



SOFTWARE TOOL ARTICLE

REVISED BioPAN: a web-based tool to explore mammalian lipidome metabolic pathways on LIPID MAPS

[version 2; peer review: 3 approved]

Caroline Gaud ¹, Bebiana C. Sousa ², An Nguyen¹, Maria Fedorova³, Zhixu Ni ³, Valerie B. O'Donnell⁴, Michael J.O. Wakelam²⁺, Simon Andrews ¹, Andrea F. Lopez-Clavijo ²

¹Bioinformatics Group, Babraham Institute, Babraham Research Campus, Cambridge, CB22 3AT, UK²Lipidomics facility, Babraham Institute, Babraham Research Campus, Cambridge, CB22 3AT, UK³Institute of Bioanalytical Chemistry, Faculty of Chemistry and Mineralogy, Center for Biotechnology and Biomedicine, Universität Leipzig, Leipzig, 04109, Germany⁴Systems Immunity Research Institute, School of Medicine, Cardiff University, Cardiff, CF14 4XN, UK

+ Deceased author

v2 First published: 06 Jan 2021, 10:4
<https://doi.org/10.12688/f1000research.28022.1>

Latest published: 09 Jun 2021, 10:4
<https://doi.org/10.12688/f1000research.28022.2>

Abstract

Lipidomics increasingly describes the quantification using mass spectrometry of all lipids present in a biological sample. As the power of lipidomics protocols increase, thousands of lipid molecular species from multiple categories can now be profiled in a single experiment.

Observed changes due to biological differences often encompass large numbers of structurally-related lipids, with these being regulated by enzymes from well-known metabolic pathways. As lipidomics datasets increase in complexity, the interpretation of their results becomes more challenging. BioPAN addresses this by enabling the researcher to visualise quantitative lipidomics data in the context of known biosynthetic pathways. BioPAN provides a list of genes, which could be involved in the activation or suppression of enzymes catalysing lipid metabolism in mammalian tissues.

Keywords

LIPID MAPS, Lipidomics, Biosynthetic pathway analysis, lipids, lipid profiling.



This article is included in the **Bioinformatics** gateway.

Open Peer Review

Approval Status

	1	2	3
version 2 (revision) 09 Jun 2021			
version 1 06 Jan 2021	 view	 view	 view

1. **Amaury Cazenave Gassiot** , Yong Loo Lin School of Medicine, Singapore, Singapore
2. **Phillip Whitfield**, University of Glasgow, Glasgow, UK
3. **Ginger Milne**, Vanderbilt University Medical Center, Nashville, USA

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Andrea F. Lopez-Clavijo (andrea.lopez@babraham.ac.uk)

Author roles: **Gaud C:** Data Curation, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **C. Sousa B:** Formal Analysis, Investigation, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Nguyen A:** Conceptualization, Methodology; **Fedorova M:** Software, Validation, Writing – Review & Editing; **Ni Z:** Software, Validation, Writing – Review & Editing; **O'Donnell VB:** Supervision, Writing – Review & Editing; **Wakelam MJO:** Conceptualization, Data Curation, Methodology, Supervision; **Andrews S:** Software, Supervision, Validation, Writing – Review & Editing; **Lopez-Clavijo AF:** Data Curation, Formal Analysis, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: We gratefully acknowledge funding from Wellcome Trust: SC25365/510342, 203014/Z/16/Z (LIPID MAPS to VOD, SA, MJOW, AFLC). LIPID MAPS is also supported by unrestricted sponsorship from Avanti Polar Lipids, Cayman Chemical and Merck. Financial support from the German Federal Ministry of Education and Research (BMBF) within the framework of the e:Med research and funding concept for SysMedOS project (to MF and ZN) are gratefully acknowledged.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Gaud C *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Gaud C, C. Sousa B, Nguyen A *et al.* **BioPAN: a web-based tool to explore mammalian lipidome metabolic pathways on LIPID MAPS [version 2; peer review: 3 approved]** F1000Research 2021, 10:4 <https://doi.org/10.12688/f1000research.28022.2>

First published: 06 Jan 2021, 10:4 <https://doi.org/10.12688/f1000research.28022.1>

REVISED Amendments from Version 1

- In the introduction to summarise the overall objective of the application note.
- In the methods to show the corresponding Z-score to $p < 0.05$.
- In the conclusions about further developments to the database.
- In the pathway calculation section to choose the level of significance in the calculation.
- In the pathway options to include logical operators
- In the Operation section to describe “unprocessed” species.
- In the export option to include other graphics formats
- Some minor typos were also corrected.

Any further responses from the reviewers can be found at the end of the article

Introduction

Lipids (fats) are essential and diverse families of molecules that play structural, energy storage and signalling roles. They are connected through complex metabolic pathways, which comprise linked series of enzymatic reactions (several are outside cells, *e.g.* PLA2 isoforms, autotaxin). Thus, lipids can be substrates, products or intermediates. It has been estimated that there are approximately 3–5000 different lipid species in mammalian cells although the true number is still unknown and extremely difficult to reliably measure^{1,2}. In recent years, great advances have been made in our ability to experimentally determine the elemental composition and quantitation of lipid levels in biological samples, with the advent of rapid scanning benchtop mass spectrometers (MS), in particular high-resolution configurations such as Orbitrap, and ToF. Liquid chromatography-MS, either tandem or high resolving power, interfaced with processing pipelines such as Lipid Data Analyser (LDA), Lipid-Finder, LipidHunter, MS-Dial, XCMS and many others makes it feasible to simultaneously monitor the dynamic changes in hundreds of lipid molecular species in biological samples^{3–7}.

As the quantity and detail of quantitative lipid data continues to grow it has become considerably more challenging to interpret the complex sets of changes within the lipidome⁸. Relevant biological perturbations usually happen not at the level of a single lipid molecular species in isolation, but as broad sets of changes over entire lipid classes and subclasses⁹. For example, a group of phosphatidylcholines will tend to change together in the same direction, since they are being regulated by the same enzyme isoform(s). This introduces characteristic patterns in the data that can reveal important clues as to the underlying level of genetic and transcriptomic regulation. Consequently, there is a need for software tools to automate and facilitate biosynthetic pathway scale analyses with lipids grouped according to structural motifs. Nguyen *et al.* and Hann *et al.* described the analysis of quantitative lipidomics datasets at pathway levels^{10–12}. Here, we have integrated the biosynthetic metabolic levels into BioPAN.

So, BioPAN combines current knowledge of lipid metabolism with a statistical analysis functionality, by comparing two biological conditions to identify activated or suppressed pathways, and presents the results in an interactive graphical display. BioPAN works with data resolved to the level of individual lipid molecular species, but it can also aggregate results at the subclass level to simplify the interpretation. Mammalian reaction information is converted into a metabolic pathway within BioPAN, allowing it to predict the most likely lipid transforming genes modulated by an experiment, to provide direct biological insight and generate hypotheses that can be experimentally tested.

BioPAN is openly available on the LIPID MAPS® Lipidomics Gateway, at <https://lipidmaps.org/bioipan/>. The tool is designed to allow users to upload and analyse their own data, and example datasets are also provided for evaluation. BioPAN is fully integrated with LipidLynxX, allowing users to convert different naming (short notation) conventions from diverse software inputs to be able to make it readable on BioPAN. Additionally, LipidLynxX will convert lipid results with fatty acid position, double bond location, and stereochemistry into a sum composition of carbon number and double bond equivalents. For example, DAG (16:0/20:1(Z11)), is converted by LipidLynxX to DG 36:1. BioPAN offers a link with LIPID MAPS Structure Database (LMSD)⁹. So, BioPAN aids for automatic integration of lipid metabolism with lipid profiles, finding strong relationships between lipid substrates and lipid products catalysed by active or suppressed enzymes.

Methods

Implementation

The statistical model used in BioPAN was originally described by Nguyen *et al.*¹⁰. In brief, BioPAN uses data from replicated quantitative lipidomics experiments with two biological experiments: a condition of interest (treated) and a control condition. BioPAN calculates statistical scores for all possible lipid pathways to predict which are active or suppressed in the treated samples compared to the control set.

The BioPAN workflow relies on the calculation of the Z-score, which considers the mean and the standard deviation of the experiment assuming a normally distributed data of lipid subclasses. A probability function (P) is computed and subtracting from one to obtain Q, which in turn is the probability that a Z-score is due to chance. The Z-score is then used to predict whether a particular reaction is significantly ($p < 0.05$) changing between control and treated conditions. Z-scores for all reactions in a pathway are combined using a cumulative function (CDF) to give a global pathway Z-score. The set of equations used to calculate Z-scores are presented and discussed elsewhere¹⁰. Changing reactions are classified as activated or suppressed depending upon the direction of change. BioPAN, by default, calls a reaction or pathway as significantly modified at a level of $p < 0.05$ (corresponding to $Z > 1.645$).

Operation

BioPAN is implemented as an open access web-based tool. The front end uses HTML, jQuery (v3.5.1) for the interface and

the Cytoscape.js library (v3.10.2)¹³ to draw and manipulate the results graphs. The server-side code uses PHP (v7.4.11) for request handling and R (v4.0.2) for data processing and statistical analysis. BioPAN runs on all modern web browsers and is available at <https://lipidmaps.org/biopan/>.

BioPAN allows uploading of lipidomics datasets containing quantitative data from two different experiments or conditions (e.g. wild type against treated/knock-out). The use of BioPAN is simple and intuitive with a user-friendly interface. The data files should have a minimum of two replicates per condition, as BioPAN compares the changes between the experiments. BioPAN takes an input file in .csv (Comma-Separated Values). The first column of the file should include the lipid abbreviation subclass followed by the notation e.g. PC 38:4, PC 18:0_20:4, PC 18:0/20:4, or PC 18:0/20:4(5Z,8Z,11Z,14Z), for glycerophospholipids, glycerolipids, and sterol lipids. Sphingolipid molecular species written as e.g. d18:1/20:4, or 18:1;O2/20:4, where 18:1;O2 is the sphingosine base and 20:4 is the N-acyl chain linked to the sphingosine base are also recognised in BioPAN. Non-conventional nomenclature like DG(aa-34:0, DG(ea-34:1) is not recognised by BioPAN and the user manually need to change the lipid subclass abbreviation¹⁴. There are additional instructions about the structure of the data file following the link <https://lipidmaps.org/biopan/doc/step1.html>, which directs the user to each lipid subclass abbreviation in LIPID MAPS® Lipidomics Gateway.

The first step in BioPAN is to load a file of quantitative lipidomics data. After uploading, BioPAN uses the LipidLynxX¹⁵ tool to cross-match some lipid names into the LIPID MAPS® Lipidomics Gateway nomenclature style according to the guidelines from COMP_DB (https://www.lipidmaps.org/resources/tools/bulk_structure_searches_documentation.php)^{9,14,16}. BioPAN then classifies the submitted lipids molecular species as either unrecognised, processed, or unprocessed. Unrecognised molecular species are lipids whose subclasses are not included into BioPAN database, while unprocessed species are part of BioPAN, but were not associated with any reactions. The unprocessed category corresponds to lipid molecular species that did not have a matching molecular species, as a reactant and product, with the same number of carbons and double bonds. For example, a substrate molecular species, such as PA 34:2 should have a matching product with the same sum composition (e.g. DG 34:2). Additionally, some biological reactions require fatty acid or fatty-acyl-CoA molecular species. So, a lack of these subclasses in the input file prevents BioPAN to build up the reaction between substrates and products. If the substrate or product is not part of other reaction within the pathway, BioPAN will report it as unprocessed, such is the case of triacylglycerides (TG) subclasses. On other instances, like phosphatidylcholine (PC) and lysophosphatidylcholine (LPC), where PC is the substrate in the reaction to produce phosphatidic acid (PA) and LPC is the substrate to produce lysophosphatidic acid (LPA), those lipid subclasses molecular species can be considered as processed by BioPAN. BioPAN's database was manually collated followed by a validation using available literature. The

database currently contains 94 lipid reactions identified in mammals, covering 41 lipid subclasses. Only processed molecular species, e.g. those which can be associated with at least one reaction, are used for downstream analysis.

The second step within BioPAN requires users to associate each sample with a condition. (e.g. sample names: control1, treated1, or control2, treated2, and so on, with their respective replicates). Each condition requires a minimum of two replicas (three or more replicas per condition are strongly recommended). Once assignments have been made, BioPAN searches for reactions and pathway changes between these. Following analysis completion, the user is directed to the results page (Figure 1). The main viewport shows an interactive graph comprising lipids and reactions assigned to the lipids included in the input file uploaded. The user can select how the pathways are displayed in the interactive graph by manually moving the nodes in the graph, while the link between product and substrate is maintained. Thus, providing a view of all possible reaction pathways. The side panel provides a legend for the graph and controls to modify the specifics of the analysis performed and the level of detail in the graph. The legend describes the shape of the nodes, which denote the lipid subclass, and the colour of the nodes represent the pathway status (active/suppressed or none). Each arrow between two nodes (lipid subclass/lipid molecular species) depicts the reaction direction between the two lipids. The colour of the arrows depends on the value of the Z-score (see the *Implementation* section), where green indicates a positive Z-score and purple negative, as shown in Figure 1. The user can click on the arrows connecting each node and a pop table is generated by BioPAN showing all the lipid molecular species selected and non-selected in the reaction between two nodes (see Extended Data, Table S-1). Selecting the arrow is only functional when selecting the lipid subclass level in the pathway options (described in the section below). Additionally, information on lipid subclasses and molecular species are available by clicking on the nodes, linking to the LMSD database (<https://www.lipidmaps.org/data/structure/index.php>).

The bottom panel shows a tabular view of statistically significant ($p < 0.05$ equivalent to Z-score > 1.645) results.

Display options. The pathway options on the left of the screen allow users to customize the display of results in the graph. BioPAN arbitrarily chooses two conditions to compare (condition of interest and control condition, see Figure 1). The user can change conditions to compare using the drop-down menu (Figure 1). Users can choose to view pathways based on either the lipid subclass or on the fatty acyl composition. They can also choose which pathways to highlight in the main graph. Users can opt for either significantly active (condition of interest $>$ control) or significantly suppressed (condition of interest $<$ control) pathways. Where the same reaction can participate in multiple pathways, users can also choose to view only the most active (or most suppressed) pathway. This will highlight only the pathway with the highest individual pathway level Z-score for pathways containing reactions where a substrate is the starting agent

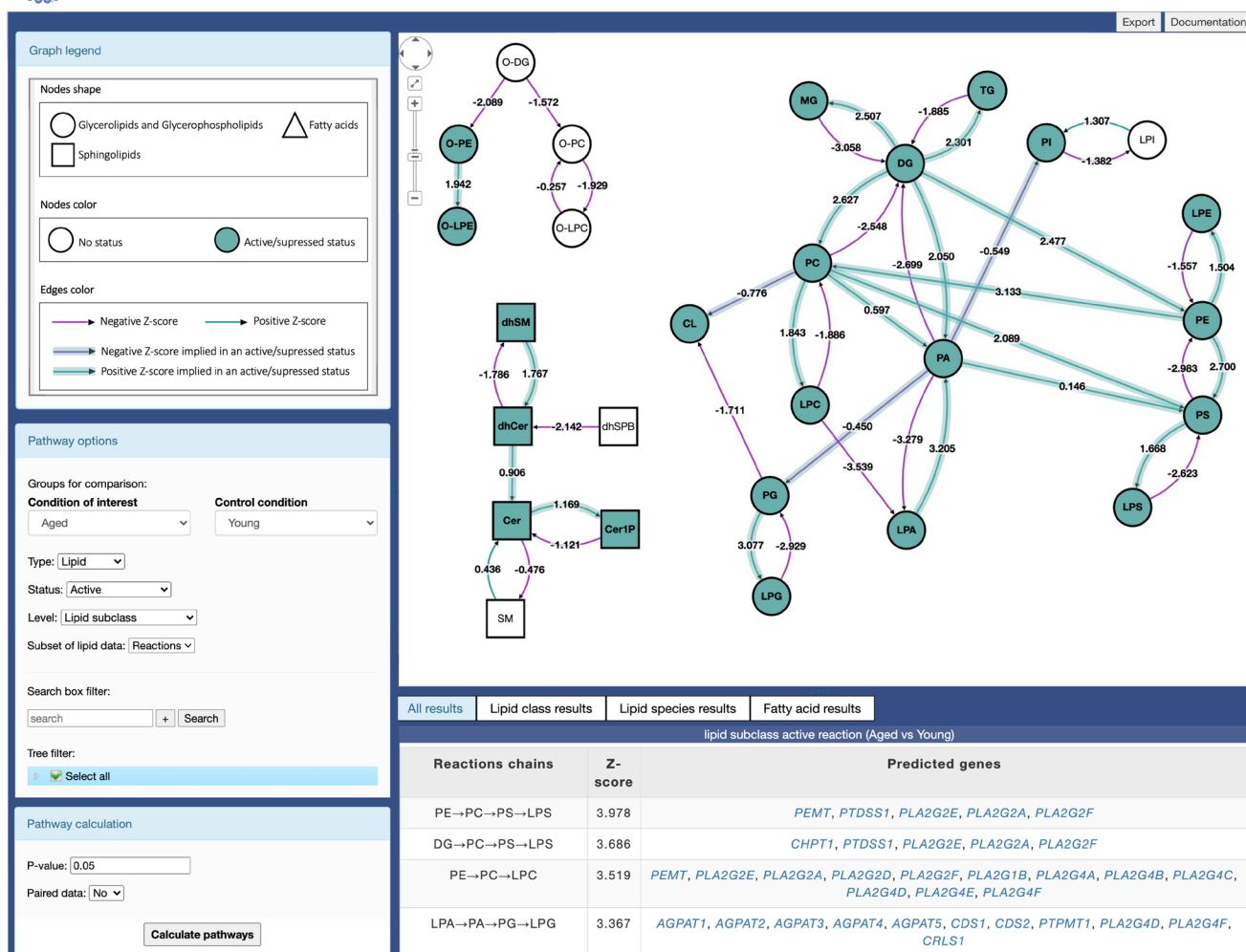


Figure 1. Main results page view for BioPAN. Pathways are displayed as an interactive graph at the centre of the screen. Users have control over the information presented in the graph using the menus on the left side panel and results tables are displayed at the bottom of the screen.

in a reaction. In other words, selecting most active or most suppressed *status* still follow the rule for the active/suppressed status in terms of colours and Z-scores as discussed above. However, the Z-scores in an active/suppressed single reaction (e.g. PC→DG) are compared for substrates, intermediates, and products following each step of the pathway. For example, in the cerebral cortex (with pathway options: active status, subclass level, reaction subset of lipid data, 0.05 *p* value, no paired-data), reactions within multiple pathways like PC→DG→MG and PC→DG→PA. PC is the initial substrate and have individual reaction Z-scores of 3.083 (PC→DG), 2.054 (DG→MG), and -0.408 (DG→PA). It can be observed above that PC→DG is common for the first step in both pathways, and then BioPAN selects the most active reaction with the highest Z-score for the second part of the pathway. So, the Z-score of DG→MG is greater than the score for DG→PA. Thus, PC→DG→MG is most active pathway compared to PC→DG→PA. There

might be more than one active pathway with one lipid-initiated substrate (e.g. DG→TG, DG→PE→PS, and DG→PC→CL), the same principle is applied, and BioPAN will consider only one of those pathways as most active.

Users can also choose to have the nodes in the graph aggregated into lipid subclasses, or they can view individual lipid molecular species. by using the level drop down menu. Alternatively, the user can opt to view only reaction chains using the subset of lipid data menu, which contains pathways stored within BioPAN (e.g. Biosynthesis of Sphingomyelin or *de novo* lipogenesis). Changing any parameter within the pathway options affects both the graph and the results tables.

An all lipid displayed view can be filtered to display only a user-selected subset of lipids. These can be chosen either from a text box, or using a tree of checkboxes for each subclass,

allowing users to focus on particular lipids of interest (see Extended Data, section S-2, Figure S-1). The search box allows the user to search for one or two queries using the logical operators AND / OR. Selecting the AND operator, the user can view lipid subclasses or molecular species for which the name implies both queries. For example, searching for “PC” and “34:0” on the lipid molecular species graph displays lipids “PC(34:0)” and “O-PC(34:0)”. Choosing the OR operator allow the user to visualise lipids whose names includes one of the two searches. For instance, searching “LPC” or “LPA” on the subclasses graph displays lipids “LPC”, “O-LPC” and “LPA”.

Calculation options. Under the Pathway Calculation section, the user can control details of the statistical calculation. They can change the threshold for significance from the default of $p < 0.05$. If replicates in the uploaded data were generated as matched pairs (e.g. control and treated sample coming from the same animal) then the calculation can also take this into account in the *t*-test step of the analysis.

Exporting results. Results of the statistical analysis of active/suppressed pathways are presented in four tables at the bottom of the page, where predicted gene changes can be visualised. Export options are also available on the top right corner in the BioPAN viewport, where the main graph can be exported in several formats (JPEG / PNG / JSON / TXT) and tables are exported as TSV files.

Example use case

Lipidomics data from cerebral cortex and liver of young and aged mice¹⁷ were used here as an example to illustrate how to use BioPAN and interpret its results. Ando *et al.*¹⁷ compared the lipid profile of different tissues in young and aged mice with the aim of detecting age-related as well as tissue-specific lipid modifications. The authors clustered alkylacyldiacylglycerols and deoxysphingolipids molecular species, reporting accumulation of deoxysphingolipids and ether-linked diacylglycerols in both tissues. Additional lipid subclasses were measured from Ando *et al.* (Additional information file 3), but those results were not discussed in the original publication¹⁰. The results from Ando *et al.* are used here to show how BioPAN can provide a wider analysis of the lipid level changes and the relationship between different lipid subclasses.

The datasets from the original paper, were downloaded as TXT files from Metabolomics Workbench (Project ID PR000713, Studies ID ST001065 and ST001066) and are available as underlying data (see *data availability* section). Each file was imported as described in the method implementation section. Pathway options and pathway calculations settings were as follows: Type: Lipid; Status: Active; Level: Lipid subclass; Subset of lipid data: Reactions; Condition of interest: Aged; Control condition: Young; *p* value: 0.05; Paired data: No.

The lipid network at the lipid subclass level is shown in **Figure 2A** for the liver and **Figure 2B** for the cerebral cortex.

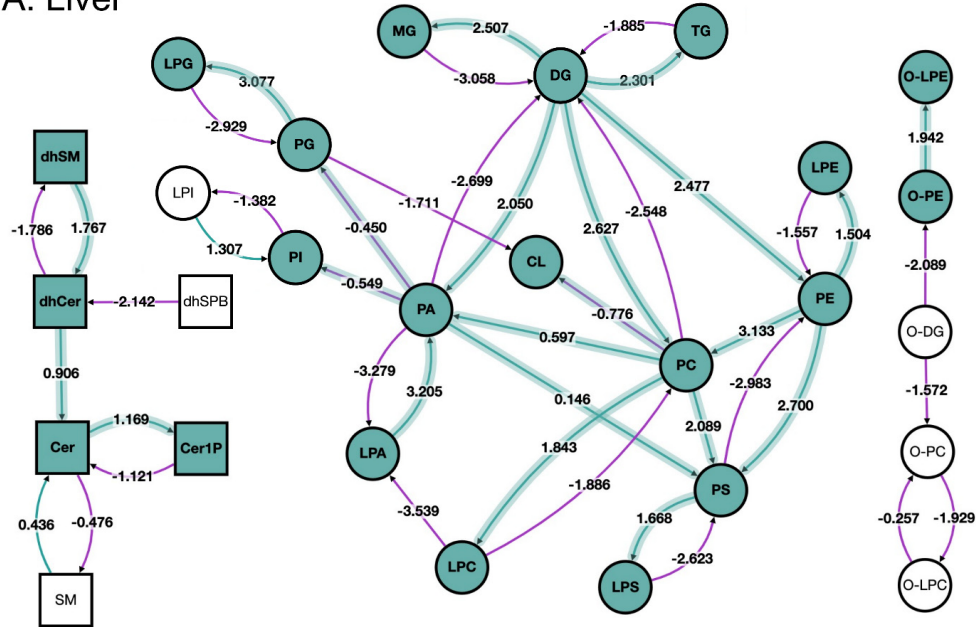
The lipid pathway graph allows the user to quickly spot trends and differences between the two sets of samples. **Figure 2** shows that reactions using ether-linked diacylglycerol (O-DG) as a substrate are suppressed in both liver and cerebral cortex in old versus young, which agrees with the reported accumulation of this particular lipid subclass by Ando *et al.*¹⁷. Here, BioPAN also found tissue-specific differences in lipid metabolism. For example, the reaction pathway results of aged mouse liver suggest active metabolism leading to accumulation of triacylglycerols (TG, Z-score = 2.301), a form of energy storage, contrary to a catabolic metabolism of TG observed in the cerebral cortex (Z-score = -2.612).

BioPAN also shows more differences in the metabolism of sphingolipids between the tissues. Specifically, in the liver the pathway is shifted towards the formation of ceramide-1-phosphate (Cer1P), while in the cerebral cortex sphingolipid metabolism points towards the formation of sphingomyelin (SM). Moreover, in the liver, glycerophospholipids and lysophospholipids synthesis is active with exception of LPA, contrary to what it is observed in the cerebral cortex, where glycerophospholipids and lysophospholipids pathways are suppressed, but production of LPA is favoured.

BioPAN provides tables showing the suggested genes known to be involved in each reaction. **Table 1** shows the lipid subclass active reaction results generated for the liver dataset. For example, the pathway PE→PC→PS→LPS designates the genes *PEMT*, *PTDSSI*, *PLA2G2E*, *PLA2G2A*, *PLA2G2F*. It should be noted that the default nomenclature of the genes in BioPAN corresponds to human species. In the example case presented here, the lipidomics data was obtained from mice, and although most often human and mouse genes are the same, there can be subtle differences between species. Thus, the user is encouraged to assign the gene according to the species in their own study. To aid the user in identifying the species each gene in **Table 1** has a link that exists in the LIPID MAPS® Proteome Database (LMPD) (<https://lipidmaps.org/resources/databases/index.php?tab=lmpd>). For example, in **Table 1** *PEMT* gene refers to a table containing the name and species of the gene, including a unique LMP ID for each gene/species pair. Thus, LMP002477 is assigned to the gene *Pemt* for mouse, which converts PE to PC. The status of this pathway might be changed by the loss of function or altered activity in any of the enzymes involved. Therefore, the list provided by BioPAN might help guide integration of lipidomics with proteomics or transcriptomics data as it directly suggests target proteins for further analysis.

Looking at fatty acids (FA) pathways at the molecular species level with BioPAN (**Figure 3**) shows the FA network obtained for liver and cerebral cortex of aged in comparison to young mice. The results show active metabolism towards the production of monounsaturated FA 16:1 and polyunsaturated FA 20:4, FA 20:5, and FA 22:4 in the cerebral cortex compared to the liver. Additionally, BioPAN has highlighted accumulation of

A. Liver



B. Cerebral cortex

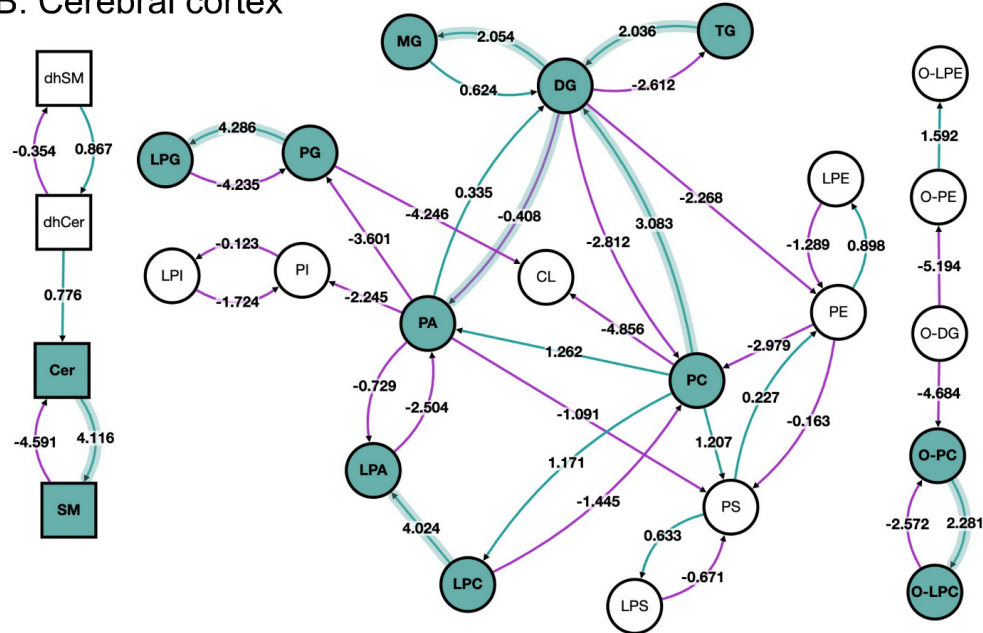


Figure 2. BioPAN lipid networks. Lipid network graphs exported from BioPAN for the liver (A) and the cerebral cortex (B) of aged mice compared to young mice¹⁷. Green nodes correspond to active lipids and green shaded arrows to active pathways. Reactions with a positive Z score have green arrows while negative Z scores are coloured purple. Pathways options: aged condition of interest, young control condition, lipid type, active status, subclass level, reaction subset of lipid data, *p* value 0.05, and no paired-data.

the long acyl chain, polyunsaturated FA 24:5 in liver, which is associated with the activation of the *ELOVL2* gene found in Table 2. Moreover, the results shown in Table 2 suggest a

general activation of elongases (*ELOVL*) rather than desaturases (*SCD*) which was found to be a common trait in both tissues.

Table 1. BioPAN predicted genes of the lipid active reactions in liver of aged vs young mice. Table contains the active reactions chains found by BioPAN according to the Z-score values. Several genes are predicted as being involved in the current status of each reaction based on their function. Pathway options: aged condition of interest, young control condition, lipid type, active status, subclass level, reaction subset of lipid data, *p* value 0.05, no paired-data.

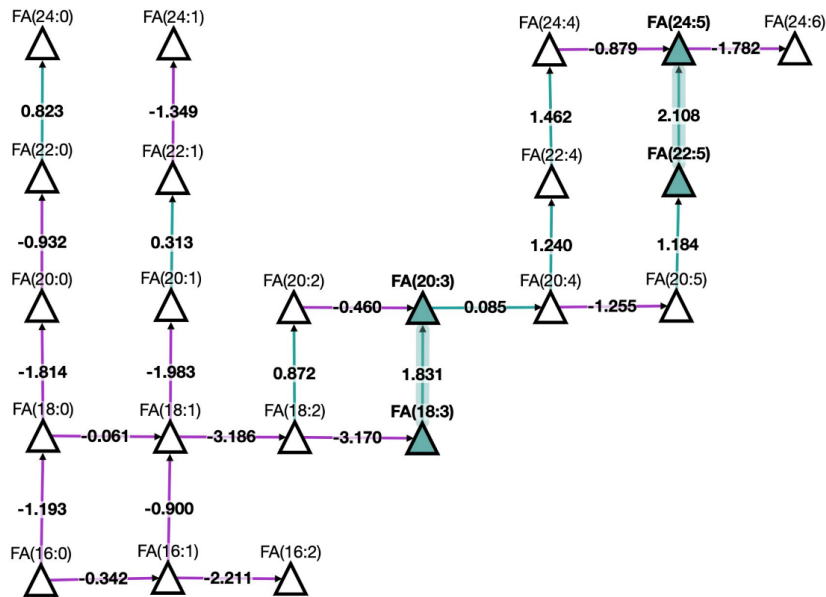
lipid subclass active reaction (Aged vs Young)		
Reaction chains	Z-score	Predicted genes
PE→PC→PS→LPS	3.978	<i>PEMT, PTDSS1, PLA2G2E, PLA2G2A, PLA2G2F</i>
DG→PC→PS→LPS	3.686	<i>CHPT1, PTDSS1, PLA2G2E, PLA2G2A, PLA2G2F</i>
PE→PC→LPC	3.519	<i>PEMT, PLA2G2E, PLA2G2A, PLA2G2D, PLA2G2F, PLA2G1B</i>
LPA→PA→PG→LPG	3.367	<i>AGPAT1, AGPAT2, AGPAT3, AGPAT4, AGPAT5, CDS1, CDS2, PTPMT1, PLA2G4D, PLA2G4F, CRLS1</i>
PE→PC→PA→PG→LPG	3.179	<i>PEMT, PLD1, PLD2, CDS1, CDS2, PTPMT1, PLA2G4D, PLA2G4F, CRLS1</i>
DG→PC→LPC	3.161	<i>CHPT1, PLA2G2E, PLA2G2A, PLA2G2D, PLA2G2F, PLA2G1B</i>
PG→LPG	3.077	<i>PLA2G4D, PLA2G4F, CRLS1</i>
LPA→PA→PS→LPS	2.898	<i>AGPAT1, AGPAT2, AGPAT3, AGPAT4, AGPAT5, CDS1, PTDSS1, PLA2G2E, PLA2G2A, PLA2G2F</i>
DG→PE→LPE	2.815	<i>CEPT1, PLA2G4C</i>
PE→PS	2.700	<i>PTDSS2</i>
PC→PS→LPS	2.657	<i>PTDSS1, PLA2G2E, PLA2G2A, PLA2G2F</i>
DG→MG	2.507	<i>PNPLA2, PNPLA3</i>
DG→TG	2.301	<i>DGAT2</i>
DG→PC→PA	2.280	<i>CHPT1, PLD1, PLD2</i>
dhSM→dhCer→Cer→Cer1P	2.218	<i>SGMS1, SGMS2, DEGS1, DEGS2, CERK</i>
DG→PA	2.050	<i>DGKA, DGKB, DGKD, DGKE, DGKG, DGKH, DGKI, DGKK, DGKQ, DGKZ</i>
O-PE→O-LPE	1.942	<i>PLA2G4A</i>
LPA→PA→PI	1.878	<i>AGPAT1, AGPAT2, AGPAT3, AGPAT4, AGPAT5, CDS1, CDS2, CDIPT</i>
PC→LPC	1.843	<i>PLA2G2E, PLA2G2A, PLA2G2D, PLA2G2F, PLA2G1B</i>
PE→PC→PA→PI	1.837	<i>PEMT, PLD1, PLD2, CDS1, CDS2, CDIPT</i>
PS→LPS	1.668	<i>PLA2G2E, PLA2G2A, PLA2G2F</i>
PE→PC→CL	1.667	<i>PEMT, TAZ</i>

Conclusion

Modern lipidomics techniques generate a huge amount of data and information on many lipid subclasses, making the interpretation of results challenging. BioPAN provides several advantages over simpler analyses of individual lipids. Mapping of the lipid network graph provides a bigger picture allowing a new perspective on the data where new connections can be made, and a study can be moved forward into particular areas of

interest. BioPAN offers a quick visualisation of differences in lipid pathways in mammalian studies and suggests genes and their enzyme products involved in those differences. Further work into the BioPAN database will be carried out to incorporate new biosynthetic pathways, such as: cholesterol, eicosanoids, and oxidised lipid subclasses, among others. Addition of glucosyl and galactosyl ceramides (Glc/Gal-Cer) to complement to the existing sphingolipids pathway in BioPAN will also be considered.

A. Liver



B. Cerebral cortex

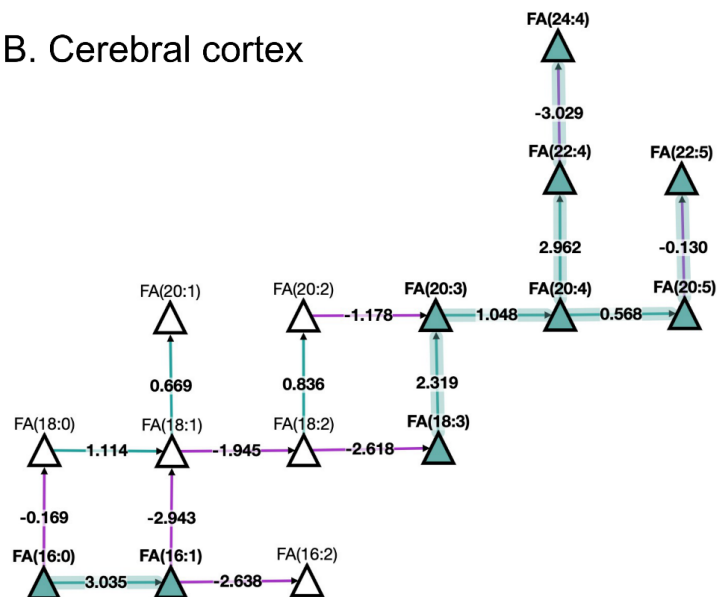


Figure 3. BioPAN fatty acids networks. FA graphs exported from BioPAN tool for the liver (**A**) and the cerebral cortex (**B**) of aged mice compared to young mice¹⁷. Green nodes correspond to active lipids and green shaded arrows to active pathways. Reactions with a positive Z score have green arrows while negative Z scores are coloured purple. Pathways options: aged condition of interest, young control condition, lipid type, active status, subclass level, reaction subset of lipid data, *p* value 0.05, and no paired-data.

Table 2. BioPAN predicted genes for fatty acid active reactions in liver of aged vs young mice. Table contains the active reaction chains found by BioPAN according to the Z-score values. Several genes are predicted as being involved in the current status of each reaction based on their function. Pathway options: aged condition of interest, young control condition, fatty acid type, active status, *p* value 0.05, no paired-data.

fa active reaction (Aged vs Young)		
Reaction chains	Z-score	Predicted genes
FA(22:5)→FA(24:5)	2.108	ELOVL2
FA(18:3)→FA(20:3)	1.831	ELOVL5, ELOVL7

Data availability

Underlying data

The datasets analysed as an example are available at the NIH Common Fund's National Metabolomics Data Repository (NMDR) website, the Metabolomics Workbench, <https://www.metabolomicsworkbench.org/> where it has been assigned Project ID PR000713 (studies ST001065 and ST001066).

The data can be accessed directly via its Project DOI: <https://dx.doi.org/10.21228/M8HM48>, <https://www.metabolomicsworkbench.org/data/DRCCMetadata.php?Mode=SetupDownloadResults&StudyID=ST001065> and <https://www.metabolomicsworkbench.org/data/DRCCMetadata.php?Mode=SetupDownloadResults&StudyID=ST001066>.

Open Science Framework: BioPAN: a web-based tool to explore mammalian lipidome metabolic pathways on LIPID MAPS, <https://doi.org/10.17605/OSF.IO/PKD6B>¹⁸.

This project contains the modified dataset from Ando *et al.*¹⁷ according to BioPAN's requirements:

- Cerebral cortex dataset: <https://osf.io/reavk/>

- Liver dataset: <https://osf.io/zd4as/>

Extended data

Open Science Framework: BioPAN: a web-based tool to explore mammalian lipidome metabolic pathways on LIPID MAPS, <https://doi.org/10.17605/OSF.IO/PKD6B>¹⁸.

This project contains the following extended data:

- Table S1, reaction table.
- Section S2, filtering and tree checkbox discussion.
- Figure S1, Filter tool example for acyl chains 38:4 and 38:5.

Software availability

Software, documentation and tutorial video are available from: <https://lipidmaps.org/biopan/>.

Archived source code at the time of the publication: <https://doi.org/10.17605/OSF.IO/PKD6B>¹⁸.

The user can try BioPAN locally by following the readme instruction file located in: <https://osf.io/7g2dx/>.

License: GNU GPL 3.0

Acknowledgements

"In memorium" The authors would like to dedicate this software application note to Michael J. O. Wakelam, who devoted his 40 years academic career to the field of lipidomics and mass spectrometry. This document is a testament to Michael's depth knowledge in lipidomics and his immense capacity to translate lipid data into biological meaning. We ought the development of BioPAN's database to Michael's dream of integrating lipids with enzymatic changes. BioPAN is Michael's legacy to the whole world as an open access tool he envisioned many moons ago. He is truly missed!

References

- Slatter DA, Aldrovandi M, O'Connor A, *et al.*: **Mapping the Human Platelet Lipidome Reveals Cytosolic Phospholipase A₂ as a Regulator of Mitochondrial Bioenergetics during Activation.** *Cell Metab.* 2016; **23**(5): 930–944.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Quehenberger O, Armando AM, Brown AH, *et al.*: **Lipidomics reveals a remarkable diversity of lipids in human plasma.** *J Lipid Res.* 2010; **51**(11): 3299–3305.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hartler J, Trötzlmüller M, Chittraju C, *et al.*: **Lipid data analyzer: Unattended identification and quantitation of lipids in LC-MS data.** *Bioinformatics.* 2011; **27**(4): 572–577.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Ni Z, Angelidou G, Lange M, *et al.*: **LipidHunter Identifies Phospholipids by High-Throughput Processing of LC-MS and Shotgun Lipidomics Datasets.** *Anal Chem.* 2017; **89**(17): 8800–8807.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Tsugawa H, Cajka T, Kind T, *et al.*: **MS-DIAL: Data Independent MS/MS Deconvolution for Comprehensive Metabolome Analysis.** *Nat Methods.* HHS Public Access. 2015; **12**(6): 523–526.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Tautenhahn R, Patti GJ, Rinehart D, *et al.*: **XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data.** *Anal Chem.* 2012; **84**(11): 5035–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Tsugawa H, Ikeda K, Takahashi M, *et al.*: **A lipidome atlas in MS-DIAL 4.** *Nat Biotechnol.* 2020; **38**(10): 1159–1163.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Spener F, Lagarde M, Gélouin A, *et al.*: **What is lipidomics?** *Eur J Lipid Sci Technol.*

- 2003; **105**(9): 481–482.
[PubMed Abstract](#) | [Publisher Full Text](#)
9. Fahy E, Subramaniam S, Murphy RC, *et al.*: **Update of the LIPID MAPS comprehensive classification system for lipids.** *J Lipid Res.* 2009; **50** Suppl(Suppl): S9–14.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 10. Nguyen A, Rudge SA, Zhang Q, *et al.*: **Using lipidomics analysis to determine signalling and metabolic changes in cells.** *Curr Opin Biotechnol.* 2017; **43**: 96–103.
[PubMed Abstract](#) | [Publisher Full Text](#)
 11. Nguyen A, Guedán A, Mousnier A, *et al.*: **Host lipidome analysis during rhinovirus replication in HBECs identifies potential therapeutic target.** *J Lipid Res.* 2018; **59**(9): 1671–1684.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 12. Hahn O, Drews LF, Nguyen A, *et al.*: **A nutritional memory effect counteracts benefits of dietary restriction in old mice.** *Nat Metab.* 2019; **1**(11): 1059–1073.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 13. Franz M, Lopes CT, Huck G, *et al.*: **Cytoscape.js: A graph theory library for visualisation and analysis.** *Bioinformatics.* 2016; **32**(2): 309–311.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 14. Liebisch G, Vizcaino JA, Köfeler H, *et al.*: **Shorthand notation for lipid structures derived from mass spectrometry.** *J Lipid Res.* 2013; **54**(6): 1523–30.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 15. Ni Z, Fedorova M: **LipidLynxX: a data transfer hub to support integration of large scale lipidomics datasets.** *bioRxiv.* 2020; 2020.04.09.033894.
[Publisher Full Text](#)
 16. Fahy E, Subramaniam S, Alex Brown H, *et al.*: **A comprehensive classification system for lipids.** *J Lipid Res.* 2005; **46**(5): 839–861.
[PubMed Abstract](#) | [Publisher Full Text](#)
 17. Ando A, Oka M, Satomi Y: **Deoxysphingolipids and ether-linked diacylglycerols accumulate in the tissues of aged mice.** *Cell Biosci.* 2019; **9**: 61.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 18. Gaud C, Sousa BC, Nguyen A, *et al.*: **BioPAN: a web-based tool to explore mammalian lipidome metabolic pathways on LIPID MAPS.** 2020.
<http://www.doi.org/10.17605/OSF.IO/PKD6B>

Open Peer Review

Current Peer Review Status:   

Version 1

Reviewer Report 09 March 2021

<https://doi.org/10.5256/f1000research.30993.r76705>

© 2021 Milne G. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Ginger Milne

¹ Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

² Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

The submitted manuscript explains a web-based tool on the Lipid Maps website that helps the user to follow biosynthetic and metabolic pathways of endogenous lipids (the lipidome) following mass spectrometric analysis. This program will help the user link lipidomics data with underlying endogenous processes. The authors have taken care to cite appropriate references and give examples of how to use this tool. Further, the tool itself has a link to an excellent web-based tutorial.

Recommendation:

1. The introduction ends abruptly. It reads as if there is a paragraph missing emphasizing the importance and expected use of this tool.

Is the rationale for developing the new software tool clearly explained?

Partly

Is the description of the software tool technically sound?

Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Lipidomics, mass spectrometry, oxidized lipid mediators, lipid metabolism.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 19 May 2021

Andrea F. Lopez-Clavijo

We thank you the reviewer Ginger Milne for her comments.

Recommendation:

- The introduction ends abruptly. It reads as if there is a paragraph missing emphasizing the importance and expected use of this tool.

--> A sentence was added: "So, BioPAN aids for automatic integration of lipid metabolism with lipid profiles, finding strong relationships between lipid substrates and lipid products catalysed by active or suppressed enzymes."

Competing Interests: None

Reviewer Report 29 January 2021

<https://doi.org/10.5256/f1000research.30993.r76703>

© 2021 Whitfield P. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Phillip Whitfield

¹ Glasgow Polyomics and Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK

² Glasgow Polyomics and Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK

This manuscript by Gaud *et al.* is focused on BioPAN, a software tool designed for the mapping of lipid networks in mammalian systems. In order to demonstrate the utility of the software an exemplar analysis is outlined using publicly available lipidomic data sets acquired from the cerebral cortex and liver of young and aged mice. This is an interesting article and the BioPAN software will be useful in visualisation outputs from lipidomic experiments. There are a few minor suggestions:

Introduction: A sentence could be added to the end of the introductory text summarising the overall objective of the presented work.

Methods: It would be useful to clarify whether $p < 0.05$ is always equivalent to $Z\text{-score} > 1.645$.

Conclusion: The authors may wish to indicate how the software could be further developed.

Is the rationale for developing the new software tool clearly explained?

Partly

Is the description of the software tool technically sound?

Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Lipidomics.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 19 May 2021

Andrea F. Lopez-Clavijo

We thank you reviewer Phillip Whitfield for his comments.

There are a few minor suggestions:

- *Introduction: A sentence could be added to the end of the introductory text summarising the overall objective of the presented work.*

--> A sentence was added: "So, BioPAN aids for automatic integration of lipid metabolism with lipid profiles, finding strong relationships between lipid substrates and lipid products catalysed by active or suppressed enzymes."

- *Methods: It would be useful to clarify whether $p < 0.05$ is always equivalent to $Z\text{-score} > 1.645$.*

--> It was added: "...corresponding to $Z > 1.645$)."

- *Conclusion: The authors may wish to indicate how the software could be further developed.*

--> A sentence was added: "Further work into the BioPAN database will be carried out to incorporate new biosynthetic pathways, such as: cholesterol, eicosanoids, and oxidised lipid subclasses, among others. Addition of glucosyl and galactosyl ceramides (Glc/Gal-Cer) to complement to the existing sphingolipids pathway in BioPAN will also be considered."

Competing Interests: None

Reviewer Report 26 January 2021

<https://doi.org/10.5256/f1000research.30993.r76702>

© 2021 Cazenave Gassiot A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Amaury Cazenave Gassiot 

¹ Department of Biochemistry, Yong Loo Lin School of Medicine, Singapore, Singapore

² Department of Biochemistry, Yong Loo Lin School of Medicine, Singapore, Singapore

The article entitled "BioPAN: a web-based tool to explore mammalian lipidome metabolic pathways on LIPID MAPS" described an online tool for visualisation of biological pathways from quantitative lipidomics data. The tool is an important contribution to lipidomics data analysis pipeline that could help make sense of ever more complex dataset.

General comments

- Well-written manuscript and an interesting contribution for those studying quantitative lipidomics datasets.
- Recommendation is to accept with minor revision and, if possible, implementation of a few suggestions.

Suggestions

1. With regards to the statistical model used by the software. Could the authors elaborate on whether correction for multiple testing is done for p-value calculations? In the "Pathway calculation" section, would it be possible to let the user input any value for the significance threshold, rather than choosing from a few drop-down options?
2. The "Filter" field in the "Pathway options" section is very useful; however, it seems to take only one value at a time. Would it be possible to include logical operators? For instance, one could filter for "PC 32:1 AND PE 32:1", or "38:5 OR 38:6".
3. Ultimately, the tool relies on information contained in the pathway database. While using an in-house dataset for testing the tool, a large number of species were unfortunately "unprocessed". Could the authors suggest how community users could help enriching the

database?

4. The export function is useful, but only allows for exports of .png and .jpg graphics, the resolution of which seems to depend on the zoom status in the display window. Would it be possible to add an option to export as vector graphics, .svg for example?

Minor comments

1. Page 1, Abstract, line 3: change "increase" to "increases".
2. Page 2, Authors roles, line 1: change "visualization" to "visualisation".
3. Page 5, right column, 3rd paragraph, line 3-4: change "visualized" to "visualised".
4. Page 9, Acknowledgements: change "In memorium" to "In memoriam".

Is the rationale for developing the new software tool clearly explained?

Yes

Is the description of the software tool technically sound?

Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Lipidomics.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 19 May 2021

Andrea F. Lopez-Clavijo

We thank you reviewer Amaury Cazenave Gassiot for his comments.

Suggestions:

- *With regards to the statistical model used by the software. Could the authors elaborate on*

whether correction for multiple testing is done for p-value calculations?

--> BioPAN does not correct for multiple testing as it calculates a Z-score to find stronger relationships within the data, hence higher Z-scores represent a more interesting result than others.

- *In the "Pathway calculation" section, would it be possible to let the user input any value for the significance threshold, rather than choosing from a few drop-down options?*

--> The drop-down menu has been replaced to a text box and it has been incorporated into BioPAN.

- *The "Filter" field in the "Pathway options" section is very useful; however, it seems to take only one value at a time. Would it be possible to include logical operators? For instance, one could filter for "PC 32:1 AND PE 32:1", or "38:5 OR 38:6".*

--> A paragraph was added in the Display options section: "The search box allows the user to search for one or two queries using the logical operators AND / OR. Selecting the AND operator, the user can view lipid subclasses or molecular species for which the name implies both queries. For example, searching for "PC" and "34:0" on the lipid molecular species graph displays lipids "PC(34:0)" and "O-PC(34:0)". Choosing the OR operator allow the user to visualise lipids whose names includes one of the two searches. For instance, searching "LPC" or "LPA" on the subclasses graph displays lipids "LPC", "O-LPC" and "LPA".".

- *Ultimately, the tool relies on information contained in the pathway database. While using an in-house dataset for testing the tool, a large number of species were unfortunately "unprocessed". Could the authors suggest how community users could help enriching the database?*

--> A paragraph was added in the operation section to address the "unprocessed" molecular species within BioPAN: "The unprocessed category corresponds to lipid molecular species that did not have a matching molecular species, as a reactant and product, with the same number of carbons and double bonds. For example, a substrate molecular species, such as PA 34:2 should have a matching product with the same sum composition (e.g. DG 34:2). Additionally, some biological reactions require fatty acid or fatty-acyl-CoA molecular species. So, a lack of these subclasses in the input file prevents BioPAN to build up the reaction between substrates and products. If the substrate or product is not part of other reaction within the pathway, BioPAN will report it as unprocessed, such is the case of triacylglycerides (TG) subclasses. On other instances, like phosphatidylcholine (PC) and lysophosphatidylcholine (LPC), where PC is the substrate in the reaction to produce phosphatidic acid (PA) and LPC is the substrate to produce lysophosphatidic acid (LPA), those lipid subclasses molecular species can be considered as processed by BioPAN.".

- *The export function is useful, but only allows for exports of .png and .jpg graphics, the resolution of which seems to depend on the zoom status in the display window. Would it be possible to add an option to export as vector graphics, .svg for example?*

--> This capability of export graphics in SVG has been added and a sentence was included in the text "in several formats (JPEG / PNG / **SVG** / JSON / TXT)".

Minor comments

1. Page 1, Abstract, line 3: change "increase" to "increases".

Done, thank you

2. Page 2, Authors roles, line 1: change "visualization" to "visualisation".

Done, thank you

3. Page 5, right column, 3rd paragraph, line 3-4: change “visualized” to “visualised”.
Done, thank you

4. Page 9, Acknowledgements: change “In memorium” to “In memoriam”.
Done, thank you

Competing Interests: None

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research