Functional analysis of omics data

Bioinformatics · Universitat Politècnica de Catalunya

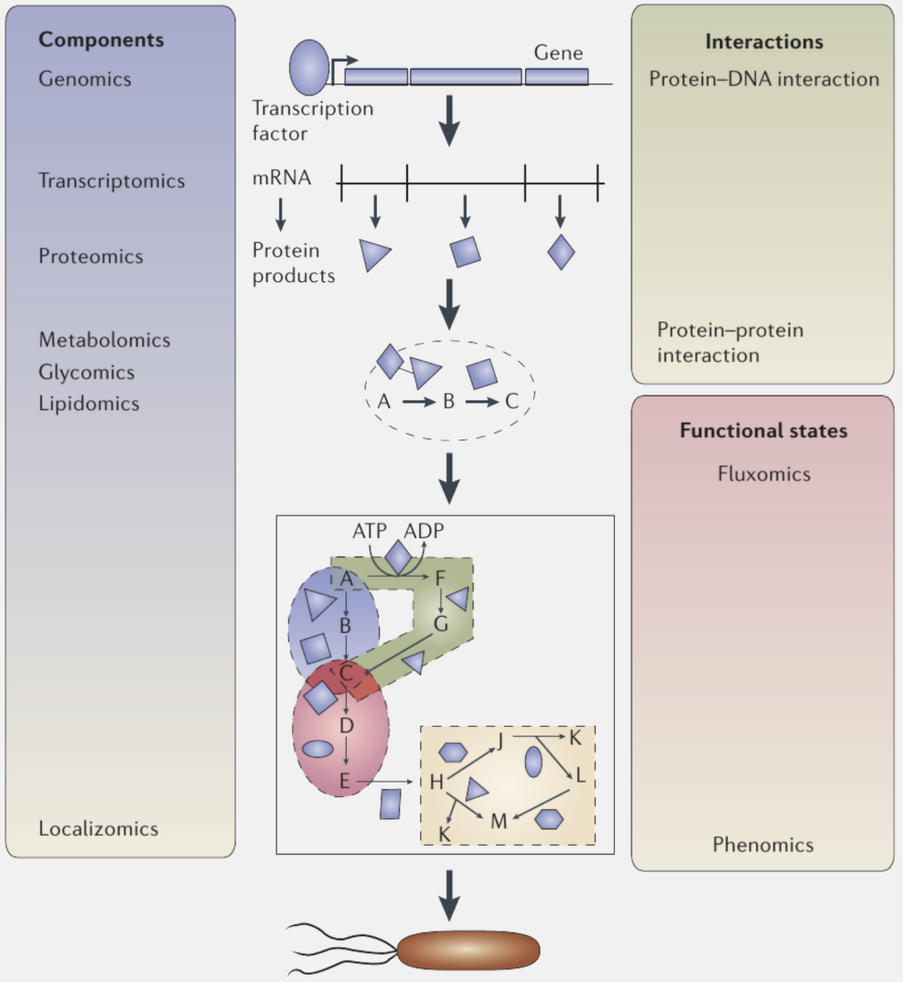
Alexandre Perera-Lluna

Omics measures different molecular layers

Layer	What is measured
Genomics	DNA sequence and variation
Transcriptomics	RNA abundance and transcript structure
Proteomics	Protein abundance, forms, and modifications
Metabolomics	Small molecules under defined conditions

No layer alone gives a complete description of biological function. Measurement coverage and interpretation depend on technology and study design.

Overview



Omics overview (Joyce and Palsson 2006).

Definition

Metabolomics is the science that studies the lightweight chemical compounds (**metabolites**) and reactions that take place in organisms.

Purpose

Metabolomics measures a context-dependent part of the molecular phenotype. Its utilities include (Nicholson et al. 2002):

- Finding of biomarkers for disease
- Understanding of diseases
- Prediction of treatment outcome

Metabolomic Analysis Workflow contributions at B2SLab (1/2)

- Fernández-Albert F., Llorach R., Garcia-Aloy M, Ziyatdinov A, Andrés-Lacueva C., Perera-Lluna A. Intensity drift removal in LC/MS metabolomics by Common Variance Compensation. Bioinformatics 30 (20): 2899-2905. (2014).
- Fernández-Albert F., Llorach R., Andrés-Lacueva C., Perera-Lluna A. An R package to analyse LC/MS metabolomic data: MAIT (Metabolite Automatic Identification Toolkit). Bioinformatics 30(13):1937-9 (2014).
- Fernández-Albert F., Llorach R., Andrés-Lacueva C., Perera-Lluna A.Peak Aggregation as an Innovative Strategy for Improving the Predictive Power of LC-MS Metabolomic Profiles. Analytical Chemistry 86 (5), 2320-2325 (2014).

Published R packages on metabolomics (available on our website http://b2slab.upc.edu) and on Bioconductor:

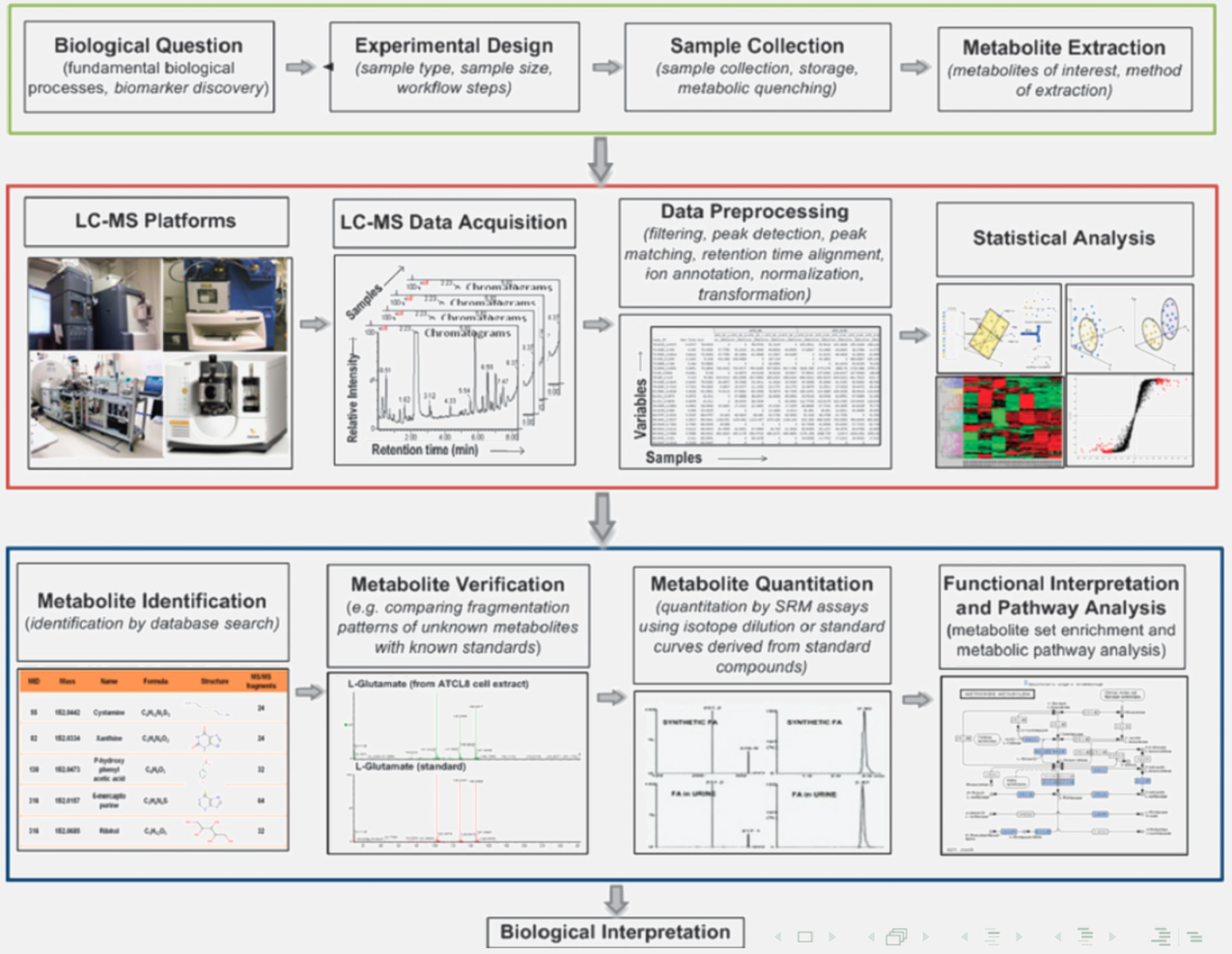
MAIT

MAIT: Metabolite Automatic Identification Toolkit. An R package to perform end-to-end LC/MS metabolomic data analyses.

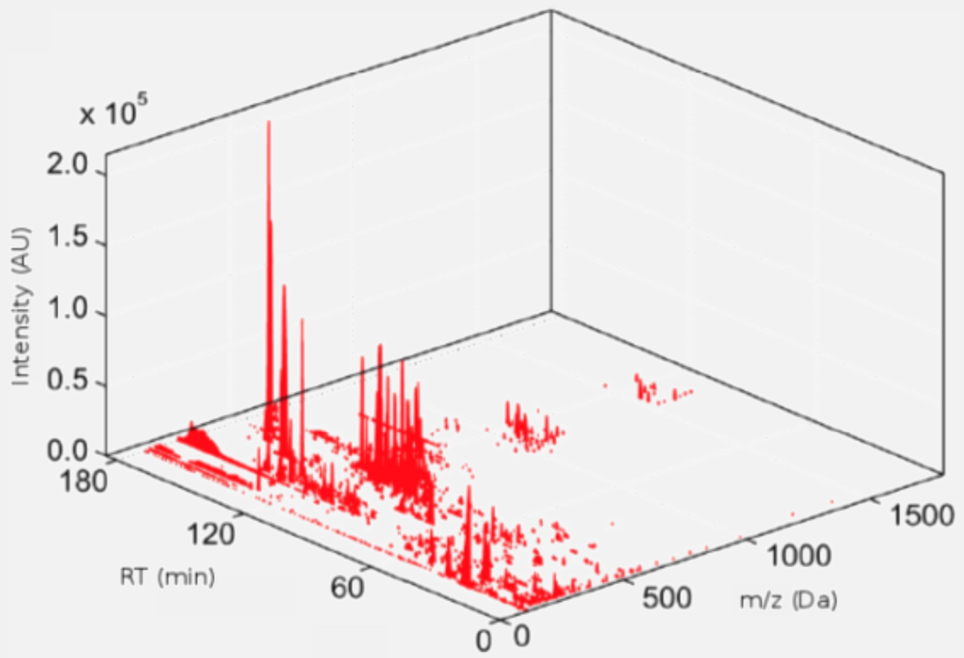
intCor

This package performs an intensity drift correction for large datasets through several methods. This library is contains the non-linear methodology to perform retention time correction in LC/MS.

Metabolomic Analysis Workflow (LC/MS)



LC/MS profile for a urine sample[2]

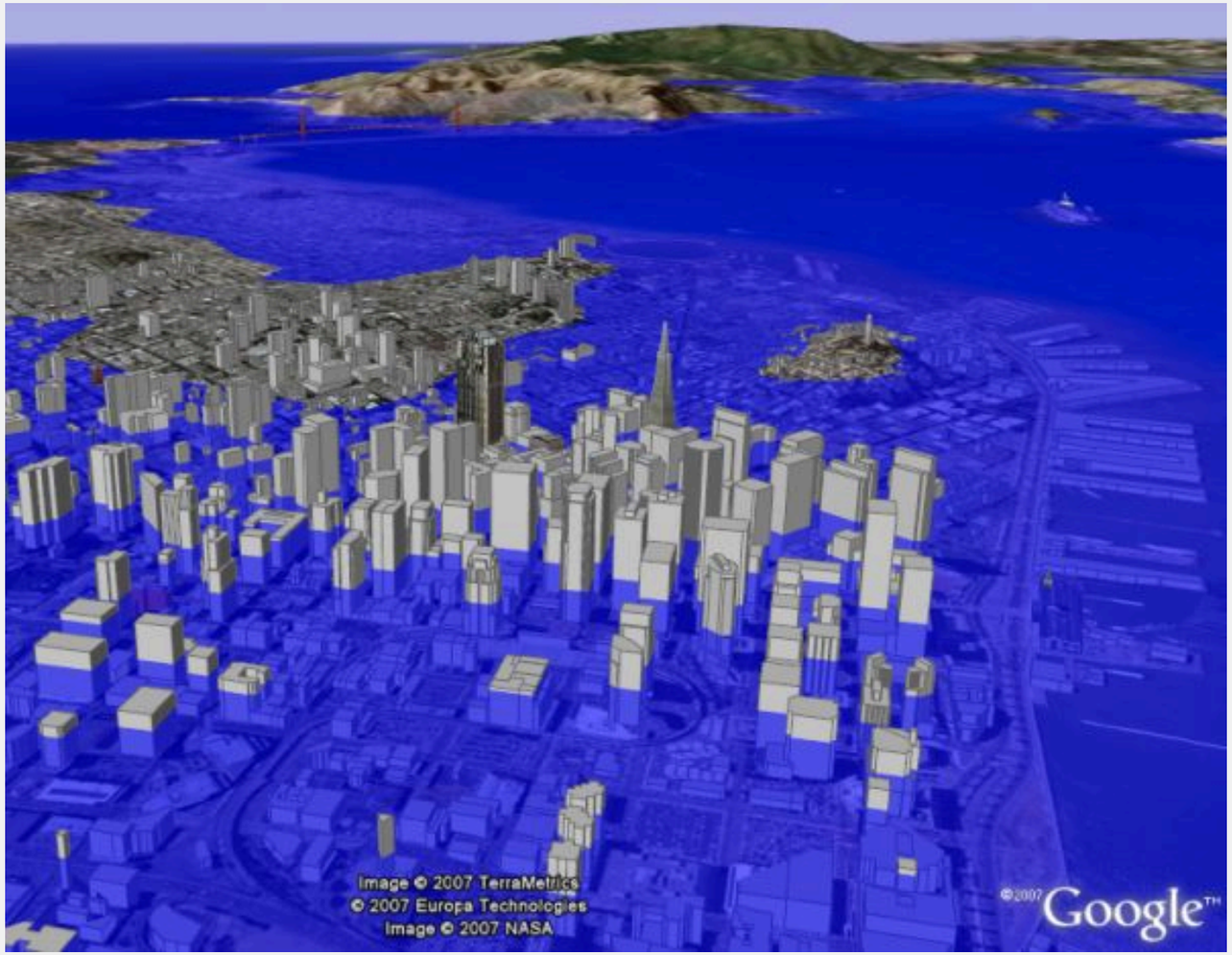


Signal Characteristics

- Given a sample, each peak has three characteristics: Intensity, mass/charge and retention time
- Highly sparse matrix
- Anisotropic behaviour
- Typical data for a sample are represented through a matrix having $\sim$10000 peaks(variables) and $\sim$10-100 samples

[2]Wei Guan et al. BMC Bioinformatics 10:259, 2009

Signal to noise?

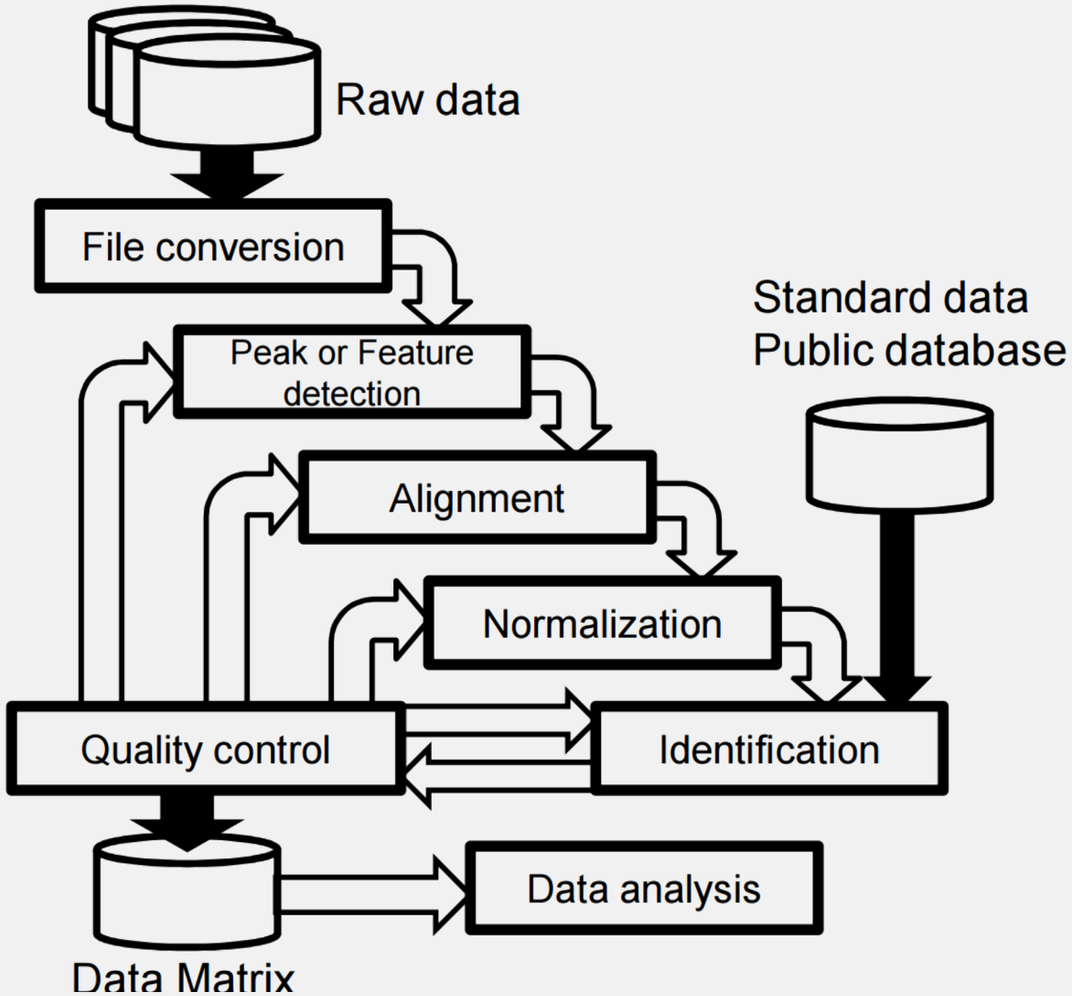


Metabolomics data (1/2)

Primary and secondary analyses

1. Primary analysis
    - Detect peaks
    - Annotate peaks
    - Quantify metabolites
2. Secondary analysis
    - Statistics
    - Machine learning
    - **Enrichment techniques**

Metabolomics data (2/2)



Metabolomics workflow (Sugimoto et al. 2012).

Enrichment techniques

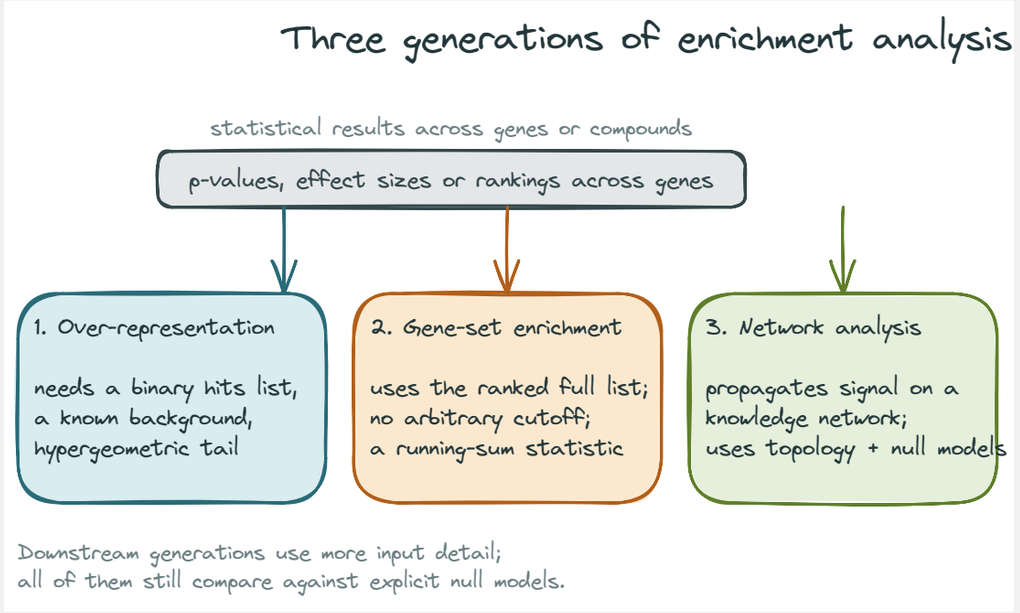
Experimental data can be understood in the framework of the **known biological context**. Enrichment techniques relate this data to **biological pathways**.

Enrichment in gene expression

> The challenge no longer lies in obtaining gene expression profiles, but rather in interpreting the results to gain insights into biological mechanisms.

(Subramanian et al. 2005)

Three generations appear on later slides: **over-representation** (binary hits list, known background), **gene-set enrichment** (ranked full list, no cutoff), **network analysis** (signal propagated over a knowledge network).



Annotation provenance

- **Computational annotation:** sequence comparison, predictive models, text mining, and other algorithms.
- **Manual curation:** expert evaluation of experimental and computational evidence.
- **Combined workflows:** computational proposals with human review.

Read the evidence and provenance attached to a claim. Annotation categories do not guarantee certainty.

Genes and gene products

- **Gene Ontology**: contains three ontologies and a hierarchical structure reflecting the specificity of the entry (Ashburner et al. 2000).

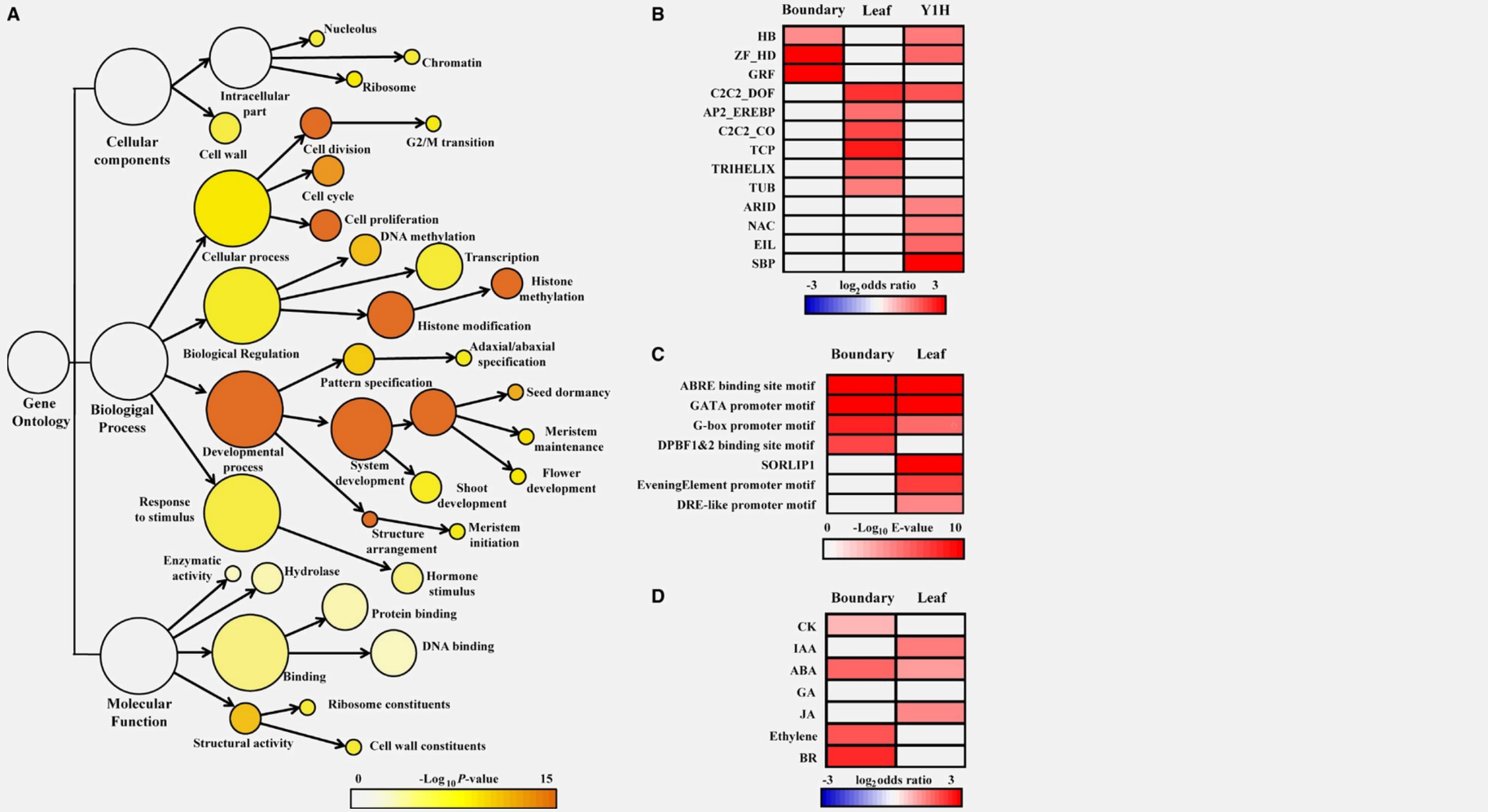
Proteomics

- **UniProt**: comprehensive database for protein sequence and annotation data (Consortium 2015).
- **STRING**: protein association database, including functional and physical associations; automatic and curated evidence (Szklarczyk et al. 2015).

Metabolomics

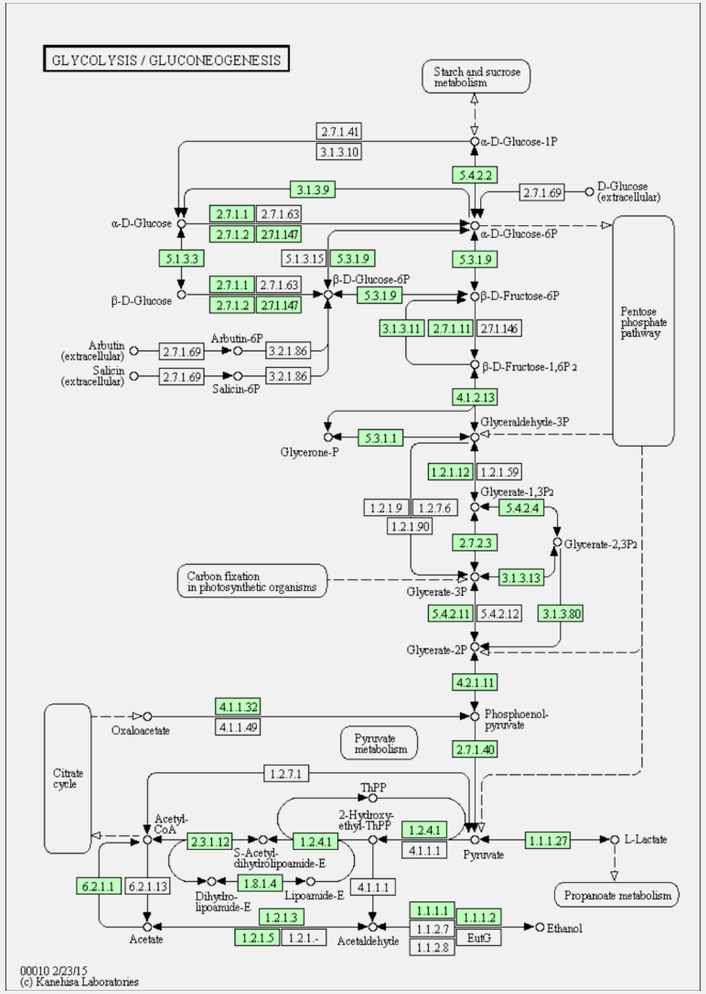
- **KEGG**: comprehensive and curated pathway database (Ogata et al. 1999).
- **Reactome, BioCyc, WikiPathways, SMPDB**

Gene Ontology



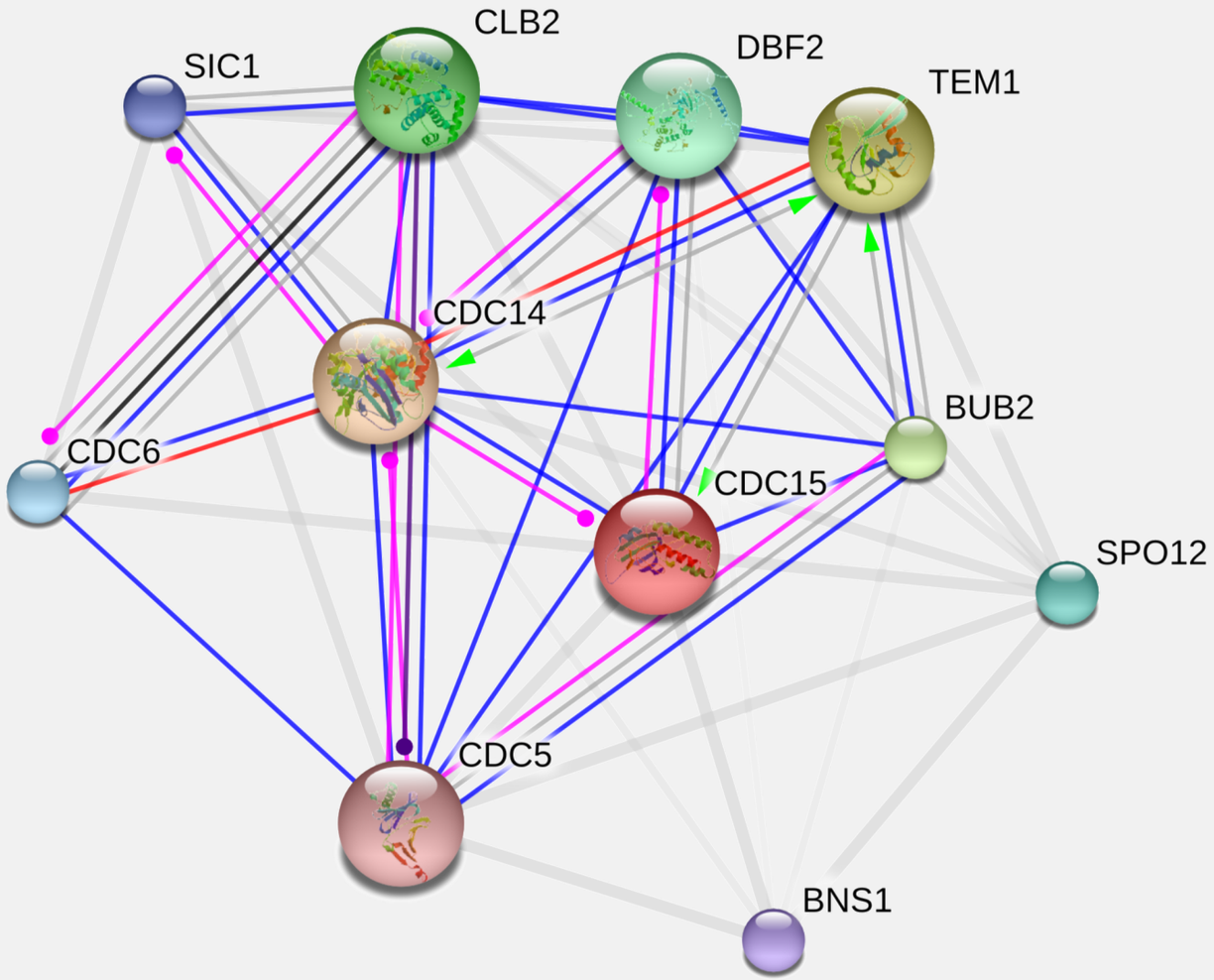
An example of GO overrepresentation analysis (ORA)(Tian et al. 2014).

Examples



Glycolysis pathway in KEGG (Ogata et al. 1999).

Databases (III) (2/2)



PPI network in STRING (Szklarczyk et al. 2015).

Define a background of N genes that could have been selected. A pathway contains K of these genes; the selected list has n genes, with overlap k.

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}.$$

For enrichment, use the **upper-tail** probability:

$$p = P(X \geq k) = \sum_{j=k}^{\min(K,n)} P(X = j).$$

Adjust across the tested gene sets. The background and selection process determine the null model.

Gene data enrichment: over representation

Advantages

- Relates genes with biological processes.
- Assesses statistical significance.
- Eases exploratory analysis.

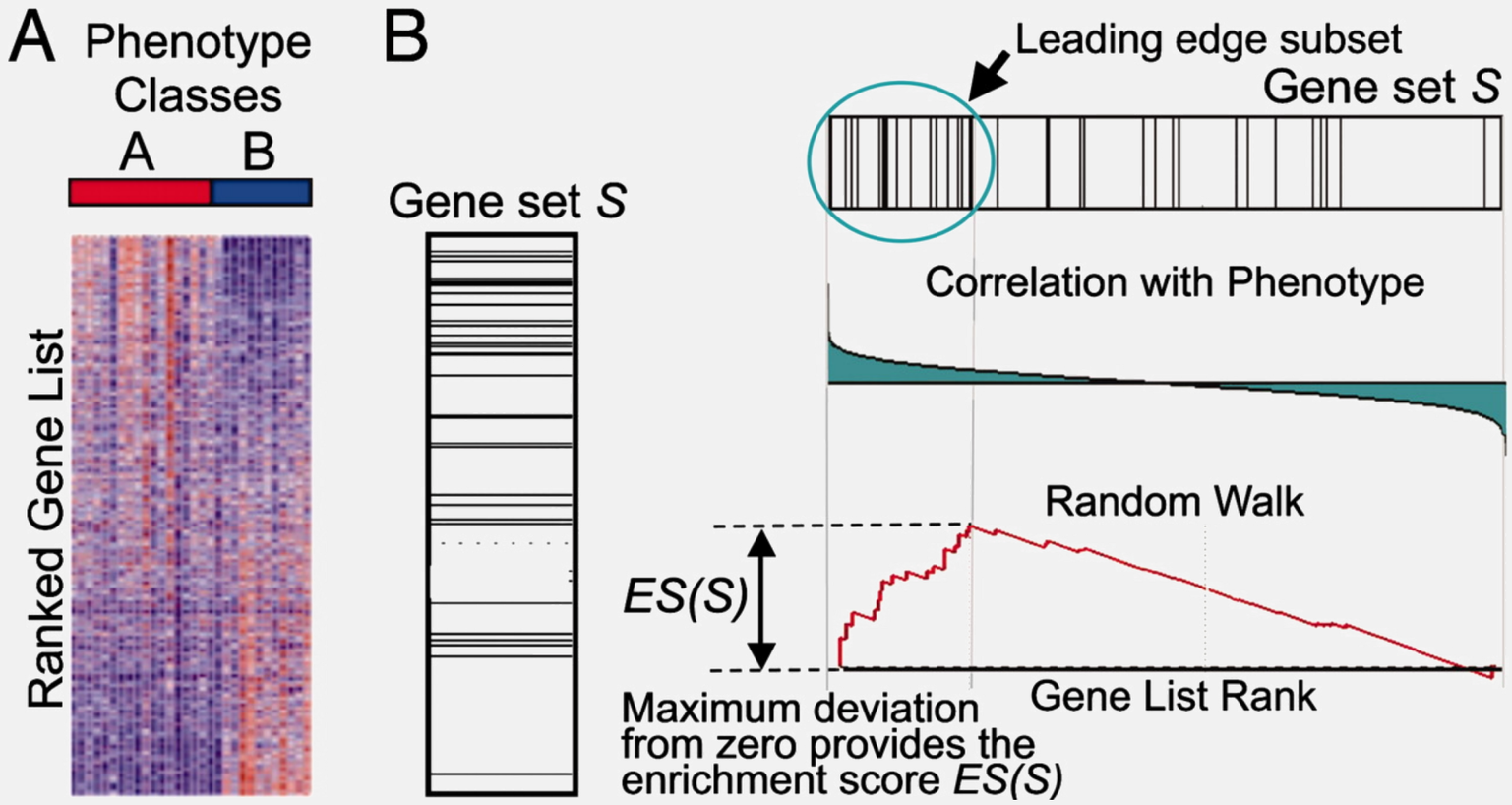
Limitations

- Relies on a prior (arbitrary) cutoff method.
- Sensitive on the input size.
- Ignores magnitude of the gene affectation.
- Simplistic context representation in a binary fashion.

Motivation

The second generation of methods intend to take advantage of **quantitative** data about gene expression.

Gene Set Enrichment Analysis (Subramanian et al. 2005)



GSEA procedure from original publication (Subramanian et al. 2005).

Gene data enrichment: set enrichment analysis (II)

Advantages

- Takes into account quantitative measures.
- Has reported to capture subtle but consistent changes in expression data.

Limitations

- Genes have complex interactions that could be accounted.
- Biological pathways exhibit substantial overlapping.

Motivation

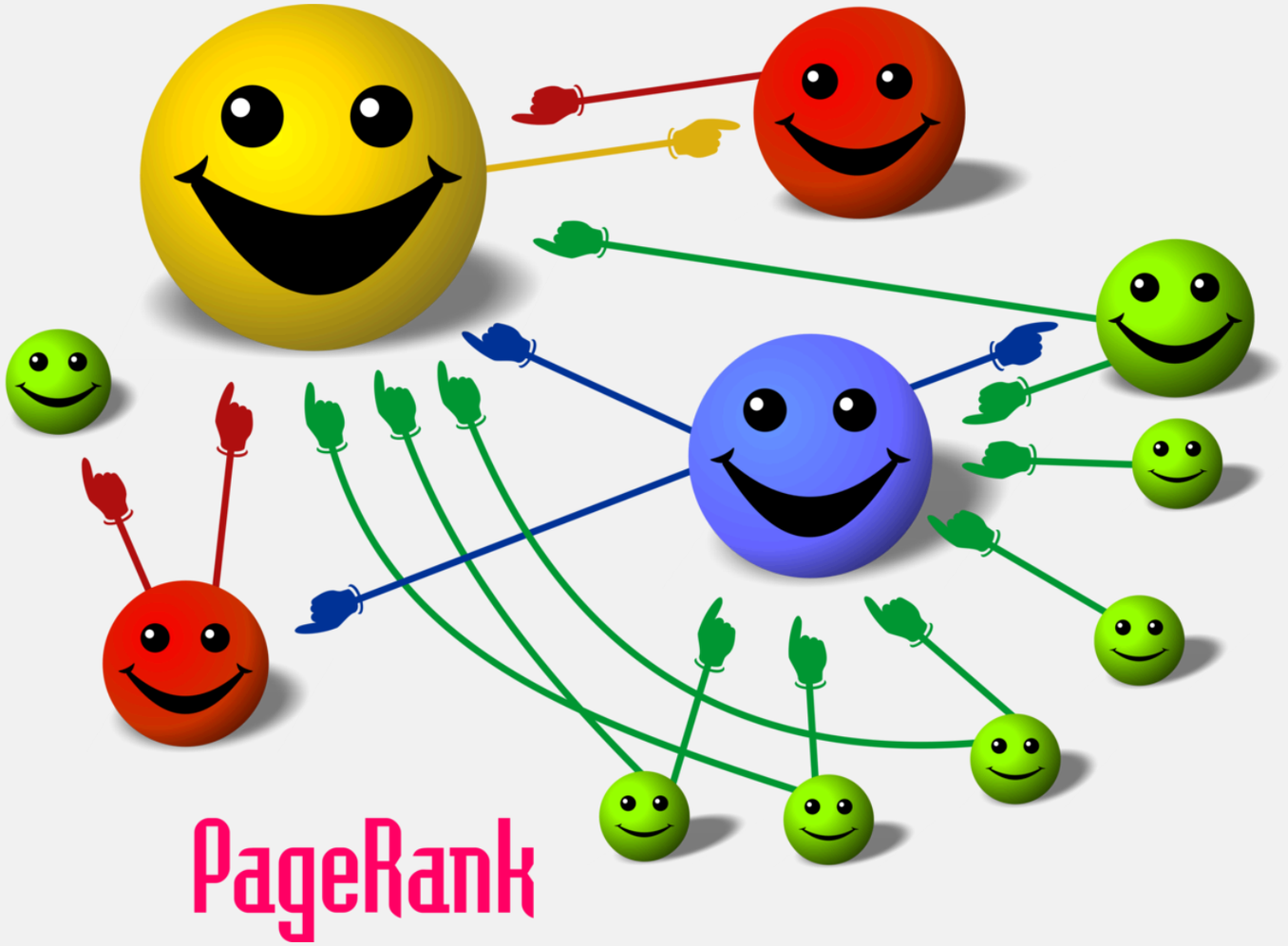
Network-based methods take into account the **topological features** of biological networks in order to test pathways.

Pathway-Express (Draghici et al. 2007)

$$IF(P) = \underbrace{\log \frac{1}{p}}_{\text{from ORA}} + \underbrace{\frac{\sum_{g \in P} |PF(g)|}{|\bar{\Delta E}| \cdot N_{de}(P)}}_{Pathwayperturbation}$$

- $\log \frac{1}{p}$ incorporates a p-value from over-representation analysis in the original Pathway-Express formulation
- $PF(g) = \Delta E(g) + \sum_{u \in US_g} \beta_{ug} \cdot \frac{PF(u)}{N_{ds}(u)}$ considers the topology of the genes within the pathway, and is similar to PageRank scores. $N_{de}(P)$: N DE genes on P_i,

Gene data enrichment: network analysis (I)



Gene data enrichment: network analysis (II)

Advantages

- Takes into account topological features.
- Benefits from current knowledge and its structure.

Limitations

- Pathway overlapping still remains an issue.
- Rising need to incorporate cell line specificity data in order to become biologically sound.

Origin

Metabolomics is a *young* science that followed Genomics and Transcriptomics trends. Concepts for data enrichment have been therefore ported from them.

Differences from genomic data enrichment

- Human reference genomes are available, but metabolite coverage depends strongly on the sample and measurement technology. The **background** for statistical tests is not obvious (Chagoyen and Pazos 2012).
- Experimental devices can exhibit low sensitivity or **noisy signals** (Hollywood et al. 2006).
- There is no technology to measure the full spectrum of compounds. Large part of the **background** is not observable.

Databases in Metabolomics

	KEGG[36]	MetaCyc [45]	PubChem [42]	ChEBI[43]	GMD[34]	HMDB[1]
Link	*www.genome.jp/kegg/*	*www.genome.jp/kegg/*	*pubchem.ncbi.nlm.nih.gov/*	*www.ebi.ac.uk/chebi/*	*gmd.mpimp-golm.mpg.de/*	*www.hmdb.ca/*
Type	Comprehensive	Comprehensive	Chemical	Chemical	Mass Spectral	Mass Spectral Comprehensive
Database Features	Genomes, genes, proteins, metabolites, drugs, diseases, pathways, visualizations	Genes, proteins, metabolites, pathways, interactive visualizations	Compound	Compound	Metabolites	Metabolites
Specificity	Generalized annotations, 2260 organism semi-specific annotations	Generalized annotations, 1939 organism specific annotations	Broad	Broad	Broad, plant heavy	Human

Main databases in Metabolomics (Booth et al. 2013). (correct link for MetaCyc is http://metacyc.org/), and I would add Reactome to the list http://reactome.org

Metabolomics data enrichment: overview (I)

Over Representation Analysis

- IMPaLA (Kamburov et al. 2011)
- MetaboAnalyst (Xia et al. 2012)
- MPEA (Kankainen et al. 2011)

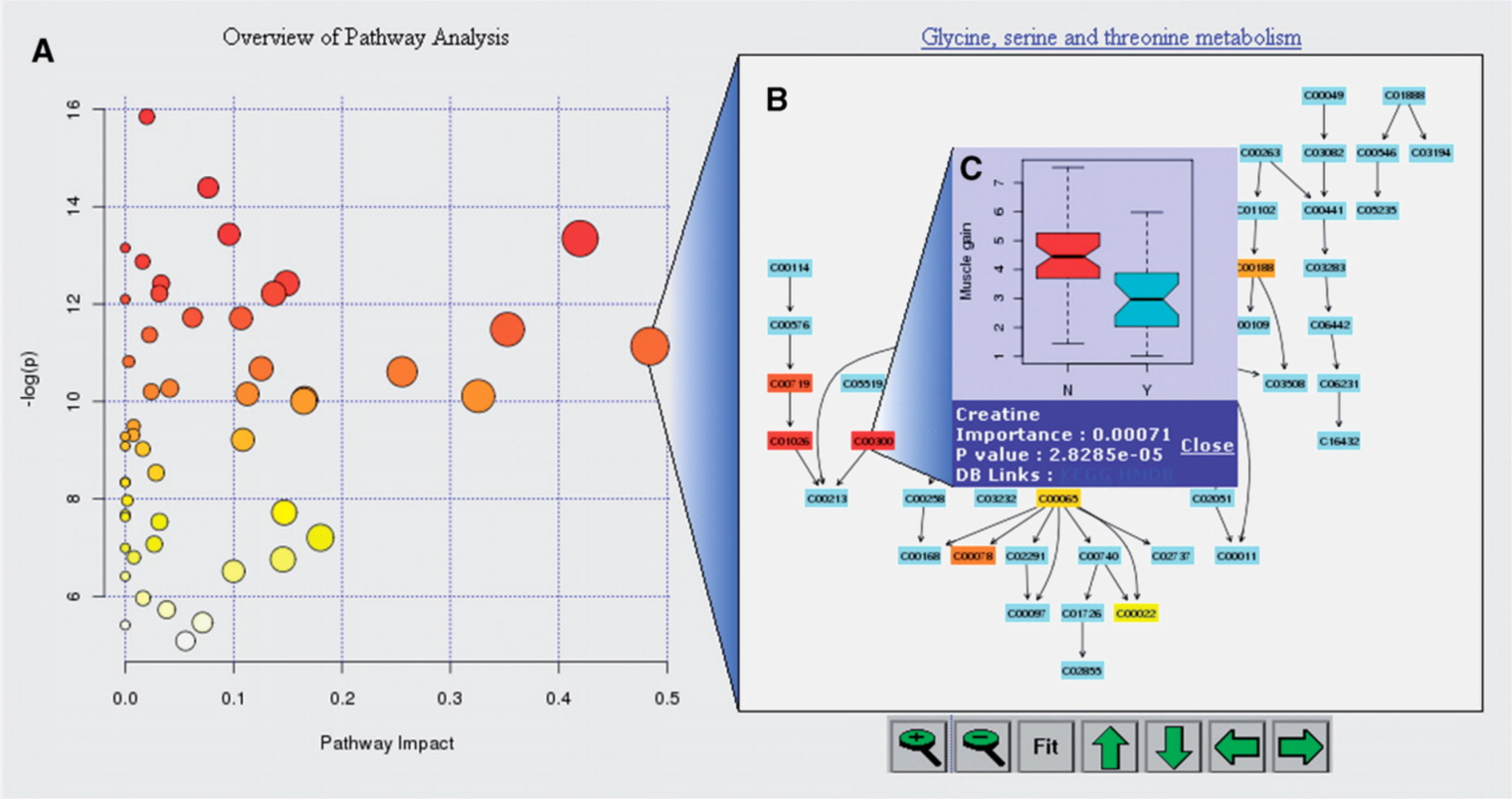
Metabolite Set Enrichment Analysis

- MSEA in MetaboAnalyst (Xia and Wishart 2010b)
- PAPi (Aggio et al. 2010)

Network Analysis

- MetPA in MetaboAnalyst (Xia and Wishart 2010a)

MetaboAnalyst (Xia et al. 2012)



MetaboAnalyst gathers topological analysis (x axis) and ORA or MSEA (y axis). It also depicts metabolite significance inside each pathway.

MetExplore [28]

Acc.: WB

Inp.: CompID, Mass IDs

DBsd: Generally BioCyc related

Funs.: Compound mapping, graph analysis of metabolism maps.

Comm: Choice of organism database, filtering options, multiple graph analysis tools, Cytoscape integration.

PAPi [24]

Acc.: R

Inp.: KEGG CompID

DBsd: KEGG

Funs.: Compares activity of metabolic pathways between sample types.

Comm: Non organism specific, more difficult/powerful command line R interface. Usable with spent media results.

MBRole [27]

Acc.: WB

Inp.: CompID

DBsd: KEGG, HMDB, PubChem, ChEBI, SMILES

Funs.: Enrichment analysis of metabolites' annotations.

Comm: Background set from known organisms or custom set. Metabolite ID converter.

MetaboAnalyst [57]

Acc.: WB

Inp.: Raw Spectra (GC and LC MS), peak lists and spectral bins (MS and NMR)

DBsd: Custom, KEGG, HMDB

Funs.: Full processing, Statistical Analysis

Comm: Comprehensive metabolomics analysis platform with easy interface, tutorials, help. Human focused though some model organisms or custom metabolite set option.

MetaboAnalyst (MSEA) [58]

Acc.: WB.

Inp.: CompID and abundances

DBsd: Custom, KEGG, HMDB

Funs.: Enrichment Analysis

Comm: Comprehensive metabolomics analysis platform with easy interface, tutorials, help. Human focused though some model organisms or custom metabolite set option.

MetaboAnalyst (MetPa [59])

Acc.: WB

Inp.: CompID and abundances

DBsd: KEGG

Funs.: Pathway Analysis

Comm: Select model organisms. Network topology analysis. Intuitive network visualization.

Metabolomics tools: historical review (2013) (7/8)

MPEA [29]

Acc.: WB

Inp.: CompID, GC-MS Spectrum as ranked list

DBsd: KEGG, GMD, SMPDB

Funs.: Pathway enrichment analysis.

Comm: Optional background set. Limited to top-down/bottom-up analysis.

MeltDB (MSEA) [30 and 60]

Acc.: WB, (L)

Inp.: Raw GC/LC-MS spectra, processed spectra, CompID and abundances

DBsd: GMD, KEGG, ChEBI, CAS

Funs.: Comprehensive preprocessing, statistical analysis and metabolite mapping, enrichment analysis.

Comm: Integrated comprehensive online system, accessible by multiple users. Many statistical tools, custom metrics and sets for enrichment analysis.

Adapted from a very nice review from (Booth et al. 2013)

Meta P-server [61]

Acc.: WB

Inp.: CompID, sample meta-data

DBsd: KEGG, HMDB, LipidMaps, PubChem

Funs.: Data quality control, statistical analysis, hypothesis testing.

Comm: No use of organismal databases. Focus mainly on global statistical analysis.

MassTrix [62 and 63]

Acc.: WB

Inp.: MS spectra

DBsd: KEGG, HMDB, LipidMaps

Funs.: Compound mapping

Comm: Choice of KEGG organism. Optional background set. Color-coding.

BioCyc (Pathway Tools) [52]

Acc.: Ins

Inp.: Annotated genome, ’omics data

DBsd: MetaCyc

Funs.: Network exploration, genome annotation, ’omics data painting.

Comm: Comprehensive systems biology network analysis.

Pathos [64]

Acc.: WB

Inp.: Simple m/z values, CompID

DBsd: KEGG

Funs.: Compound mapping

Comm: Choice of limited organism databases.

PaintOmics [65]

Acc.: WB

Inp.: KEGG formatted metabolites and/or genes

DBsd: KEGG

Funs.: Compound mapping

Comm: Choice of 100 hundred top species. Colours pathway metabolites and genes according to increase/decrease.

IMPaLA [66]

Acc.: WB

Inp.: Gene IDs and/or CompID

DBsd: KEGG, HMDB, CAS, ChEBI, PubChem, Reactome, Wikipathways

Funs.: Enrichment Analysis

Comm: Combined analysis with proteins or transcripts. Organism independent. Optional background set.

MetaMapp [25]

Acc.: WB

Inp.: CompID

DBsd: KEGG

Funs.: Metabolite networking

Comm: Organism independent. Network construction based on chemical similarity.

VANTED [67]

Acc.: Ins

Inp.: Compound abundances

DBsd: KEGG

Funs.: Metabolite networking, compound mapping, statistical analysis

Comm: Combined analysis with proteins and transcripts. Organism independent. Direct visualization of results on networks. Time course analysis. Statistical analysis.

Metabolomics tools: historical review (2013) (9/9)

TICL[68]

Acc.: WB

Inp.: Compound IDs

DBsd: KEGG

Funs.: Enrichment analysis.

Comm: No choice of organism. Reported as non-functional in the source review

Adapted from a very nice review from (Booth et al. 2013)

Subnetwork methods combine molecular measurements with a network to identify connected groups of relevant entities.

Scored modules

Methods such as jActiveModules search for subnetworks with strong aggregate evidence. Search procedures need not find a guaranteed global optimum. (Ideker et al. 2002)

Diffusion methods

Methods such as HotNet propagate influence through a network and identify affected modules. Heat diffusion is not the same as maximum-flow optimisation. (Vandin et al. 2011)

Problem addressed: enrichment techniques

- LC/MS provides data in terms of **compounds** of statistical relevance, but…
- The researcher needs to understand what is going on the underlying biology.
- A key concept are the **metabolic pathways**.
- Pathway enrichment techniques bridge between both worlds.

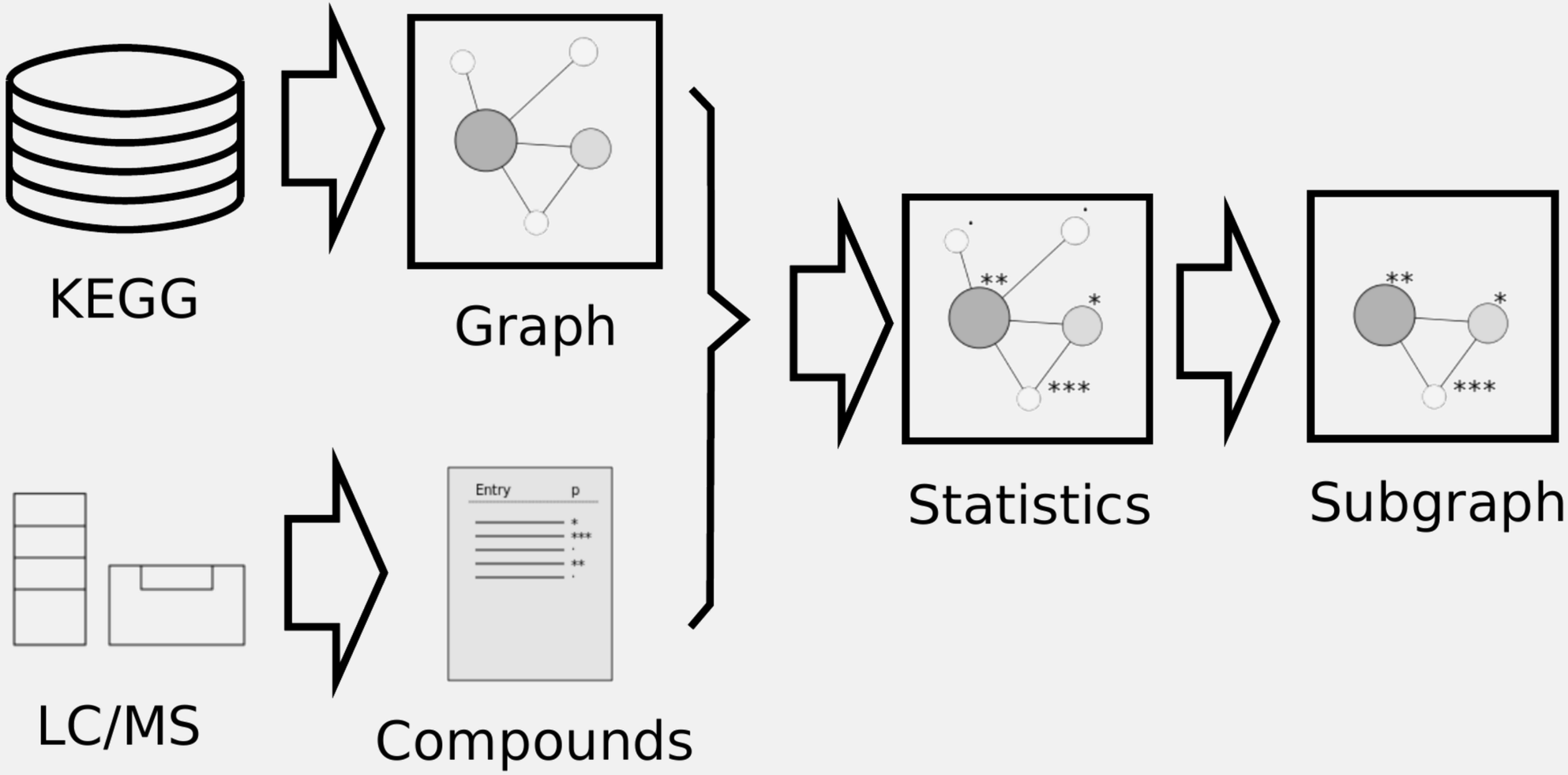
Goal

Integrate existing knowledge Base in form of a **database** with statistical results from **compound analyses** to perform a LC/MS pathway enrichment.

In addition…

- Offer a complete interpretation of the biological elements involved in the experiment.
- Allow further integration with other omics.

Adding a Knowledge Base



Pathway hsa00010 in the cell

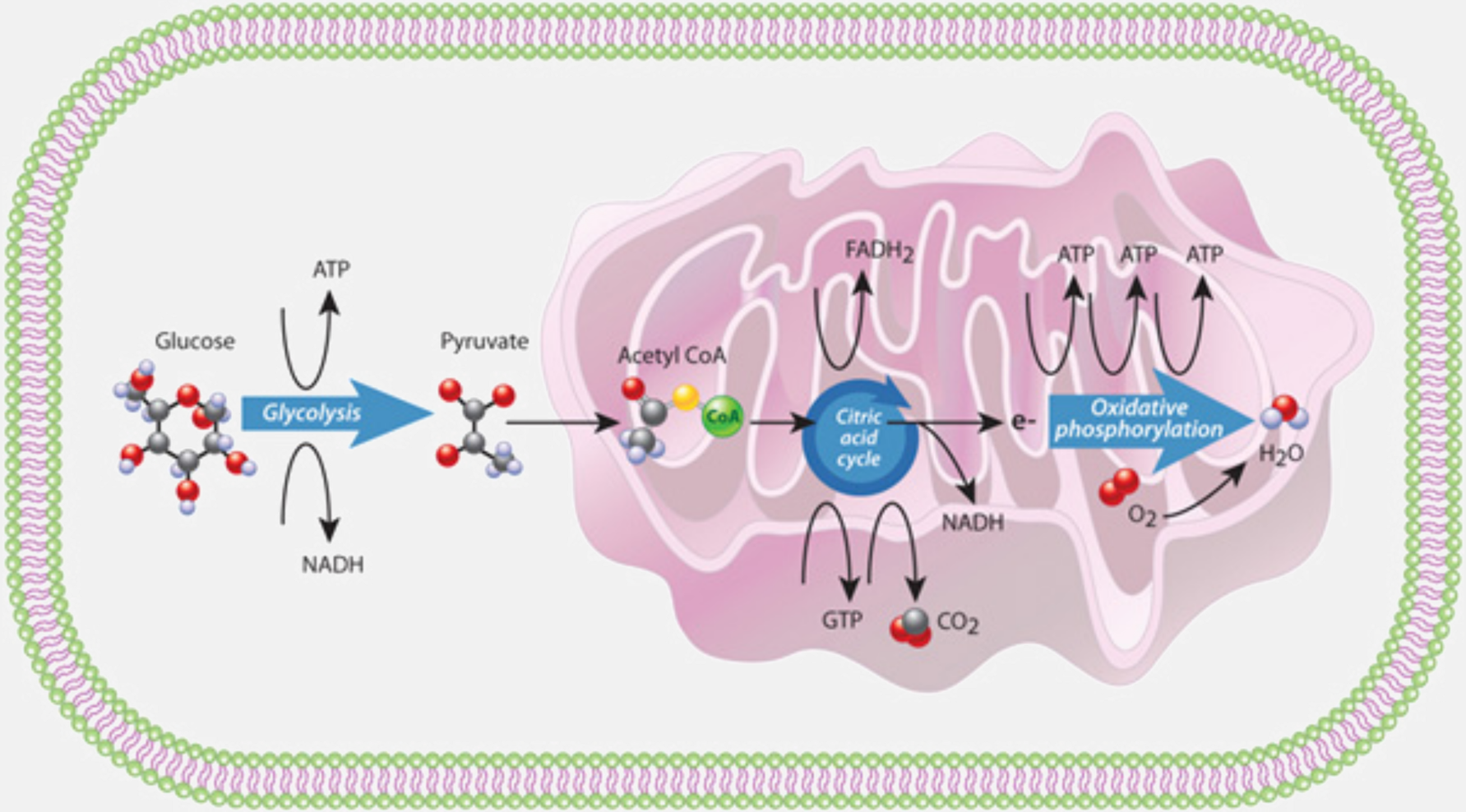
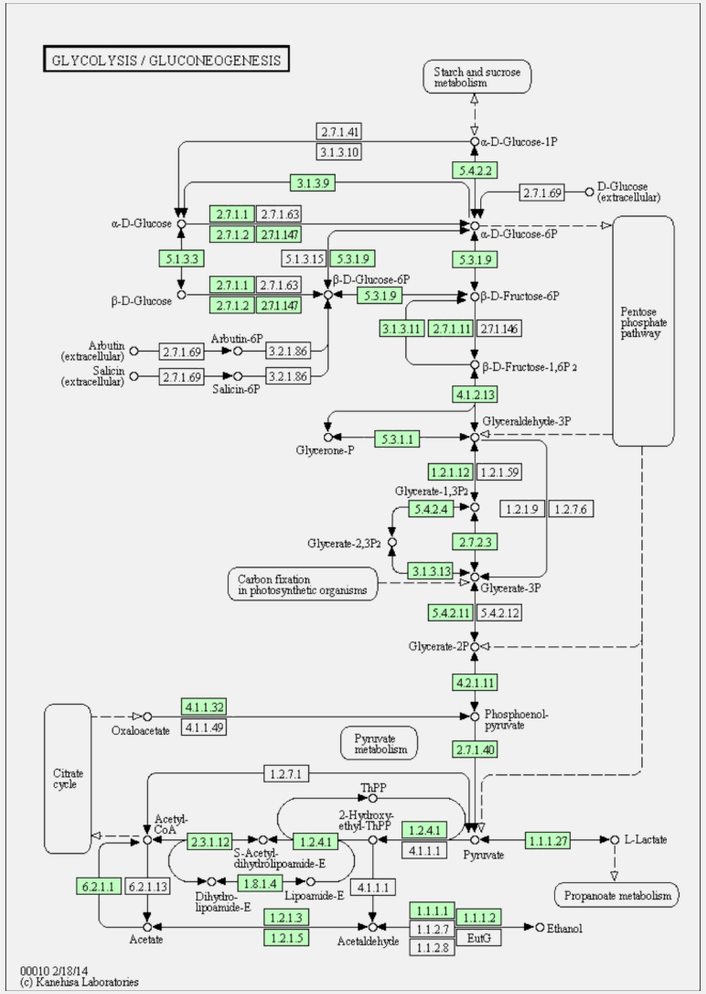


Image extracted from www.nature.com

Pathway hsa00010 in KEGG

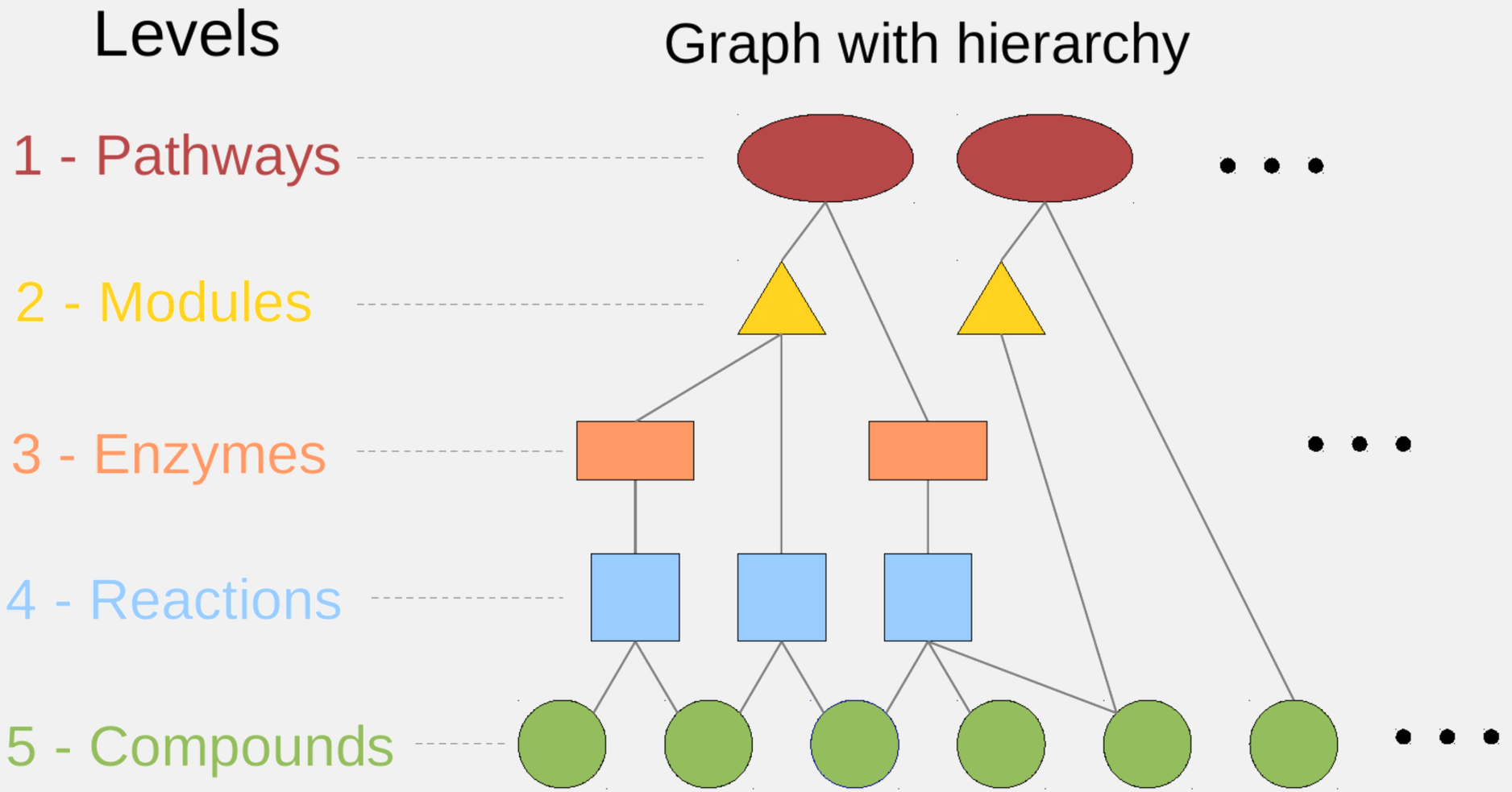


Knowledge Base: KEGG contents

Current Statistics

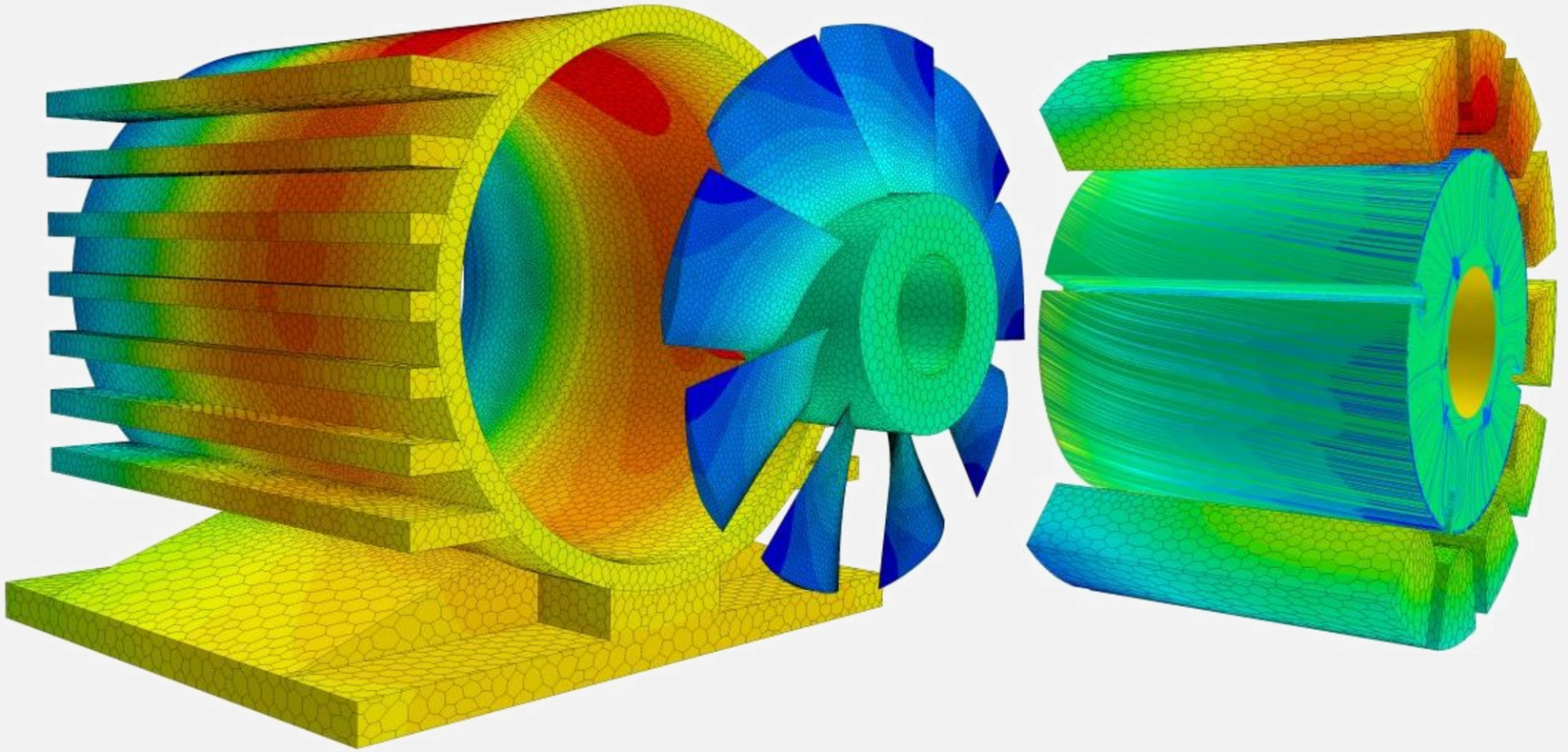
KEGG Database as of 2015/1/22

Systems information		
KEGG PATHWAY	Pathway maps, reference (total)	468 (346,221)
KEGG BRITE	Functional hierarchies, reference (total)	160 (120,717)
KEGG MODULE	KEGG modules, reference (total)	653 (279,580)
Genomic information		
KEGG ORTHOLOGY	KEGG Orthology (KO) groups	18,379
KEGG GENOME	KEGG Organisms	3,561
KEGG GENES	Genes in high-quality genomes (299 eukaryotes, 3052 bacteria, 178 archaea)	15,768,634
KEGG SSDB	Best hit relations within GENES Bi-directional best hit relations within GENES	
KEGG DGENES	Genes in draft genomes (26 eukaryotes)	654,883
KEGG MGENES	Genes in metagenomes (796 samples)	131,900,984
Chemical information		
KEGG COMPOUND	Metabolites and other small molecules	17,361
KEGG GLYCAN	Glycans	10,987
KEGG REACTION	Biochemical reactions	9,790
KEGG RPAIR	Reactant pair chemical transformations	14,874
KEGG RCLASS	Reaction class	2,949
KEGG ENZYME	Enzyme nomenclature	6,415
Health information		
KEGG DISEASE	Human diseases	1,402
KEGG DRUG	Drugs	10,139
KEGG DGROUP	Drug groups	1,721
KEGG ENVIRON	Crude drugs and health-related substances	850


Levels
Graph with hierarchy
1 - Pathways
2 - Modules
3 - Enzymes
4 - Reactions
5 - Compounds

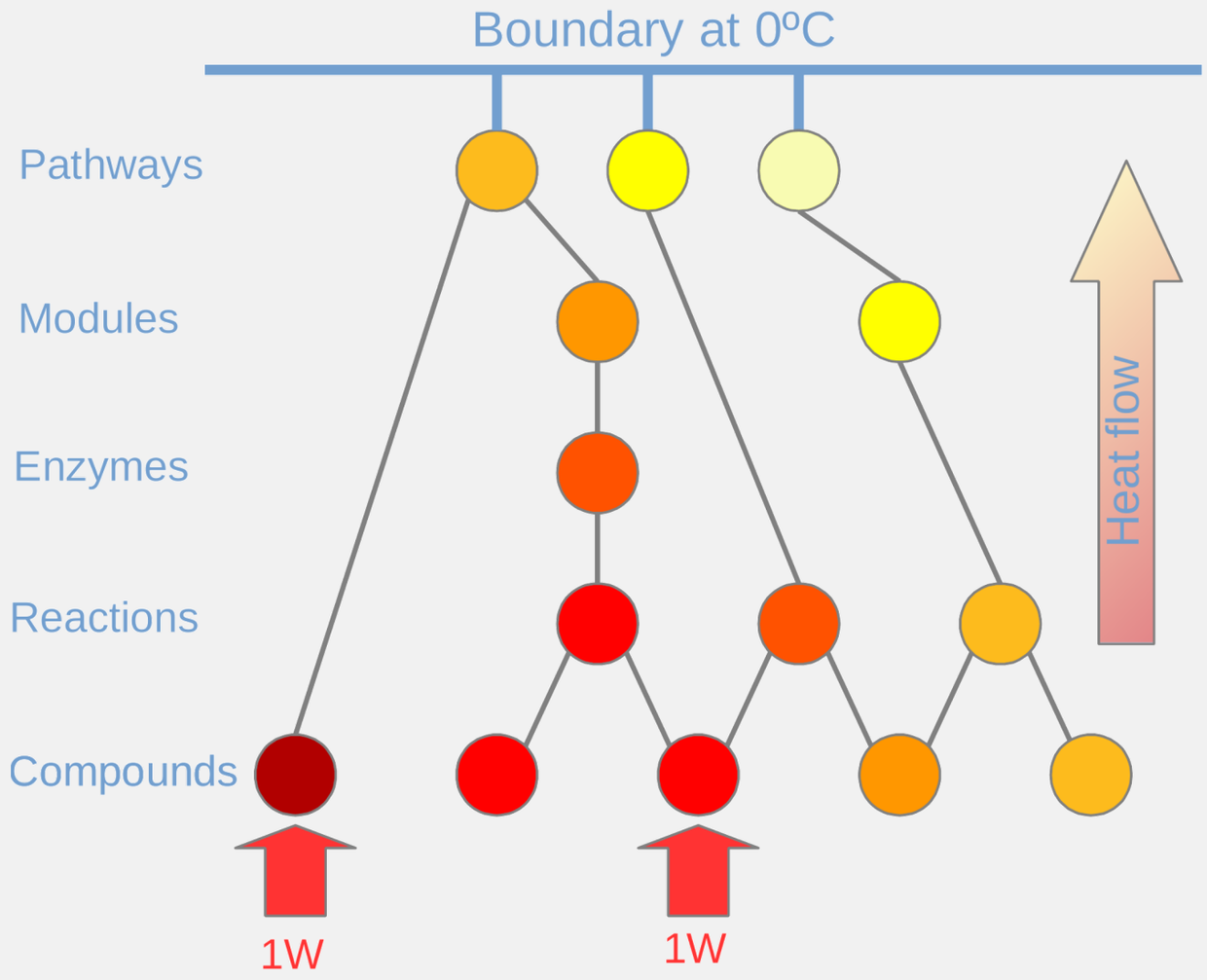
Motivation

How can we find relevant parts of a graph? We will generate heat on certain nodes and analyse the hottest subgraph.



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

Diffusion on graphs (I) (2/2)



Heat diffusion in a graph: some nodes introduce heat to the structure.

Basics

We apply the finite difference formulation of the heat equation:

$$T^{n+1} = T^n + DTC \cdot [KI \cdot T^n + KC \cdot TC + G]$$

T^k

are the node temperatures in the k-th instant of time.

DTC

is a diagonal matrix relating the time step and heat capacities of the nodes.

KI

is the matrix containing the conductances between nodes.

KC

is the matrix containing the conductances between the nodes in the graph and the boundary nodes.

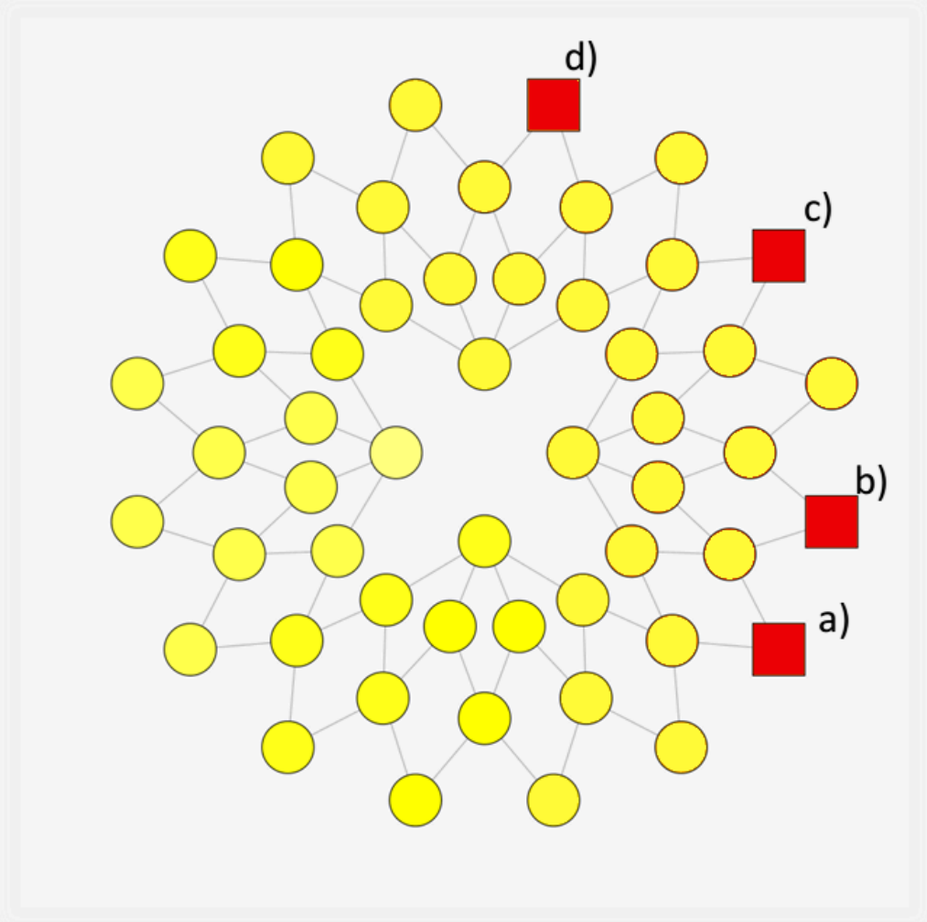
TC

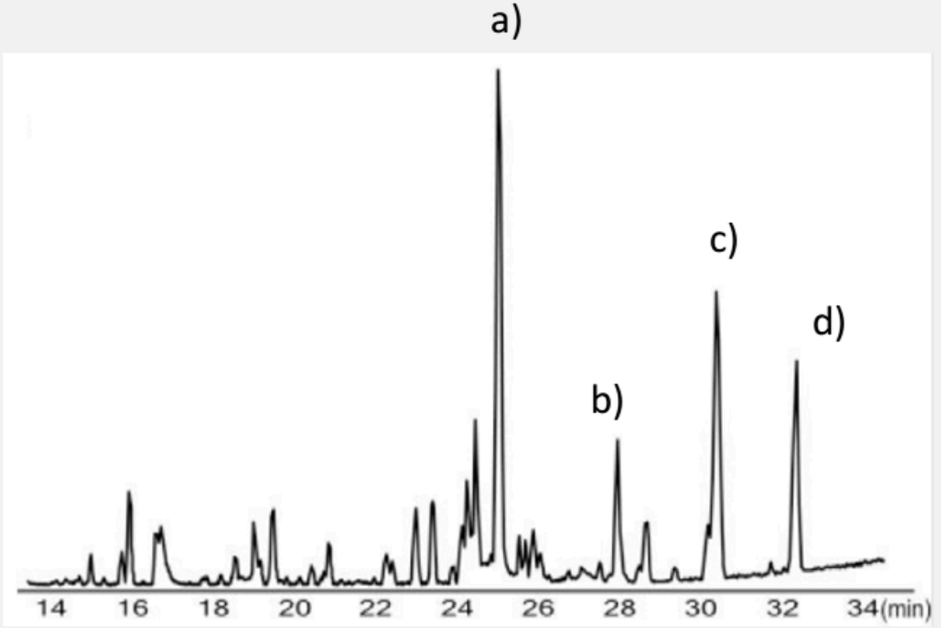
are the temperatures of the boundary nodes.

G

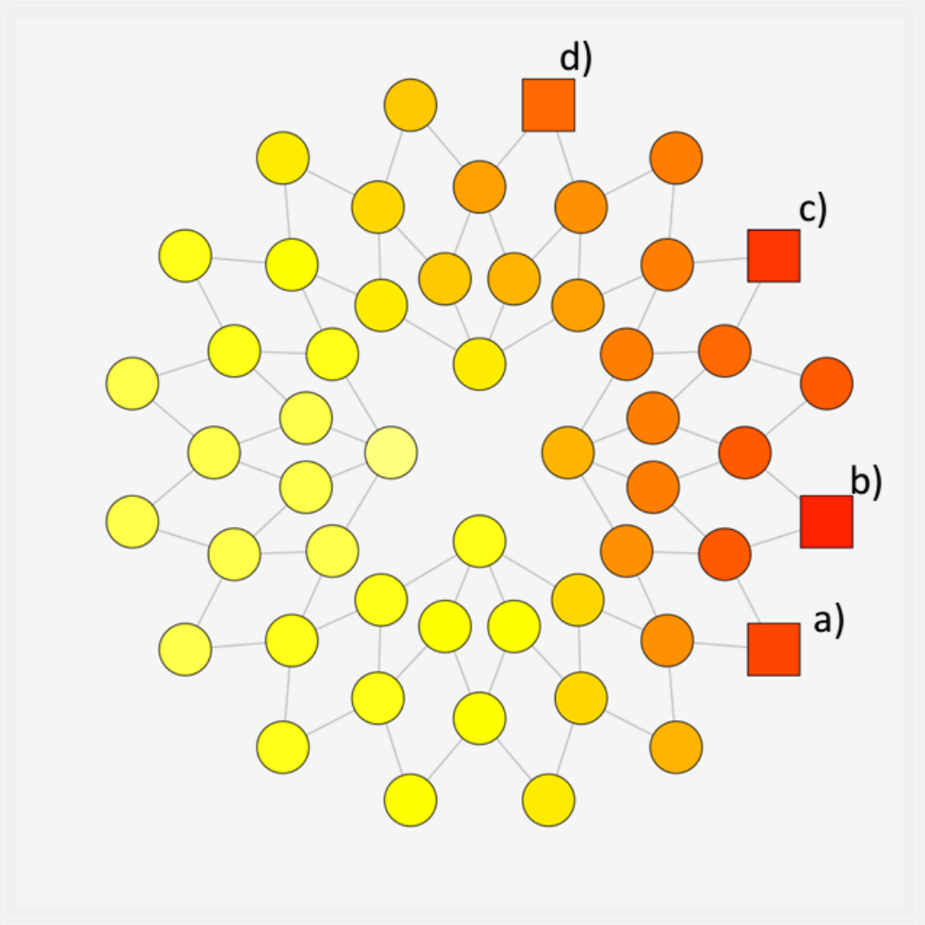
is the inner heat generation of the graph nodes.

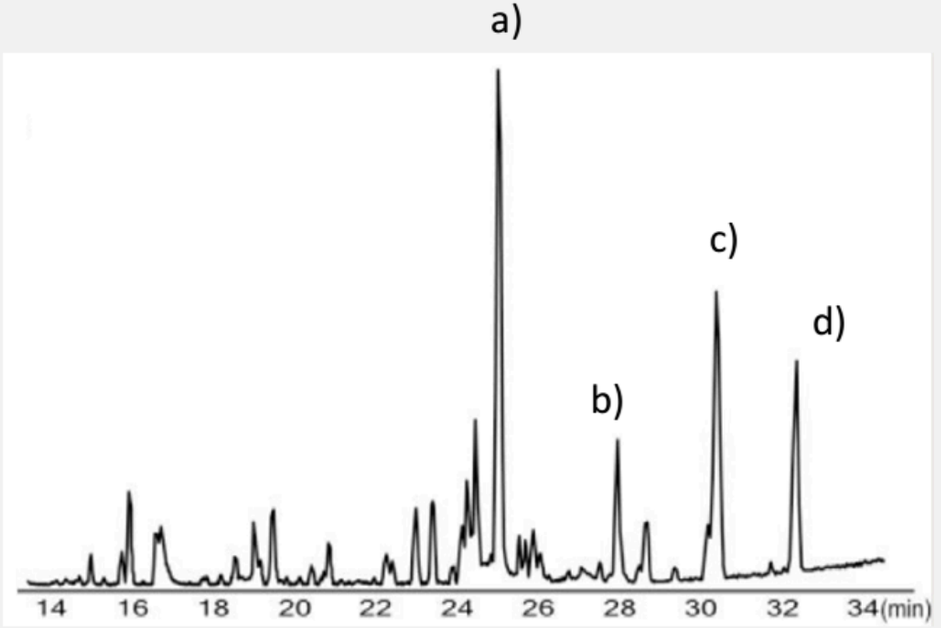
Input set





Input set





Deterministic input

In this case, we know who the affected nodes are. Thus, we get a deterministic temperature:

$$T_{input} = R \cdot q$$

… but how can we know if a node is hot or not?

- Temperatures are not explicative by themselves
- Temperatures are **not** comparable between nodes

Random input

However, what happens if the input contains no signal, i.e. is random?

$$T_{null} = R \cdot X$$

Where **X** is a random variable that randomly selects compounds.

Meaning of the null model

Analysing the distribution of temperatures for random cases allows the comparison between nodes in terms of their **distributions**. *Is the temperature of node A as extreme as node B?*

Statistical moments

We have determined the **exact** first and second moments of the temperature distributions:

$$\mathbb{E}(T(X)_i) = \frac{n_{in}}{n_p}\left[\sum_j r_{ij}\right]$$

$$V(T(X)_i) = \frac{n_{in}(n_p - n_{in})}{n_p(n_p-1)}\left[\sum_j r_{ij}^2 - \frac{1}{n_p}\left(\sum_j r_{ij}\right)^2\right]$$

Parametric approximation

Choose and assess a distributional approximation for the null temperature, then calculate an upper-tail probability.

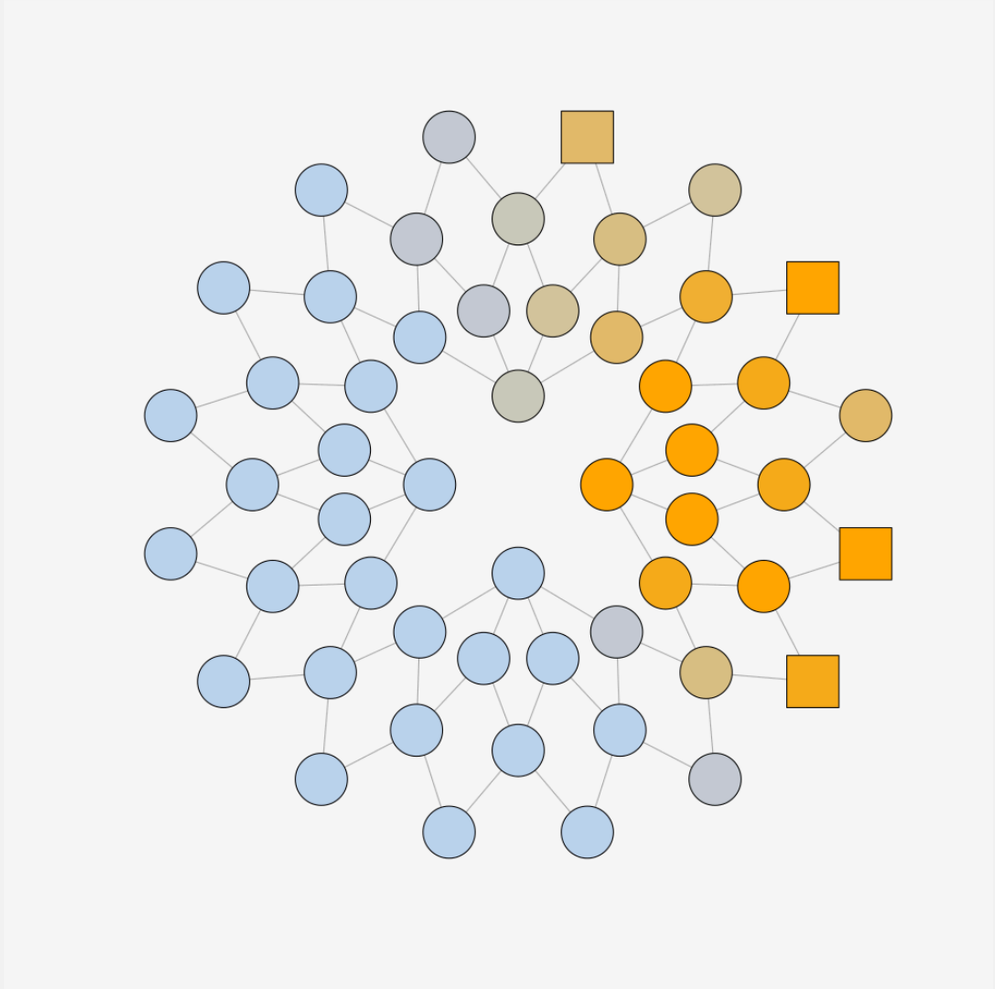
Monte Carlo estimate

For B null simulations, a common estimate is

$$\hat{p} = \frac{1 + \#\{b : T_i^{(b)} \geq T_i^{\mathrm{obs}}\}}{B + 1}.$$

The null sampling scheme, ties, and number of simulations matter. Adjust for the family of nodes tested.

Node-base test



Motivation

In practice, not all the nodes of interest can be **measured**. Technology only allow us to measure a subset of possible compounds in the database. Thus, there is a **portion of nodes that introduce uncertainty**. Assuming that they are not affected may be altering the results, depending on their amount…

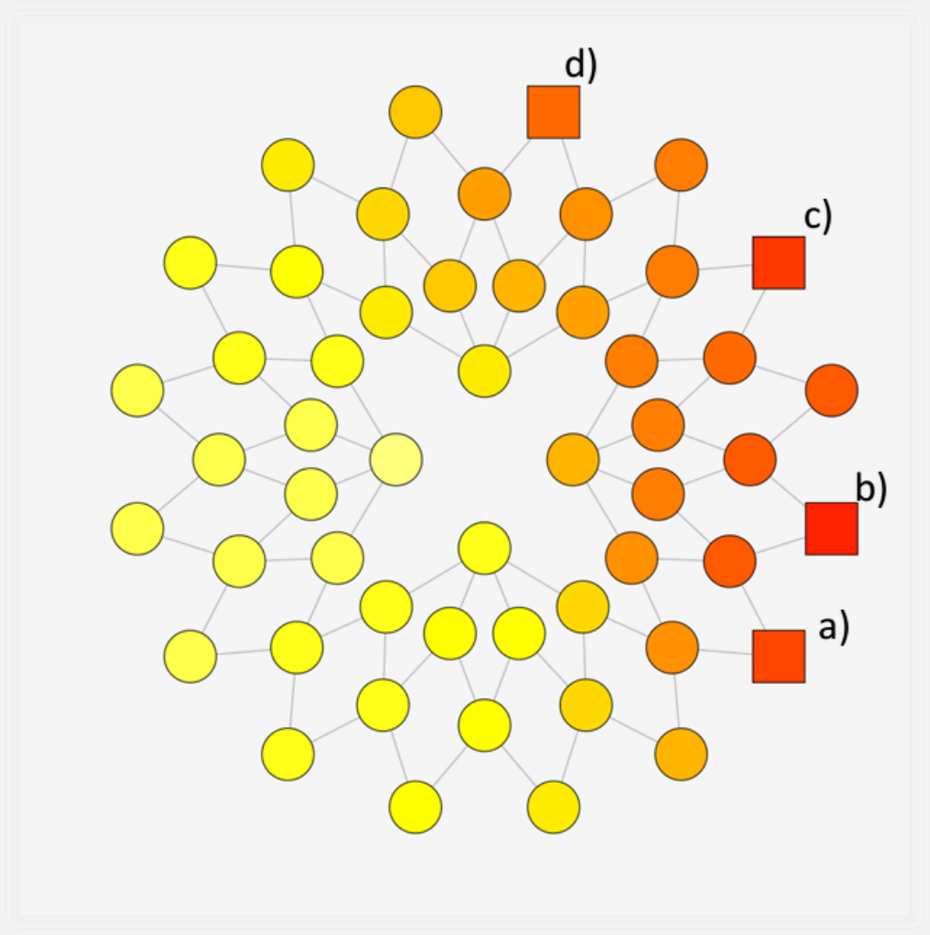
Proposed solution

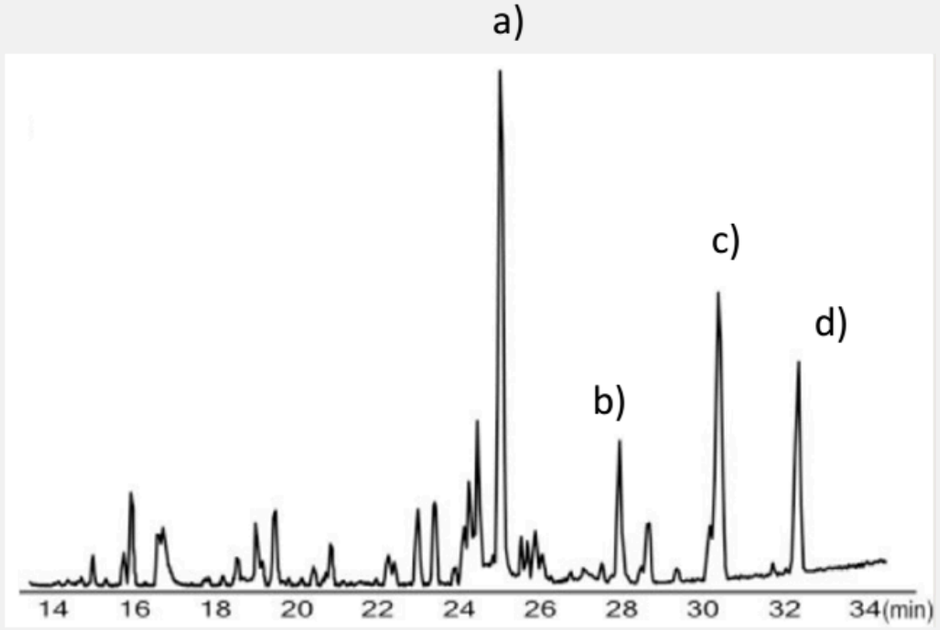
We suggest splitting a null model consisting of two parts:

- **Observable pool**: contains the nodes that can be measured. The affected nodes are a subset of the observable pool.
- **Latent pool**: contains the nodes that **cannot be measured** through current technology.

We want to see how the presence of the **latent pool** modifies our statistical tests.

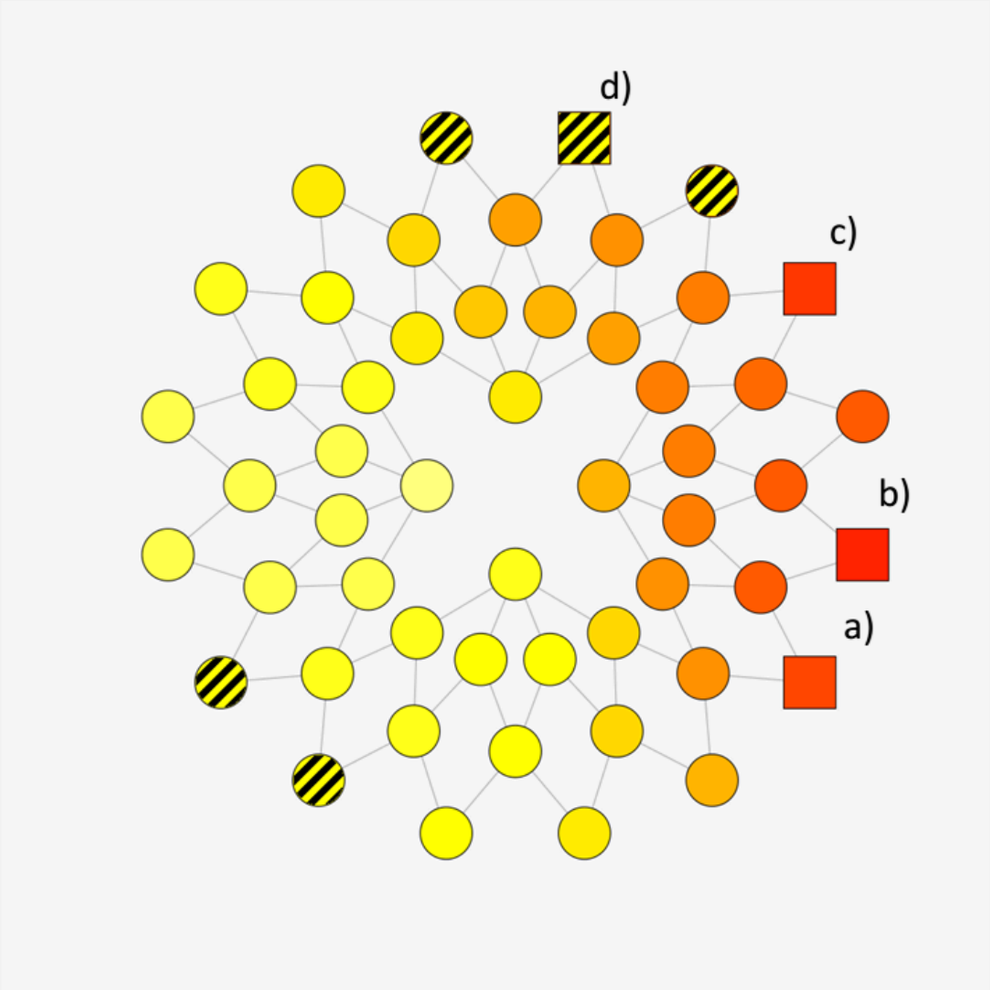
Assuming Latent Variance

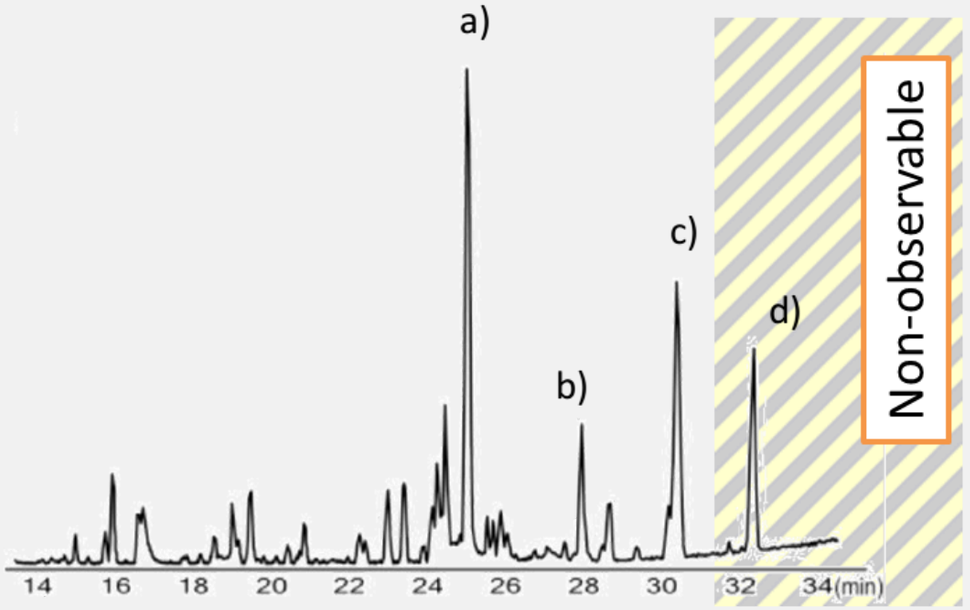




Our original input

Assuming Latent Variance





Barred nodes are **not observable**

Modification of statistical moments

We will call **ALL** the full model, **OBS** the observable model and **LAT** the latent model.

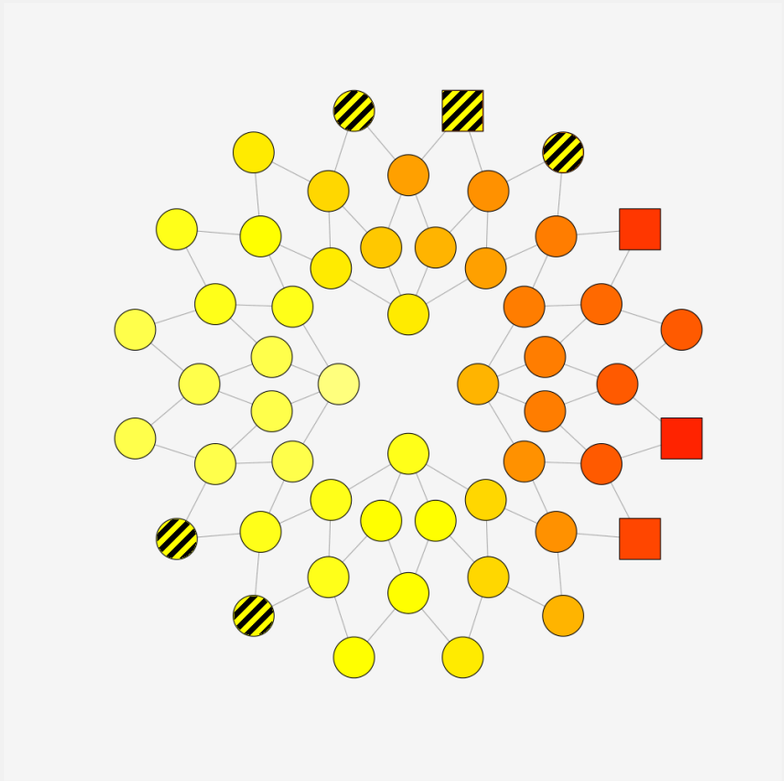
$$\mu_{ALL} \text{ remains unaffected}$$

$$\Sigma_{ALL} = \Sigma_{OBS} + 2 \cdot \Sigma_{LAT}$$

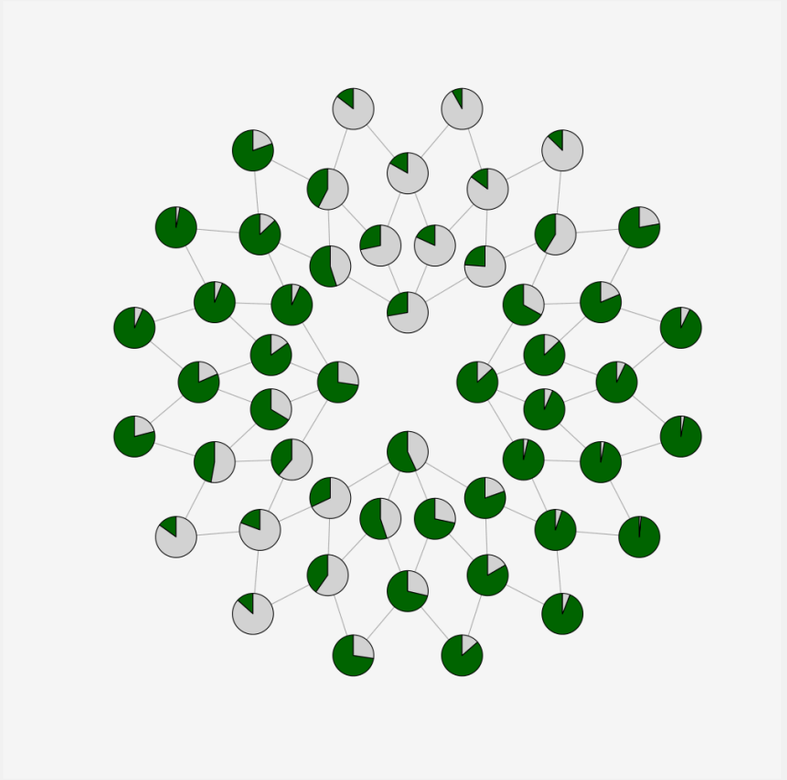
Interpretation

By introducing the latent pool all the variances grow. The magnitude of this enlargement will assess **how sensitive a particular node is** to experimental uncertainty.

Assuming Latent Variance

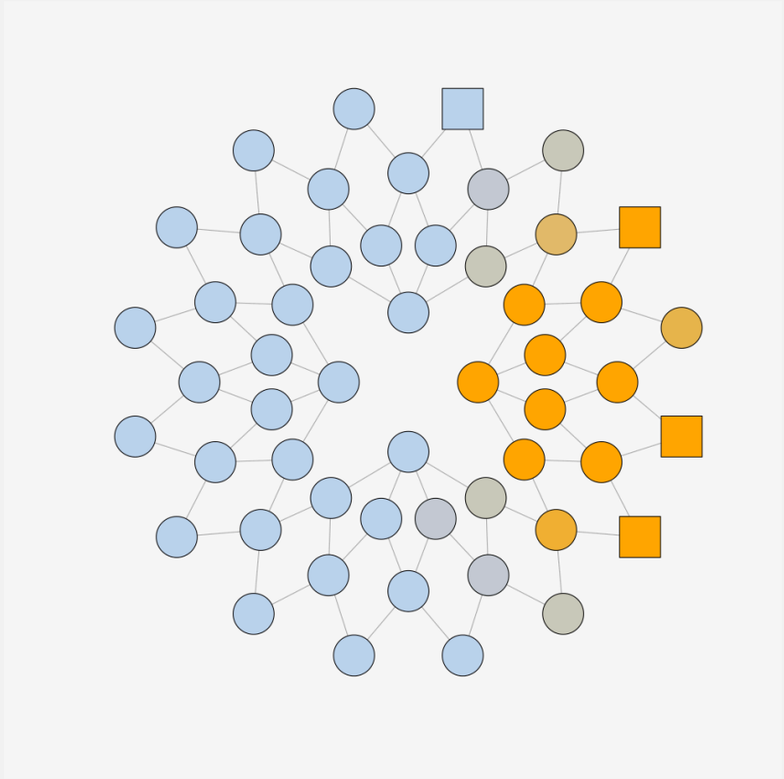


Barred nodes are **not observable**

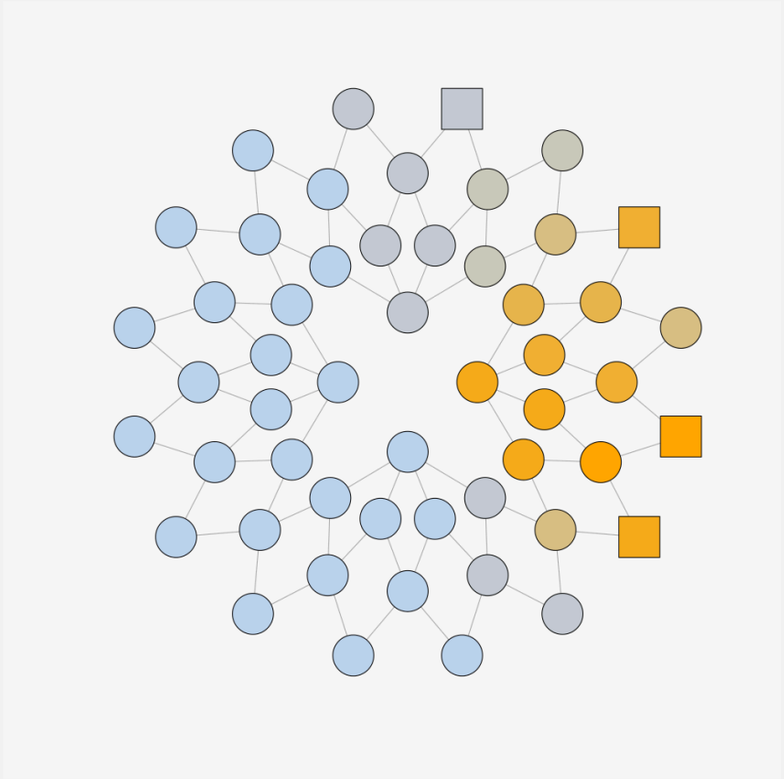


Proportion of variance explained by non-*observability* (red)

Assuming Latent Variance

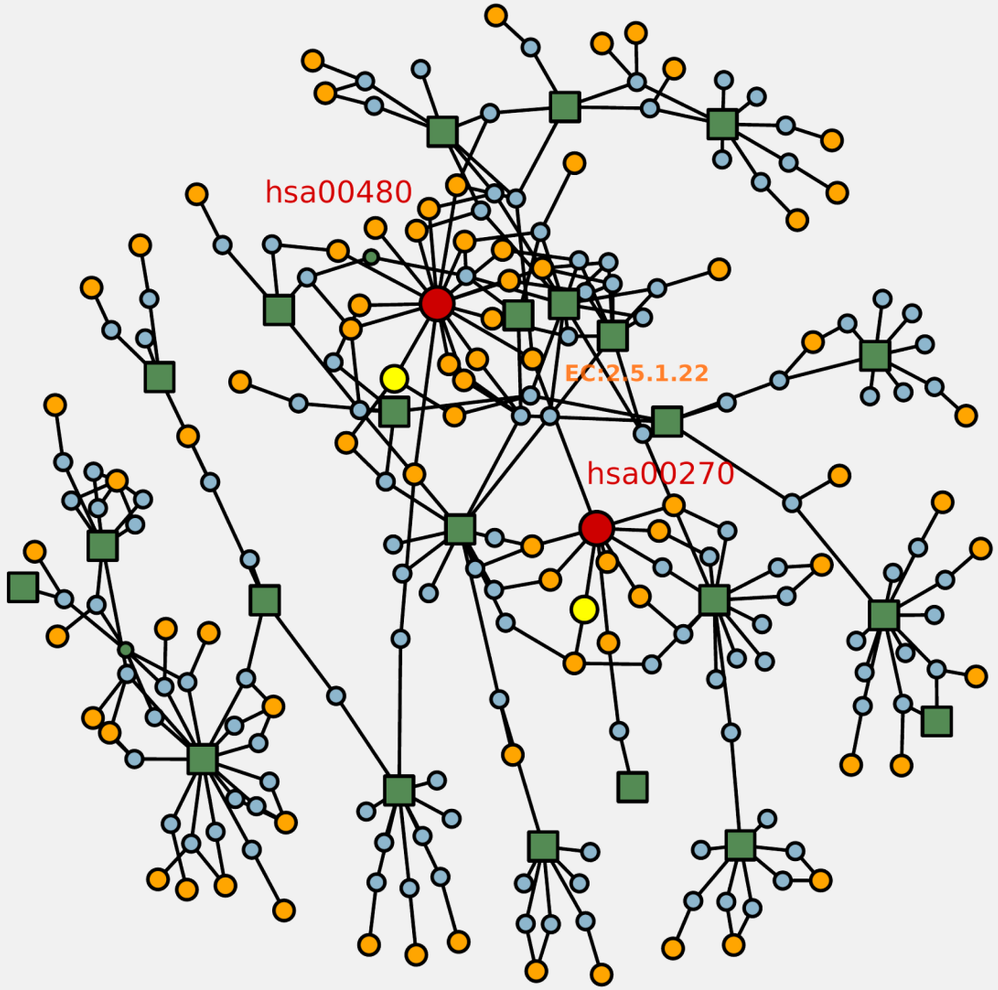


All graph is observable. Including the 5 non-observable nodes from the null model as non-affected (not fair)



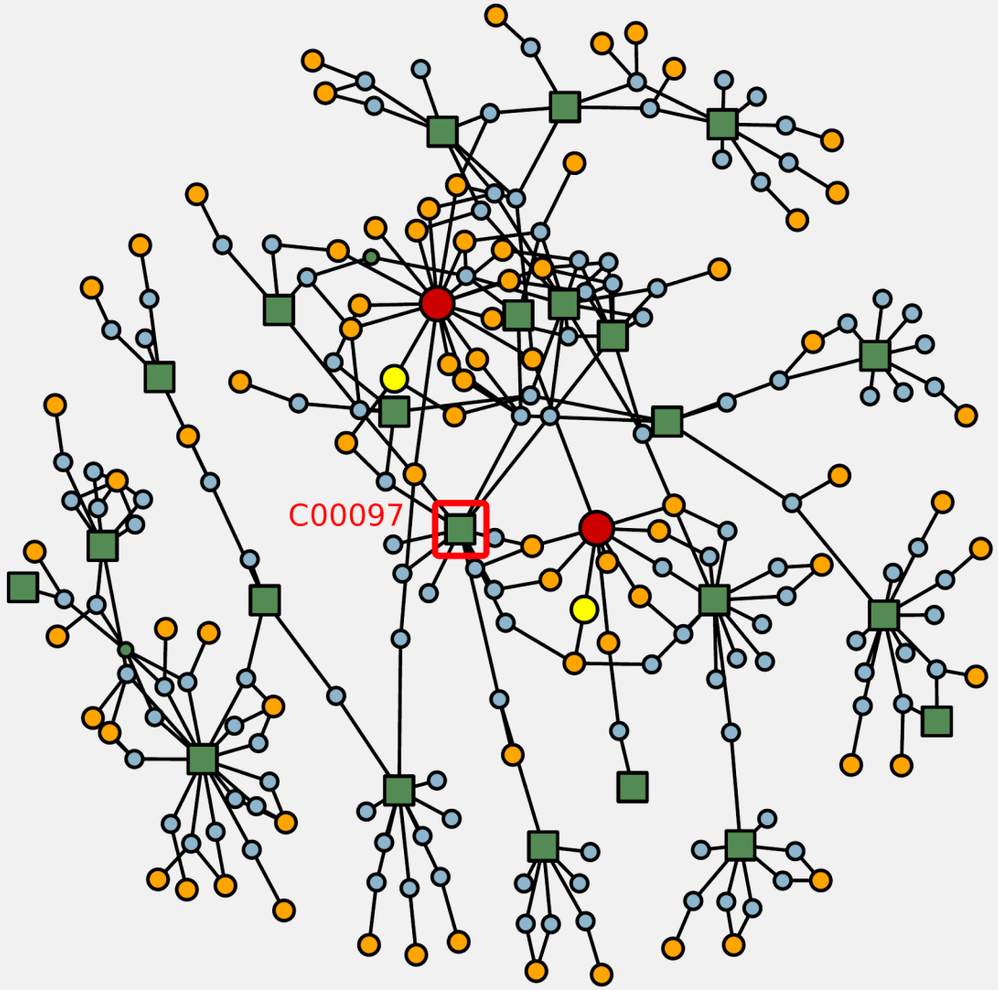
Part of the graph is not observable. There is a compound which is affected but not observed by the experiment (fair statistics)

Null Heat Diffusion Model Test relevant nodes (1/6)



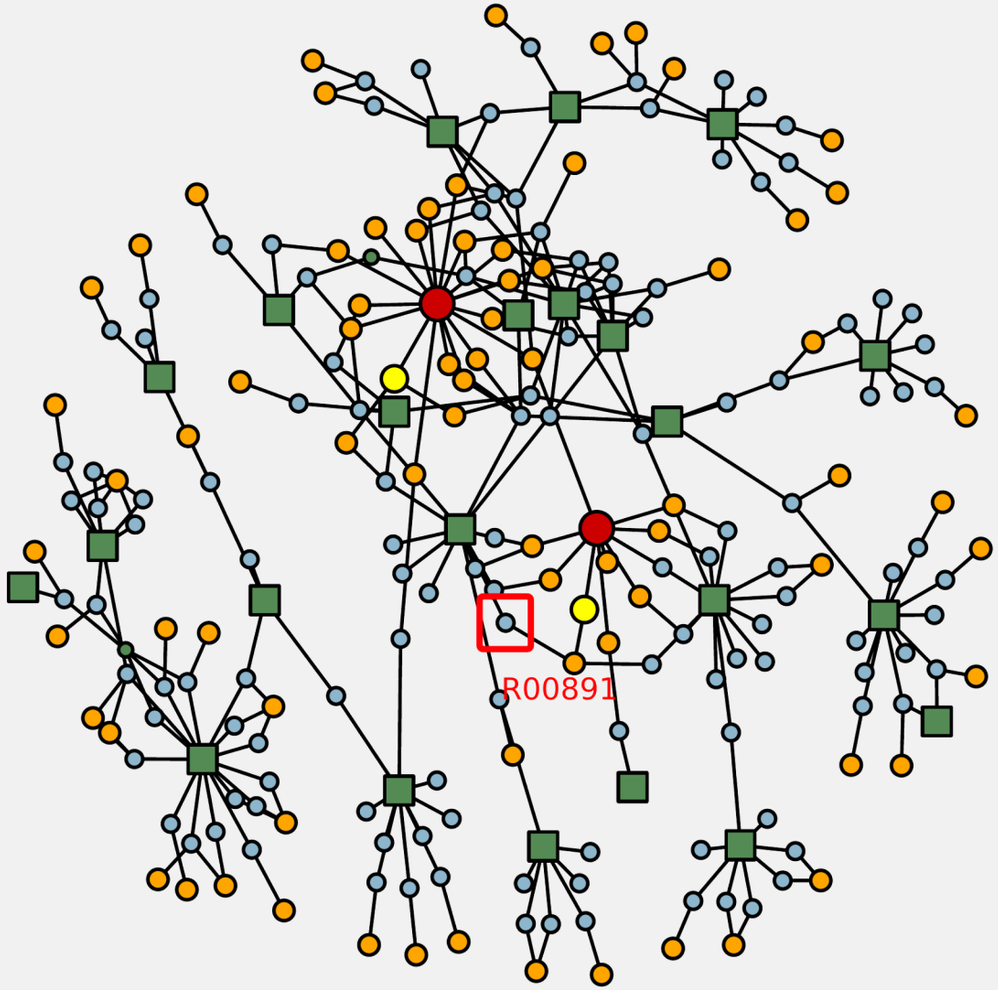
Relevant nodes interpretation.

Null Heat Diffusion Model Test relevant nodes (2/6)



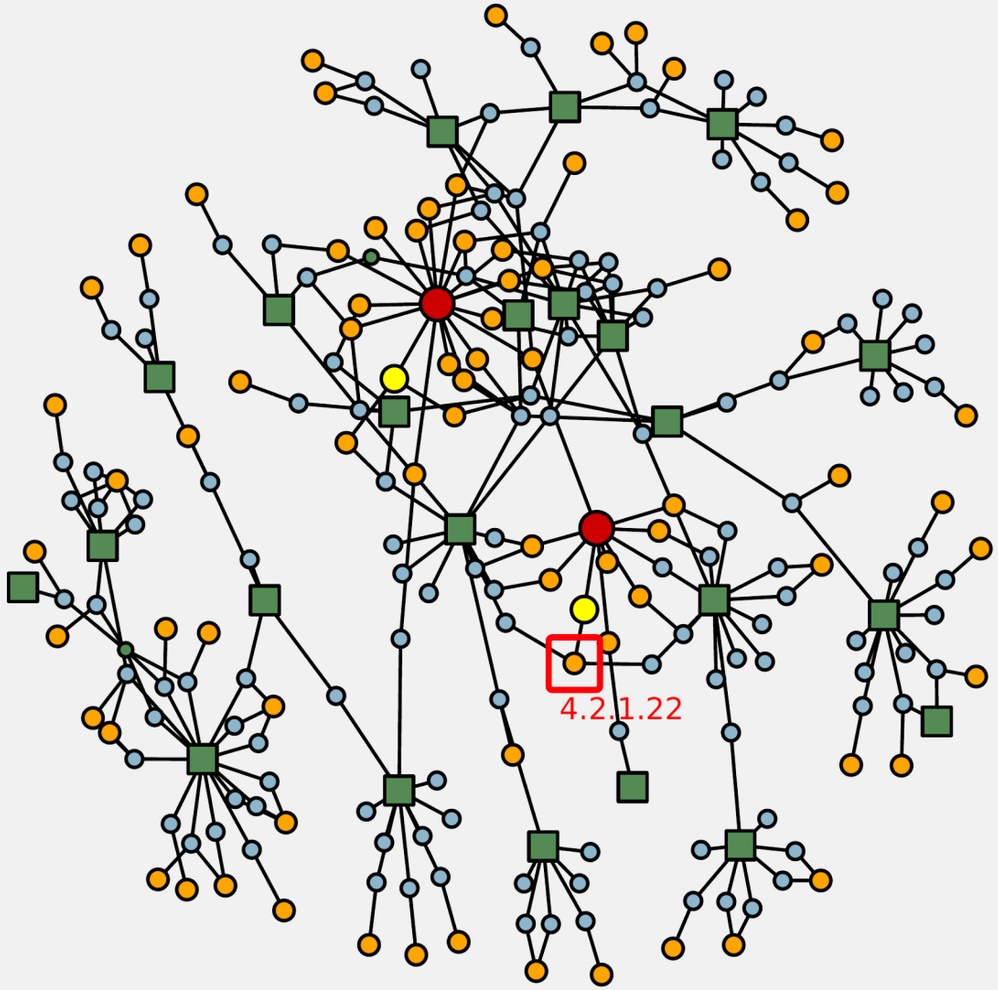
Relevant nodes interpretation.

Null Heat Diffusion Model Test relevant nodes (3/6)



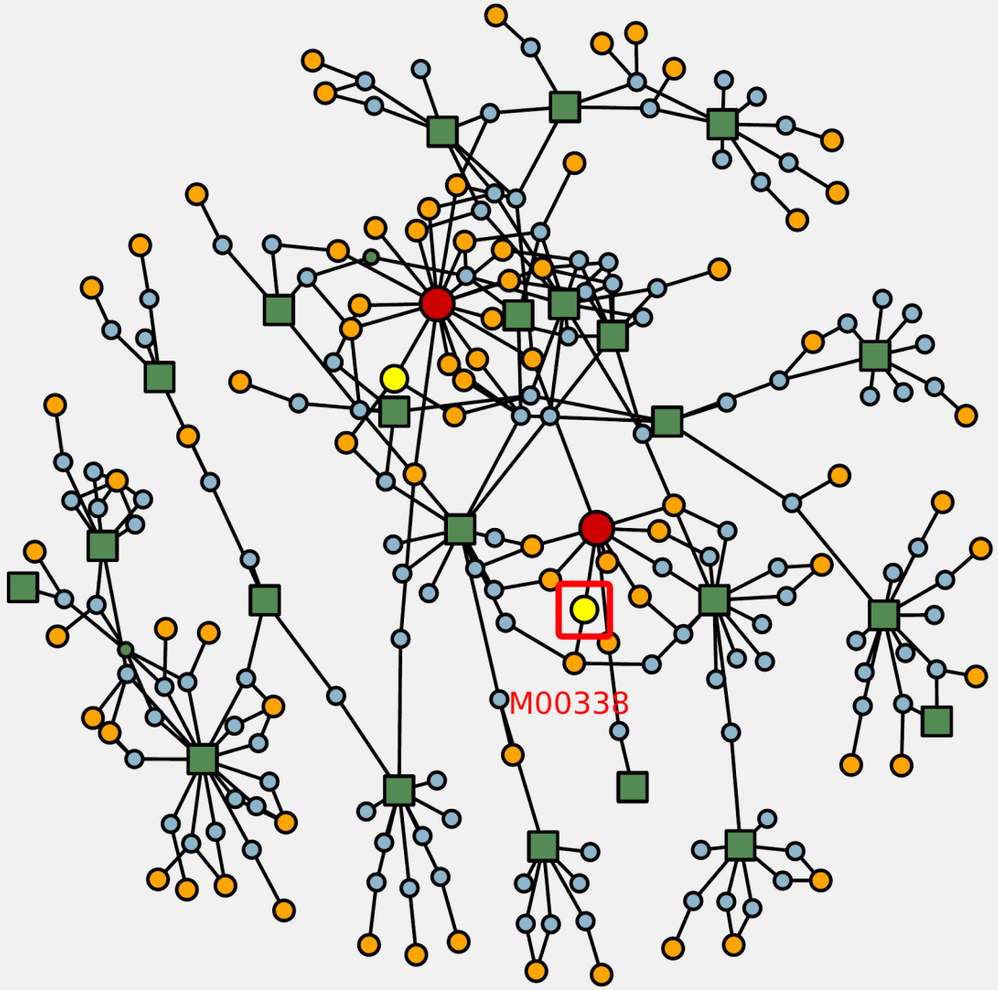
Relevant nodes interpretation.

Null Heat Diffusion Model Test relevant nodes (4/6)



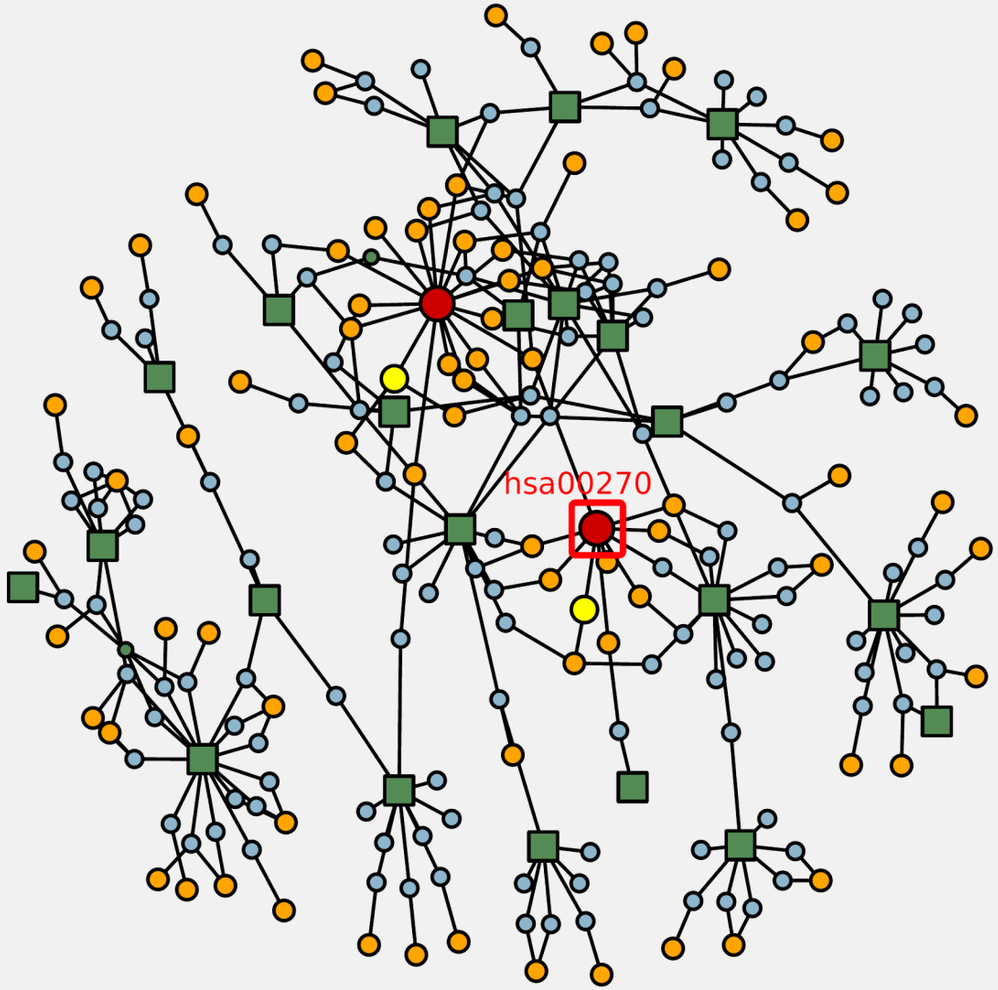
Relevant nodes interpretation.

Null Heat Diffusion Model Test relevant nodes (5/6)



Relevant nodes interpretation.

Null Heat Diffusion Model Test relevant nodes (6/6)



Relevant nodes interpretation.

Heat Diffusion scores (1/4)

Pathway	Pathway description	T [ºC]	Related enzymes
hsa00480	Glutathione metabolism	2.080	2.3.2.4 ; 1.8.1.13
hsa00330	Arginine and proline metabolism	1.458	2.3.1.1 ; 2.7.2.8
hsa00270	Cysteine and methionine metabolism	1.424	3.5.1.31 ; 1.8.4.1
hsa00564	Glycerophospholipid metabolism	1.317	3.1.1.5 ; 3.1.4.46
hsa01100	Metabolic pathways	1.231	2.3.1.1 ; 6.3.2.3

Heat Diffusion scores (2/4)

Pathway	Pathway description	T [ºC]	Related enzymes
hsa00250	Alanine, aspartate and glutamate metabolism	1.226	2.3.1.17 ; 3.5.1.15
hsa00770	Pantothenate and CoA biosynthesis	1.225	2.7.1.33 ; 3.5.1.22
hsa00600	Sphingolipid metabolism	1.205	3.2.1.18 ; 3.1.4.41
hsa00920	Sulfur metabolism	1.158	4.4.1.2 ; 2.5.1.48
hsa00520	Amino sugar and nucleotide sugar metabolism	1.156	2.7.7.43 ; 4.1.3.3

Heat Diffusion scores (3/4)

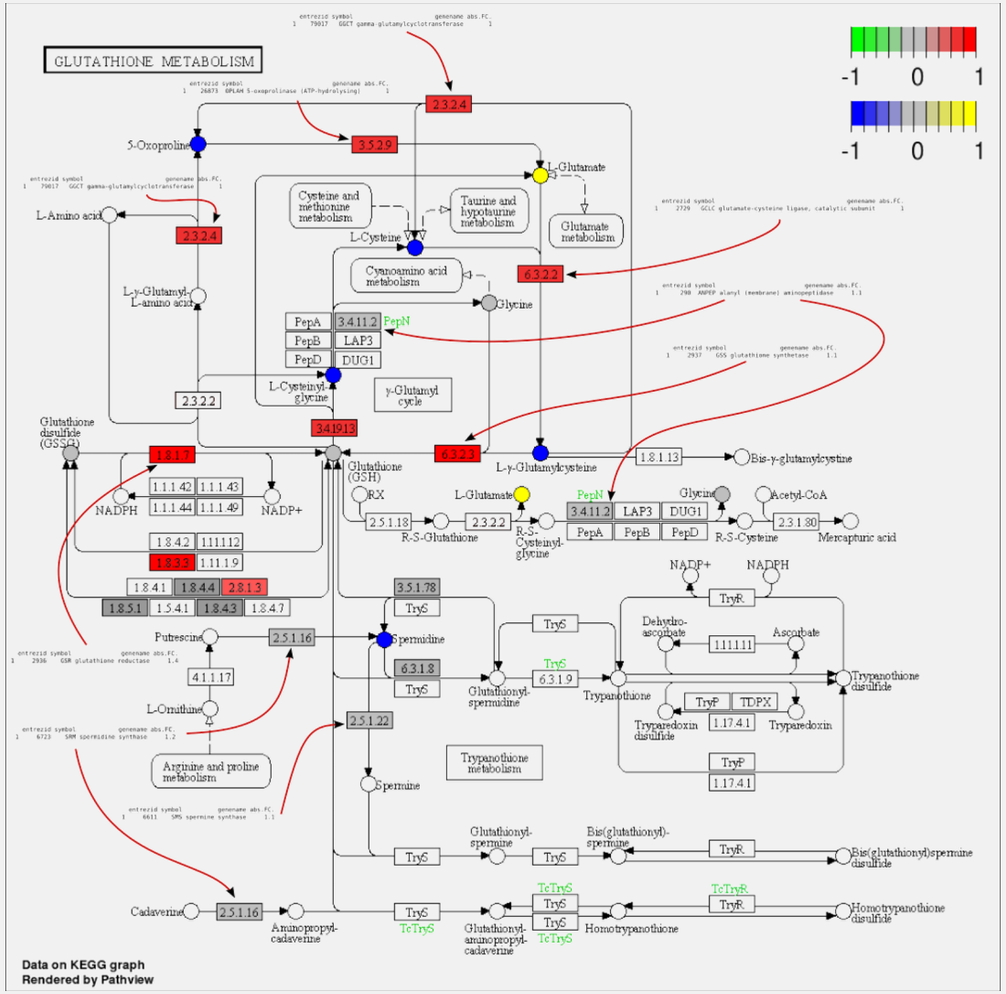
HD algorithm on LIN UCG and 10ºC boundary.

Pathway	Pathway description	T [ºC]	Related enzymes
hsa00500	Starch and sucrose metabolism	1.135	2.4.1.1 ; 2.4.1.17
hsa00360	Phenylalanine metabolism	1.128	1.1.1.222 ; 1.1.1.237
hsa00260	Glycine, serine and threonine metabolism	1.117	1.2.1.31 ; 1.1.1.95
hsa00410	beta-Alanine metabolism	1.115	1.2.1.31 ; 2.5.1.16
hsa00650	Butanoate metabolism	1.103	1.1.99.2 ; 2.8.3.12

Heat Diffusion scores (4/4)

Glutathione metabolism involvement has been validated through isotope tagging at Center for Omics Sciences (Dr. Yanes, YanesLab http://www.yaneslab.com)

Validation


GLUTATHIONE METABOLISM
-1 0 1
-1 0 1
2.3.2.4
3.5.2.9
5-Oxoproline
L-Glutamate
L-Amino acid
2.3.2.4
Cysteine and methionine metabolism
Taurine and hypotaurine metabolism
Glutamate metabolism
L-Cysteine
Cyanoamino acid metabolism
6.3.2.2
Glycine
L-γ-Glutamyl-L-amino acid
PepA 3.4.11.2 PepN
PepB LAP3
PepD DUG1
L-Cysteinyl-glycine
γ-Glutamyl cycle
2.3.2.2
3.4.19.13
Glutathione disulfide (GSSG)
1.8.1.7
6.3.2.3
Glutathione (GSH)
L-γ-Glutamylcysteine
1.8.1.13
Bis-γ-glutamylcystine
NADPH
NADP+
1.1.1.42 1.1.1.43
1.1.1.44 1.1.1.49
RX
2.5.1.18
R-S-Glutathione
L-Glutamate
2.3.2.2
R-S-Cysteinyl-glycine
PepN
3.4.11.2 LAP3 DUG1
PepA PepB PepD
Glycine
Acetyl-CoA
2.3.1.80
R-S-Cysteine
Mercapturic acid
1.8.4.2 1.11.1.12
1.8.3.3 1.11.1.9
1.8.4.1 1.8.4.4 2.8.1.3
1.8.5.1 1.5.4.1 1.8.4.3 1.8.4.7
3.5.1.78
TryS
Putrescine
2.5.1.16
Spermidine
6.3.1.8
TryS
4.1.1.17
L-Ornithine
2.5.1.22
Arginine and proline metabolism
Spermine
Trypanothione metabolism
Glutathionyl-spermidine
TryS
6.3.1.9
Trypanothione
NADP+
NADPH
TryR
Dehydro-ascorbate
1.11.1.11
Ascorbate
Trypanothione disulfide
TryP TDPX
Tryparedoxin disulfide
1.17.4.1
Tryparedoxin
TryP
1.17.4.1
Glutathionyl-spermine
Bis(glutathionyl)-spermine
TryR
Bis(glutathionyl)spermine disulfide
Cadaverine
2.5.1.16
Aminopropyl-cadaverine
TryS
TcTryS
Glutathionyl-aminopropyl-cadaverine
TcTryS
TryS
TryS
TcTryS
Homotrypanothione
TcTryR
TryR
Homotrypanothione disulfide
Data on KEGG graph
Rendered by Pathview

About b2slab.upc.edu

Alexandre Perera Lluna



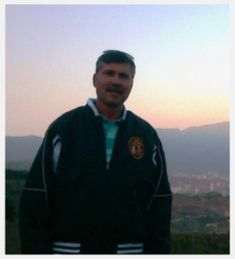
Pere Caminal i Magrans



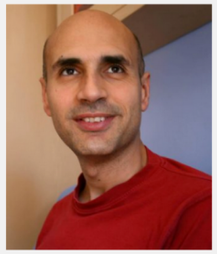
Montserrat Vallverdú



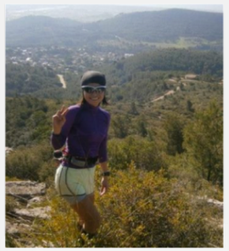
Pedro Gomis



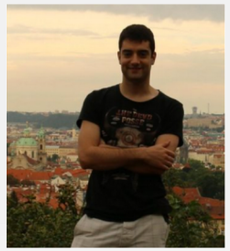
Samir Kanaan Izquierdo



Giovana Gavidia



Sergi Picart Armada



María Maqueda



Rudys Magrans



Jesus Escrivá



Carmen González



Eloy Félix



Imanol Morata



Collaborators in the source edition (1/5)



Takeda Cambridge Ltd.

Collaborators in the source edition (2/5)

COS

Centre for Omic Sciences (COS), Rovira i Virgili University, Avinguda Universitat 1, 43204 Reus, Spain (Dr. Oscar Yanes)

Collaborators in the source edition (3/5)



Psykiatri

Mental Health Center Sct. Hans, Roskilde, Denmark (Dr. Buil)

Collaborators in the source edition (4/5)



The Biomarkers and Nutritional & Food Metabolomics research group from the Department of Nutrition and Food Science of the University of Barcelona (Dr. Rafael Llorach and Dr. Cristina Andrés-Lacueva)

Collaborators in the source edition (5/5)



Genomic Instability and Cancer Laboratory (Dr. Travis Stracker)

About Us b2slab.upc.edu



Questions are welcome



References

References are cited with the relevant figures and methods. Full bibliographic details follow.

References (1)

Tian, Zhang, He et al. (2014). **An organ boundary-enriched gene regulatory network uncovers regulatory hierarchies underlying axillary meristem initiation**. Molecular systems biology.

Booth, Weljie, Turner (2013). **Computational tools for the secondary analysis of metabolomics experiments**. Computational and structural biotechnology journal.

References (2)

Ogata, Goto, Sato et al. (1999). **KEGG: Kyoto Encyclopedia of Genes and Genomes.**. Nucleic Acids Research. Source

Szklarczyk, Franceschini, Wyder et al. (2015). **STRING v10: protein-protein interaction networks, integrated over the tree of life.**. Nucleic Acids Research. Source

References (3)

Consortium (2015). **UniProt: a hub for protein information**. Nucleic Acids Research. Source

Ashburner, Ball, Blake et al. (2000). **Gene Ontology: Tool for the unification of biology**. Nature Genetics.

References (4)

Kamburov, Cavill, Ebbels et al. (2011). **Integrated pathway-level analysis of transcriptomics and metabolomics data with IMPaLA.**. Bioinformatics. Source

Subramanian, Tamayo, Mootha et al. (2005). **Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles**. Proceedings of the National Academy of Sciences. Source

References (5)

Xia, Wishart (2010). **MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data.**. Nucleic Acids Research. Source

Xia, Wishart (2010). **MetPA: a web-based metabolomics tool for pathway analysis and visualization.**. Bioinformatics. Source

References (6)

Xia, Mandal, Sinelnikov et al. (2012). **MetaboAnalyst 2.0 - a comprehensive server for metabolomic data analysis.**. Nucleic Acids Research. Source

Nicholson, Connelly, Lindon et al. (2002). **Metabonomics: a platform for studying drug toxicity and gene function**. Nature Reviews Drug Discovery. Source

References (7)

Hollywood, Brison, Goodacre (2006). **Metabolomics: Current technologies and future trends**. PROTEOMICS. Source

Hosack, Dennis Jr, Sherman et al. (2003). **Identifying biological themes within lists of genes with EASE**. Genome Biol.

References (8)

Chagoyen, Pazos (2012). **Tools for the functional interpretation of metabolomic experiments**. Briefings in Bioinformatics.

Barabási, Gulbahce, Loscalzo (2011). **Network medicine: a network-based approach to human disease**. Nature Reviews Genetics.

References (9)

Ideker, Ozier, Schwikowski et al. (2002). **Discovering regulatory and signalling circuits in molecular interaction networks**. Bioinformatics.

Draghici, Khatri, Tarca et al. (2007). **A systems biology approach for pathway level analysis**. Genome Research.

References (10)

Kankainen, Gopalacharyulu, Holm et al. (2011). **MPEA—metabolite pathway enrichment analysis**. Bioinformatics.

Aggio, Ruggiero, Villas-Bôas (2010). **Pathway Activity Profiling (PAPi): from the metabolite profile to the metabolic pathway activity**. Bioinformatics.

References (11)

Joyce, Palsson (2006). **The model organism as a system: integrating'omics' data sets**. Nature Reviews Molecular Cell Biology.

Vandin, Upfal, Raphael (2011). **Algorithms for detecting significantly mutated pathways in cancer**. Journal of Computational Biology.

References (12)

Sugimoto, Kawakami, Robert et al. (2012). **Bioinformatics tools for mass spectroscopy-based metabolomic data processing and analysis**. Current Bioinformatics.