LipidSuite: interactive web server for lipidomics differential and enrichment analysis

Ahmed Mohamed [1,*] **and Michelle M. Hill** [1,2,*]

[1]Precision & Systems Biomedicine Laboratory, QIMR Berghofer Medical Research Institute, Herston, QLD 4006, Australia and [2]Centre for Clinical Research, Faculty of Medicine, The University of Queensland, Herston, QLD 4006, Australia

ABSTRACT

Advances in mass spectrometry enabled high throughput profiling of lipids but differential analysis and biological interpretation of lipidomics datasets remains challenging. To overcome this barrier, we present LipidSuite, an end-to-end differential lipidomics data analysis server. LipidSuite offers a step-by-step workflow for preprocessing, exploration, differential analysis and enrichment analysis of untargeted and targeted lipidomics. Three lipidomics data formats are accepted for upload: mwTab file from Metabolomics Workbench, Skyline CSV Export, and a numerical matrix. Experimental variables to be used in analysis are uploaded in a separate file. Conventional lipid names are automatically parsed to enable lipid class and chain length analyses. Users can interactively explore data, choose subsets based on sample types or lipid classes or characteristics, and conduct univariate, multivariate and unsupervised analyses. For complex experimental designs and clinical cohorts, LipidSuite offers confounding variables adjustment. Finally, data tables and plots can be both interactively viewed or downloaded for publication or reports. Overall, we anticipate this free, user-friendly webserver to facilitate differential lipidomics data analysis and reanalysis, and fully harness biological interpretation from lipidomics datasets. LipidSuite is freely available at http://suite.lipidr.org.

GRAPHICAL ABSTRACT
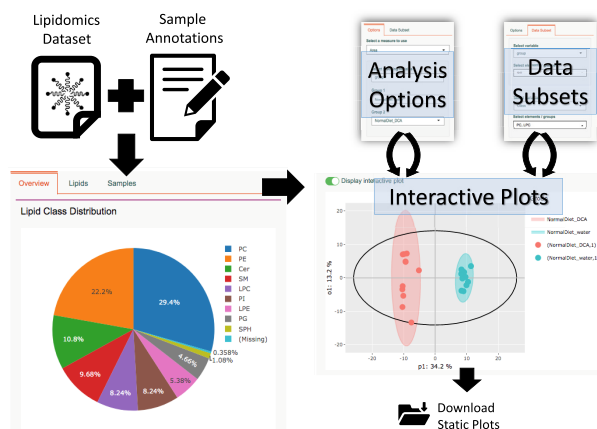
INTRODUCTION

Biological lipids play important roles in cell and organismal function, ranging from cellular membrane structure and dynamics, to intercellular communication in the form of hormones and lipid mediators, and finally, as energy storage molecules. These varied lipid functions are enabled through a diversity of lipid classes and structures. Dysregulation of specific structures have been implicated in pathogenesis of a range of conditions including metabolic diseases, neurodegeneration and cancers. Advances in mass spectrometry instrumentation and methods for lipid structure elucidation have enabled high throughput profiling of hundreds to thousands of lipid species in a single analysis. Mass spectrometry-based lipidomics has become a vital part of biomedical and life science research (1), complementing information from genomics, proteomics and metabolomics.

In a generic lipidomics workflow, lipids extracted from biological sources are subjected to mass spectrometry with or without liquid chromatography separation. Internal standards should be included to ensure similar extraction efficiency, mass spectrometry loading, and for absolute quanti-

*To whom correspondence should be addressed. Tel: +61 7 3845 3020; Fax: +61 7 3362 0102; Email: michelle.hill@qimrberghofer.edu.au
Correspondence may also be addressed to Ahmed Mohamed. Tel: +61 3 9345 2555; Fax: +61 3 9347 0852; Email: mohamed.a@wehi.edu.au
Present address: Ahmed Mohamed, Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3052, Australia.

Table 1. Comparison of features of web servers for lipidomics data analysis

	Untargeted lipidomics	Targeted lipidomics	Lipid name parsing	Quality control	Normalization	Two-group comparison	Multigroup comparison	Factorial analyis	Unsupervised	Orthogonal	Lipid set enrichment analysis	Lipid chain analysis	Interactive Data subsetting	Interactive plotting
LipidSuite	■	■	■	■	■	■	■	■	■	■	■	■	■	■
MetaboAnalyst	■	*		■	■	■	■		■	■				**
Workflow4Metabolomics	■	*		■	■	■	■	■	■	■				
MetaBox	■	*		■	■	■	■		■		■	■		■
Lipid Mini-On	■	*	■								■	■		■
LION/web	■	*	■								■	■		**
Goslin			■											
LipidLynxX			■											

Input | Pre-processing | Univariate analysis | Multivariate analysis | Lipid-focussed analysis | Interactivity

Grey filled box = feature available. Empty box = not available.
* Targeted datasets not supported as Skyline export format, but when formatted as numerical matrix
** Limited to some plots

tation in targeted workflows. The relative levels of observed lipid species are measured by the intensity or area under the peak. In untargeted workflows, lipid annotation is made by matching tandem mass spectra (MS/MS) to spectral libraries. Software tools developed for lipid annotation were recently reviewed by Züllig and Köfeler (2). Most of the software tools require a level of coding proficiency by the user, there are a few web servers that enable wider accessibility by researchers (Table 1). Metabolomics web servers such as MetaboAnalyst (3), Workflow4metabolomics (4) and MetaBox (5) offer comprehensive statistical analyses but do not offer lipid-specific features such as lipid name parsing, enrichment and chain length analyses. On the other hand, lipidomics analysis web servers goslin (6) and Lipid-LynxX (7) address lipid name parsing, while Lipid Mini-On (8) and LION/web (9) also offer ontology enrichment and lipid chain length analyses (Table 1). However, none offer the ability to integrate lipid class enrichment and chain length analyses with a statistical workflow including normalization, to provide lipid-specific biological interpretation of the differential analysis results.

Here, we present an integrated webserver for differential lipidomics which offers tools for the each step of the analysis workflow. Starting from annotated lipidomics data with the experimental groupings, LipidSuite offers options for all steps in lipidomics data analytics, from data quality control (QC), normalization, imputation to differential analysis. Uniquely, LipidSuite allows users to interactively process and visualize subsets of their data; users can choose or exclude certain samples or lipid molecules. The data analytics components of LipidSuite uses lipidr R-package (10), which includes univariate, multivariate statistics, lipid set enrichment and chain trend analysis. LipidSuite is freely accessible at http://suite.lipidr.org.

Overview of LipidSuite web server

LipidSuite is a user-friendly web application that allows advanced lipidomics analysis in a user-interactive fashion. The app main interface displays the analysis steps on an expandable menu on the left-hand side (Figure 1), with a guided help section on the right. The analysis steps include: Input, Data QC, Preprocessing, Data Exploration, Differential Analysis and Heatmap. The complete analysis workflow is demonstrated with two example datasets available on LipidSuite, as well as in a detailed tutorial video available on the landing page. We describe each step with the available options in detail below.

Input

Lipidsuite analysis begins with upload of two files, containing the lipidomics measurements and the experimental sample annotation, respectively (Figure 2A). LipidSuite accepts three lipidomics data formats, (i) a Skyline CSV Export, useful for targeted lipidomics, (ii) a numerical matrix where lipids are rows and samples are columns and iii) mwTab file from Metabolomics Workbench (11). Notably, LipidSuite can utilize additional measures exported by Skyline, where users can include retention time, peak height or background in addition to the default peak area. Instructions for each file type is displayed when the file type is selected, with example files available for download. Lipid names should be provided in a LIPIDMAPS convention or in 'Class XX:YY'
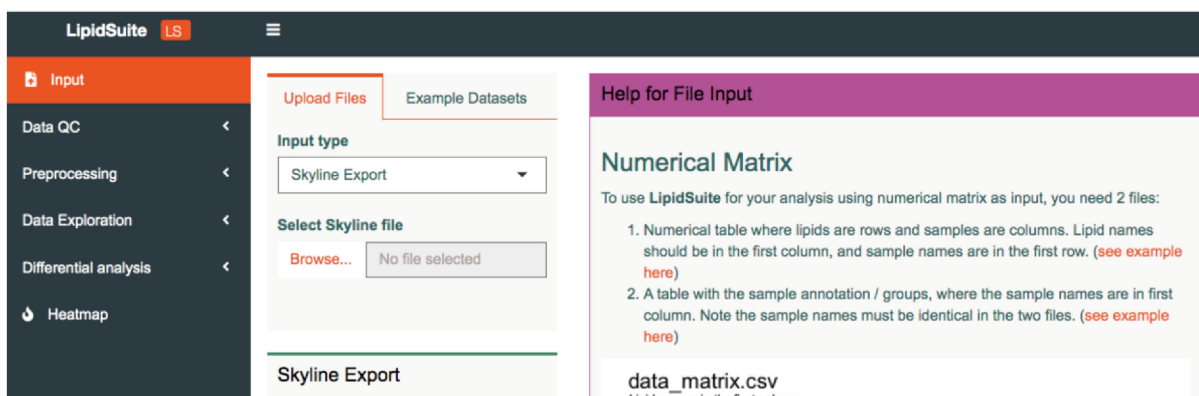
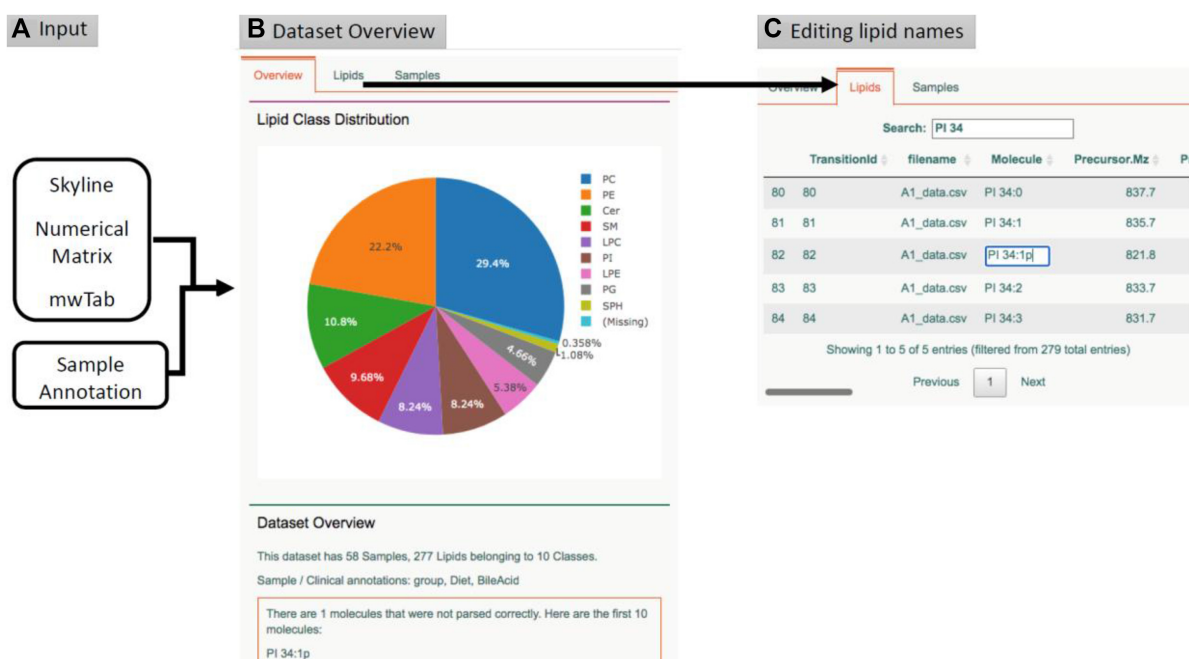**Figure 1.** LipidSuite web interface.

Figure 2. Dataset input process in LipidSuite. (**A**) Two files required as input, a file of lipidomics measurements (in Skyline, numerical matrix or mwTab format), and a file with experimental sample annotations (in CSV format). (**B**) Overview of uploaded datasets, showing lipid class distribution and lipid name errors. (**C**) Interactive editing of mismatching lipid names.

format (12) for LipidSuite to automatically extract class and chain information from lipid molecules.

The experimental sample annotation file should be provided as a CSV-formatted file, where first column is matching the sample names provided in the lipidomics dataset. Additional columns represent sample grouping or annotations that can be subsequently used in plots and statistical analyses.

Data QC

Three functionalities are implemented in LipidSuite to allow thorough data inspection. *Overview* provides the user an opportunity to correct any parsing errors, while *Samples* and *Lipids* allows data inspection through interactive plotting and data subsetting focusing on particular sample(s) or lipid(s).

Overview. Total number of samples, number of lipids and number of lipid classes are summarized, and the distribution of lipid classes are displayed as a pie chart (Figure 2B). Errors in lipid name parsing are highlighted, which can be corrected by navigating to the 'Lipids' tab and using the Search function to locate the misnamed lipid molecule (Figure 2C). Interactive editing of sample annotation can be performed similarly from the 'Samples' tab.

Samples. Interactive plots allow visual inspection of sample quality, and ability to choose a subset of data, e.g. to inspect QC samples only, via the 'Data Subset' tab. Data subsetting (Figure 3) can also be utilized to explore intensity distribution of certain lipid classes across samples. Users can customize the displayed plots by choosing plot type,

Figure 3. Interactive user-driven analysis workflow. LipidSuite allows users to try out different processing and analysis options, and directly assess the results with interactive visualization. Then, with the novel data subsetting functionality, users can focus on, or exclude subsets of their datasets, and redo the analysis and visualization on sliced datasets. Once done, static plots can be exported as images for publication and reports.

measure and color annotation. Plots are interactive, and can also be downloaded for record.

Lipids. Similarly, lipid QC plots allow visual inspection of lipid molecules or classes, to determine any data quality issues. By default, plots show the variability (%CV) for each lipid molecule across all samples. Powered by the data subsetting functionality, users can explore the technical variability in intensity or retention time of lipid molecules by focusing only on QC samples. Low quality lipids or lipid classes can be removed from analysis if required.

Preprocessing

Before differential analysis, LipidSuite requires three preprocessing steps to be performed, namely summarization, imputation and normalization. In cases where a Skyline export file is used as input, the preprocessing steps can be applied on any of the exported measures. While this functionality is kept for versatility of the LipidSuite, it is often logical to apply preprocessing only to the measure representing lipid concentration.

Summarization. This is an essential step only for targeted lipidomics that report multiple transitions per lipid molecule, where it summarizes all transitions to calculate a single intensity value for each molecule. In other cases where summarization is irrelevant, users can choose 'Don't Summarize' option from the drop-down list to skip this step. LipidSuite displays the number of lipid measurements before and after summarization.

Imputation. Since dealing with missing values is an essential step of analyzing lipidomics datasets, LipidSuite implements a wide range of imputation methods addressing different types of missingness (13). If the data is missing because of low abundance (missing not at random), users can use sophisticated QRILC (Quantile Regression Imputation of Left-Censored data) imputation method, impute missing values with minimum deterministic or probabilistic values, or replace them with zeros (14). On the other hand, if the data is missing at random, for example due to ion interference, users can utilize K-Nearest Neighbours, Singular Value Decomposition (13). Following imputation, Lipid-Suite displays a histogram of imputed intensities versus the true measurements, allowing users to assess the effect of imputation of the data distribution.

Normalization. To prepare lipidomics datasets for comparative analyses, LipidSuite allows normalization using Probabilistic Quotient Normalization (15,16), a widely used method for metabolomics data, which corrects for dilution factors across samples. Alternatively, users can normalize each lipid class against its spiked-in internal standards. To prevent low-intensity samples from skewing the normalization process, LipidSuite automatically detects and excludes 'Blank' samples. Similarly, users can exclude additional samples, such as outliers, from the normalization process. In cases where users provide a pre-normalized data, this step can be bypassed by selecting 'Dataset is already normalized'. User can easily assess the effect of normalization methods on their datasets though boxplots of intensities before and after normalization.

Data exploration

Users can explore their datasets in unsupervised or supervised manner, using Principal Component Analysis (PCA) and Orthogonal Partial Least-Squares Discriminant Analysis (OPLS-DA), respectively. PCA offers a simple visual exploration for separation between sample groups. Users can color samples on the plot using any of the annotations. Data subsetting enables users to replot PCA without outlier samples and focus on certain lipid classes to show how much they contribute to the separation between sample groups. On the other hand, OPLS-DA reveals which groups of lipids are discriminative between two sample groups (17). Top associated lipids can be extracted from the loading plot.

Differential analysis

LipidSuite supports both two- and multi-group differential analysis using moderated Student's *t*-test and ANOVA, respectively (18). Results from the univariate Analysis is used for lipid set enrichment and chain trend analysis, therefore it must be conducted first.

Univariate analysis. The groups to be compared are selected using tickboxes. For two-group comparison, moderated Student t-test is applied and the outputs include a volcano plot, showing $\log_2$-fold changes (LogFC) and adjusted *P*-values, and a result data table. For multi-group comparison, moderated ANOVA is applied and LipidSuite displays a modified volcano plot, with the average intensity of lipids instead of LogFC, since mean differences are not calculated in ANOVA. Furthermore, LipidSuite allows users to easily adjust for confounding factors such as age, using the 'Adjust by' dropdown menu.

Lipid set enrichment. To assess the lipid sets that are preferentially altered in the chosen comparison, individual lipids are ranked by LogFC, enrichment scores are calculated for each lipid set as previously described (10). Enrichment plot is displayed in interactive or static mode, and the raw table can be exported as CSV or Excel.

Chain trend analysis. To investigate the correlation between chain characteristics and differential regulation, trend plots showing the trend in logFC obtained in the univariate analysis and chain lengths or total unsaturations.

Heatmap

The final data visualization offered by LipidSuite is an interactive heatmap of the selected dataset. The hierarchical clustering with customizable sample and lipid annotations allow users to generate a heatmap to show co-clustering of samples and lipids.

CONCLUSION

LipidSuite webserver is designed to facilitate differential lipidomics data analysis and datamining, after lipid annotation has been made using other tools. With a focus on biological interpretation, LipidSuite provides multiple options for data quality control, imputation, normalization, data subsetting and confounder adjustment. In addition to univariate statistical analysis and OPLS multivariate analysis. Finally, a customisable heatmap from unsupervised clustering is generated. Together, LipidSuite offers a convenient interface for life science researchers to conduct sophisticated lipidomics data analytics with biological interpretation. We anticipate this new free webserver to accelerate lipidomics research.

DATA AVAILABILITY

Any required links are present in the manuscript as described.

REFERENCES

1. Lam,S.M., Wang,Z., Li,B. and Shui,G. (2021) High-coverage lipidomics for functional lipid and pathway analyses. *Anal. Chim. Acta*, **1147**, 199–210.
2. Zullig,T. and Kofeler,H.C. (2021) High resolution mass spectrometry in lipidomics. *Mass Spectrom. Rev.*, **40**, 162–176.
3. Chong,J., Soufan,O., Li,C., Caraus,I., Li,S., Bourque,G., Wishart,D.S. and Xia,J. (2018) MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.*, **46**, W486–W494.
4. Guitton,Y., Tremblay-Franco,M., Le Corguille,G., Martin,J.F., Petera,M., Roger-Mele,P., Delabriere,A., Goulitquer,S., Monsoor,M., Duperier,C. *et al.* (2017) Create, run, share, publish, and reference your LC-MS, FIA-MS, GC-MS, and NMR data analysis workflows with the Workflow4Metabolomics 3.0 Galaxy online infrastructure for metabolomics. *Int. J. Biochem. Cell Biol.*, **93**, 89–101.
5. Wanichthanarak,K., Fan,S., Grapov,D., Barupal,D.K. and Fiehn,O. (2017) Metabox: a toolbox for metabolomic data analysis, interpretation and integrative exploration. *PLoS One*, **12**, e0171046.
6. Kopczynski,D., Hoffmann,N., Peng,B. and Ahrends,R. (2020) Goslin: a grammar of succinct lipid nomenclature. *Anal. Chem.*, **92**, 10957–10960.
7. Ni,Z. and Fedorova,M. (2020) LipidLynxX: a data transfer hub to support integration of large scale lipidomics datasets. bioRxiv doi: https://doi.org/10.1101/2020.04.09.033894, 26 July 2020, preprint: not peer reviewed.
8. Clair,G., Reehl,S., Stratton,K.G., Monroe,M.E., Tfaily,M.M., Ansong,C. and Kyle,J.E. (2019) Lipid Mini-On: mining and ontology tool for enrichment analysis of lipidomic data. *Bioinformatics*, **35**, 4507–4508.
9. Molenaar,M.R., Jeucken,A., Wassenaar,T.A., van de Lest,C.H.A., Brouwers,J.F. and Helms,J.B. (2019) LION/web: a web-based ontology enrichment tool for lipidomic data analysis. *Gigascience*, **8**, giz061.
10. Mohamed,A., Molendijk,J. and Hill,M.M. (2020) lipidr: a software tool for data mining and analysis of lipidomics datasets. *J. Proteome Res.*, **19**, 2890–2897.
11. Sud,M., Fahy,E., Cotter,D., Azam,K., Vadivelu,I., Burant,C., Edison,A., Fiehn,O., Higashi,R., Nair,K.S. *et al.* (2016) Metabolomics workbench: an international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res.*, **44**, D463–D470.
12. Sud,M., Fahy,E., Cotter,D., Brown,A., Dennis,E.A., Glass,C.K., Merrill,A.H. Jr, Murphy,R.C., Raetz,C.R., Russell,D.W. *et al.* (2007) LMSD: LIPID MAPS structure database. *Nucleic Acids Res.*, **35**, D527–D532.

13. Wei,R., Wang,J., Su,M., Jia,E., Chen,S., Chen,T. and Ni,Y. (2018) Missing value imputation approach for mass spectrometry-based metabolomics data. *Sci. Rep.*, **8**, 663.

14. Lazar,C. (2015) imputeLCMD: a collection of methods for left-censored missing data imputation. *R package*.

15. Dieterle,F., Ross,A., Schlotterbeck,G. and Senn,H. (2006) Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Anal. Chem.*, **78**, 4281–4290.

16. Välikangas,T., Suomi,T. and Elo,L.L. (2018) A comprehensive evaluation of popular proteomics software workflows for label-free proteome quantification and imputation. *Brief. Bioinform.*, **19**, 1344–1355.

17. Trygg,J. and Wold,S. (2002) Orthogonal projections to latent structures (O-PLS). *J. Chemom.*, **16**, 119–128.

18. Ritchie,M.E., Phipson,B., Wu,D., Hu,Y., Law,C.W., Shi,W. and Smyth,G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47.