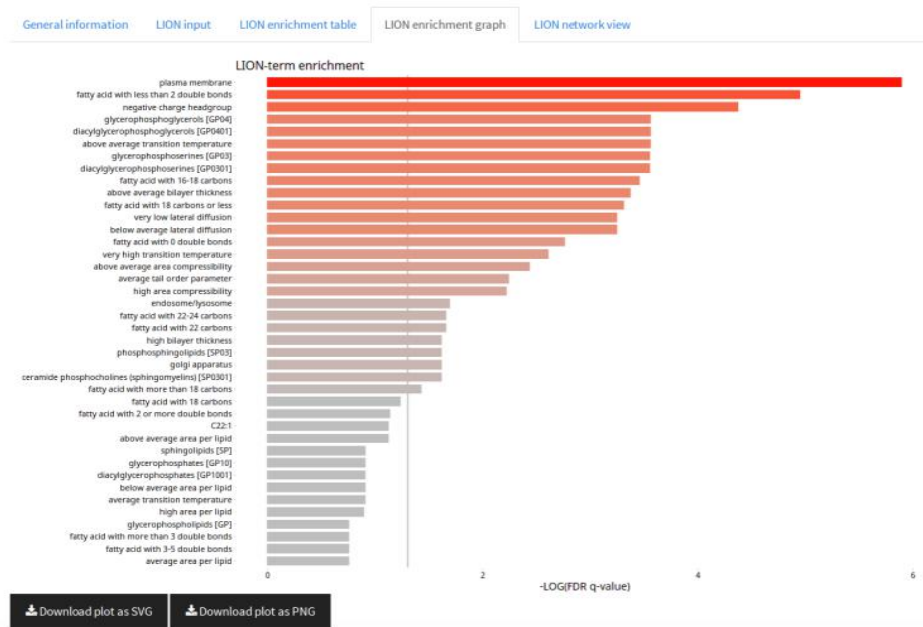
Below the input field, there is a 'Submit' button. At the bottom, there is a section for 'Kolmogorov-Smirnov ranking' with two radio buttons: 'from low to high' (selected) and 'from high to low'. There are also two links: 'example 1 (plasma membrane fraction)' and 'example 2 (mitochondrial fraction)'.

Where *CLASS* is a prefix indicating the lipid class (i.e. PC, PE, SM), *#C* is the number of fatty acid carbons and *#DB* the number of fatty acid unsaturations. For convenience, some other notation are also supported (shorthand notations, LIPIDMAPS- and SwissLipids-identifiers).

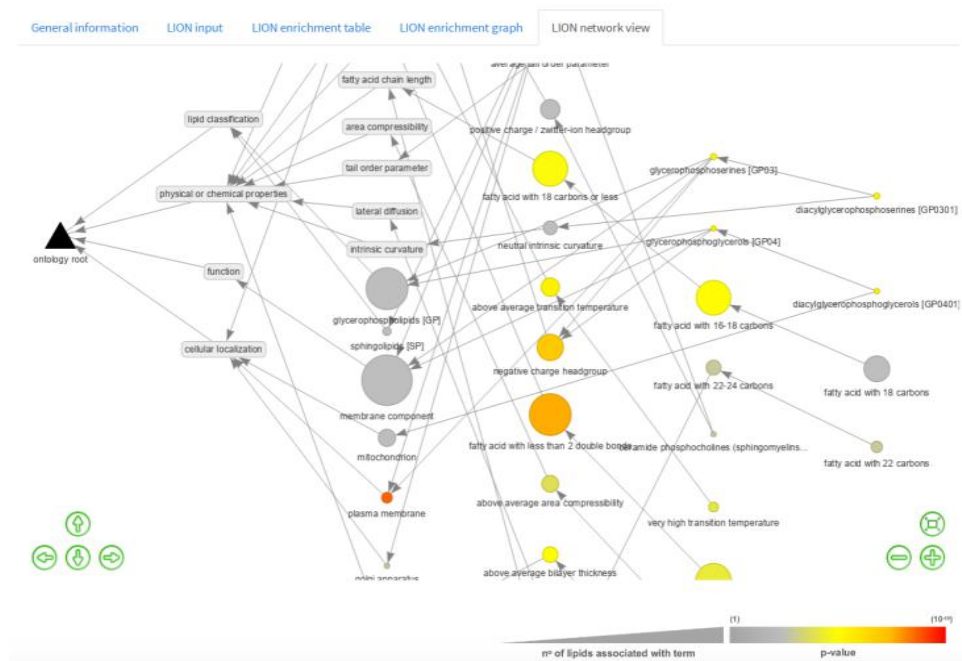
- 6) Click on 'example 1: plasma membrane fraction' on the bottom of the field to enter the example set.
- 7) Depending on the data type, choose the ranking direction. In this case, the values are p-values. P-values should be ranked 'from low to high'. Click on 'submit'.
- 8) Now, lipids are matched with the LION-database. Progression of this process is monitored at the bottom of the page.
- 9) On the right side of the page, under the tab 'LION-input', a table will appear that provides information about the mapping of lipids to LION. In addition, the percentage of lipids that could be mapped is given. Check if the input is correctly mapped to LION. Small typos might affect appropriate mapping. The table can be downloaded at the bottom of the page.
- 10) The tab 'LION enrichment table' shows information about enrichment of LION-terms. The table consists of information about terms (LION ID, description) and statistics (number of annotated lipids in the term, p-value and FDR-corrected p-value). This table, together with a more detailed table that includes lipids associated with the reported terms, can be downloaded at the bottom of the page.

General information		LION input	LION enrichment table	LION enrichment graph	LION network view
Term ID	Discription	Annotated	p-value	FDR q-value	
LION:0012082	plasma membrane	41	2.6e-09	1.28e-06	
LION:0002966	fatty acid with less than 2 double bonds	160	4.6e-08	1.13e-05	
LION:0000093	negative charge headgroup	101	2.6e-07	4.27e-05	
LION:0000014	glycerophosphoglycerols [GP04]	21	3.1e-06	2.79e-04	
LION:0000059	diacylglycerophosphoglycerols [GP0401]	21	3.1e-06	2.79e-04	
LION:0001740	above average transition temperature	71	3.4e-06	2.79e-04	
LION:0000013	glycerophosphoserines [GP03]	24	4.6e-06	2.83e-04	
LION:0000053	diacylglycerophosphoserines [GP0301]	24	4.6e-06	2.83e-04	
LION:0002948	fatty acid with 16-18 carbons	133	6.4e-06	3.51e-04	
LION:0080974	above average bilayer thickness	60	8.6e-06	4.24e-04	
LION:0000100	fatty acid with 18 carbons or less	134	1.1e-05	4.93e-04	
LION:0080976	very low lateral diffusion	34	1.4e-05	5.69e-04	
LION:0080981	below average lateral diffusion	63	1.5e-05	5.69e-04	
LION:0002968	fatty acid with 0 double bonds	136	4.9e-05	1.73e-03	
LION:0001739	very high transition temperature	38	7.5e-05	2.46e-03	
LION:0080958	above average area compressibility	64	0.00012	3.70e-03	
LION:0080986	average tail order parameter	27	0.00020	5.80e-03	
LION:0080955	high area compressibility	34	0.00022	6.03e-03	
LION:0012086	endosome/lysosome	17	0.00079	2.05e-02	

- 11) The tab 'LION enrichment graph' shows a barplot of the $-\text{LOG}(\text{FDR q-value})$ of the 40 most enriched terms. This plot can be downloaded as SVG- or PNG-file at the bottom of the page.



- 12) The tab 'LION network view' provides an interactive network of the enriched LION-terms. All terms and relations on the path between the ontology root (left) and significant terms ($\text{FDR q-value} < 0.05$) are shown. Colors are scaled by the raw p-value of the terms, while the node sizes are determined by the number of lipids associated with the given terms. The network can be browsed by the navigation buttons at the bottom of the page.



Interpretation of LION enrichment analysis

LION/web provides information about LION-term enrichment by making use of p-values. These values describe the probability that terms are not enriched. Hence, terms with low p-values can be considered as enriched. However, choosing a classical p-value cut-off of 0.05 to select significant hits would result in a number of false positives. To circumvent this, LION/web corrects p-values for multiple testing (by the false discovery rate method, FDR), resulting in FDR q-values. Technically, 100 randomized inputs with a q-value cut-off of 0.05 would result in 5 false positives. However, LION is constructed as a hierarchical structure and many terms are not completely independent, resulting in an over-correction by the FDR-method. Hence, also terms with q-values higher than the cut-off should be included in downstream interpretations.

In this example, we analyzed the plasma membrane (PM) fraction of macrophages (Andreyev *et al.* 2010). Analysis resulted in hits that include terms associated with sub-cellular localization ('plasma membrane'), low membrane fluidity (i.e., 'above average transition temperature', 'very low lateral diffusion', 'high area compressibility'), bilayer thickness ('above average bilayer thickness', 'high bilayer thickness'), lipid saturation ('fatty acid with less than 2 double bonds'), headgroup charge ('negative charge headgroup') and lipid classification (i.e., 'glycerophosphoserines', 'phosphosphingolipids'). These findings are in line with current knowledge of PM-properties.

References:

Andreyev AY, Fahy E, et al. 2010. Subcellular organelle lipidomics in TLR-4-activated macrophages. *J Lipid Res.* 2010 Sep;51(9):2785-97.

