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.
a a h h b r b h h r b b h ba h h a h

How to cite this article: Gaud C, C. Sousa B, Nguyen A BioPAN: a web-based tool to explore mammalian lipidome metabolic pathways on LIPID MAPS [version 1; peer review: 2 approved] F1000Research 2021, 10:4 https://doi.org/10.12688/f1000research.28022.1

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

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 To F Liquid chromatography-MS, either tandem or high resolving power, interfaced with processing pipelines such as Lipid Data Analyser (LDA), LipidF inder, 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 continue to grow it has become considerably more challenging to interpret the complex sets of changes within the lipidome[8]. 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 subclasses[9]. F or 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

-

classes. 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 (CD F) to give a global pathway Z-score The set of equations used to calculate Z-scores are presented and discussed elsewhere[10]. 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 $p < 0.05$ ($Z > 1.6\ \ 5$).

Operation

tion, double bond location, and stereochemistry into a sum experiment assuming a normally distributed data of lipid sub

The first step in BioPAN is to load a file of quantitative lipidomics data. After uploading, BioPAN uses the LipidLynxX[15] 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)[9,1 ,16]. BioPAN then classifies the submitted lipids 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. BioPAN's database was manually collated followed by a validation using available literature. The database currently contains 9 lipid reactions identified in mammals, covering 1 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. ollowing analysis completion, the user is directed to the results page (

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

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, PTDSS1, 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,

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 (A) pathways at the molecular species level with BioPAN (igure 3) shows the A network obtained

identification and quantitation of lipids in LC-MS data. Bioinformatics. 2011; (4): 572–577.

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 1

Reviewer Report 29 January 2021

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

Phillip Whitfield
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-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.

Reviewer Report 26 January 2021

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

Amaury Cazenave Gassiot
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.