

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

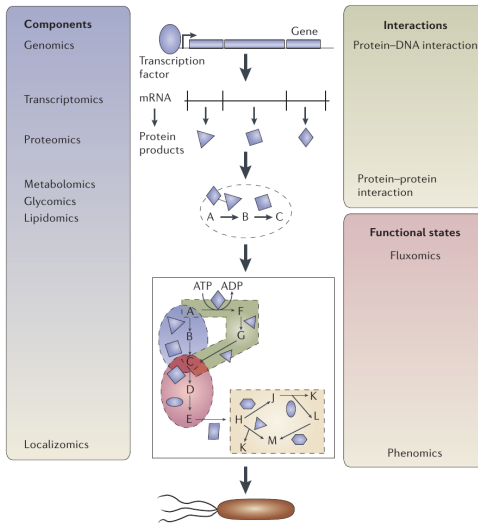
Sergi Picart-Armada
Dr. Alexandre Perera-Lluna

Bioinformatics and Biomedical Signals Laboratory
Polytechnic University of Catalonia

January 14, 2019



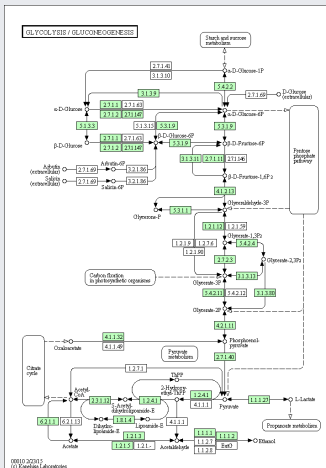
Omic sciences



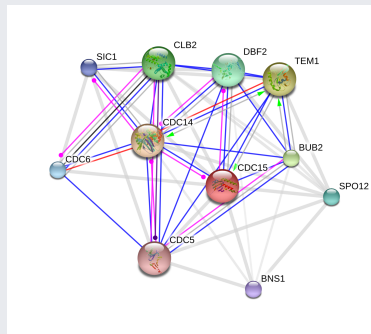
Omic overview [Joyce and Palsson, 2006].

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

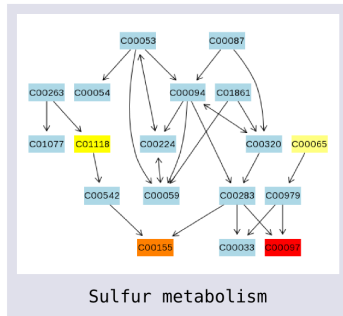
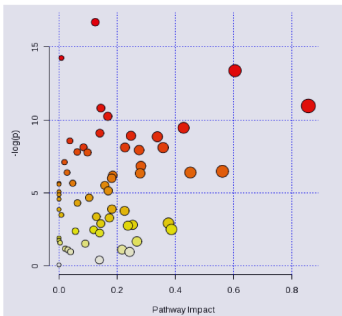
- ...by finding their most relevant **biological functions**.
- Takes advantage of the annotated **biological context**.
- **Over Representation Analysis**: statistical tests for pathways.

Application to other Omics

This approach was exported to understand lists of:

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

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



Metabolic pathway

Thiamine metabolism

Nitrogen metabolism

Cysteine and methionine metabolism

Taurine and hypotaurine metabolism

Lysine biosynthesis

Glutathione metabolism

FDR

3.51E-6

2.03E-5

3.23E-5

2.52E-4

3.68E-4

6.93E-4

Impact

0.125

0.008

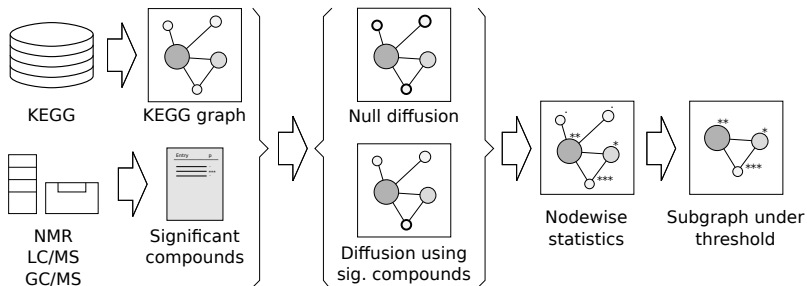
0.605

0.144

0.168

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

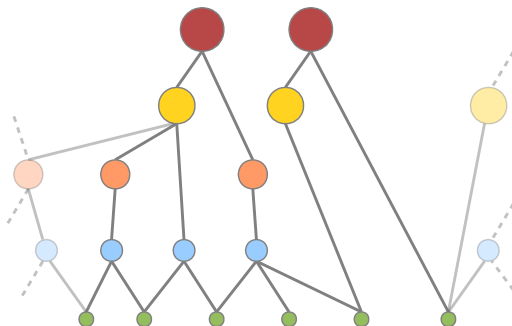
1 - Pathways

2 - Modules

3 - Enzymes

4 - Reactions

5 - Compounds

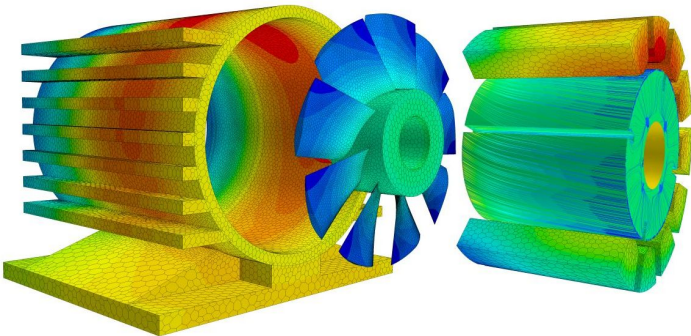


Knowledge graph in FELLA [Picart-Armada et al., 2017] is built from KEGG [Kanehisa et al., 2008].

Heat Diffusion (I)

Overview

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

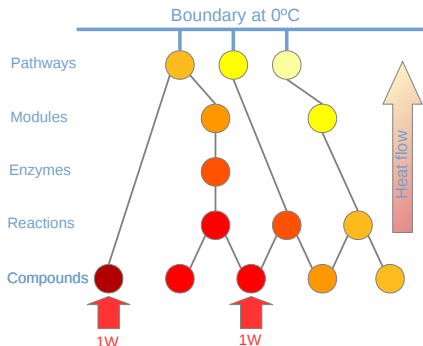


Heat diffusion in a meshed object. Image extracted from www.cd-adapco.com.

Heat Diffusion (II)

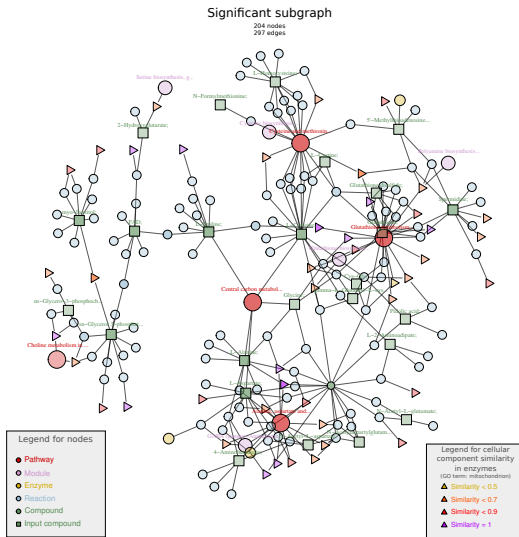
Intuitive idea

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].

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

Interactive app (I)



Pathway enrichment from KEGG compounds

[1. Upload your compounds](#)[2. Advanced options](#)[3. Examine your results](#)[4. Export your results](#)

Overview and instructions

SHELLA online tool allows you to perform a pathway enrichment for Metabolomics using KEGG database.

1. Upload your compounds

You can upload your data in terms of affected compounds, alternatively you can use our sample data. Once your compounds are uploaded, you will be shown the matches and the mismatches (excluded compounds).

Upload your compounds

Define your significant compounds:

- ☒ Example 1
- ☐ Example 2
- ☐ Example MetaboAnalyst
- ☐ I will upload my compounds

Check your uploaded compounds

Remember to place your compounds as a column in the csv file.

The mapped compounds should appear below:

KEGG name

1,3-Diaminopropane

N-Carbamoylsarcosine

Urea

Excluded compounds:

Interactive app (II)

[1. Upload your compounds](#)[2. Advanced options](#)[3. Examine your results](#)[4. Export your results](#)

Adjust arguments (optional)

The database

Local databases choice

human

T01001	Homo sapiens (human) KEGG C
hsa	Release 84.0+10-08, Oct 1
	Kanehisa Laboratories
	39,519 entries
linked db	pathway
	brite
	module
	ko
	genome
	enzyme
	disease
	drug
	dgroup
	ncbi-geneid
	ncbi-proteinid
	uniprot

Graphical parameters

Threshold (p-score) for the nodes in the solution graph



Limit for the number of nodes in the solution graph



Threshold for the size of a connected component to be shown



Limit for the label length in the plot

GO Cellular Component for enzymes

You can add a cellular component filter for the enzymes in your output. For example, if you decide to filter by 'mitochondrion', use the example GO label. For each enzyme family, its GO term will be determined by the best semantic similarity using the genes in that family. The GO cellular component of the best hit will be appended to the node label and the node shape will become triangular. If GO annotations are unavailable for an enzyme, it will be left with a circular shape.

Adding a GO term takes some time, please be patient

Specify a GO term

GO data options

GO orgDb

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 ▼

se (deaminating)

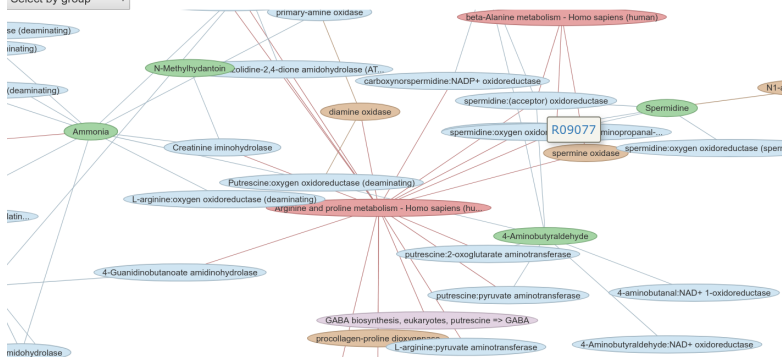
inating)

(deaminating)

Ammonia

latin...

midohydrolase



Interactive app (IV)

[1. Upload your compounds](#)[2. Advanced options](#)[3. Examine your results](#)[4. Export your results](#)

Export tables

Export the whole results table as csv

[!\[\]\(51514032c8ca341817228f39f1307b05_img.jpg\) Download table with results](#)

Export the enzymes in the solution with related genes and GO terms as csv

[!\[\]\(0d7ca0919e6c47bbd874bfa0189fe22e_img.jpg\) Download table with enzymes](#)

Export the genes (entrez) that belong to the enzyme EC numbers as a text file

[!\[\]\(f219cfc00b8db0cd1a81ae1fc9afaf28_img.jpg\) Download text file with genes](#)

Export graph solution to R igraph

[!\[\]\(683dba75afe26e28cd4de5730b776760_img.jpg\) Download R igraph solution](#)

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.