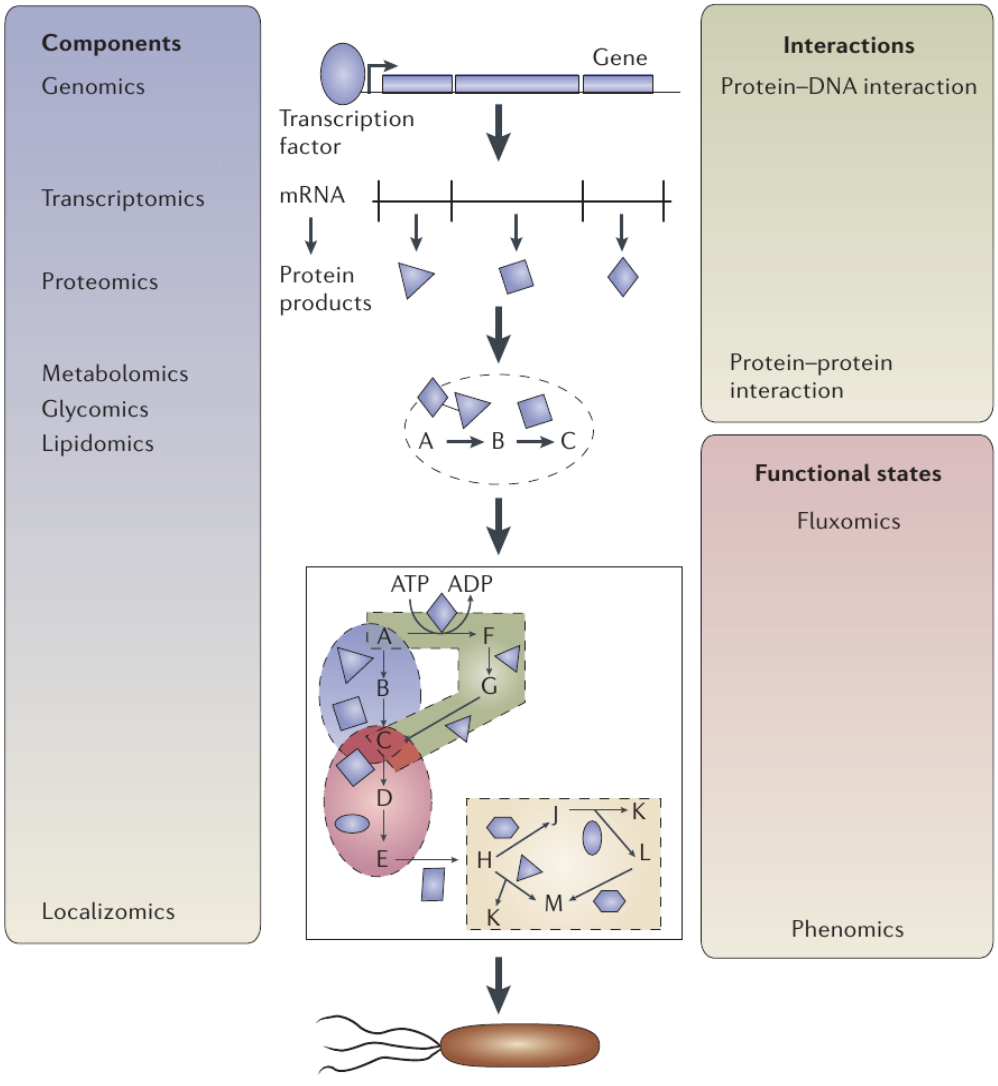
FELLA: an R package for network-based enrichment of metabolomics data

Sergi Picart-Armada
Dr. Alexandre Perera-Lluna

Bioinformatics and Biomedical Signals Laboratory
Polytechnic University of Catalonia

January 14, 2019

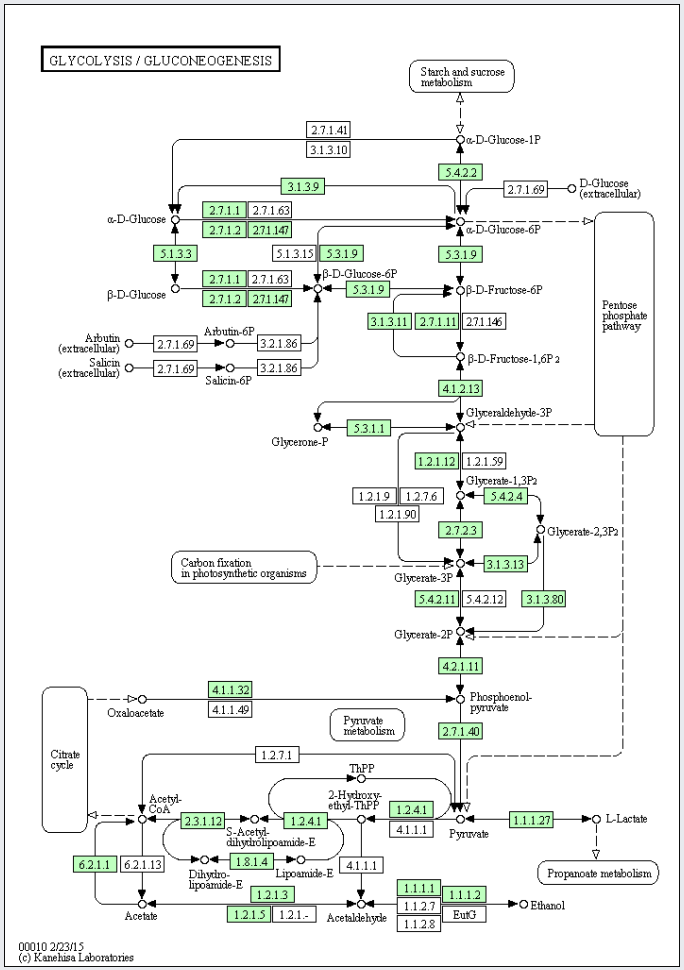
Omic sciences



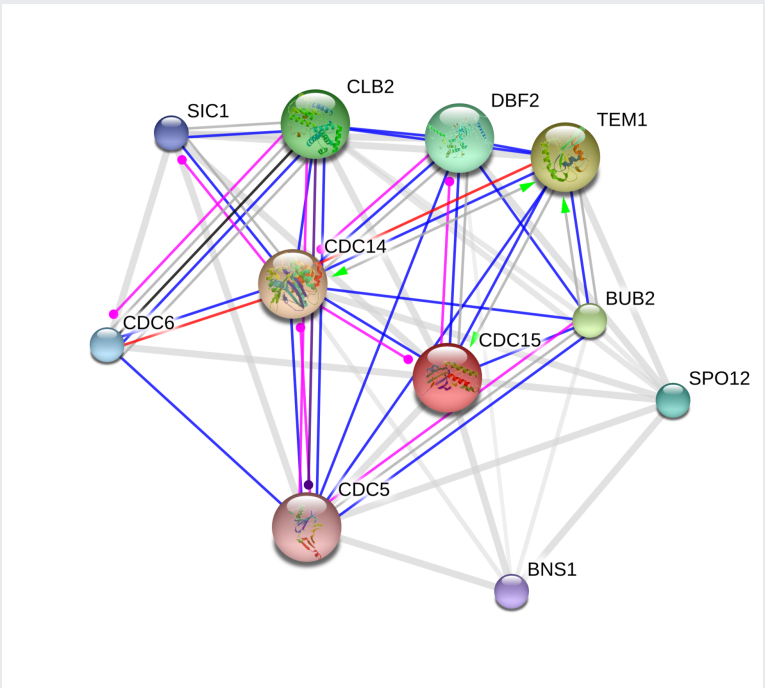
Omics overview [Joyce and Palsson, 2006].

Databases

Examples



Glycolysis pathway in KEGG [Kanehisa et al., 2008].



PPI network in STRING [Szklarczyk et al., 2015].

Enrichment techniques

Motivation

- In the early 2000's gene expression studies led to **large lists of differentially expressed genes**...
- How could we interpret them...?

Functional analysis

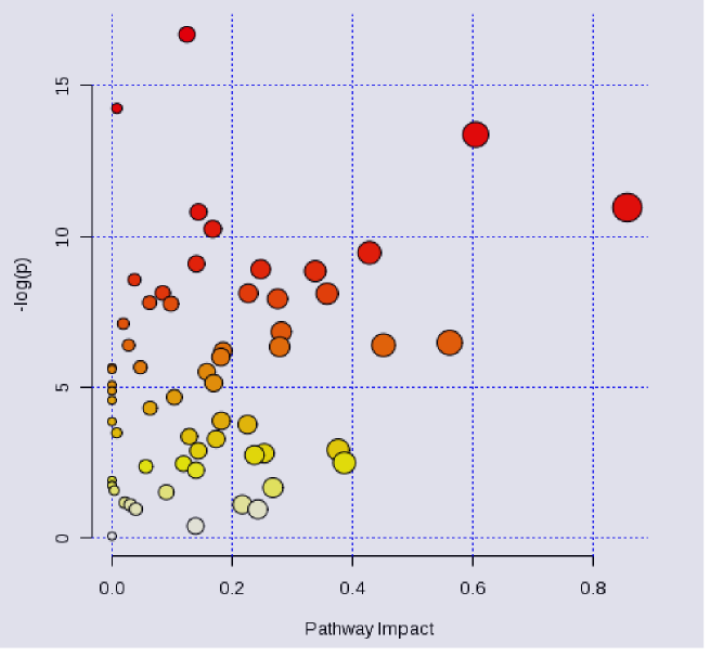
- ...by finding their most relevant **biological functions**.
- Takes advantage of the annotated **biological context**.
- **Over Representation Analysis**: statistical tests for pathways.

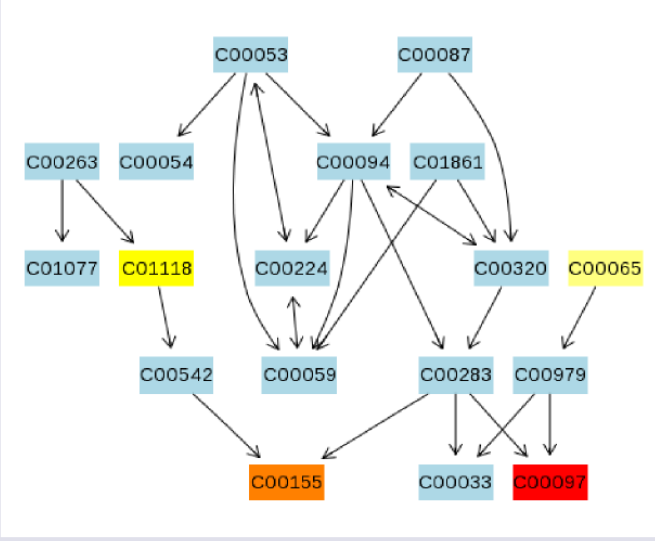
Application to other Omics

This approach was exported to understand lists of:

- Proteins (**Proteomics**)
- Metabolites (**Metabolomics**)

Current tools: MetaboAnalyst [Xia et al., 2012]

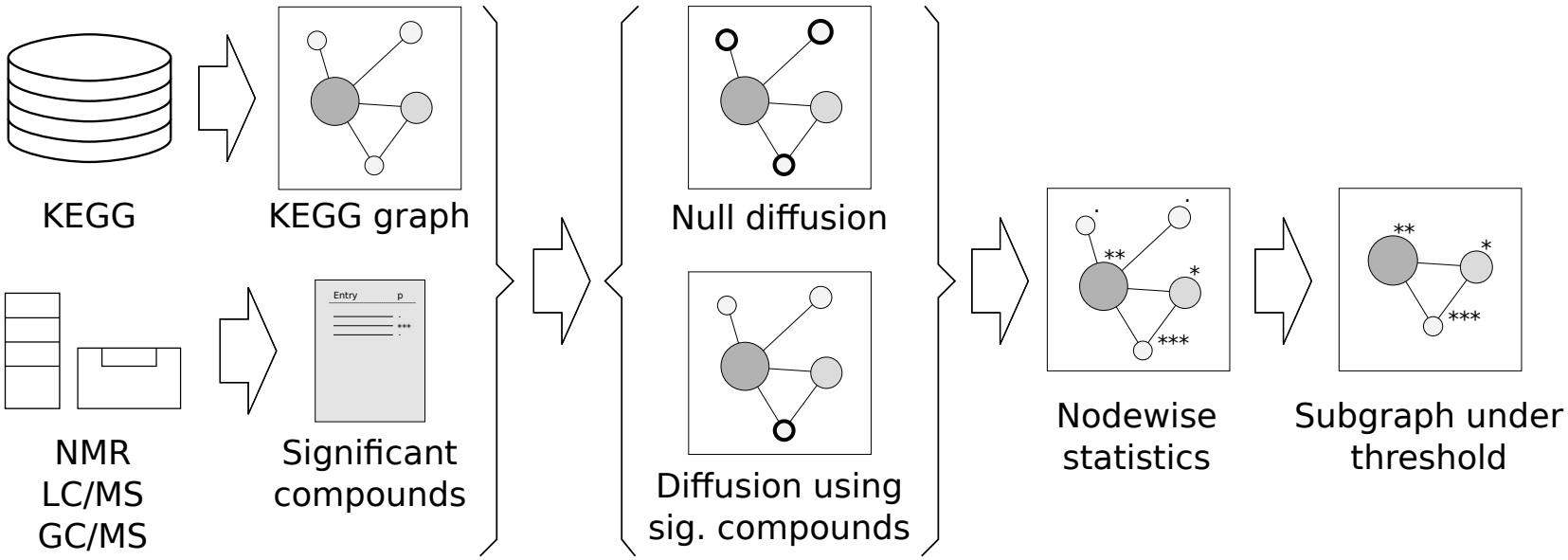




Sulfur metabolism

Metabolic pathway	FDR	Impact
Thiamine metabolism	3.51E-6	0.125
Nitrogen metabolism	2.03E-5	0.008
Cysteine and methionine metabolism	3.23E-5	0.605
Taurine and hypotaurine metabolism	2.52E-4	0.144
Lysine biosynthesis	3.68E-4	0.168
Glutathione metabolism	6.93E-4	0.428

Workflow in FELLA



Workflow to enrich a list of metabolites using FELLA [Picart-Armada et al., 2017].

Building a graph object from KEGG

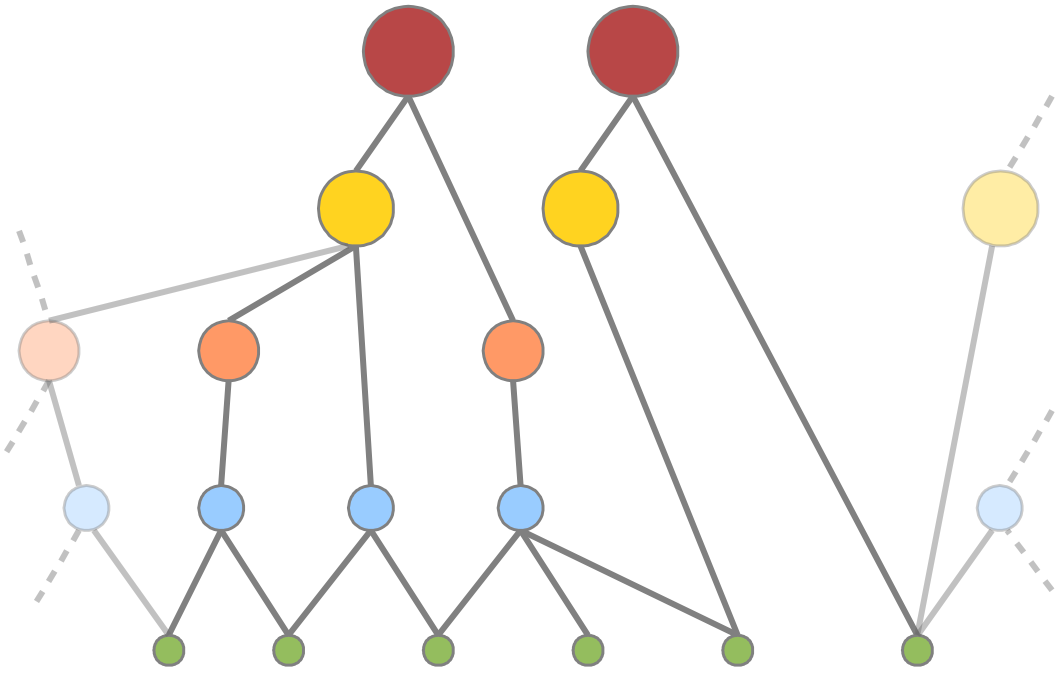
1 - Pathways

2 - Modules

3 - Enzymes

4 - Reactions

5 - Compounds

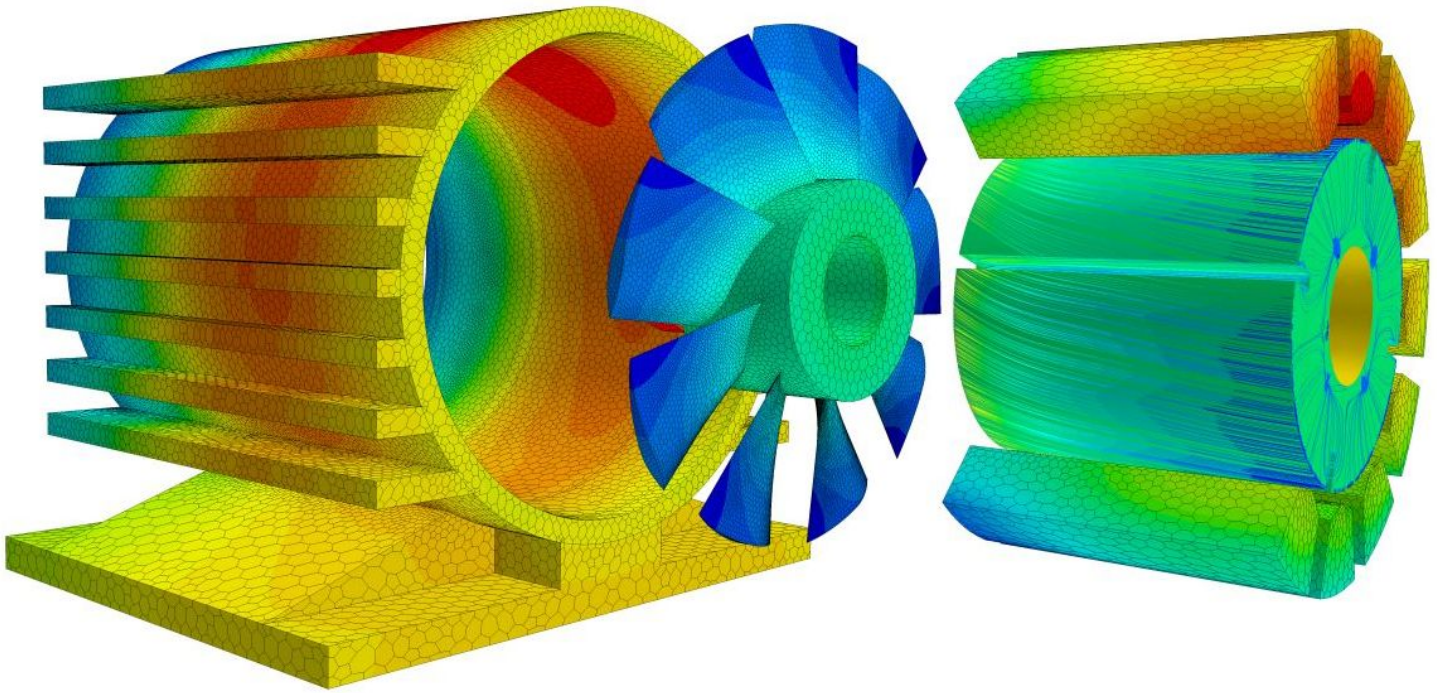


Knowledge graph in FELLA [Picart-Armada et al., 2017] is built from KEGG [Kanehisa et al., 2008].

Heat Diffusion (I)

Overview

KEGG graph is regarded as a meshed object on which we diffuse heat.

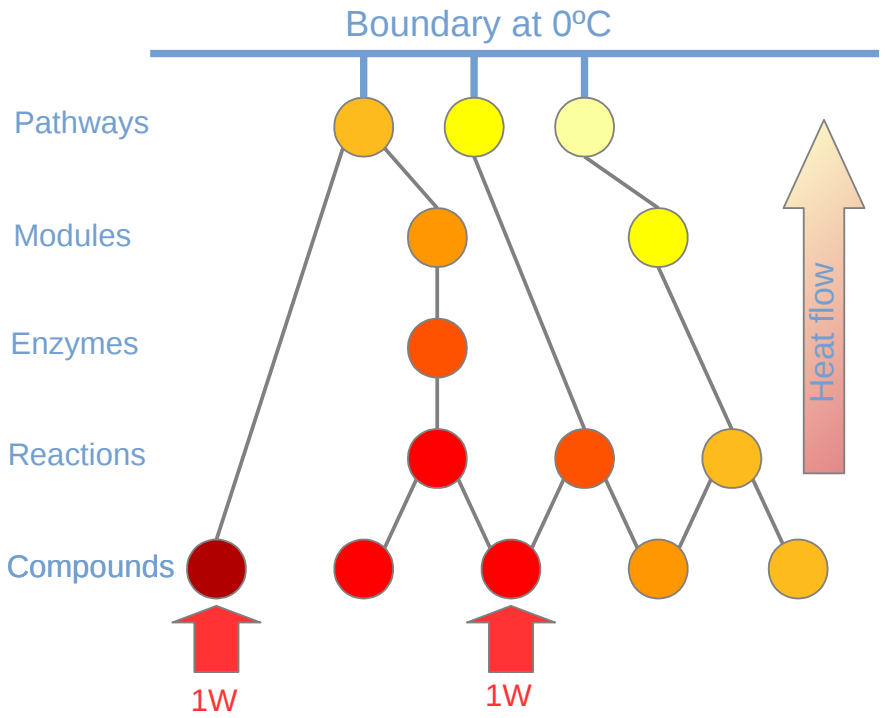


Heat diffusion in a meshed object. Image extracted from www.cd-adapco.com.

Heat Diffusion (II)

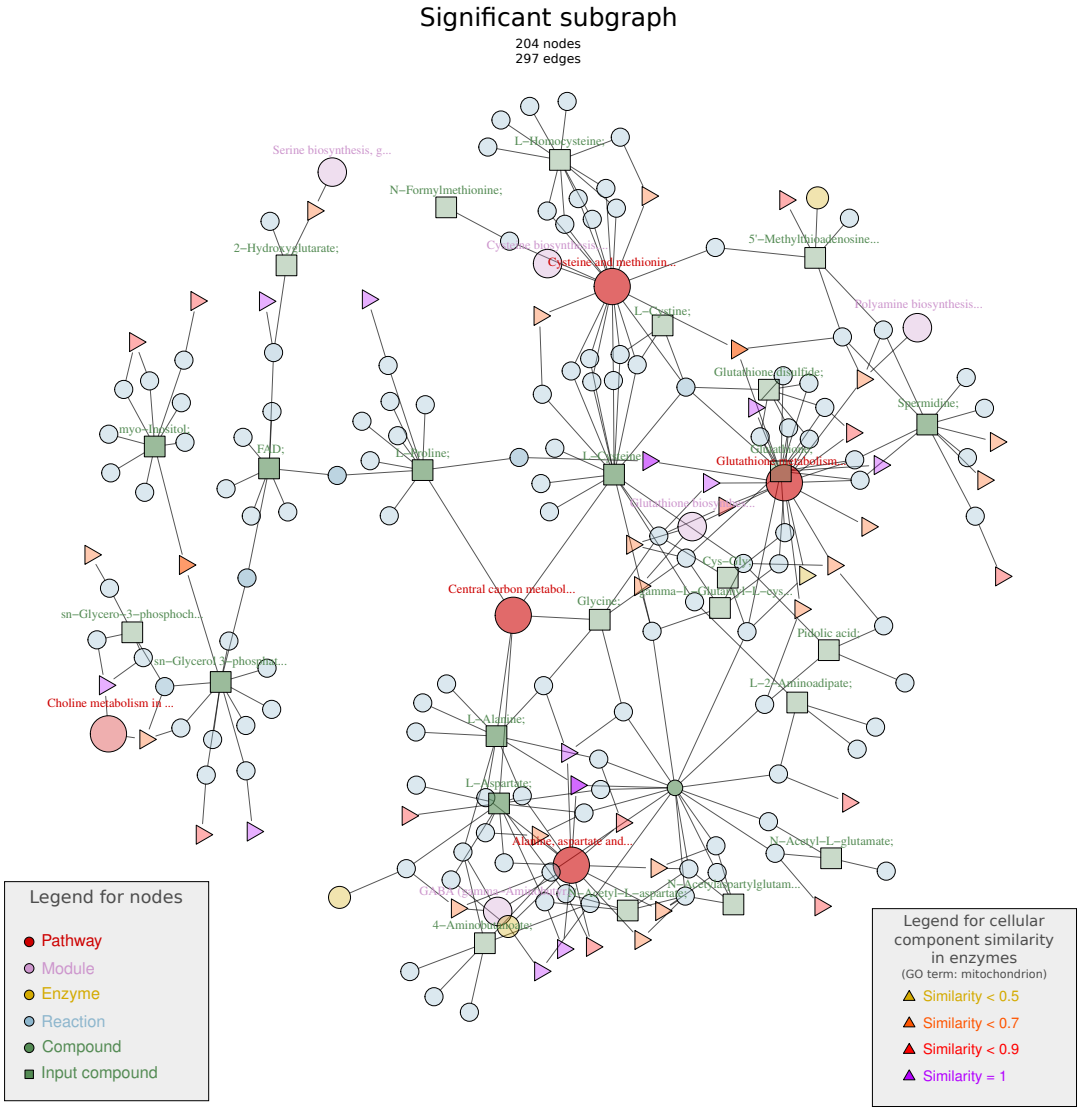
Intuitive idea

We force significant metabolites to be warm. Heat flows through metabolites, reactions, enzymes, modules and pathways. Afterwards, we select the *warmest* nodes.



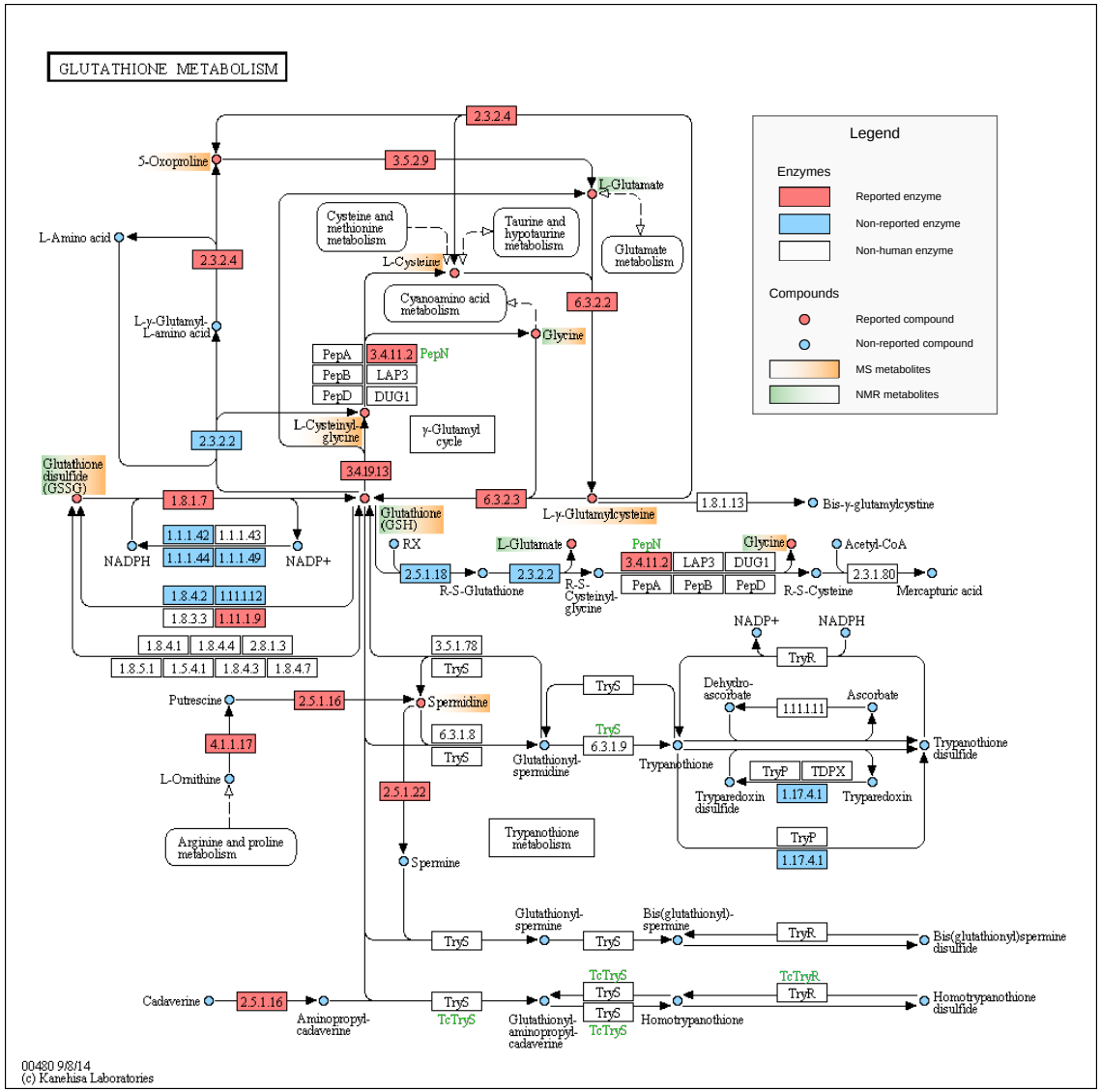
Diffusion in the KEGG graph. Affected metabolites introduce heat flow and pathways dispel it. Diffusion scores undergo a statistical normalisation afterwards [Picart-Armada et al., 2017].

Case study (I)



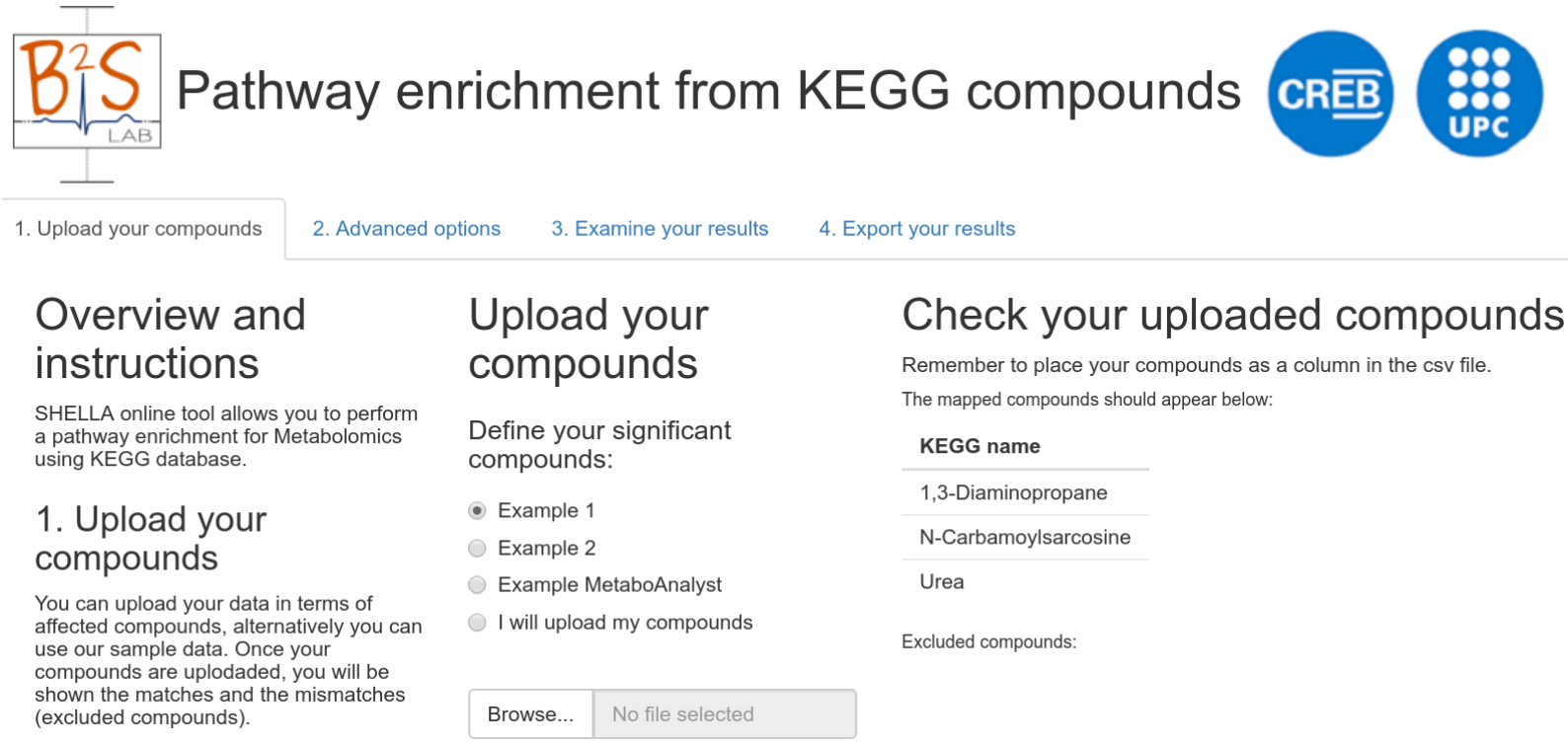
Diffusion applied to a metabolomics study. 38 metabolites were affected by knocking out an uncharacterised mitochondrial protein [Picart-Armada et al., 2017].

Case study (II)

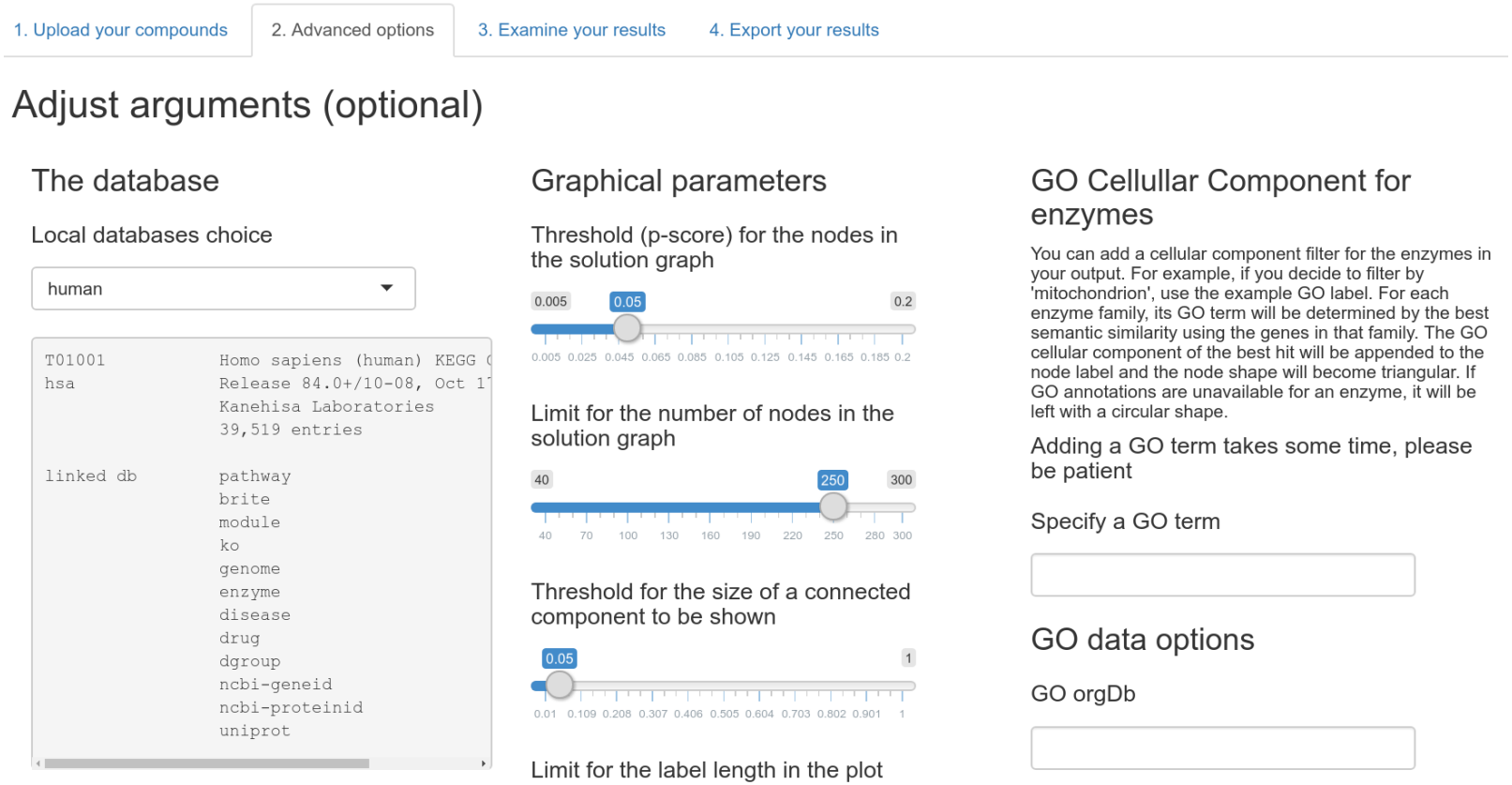


Relevant metabolites and enzymes, overlaid in the Glutathione metabolism pathway from KEGG [Picart-Armada et al., 2017].

Interactive app (I)



Interactive app (II)



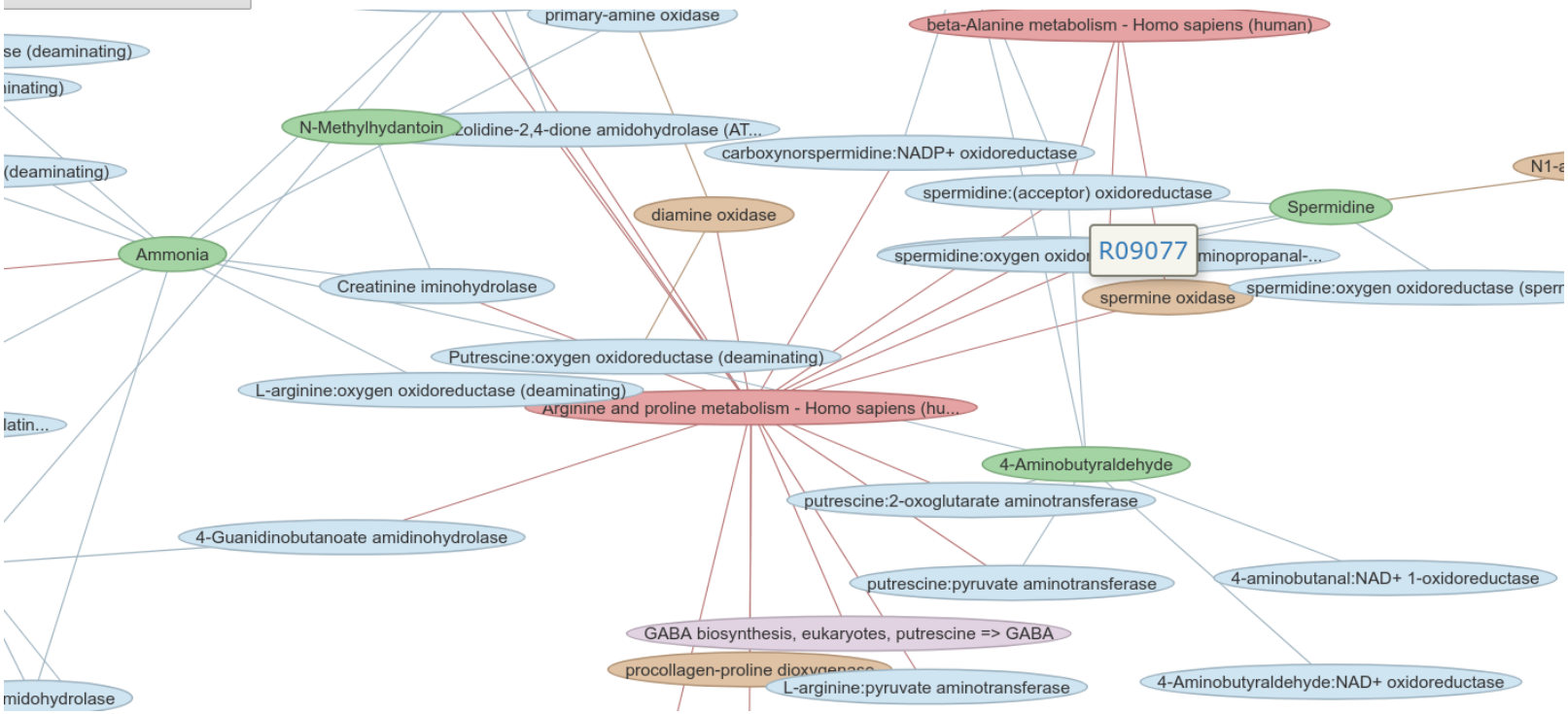
Interactive app (III)

1. Upload your compounds | 2. Advanced options | 3. Examine your results | 4. Export your results

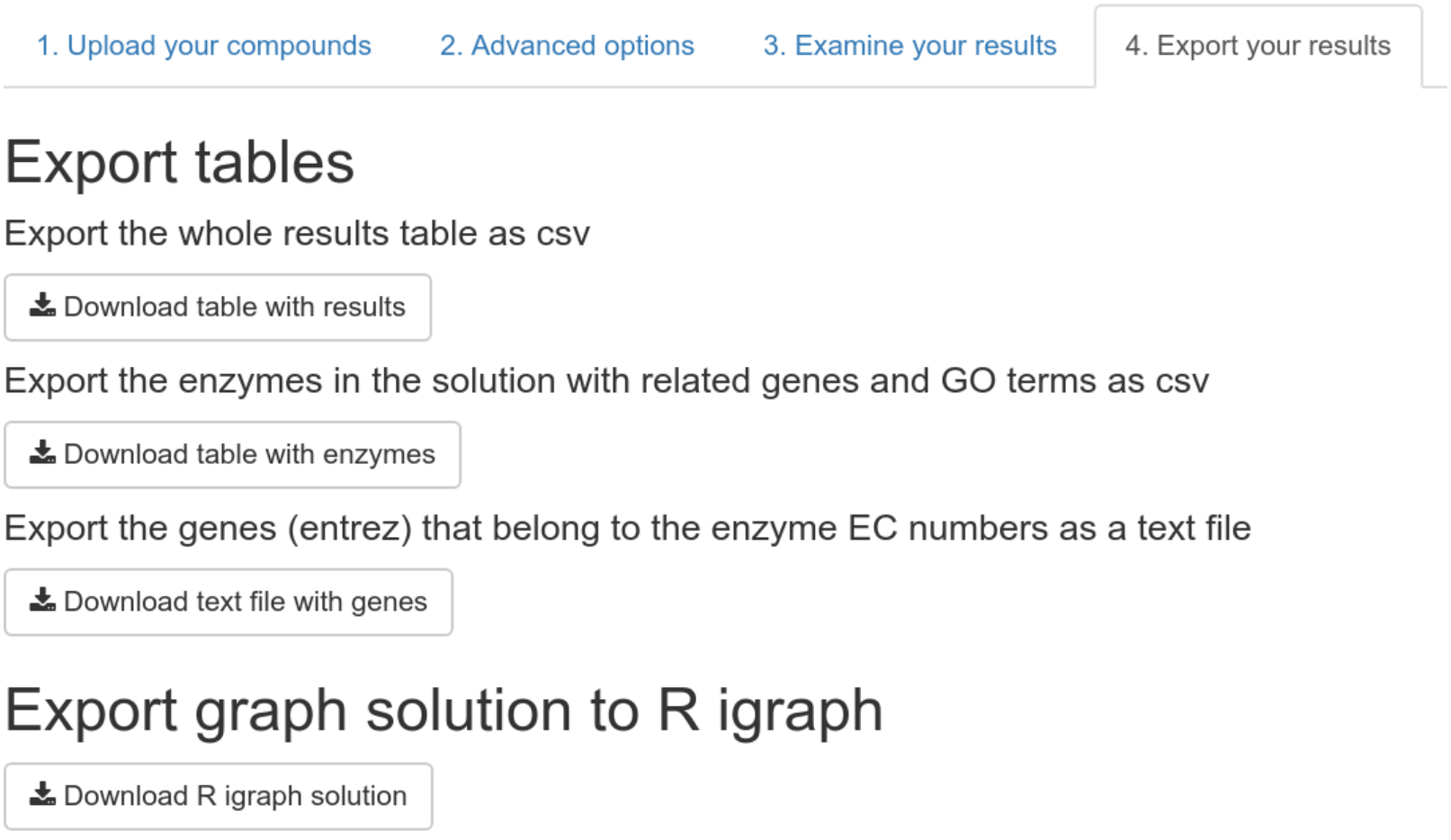
Number of nodes: 78

Select by id

Select by group



Interactive app (IV)



Availability

Bioconductor

FELLA is actively maintained and available in Bioconductor:
`https://doi.org/doi:10.18129/B9.bioc.FELLA`

Vignettes with toy and real examples

- Quick start: toy dataset to illustrate the essentials
- FELLA: main vignette, thorough description and three datasets
- Mus musculus study: example on an animal disease model
- Gilt-head bream: example on a non-human drug exposition study

Questions

`FELLA`'s status:
Algorithm: published [Picart-Armada et al., 2017]
R package: published [Picart-Armada et al., 2018]

Interactive app (small scale):
E-mail me and I will share the URL

Maintainer:
`sergi.picart@upc.edu`

FELLA: an R package for network-based enrichment of metabolomics data

Sergi Picart-Armada
Dr. Alexandre Perera-Lluna

Bioinformatics and Biomedical Signals Laboratory
Polytechnic University of Catalonia

January 14, 2019

References I


Joyce, A. R. and Palsson, B. Ø. (2006).
The model organism as a system: integrating'omics' data sets.
Nature Reviews Molecular Cell Biology, 7(3):198–210.


Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., and Yamanishi, Y. (2008).
KEGG for linking genomes to life and the environment.
Nucleic Acids Research, 36(Database-Issue):480–484.


Picart-Armada, S., Fernández-Albert, F., Vinaixa, M., Rodríguez, M. A., Aivio, S., Stracker, T. H., Yanes, O., and Perera-Lluna, A. (2017).
Null diffusion-based enrichment for metabolomics data.
PloS one, 12(12):e0189012.

References II

Picart-Armada, S., Fernández-Albert, F., Vinaixa, M., Yanes, O., and Perera-Lluna, A. (2018).
FELLA: an R package to enrich metabolomics data.
BMC bioinformatics, 19(1):538.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J., and von Mering, C. (2015).
STRING v10: protein-protein interaction networks, integrated over the tree of life.
Nucleic Acids Research, 43(Database-Issue):447–452.

Xia, J., Mandal, R., Sinelnikov, I., Broadhurst, D., and Wishart, D. S. (2012).
MetaboAnalyst 2.0 - a comprehensive server for metabolomic data analysis.
Nucleic Acids Research, 40(Web-Server-Issue):127–133.