Mapping metabolomics data: Complexity and issues

Mathieu Umec, Marie Lefebvre, Ghina Hajjar, Franck Giacomoni, Blandine Comte, Estelle Pujos-Guillot

Mapping metabolomics data: Complexity and issues

Mathieu UMEC, Marie Lefebvre, Ghina Hajjar, Franck Giacomoni , Blandine Comte, Estelle Pujos-Guillot

Université Clermont Auvergne, INRAE, UNH, Plateforme d'Exploration du Métabolisme, MetaboHUB Clermont, Clermont-Ferrand, France

CONTEXT

To optimize the translation of large-scale metabolomics by defining meaningful results, data contextualization is mandatory. **Although a number of tools and methods have been developed, there is still no standardization of practices**. In this context, the objective of the work was to evaluate pathway analysis to biologically contextualize metabolomics data, identify bottlenecks and optimize workflows to provide reproducible information able to guide biological interpretation.

MATERIAL AND METHODS

Different alternative tools using neither the same methods nor the same databases were first evaluated. In this analysis, four different tools, namely **ConsensusPAthDB, MetaboAnalyst, MetExplore and RaMP**, all enabling metabolite mapping using the **Over-Representation Analysis (ORA)** method, were used. This is the easiest and most commonly used method, as it requires only a list of metabolic pathways, a list of identified metabolites and statistical analysis.

To fulfil our comparison objective, a published dataset including a list of identified metabolites modulated with metabolic syndrome in elderly men was used (Comte *et al.*, 2021). This list contained **119 metabolites with a CheBI identifier, including 110 unique ones**. In order to assess the influence of using one type of identifier rather than another, we selected all metabolites with **ChEBI, HMDB and KEGG identifiers**, resulting in a set of only 17 metabolites.

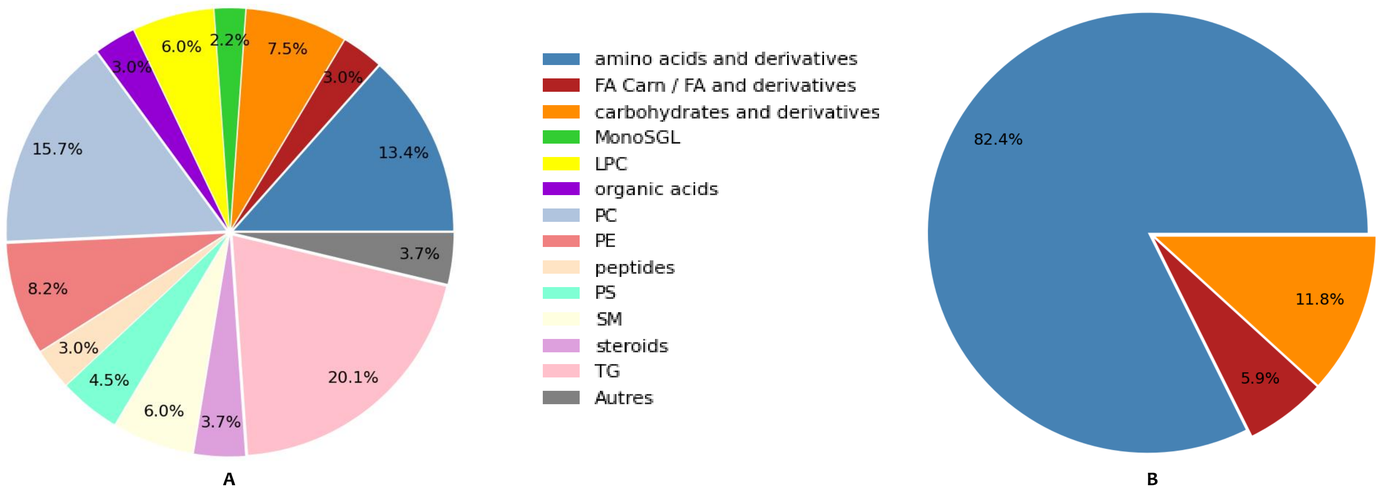

Figure 1: *Chemical characterization of the use case list* ***(Figure A)*** *and sub-dataset presenting all KEGG, HMDB and* ChEBI *ID* ***(Figure B)***

Before mapping: Definition of Pathways

Depending on the tools used (databases or network), pathways with the **same name may not contain the same metabolites**, a pathway name can sometimes contain only a few metabolites, and in other database a hundred metabolites. Therefore, it is not particularly meaningful to exactly compare pathway analyses across different databases unless the results undergo thorough manual curation. However, **using multiple pathway databases was found to be a good strategy to derive a consensus pathway signature and increase the metabolome coverage**

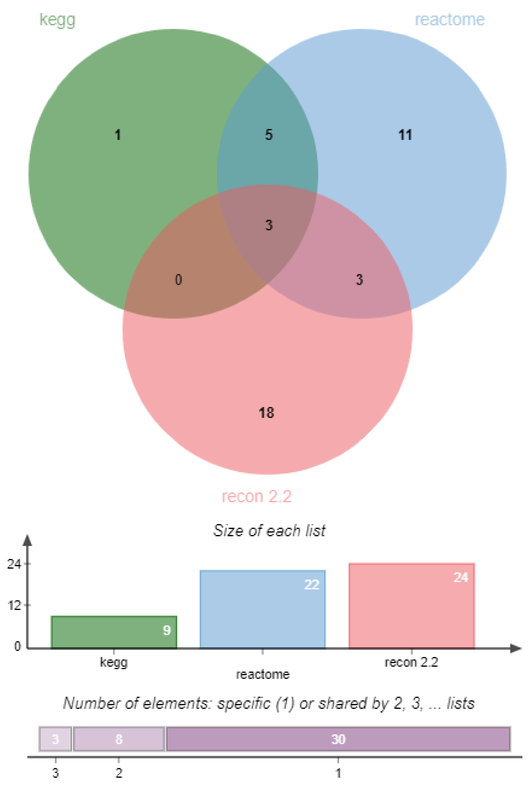

Figure 2: *Venn chart of Urea Cycle pathways defined on RECON 2.2 Network, KEGG (version 110) and Reactome Database (version 88)* **(Bardou *et al.*, BMC Bioinformatics 2014, 15:293 doi:10.1186/1471-2105-15-293)**

Database or Network?

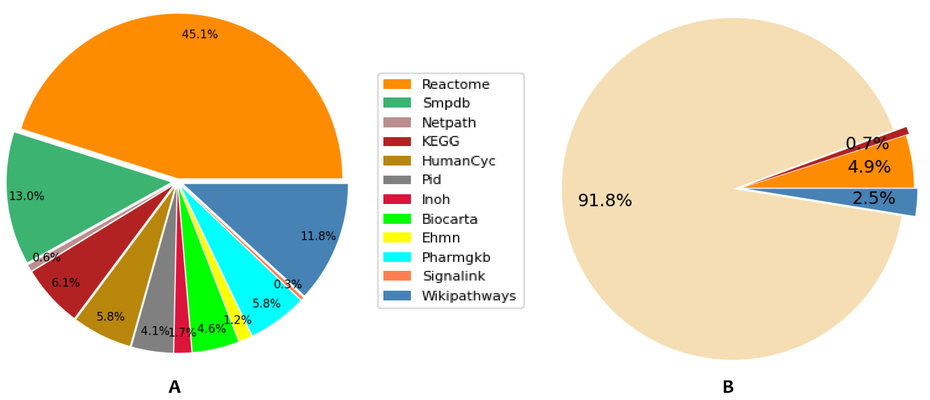

Figure 3: *Pathways origin for ConsensusPathDB (5 578 pathways)* ***(Figure A)*** *and RaMP (54 024 pathways)* ***(Figure B)***

ConsensusPathDb and **RaMP** allowed to **simultaneously map on all the available databases,** unlike **MetaboAnalyst** and **MetExplore**, which **require the user to select the biological source. MetaboAnalyst offers only 4 choices**: the smallest resource is a set of 80 metabolite sets from KEGG, while the largest includes 3 694 from RaMP. **MetExplore offers a wide variety of metabolic networks** from different species, and also allows adding other networks for analysis.

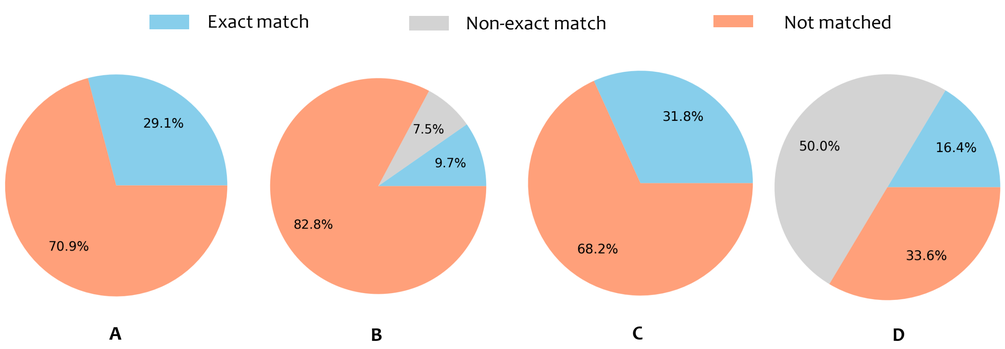

Figure 4 : *Percentage coverage when mapping to ConsensusPathDB,* ***(Figure A)*** *MetaboAnalyst,* ***(Figure B)*** *,RaMP* ***(Figure C)*** *and MetExplore Recon 2.2* ***(Figure D)***

For improving the amount of metabolites mapped, the **ChEBI ontology** was used in order to add the identifiers of associated metabolites. For fully identified metabolites, the use of metabolic network and subnetwork extraction appeared to be more pertinent to go deeper into metabolic exploration.

Mapping results, visualization and interpretation

	ChEBI (17)[a]	KEGG (17)[a]	HMDB (17)[a]	ChEBI (110)	KEGG (31)	HMDB (38)
ConsensusPathDB (All)	325	253	272	326	276	334
MetaboAnalyst (KEGG)	28	28	28	31	31	31
MetExplore (Recon 2.2)	44	NA	NA[b]	57	NA[b]	NA[b]
RaMP (All)	910	883	910	311	293	282

[a] The sub-dataset is used

[b] These identifiers type can not be use in this tool

Figure 6: *Number of metabolic pathways obtained when mapping with different datasets on the various tools and according to the type of identifier used. The number in brackets is the number of metabolites used.*

Using multiple pathway databases was found to be a good strategy to derive a **consensus pathway signature and increase the metabolome coverage.**

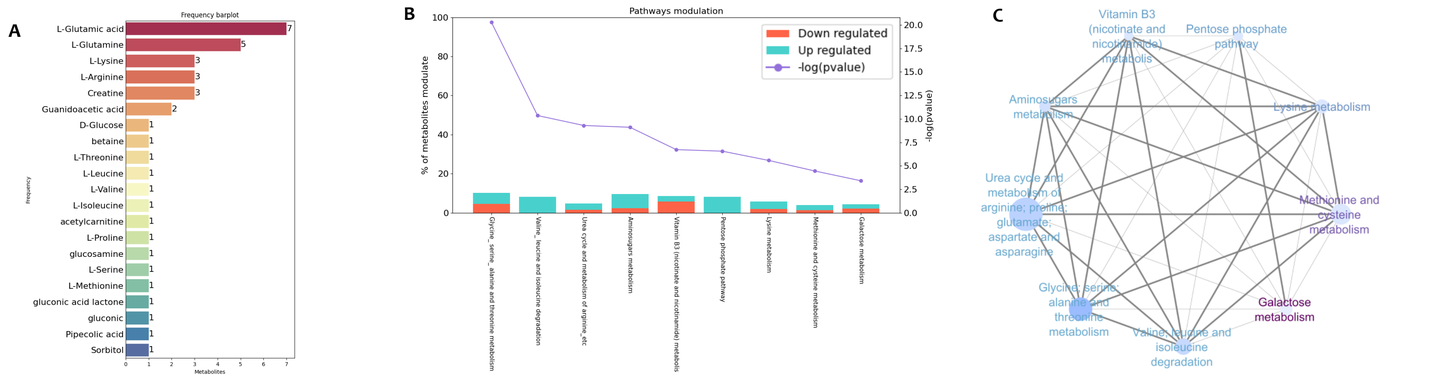

Figure 7: *Visualization of mapping results on EHMN database of ConsensusPathDB of frequencies of metabolites of interest* ***(Figure A),*** *Regulation plot of pathways* ***(Figure B),*** *and global networks seen at pathway level* ***(Figure C)***

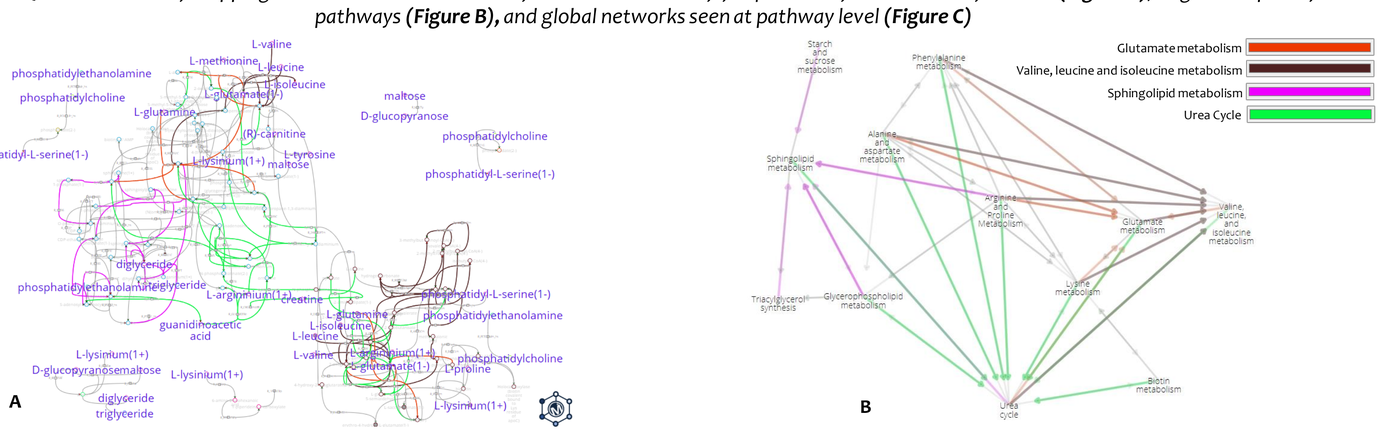

Figure 8: *Representation of mapping results in the form of subnetworks viewed at the metabolite level* ***(Figure A)*** *or at the metabolic pathway level* ***(Figure B)***.

For fully identified metabolites, the use of metabolic network and subnetwork extraction appeared to be more pertinent to go deeper into metabolic exploration.

TO CONCLUDE…

Because metabolite mapping opens up so many possibilities, it is critical to **modify and optimize workflows to fit the objective of the contextualization and the characteristics of the input data in accordance with the used tool**. Furthermore, it is important to have in mind the **limitations of relying just on one mapping tool.**

References: Comte *et al. EBioMedicine 2021 69:103440. , doi: 10.1016/j.ebiom.2021.103440*

Acknowledgments: This work is supported by the French Ministry of Research and National Research Agency as part of the French metabolomics and fluxomics infrastructure (MetaboHUB, ANR-11--INBS-0010).

Mathieu.umec@inrae.fr

Centre Auvergne Rhône Alpes
Site de Theix
63122 Saint Genès Champanelle