

CONNECTING THE DOTS: METABOLIC NETWORKS FOR METABOLOME MINING

Fabien JOURDAN

Senior Researcher

French Agronomic Research Institute (INRA) Toulouse, FRANCE

 @MetExplore

 Fabien.Jourdan@inra.fr

Overview of the course

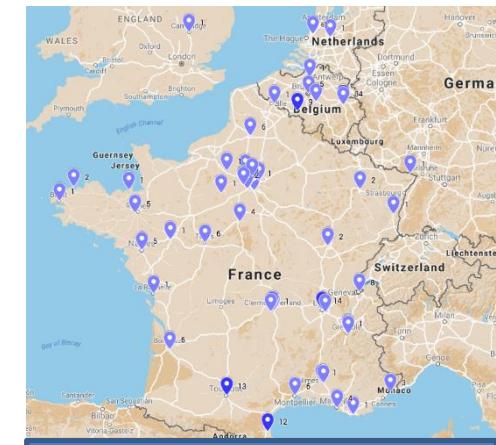
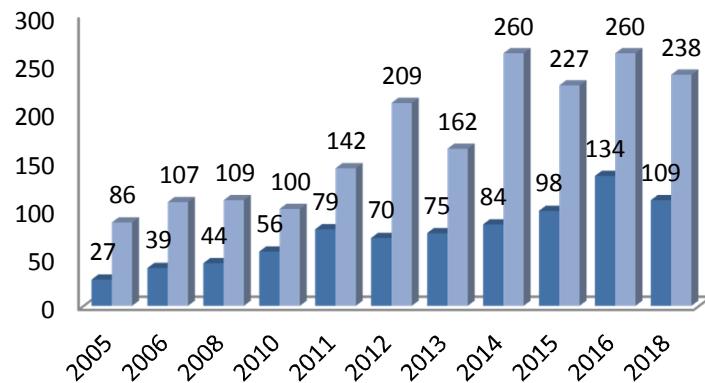
- ” Some words of Metabolomics in France ... and in Toulouse
- ” What is metabolomics, from raw data to metabolic profiles
- ” Metabolomics data analysis in the context of metabolic networks
 - . Gather metabolic knowledge - *Reconstruction*
 - . Integrate metabolomics data – *Mapping*
 - . **PRACTICE 1**
 - . Model global metabolism - *Graphs*
 - . Suggest interpretation – *Algorithms*
 - . **PRACTICE 2**
 - . Suggest metabolites – *Algorithms*
 - . **PRACTICE 3**
- ” **QUIZZ**
- ” Getting closer to phenotype: transcriptomics and human cellular metabolism

METABOLOMICS IN FRANCE ... AND IN TOULOUSE

French-speaking Metabolomics and Fluxomics network (since 2005, affiliated to Met Soc since 2013)

- “ to make an inventory and promote French skills in the fields of metabolomics and fluxomics
- “ to provide and support scientific meetings or workshops in metabolomics and fluxomics
- “ to facilitate knowledge transfer to students and newcomers in the field and help students to promote their work

1 conference/year
 >300 membres
 >600 scientists registered to the mailing list
 >10 travelling grants/year
 1 PhD best PhD award/year
 1 Mantra : « Good food and Good science »



Mailing list (>600 persons): rwmf@listes.inra.fr
 Travel grants for young French Speaking scientists



Fabien JOURDAN © 2019



21 AU 23 MAI 2019

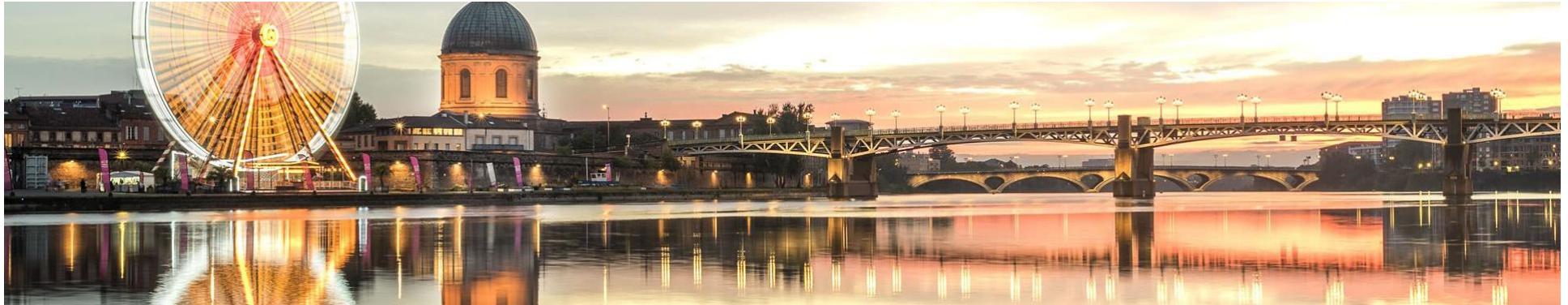
POLYDOME, CLERMONT-FERRAND

www.rfmf.fr

Guest speakers:

- ” Konstantinos Aliferis
- ” Coral Barbas
- ” Pieter Dorrestein
- ” Claudio Luchinat





Institute: INRA Division: Human Nutrition

Laboratory: ToxAlim, research centre on food toxicology, Toulouse

Modeling group:

Fabien JOURDAN DR INRA – Math/Comp. science...Word/Excel

Nathalie POUPIN CR INRA – Modelling/Biology

Florence VINSON IE INRA - *Comp. science*

Clément FRAINAY CR INRA- Math/Comp. science

Maxime CHAZALVIEL IE CDI medDay - *Comp. science*

Pablo RODRIGUEZ MIER Postdoc INRA - *Comp. science/modelling*

Joran VILLARET PhD candidate – *Biology*

A map of Southern France and Northern Spain, centered on Toulouse. The map shows the coastline of the Mediterranean Sea to the south and the Atlantic Ocean to the west. Major cities like Paris, Lyon, and Marseille are visible. The road network is depicted with a dense grid of yellow lines. A red dot marks the location of Toulouse. The Pyrenees mountain range is shown as a dark line along the southern border.

mainly dry lab

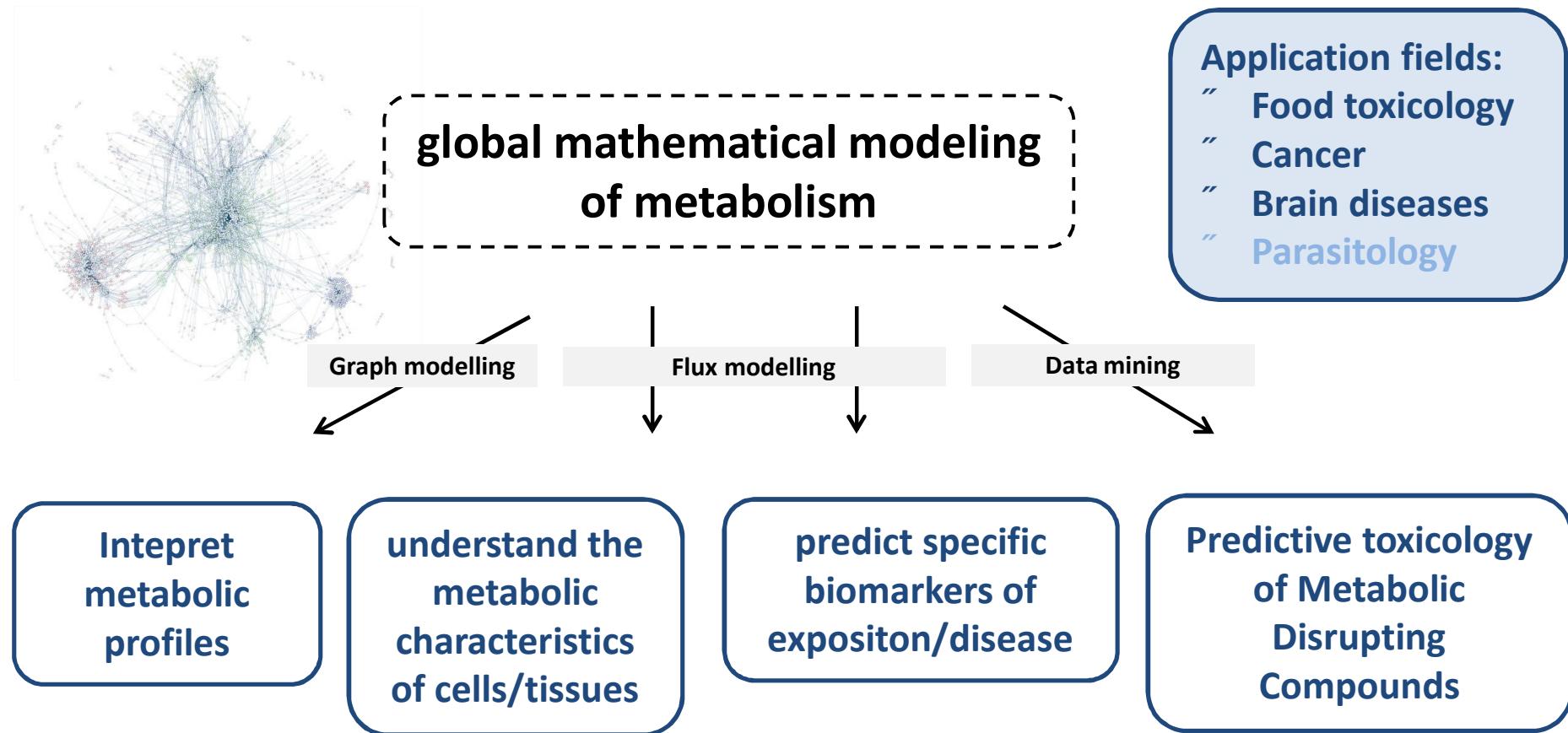
```
public TransformationTable(Vector allTransfo,  
    super(new GridLayout(1,0));  
    massNetwork=m;  
    allTransformationMasses=allTransfo;  
    data=new Object[allTransformationMasses.size()  
        : allTransformationMasses.size();  
    11Transfor
```



Positions available in H2020 project:

- Software engineer
 - Research Assistant in MS Metabolomics

Modeling approaches for global metabolic studies



Web server development (www.metexplore.fr)



MetExplore : omics data analysis in genome scale networks

Nucleic Acids Research

Issues Section browse ▾ Advance articles Submit ▾ Purchase About ▾ All Nucleic Acids Re



Volume 46, Issue W1
2 July 2018

MetExplore: collaborative edition and exploration of metabolic networks

Ludovic Cottret , Clément Frainay, Maxime Chazalviel, Floréal Cabanettes, Yoann Gloaguen, Etienne Camenen, Benjamin Merlet, Stéphanie Heux, Jean-Charles Portais, Nathalie Poupin, Florence Vinson, Fabien Jourdan 

Nucleic Acids Research, Volume 46, Issue W1, 2 July 2018, Pages W495–W502,
<https://doi.org/10.1093/nar/gky301>

Published: 30 April 2018 Article history ▾



The project

“Publications :

- “Cottret et al (2018). *Nucleic Acids Research*
- “Chazalviel et al (2017). *Bioinformatics*
- “Frainay et al (2018). *Bioinformatics*

“Number of citations: >140

“Metrics:

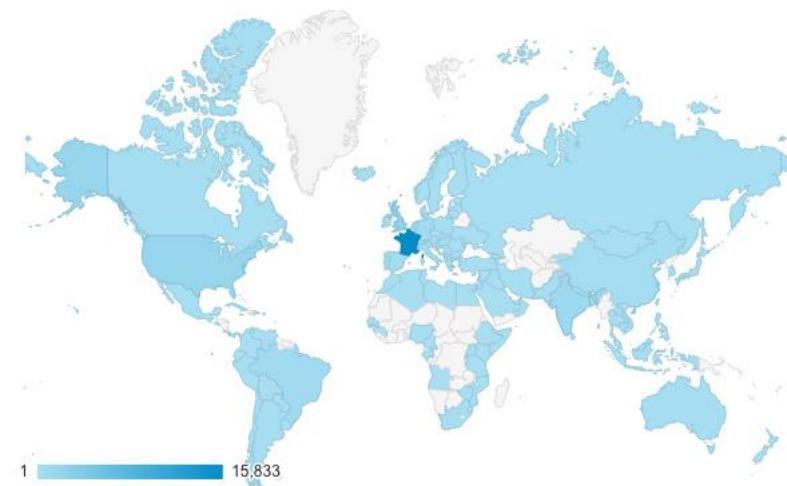
- “**842** registered users,
- “> **1300** networks
- “> **540** persons trained

“Involved in several national and EU grants

“1 industrial partner (MedDay pharma)

Website

<http://www.metexplore.fr/>

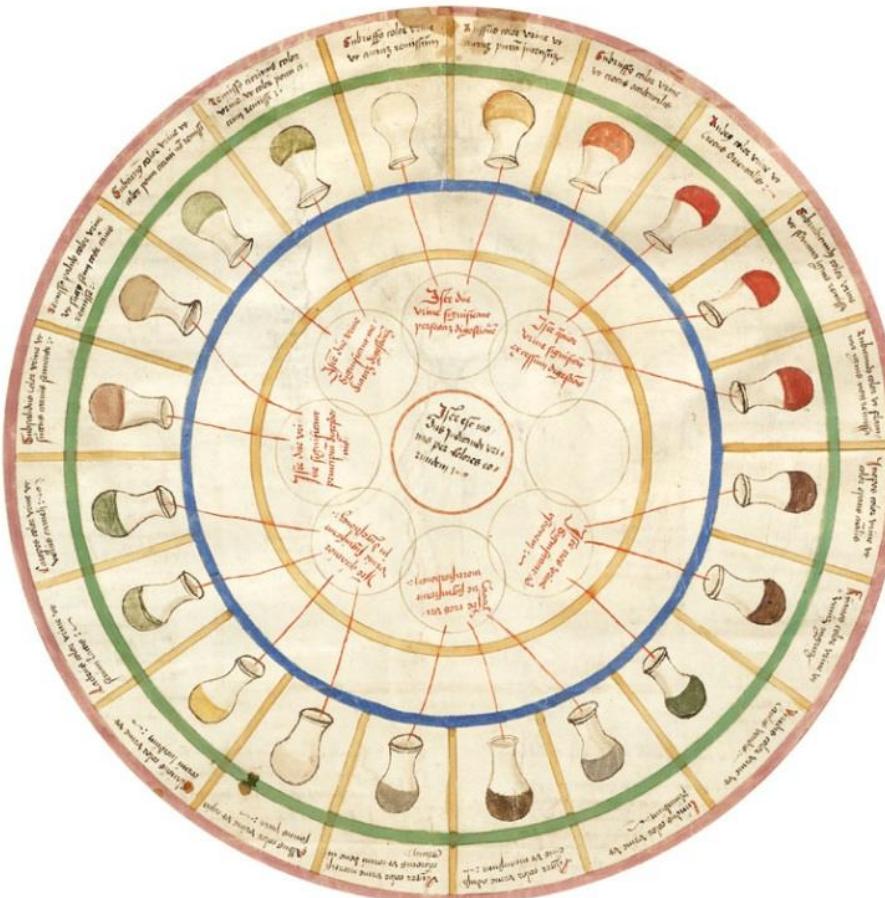


OVERVIEW ON METABOLOMICS: FROM RAW DATA TO METABOLIC PROFILES

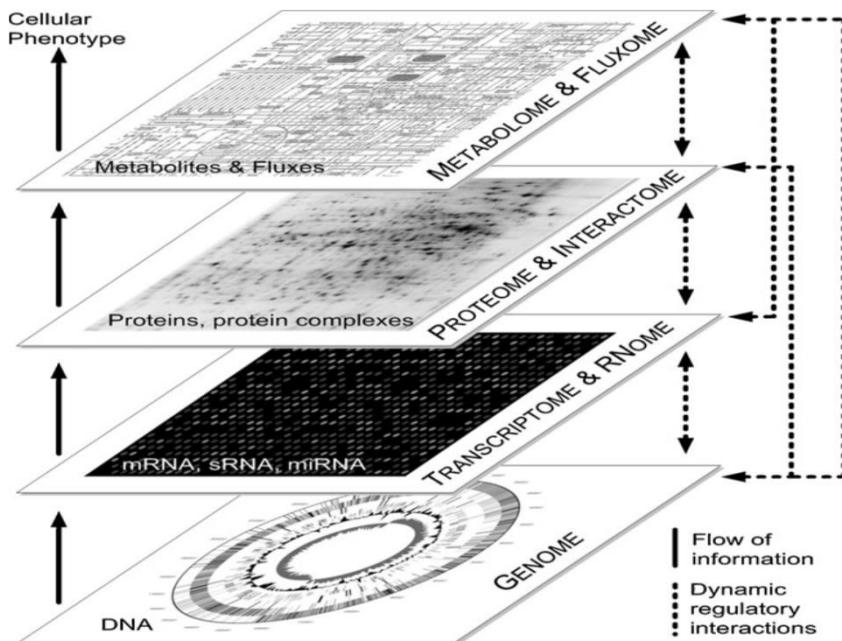
From metabolome to interpretation/prediction

"This urine wheel was published in 1506 by Ullrich Pinder, in his book Epiphanie Medicorum. It describes the possible colours, smells and tastes of urine, and uses them to diagnose disease."
THE ROYAL LIBRARY, COPENHAGEN

Nicholson JK, Lindon JC. Nature 2008; 455:1054–1056



Comprehensive analysis of metabolic networks



Kohlstedt et al, 2010; Sauer et al, 2004

- Readout of the metabolic network
- Response to genetic/environmental perturbations
- Systems biology (& synthetic biology)
- Identification of key regulated sites in networks
- Investigation of gene function
- Phenotyping: next to any observable phenotype: diagnostics, functional genomics....

Meet the human metabolome

Imagine that at a routine medical check-up your doctor takes a urine sample, then reports a few days later that your risk of type 2 diabetes is normal, but there are hints that your arteries are furring up.

A similar scenario has been promised for the past 20 years by those working in genomics and proteomics, but has not yet materialized. Now, however, an increasing number of researchers are claiming that metabolomics — the study of all the body's metabolites — will finally come up with the goods.

Supporters of this burgeoning branch of molecular medicine are gung-ho about their chances of success. "In retrospect, we wonder why we spent millions on the genome," says Bruce German, who studies lipid metabolism at the University of California, Davis. With the knowledge we have today, he reckons, scientists should have gone straight for the metabolome. But can it deliver?

Metabolomics is the study of the raw materials and products of the body's bio-chemical reactions, molecules that are smaller than most proteins, DNA and other macromolecules. The aim is to be able to take urine, blood or some other body fluid, scan it in a machine and find a profile of tens or hundreds of chemicals that can predict whether an individual is on the road to a disease, say, or likely to experience side-effects from a particular drug.

Researchers are already trying to flag impending disease by measuring levels of gene expression or proteins, but supporters of metabolomics say they should be able to do it better. Small changes in the activity of a gene or protein (which may have an unknown impact on the workings of a cell) often create a much larger change in metabolite levels. The approach has already proved its worth: cholesterol and glucose have long been chemical canaries for heart disease and diabetes.

Data fingerprinting

But realizing this vision isn't straightforward. One of the first tasks is to create a catalogue of compounds in the human body, and this is proving hard to define. David Wishart at the University of Alberta, Edmonton, and his colleagues have taken an initial step forward by producing something they rather grandly call the first draft of the human metabolome¹. They searched the published literature for known human metabolites, and have collected around 2,500 of them into a public database (www.hmdb.ca) along with other information

such as known links to disease. The researchers also used nuclear magnetic resonance and mass spectroscopy to produce characteristic 'fingerprints' for more than 400 compounds, and have added these to the database.

It's the most comprehensive collection of metabolite data to be made publicly available. But others in the field point out that Wishart's catalogue is far from complete because the number and nature of compounds in the human metabolome will vary depending on which body fluid is looked at and the method used for the analysis. There is also no clear division between compounds produced by the human body, those produced by our gut bacteria and fleeting products generated by food or drugs swallowed that day.

"The notion that this is a first draft of the human metabolome is nonsense," says Jeremy Nicholson of Imperial College London, one of the pioneers of the field. "I agree that it covers a helluva lot of important metabolites, but it's a very arbitrary guess at what might be useful and what might not."

An added complication is that one person's profile of metabolites is likely to be dramatically different from another's, and each may fluctuate markedly depending on the time of day, what they last ate and other aspects of their lifestyle. To get a handle on this variation, Nicholson has studied tens of thousands of urine samples from many ethnic groups around the world and found that each group is remarkably different. A separate study showed that the metabolic profile of meat eaters is very different from that of vegetarians². This means that a person's metabolome might need to be measured many times during their lives in order to be able to pick out changes that might

PHOTO: DAVID MCKEE/PHOTOBEST

Hopes are high that in future a urine sample might reveal our health profile from our metabolites.

signal disease. Also, any one metabolite will have to be assessed relative to the pattern of many others.

Whatever the total metabolite tally, researchers will have to prove that particular concentrations and combinations can reveal something about drugs or disease. Preliminary studies suggest that this can be done. Last year, for example, Nicholson and his colleagues showed that a fingerprint of the metabolites in urine could predict which rats would suffer liver damage from the drug paracetamol³. He says he has now shown the same in humans, and is preparing the study for publication. A team led by Douglas Kell at the University of Manchester, UK, has developed a computer model based on metabolite profiles in blood plasma that can identify pregnant women with the dangerous condition called pre-eclampsia⁴.

Lab on a chip

But just as with genomics and proteomics, finding profiles that reliably predict the onset of disease will be a major undertaking, because it will typically require sampling regular profiles from many thousands of people and then following them for years to see which ones develop a particular condition. Researchers are hopeful that biobanks — large collections of people's biological samples and medical records — will in future supply this information, but such studies could take decades. They will also have to learn from those working on gene-expression or proteomic profiles, who have sometimes struggled to show that a test that works in one group also works in another, or that the changes they see are actually involved in a disease. "In my opinion we should know why this metabolite is going up or down," says William Bigbee, an expert in biomarkers and proteomics at the University of Pittsburgh Cancer Institute, Philadelphia.

In the long run, the best way to predict an individual's disease risk is likely to come from understanding the biology behind each disease — and that will come from a combination of genomics, proteomics and metabolomics. "I don't want to buy six machines," says Ben van Ommen of the Netherlands Organization for Applied Scientific Research in Zeist. "I want a lab on a chip that measures metabolites, proteins and gene expression." ■

Helen Pearson

1. Wishart, D. S. et al. *Nucleic Acids Res.* **35**, D521–D526 (2007).
2. Stella, C. et al. *J. Proteome Res.* **5**, 2780–2788 (2006).
3. Clayton, T. A. et al. *Nature* **440**, 1073–1077 (2005).
4. Kenny, L. C. et al. *Metabolomics* **1**, 227–234 (2005).

The approach has already proved its worth: cholesterol and glucose have long been chemical canaries for heart disease and diabetes.

Recent omics...old science

**pect, we
hy we
lions on
ne."**

human metabolome will vary depending on which body fluid is looked at and the method used for the analysis. There is also no clear division between compounds produced by the human body, those produced by our gut bacteria and fleeting products generated by food or drugs swallowed that day.

Strong variability

But just as with genomics and proteomics, finding profiles that reliably predict the onset of disease will be a major undertaking, because it will typically require sampling regular profiles from many thousands of people and then following them for years to see which ones develop a particular condition. Researchers

Requires biobanks and statistics

In the long run, the best way to predict an individual's disease risk is likely to come from understanding the biology behind each disease

Biological interpretation is essential



Applications



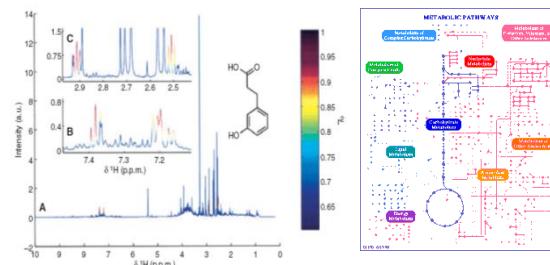
**Health
Personalized medicine**



Biotechnology



Pharmacology



Basic & Applied Microbiology



Nutrition & agrofood industry



Agronomy

Practical clinical usage



Science Translational Medicine

Home News Journals Topics Careers

Science Science Advances Science Immunology Science Robotics Science Signaling Science Translational Medicine

SHARE RESEARCH ARTICLE CANCER DIAGNOSTICS



8

Intraoperative Tissue Identification Using Rapid Evaporative Ionization Mass Spectrometry

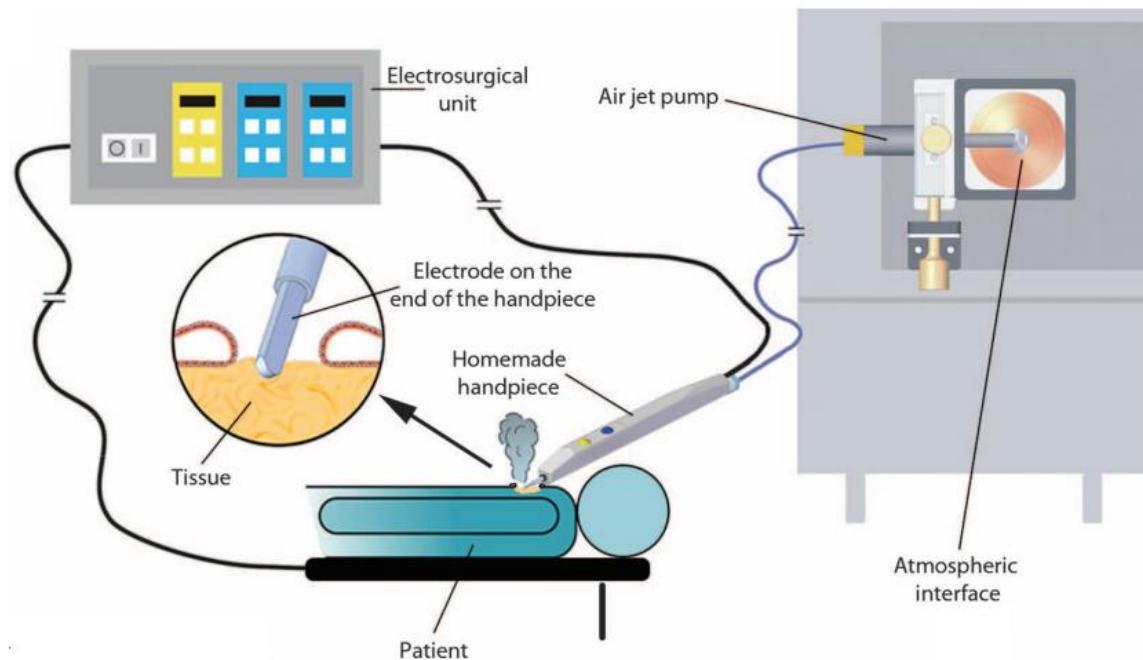
Júlia Balog^{1,*}, László Sasi-Szabó^{2,*}, James Kinross^{3,4}, Matthew R. Lewis³, Laura J. Muirhead^{3,4}, Kirill Veselkov³, Reza Mirnezami⁴, Balázs Dezső⁵, László Damjanovich², Ara Darzi⁴, Jeremy K. Nicholson^{3,†} and Zoltán Takáts^{3,‡}

+ Author Affiliations

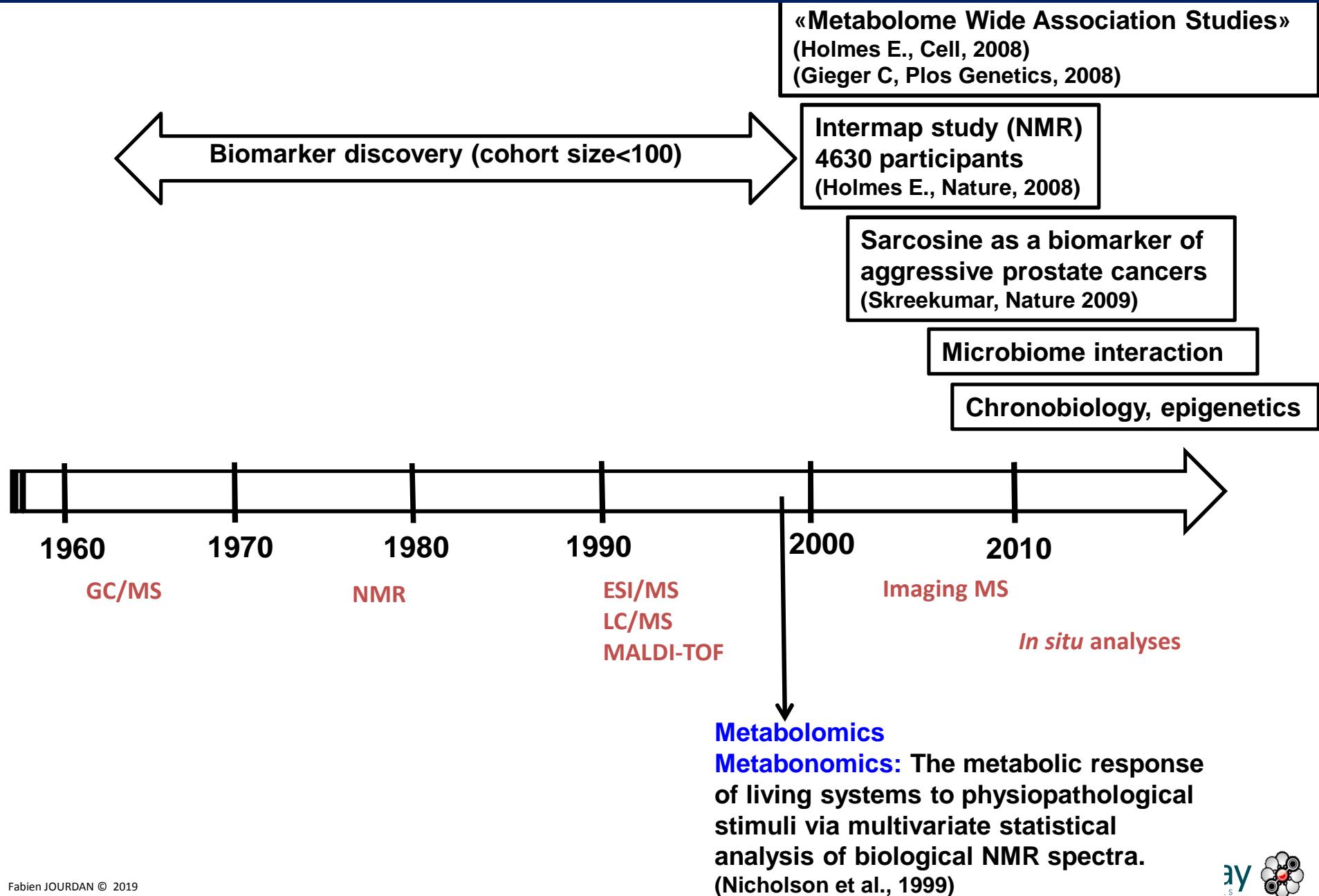
*Corresponding author. E-mail: z.takats@imperial.ac.uk (Z.T.); j.nicholson@imperial.ac.uk (J.K.N.)

‡ These authors contributed equally to this work.

Science Translational Medicine 17 Jul 2013;
Vol. 5, Issue 194, pp. 194ra93
DOI: 10.1126/scitranslmed.3005623



History



The world of small molecules

Metabolome

Stephen Oliver (1998, UK)

- Metabolome : “the set of all low-molecular weight compounds synthesized by an organism”.

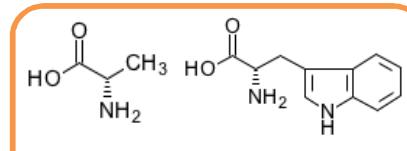
Oliver Fiehn (Germany, 2002)

- Metabolomics (Strict definition): comprehensive analysis to identify and quantify all metabolites of a biological system.

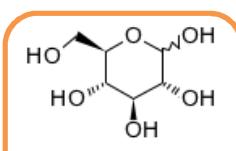
The metabolome include peptides, lipids, amino acids, nucleic acids, carbohydrates, organic acids, vitamins, minerals, food additives, drugs, toxins, pollutants etc that biological systems ingest, metabolize, catabolize or come into contact with.

Metabolome diversity

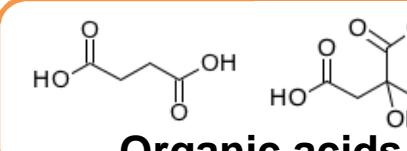
Primary metabolites



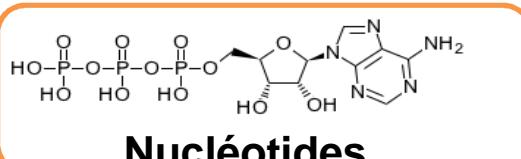
Aminoacids



Sugars

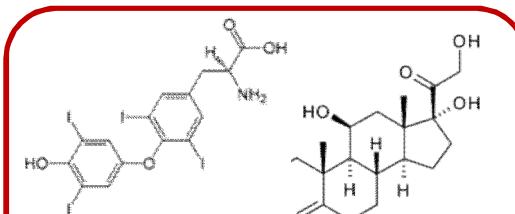


Organic acids



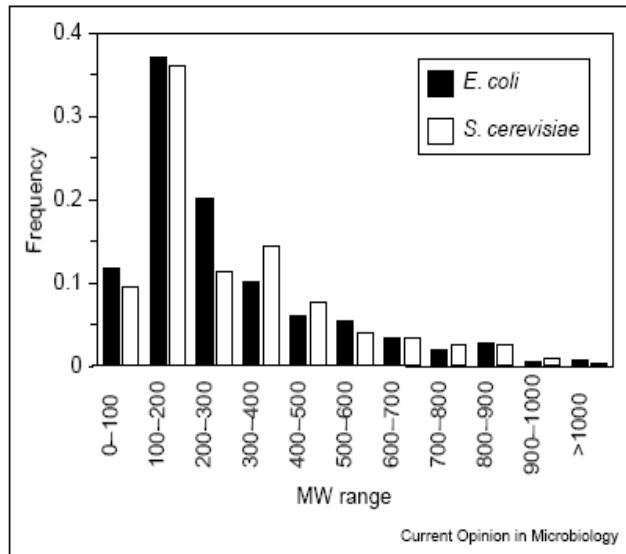
Nucléotides

Secondary metabolites



The world of small molecules

Molecular mass range



80% of hydrophilic metabolites have a molecular mass ≤ 600 (*E. coli*, *S. cerevisiae*)

Physico-chemical diversity

- Comparison with transcriptomics/proteomics
 - ADN/ARN: 4 bases
 - Proteins: 20 amino-acids
 - Common physico-chemical properties
 - Extraction/analysis: easiest automation

Metabolites

Number $\geq 150\,000$ in Nature (most are unknown)

Most of them not identified
broad chemical diversity
High turn-over rates (<sec)
Dynamic range:

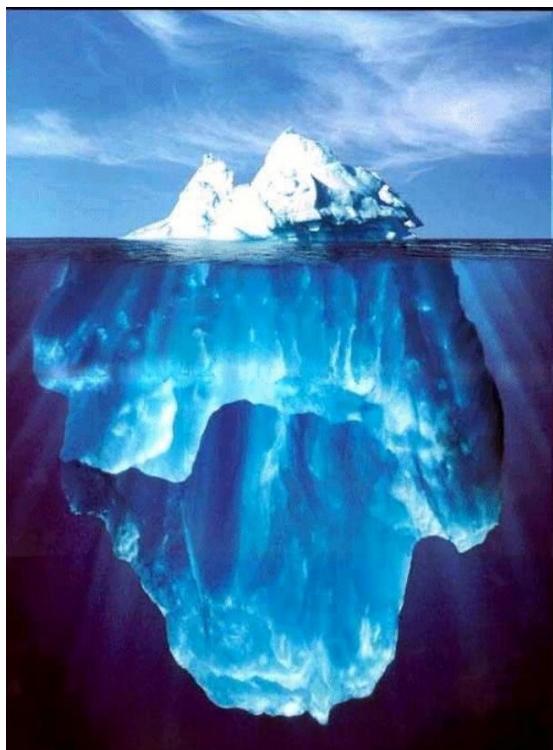
Major constituents (sugars, etc.): 10-100 mM
Minor constituents (vitamins, etc.): pM



The metabolome analysis relies on combinations of approaches

The world of small molecules

Dynamic range of the metabolome (concentration range)



$\sim 0,1 \text{ M}$

(major components, carbon sources, sugars, etc)

9 decades

= dynamic range of the metabolome in a single sample

(the dynamic range of a MS detector is 4-5 decades)

$\sim 10^{-10} \text{ M}$

(minor components, bioeffective compounds: signal molecules, etc)

NMR & MS are the main platforms for metabolomics

NMR



Advantages

- Easy sample preparation (biofluids)
- Analysis of complex mixtures without separation or a priori consideration
- Detailed structural information
- Multi-nuclear: ^1H , ^{13}C , ^{31}P , ^{15}N , etc
- Isotopic analysis (stable isotopes)
- *in vivo /in situ* analysis (non invasive & non destructive)
- (Can be coupled to LC)
- Robust, reliable (HT analysis)

Drawbacks

- Limited sensitivity (micro to nano-moles):
access to major metabolites only
- Cost of equipments
(analysis not necessarily expensive)
- Dynamic range

MS



Advantages

- Mass information (identification)
- Flexibility brought by the various types of MS detectors
- High sensitivity (10s attomole)
- Isotopic analysis (stable/radioactive)
- Complex mixtures: need for separation: GC/LC/CE

Drawbacks

- Complex mixtures: need for separation: GC/LC/CE
(Cost of equipments)
- Lack of robustness for fingerprinting
- Lack of reproducibility between equipments (LC-MS)

Combining analytical platforms: improved data consistency

frontiers in
PLANT SCIENCE

TECHNLOGY REPORT
published: 10 February 2012
doi: 10.3389/fpls.2012.00015

Metabolomics as a hypothesis-generating functional genomics tool for the annotation of *Arabidopsis thaliana* genes of "unknown function"

Stephanie M. Quanbeck¹, Libuse Brachova¹, Alexis A. Campbell¹, Xin Guan¹, Ann Perera¹, Kun He², Seung Y. Rhee², Preeti Bais³, Julie A. Dickerson³, Philip Dixon⁴, Gert Wohlgemuth⁵, Oliver Fiehn⁶, Lenore Barkan⁶, Iris Lange⁶, B. Markus Lange⁶, Insuk Lee⁷, Diego Cortes⁸, Carolina Salazar⁹, Joel Shuman¹⁰, Vladimir Shulaev⁹, David V. Huhman¹¹, Lloyd W. Sumner¹¹, Mary R. Roth¹², Ruth Welti¹², Hilal Ilarslan¹³, Eve S. Wurtele¹³ and Basil J. Nikolau^{1*}



Analysis of *A. thaliana* metabolome

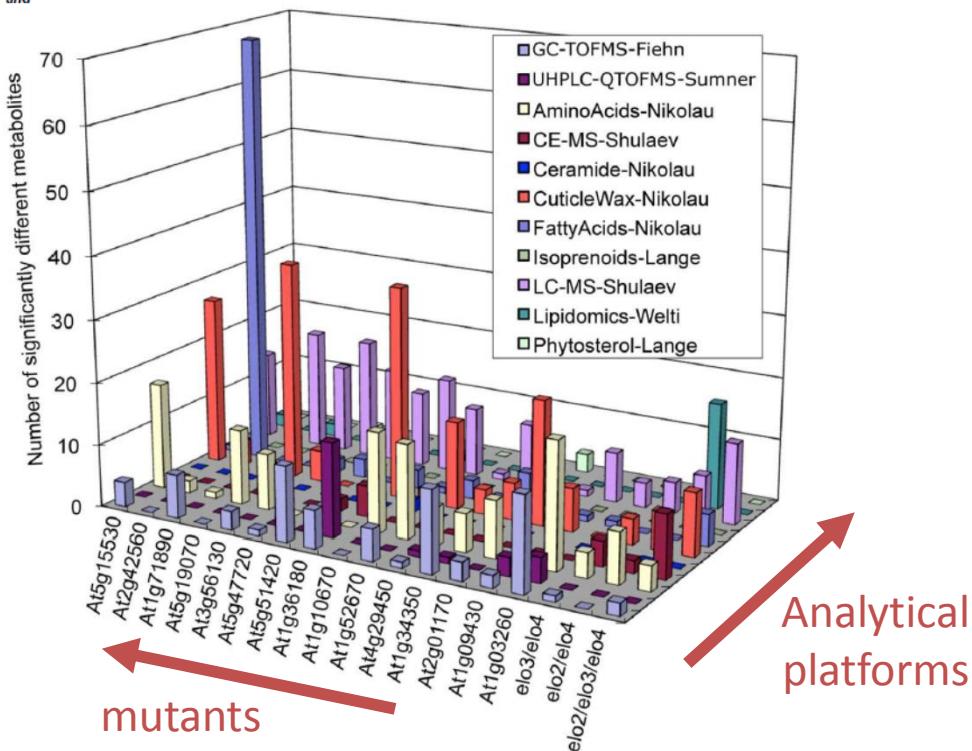
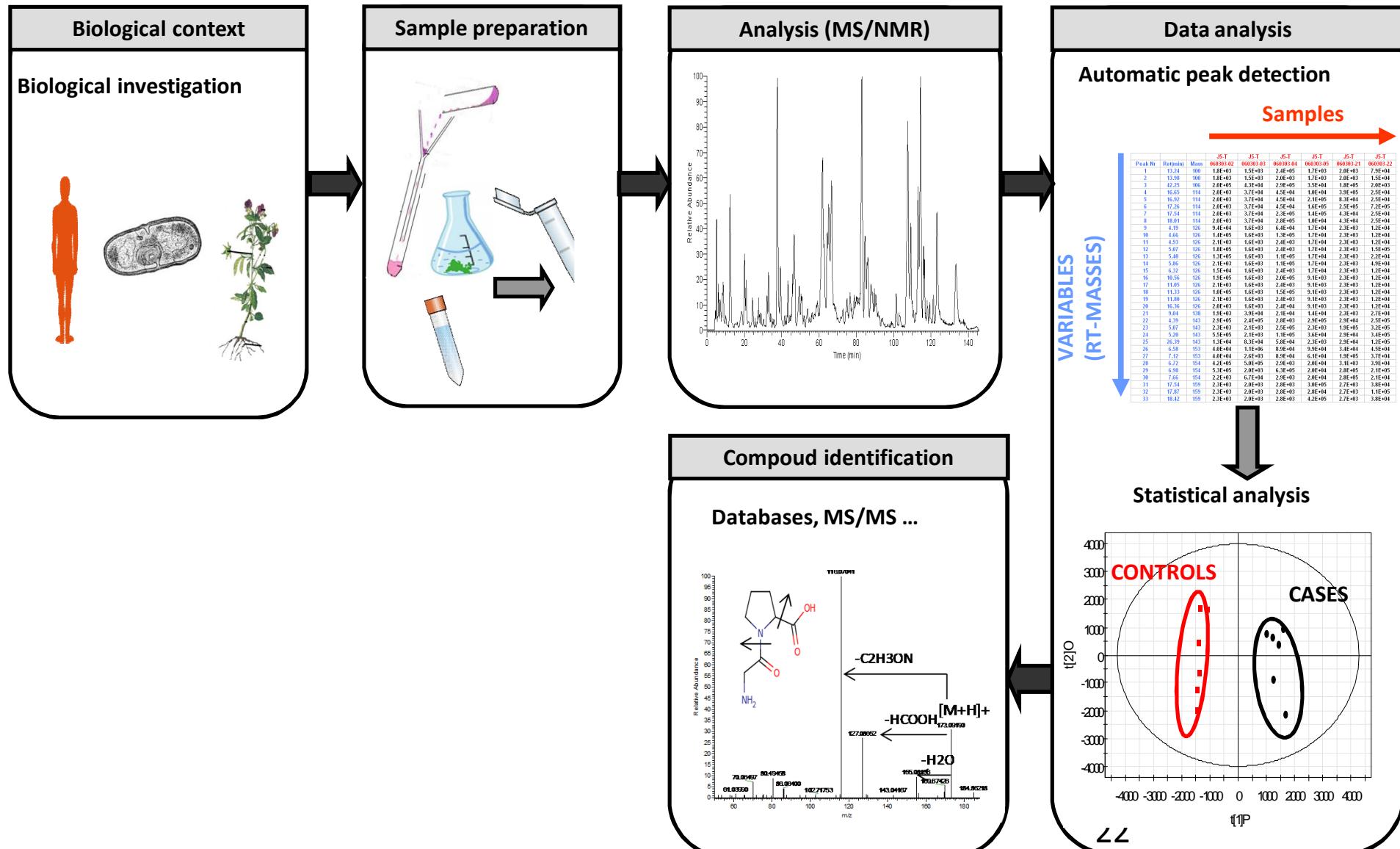


FIGURE 1 | Distribution of significantly altered metabolites among different mutants as detected by different analytical platforms (identified in the insert).

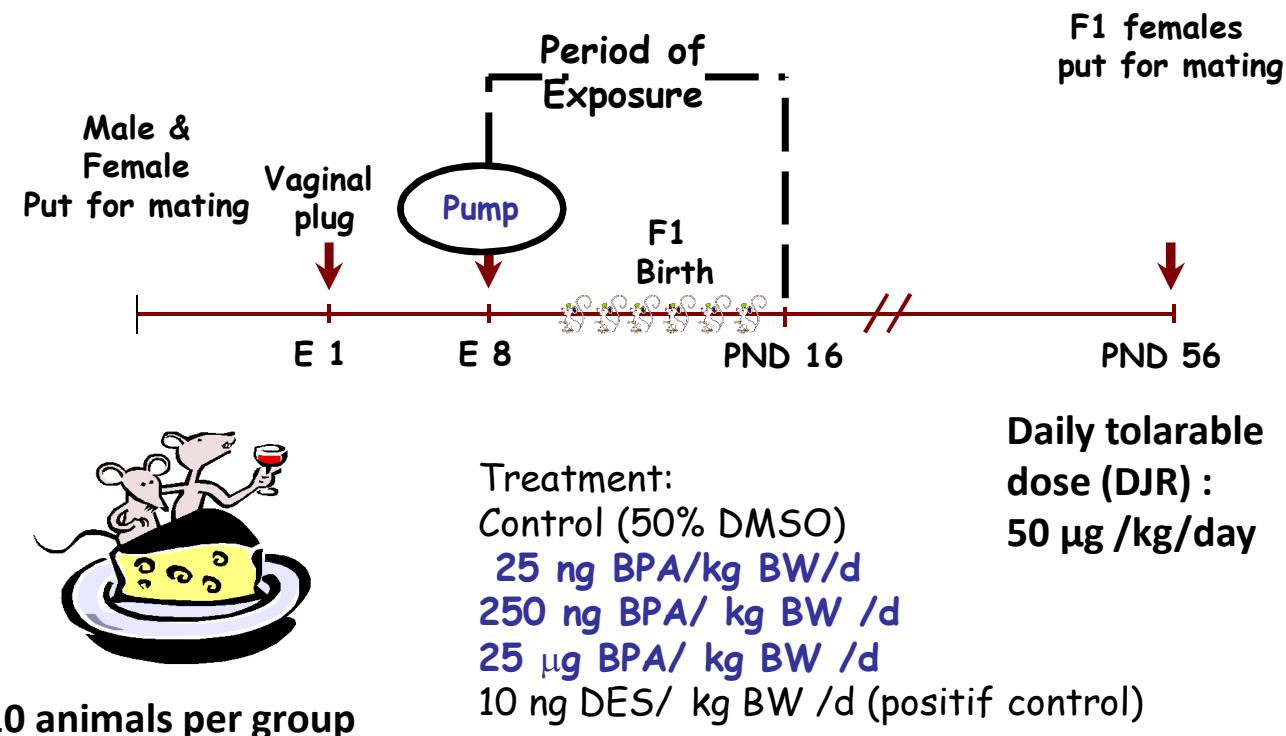
Metabolomics workflow



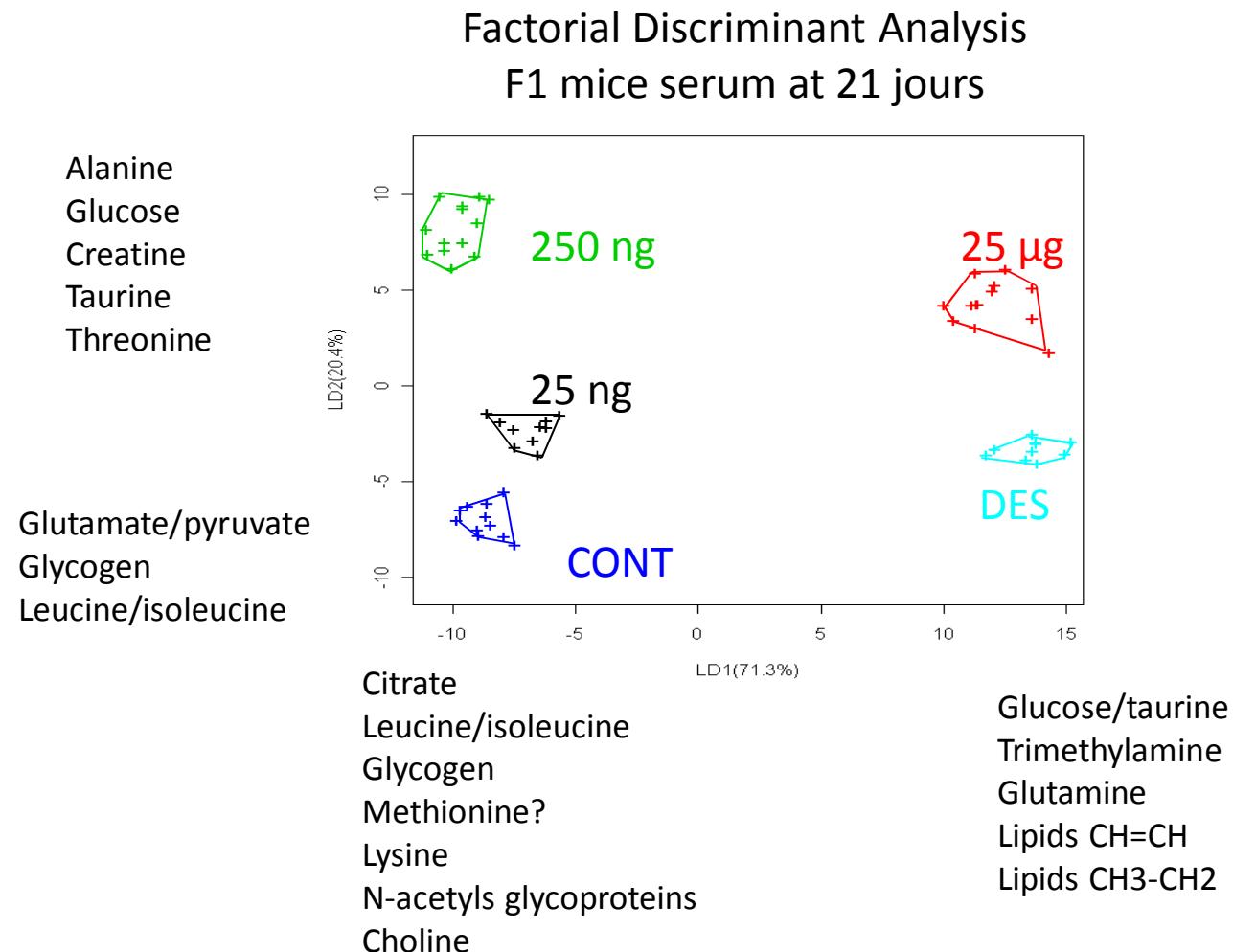
Impact of perinatal exposure to low doses of bisphenol A

Partenaires: D. Zalko, TOXALIM
A. Soto, Tufts Univ. School of Med., Boston

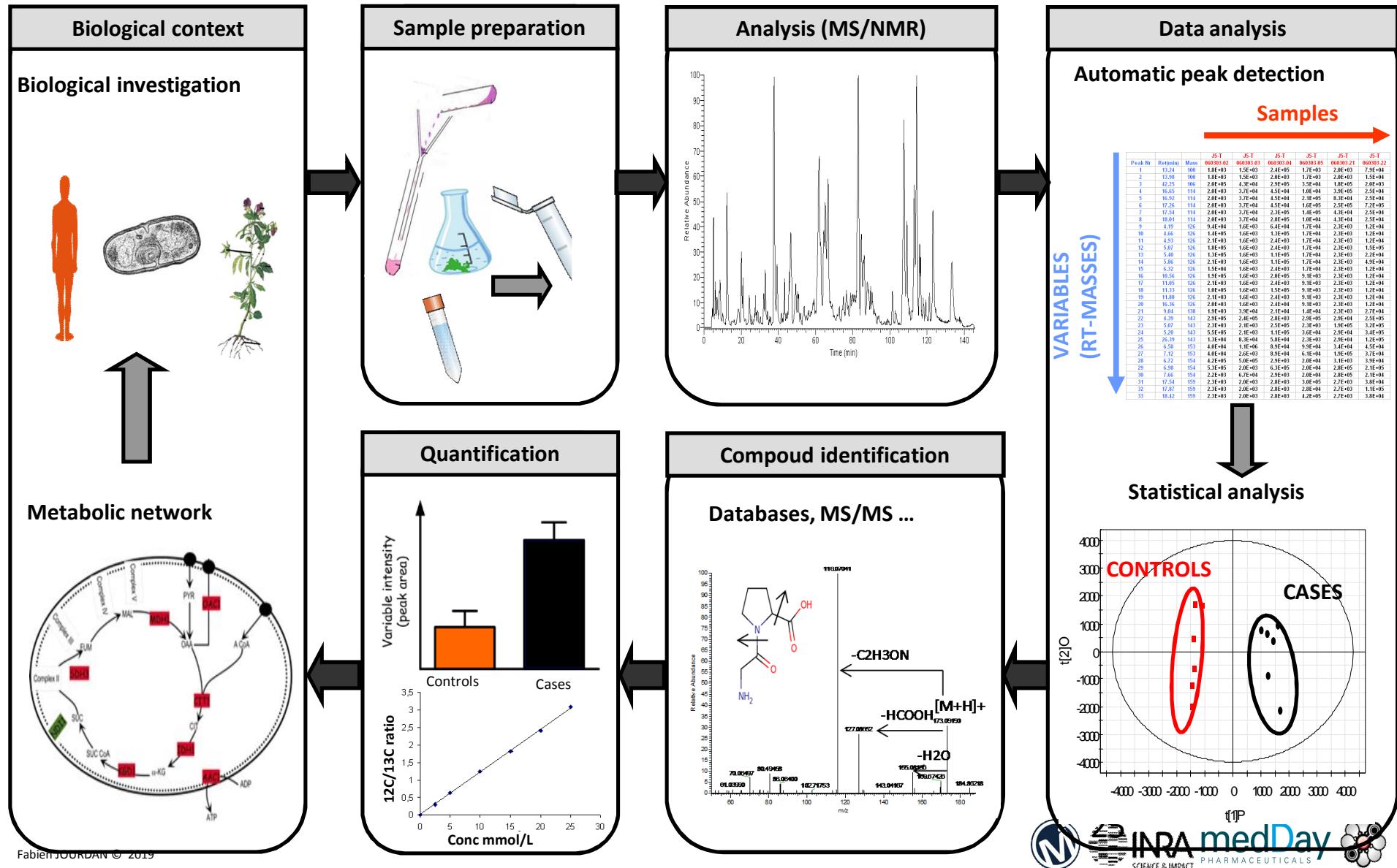
Perinatal exposure of CD1 mice to BPA



Impact of perinatal exposure to low doses of bisphenol A

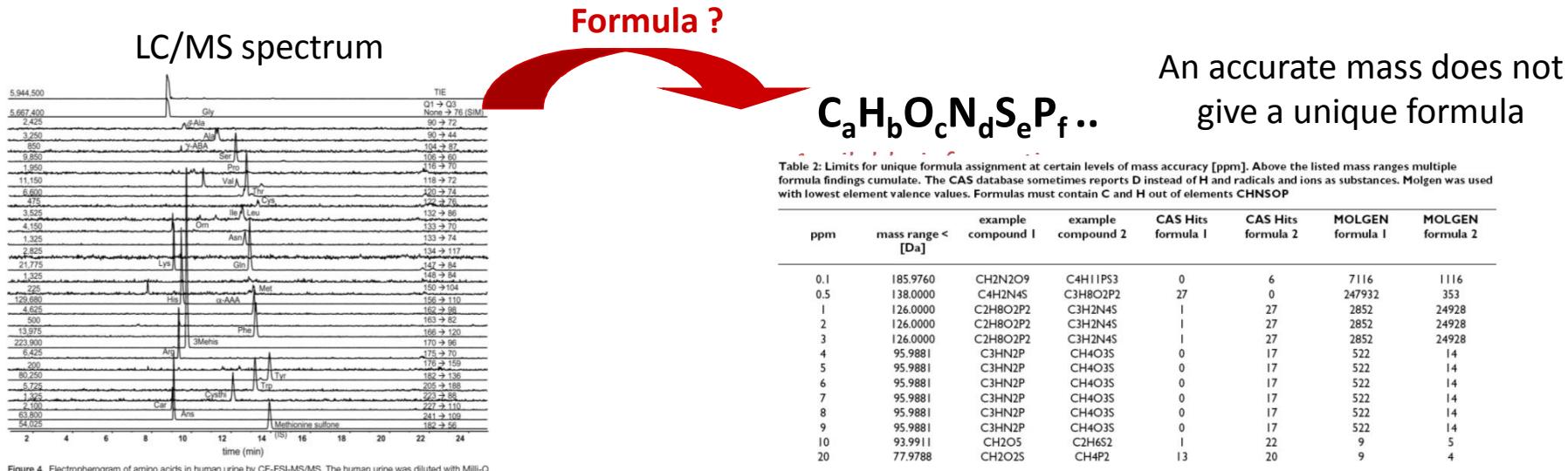


Metabolomic workflow



MS-based metabolomics: identification

The challenge: identification of compounds in (highly) complex mixtures



A formula does not give a unique compound

Structure ?

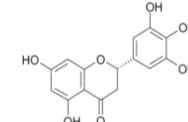
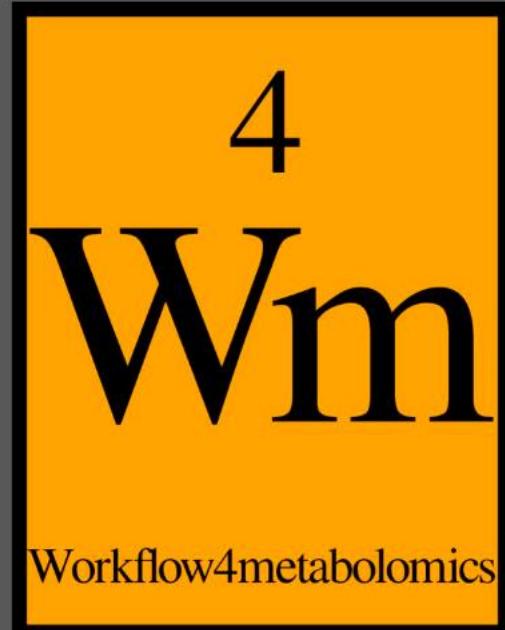


Table I: Example of a molecular formula search for $C_{15}H_{12}O_7$ in different chemical databases. Search date: July 2007

Database name	Compounds found	Total database entries
Chemical Abstracts (CAS)	181	24,000,000
Beilstein Database (MDL)	166	8,000,000
Dictionary of Natural Products (DNP)	129	170,000
PubChem (NIH)	19	800,000
Available Chemicals Directory (MDL)	6	400,000
ChEMBL (NIH)	6	370,000
KEGG (Kyoto University)	3	13,000
NIST05 (NIST mass spectral database)	2	163,000
MOLGEN molecular isomer generator (allowing 2 benzene groups; 1 ether group, 1 keto group; 5 hydroxy groups)	788,000	-

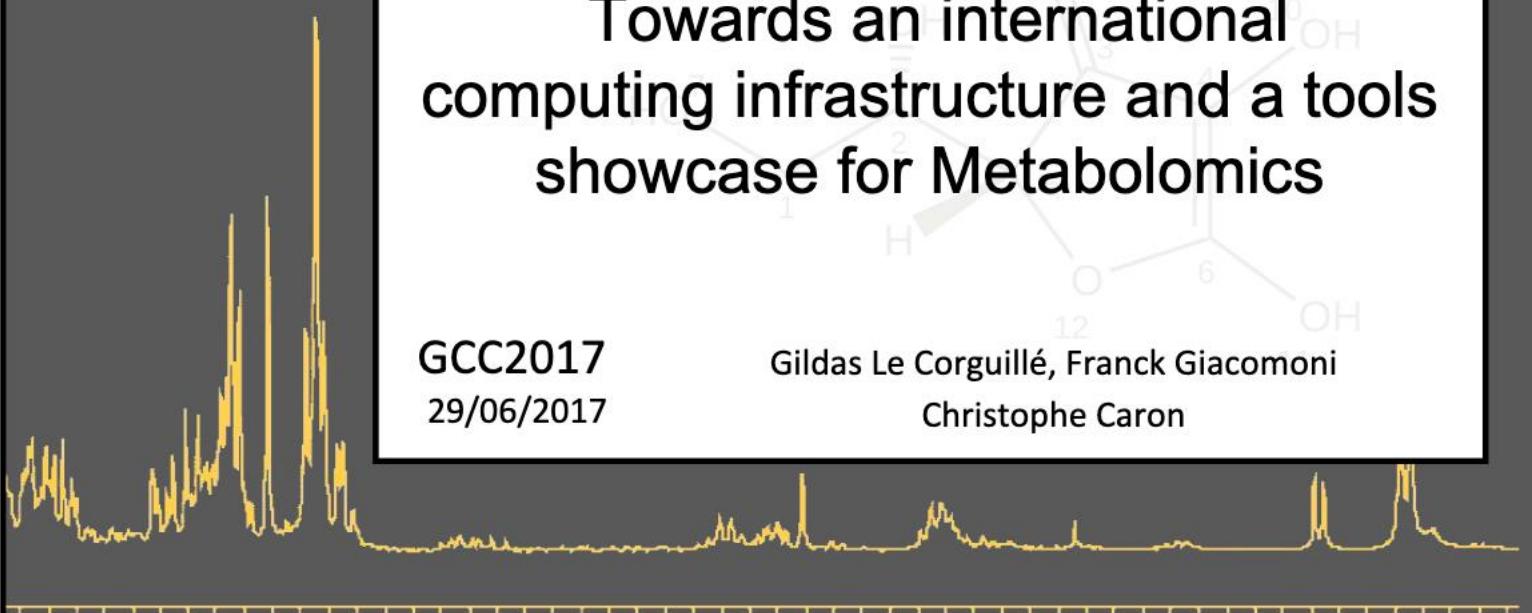
Kind & Fiehn, BMC Bioinformatics 2006



Towards an international computing infrastructure and a tools showcase for Metabolomics

GCC2017
29/06/2017

Gildas Le Corguillé, Franck Giacomoni
Christophe Caron

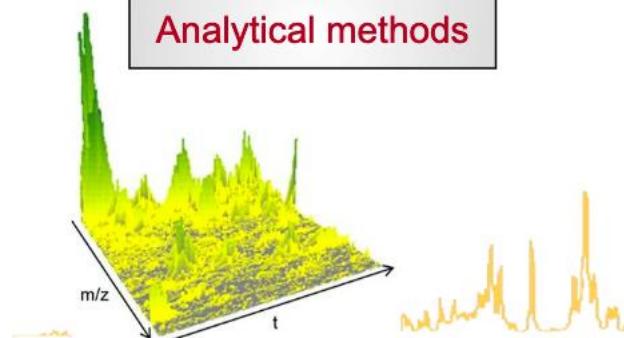


Metabolomics: a data driven approach

name	namecustom	mz	mzmin	mzmax
M100T183	M100.07529T183	100.075290217332	100.073962458109	100.075856347639
M101T64	M101.017412T64	101.017411813327	101.016363119337	101.025598611889
M101T1090	M101.094497T1090	101.094497373251	101.093825596321	101.094931330369
M102T67	M102.121467T67	102.121466917005	102.118620434328	102.128045805211
M103T60	M103.032369T60	103.032369394794	103.031358949596	103.040632681061
M103T61	M103.120075T61	103.120074855461	103.118630961785	103.129325340649
M104T1162	M103.950053T1162	103.950052934034	103.948946676433	103.95133690961
M104T60	M104.037434T60	104.037433793914	104.036015958492	104.046229685927
M104T47	M104.100665T47	104.100665065741	104.099665732044	104.108679356773
M104T88	M104.095422T88	104.095422407145	104.093036006622	104.100116358936
M104T42	M104.195599T42	104.195598628802	104.187807623186	104.197644370905
M105T61	M105.045229T61	105.045228536653	105.036480671616	105.046334818564
M105T242	M105.069935T242	105.069934637469	105.06894296829	105.071143736243
M105T48	M105.103611T48	105.103611067744	105.102427529644	105.111698113321

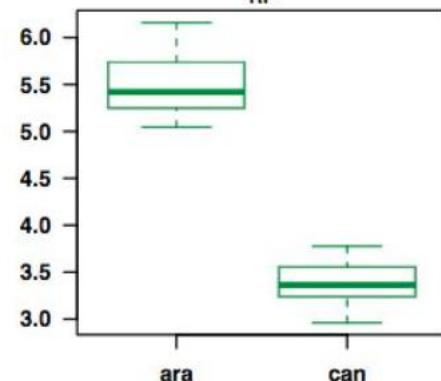
Data extraction

Analytical methods



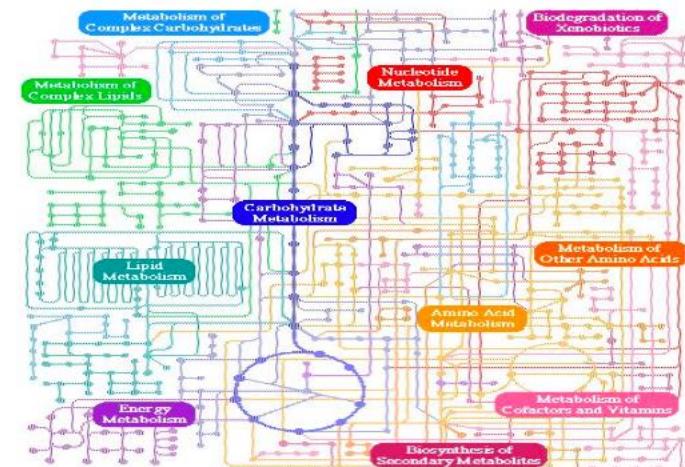
Statistical analysis

M585T579



Identification

Interpretation





A main public instance

<http://workflow4metabolomics.org>

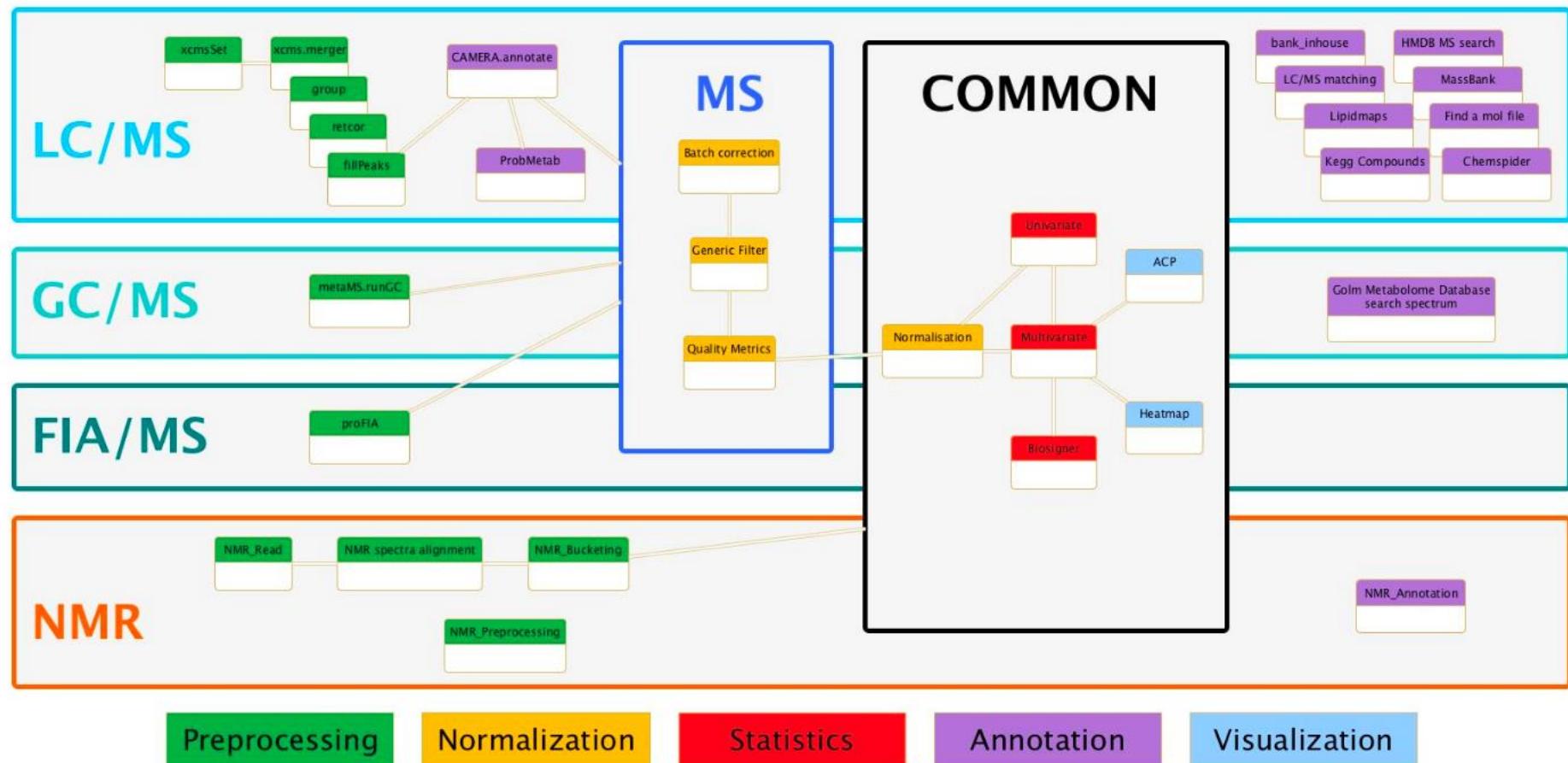
The screenshot shows the Galaxy interface for the Workflow4metabolomics instance. The left sidebar lists various bioinformatics tools categorized by platform: LC-MS, GC-MS, NMR, and COMMON TOOLS. The main content area features a banner for 'Workflow4metabolomics' with a 'Current version : 3.0' badge. It includes a publication summary, a 'Latest news' section with two items, and a 'Common' workflow diagram. The right side shows a 'History' panel listing recent jobs, such as 'xcms summary h tml' and 'xset.merged.grou p.retcov.group.fillPeak s.annotateDiffreport (sample-vs-b_tsv)'.

800 accounts
25 000 jobs/year

19 related papers

HelpDesk support@workflow4metabolomics.org

Tools



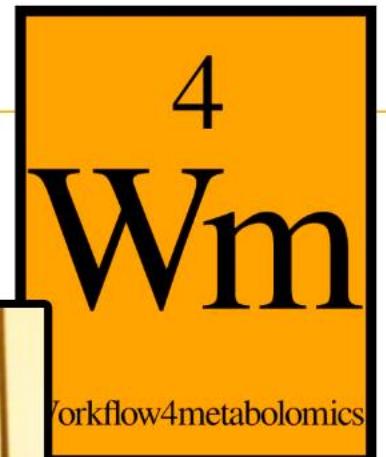
36 tools

Trainings

Workflow4Experimenters

- 3 sessions since 2014
- 10 Trainers for 20 Trainees
- “Bring your own data” tutoring sessions





MERCI !



...



Example: Yeast cadmium exposure study

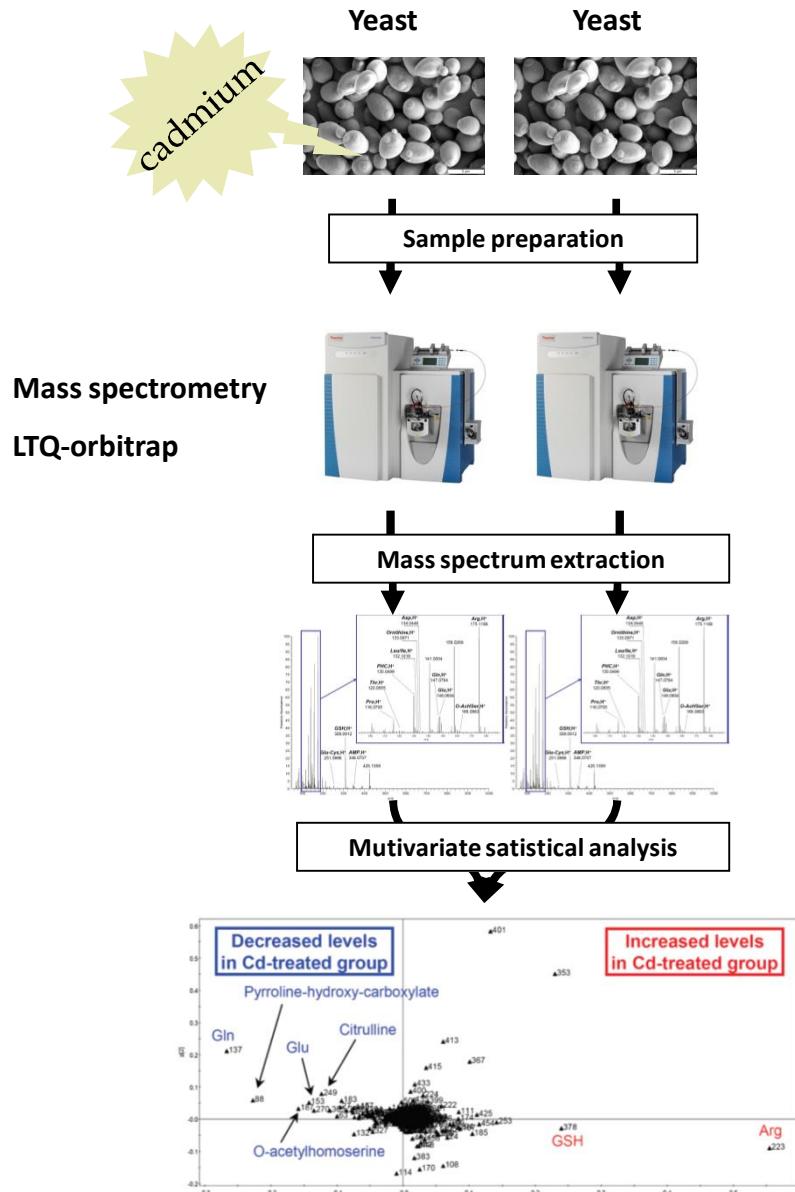


Table 2. List of 21 Discriminating Variables (i.e., $[M+H]^+$ Ions) Highlighted by PLS-DA

$[M+H]^+$ (<i>m/z</i>)	intensity ratio ^a	metabolite ID	conclusion
175.119	1.9	arginine	identified
308.091	33.9	reduced glutathione	identified
162.076	0.5	O-acetylhomoserine and/or 2-aminoadipate	to be confirmed ^c
123.055	4.8	niacinamide and/or pyridine-3-aldoxime	to be confirmed ^d
130.050	0.7 (NS) ^b	pyrrole-hydroxy-carboxylate	identified
150.058	0.3	methionine	identified
176.103	0.7	citrulline	to be confirmed ^c
120.065	0.6	threonine and/or homoserine	identified
147.076	0.7	glutamine	identified
148.060	0.8	glutamate	identified
251.070	192.2	glutamylcysteine	identified
298.096	11.0	5-methylthioadenosine	identified
106.050	0.2	serine	identified
76.039	0.3	glycine	to be confirmed ^e
223.075	50.5	cystathione	identified
147.113	0.7	lysine	identified
179.048	35.9	cysteinylglycine	to be confirmed ^d
132.102	1.2	leucine/isoleucine	identified
182.081	2.9	tyrosine	identified
156.077	1.2	histidine	identified
90.055	0.8 ^b	alanine	identified

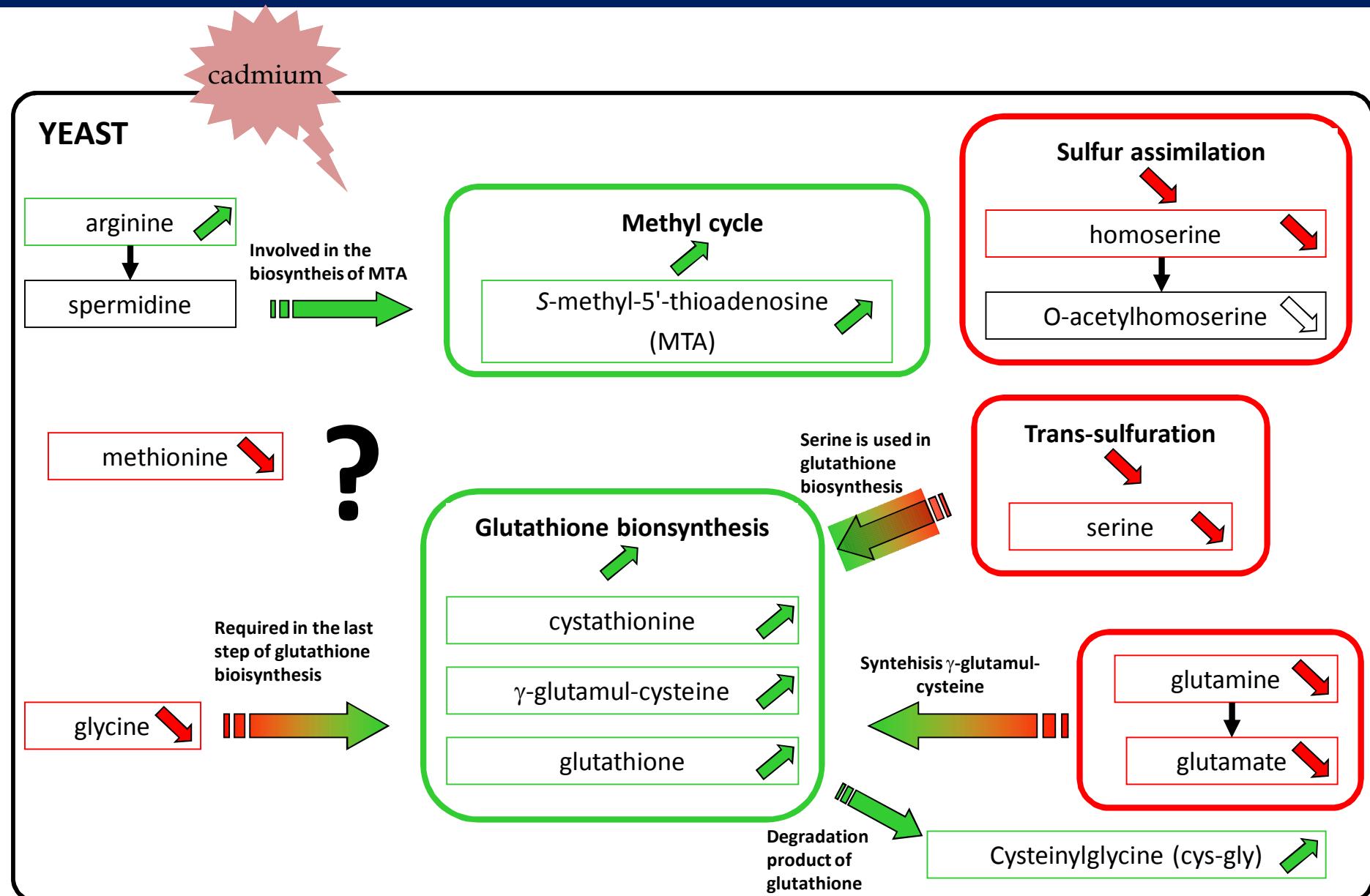
^a Cd to control. ^b NS: not statistically significant ($p > 0.05$, t-test).
^c Interference from isobaric ion in the CID spectra. ^d Lack of diagnostic ions from the CID spectra. ^e No CID spectra due to low signal intensity.

What are the metabolic processes involved ?

Madalinski G, Godat E, Alves S, et al. Direct introduction of biological samples into a LTQ-Orbitrap hybrid mass spectrometer as a tool for fast metabolome analysis. *Anal. Chem.*



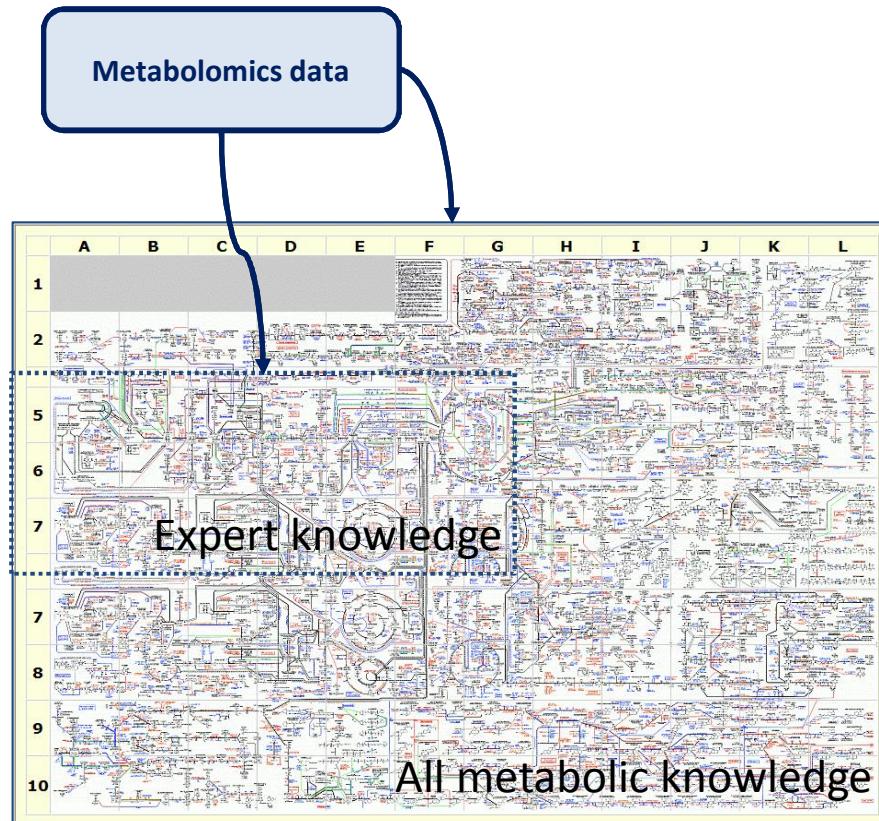
Expertise based analysis



Untargeted observations require a more holistic analysis

"Rationality is bounded when it falls short of omniscience. And the failures of omniscience are largely failures of knowing all the alternatives, uncertainty about relevant exogenous events, and inability to calculate consequences."

Herbert A. Simon, "Rational decision making in business organizations"
Nobel Memorial Lecture 1978.



Gather metabolic knowledge - *Reconstruction*
Integrate metabolomics data - *Mapping*
Model global metabolism - *Graphs*
Suggest interpretation - *Algorithms*

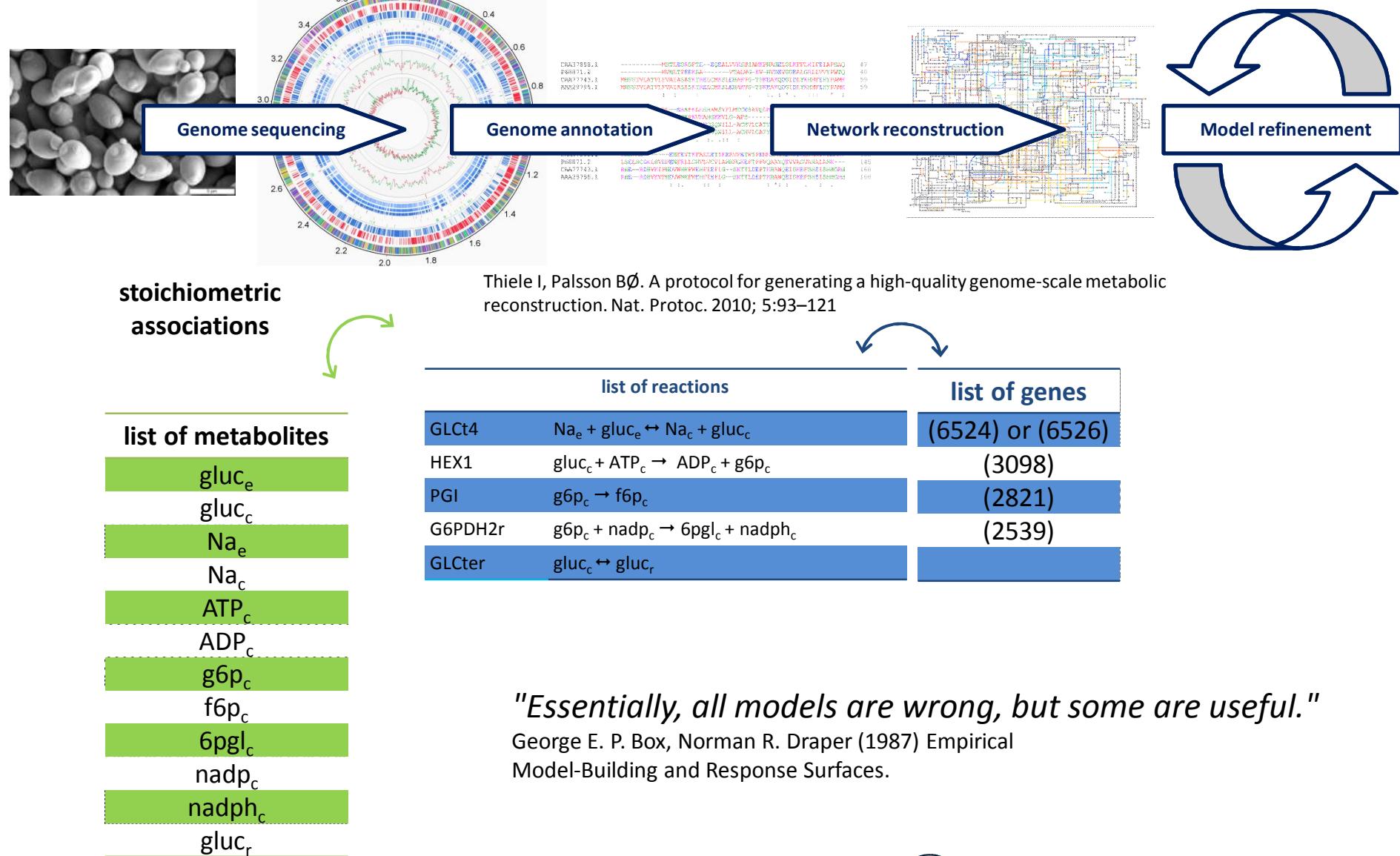


Visualization



GATHER METABOLIC KNOWLEDGE - *RECONSTRUCTION*

Gathering metabolic knowledge – genome scale models



Metabolic network and model reconstruction...in practice

2005 *T. brucei* TREU 927 genome sequenced

2008 Annotation for 9,068 protein-coding genes and first reconstruction

2012 An international consortium of 40 investigators, expert in various aspects of trypanosome metabolism, is gathered for a Two-day "jamboree".

2014 The TrypanoCyc project published (Shameer et al. NAR 2014)

Overview of the annotation:

1,368 editing events

<http://www.metexplore.fr/trypanocyc/>

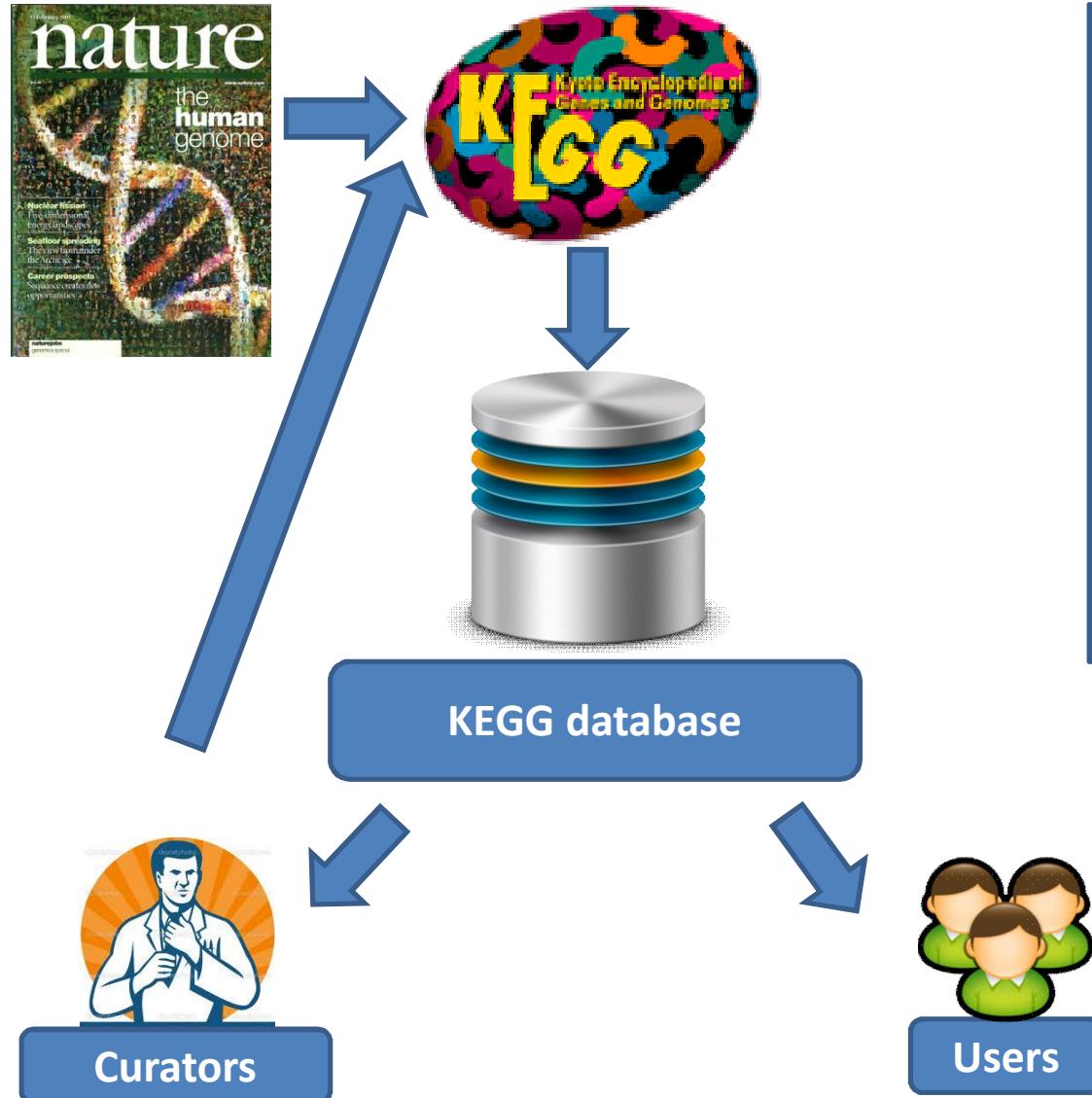
- ” 653 annotations made on 464 reactions.
- ” 17 pathways added
- ” 35 enzymatic-reactions added
- ” 10 transport reactions added
- ” 41 enzymes added
- ” 2 protein complexes added
- ” 104 metabolites have been added

But still a long way to go!

A screenshot of the TrypanoCyc web interface. At the top, there's a navigation bar with links for Home, Search, Genome, Metabolism, Analysis, and Help. Below the navigation is a search bar with the placeholder text 'Searching Trypanosome brucei'. The main content area is titled 'TrypanoCyc Overview' and contains a detailed description of the database's purpose and features. It highlights the ability to map metabolites, enzymes, and reactions across different developmental stages and compartments. Below this is a section titled 'TrypanoCyc Annotation' with a bulleted list of objectives: localization of enzyme activity and presence/absence of enzyme activity at various developmental stages of the parasite. Further down is a 'Mapping, Model Analysis and SBML Export' section. At the bottom of the interface is a large, complex metabolic network diagram represented as a grid of nodes connected by lines, showing the relationships between various metabolites and enzymes in the trypanosome.



KEGG



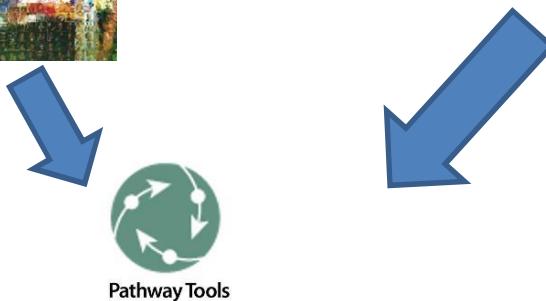
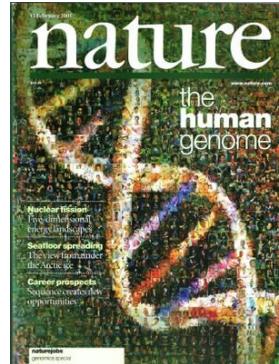
Pros:

- “easy to use
- “large range of organisms
- “large range of tools

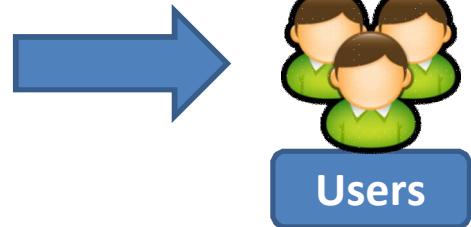
Cons:

- “no information on the “quality of a reconstruction”
- “not designed for modelling
- “downloading data requires a licence

BioCyc



BioCyc organism specific database



Pros:

- “ large range of organisms
- “ large range of tools
- “ information on the quality of reconstructions
- “ reconstructions made by consortium

Cons:

- “ less user friendly than KEGG
- “ they are moving to a licencing policy



Systems Biology Markup Language

- “ Formalised way to describe metabolic networks
- “ First designed for computation
- “ Generally attached to publications
- “ Single file per network
- “ Easy to read for computers



The Systems Biology Markup Language



```

<sbml xmlns="http://www.sbml.org/sbml/level2" xmlns:html="http://www.w3.org/1999/xhtml" level="2" version="1">
  <model id="AraC4_2009" name="C4GEM_v2">
    <listOfCompartments>
      <compartment id="Extracellular" />
      <compartment id="Cytosol" outside="Extracellular" />
      <compartment id="Mitochondria" outside="Cytosol" />
      <compartment id="Plastid" outside="Cytosol" />
      <compartment id="Vacuole" outside="Cytosol" />
      <compartment id="Peroxisome" outside="Cytosol" />
    </listOfCompartments>
    <listOfSpecies>
      [...]
      <species id="Pyrophosphate_m" name="C00013_m" compartment="Mitochondria" charge="0" boundaryCondition="false" />
      <species id="Pyrophosphate_p" name="C00013_p" compartment="Plastid" charge="0" boundaryCondition="false" />
      <species id="Pyrophosphate_x" name="C00013_x" compartment="Peroxisome" charge="0" boundaryCondition="false" />
      <species id="Pyruvate_c" name="C00022_c" compartment="Cytosol" charge="0" boundaryCondition="false" />
      <species id="Pyruvate_m" name="C00022_m" compartment="Mitochondria" charge="0" boundaryCondition="false" />
      <species id="Pyruvate_p" name="C00022_p" compartment="Plastid" charge="0" boundaryCondition="false" />
      [...]
    </listOfSpecies>
    <listOfReactions>
      [...]
      <reaction id="R00006_c" name="acetolactate synthase, chloroplast / acetohydroxy-acid synthase (ALS)" reversible="false">
        <listOfReactants>
          <speciesReference species="Pyruvate_c" stoichiometry="2" />
        </listOfReactants>
        <listOfProducts>
          <speciesReference species="2-Acetolactate_c" stoichiometry="1.0" />
          <speciesReference species="CO2_c" stoichiometry="1.0" />
        </listOfProducts>
        <notes>
          <html:listOfGenes>
            <html:p>ENTRY: AT3G48560</html:p>
            <html:p>ID:</html:p>
            <html:p>ENZYME: acetolactate synthase, chloroplast / acetohydroxy-acid synthase (ALS)</html:p>
            <html:p>EC: 2.2.1.6</html:p>
            <html:p>PATHWAY: Pantothenate and CoA biosynthesis</html:p>
          </html:listOfGenes>
        </notes>
      </reaction>
      [...]
    </listOfReactions>
  </model>
</sbml>

```

Compartments

```
<compartment id="c"
    constant="false"
    spatialDimensions="3"
    name="cytoplasm"
    metaid="_cbd1dd46_c958_4b53_9667_27890b8ac164"
    sboTerm="SBO:0000290"
    size="1">
<annotation>
    <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:bqmodel="http://biomodels.net/model-qualifiers/"
        xmlns:bqbiol="http://biomodels.net/biology-qualifiers/">
        <rdf:Description rdf:about="#_cbd1dd46_c958_4b53_9667_27890b8ac164">
            <bqbiol:is>
                <rdf:Bag>
                    <rdf:li rdf:resource="http://identifiers.org/obo.go/GO:0005737"/>
                </rdf:Bag>
            </bqbiol:is>
        </rdf:Description>
    </rdf:RDF>
</annotation>
</compartment>
```

Species

```
<species id="M_gip_c"
    initialConcentration="1"
    constant="false"
    charge="-2"
    hasOnlySubstanceUnits="false"
    name="D-Glucose 1-phosphate"
    metaid="_metaM_gip_c"
    boundaryCondition="false"
    sboTerm="SBO:0000247"
    compartment="c">
<notes>
    <body xmlns="http://www.w3.org/1999/xhtml">
        <p>FORMULA: C6H11O9P</p>
        <p>CHARGE: -2</p>
        <p>HEPATONET_1.0_ABBREVIATION: HC00103</p>
        <p>EHMN_ABBREVIATION: C00103</p>
        <p>INCHI: InChI=1S/C6H13O9P/c7-1-2-3(8)4(9)5(10)6(14-2)15-16(11,12)13/h2-10H,1H2,(H2,11,12,13)/p-2/t2-,3-,4+,5-,6-/m1/s1</p>
    </body>
</notes>
<annotation>
    <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:bqmodel="http://biomodels.net/model-qualifiers/"
        xmlns:bqbiol="http://biomodels.net/biology-qualifiers/">
        <rdf:Description rdf:about="#_metaM_gip_c">
            <bqbiol:is>
                <rdf:Bag>
                    <rdf:li rdf:resource="http://identifiers.org/chebi/CHEBI:58601"/>
                </rdf:Bag>
            </bqbiol:is>
        </rdf:Description>
    </rdf:RDF>
</annotation>
</species>
```



Reaction

```
<reaction id="R_10FTHF5GLUtl"
    name="5-glutamyl-10FTHF transport, lysosomal"
    metaid="_metaR_10FTHF5GLUtl"
    reversible="false"
    sboTerm="SBO:0000185">
<notes>
    <body xmlns="http://www.w3.org/1999/xhtml">
        <p>GENE_ASSOCIATION: </p>
        <p>SUBSYSTEM: Transport, lysosomal</p>
        <p>EC Number: </p>
        <p>Confidence Level: 2</p>
        <p>AUTHORS: PMID:11375437</p>
        <p>NOTES:carrier-mediated transport - but which and how is unknown IT</p>
    </body>
</notes>
<annotation>
    <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:bqmodel="http://biomodels.net/model-qualifiers/"
        xmlns:bqbiol="http://biomodels.net/biology-qualifiers/">
        <rdf:Description rdf:about="#_metaR_10FTHF5GLUtl">
            <bqbiol:isVersionOf>
                <rdf:Bag>
                    <rdf:li rdf:resource="http://identifiers.org/obo.eco/ECO:0000000"/>
                </rdf:Bag>
            </bqbiol:isVersionOf>
        </rdf:Description>
    </rdf:RDF>
</annotation>
<listOfReactants>
    <speciesReference species="M_10fthf5glu_c" stoichiometry="1"/>
</listOfReactants>
<listOfProducts>
    <speciesReference species="M_10fthf5glu_l" stoichiometry="1"/>
</listOfProducts>
<kineticLaw>
    <math xmlns="http://www.w3.org/1998/Math/MathML">
        <ci> FLUX_VALUE </ci>
    </math>
</kineticLaw>
<listOfParameters>
```

Annotation scoring system

Evidence type	Confidence score	Description
Biochemical data	4	Direct evidence for gene product function and biochemical reaction: Protein purification, biochemical assays, experimentally solved protein structures and comparative gene-expression studies.
Genetic data	3	Direct and indirect evidence for gene function: Knock-out characterization, knock-in characterization and over expression.
Physiological data	2	Indirect evidence for biochemical reaction based on physiological data: secretion products or defined medium components serve as evidence for transport and metabolic reactions.
Sequence data	2	Evidence for gene function: Genome annotation, SEED annotation.
Modelling data	1	No evidence is available but reaction is required for modelling. The included function is a hypothesis and needs experimental verification. The reaction mechanism may be different from the included reaction(s).
Not evaluated	0	
Negative hypothesis	-1	Although there is no evidence against this reaction, it is expected to not exist
Evidence against the reaction	-2	Direct/indirect evidence against the hypothesis is available

Supplementary information

The screenshot shows a journal article page from *Nature Biotechnology*. At the top, there's a navigation bar with links to nature.com, Publications A-Z index, Browse by subject, Login, Register, and Cart. A banner for "nature collections Multicolor flow cytometry" is displayed. Below the header, the journal title "nature biotechnology" is shown, along with a search bar and a link to Advanced search.

The main content area features the article title "A community-driven global reconstruction of human metabolism" by Ines Thiele, Neil Swainston, Ronan M T Fleming, Andreas Hoppe, Swagatika Sahoo, Maike K Aurich, Hulda Haraldsdottir, Monica L Mo, Ottar Rolfsson, Miranda D Stobbe, Stefan G Thorleifsson, Rasmus Agren, Christian Bölling, Sergio Bordel, Arvind K Chavali, Paul Dobson, Warwick B Dunn, Lukas Endler, David Hala, Michael Hucka, Duncan Hull, Daniel Jameson, Neema Jamshidi, Jon J Jonsson, Nick Juty, and others. The article is categorized under NATURE BIOTECHNOLOGY | COMPUTATIONAL BIOLOGY | RESOURCE. It includes links for Affiliations, Contributions, Corresponding author, and a DOI: 10.1038/nbt.2488.

Below the article, there's a sidebar for "Supplementary information" which includes PDF files, Zip files, and Excel files. A red box highlights the "Zip files" section, which contains a link to "Supplementary Data (30.4 MB)" described as "Recon 2 and cell type-specific models in SBML format".

SBML files as supplementary material

Databases:

- " BIGG
- " Human Atlas
- " Biomodels
- " MetExplore
- "

SBML file

Files and databases for *Arabidopsis thaliana*

Model/database	URL	Reference	Curated	SBML
Poolman2009	http://www.plantphysiol.org/content/151/3/1570/suppl/DC1	Poolman et al. 2009	YES	YES
AraGEM	http://www.plantphysiol.org/content/152/2/579/suppl/DC1	Gomes et al. 2010	YES	YES
BMID000000140799	http://www.ebi.ac.uk/biomodels-main/BMID000000140799	Chen et al. 2010	no	YES
Radrich2010	http://www.biomedcentral.com/1752-0509/4/114/additional	Radrich et al. 2010	no	YES
KEGG	http://www.genome.jp/kegg/	Kanehisa et al. 2008	YES	no
AraCyc	http://www.arabidopsis.org/biocyc/	Zhang et al. 2005	YES	no

F. Jourdan, Metabolomics and Polyphenols review, 2013

Pros:

- “ large range of organisms
- “ all in one file
- “ allows modelling
- “ can be imported in various software

cons:

- “ cannot be used without a software
- “ quality and consistency of information vary



Human genome-scale metabolic networks

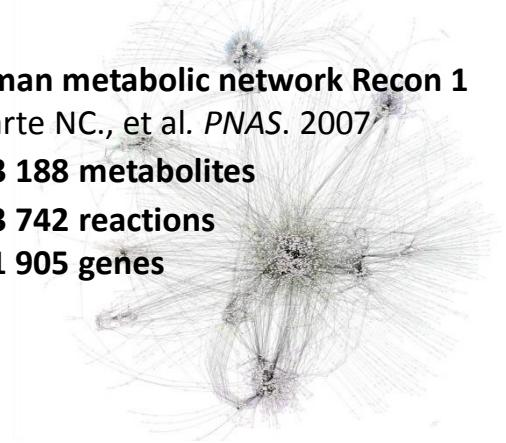
Human metabolic network Recon 1

Duarte NC., et al. *PNAS*. 2007.

3 188 metabolites

3 742 reactions

1 905 genes



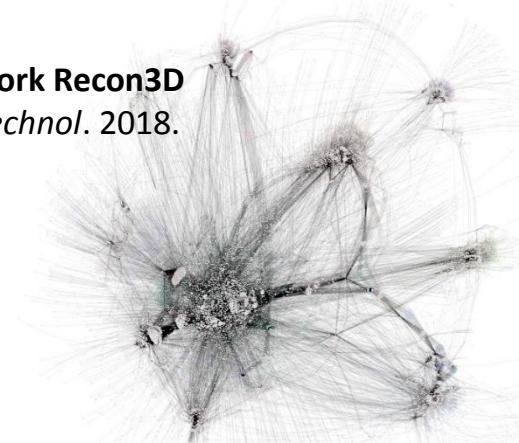
Human metabolic network Recon3D

Brunk E., et al. *Nat Biotechnol.* 2018.

8 399 metabolites

13 543 reactions

3 697 genes



2007

2013

2018

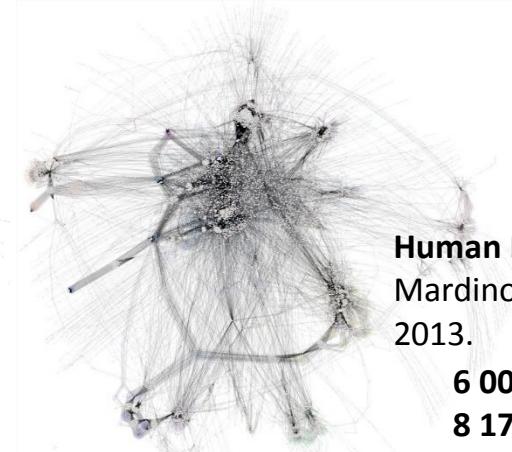
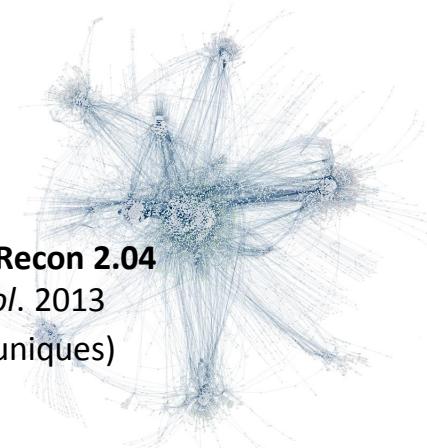
Human metabolic network Recon 2.04

Thiele I., et al. *Nat Biotechnol.* 2013

5063 metabolites (2626 uniques)

7440 reactions

2194 genes



Human Metabolic Reaction (HMR)

Mardinoglu A., et al. *Mol Sys Biol.* 2013.

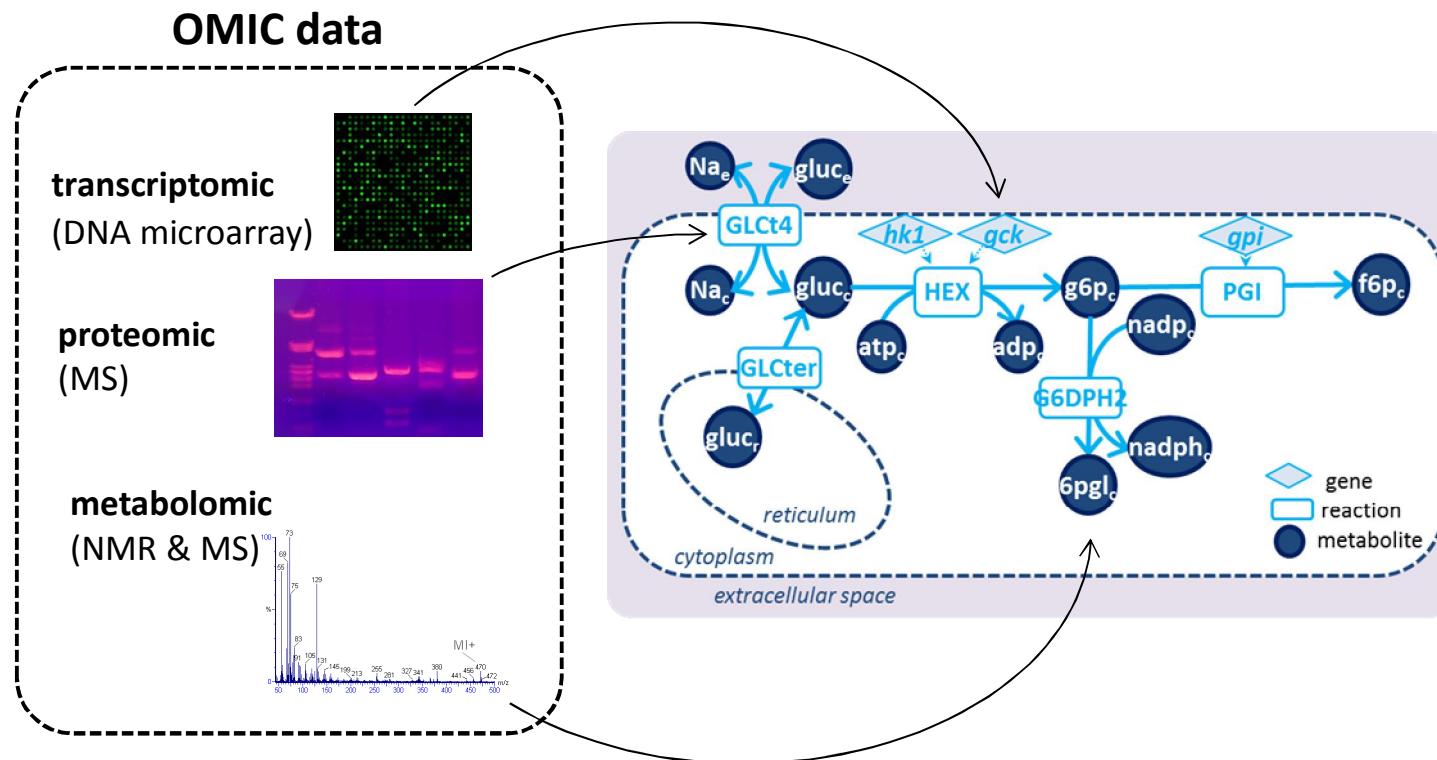
6 005 metabolites

8 174 reactions

genes



Metabolic networks for multi-omic interpretation



SBML



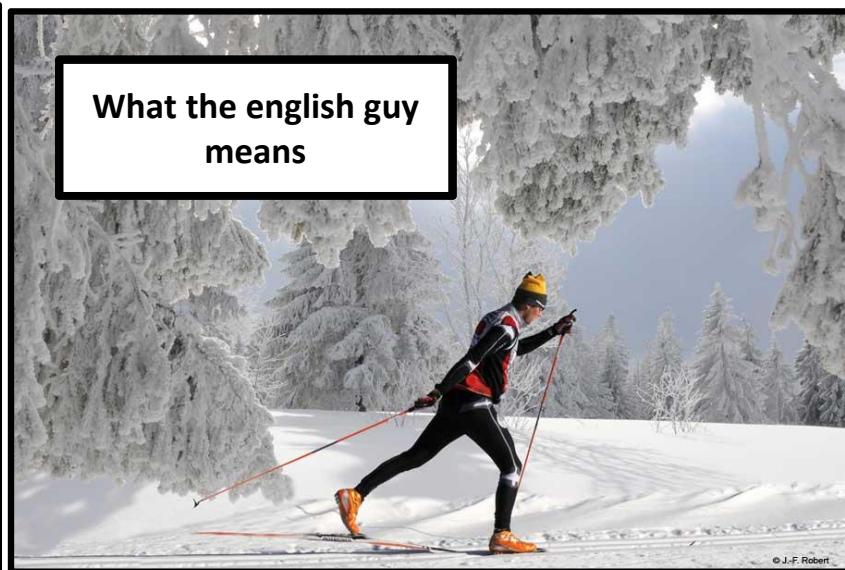
Take home message

- “ Metabolic networks unify genes-proteins-reactions and metabolites
- “ Curation is mandatory (and long) to have a good network
- “ For the same organism many networks are available...for other organisms no network is available...

Now that we have the metabolomics data and the network, how can we map the data?

INTEGRATE METABOLOMICS DATA - MAPPING

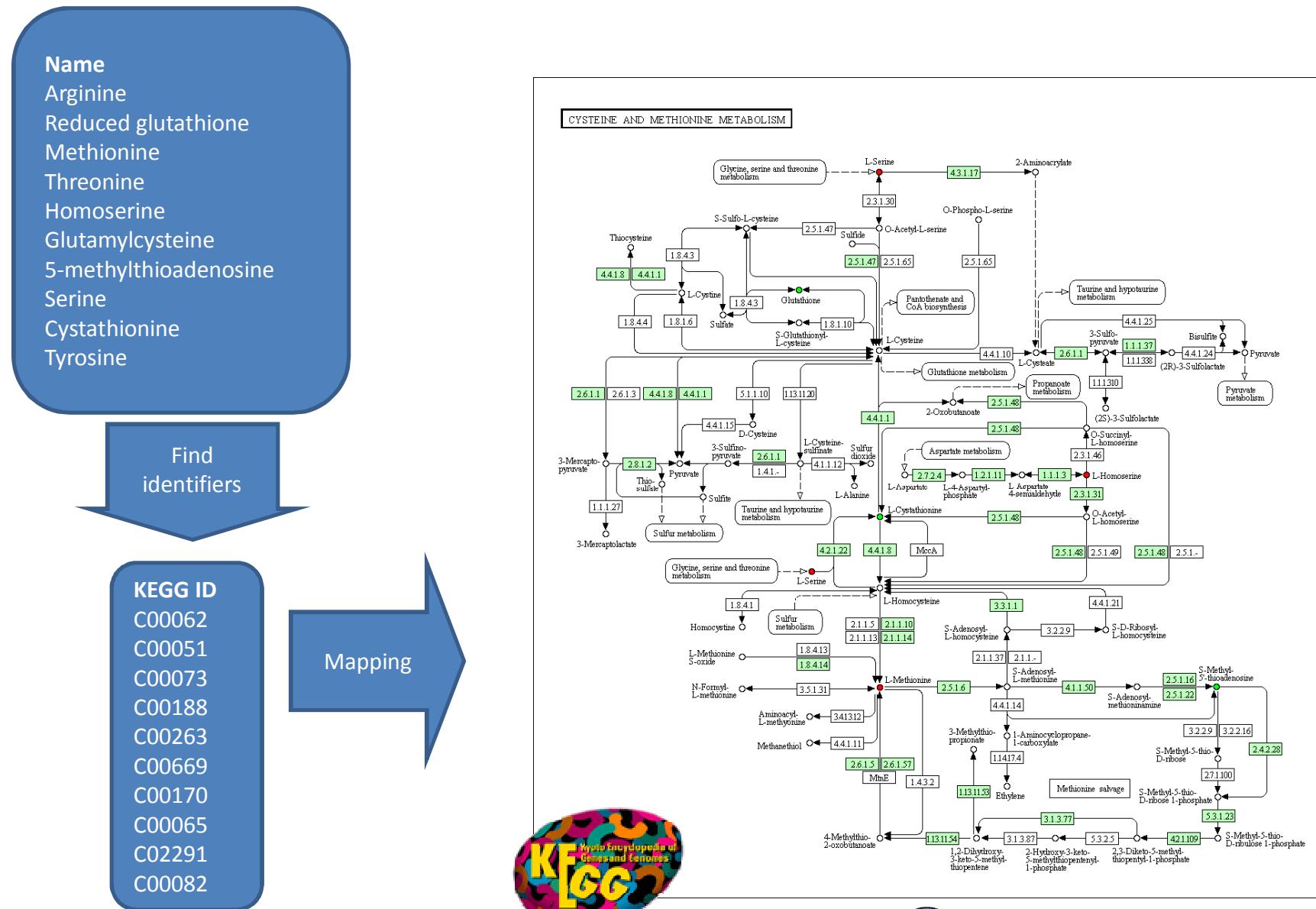
Fabien JOURDAN ©



Wrong understanding of names may lead to wrong conclusions!



From names to identifiers to pathways/networks



<https://goo.gl/xuPZLH>

PRACTICE 1

Mapping result

Mapping

Chemical Name	KEGG identifiers
Arginine	C00062
Reduced glutathione	C00051
Methionine	C00073
Threonine	C00188
Homoserine	C00263
Glutamylcysteine	C00669
5-methylthioadenosine	C00170
Serine	C00065
Cystathione	C02291
Tyrosine	C00082

MetExplore v2.13.25

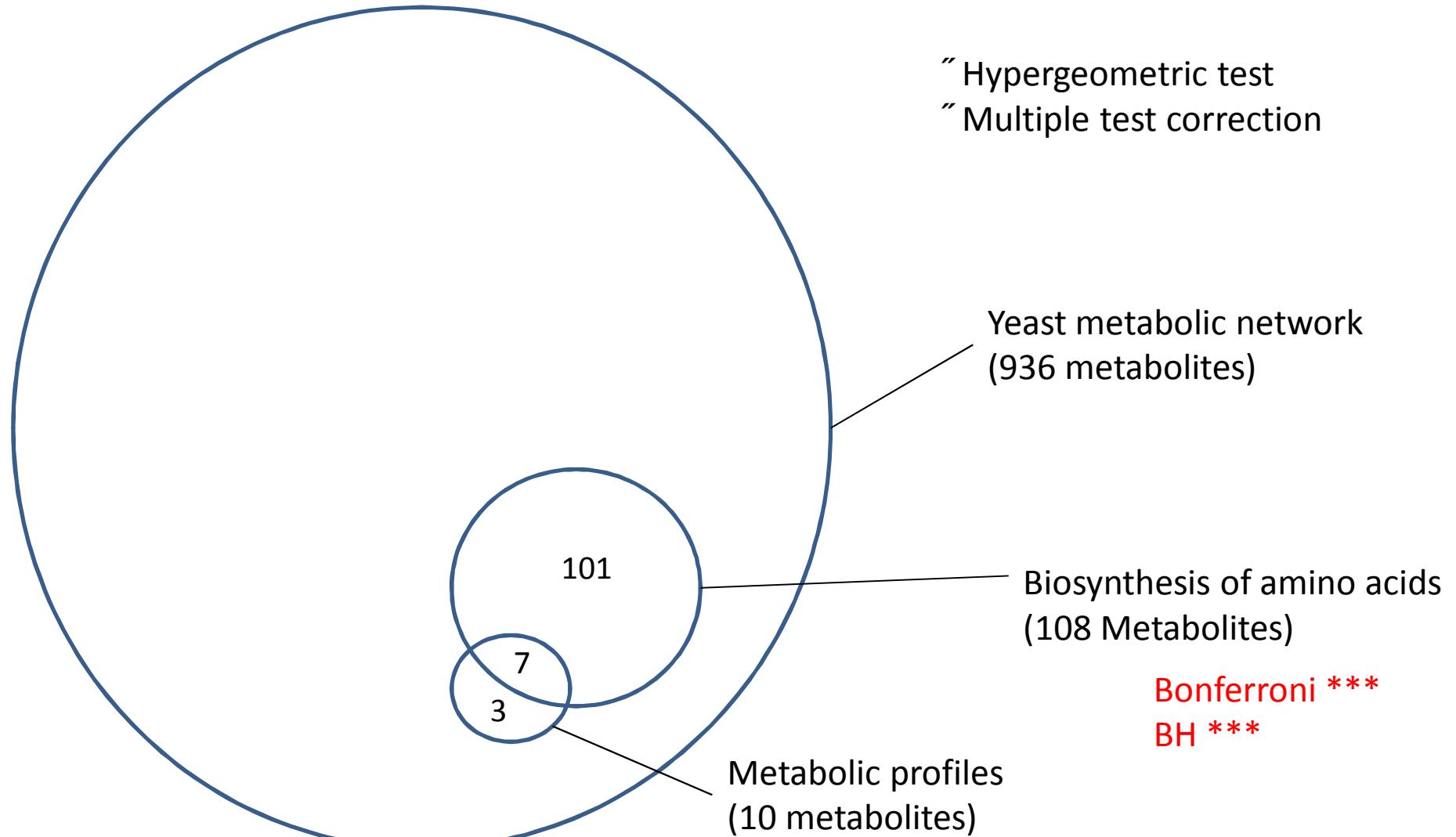
User Profile Network Data Network Curation Network Viz Mapping BioSources Compartments (1/1) Pathways (79/79) Reactions (1135/1135) Metabolites (936/936) Enzymatic Complexes + Add Edit Delete Curation Statistics Curation Votes

	Name	Nb Reactions	Coverage	Nb of ...
1	Metabolic pathways - <i>Saccharomyces cerevisiae</i> (budding yeast)	799	1.54	10
2	Biosynthesis of amino acids - <i>Saccharomyces cerevisiae</i> (budding yeast)	105	6.48	7
3	Biosynthesis of secondary metabolites - <i>Saccharomyces cerevisiae</i> (budding yeast)	285	2.78	7
4	Biosynthesis of antibiotics - <i>Saccharomyces cerevisiae</i> (budding yeast)	201	3.57	6
5	Cysteine and methionine metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	40	12.5	5
6	Aminoacyl-tRNA biosynthesis - <i>Saccharomyces cerevisiae</i> (budding yeast)	24	11.11	5
7	Glycine, serine and threonine metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	29	12.9	4
8	Glutathione metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	20	10	3
9	Phenylalanine, tyrosine and tryptophan biosynthesis - <i>Saccharomyces cerevisiae</i> (budding yeast)	21	9.09	2
10	Arginine and proline metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	27	7.41	2
11	Ubiquinone and other terpenoid-quinone biosynthesis - <i>Saccharomyces cerevisiae</i> (budding yeast)	1	50	1
12	Cyanoamino acid metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	4	11.11	1
13	beta-Alanine metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	8	8.33	1
14	Sphingolipid metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	14	7.69	1
15	Lysine biosynthesis - <i>Saccharomyces cerevisiae</i> (budding yeast)	14	6.25	1
16	Tyrosine metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	16	5.88	1
17	Arginine biosynthesis - <i>Saccharomyces cerevisiae</i> (budding yeast)			1
18	One carbon metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)			1
19	Methane metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	16	5	1
20	Valine, leucine and isoleucine biosynthesis - <i>Saccharomyces cerevisiae</i> (budding yeast)	17	5	1
21	Glyoxylate and dicarboxylate metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	15	4.55	1
22	Alanine, aspartate and glutamate metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	24	4.17	1
23	Glycerophospholipid metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	42	2.86	1
24	Carbon metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	73	1.72	1

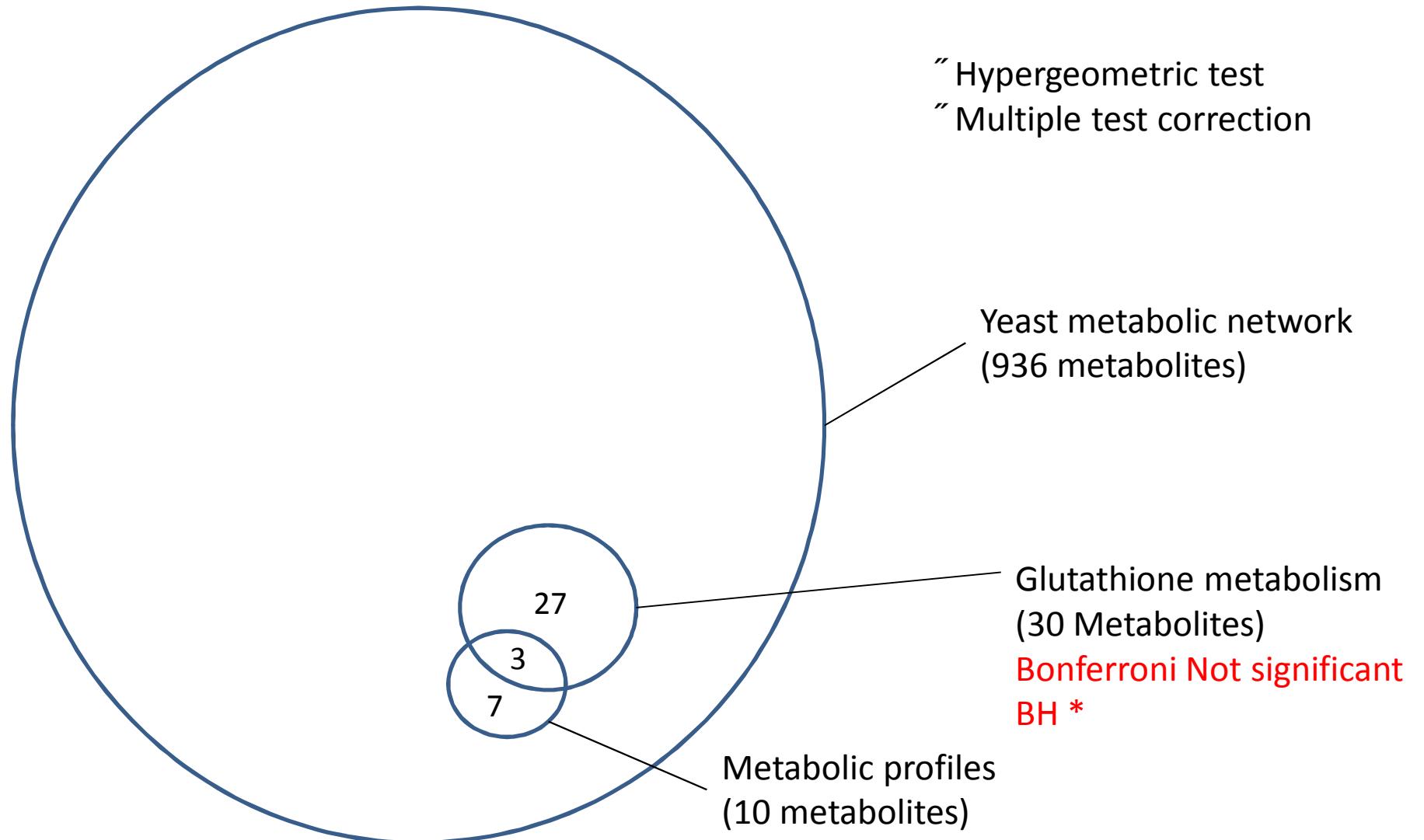
Which pathways are significant?



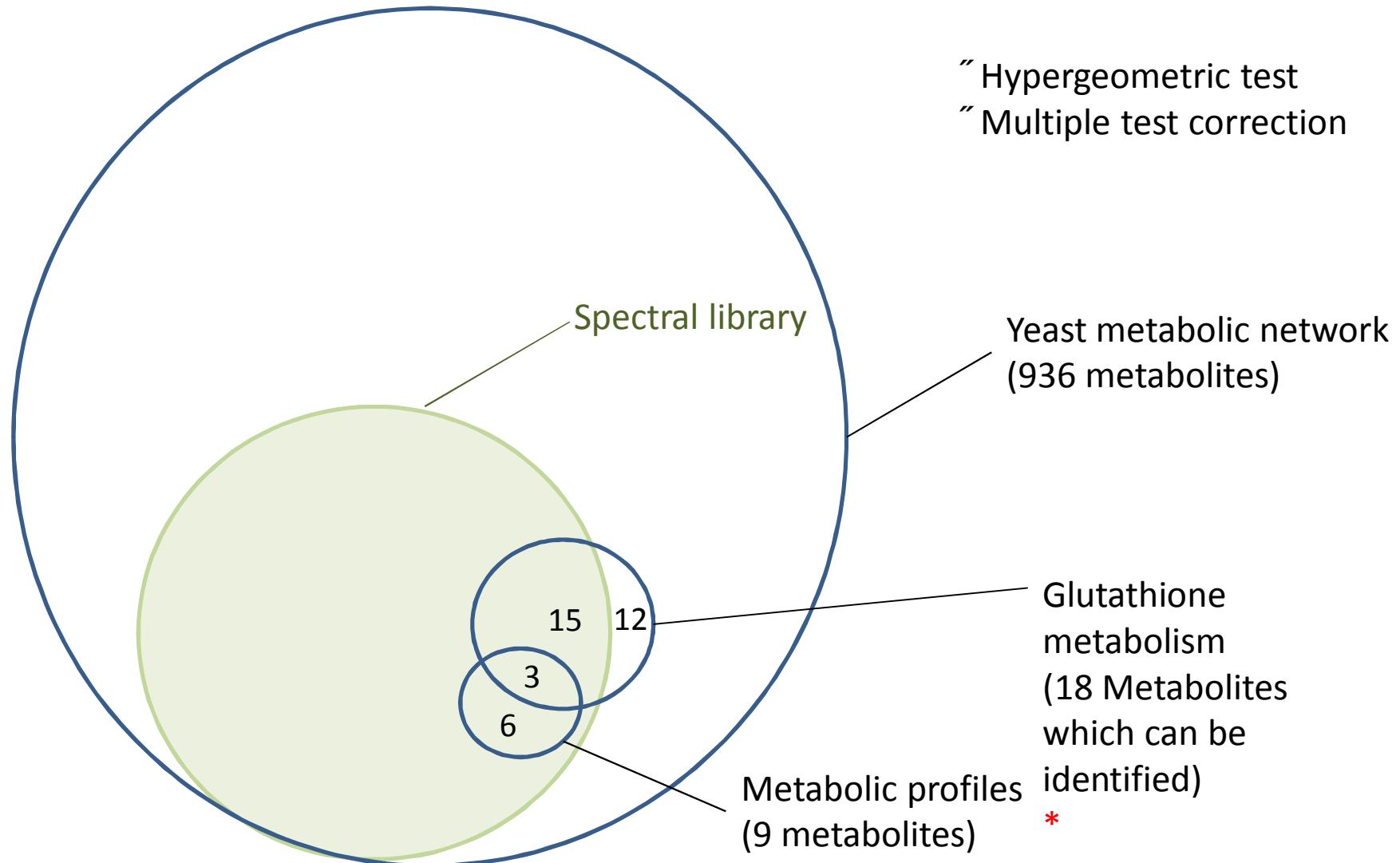
Pathway enrichment



Pathway enrichment



Pathway enrichment with background set



Pathway Enrichment in MetExplore

Mapping												
User Profile		Network Data		Network Curation		Network Viz		Mapping				
BioSources		Compartments (1/1)		Pathways (79/79)		Reactions (1135/1135)		Metabolites (936/936)				
Add		Edit		Curation Statistics		Curation Votes		(?)				
						Mapping on Metabolite (?)						
		Name		Identifier		Nb Reactions	Coverage	Nb of Mapped	p-value	Bonferroni correcte	BH-corrected p-val	
1		Cysteine and methionine metabolism - S...		sce00270		40	12.5	5	2.39e-5	*** (5.73e-4)	*** (2.87e-4)	
2		Aminoacyl-tRNA biosynthesis - Saccharo...		sce00970		24	11.11	5	4.33e-5	** (1.04e-3)	*** (3.47e-4)	
3		Biosynthesis of amino acids - Sacch...		sce01230	105	6.48	7	2.03e-5	*** (4.87e-4)	*** (4.87e-4)		
4		Glycine, serine and threonine metabolis...		sce00260		29	12.9	4	1.81e-4	** (4.34e-3)	** (1.08e-3)	
5		Biosynthesis of antibiotics - Saccharomy...		sce01130		201	3.57	6	3.42e-3	(8.22e-2)	*	(1.37e-2)
6		Glutathione metabolism - Saccharomyce...		sce00480		20	10	3	3.07e-3	(7.36e-2)	*	(1.47e-2)
7		Biosynthesis of secondary metabolites - ...		sce01110		285	2.78	7	5.28e-3	(1.27e-1)	*	(1.81e-2)
8		Ubiquinone and other terpenoid-quinone...		sce00130		1	50	1	2.13e-2	(5.10e-1)	(5.67e-2)	
9		Metabolic pathways - Saccharomyces ce...		sce01100		799	1.54	10	2.59e-2	(6.22e-1)	(6.22e-2)	
10		Phenylalanine, tyrosine and tryptophan ...		sce00400		21	9.09	2	2.12e-2	(5.08e-1)	(6.35e-2)	

Pathway enrichment

Good side

- . Provides information on functions where data are found
- . Allows prioritizing analysis
- . They provide a p-value!

Dark side

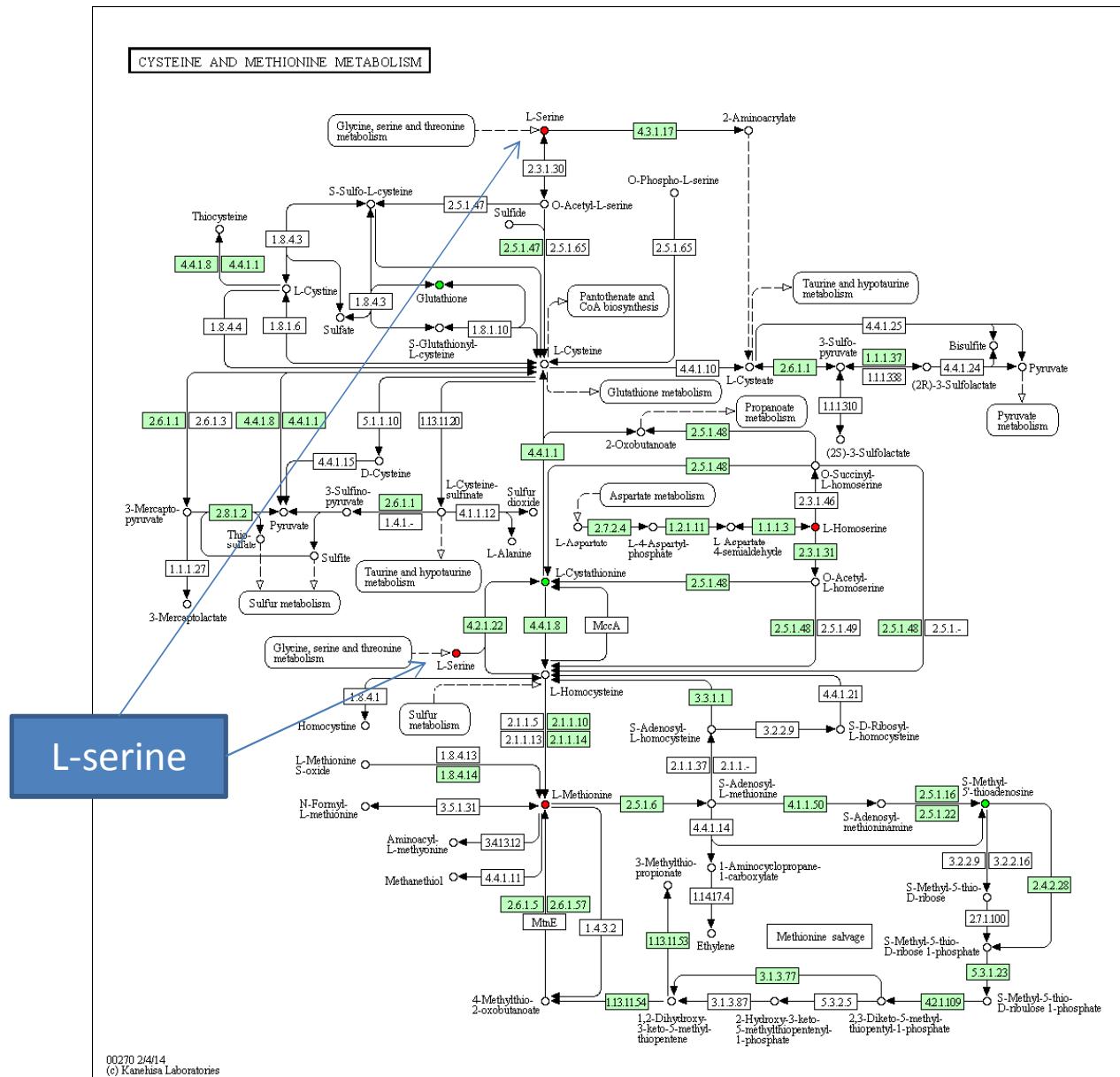
- . Doesn't take into account quantitative information (there are some solutions for that)
- . Depends on the test (less or more strict)
- . Depends on the definition of pathways (we will see it is quite subjective)
- . In metabolomics some metabolites cannot be detected
- . Pathways overlap



	Name	Nb Reactions	Coverage	Nb of ...
1	 Metabolic pathways - <i>Saccharomyces cerevisiae</i> (budding yeast)	799	1.54	10
2	 Biosynthesis of amino acids - <i>Saccharomyces cerevisiae</i> (budding yeast)	105	6.48	7
3	 Biosynthesis of secondary metabolites - <i>Saccharomyces cerevisiae</i> (budding yeast)	285	2.78	7
4	 Biosynthesis of antibiotics - <i>Saccharomyces cerevisiae</i> (budding yeast)	201	3.57	6
5	 Cysteine and methionine metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	40	12.5	5
6	 Aminoacyl-tRNA biosynthesis - <i>Saccharomyces cerevisiae</i> (budding yeast)	24	11.11	5
7	 Glycine, serine and threonine metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	29	12.9	4
8	 Glutathione metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	20	10	3
9	 Phenylalanine, tyrosine and tryptophan biosynthesis - <i>Saccharomyces cerevisiae</i> (budding yeast)	21	9.09	2
10	 Arginine and proline metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	27	7.41	2
11	 Ubiquinone and other terpenoid-quinone biosynthesis - <i>Saccharomyces cerevisiae</i> (budding yeast)	1	50	1
12	 Cyanoamino acid metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	4	11.11	1
13	 beta-Alanine metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	8	8.33	1
14	 Sphingolipid metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	14	7.69	1
15	 Lysine biosynthesis - <i>Saccharomyces cerevisiae</i> (budding yeast)	14	6.25	1
16	 Tyrosine metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	16	5.88	1
17	 Arginine biosynthesis - <i>Saccharomyces cerevisiae</i> (budding yeast)	16	5.56	1
18	 One carbon pool by folate - <i>Saccharomyces cerevisiae</i> (budding yeast)	21	5.26	1
19	 Methane metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	16	5	1
20	 Valine, leucine and isoleucine biosynthesis - <i>Saccharomyces cerevisiae</i> (budding yeast)	17	5	1
21	 Glyoxylate and dicarboxylate metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	15	4.55	1
22	 Alanine, aspartate and glutamate metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	24	4.17	1
23	 Glycerophospholipid metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	42	2.86	1
24	 Carbon metabolism - <i>Saccharomyces cerevisiae</i> (budding yeast)	73	1.72	1

Which pathways are significant?

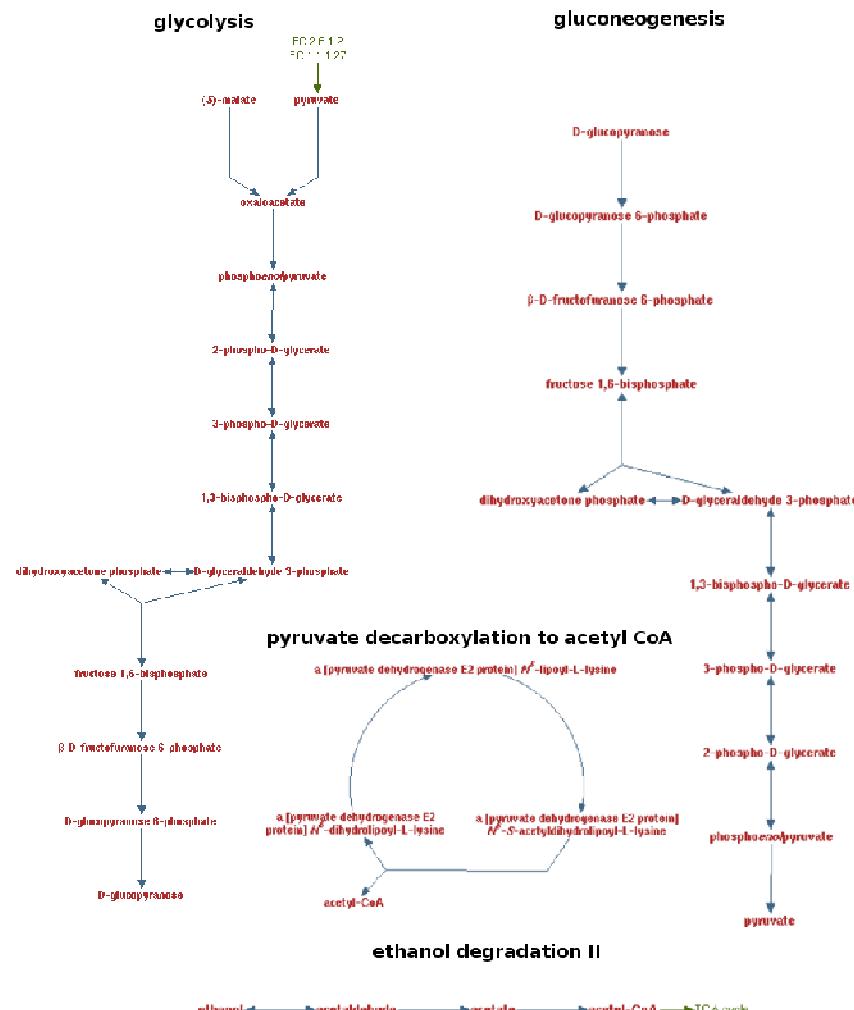
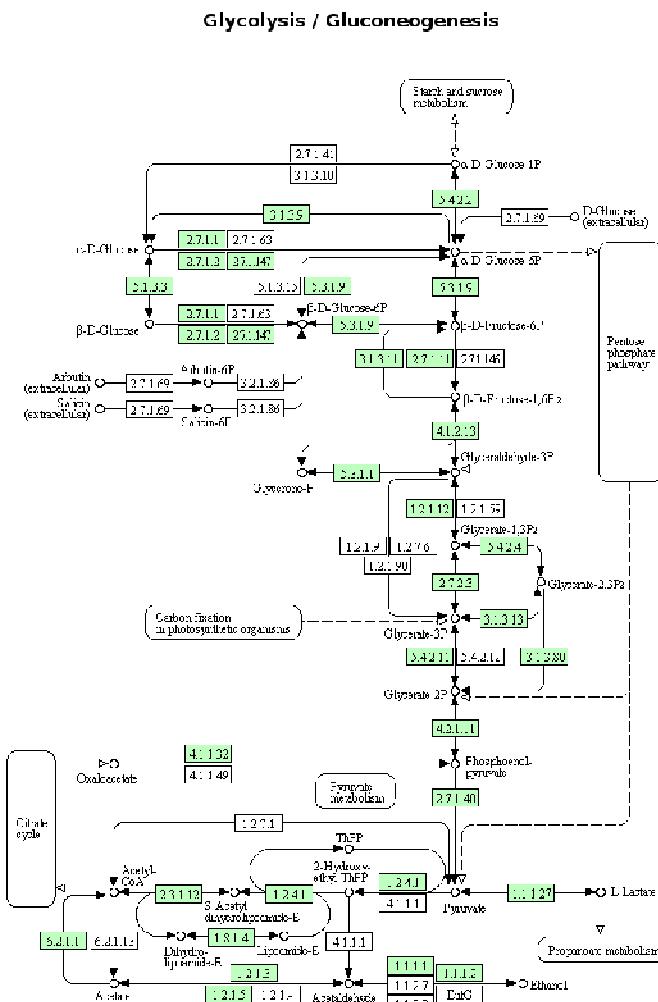
Limits of pathway mapping: connectivity



“ Even at the pathway level,
not easy to identify which
reactions
procude/consume L-serine

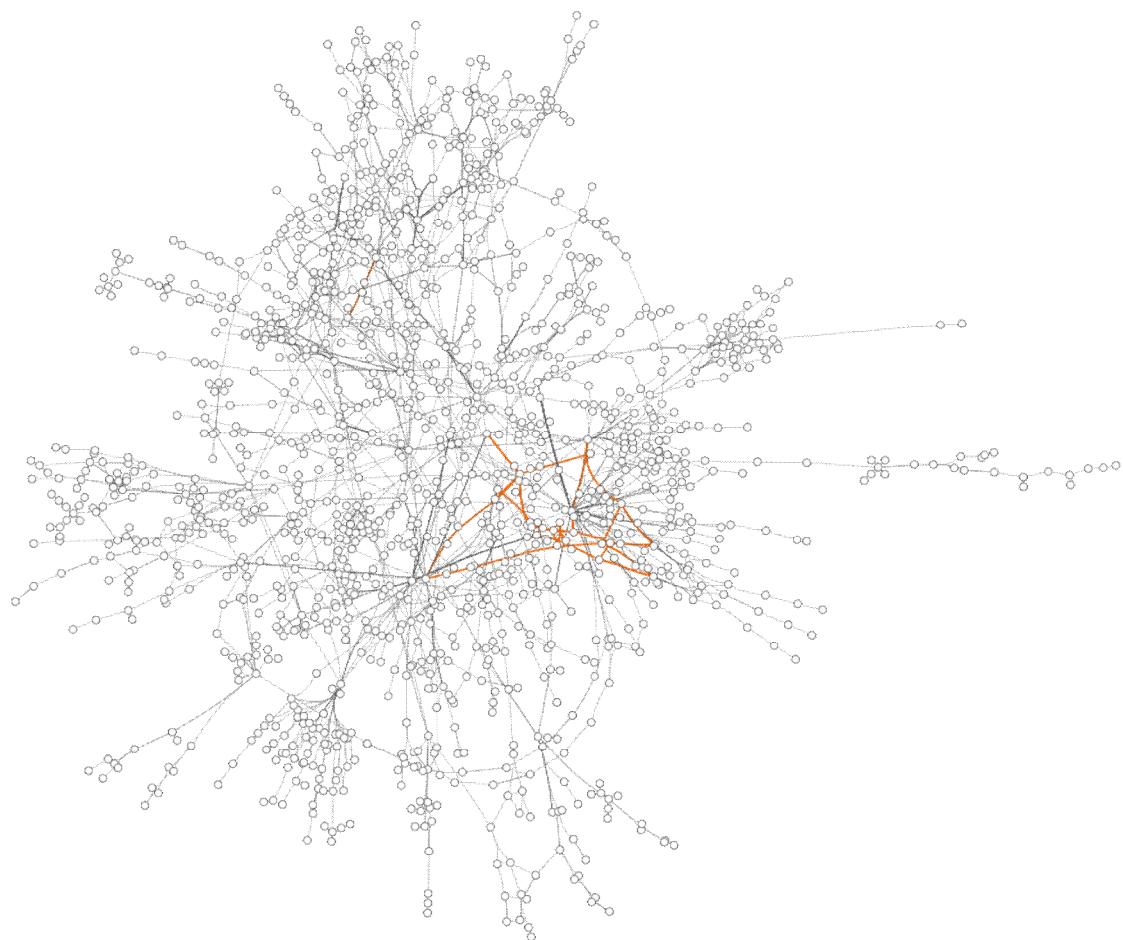
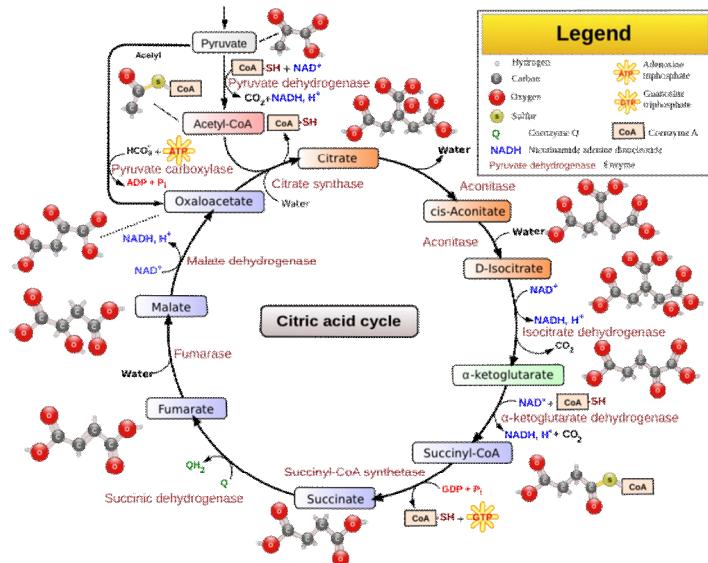
“ How to incorporate
arginine in the picture?

Pathway boundary is a versatile notion in databases



Taking into account full complexity

Pathways vs. Networks

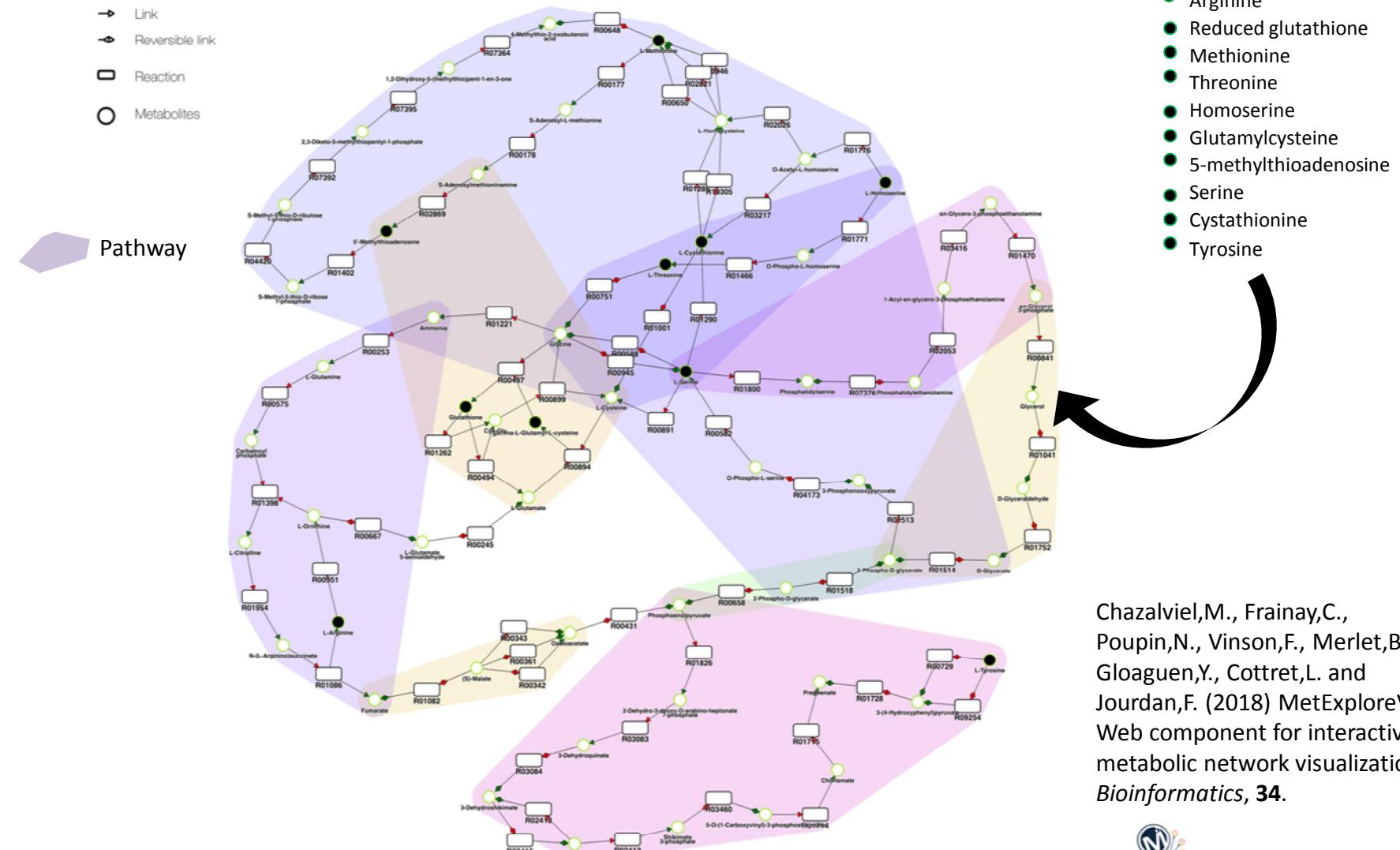


Credit: Oscar Yanes



Analyse metabolites in the context of networks

Metabolic modulations may span several pathways

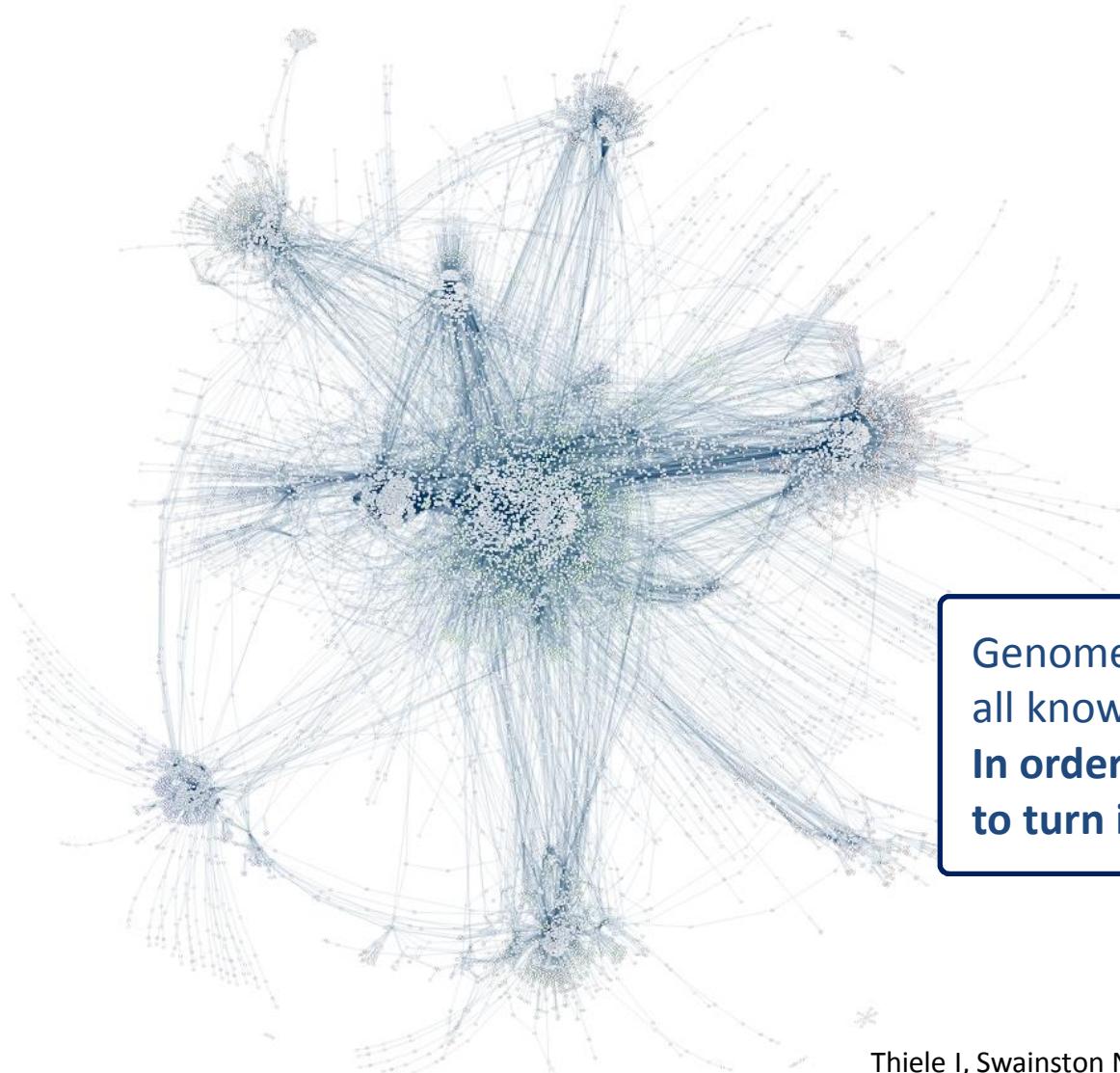


Chazalviel,M., Frainay,C.,
Poupin,N., Vinson,F., Merlet,B.,
Gloaguen,Y., Cottret,L. and
Jourdan,F. (2018) MetExploreViz:
Web component for interactive
metabolic network visualization.
Bioinformatics, **34**.



MODEL GLOBAL METABOLISM - *GRAPHS*

Genome-scale metabolic networks



Recon 2 (human)
7440 reactions
2626 metabolites
1733 genes

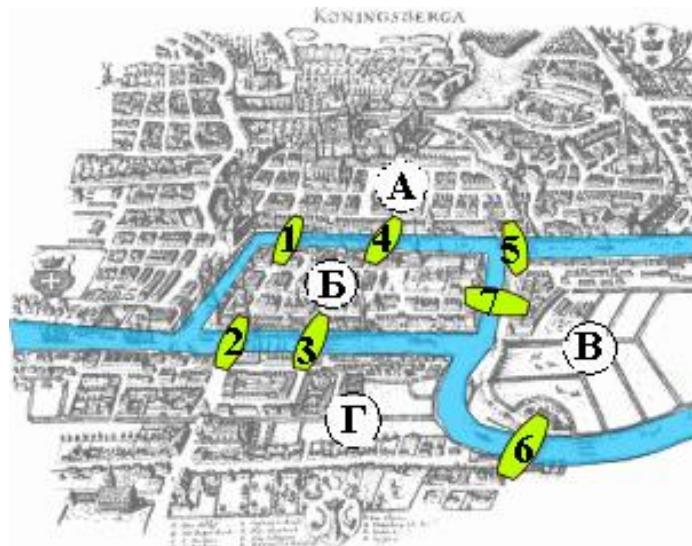
Genome scale metabolic network contains all knowledge.
In order to perform computation we need to turn it into a mathematical model.

Thiele I, Swainston N, Fleming RMT, et al. A community-driven global reconstruction of human metabolism. Nat. Biotechnol. 2013; 31:419–25



From knowledge to model

Leonhard Euler “Seven Bridges of Königsberg”, published in 1735.



« If we compared the Bernoullis to the Bach family, then Leonhard Euler is unquestionably the Mozart of mathematics »

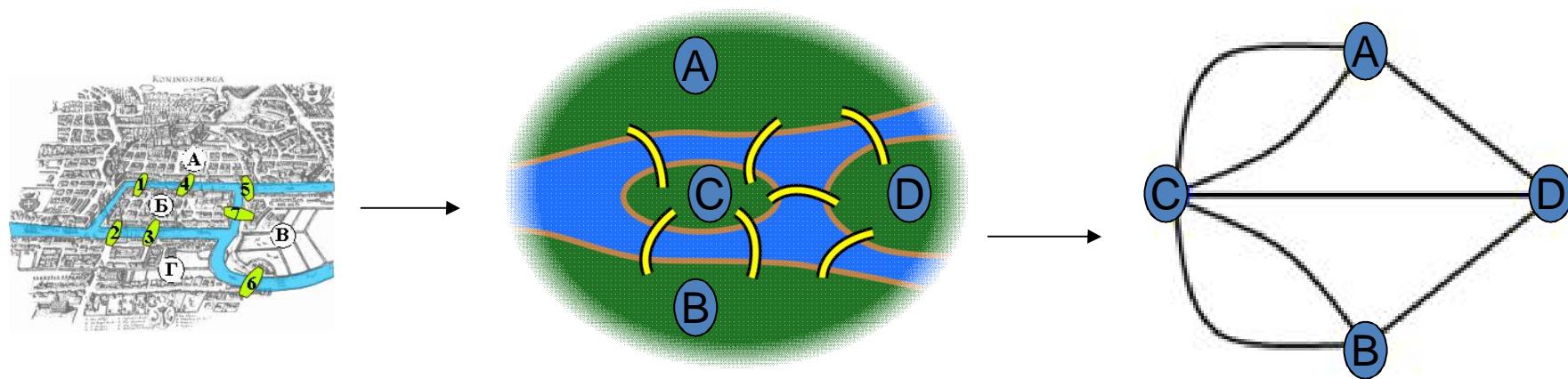
Eli Maor



Is there a path which starts from one point, finishes in another point and goes through all bridges without crossing the same bridge twice?

From knowledge to model

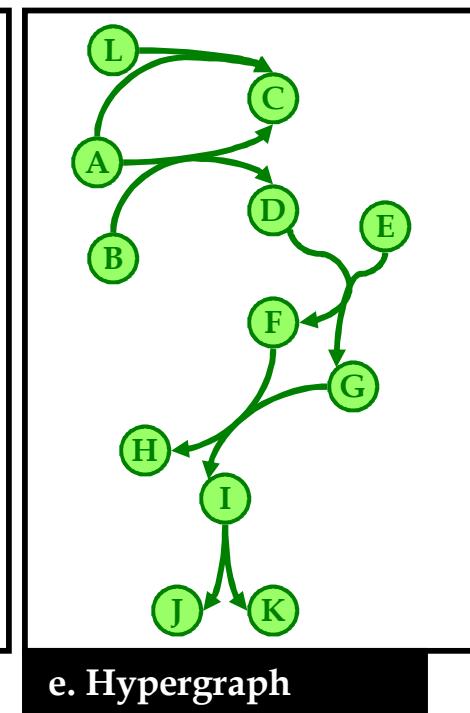
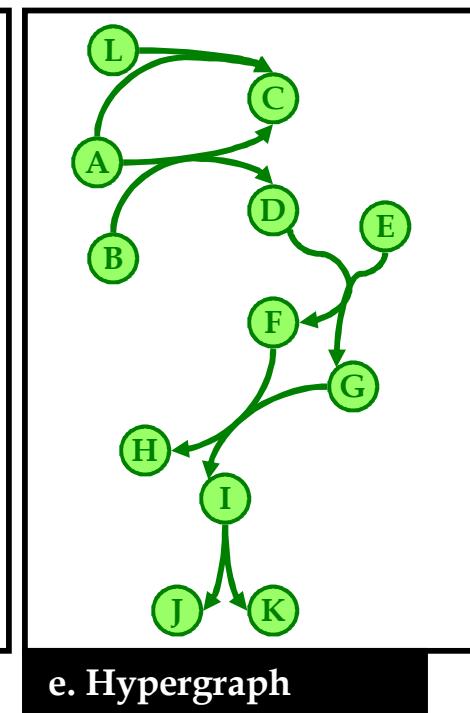
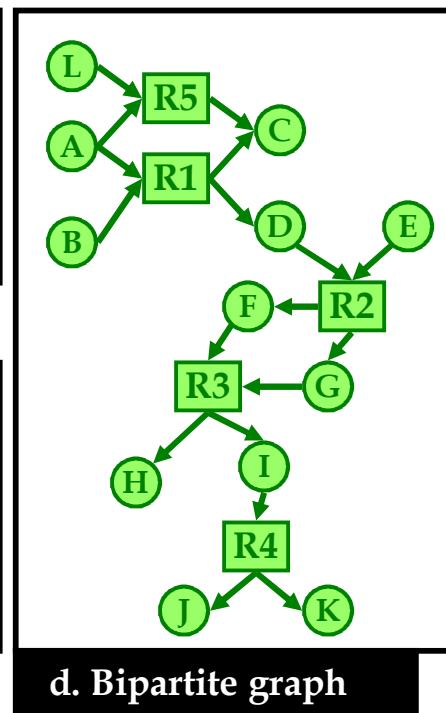
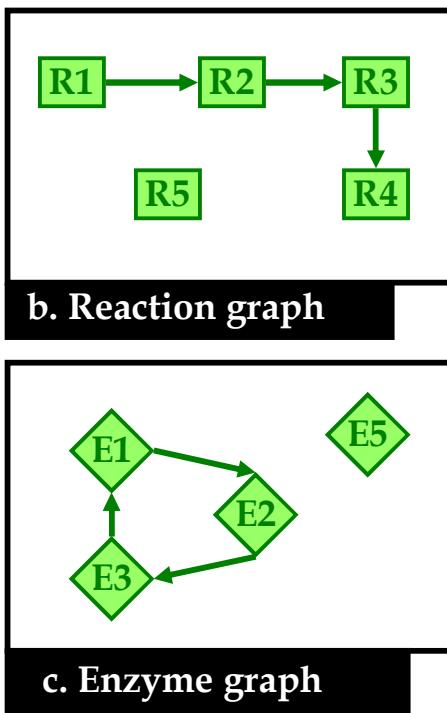
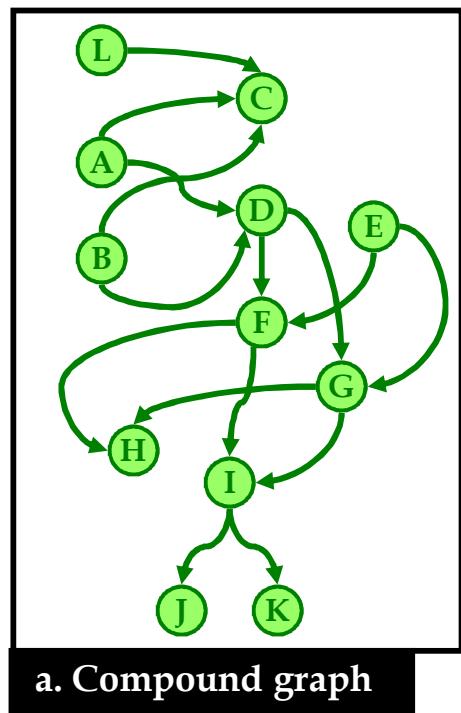
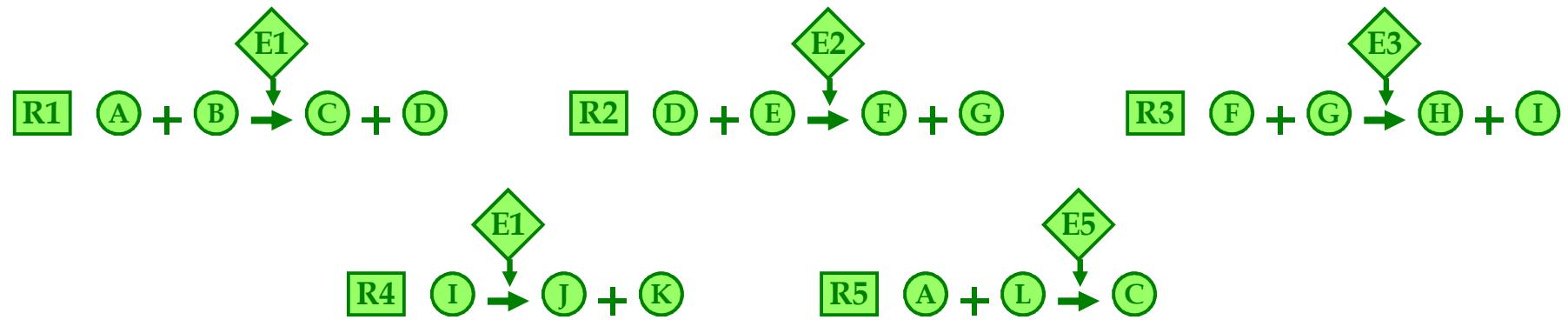
Leonhard Euler “Seven Bridges of Königsberg”, published in 1735.



A necessary condition for the walk of the desired form is that the graph be connected and have exactly zero or two nodes of odd degree.

Graph modelling(s) of metabolic networks

Textual description of the network

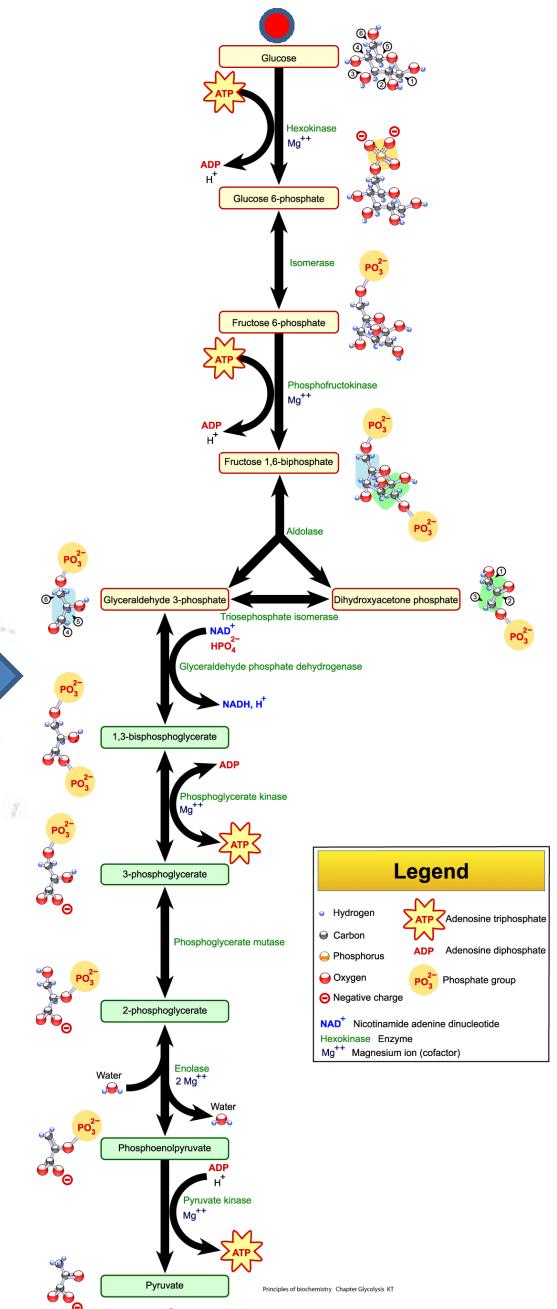
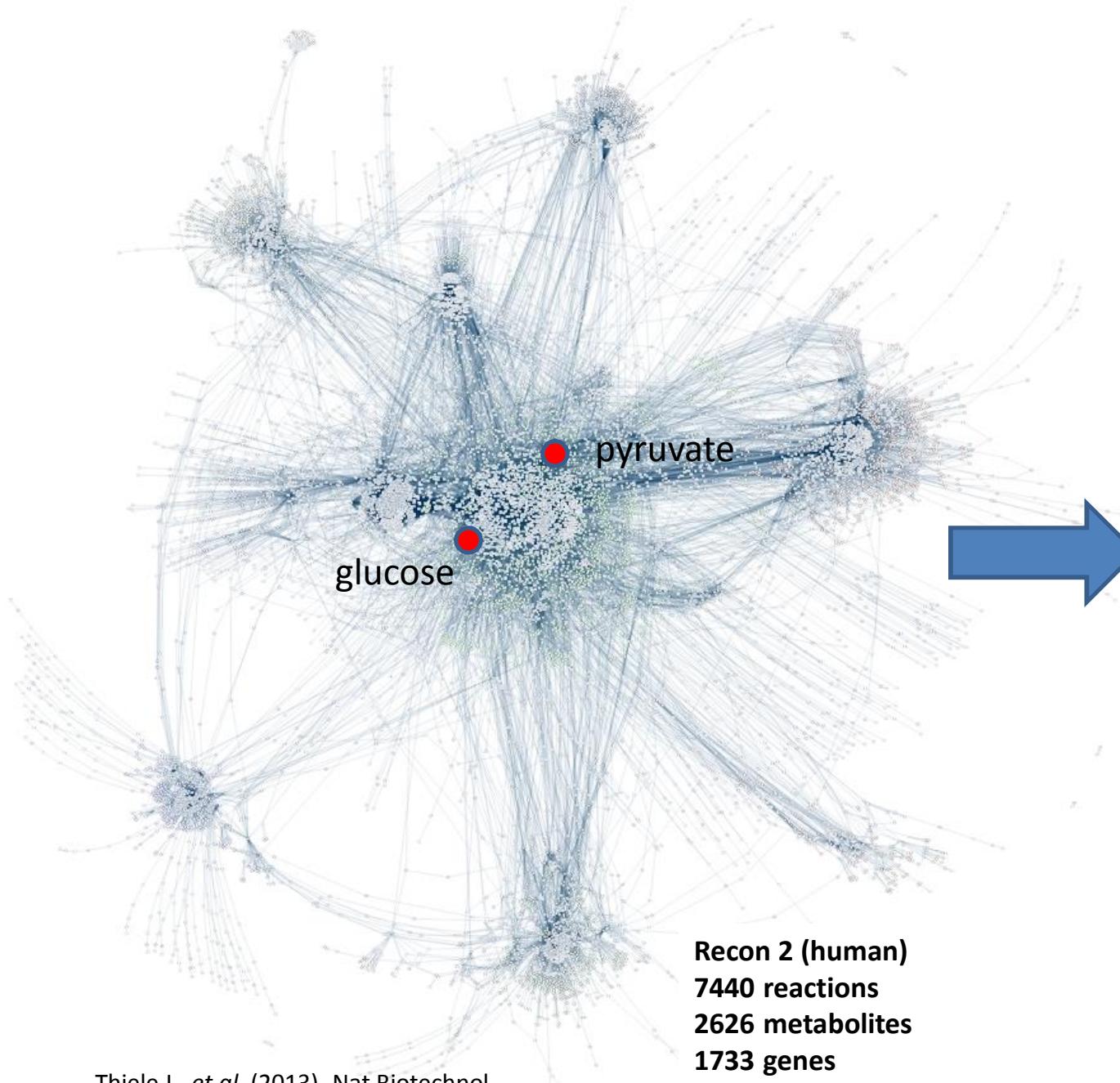


Network jungle

- „ Biological networks:
 - . Genome scale metabolic networks
 - „ Reaction networks
 - „ Metabolite networks
 - „
 - . Other biological networks (regulatory, trophic ...)
- „ Molecular networks:
 - . ab initio (mass differences)
 - . Spectral similarity networks (GNPS)
- „ Statistical networks:
 - . Correlation networks
 - . Gaussian model networks

Important to know what nodes and edges are!!

SUGGEST INTERPRETATION - *ALGORITHMS*

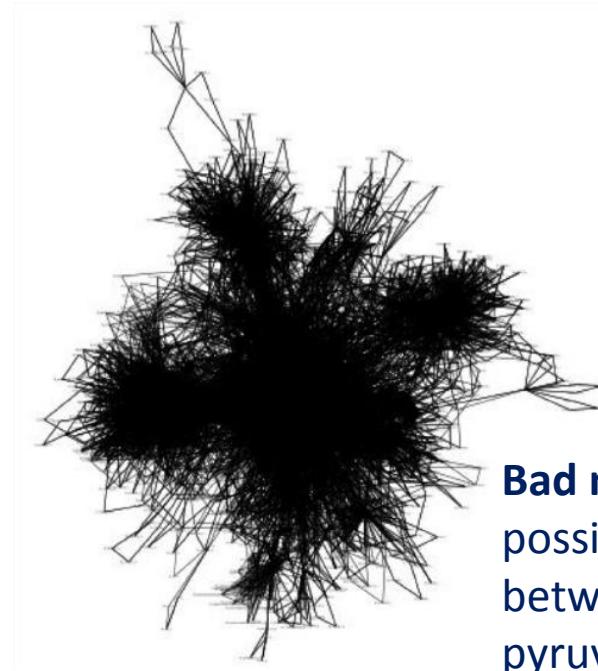


Thiele I., et al. (2013). Nat Biotechnol.

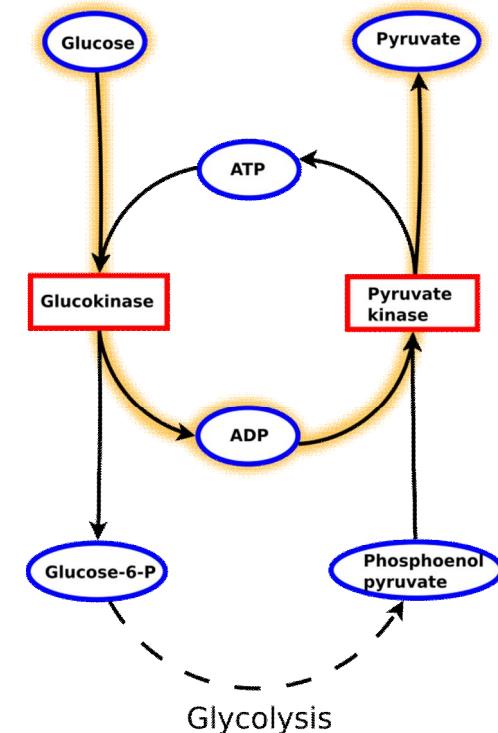


Problem complexity: going from glucose to pyruvate

Good news: using graph algorithms we can compute paths between metabolites in the network.

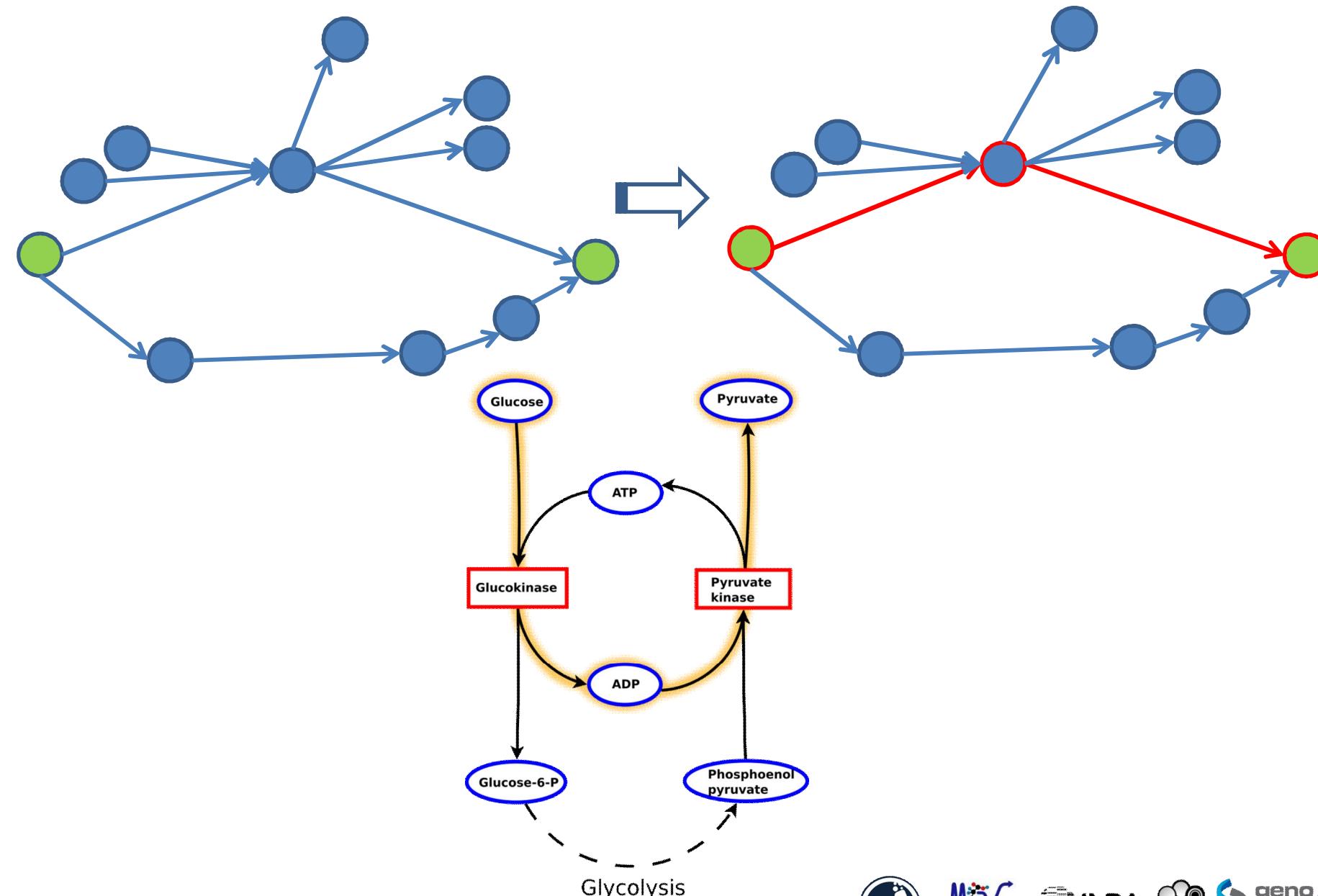


Bad news: 500 000
possible paths
between glucose and
pyruvate!

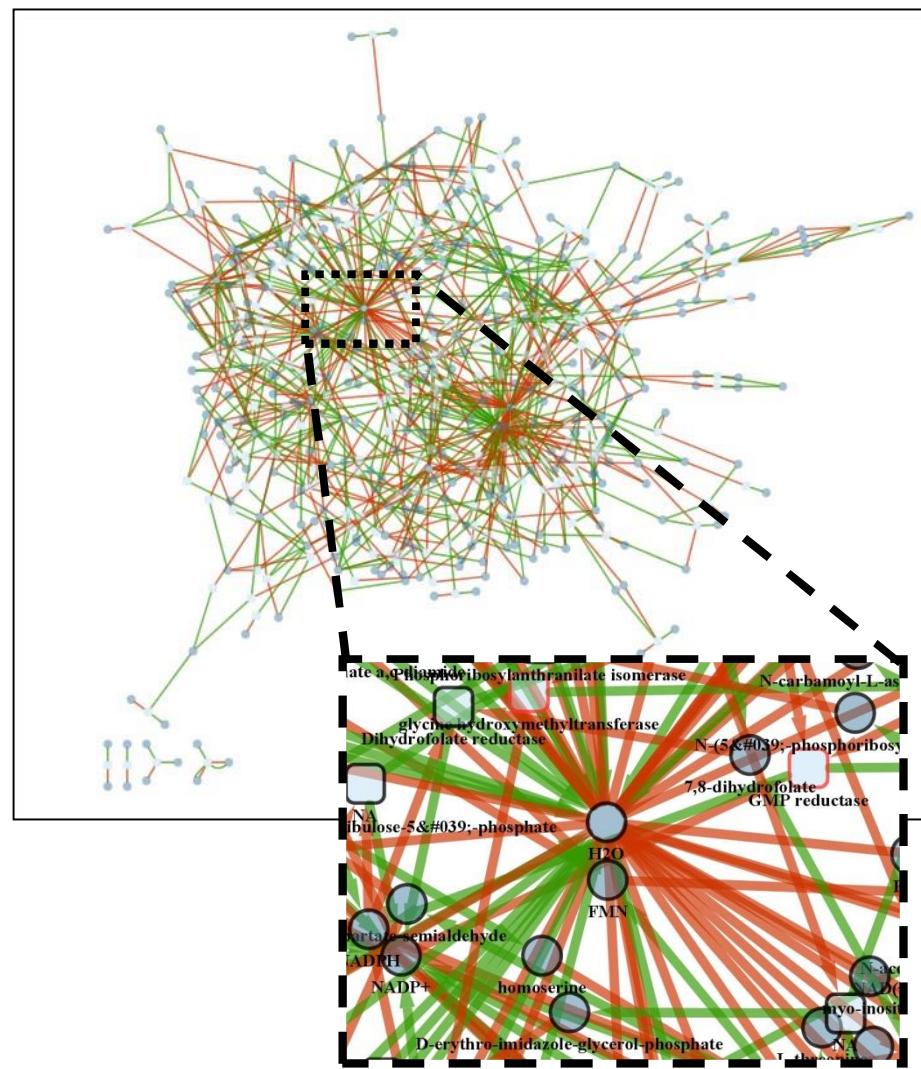


3. Küffner et al. (2000). Pathway analysis in metabolic databases via differential metabolic display (DMD).
Bioinformatics.

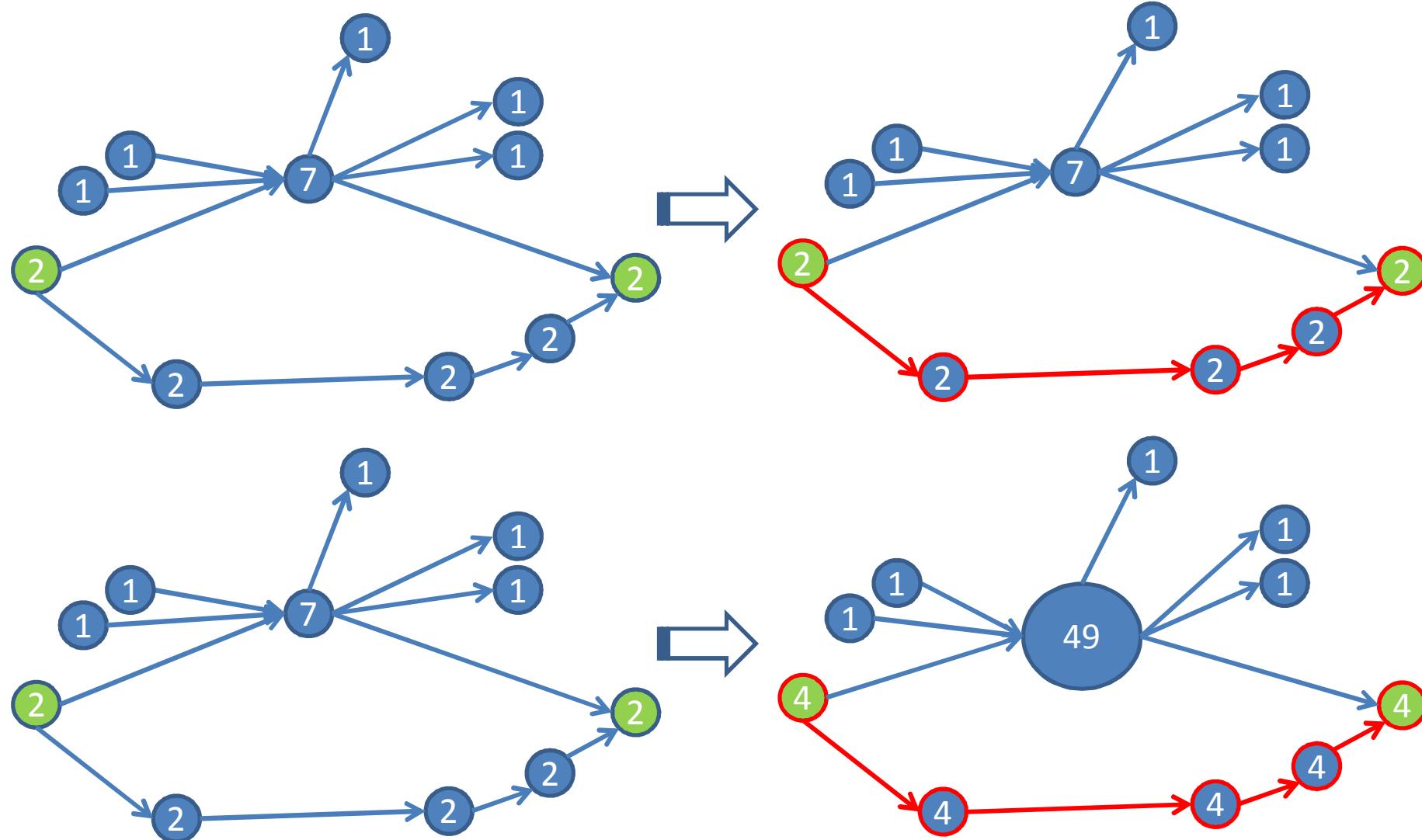
Shortest path



Using the topology to avoid side compounds



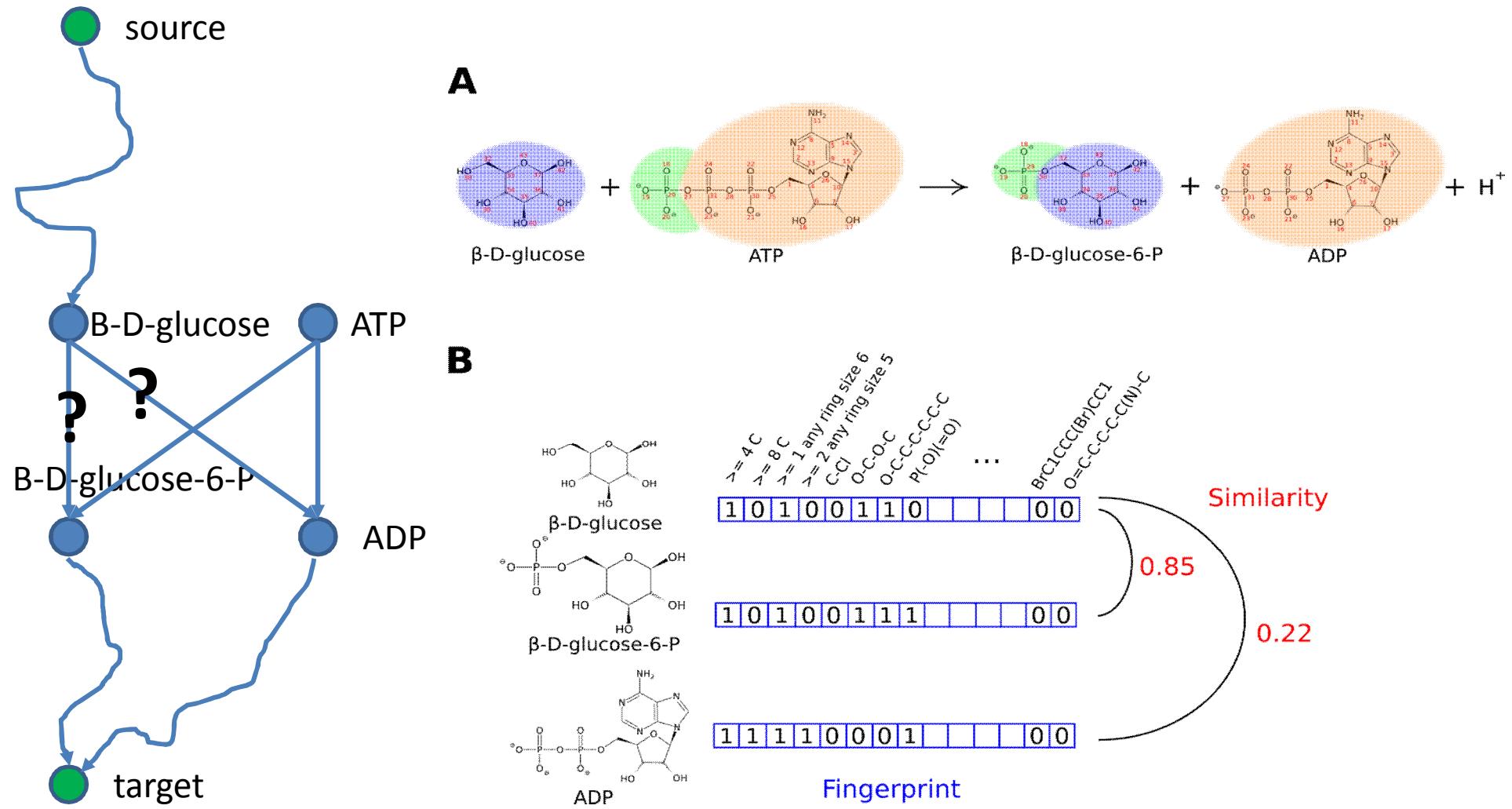
Lightest path and Lightest path²



Faust K, van Helden J. Predicting metabolic pathways by sub-network extraction. Methods Mol. Biol. 2012; 804:107–30

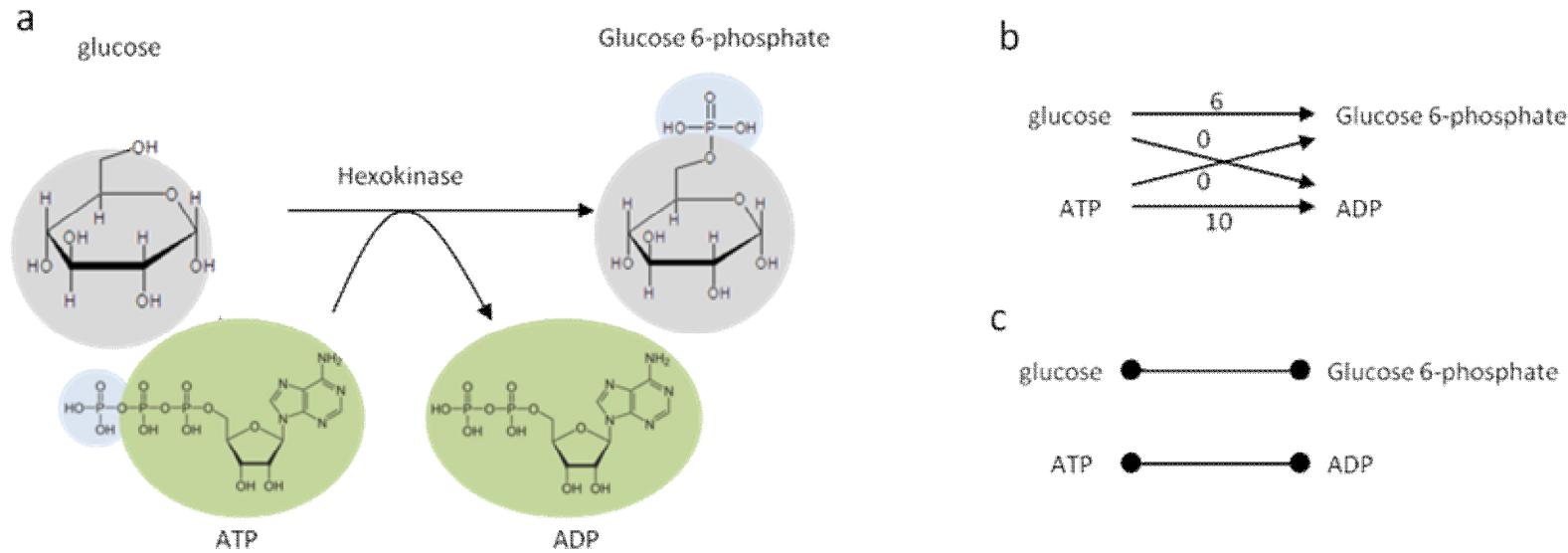


Using the chemistry to improve path search



Frainay C. & Jourdan F. Computational methods to identify metabolic sub-networks based on metabolomic profiles. 2016.
Briefings in Bioinformatics.

Adapting graph methods to take into account biochemistry



Atom transfer graph allows avoiding side compounds providing better topological analysis

Rahman,S.A., Torrance,G., Baldacci,L., Martínez Cuesta,S., Fenninger,F., Gopal,N., Choudhary,S., May,J.W., Holliday,G.L., Steinbeck,C., et al. (2016) Reaction Decoder Tool (RDT): extracting features from chemical reactions. *Bioinformatics*, **32**, 2065–2066.

Cerebellar ataxia with elevated cerebrospinal free sialic acid (CAFSA)

CSF metabolic fingerprint

Citric acid
Acetyl-alanine
N-acetyl-methionine
Acetylcarnitine
Aspartic acid
Malic acid
Cis-aconitic acid
Adipic acid
Choline
3-hydroxyisovalerylcarnitine
Tiglylcarnitine
Isovalerylcarnitine
Propionylcarnitine
Butyrylcarnitine
Carnitine
Adipoyl carnitine
Acetyl-glucosamine
N-acetylneurameric acid
Glyceric acid
Kynurenine
Tryptophan
Glutamine
Phenylacetil-glutamine
3-methyl-2-oxovaleric acid
Porphobilinogen
Glycocholic acid
Tyrosine
Phenylalanine
Dopamine sulfate
Creatine
Hippuric acid
Indole-3-acetate
Creatinine
Inosine
Methylthioadenosine
Deoxyribose
Uric acid
Xanthine
Panthothentic acid
Folate
Threonic acid
4-Acetamidobutanoic acid
Glucitol



[Journal of Inherited Metabolic Disease](#)

May 2018, Volume 41, Issue 3, pp 447–456 | [Cite as](#)

Targeted versus untargeted omics — the CAFSA story

Authors

Authors and affiliations

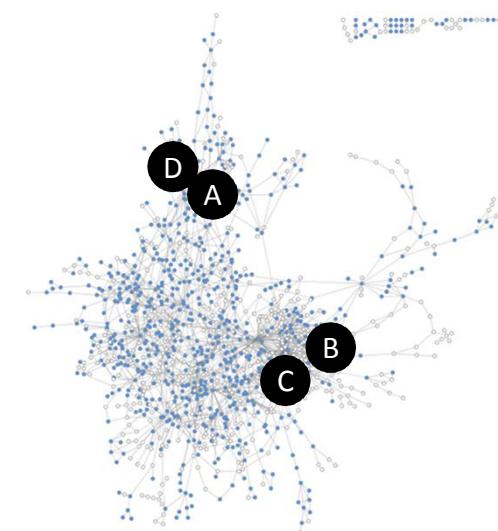
Maria del Mar Amador, Benoit Colsch, Foudil Lamari, Claude Jardel, Farid Ichou, Agnès Rastetter, Frédéric Sedel, Fabien Jourdan, Clément Trainay, Ronald A. Wevers, Emmanuel Roze, Christel Depienne, Christophe Junot, Fanny Mochel [✉](mailto:moche@inra.fr)

Is the disease affecting globally metabolism?

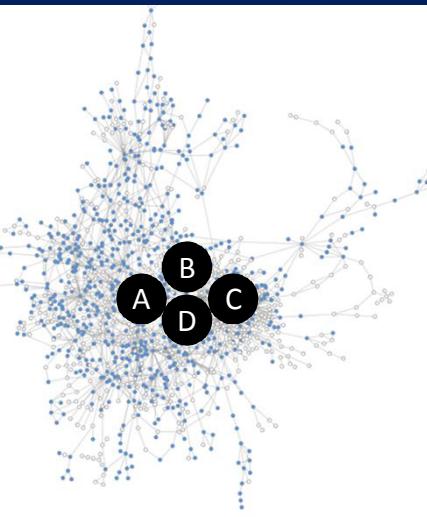
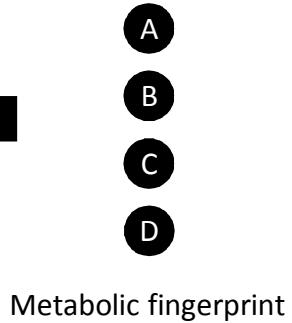
Can we divide the fingerprint into various metabolic modules?

Can we detect a small set of biochemical reactions connecting some metabolites in the fingerprint?

In which part of the metabolic network is the modulation taking place?

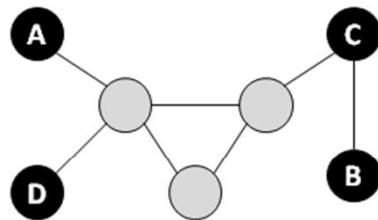


Modulation is spread in the network



Modulation is located in the same « area » of the network

a



✗ Metabolite in fingerprint

○ Metabolite not in fingerprint

— Compound to compound connection

b

0	4	3	2
4	0	1	4
2	1	0	3
2	4	3	0

A
B
C
D

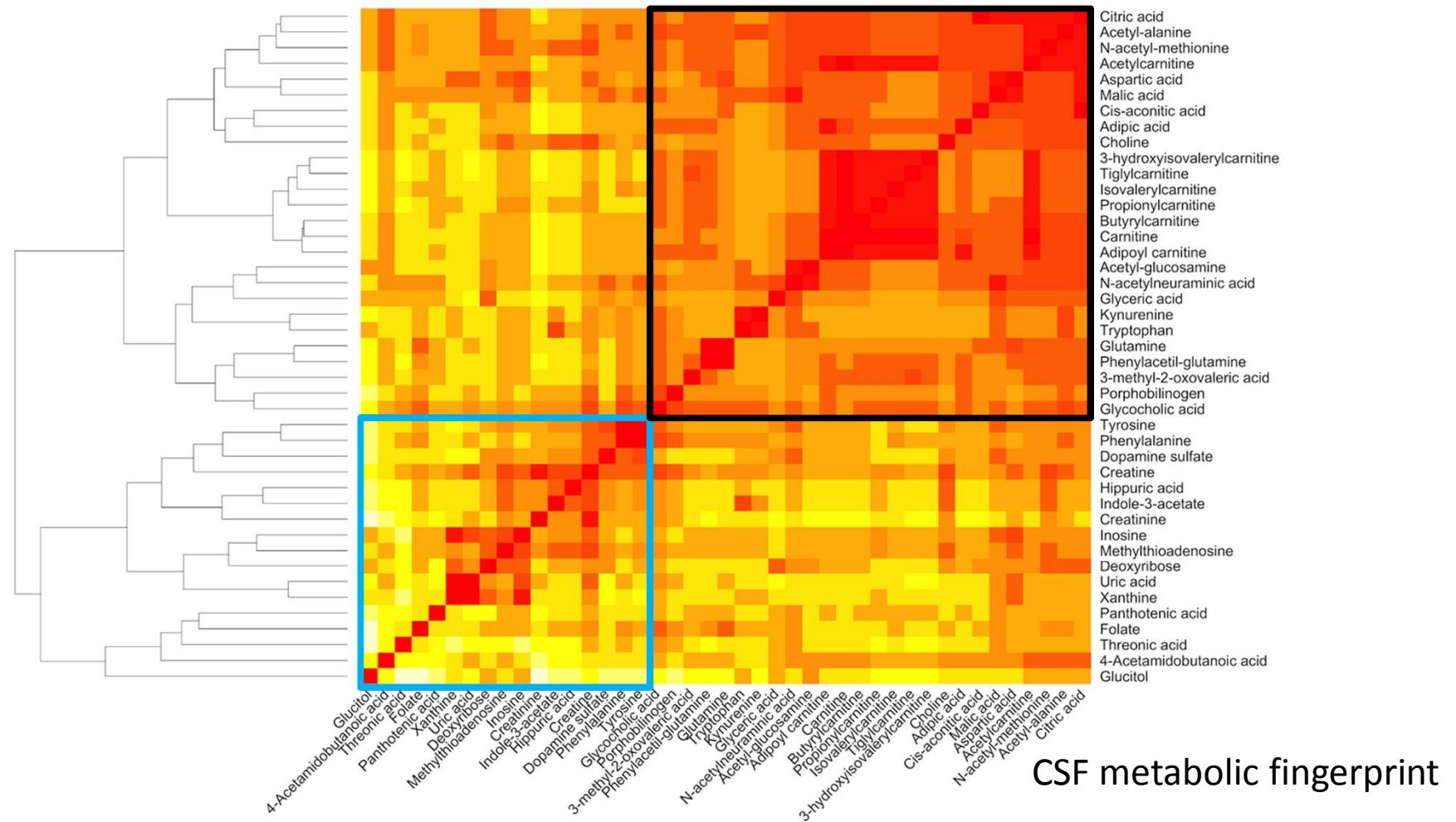
A
B
C
D

c

0	2	3	4	A
2	0	3	4	D
3	3	0	1	C
4	4	1	0	B

distances
0 1 2 3 4

Cerebellar ataxia with elevated cerebrospinal free sialic acid (CAFSA)



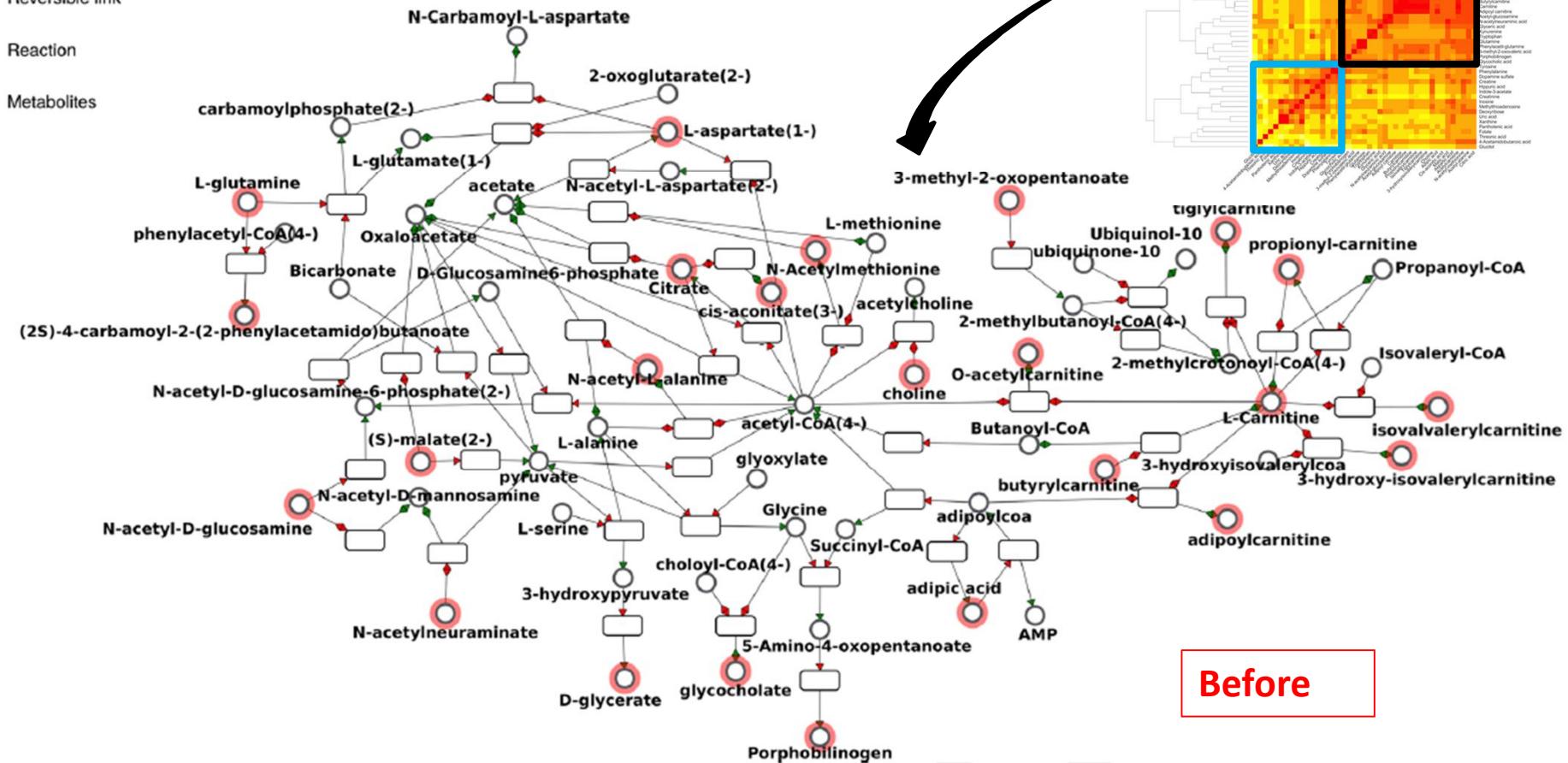
Based on biochemical distance in the network, it is possible to identify clusters in the fingerprint.

We now focus on the black cluster.

Visualisation proposed in the article

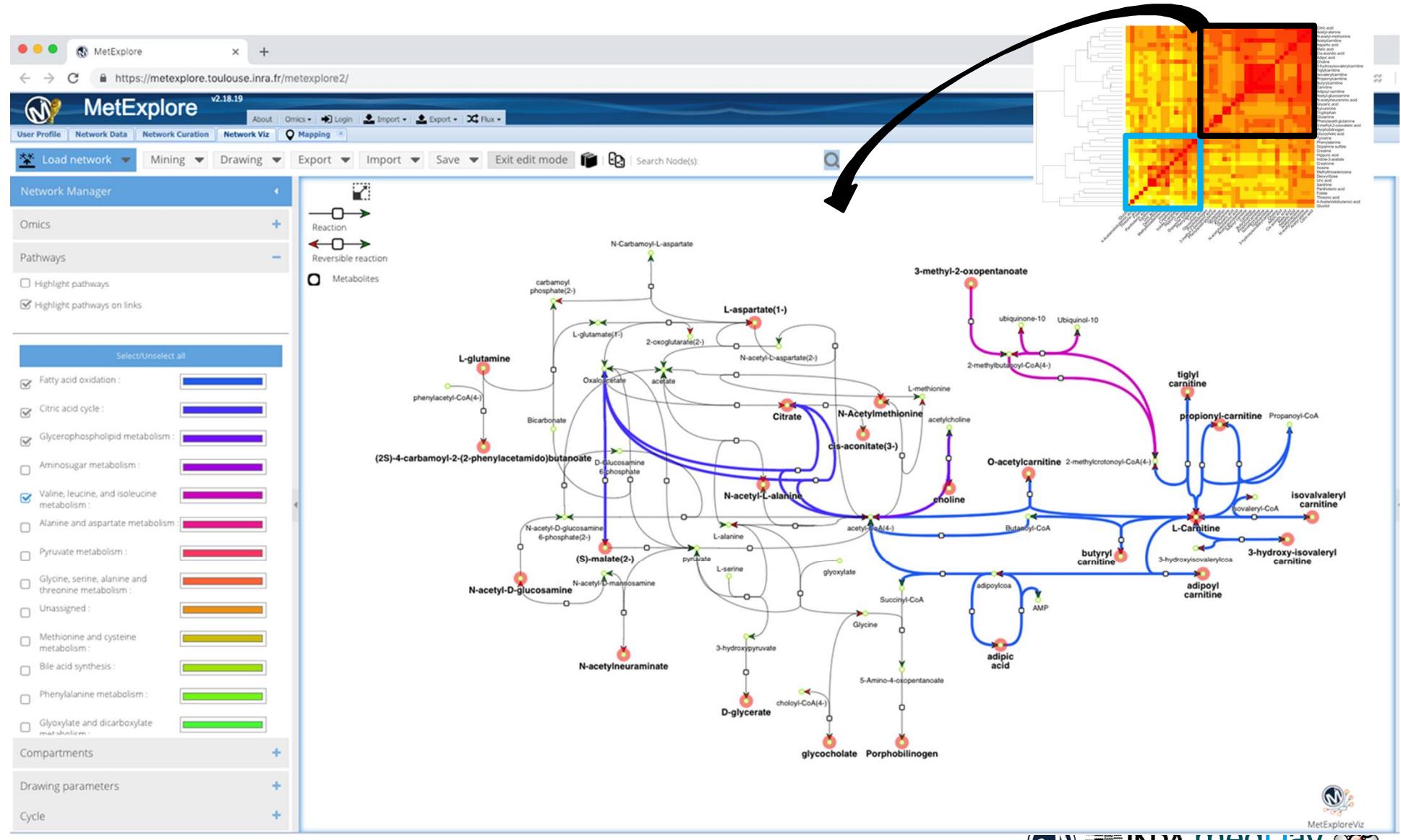
J Inherit Metab Dis

- Link
- ↔ Reversible link
- Reaction
- Metabolites

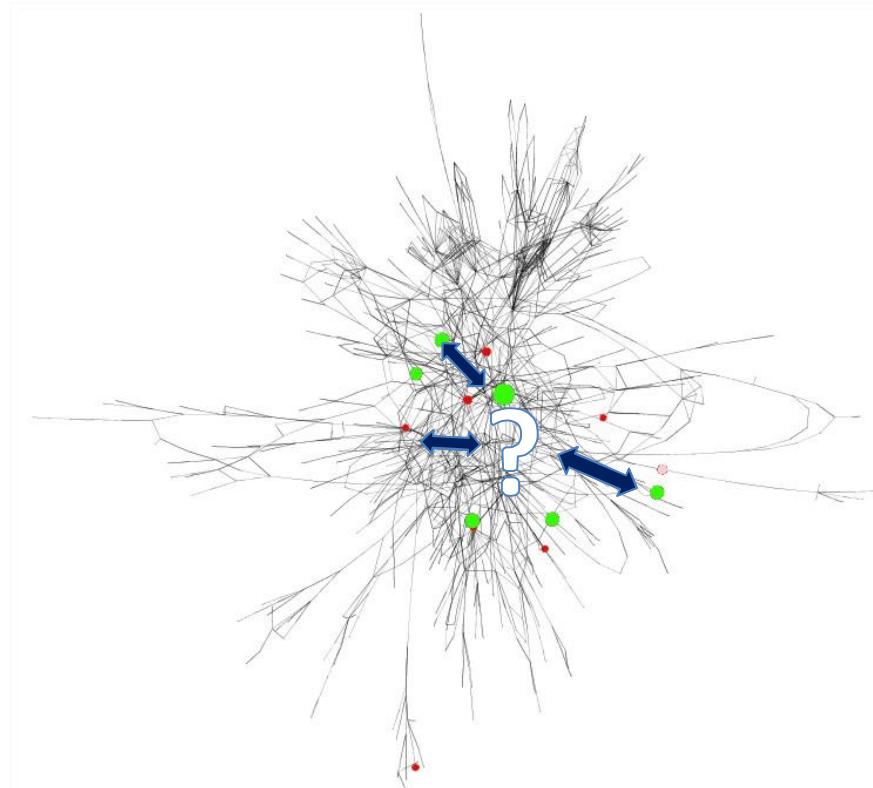


Before

Work on visualisation to get it closer to text book representations



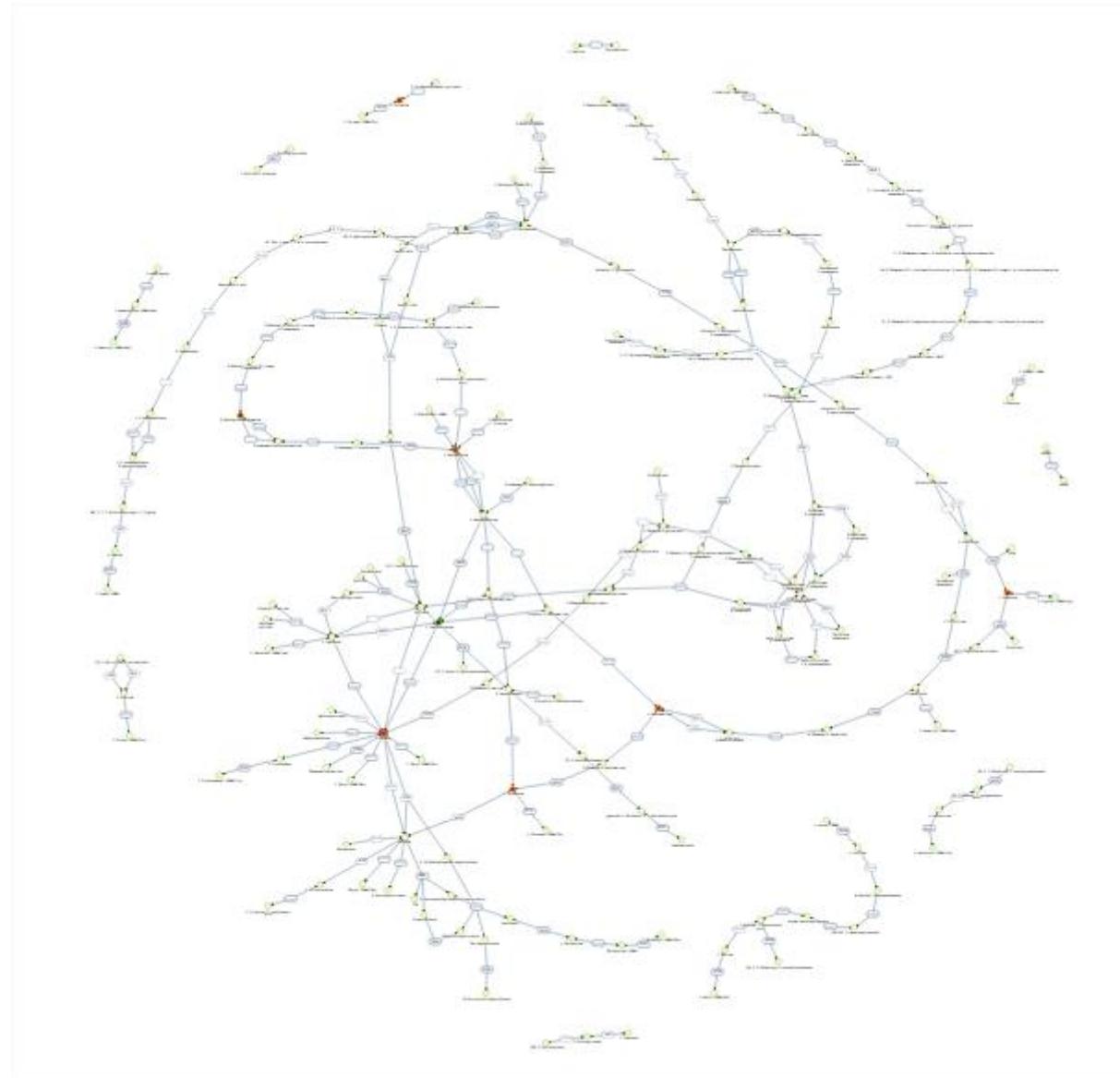
Connecting metabolites from a fingerprint



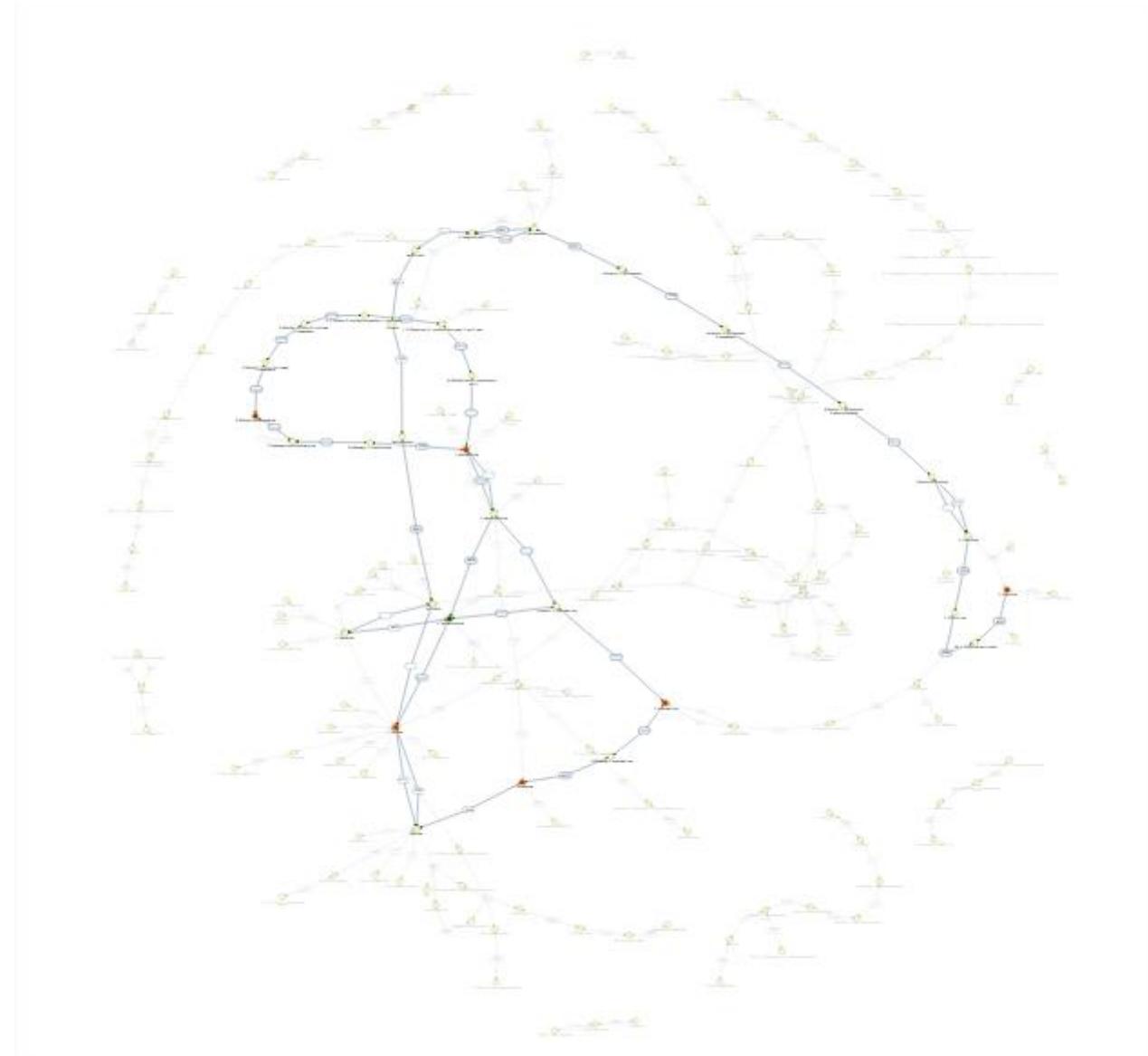
« Making sense of the soup »!
How can we connect metabolites?



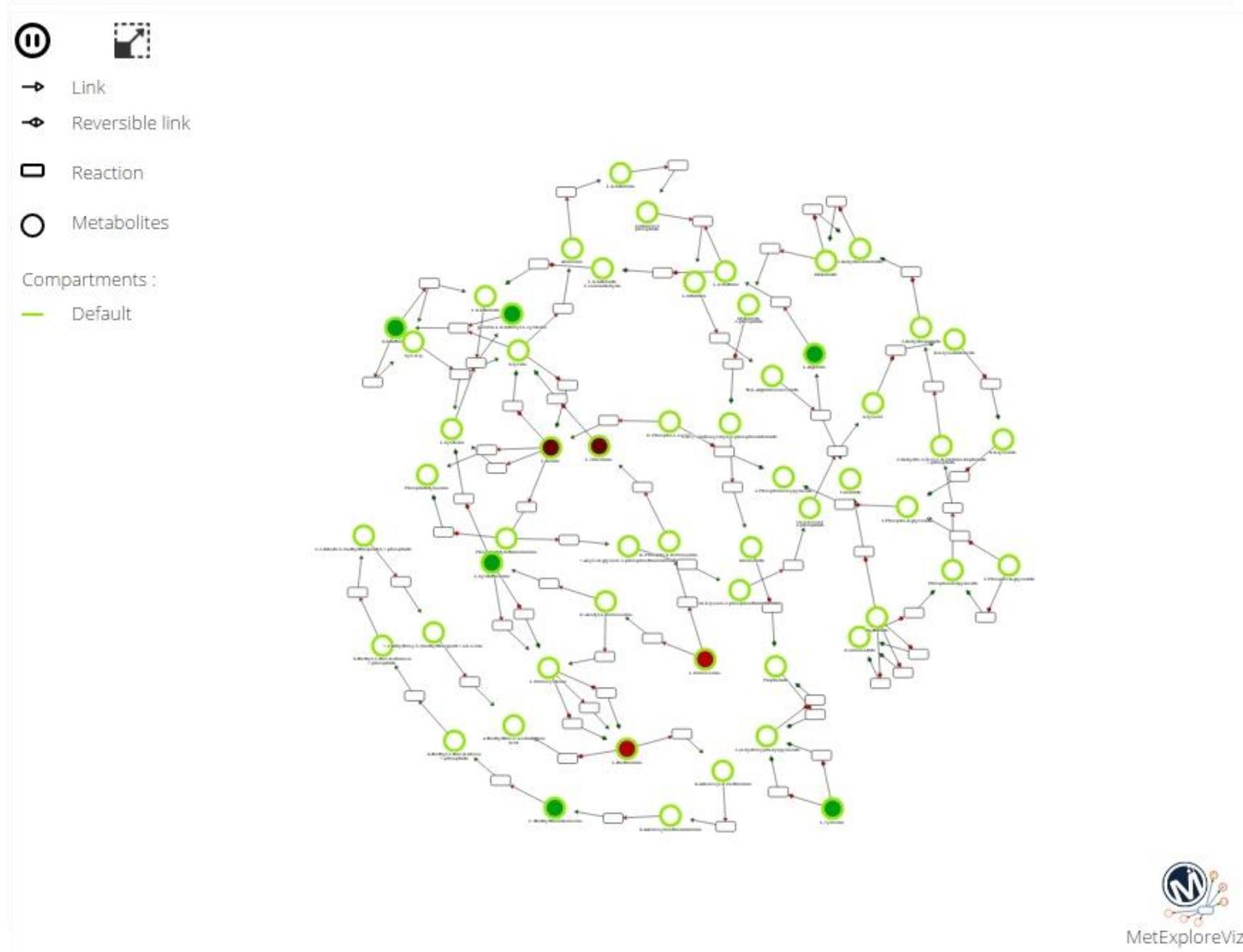
Union of lightest paths



Union of lightest paths

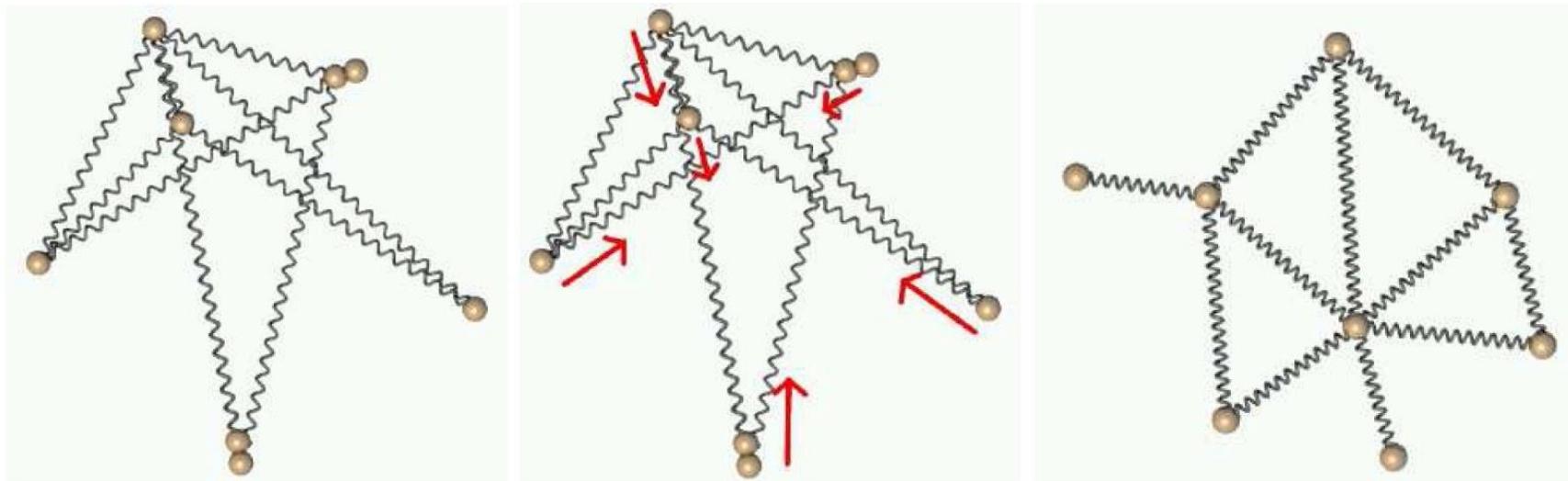


Sub-network extraction: union of all paths



VISUALISATION

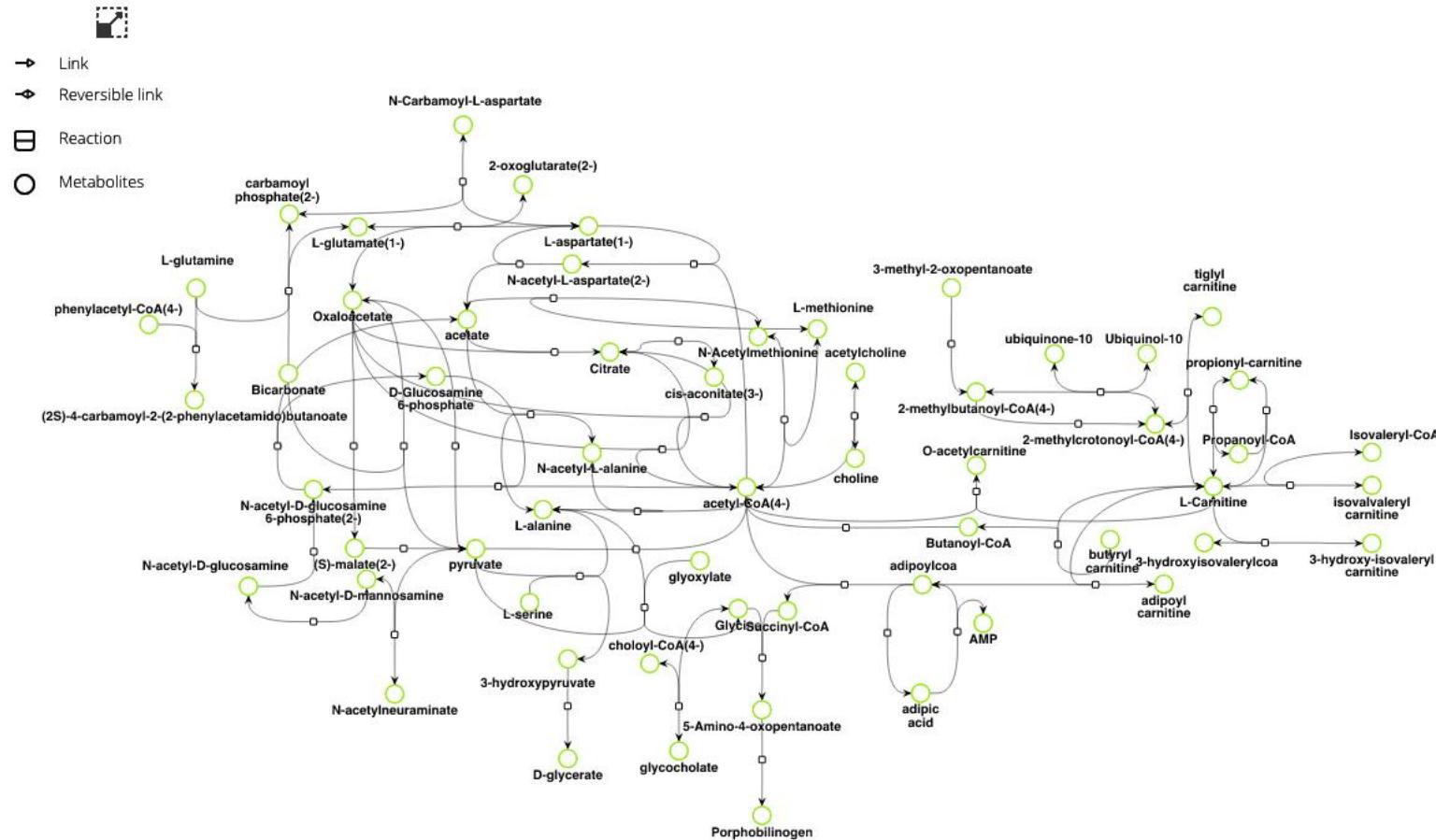
Force Directed Algorithm



Warning: non deterministic algorithm...means not always the main drawing !!

PRACTICE 2

New visualisation features in MetExplore



Data mapping

MetExplore version

User Profile Network Data Network Curation Network Viz

Load network Mining Drawing Export Import Save Edit mode Search Node(s):

Network Manager

Link Reversible link Reaction Metabolites

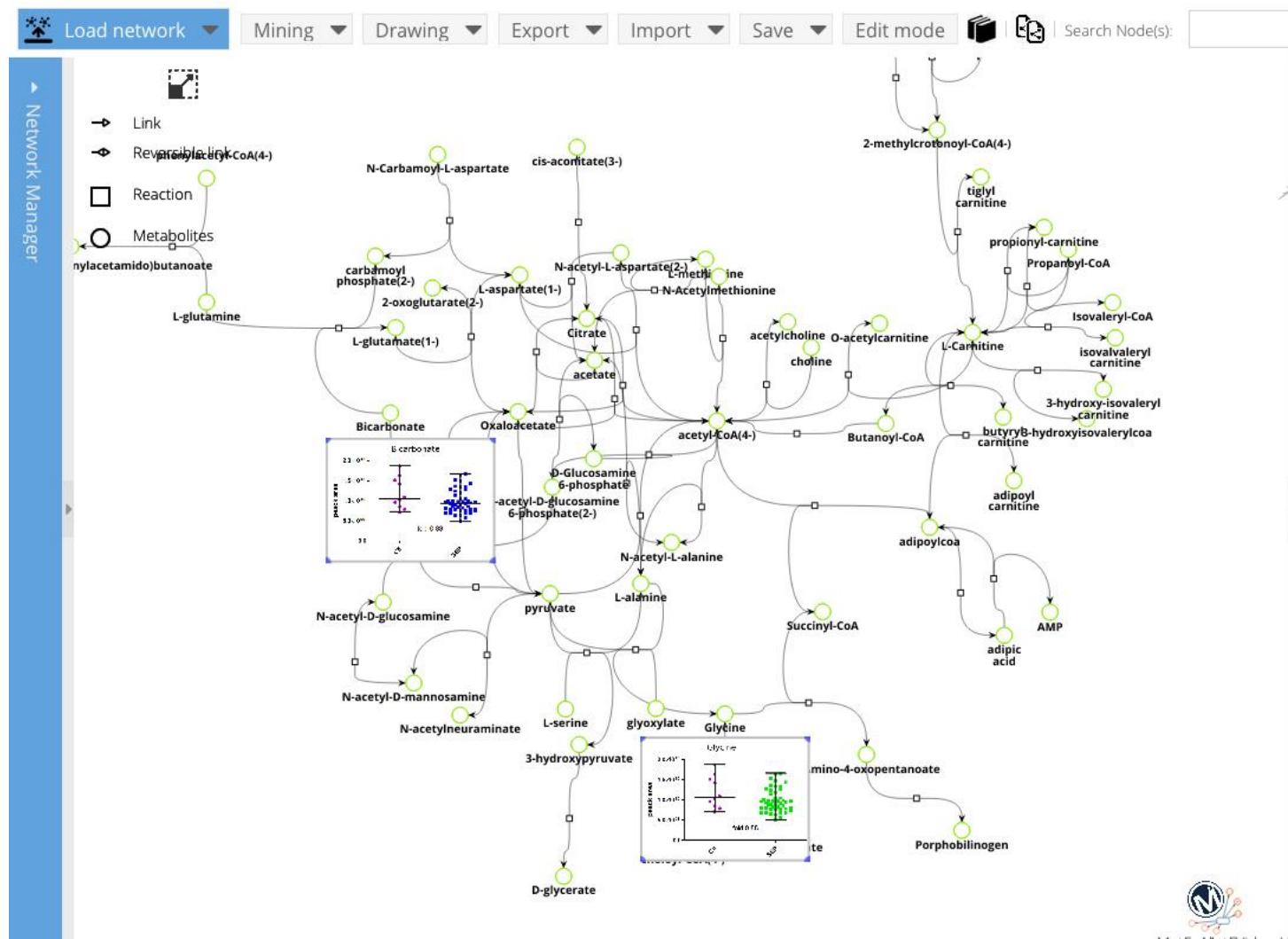
Favoris Nom Date de modification Tail

- Récents Tuto.txt hier à 11:29
- Téléchargeme... choline.png hier à 10:30
- iCloud Drive Bicarbonate.png hier à 10:27
- Bureau Glycine.png hier à 10:21
- Documents Glycine.jpg hier à 10:21
- Applications M_gly.jpg hier à 10:21
- Périphériques Bicarbonate.jpg 25 mai 2018 à 15:12
- MacBook Pro... choline.jpg 25 mai 2018 à 15:10
- Disque distant M_chol.jpg 25 mai 2018 à 15:10
- Glycine.svg 15 mai 2018 à 10:04
- M_hco3.svg 15 mai 2018 à 10:04
- M_hco3.jpg 9 mai 2018 à 16:58
- Test.json 16 avril 2018 à 09:43

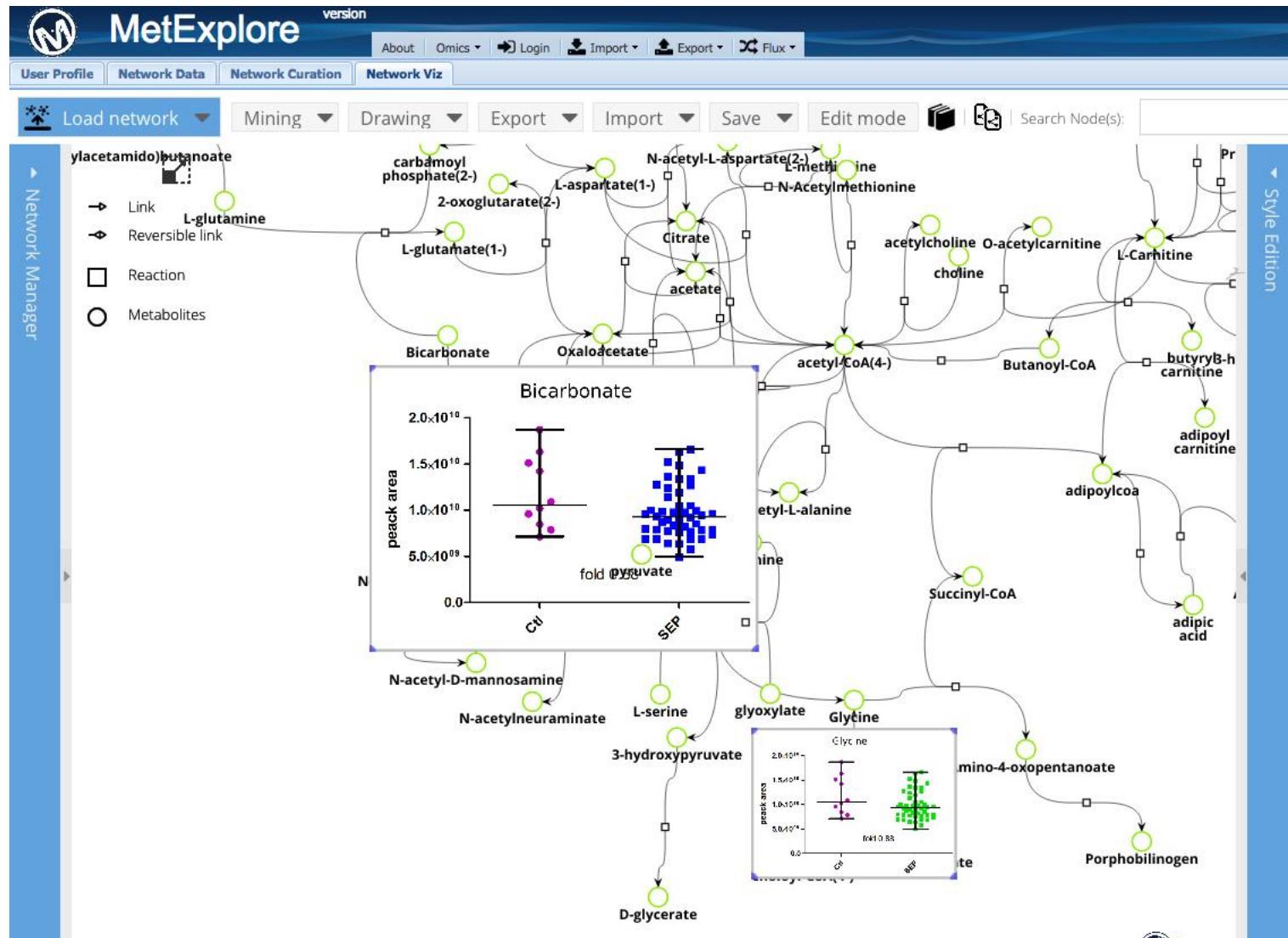
Masquer l'extension Annuler Ouvrir



Data mapping

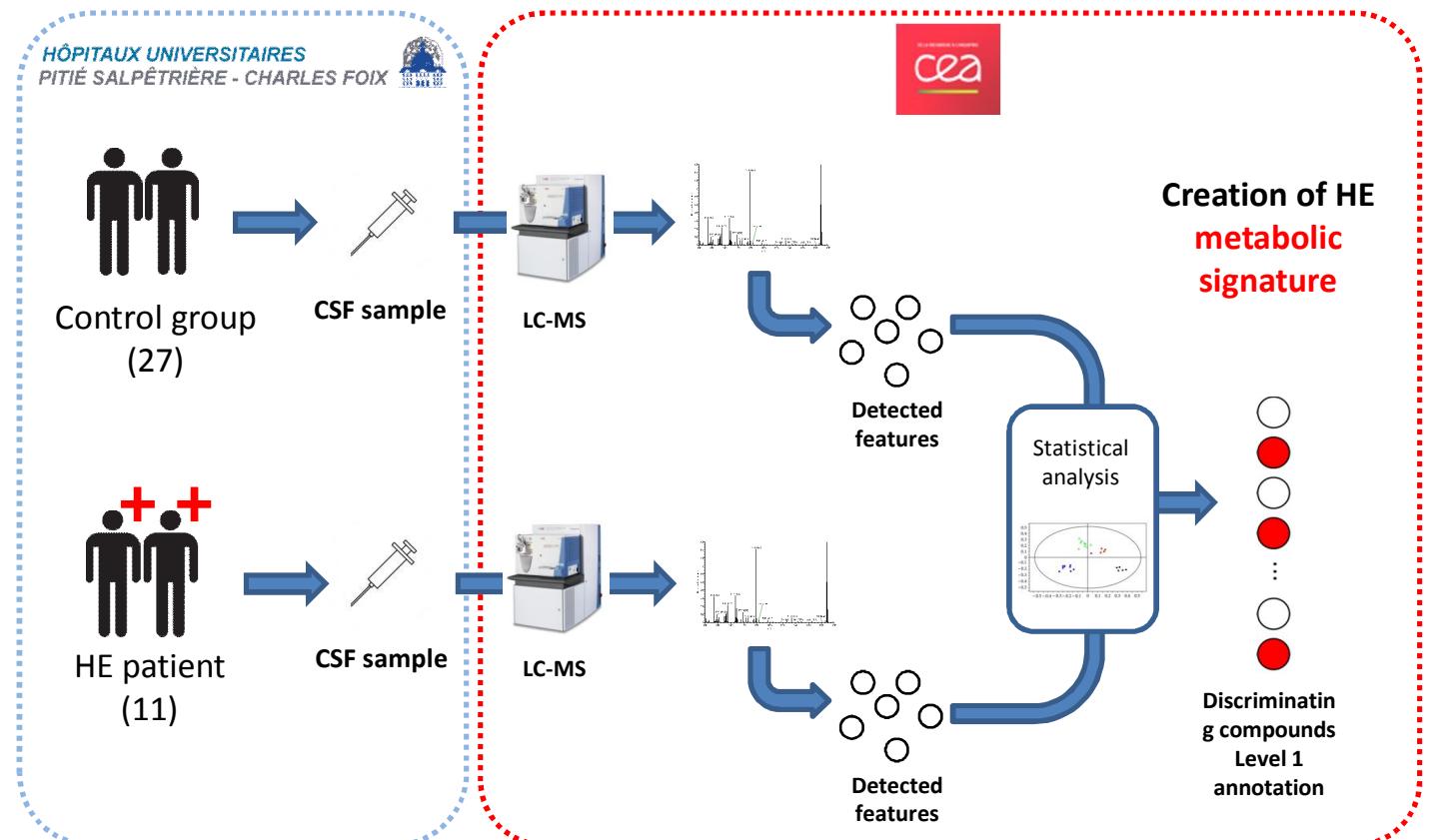


Data mapping



METABORANK

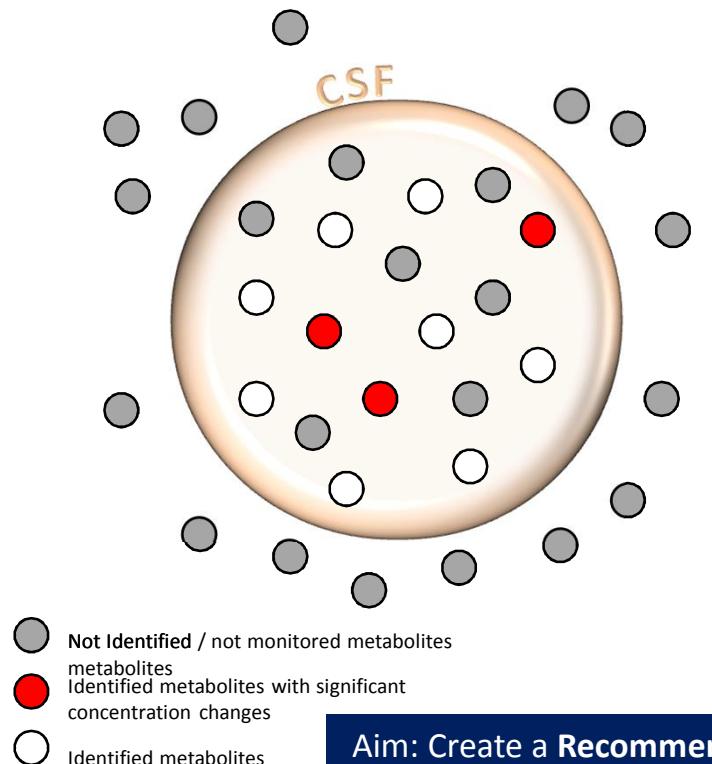
Experimental protocol



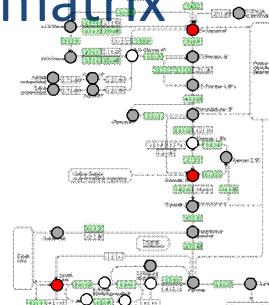
Weiss N, Barbier Saint Hilaire P, Colsch B, et al. Cerebrospinal fluid metabolomics highlights dysregulation of energy metabolism in overt hepatic encephalopathy. *J. Hepatol.* 2016; 65:1120–1130



Incompleteness of metabolic signatures



- Not identified
- Not detected
- Out of biological matrix



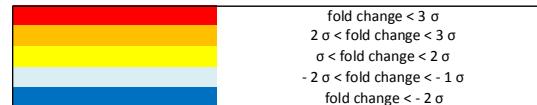
Aim: Create a **Recommendation System** to complete signatures

- Based on metabolic **network** architecture
- Using knowledge available in **literature**



Context HE: signature obtained from LC/HRMS analysis

MetaboliteID (in-house database)	Biochemical pathway or chemical class	HE	p-values
Acetyl-glucosamine	acetylated compounds	2,26	4,30E-06
Acetyl-alanine		2,24	1,92E-05
Glyceric acid	Alcohols and polyols	1,43	3,11E-03
4-Acetamidobutanoic acid	Amino Fatty Acids	2,82	2,03E-05
Asparagine		1,68	1,05E-03
Citrulline	Amino Acids and Derivatives	1,88	3,32E-03
Threonine		1,54	2,22E-02
Cystine		2,38	1,14E-04
Glycocholic acid	Bile Acids	241,78	5,68E-06
Taurocholic acid		3e+5	NA
Acetyl-L-carnitine	Fatty acid metabolism	2,27	1,57E-03
Octanoylcarnitine		4,10	2,29E-04
Propionylcarnitine		2,48	2,49E-03
3-Hydroxyisovalerylcarnitine		5,49	1,24E-02
Glutamic acid	Glutamate/Glutamine metabolism	2,30	1,98E-04
Glutamine		2,39	3,11E-05
Pyroglutamic acid		2,29	3,38E-06
Methionine	Methionine metabolism	7,08	1,70E-05
Dihydrothymine	Nucleosides and derivates	3,45	6,52E-06
Leu-Ala	Peptides	5,45	3,38E-06
phenylalanine	Phenylalanine metabolism	3,04	4,40E-06
p-hydroxyphenyllactic acid		8,42	1,29E-05
Tyrosine		5,12	3,61E-05
Cortisol	Steroid	3,55	1,59E-05
5-hydroxyindoleacetic acid	Tryptophan metabolism	2,60	2,08E-03
Hydroxytryptophan		4,02	1,50E-06
Indolelactic acid		6,00	5,68E-06
kynurenone		5,91	2,59E-06
Tryptophan		3,57	2,03E-05
Pyridoxic acid	Vitamin B6 metabolism / Pyridines catabolism	15,93	1,32E-04

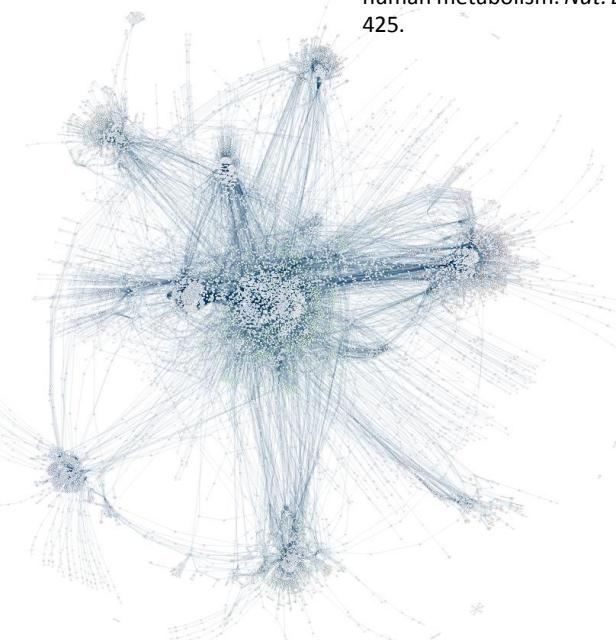


→ 30 relevant significant compound
 p -value < 0.005 (Mann-Whitney test) fold change > 2 σ
Found in human metabolic network

(Weiss N. et al., J. Hepathol, 2016)

7440 reactions
 2626 metabolites
 1733 genes

Thiele,I., Swainston,N., Fleming,R.M.T., Hoppe,A., Sahoo,S., Aurich,M.K., Haraldsdottir,H., Mo,M.L., Rolfsson,O., Stobbe,M.D., et al. (2013) A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.*, **31**, 419–425.

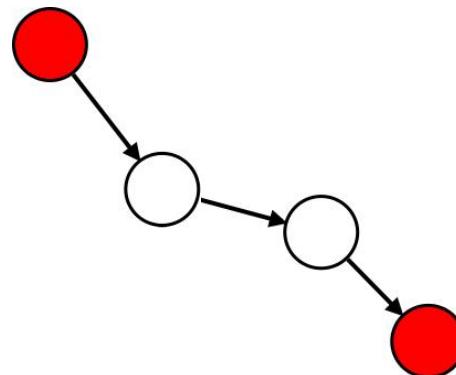


Genome scale metabolic network



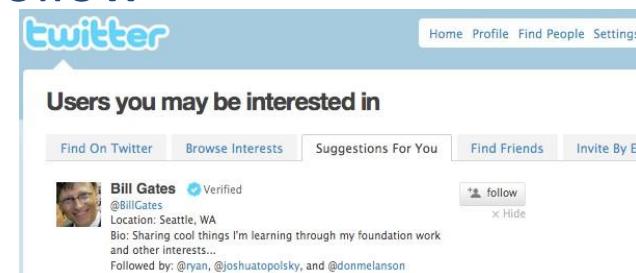
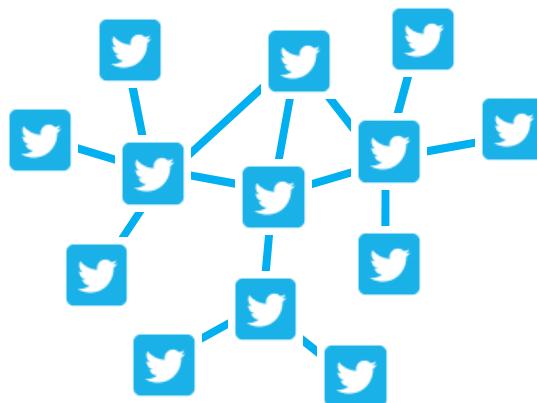
Network based suggestion system principle

- “ Hypothesis :
 - perturbations are propagated over the network
- “ Algorithm :
 - Estimation of the likelihood of a compound to be reached from signature's compounds



Approach used in recommendation system

” Twitter’s « Who To Follow »



“We are leaving the age of information and entering the age of recommendation”

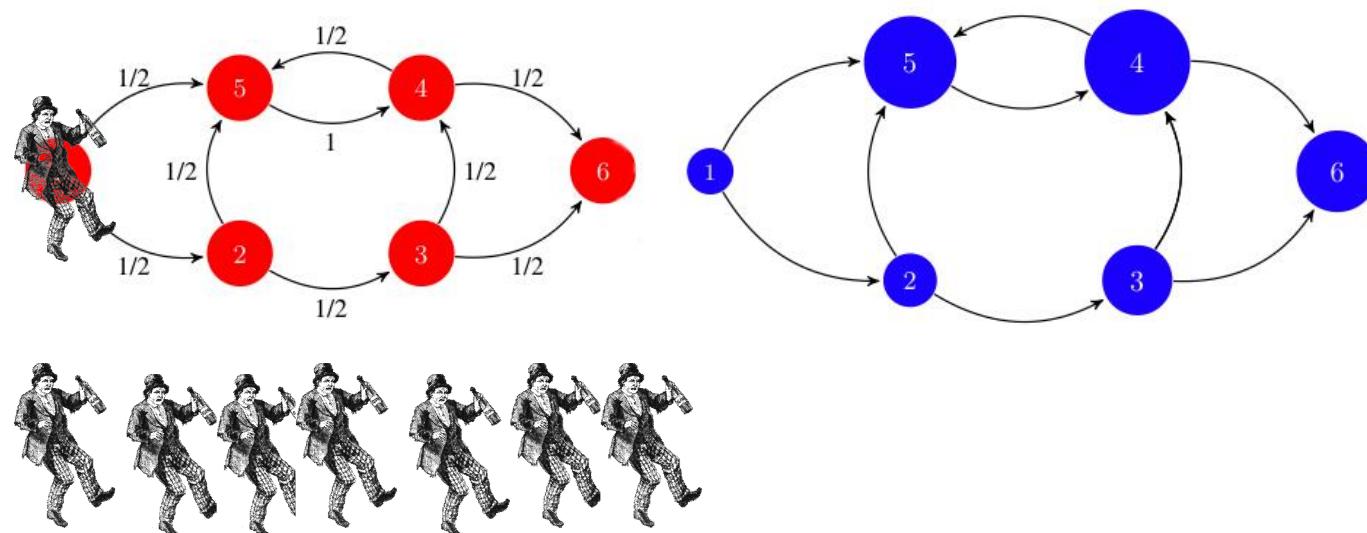
Chris Anderson - The Long Tail

Gupta P, Goel A, Lin J, et al. WTF: The Who to Follow Service at Twitter. 2013; 505–514

PageRank algorithm, predicting important nodes

“ Random walks principia:

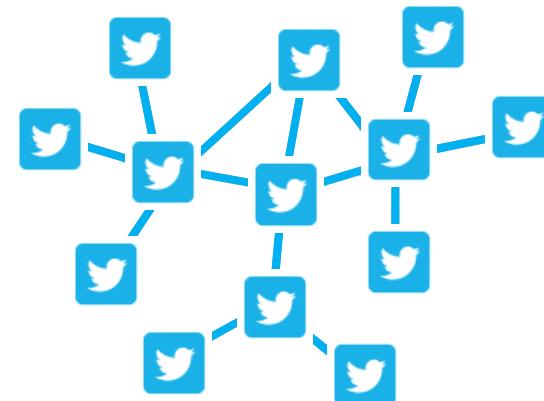
- . The probability to encounter a node during a random walk define its importance in the network



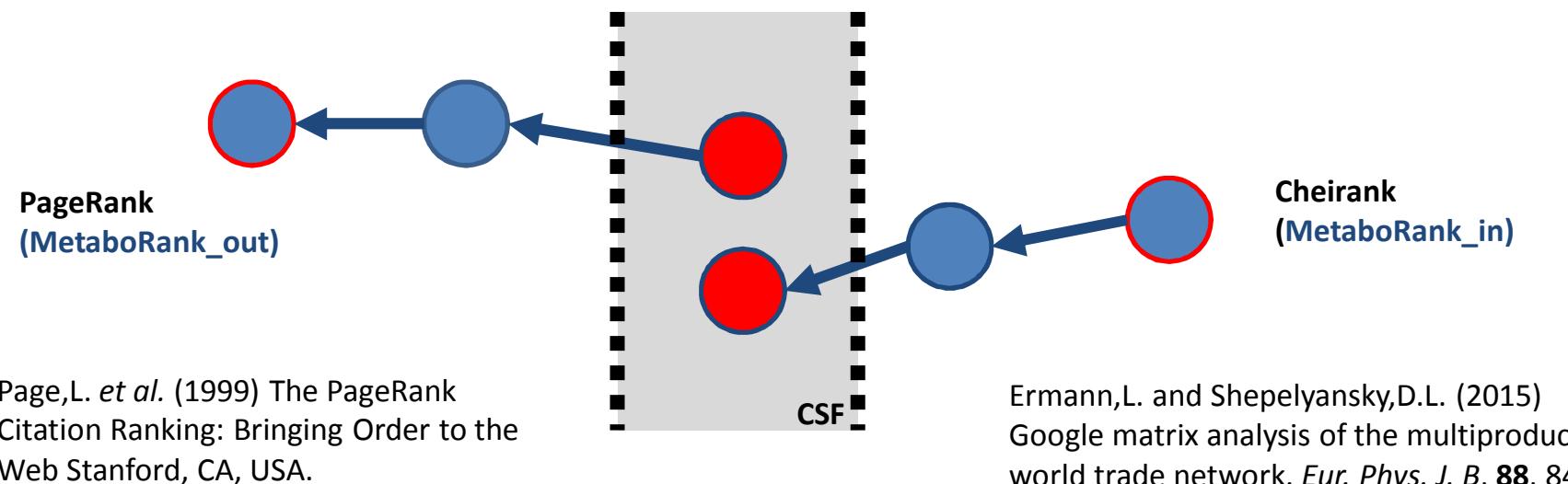
Personalized PageRank

- “ PageRank = « global » importance
- “ **Personalized PageRank** = defined **relatively** to a set of nodes of interest

Personalized PageRank using metabolite from profile as starting points

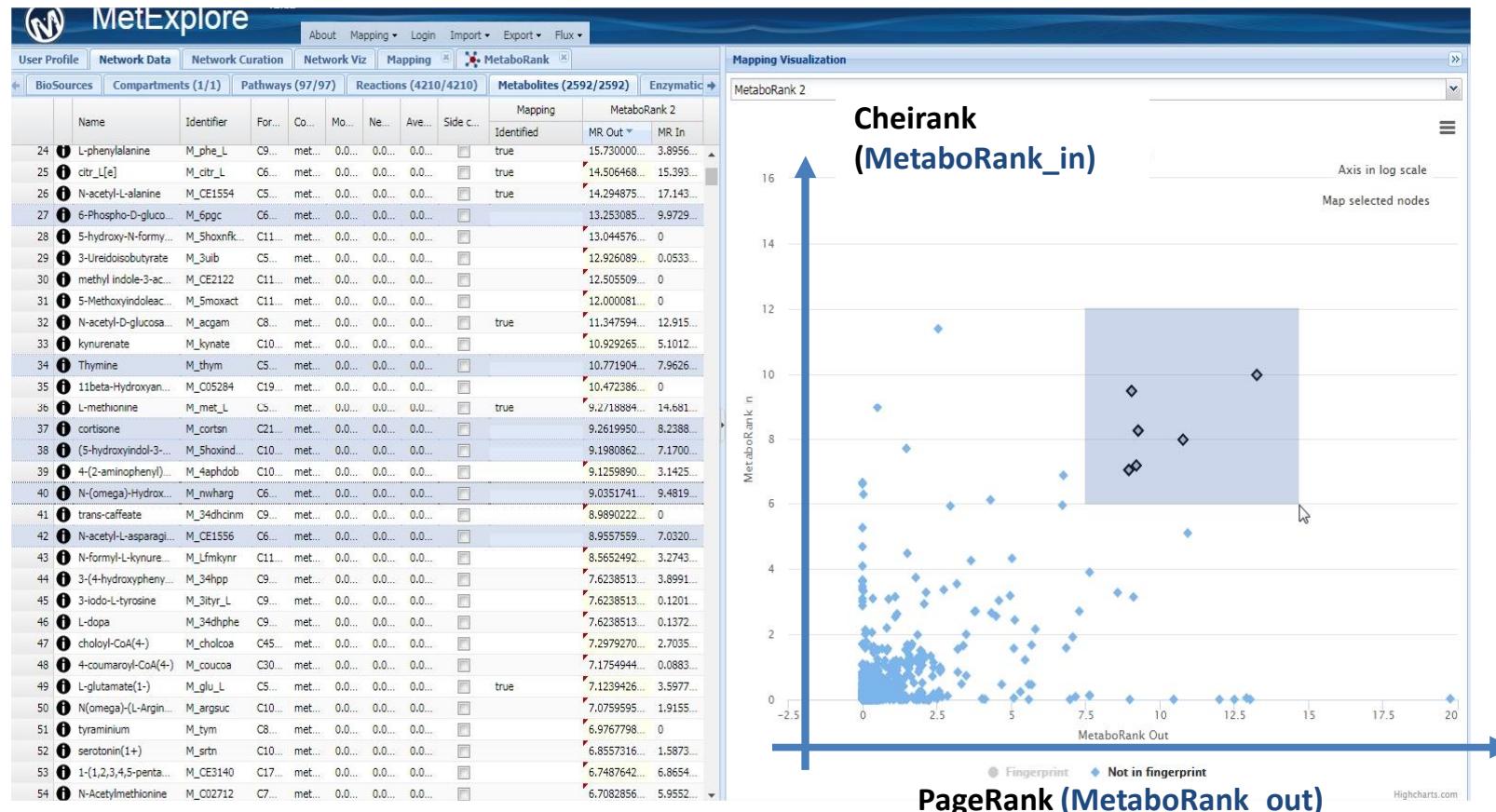


Looking downstream and upstream metabolites



MetaboRank : combination of **MetaboRank_in** and **MetaboRank_out**

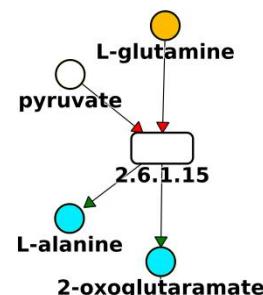
MetaboRank implementation: MetExplore (www.metexplore.fr)



ω -Amidase: an underappreciated, but important enzyme in L-glutamine and L-asparagine metabolism; relevance to sulfur and nitrogen metabolism, tumor biology and hyperammonemic diseases

Arthur J. L. Cooper¹ · Yevgeniya I. Shurubor^{2,3} · Thambi Dorai⁴ · John T. Pinto¹ · Elena P. Isakova³ · Yulia I. Deryabina³ · Travis T. Denton⁵ · Boris F. Krasnikov^{1,3}

Received: 31 March 2015 / Accepted: 24 July 2015
© Springer-Verlag Wien 2015



Published in final edited form as:
Metab Brain Dis. 2014 December ; 29(4): 991–1006. doi:10.1007/s11011-013-9444-9.

α -Ketoglutaramate: An overlooked metabolite of glutamine and a biomarker for hepatic encephalopathy and inborn errors of the urea cycle

Arthur J. L. Cooper¹ and Tomiko Kuhara^{2,*}

Highlighting metabolites which may be hidden in the litterature



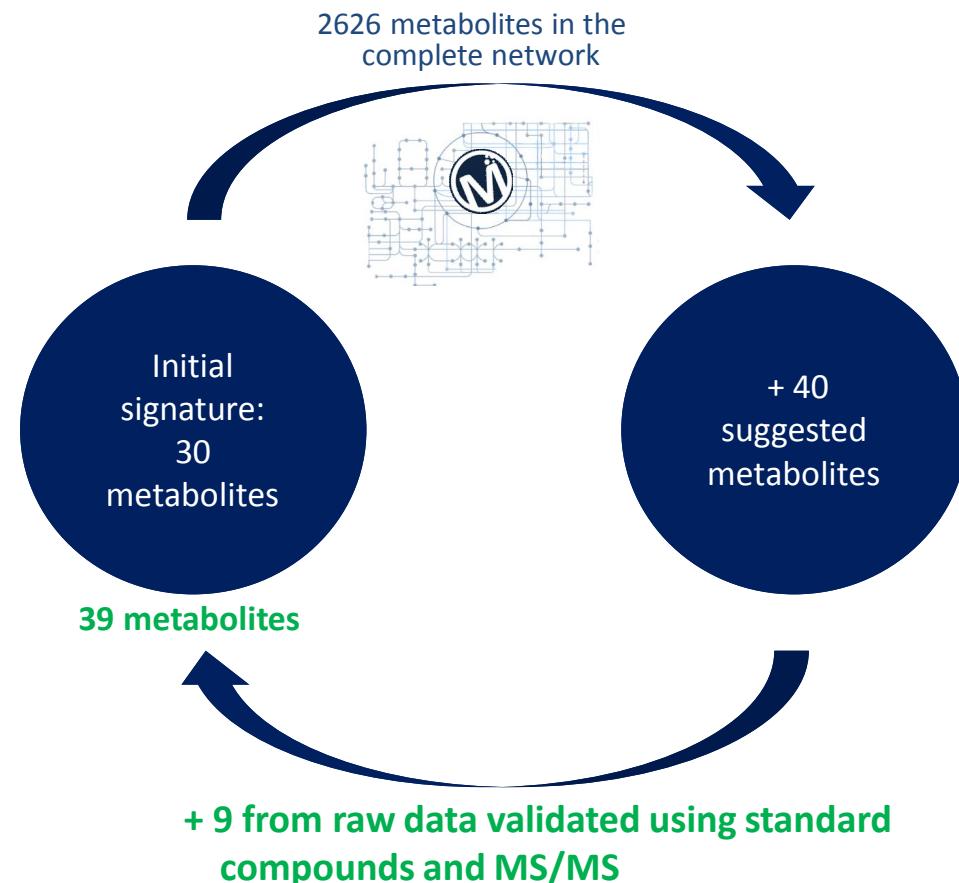
« Bonjour

Je viens seulement de lire les articles (très intéressants) sur l'**alphacetoglutaramate** que vous m'avez adressés. Malheureusement je ne peux guère vous aider car(à ma honte) j'ignorais les travaux de Cooper sur cet intéressant composé et son accumulation dans les hyperammoniemies avec hyperglutaminémie.Au passage j'ai noté (avec intérêt et satisfaction) que ce composé ne s'accumule pas dans les hyperammoniémes des aciduries organiques dans lesquelles la glutamine est normale et ou j'ai toujours observé une bien meilleure tolérance clinique à taux comparables d'ammoniéme.D'une façon générale il est vrai que la voie réversible de la transamination des AA en alphacetoacides a été peu explorée récemment et semble être "passée de mode ». Dans les années 60/70 je me souviens avoir utilisé des mélanges d'acides alphacétioniques d'AA essentiels dans le traitement des hyperammoniémes avec élévation de la glutamine mais c'était cher et difficile à obtenir.La seule application thérapeutique a été le phenylbutyrate dans les déficits du cycle de l'urée , qui se métabolise en phenylacétate puis en phenylacetylglutamine. En tout cas ce composé devrait être intégré dans la liste des marqueurs métaboliques intéressants.

Bien à vous

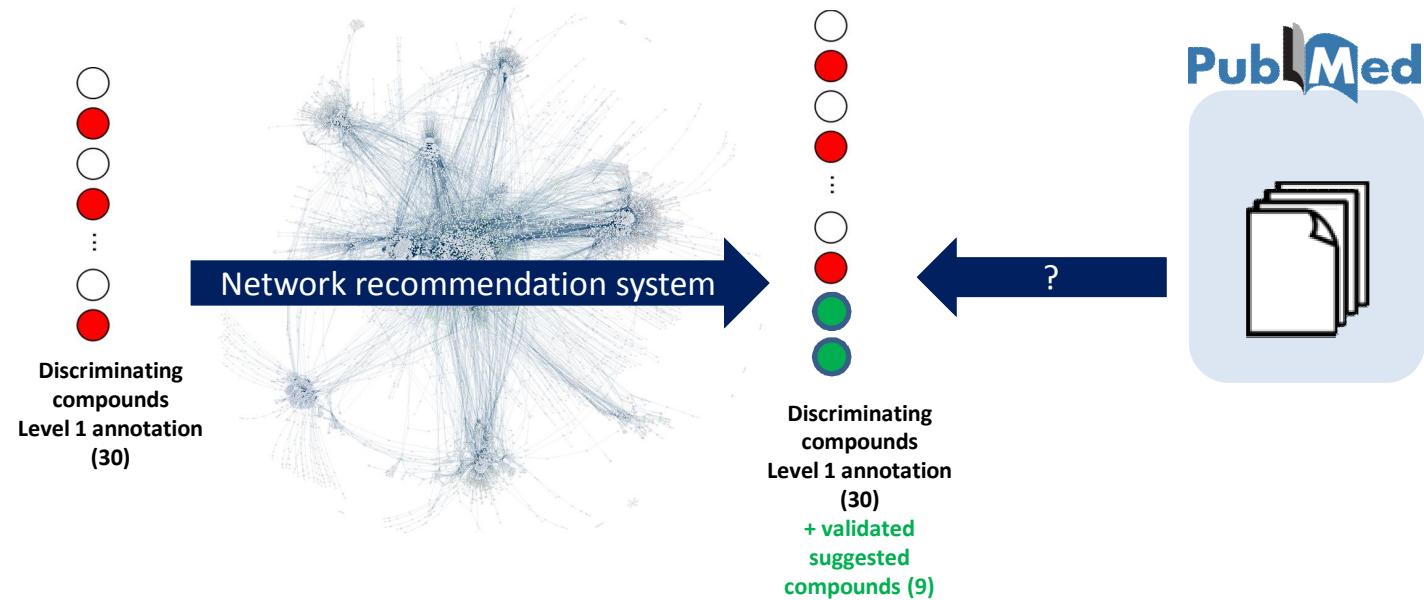
Jean-Marie Saudubray »

Recommendation system: application to EH

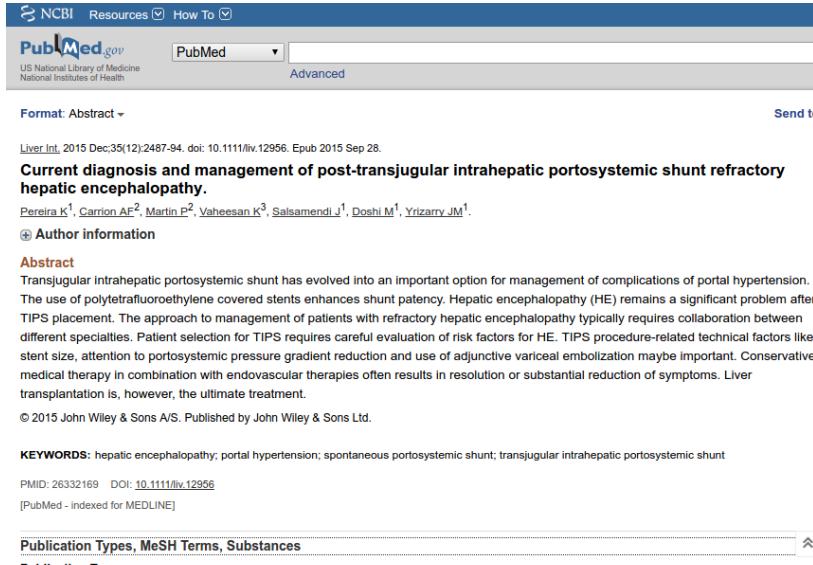


PRACTICE 3

Ordering suggested metabolites based on litterature



Exploiting knowledge in literature



The screenshot shows a PubMed search result for the article "Current diagnosis and management of post-transjugular intrahepatic portosystemic shunt refractory hepatic encephalopathy". The abstract discusses the evolution of TIPS for portal hypertension and the challenges of managing refractory HE. It highlights the use of polytetrafluoroethylene covered stents to enhance shunt patency. The paper is from Liver Int., 2015 Dec;35(12):2487-94, doi: 10.1111/liv.12956, Epub 2015 Sep 28.

Article + Meta Data



MeSH



Edema

Publication Types, MeSH Terms, Substances

Publication Types

Review

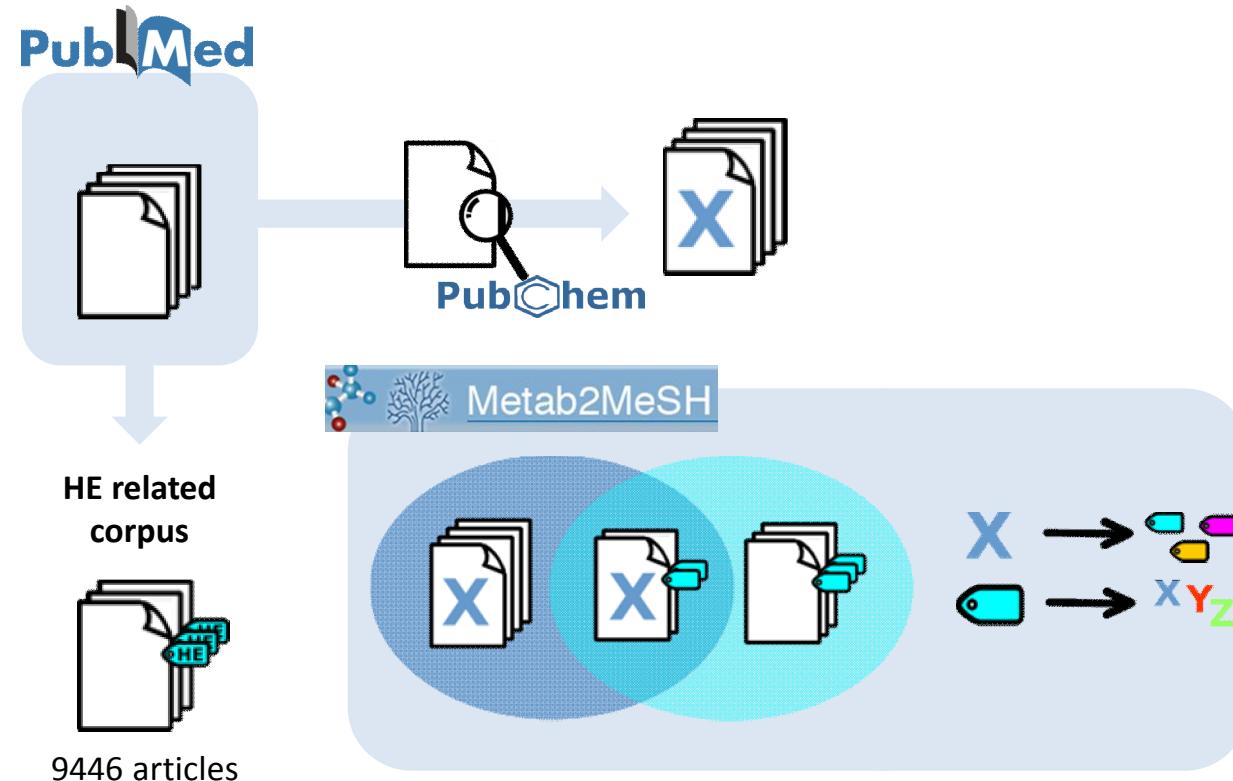
MeSH Terms

- Disease Management
- Hepatic Encephalopathy/diagnosis
- Hepatic Encephalopathy/etiology
- Hepatic Encephalopathy/prevention & control
- Humans
- Hypertension, Portal/surgery*
- Polytetrafluoroethylene/pharmacology
- Portasystemic Shunt, Transjugular Intrahepatic/adverse effects*
- Portasystemic Shunt, Transjugular Intrahepatic/instrumentation
- Portasystemic Shunt, Transjugular Intrahepatic/methods
- Risk Adjustment

Substances

- Polytetrafluoroethylene

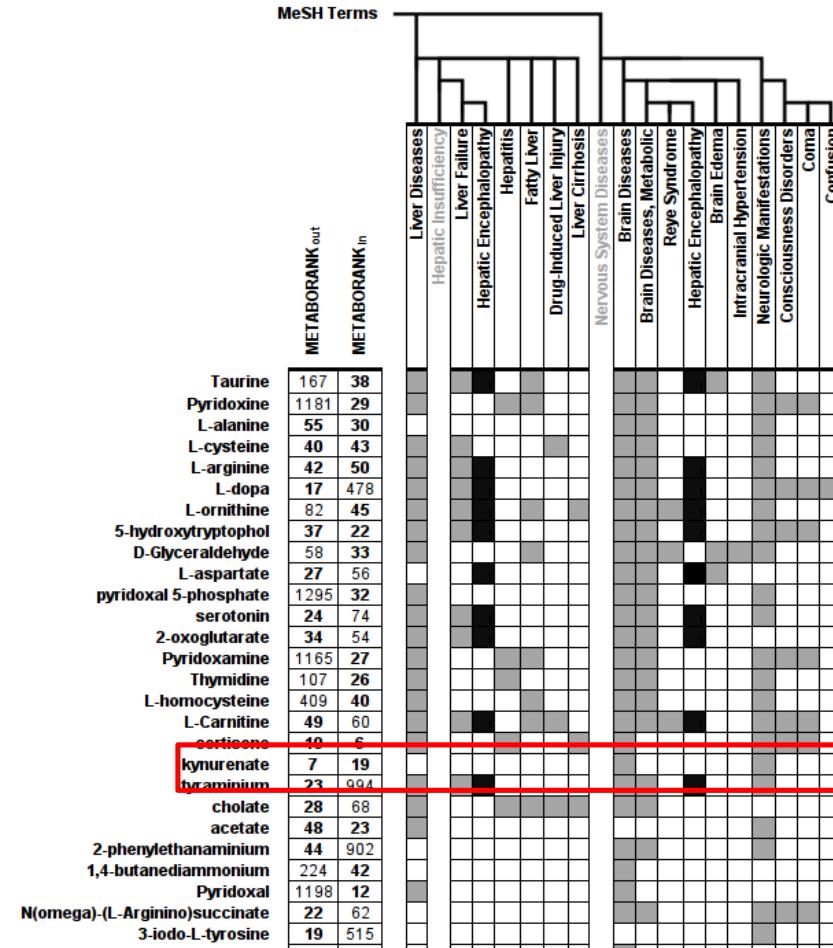
Automatic literature analysis



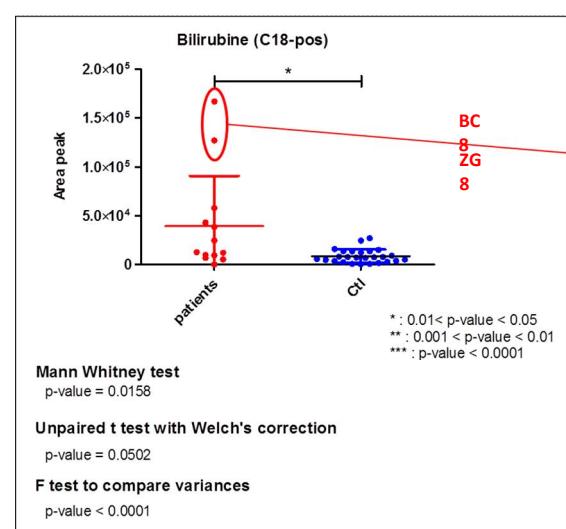
Sartor MA, Ade A, Wright Z, et al. Metab2MeSH: annotating compounds with medical subject headings. Bioinformatics 2012; 28:1408–10

Litterature based suggestion results

	CFP (Core Fingerprint)	CFP + Top MetaboRank _{in}	CFP + Top MetaboRank _{out}
	MetaboRank _{in}	MetaboRank _{out}	
5-Hydroxyindoleacetate	✓ ✓ ✓	✓ ✓ ✓	
L-Citrulline	✓ ✓ ✓	✓ ✓ ✓	
L-glutamate	✓ ✓ ✓	✓ ✓ ✓	
L-glutamine	✓ ✓ ✓	✓ ✓ ✓	
L-methionine	✓ ✓ ✓	✓ ✓ ✓	
L-phenylalanine	✓ ✓ ✓	✓ ✓ ✓	
L-tryptophan	✓ ✓ ✓	✓ ✓ ✓	
L-tyrosine	✓ ✓ ✓	✓ ✓ ✓	
O-acetyl carnitine	✓ ✓ ✓	✓ ✓ ✓	
octanoyl carnitine	✓ ✓ ✓	✓ ✓ ✓	
L-ornithine		✓	
Taurine		✓	
5-hydroxytryptophol		✓ ✓	
L-arginine		✓ ✓	
2-oxoglutarate		✓	
L-aspartate		✓	
L-Carnitine		✓	
L-dopa		✓	
serotonin		✓	
tyramine		✓	
bilirubin		✓	
choline		✓	
creatinine		✓	
D-aspartate		✓	
Diazepam		✓	
lipote		✓	
lithocholate		✓	
N-acetyl-L-aspartate		✓	
N-acetyl-L-cysteine		✓	
pantetheine		✓	
phenylacetate		✓	
quinolinate		✓	
Thiamine diphosphate		✓	
Urea		✓	
D-ornithine		✓	
D-arginine		✓	
benzoate		✓	
ammonium		✓	
	NC NC	NC NC	



Automatic literature analysis: e.g. of bilirubin

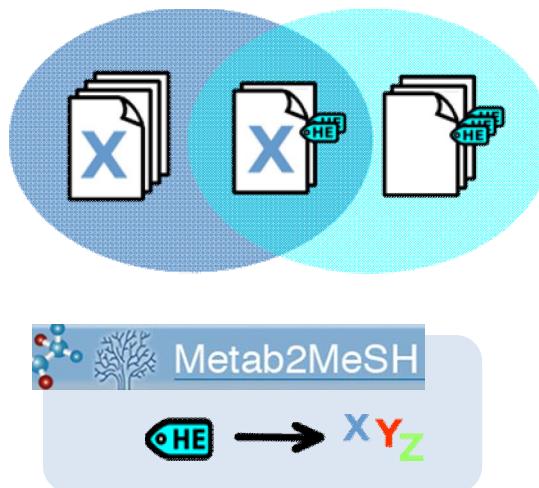


ID	BC8	ZG8
age	64	51
ddn	09.03.48	05.03.62
West-Haven	2	1
GCS	13	15
NH3	49	34
TP	28	37
bili tot	325	306
ALT	44	63
AST	63	100
creat	49	108
Ascite (pts)	3	3
Child-P	14	15
MELD	29	27

Patients BC8 and ZG8 :
High bilirubin plasma rate
Among highest MELD score
in the HE cohort



Automatic literature analysis: HE-related compound extraction



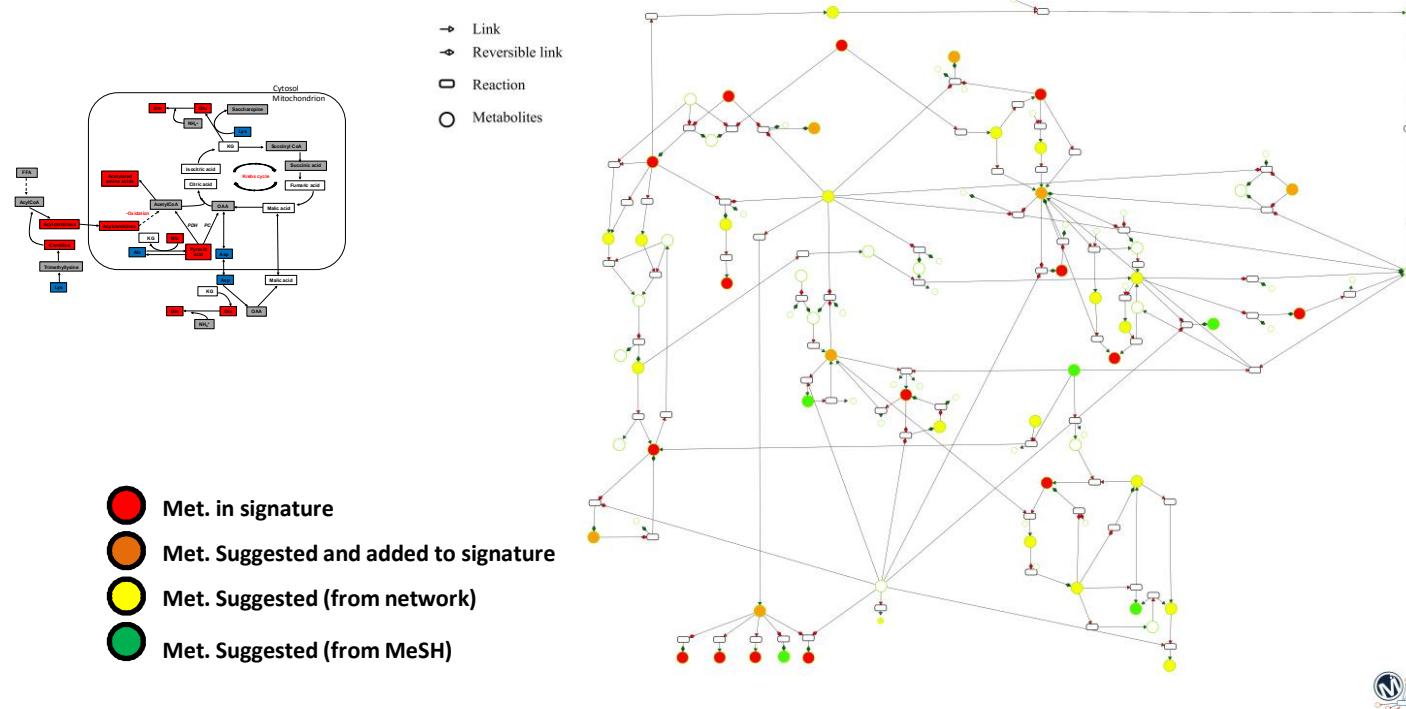
- ” 38 metabolites in the network associated with HE
- ” 10 found in the 30 metabolites of the signature
- ” 10 more found in the metabolites of the suggestion list
- ” 4 cannot be biochemically linked to the signature
- ” 3 added to the signature

HE-related compounds
5-Hydroxyindoleacetate
Citruline
L-glutamate
L-glutamine
L-methionine
L-phenylalanine
L-tryptophan
L-tyrosine
O-acetyl carnitine
octanoyl carnitine
Ornithine
Taurine
5-hydroxytryptophol
L-arginine
2-oxoglutarate
L-aspartate
L-Carnitine
L-dopa
serotonin
tyramine
bilirubin
choline
creatinine
D-aspartate
Diazepam
lipoate
lithocholate
N-acetyl-L-aspartate
N-acetyl-L-cysteine
pantetheine
phenylacetate
quinolinate
Thiamine diphosphate
Urea
D-ornithine
D-arginine
benzoate
ammoniac



Conclusion

- “ Network+Mesh allowed adding 12 metabolites to the 30 in the initial signature
- “ Emphasize links with similar pathologies
- “ New biological scenarios taking into account 15 more compounds from the extended signature thanks to suggested compounds

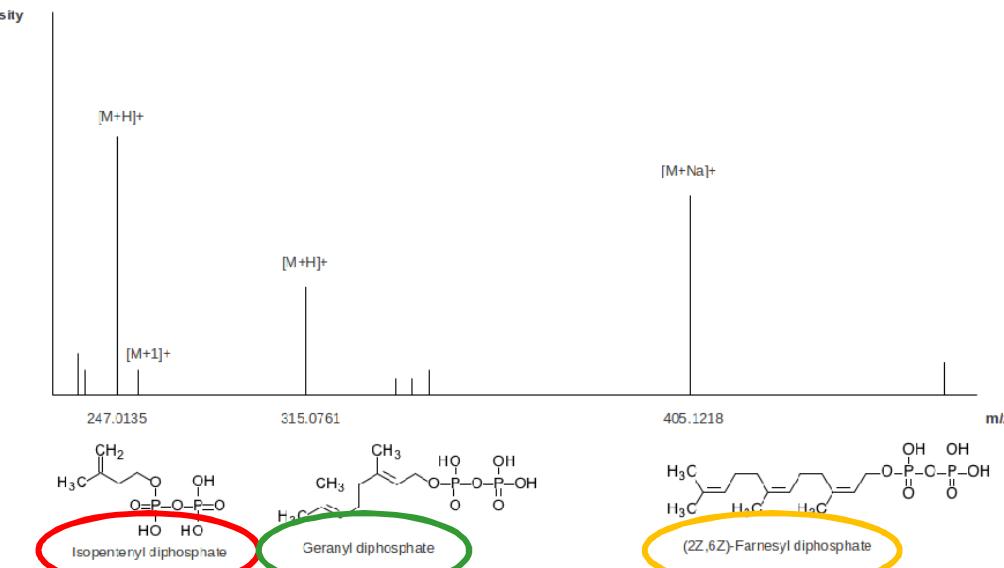


Helping in metabolite identification

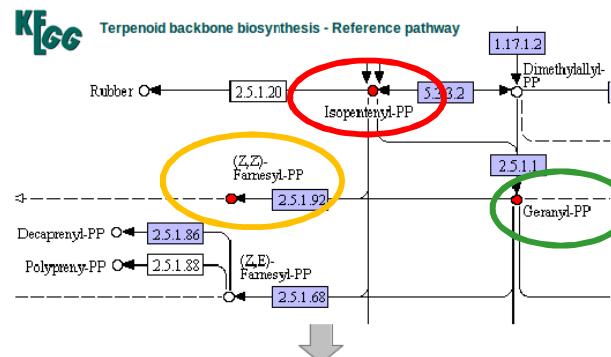
Taking biology into account to enhance metabolite identification:

- “ Use network structure
- “ Provide a probabilistic annotation by ranking identifications

Usual exact mass search



Introduction of prior Knowledge



Silva et al. *ProbMetab: an R package for Bayesian probabilistic annotation of LC-MS based metabolomics*. (2014). *Bioinformatics*.

Probabilistic annotation

Mass Peak	Candidate Compounds	Probability
247.0135	Isopentenyl diphosphate	1
315.0761	Geranyl diphosphate	0.7
	(+)-Bomyl-diphosphate	0.3
405.1218	(2Z,6Z)-Farnesyl diphosphate	1

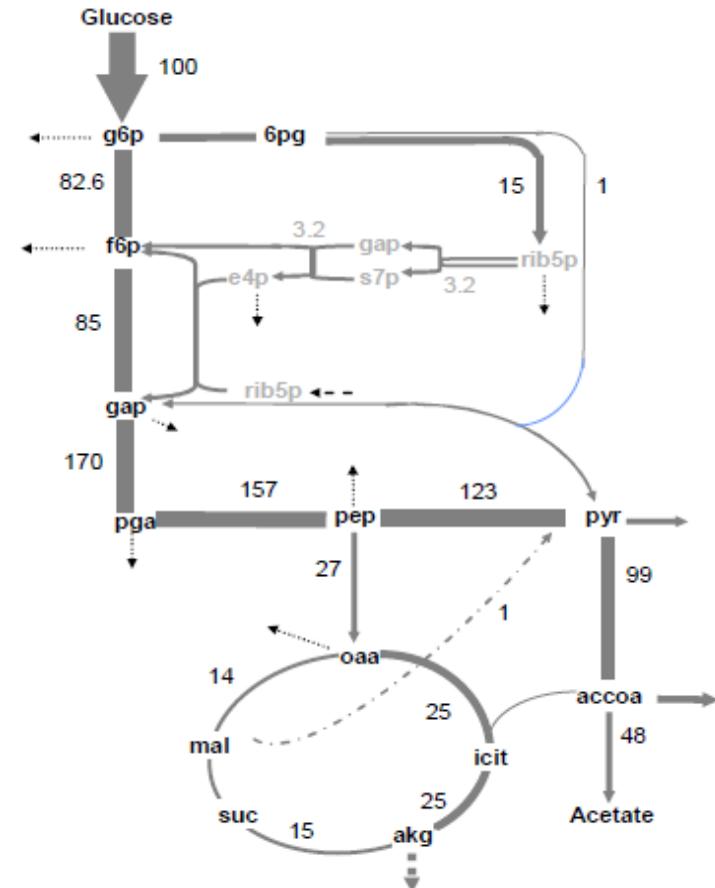
Network modelling to predict metabolic behaviors



Dynamic / semi-quantitative analysis

→ understand the behavior of the system under specific conditions

Predict metabolic fluxes, cell growth, drug targets...



Metabolic networks for metabolome mining

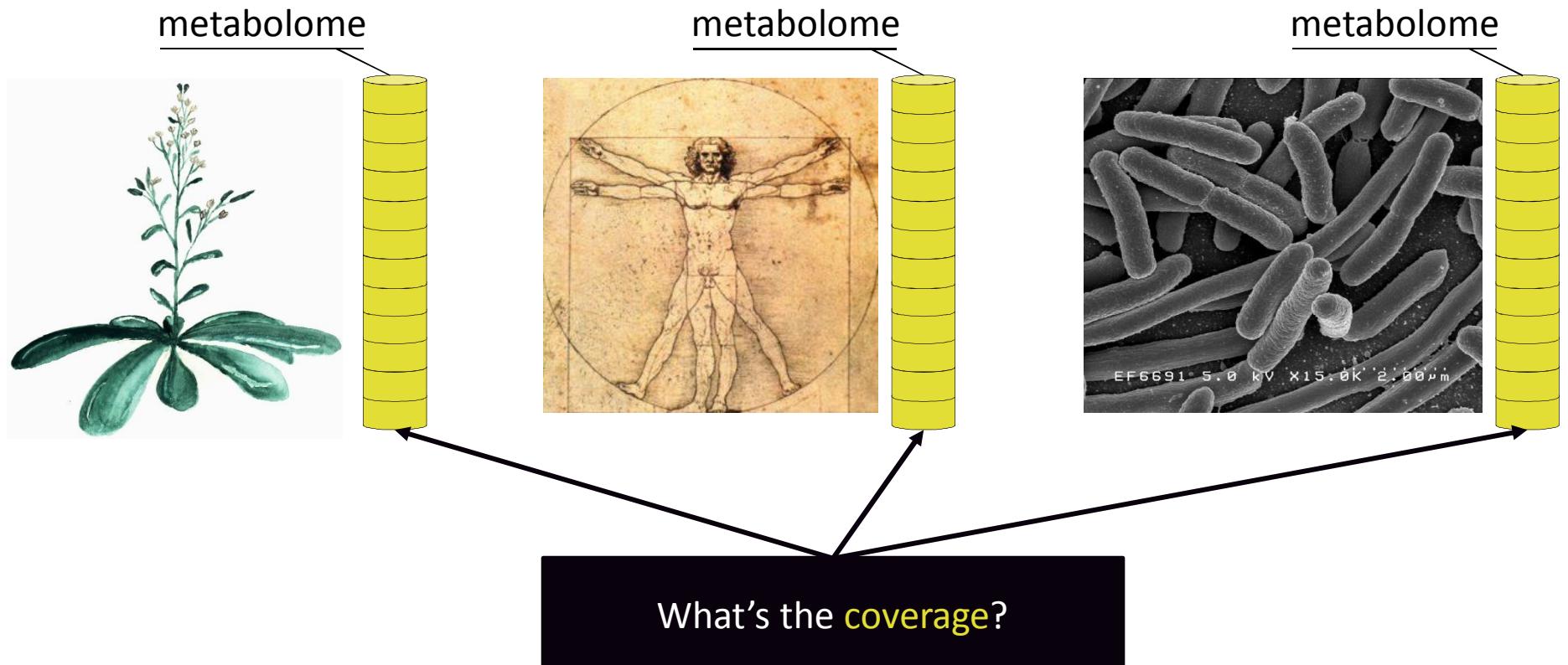
- “ How do spectral libraries cover known endogenous metabolism?
« Mind the gap »

Clément Frainay , Emma L. Schymanski , Steffen Neumann , Benjamin Merlet , Reza M Salek , Fabien Jourdan and Oscar Yanes. *Mind the gap: mapping mass spectral databases in genome-scale metabolic networks reveals poorly covered area.* (2018). **Metabolites.** 8(3), 51.

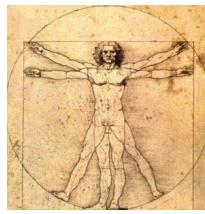
- “ How can we use network topology to help in Metabolic fingerprinting? **MetaboRank**

Clément Frainay, Sandrine Aros, Maxime Chazalviel, Thomas Garcia, Florence Vinson, Nicolas Weiss, Benoit Colsch, Frédéric Sedel, Dominique Thabut, Christophe Junot and Fabien Jourdan. *MetaboRank: network-based recommendation system to interpret and enrich metabolomics results.* (2018). **Bioinformatics.** Open access ahead of print.

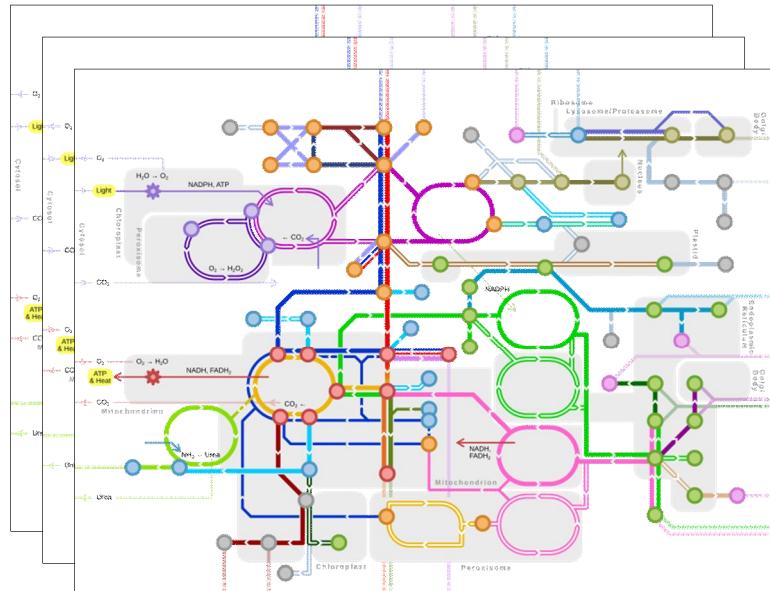
How much of endogenous metabolism can we monitor?



Use metabolic network as a context

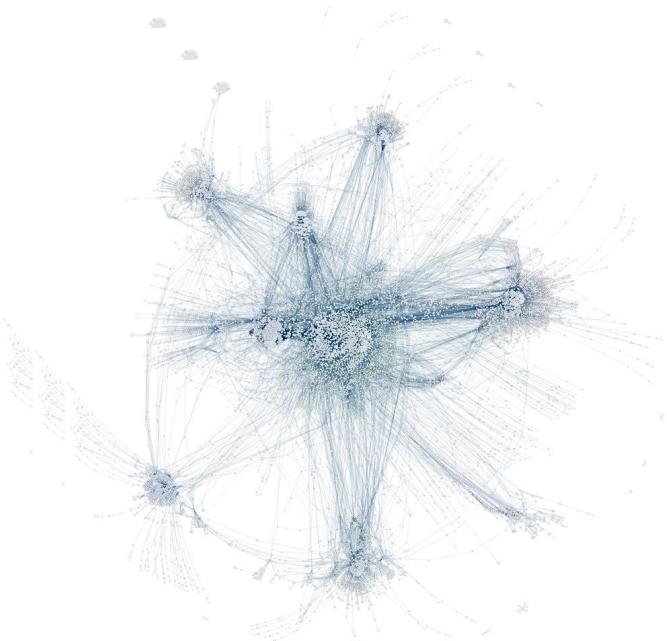


<https://en.wikipedia.org/wiki/Portal:Metabolism>



Pathways

<https://www.metexplore.fr>



Network

Metabolic network encodes endogenous metabolic capacities of an organism.



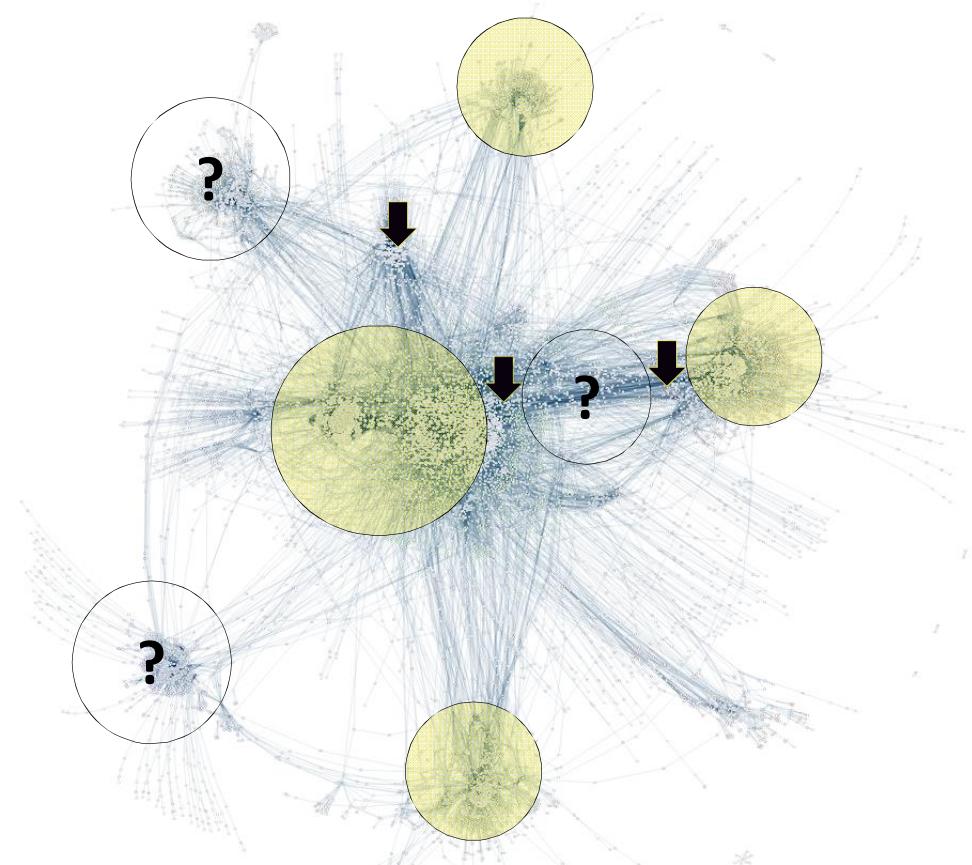
Network topology allows addressing several questions

How much of metabolic networks do MS/MS spectral libraries cover?

What are the blind spots of the network?

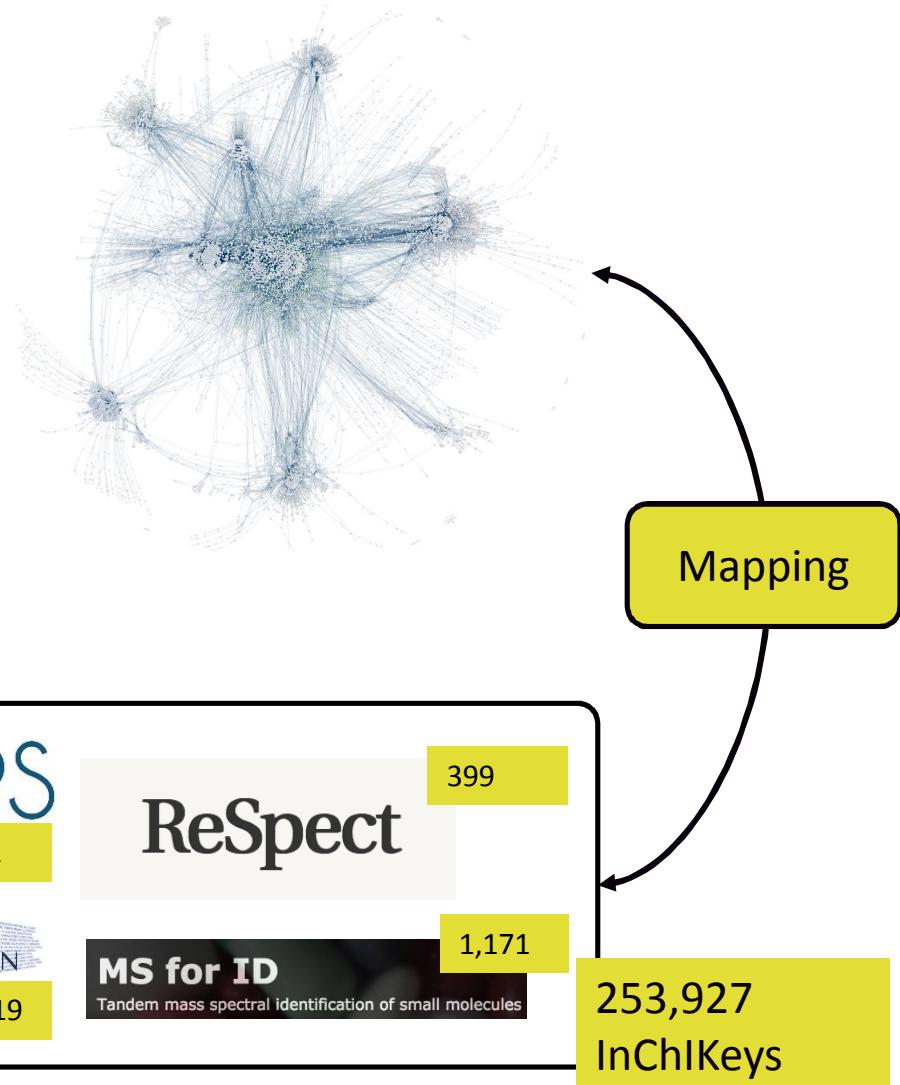
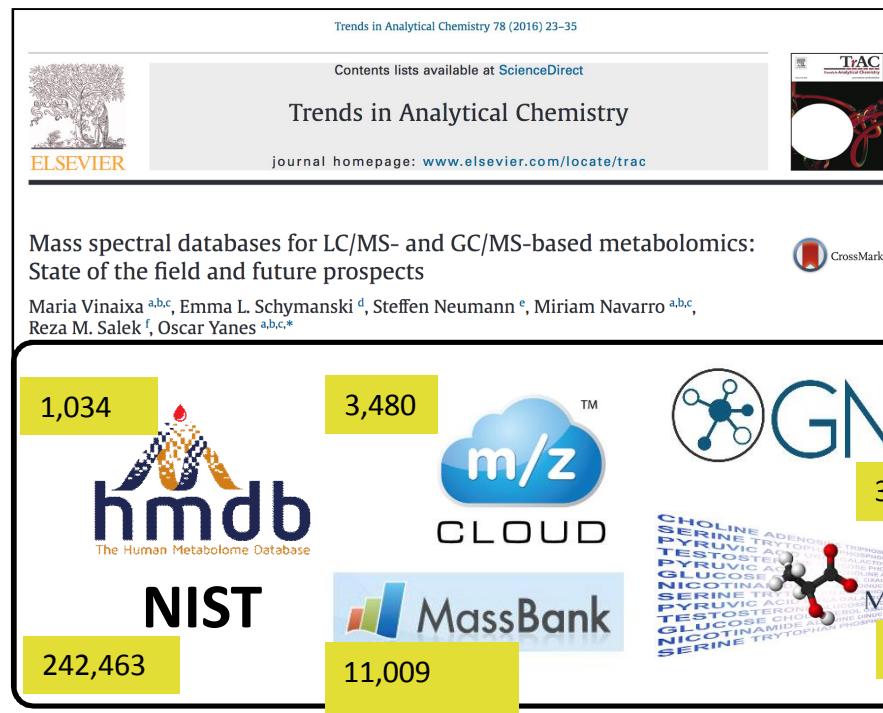
Can we suggest metabolites which could fill the gaps?

Networks may contain thousands of metabolites and reactions, it requires algorithms



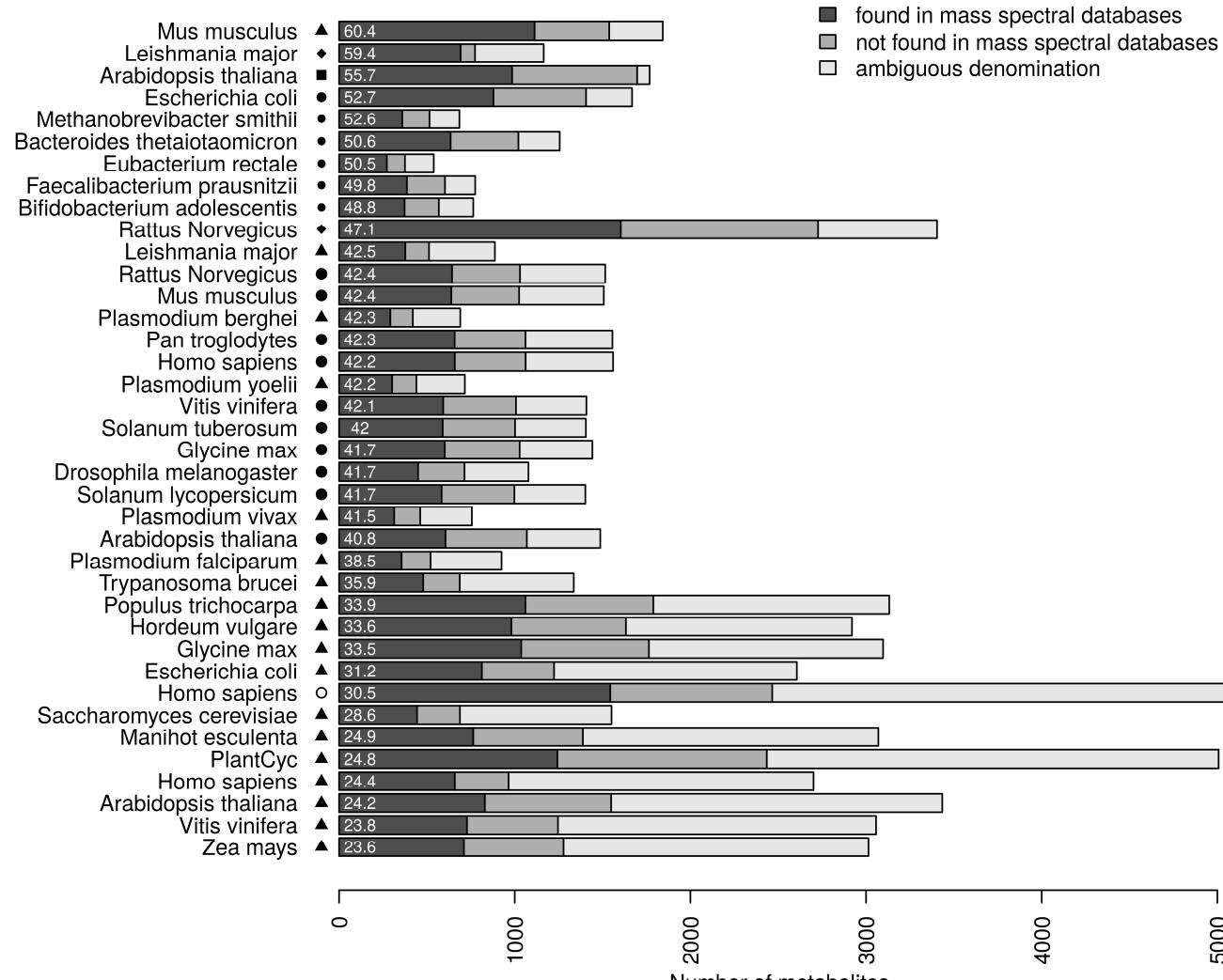
Mass spectral data – InChIKey based mapping

Spectral libraries (2016)



Mass spectral databases for LC/MS- and GC/MS-based metabolomics: State of the field and future prospects Vinaixa M, Schymanski E, Neumann S, Navarro M, Salek R et. al. TrAC Trends in Analytical Chemistry 2016 vol: 78 pp: 23-35

Coverage of organism metabolic networks



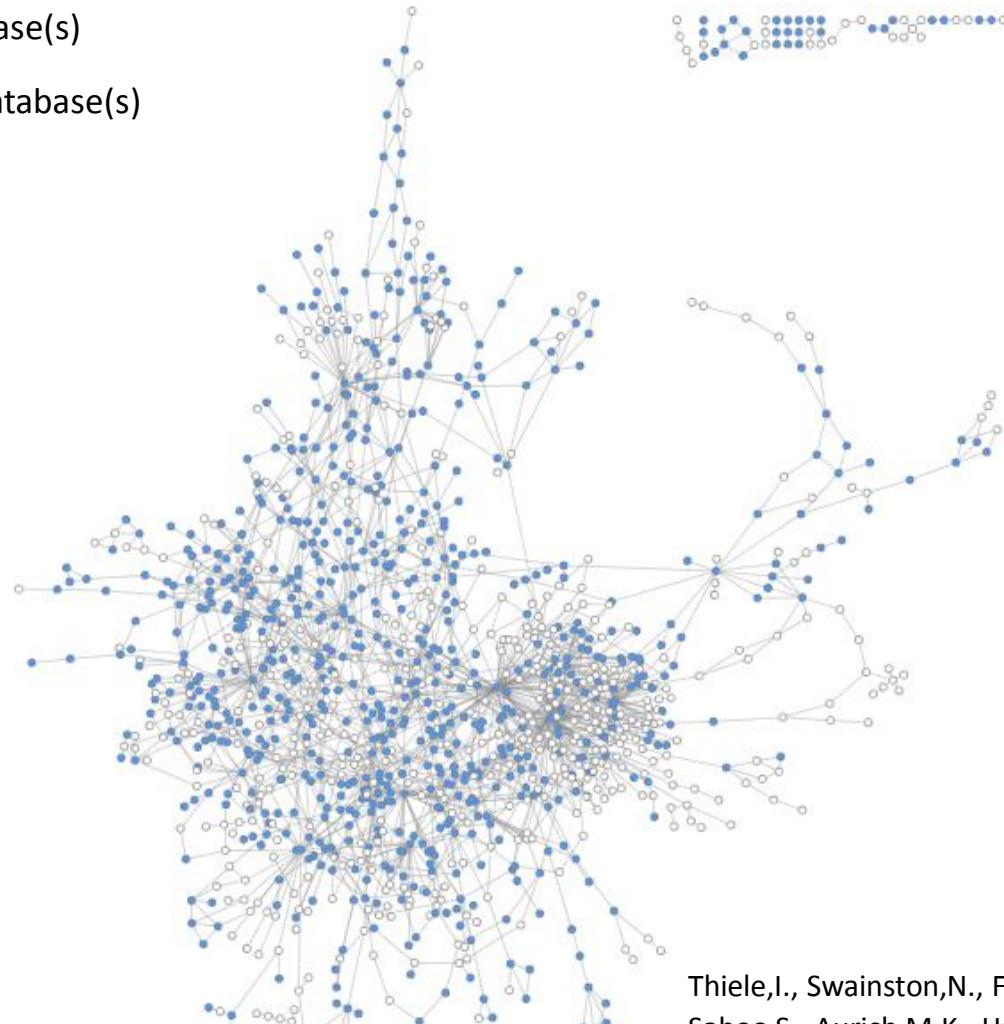
- „ Relatively homogenous
- „ Depends of the reconstruction

▲ BioCyc
◆ BioModels
● Kegg
● Metabolic Atlas
○ Recon2



Human metabolic network coverage

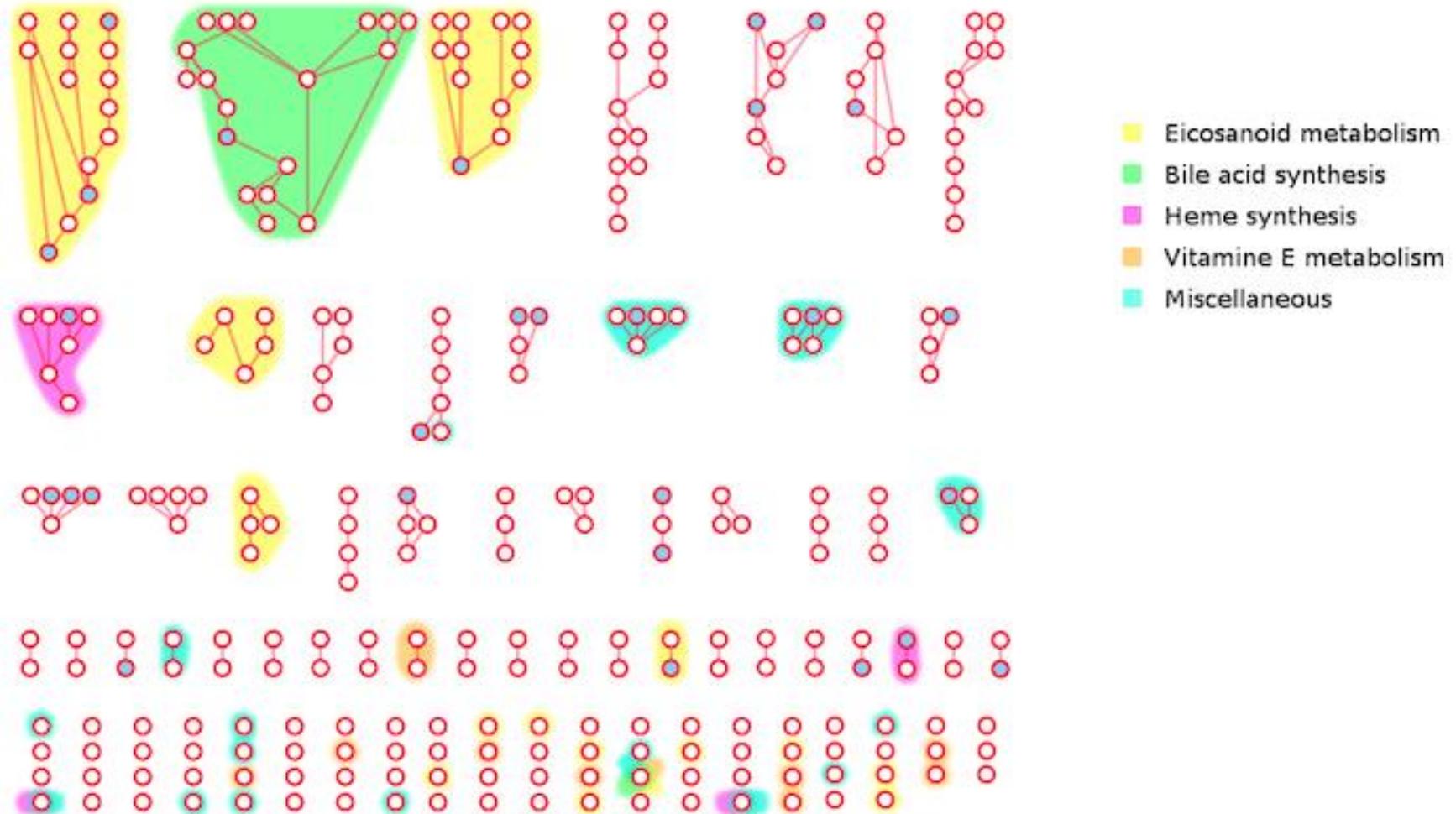
- Metabolite found in database(s)
- Metabolite not found in database(s)
- Reaction (carbon transfer)



Good spread in the Recon2 network core

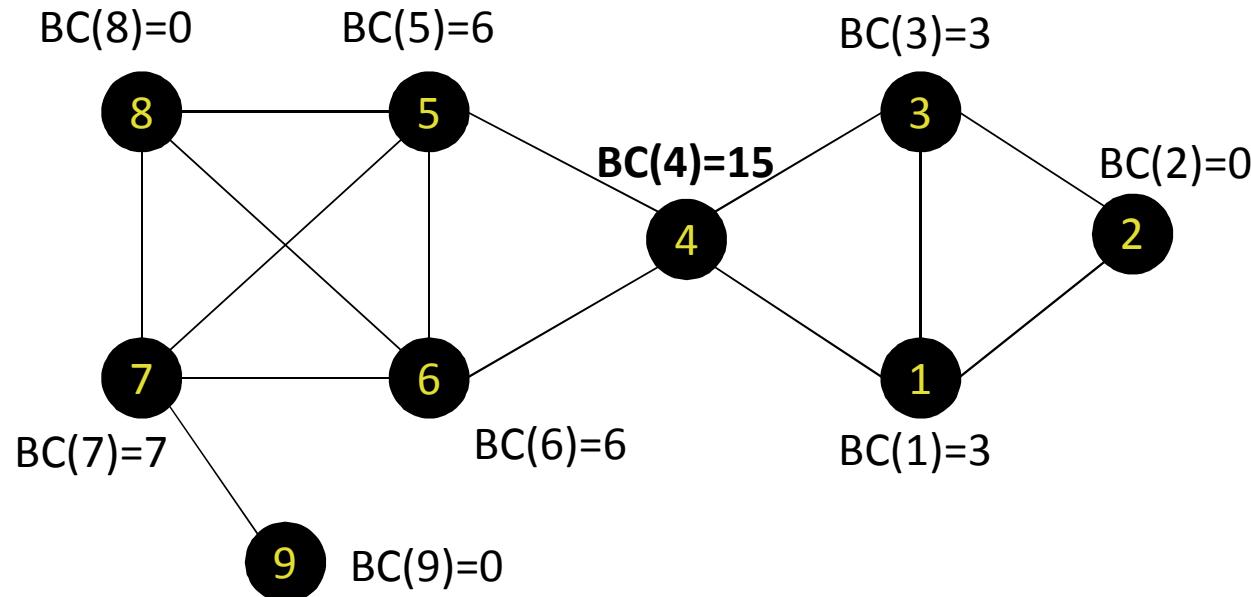
Thiele,I., Swainston,N., Fleming,R.M.T., Hoppe,A., Sahoo,S., Aurich,M.K., Haraldsdottir,H., Mo,M.L., Rolfsson,O., Stobbe,M.D., et al. (2013) A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.*, **31**, 419–425.

... but some parts are poorly covered



Sub-networks obtained using **Label Propagation Algorithm (LPA)** for community detection

Centrality based search for key metabolites to be added



Frainay,C. and Jourdan,F. (2017) Computational methods to identify metabolic sub-networks based on metabolomic profiles. *Brief. Bioinform.*, **18**, 43–56.

Betweenness centrality quantifies the number of times a node acts as a bridge along the shortest path between two other nodes in the network

Suggestion of metabolites to add to the libraries

Name (from network)	PubChem CID	InChIKey
(25R)-3alpha,7alpha,12alpha-trihydroxy-5beta-cholestan-26-oyl-CoA(4-)	15942889	MNYDLIUNNOCPHG-FJWDCHQMSA-N
12-oxo-c-LTB3	122164853	ZFHPYBQKHVEFHO-LECUDPRGSA-N
3alpha,7alpha,12alpha-Trihydroxy-5beta-cholestanoate	440460	CNWPIIOQKZNXBB-SQZFNYHNSA-N
3alpha,7alpha,12alpha-trihydroxy-5beta-cholestan-26-al	193321	XJZGNVBLVFOSKJ-XZULNKEGSA-N
12-oxo-leukotriene B4	5280876	SJVWVCVZWMJXOK-NOJHDUNKSA-N
20-CoA-20-oxo-leukotriene B4	53481505	WLWKYZHFLKRKEU-WCOJVGLOSA-J
5beta-cholestane-3alpha,7alpha,12alpha,26-tetrol	439479	USFJGINJGUFSY-XZULNKEGSA-N
(4R,5S)-4,5,6-trihydroxy-2,3-dioxohexanoate	440390	GJQWCDSAOUUMKSE-STHAYSLISA-N
20-carboxy-leukotriene-B4	5280877	SXWGPVJGNOLNHT-VFLUTPEKSA-N
5beta-cholestane-3alpha,7alpha,12alpha-triol	160520	RIVQQZVHIVNQFH-XJZYBRFWSA-N
3-oxo-tetracosa-12,15,18,21-all-cis-tetraenoyl-CoA	131769900	HPMVBGKWFWCZAY-JDTXFHFDSA-N
6-pyruvoyl-5,6,7,8-tetrahydropterin	128973	WBJZXBUVECZHCE-UHFFFAOYSA-N
Hydroxymethylbilane	788	WDFJYRZCZIUBPR-UHFFFAOYSA-N
5beta-cholestane-3alpha,7alpha,12alpha,25-tetrol	160520	RIVQQZVHIVNQFH-XJZYBRFWSA-N
3(S)-hydroxy-tetracosa-12,15,18,21-all-cis-tetraenoyl-CoA	53477712	NTIXPPFPXLJCT-OWOWEXKPSA-N
Uroporphyrinogen III	1179	HUHWZXWWOFSFKF-UHFFFAOYSA-N
12-oxo-20-hydroxy-leukotriene B4	53481459	CZWPUWRHQBAXJS-PABROBRYSA-N
3-oxo-all-cis-6,9,12,15,18-tetracosapentaenoyl-CoA	131769894	UQPANOFGYCZRAV-UWOIJHEUSA-N
all-cis-10,13,16,19-docosatetraenoyl-CoA	71627222	BEEQBBPNTYBGDP-BUSXXEPMSA-J
kinetensin	53481569	PANUJGMSOSQAAY-HAGIGRARSA-N



Conclusions

- ” Coverage of metabolic networks range from 20% to 60%
- ” On human network there is a relatively homogen covering
- ” Some parts are nevertheless not covered enough
- ” Network topology can help in finding key gaps
- ” Network « quality » is impacting coverage analysis
- ” Method is generic can be applied if there are updates of networks and databases
- ” Not taking into account unknown parts of metabolism ... so the story is not over!

Usage of these networks

- “ How do spectral libraries cover known endogenous metabolism?
« Mind the gap »

Clément Frainay , Emma L. Schymanski , Steffen Neumann , Benjamin Merlet , Reza M Salek , Fabien Jourdan and Oscar Yanes. *Mind the gap: mapping mass spectral databases in genome-scale metabolic networks reveals poorly covered area.* (2018). **Metabolites.** 8(3), 51.

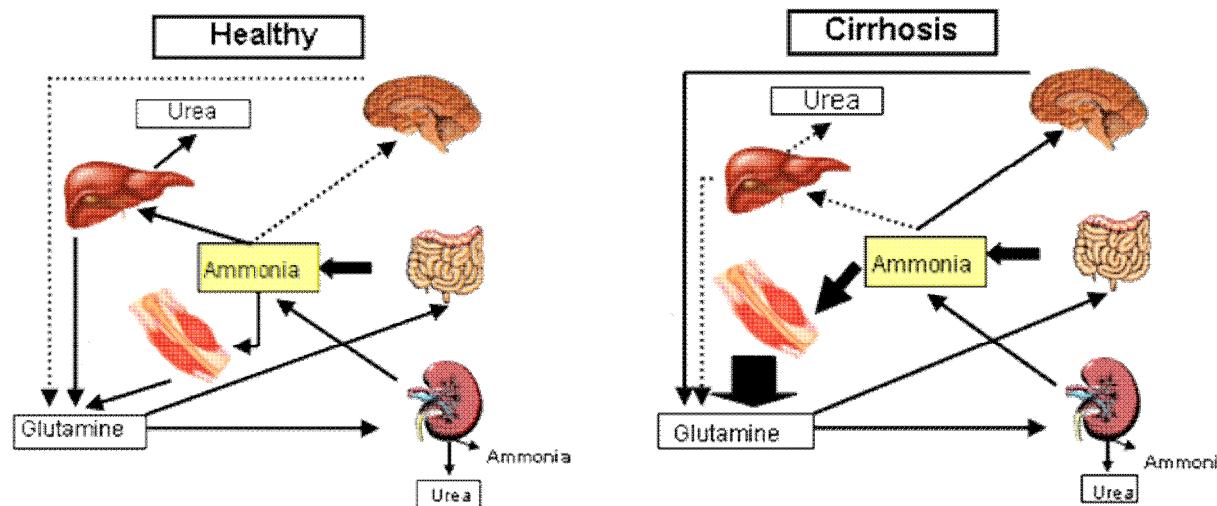
- “ How can we use network topology to help in Metabolic fingerprinting? **MetaboRank**

Clément Frainay, Sandrine Aros, Maxime Chazalviel, Thomas Garcia, Florence Vinson, Nicolas Weiss, Benoit Colsch, Frédéric Sedel, Dominique Thabut, Christophe Junot and Fabien Jourdan. *MetaboRank: network-based recommendation system to interpret and enrich metabolomics results.* (2018). **Bioinformatics.** Open access ahead of print.

Hepatic Encephalopathy (HE): complex syndrome

Physiopathology

- Hyperammonemia
- Glutamine accumulation in the brain



Weiss N, Barbier Saint Hilaire P, Colsch B, et al. Cerebrospinal fluid metabolomics highlights dysregulation of energy metabolism in overt hepatic encephalopathy. *J. Hepatol.* 2016; 65:1120–1130

Take home messages

- “ Agreement on metabolite identifiers can save (a lot of) time to perform mapping and analysis
- “ Using metabolic knowledge can help thinking out of the box
- “ Metabolic pathways are useful for functional analysis ...
- “ ...but may overshadow more complex processes
- “ Genome scale metabolic networks provide an holistic context ...
- “ ... but require modelling and algorithms
- “ Same networks can be used for flux modeling...
- “ ... but will require extra parameters
- “ Recommendation system can help in improving fingerprinting

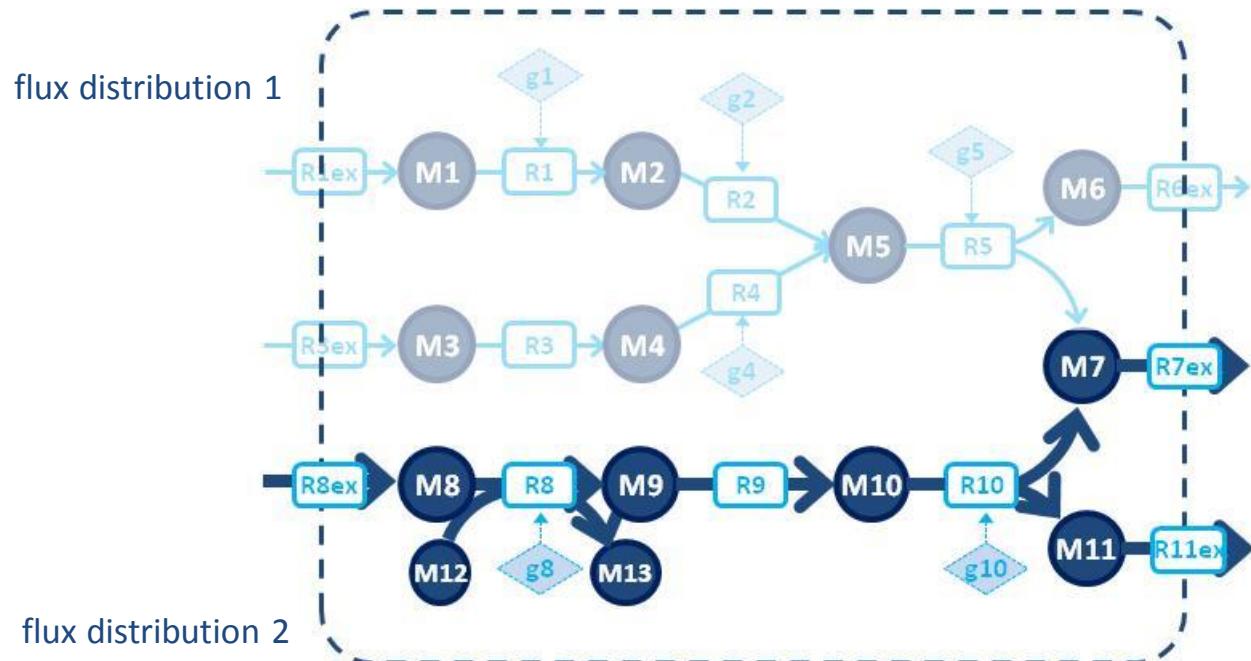


kahoot.it



GETTING CLOSER TO PHENOTYPE: TRANSCRIPTOMICS AND HUMAN CELLULAR METABOLISM

One network, several phenotypes



flux $\neq 0$ \Leftrightarrow « active » reactions

flux = 0 \Leftrightarrow « inactive » reactions

$\hookrightarrow \left\{ \begin{array}{l} \text{1 flux} \\ \text{distribution} \end{array} \right. \Leftrightarrow \text{1 sub-network of active reactions} \Leftrightarrow \text{1 phenotype}$

A global network model = an infinity of possible phenotypes

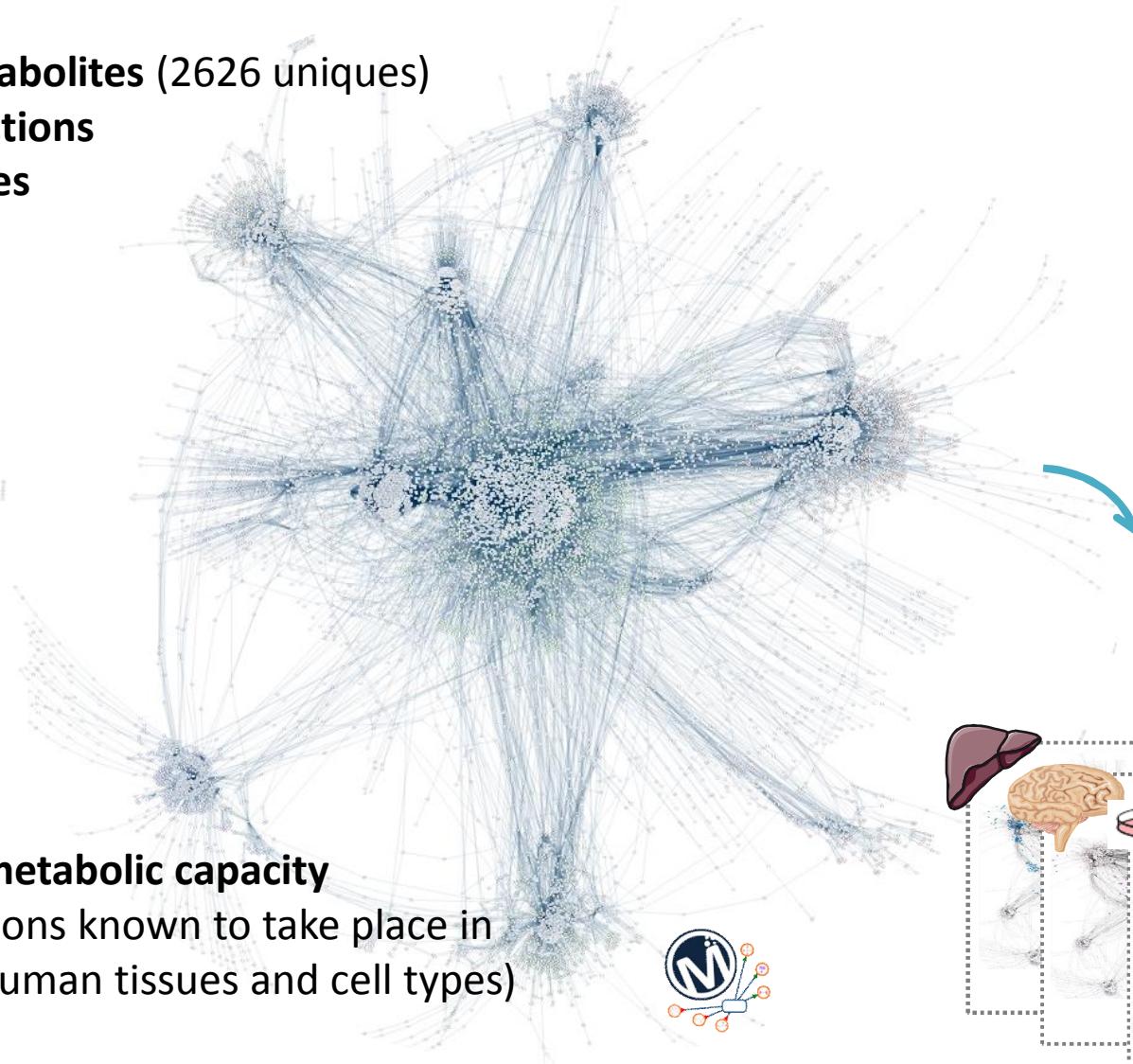


Extract active parts: create network for each phenotype

5063 metabolites (2626 uniques)

7440 reactions

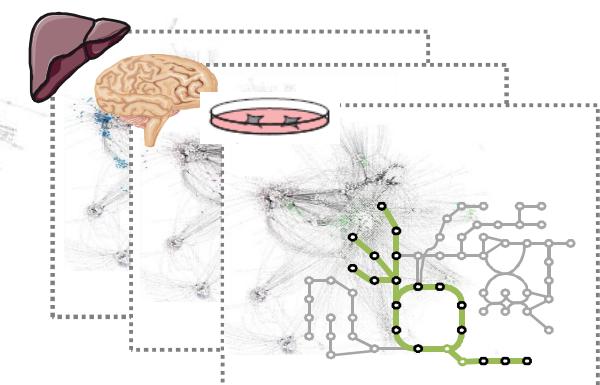
2194 genes



→ **generic metabolic capacity**

(all reactions known to take place in various human tissues and cell types)

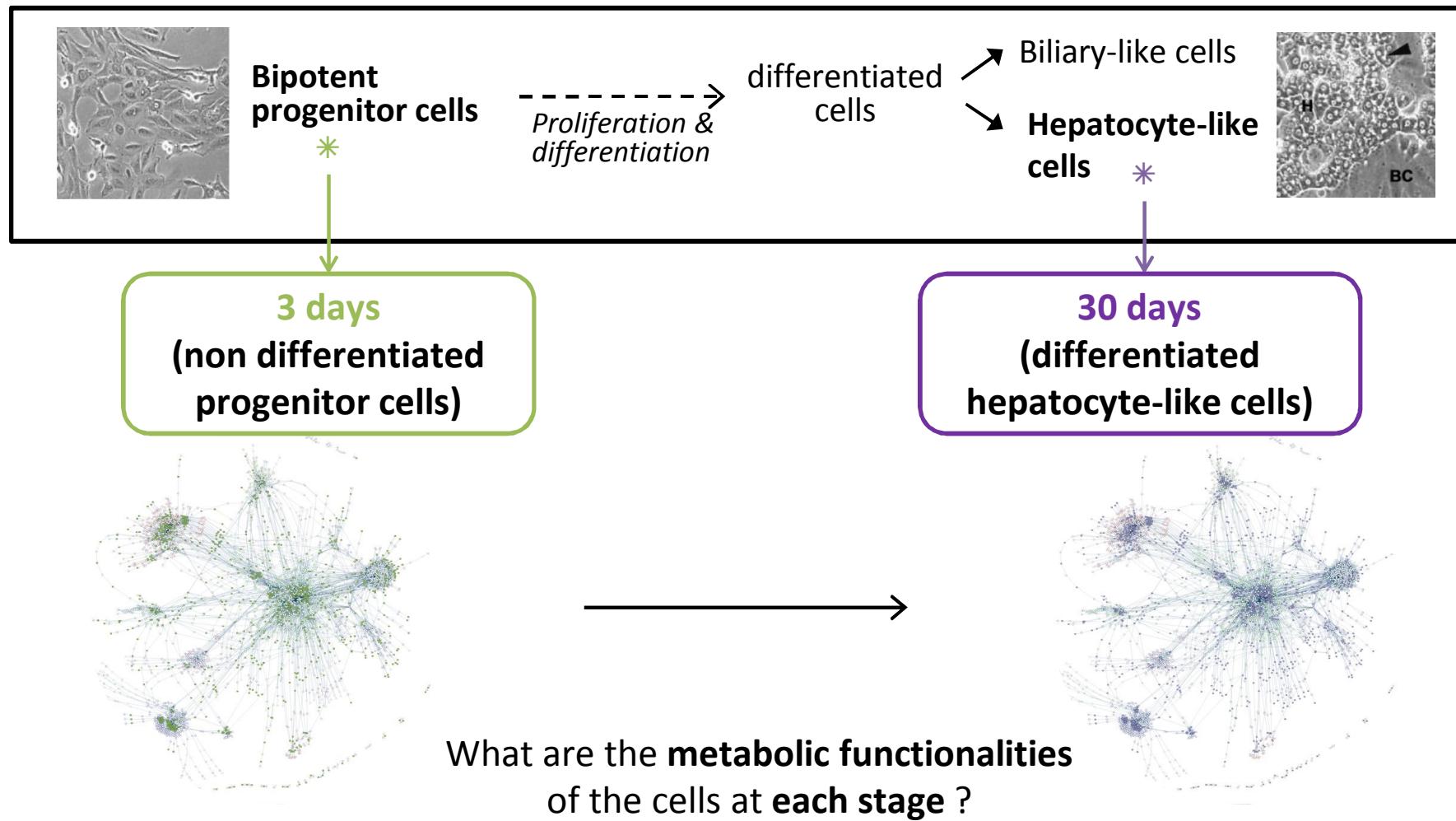
Which reactions are specifically active?



Large-Scale Modeling Approach Reveals Functional Metabolic Shifts during Hepatic Differentiation

Nathalie Poupin,^{*,†,¶} Anne Corlu,[‡] Nicolas J. Cabaton,[†] Hélène Dubois-Pot-Schneider,^{‡,§}
Cécile Canlet,[†] Elodie Person,[†] Sandrine Bruel,[†] Clément Frainay,[†] Florence Vinson,[†]
Florence Maurier,^{†,||} Fabrice Morel,[‡] Marie-Anne Robin,[‡] Bernard Fromenty,[‡] Daniel Zalko,[†]
and Fabien Jourdan[†]

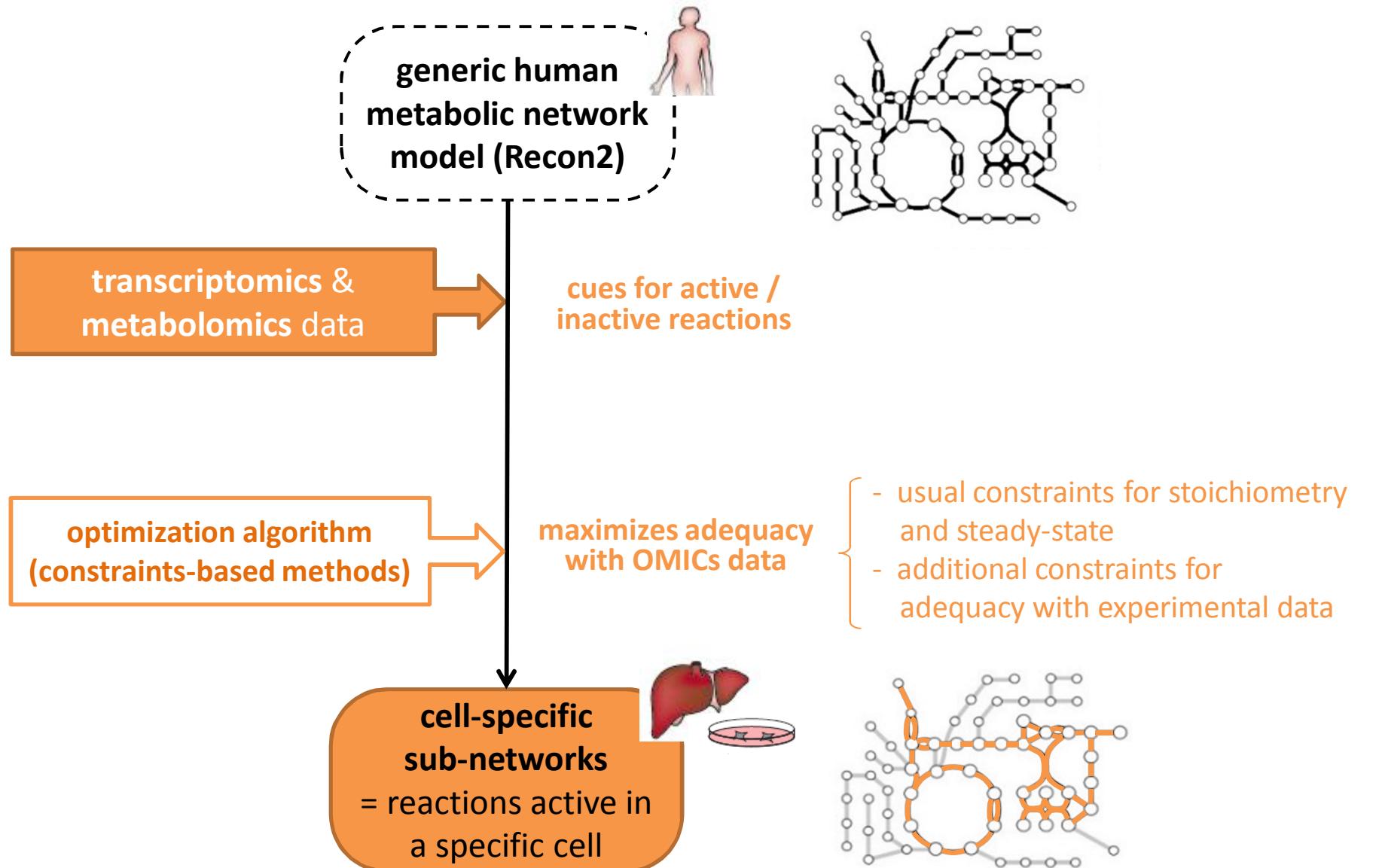
HepaRG cells



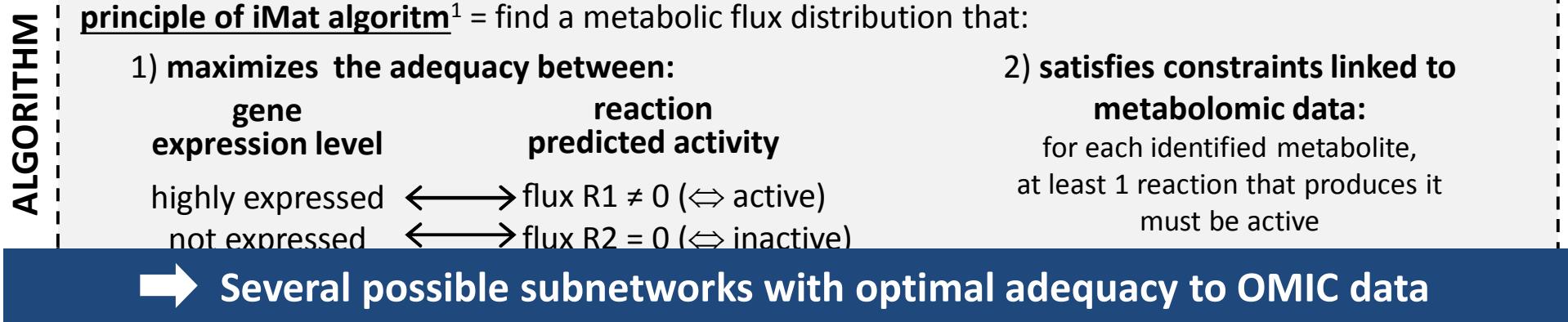
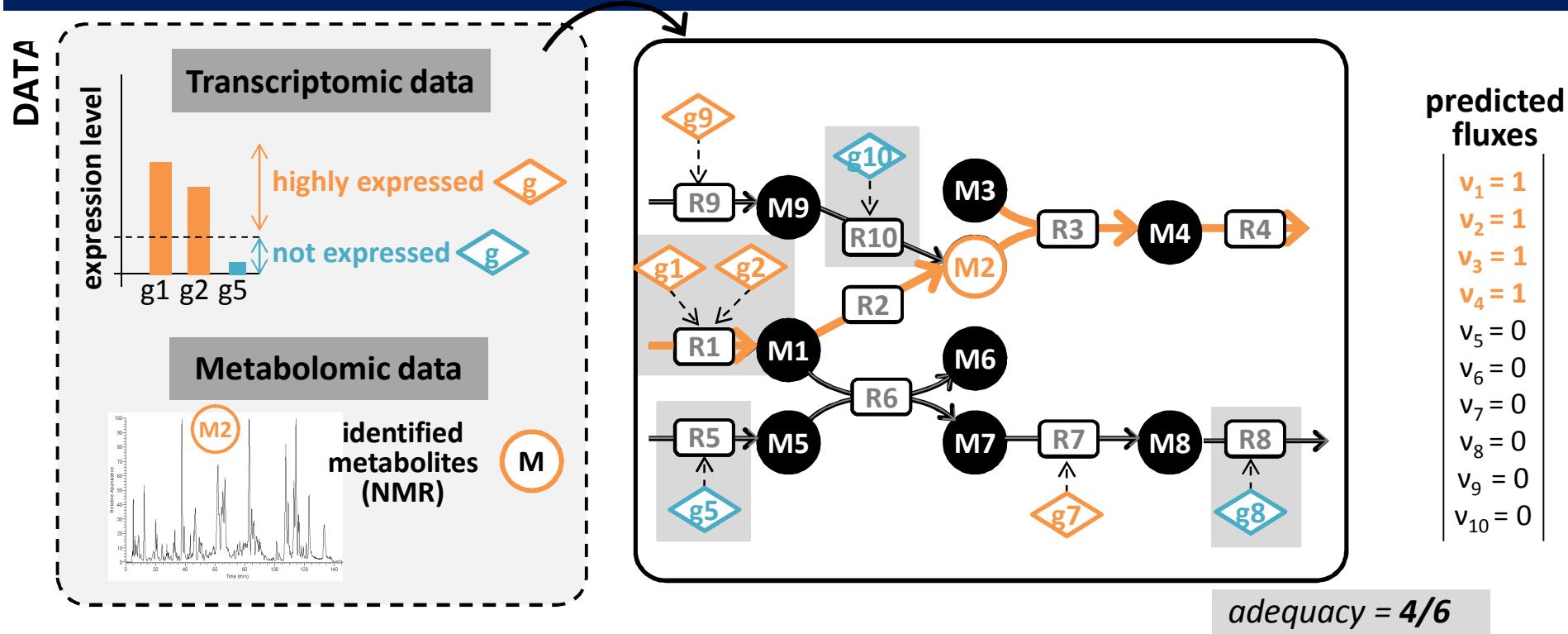
→ Reconstruction of the metabolic network of HepaRG cells at 2 differentiation stages



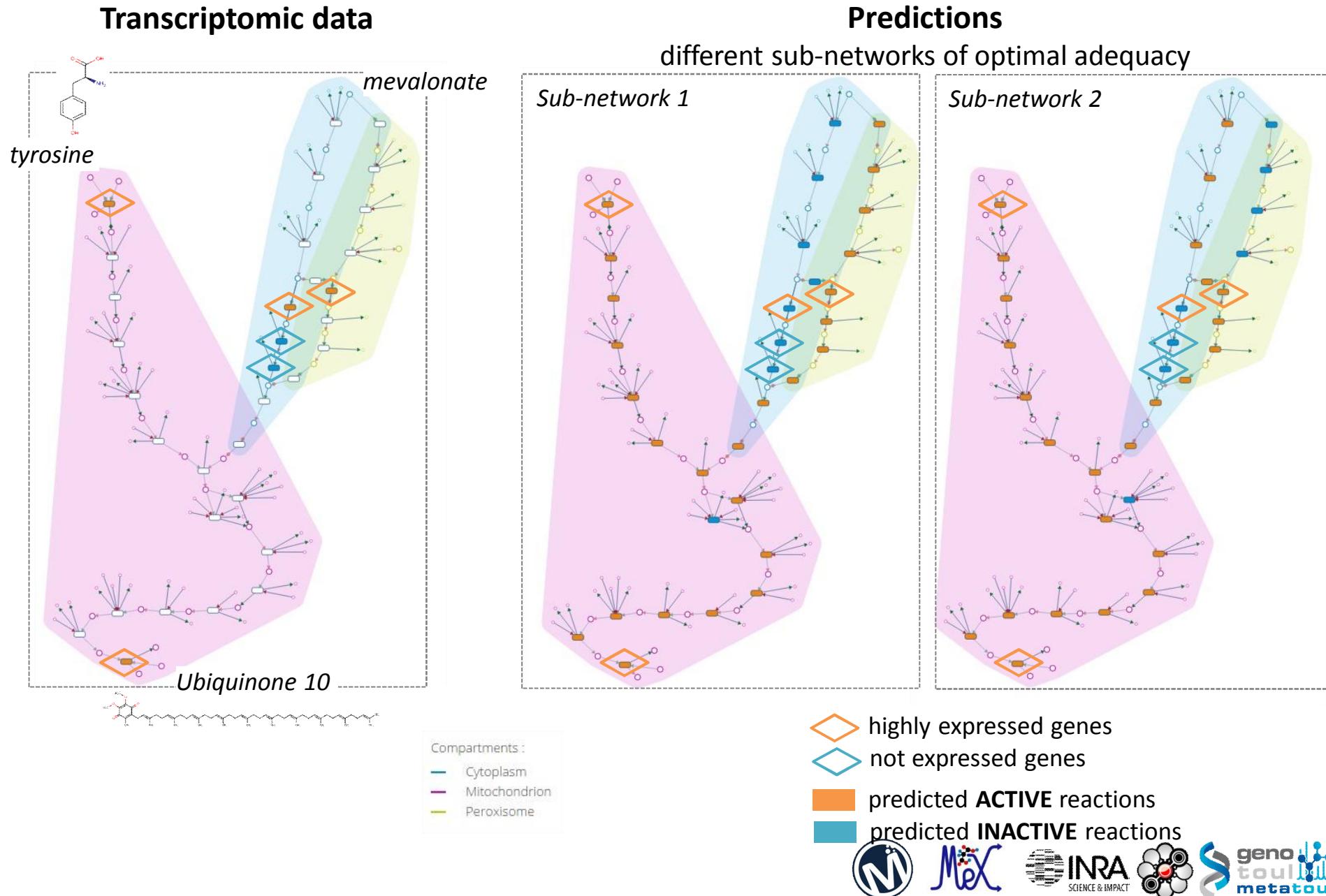
Building cell-specific networks from OMICs data: in THEORY



Building cell-specific networks from OMICs data

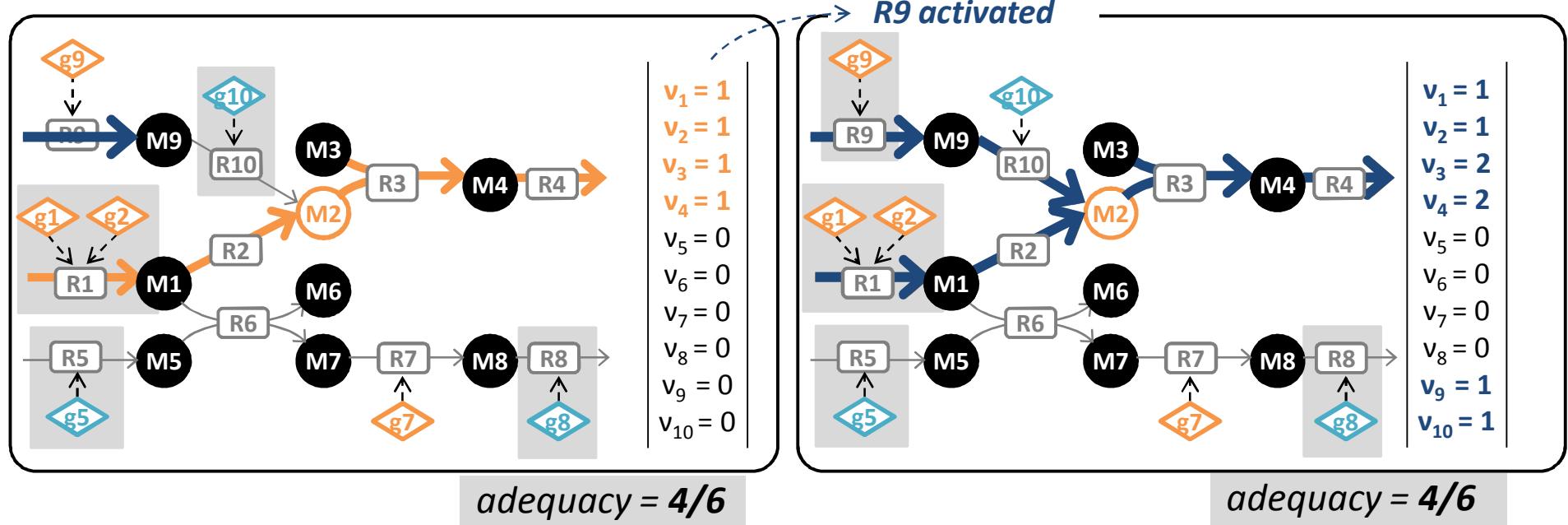


Example: Ubiquinone synthesis pathway



Finding alternative solutions

→ All reactions are successively activated or inactivated



→ several selected flux distributions
with optimal adequacy
= sub-networks of active reactions

$v_1 = 1$	$v_1 = 1$
$v_2 = 1$	$v_2 = 1$
$v_3 = 1$	$v_3 = 2$
$v_4 = 1$	$v_4 = 2$
$v_5 = 0$	$v_5 = 0$
$v_6 = 0$	$v_6 = 0$

"required" reactions
(flux $\neq 0$ in all solutions)

"inactive" reactions

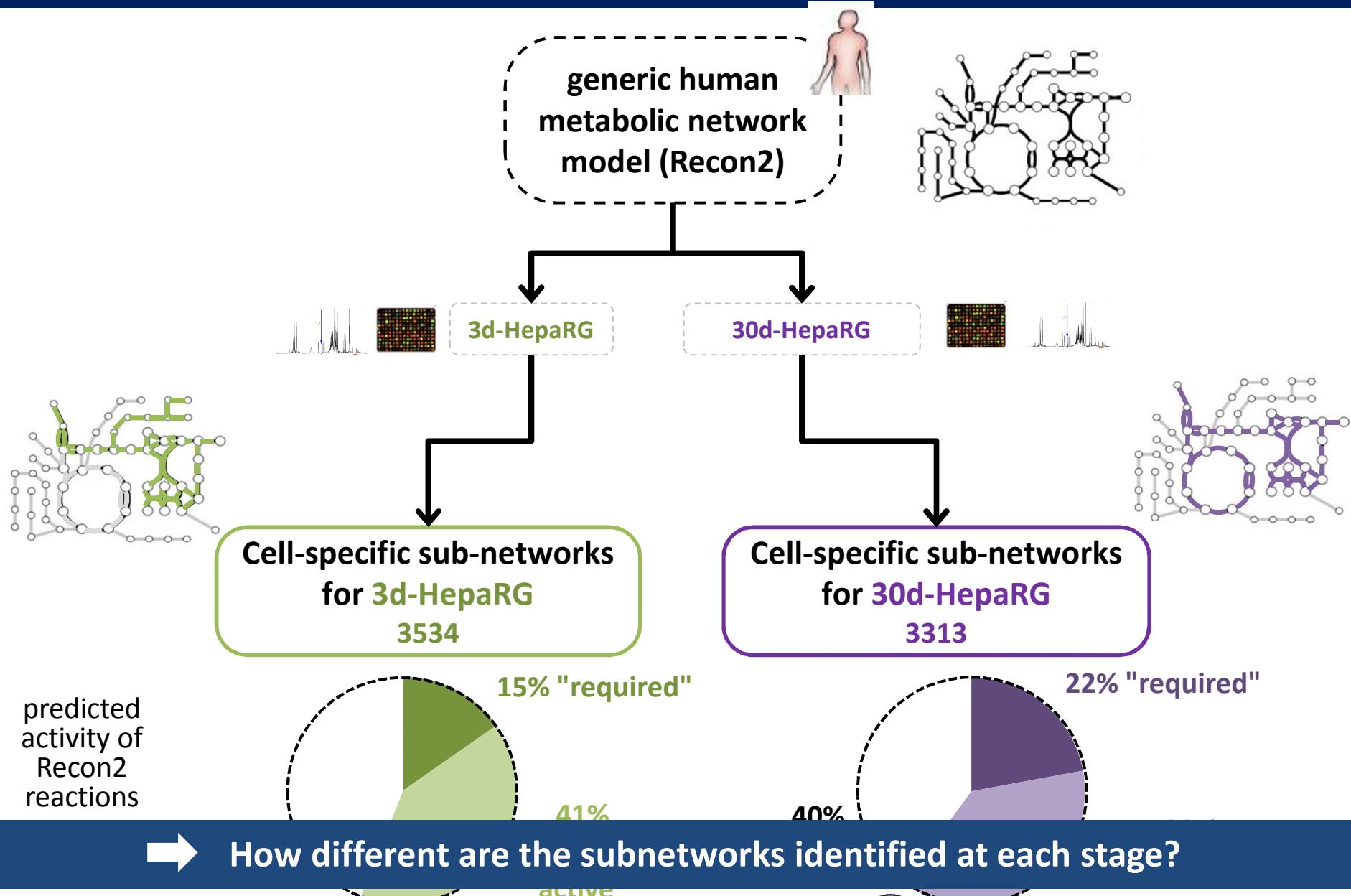
→ How to deal with all these optimal alternative subnetworks?

$v_9 = 0$	$v_9 = 1$
$v_{10} = 0$	$v_{10} = 1$

"potentially active" reactions



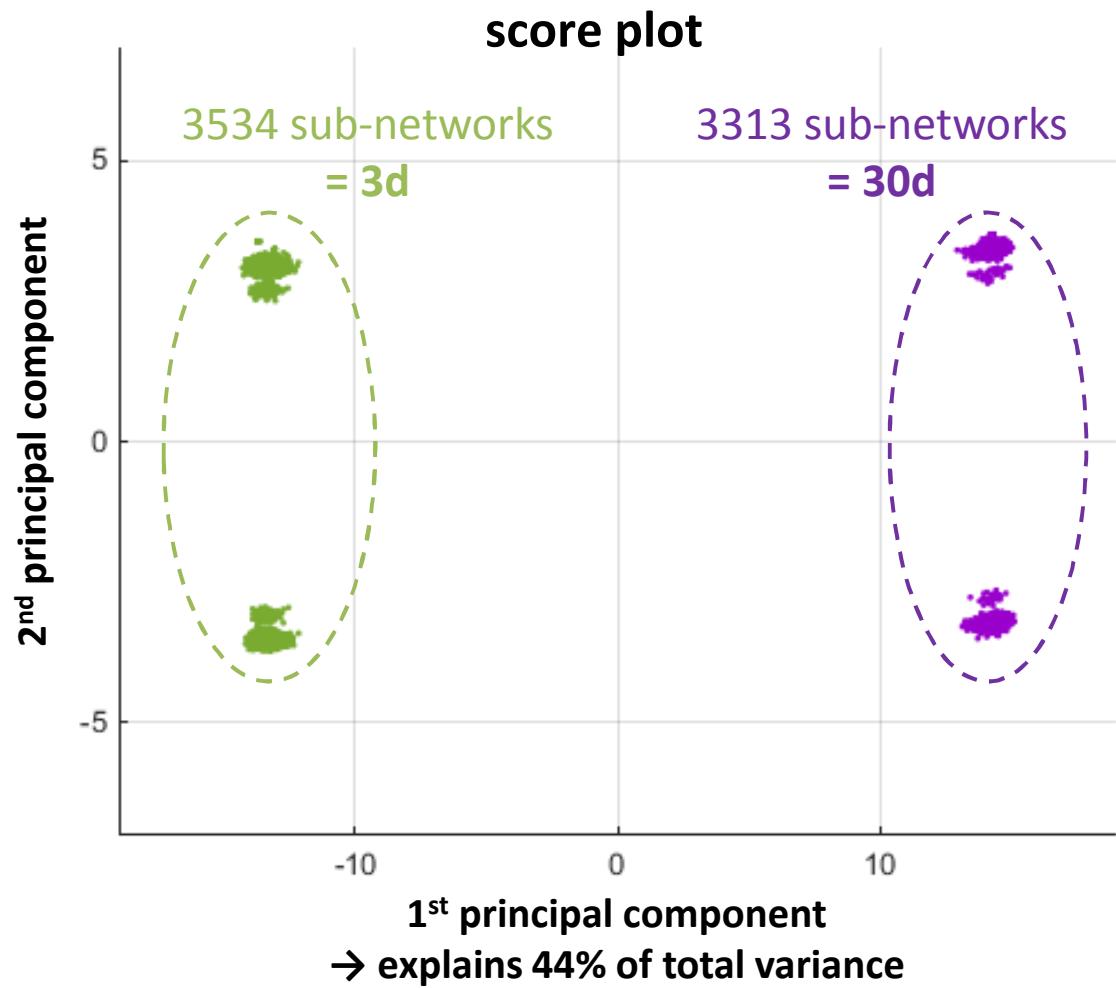
Building HepaRG-specific metabolic networks



comparison d03 vs. d30

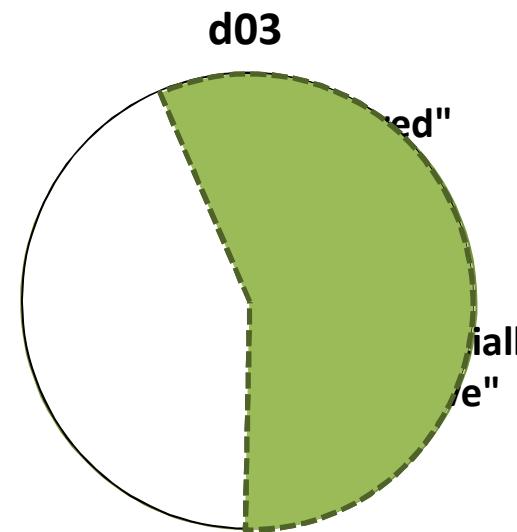
Principal component analysis : d03 & d30 – all sub-networks

	3 days			30 days		
Reactions	Network ₁	Network ₂	Network _n	Network ₁	Network ₂	Network _n
R ₁	1	1	1	1	0	0
R ₁	1	0	1	1	1	1
R ₁	1	1	0	0	0	0
R ₁	0	0	0	1	1	1
R ₁	0	0	0	1	0	1
...
R ₇₄₄₀	1	1	1	0	0	1



comparison d03 vs. d30

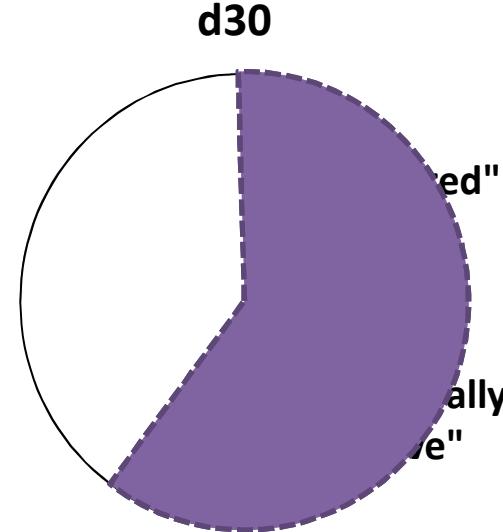
identification of ACTIVATED and INACTIVATED reactions



4178 predicted
active reactions

Reactions
predicted to be
INACTIVATED

200



4462 predicted
active reactions

484

•----->

Reactions
predicted to be
ACTIVATED



comparison d03 vs. d30

Pathway enrichment analysis on reactions:

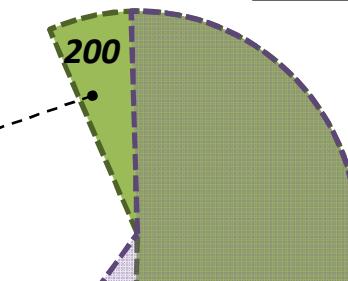
predicted to be **INACTIVATED**

Pathway	P-value
Transport, extracellular	5.60e ⁻³¹
Fatty acid synthesis	1.30e ⁻¹⁵

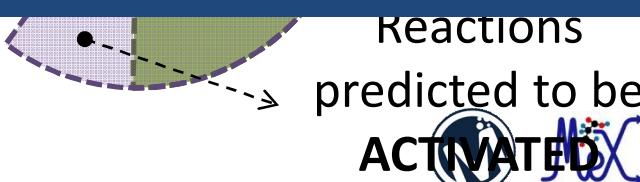
predicted to be **ACTIVATED**

Pathway	P-value
Biotin metabolism	1.31e ⁻¹¹
Fatty acid oxidation	6.80e ⁻⁰⁷
Bile acid synthesis	5.13e ⁻⁰⁵
Tryptophan metabolism	2.44e ⁻⁰⁴
Cytochrome metabolism	3.76e ⁻⁰⁴
Blood group synthesis	1.59e ⁻⁰³
Lysine metabolism	4.44e ⁻⁰³
Limonene and pinene degradation	3.45e ⁻⁰²

Reactions
predicted to be
INACTIVATED



→ Further investigation of the pathways using visualization tools



Reactions
predicted to be
ACTIVATED



Fatty acid oxidation pathway



MetExploreViz

Comparison
d03 vs. d30

Endoplasmic
reticulum
& cytoplasm

peroxisome

mitochondria



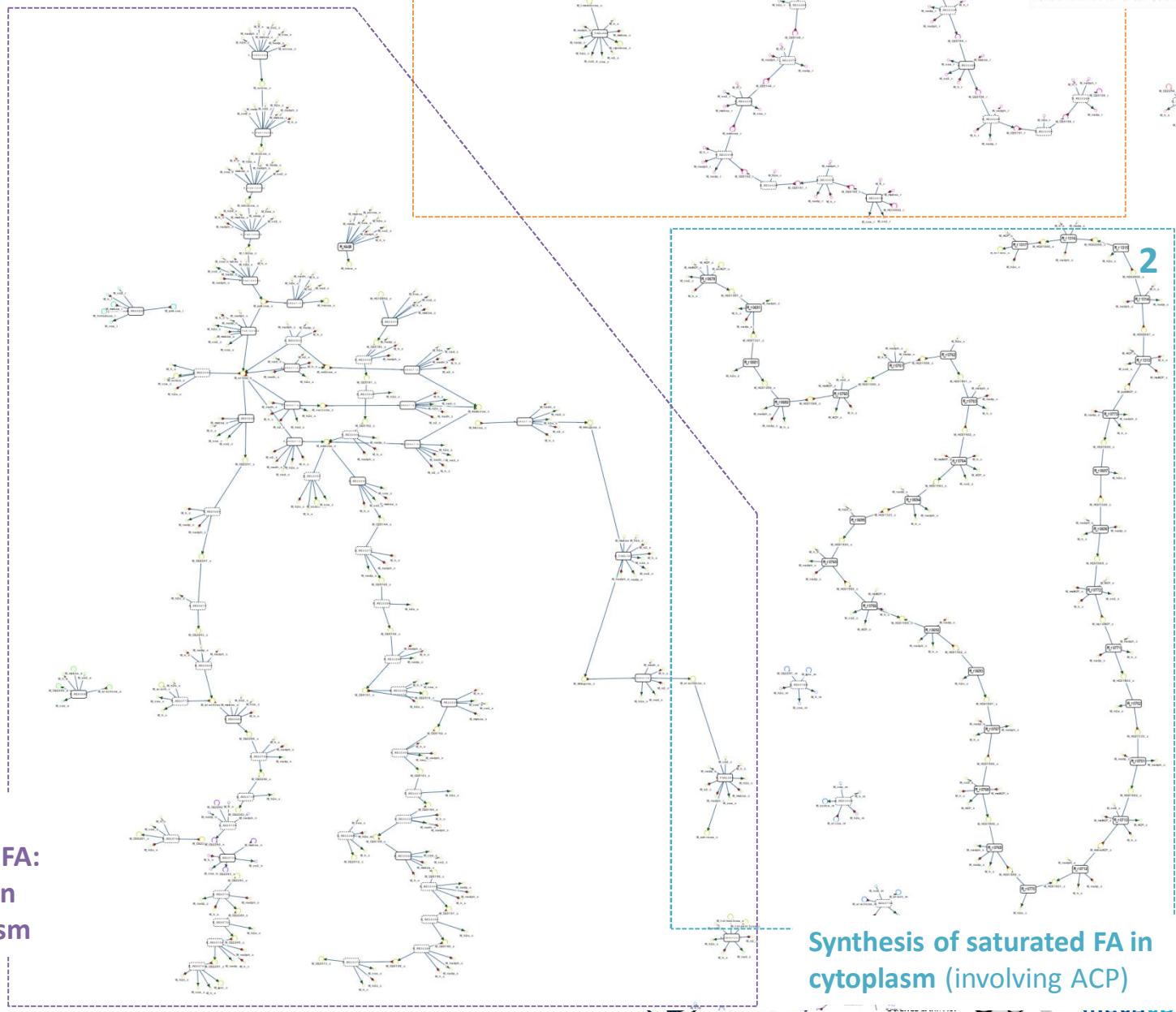
predicted ACTIVATED reactions

FA synthesis in Recon2

FA long chain synthesis in ER:
elongation from palmitoylCoA (C16:1)
to cis15-tetracosenoylCoA (C24:1)



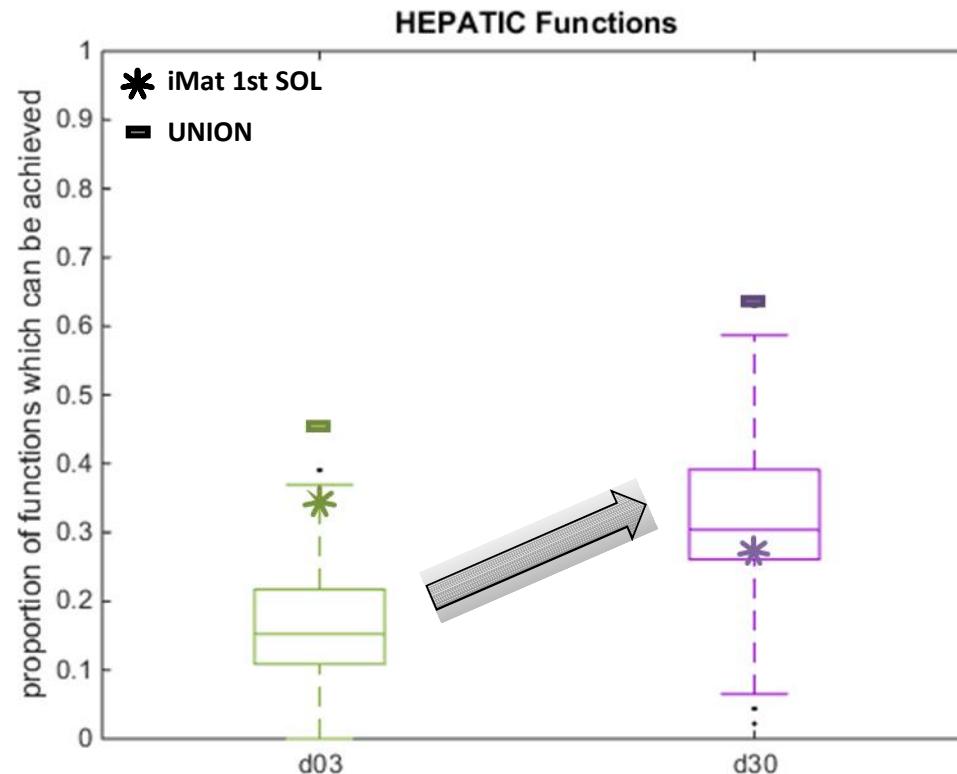
MetExploreViz



Evaluating the capacity to perform hepatic functions

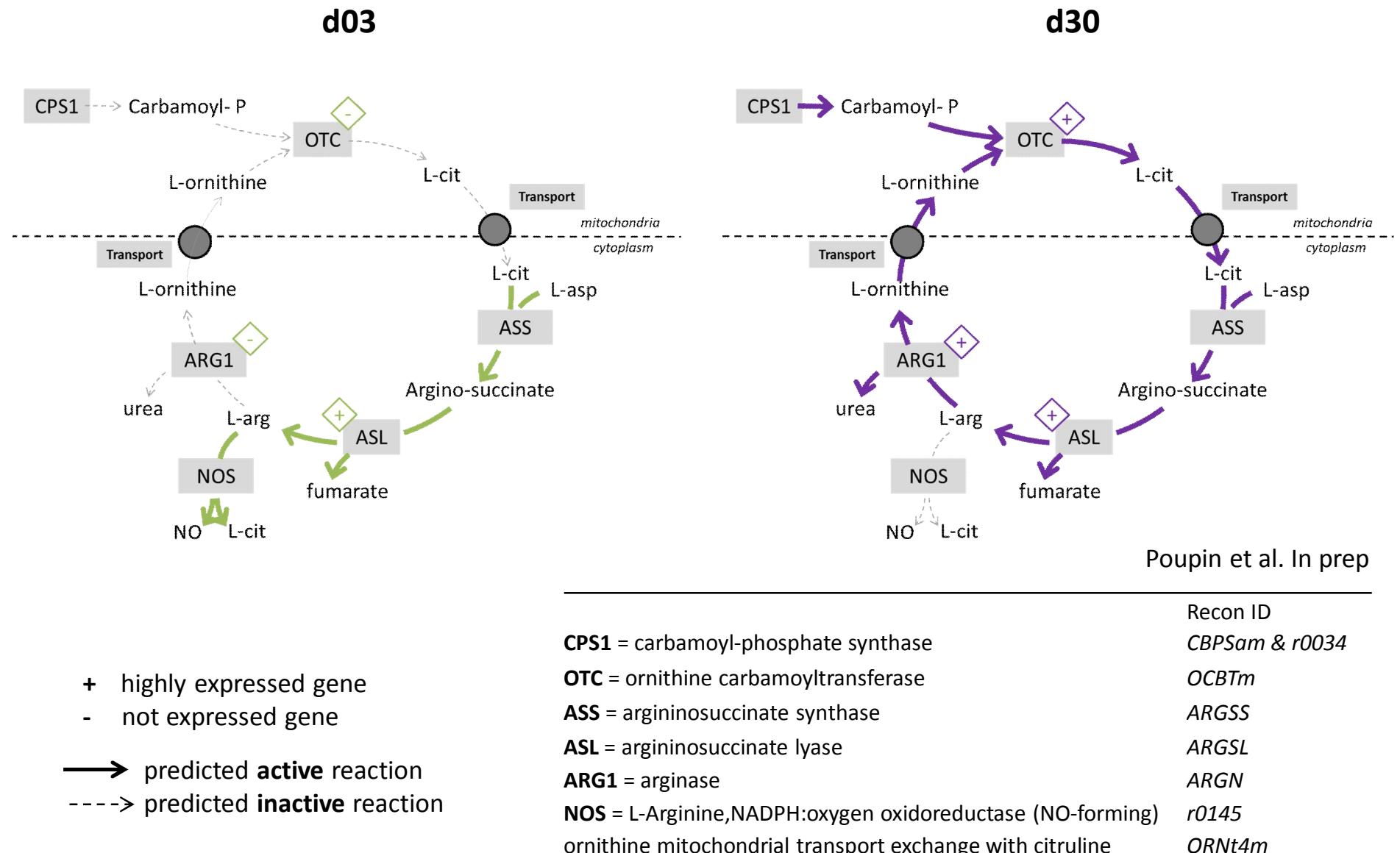
Simulation of 44 defined HEPATIC functions

(functions known to specifically take place in liver cells, such as ammonia detoxification through ureogenesis, ketogenesis and bile formation).



→ On average, d30-cells are predicted to be able to perform a higher number of hepatic functions → Focus on the urea synthesis function

Network analysis highlights the activation of urea cycle





N. Poupin
L. Fernando
M. Chazalviel
F. Vinson
B. Merlet
C. Frainay
N. Cabaton
D. Zalko



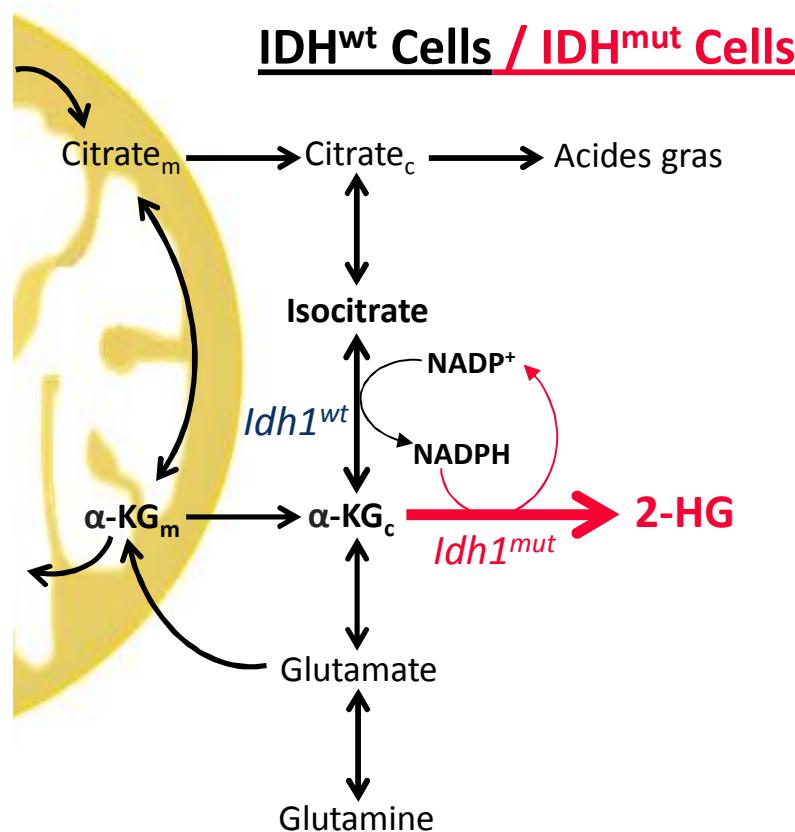
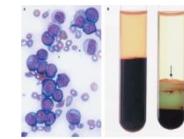
T. Palama
P. Millard
F. Bellvert
L. Peyriga
E. Cahoreau
L. Gales
J.C. Portais



MODELLING THE METABOLIC NETWORK OF AML CELLS TO UNDERSTAND GLOBAL METABOLIC SHIFTS INDUCED BY IDH1 MUTATION

Acute Myeloid Leukemia (AML)

- “ Heterogeneous hematological malignancies with **accumulation of transformed hematopoietic stem cells** in the bone marrow and blood.
- “ **prognosis of this disease remains unfavorable** with current treatments due to **frequent relapses**.



- “ **mutation of the metabolic enzyme IDH1/IDH2** (isocitrate deshydrogenases) observed in 15% of AML patients
- “ The mutation drives the production of an oncometabolite 2-HG (2-hydroxyglutarate)
- “ A corpus of evidences for a higher metabolic flexibility of mutant cells.



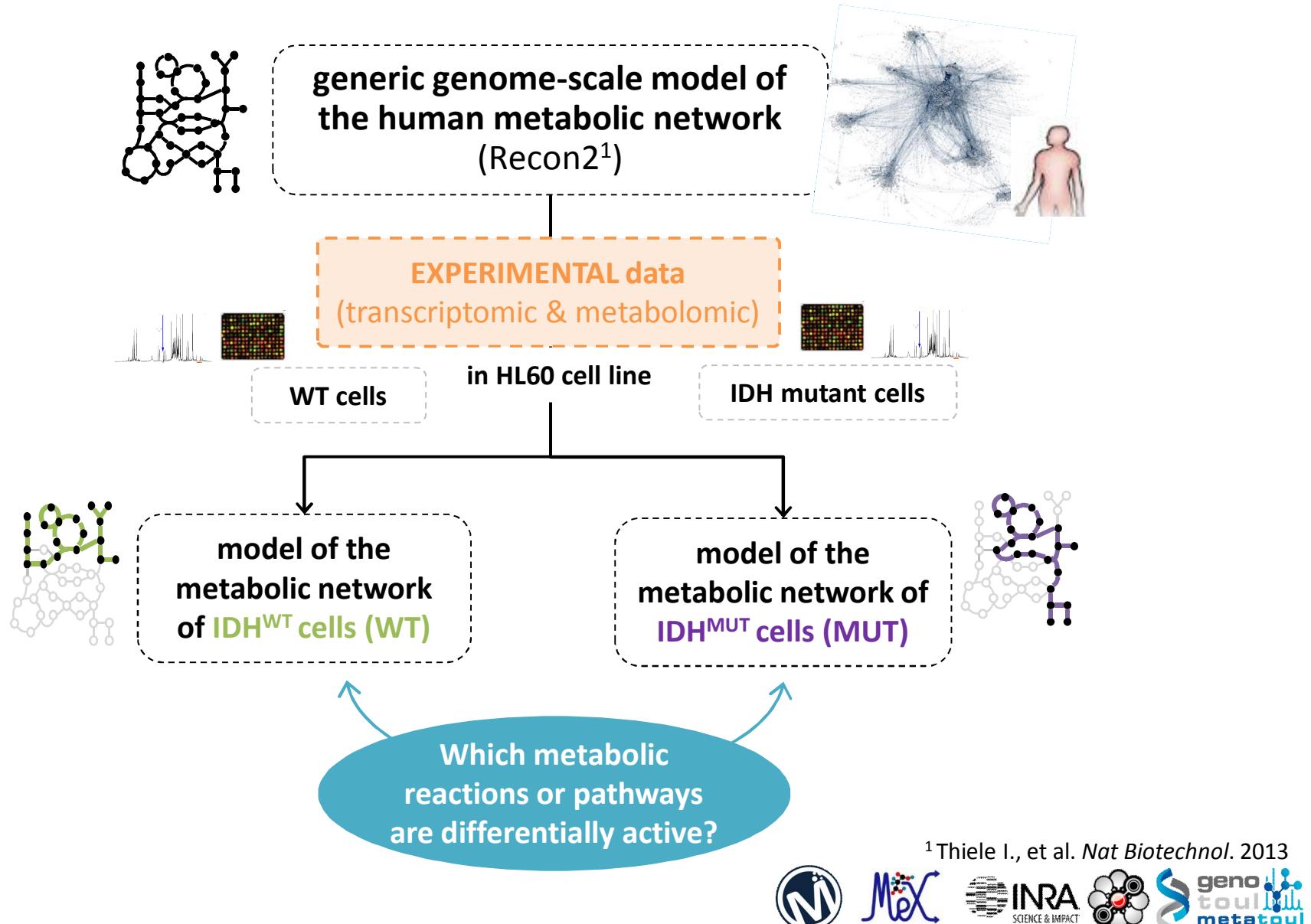
Team RESIST@ML - JE. Sarry
CRCT – UMR1037 – INSERM & UT3

→ *in silico* global analysis to better understanding the metabolism reprogramming



General pipeline

Objective: building *in silico* genome-scale models for IDH^{wt} and IDH^{mut} cells



Constraints from TRANSCRIPTOMIC data

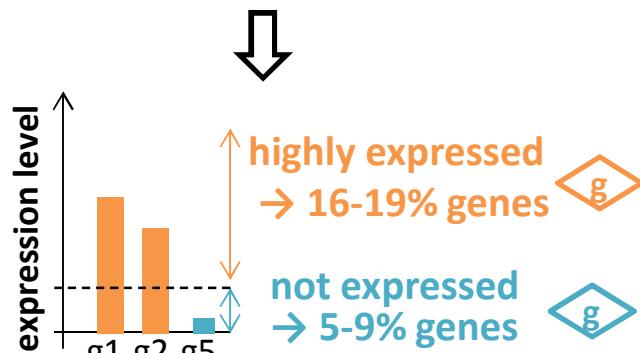


TRANSCRIPTOMIC DATA

Micro-array:

t1-16650001	4.906	4.906
t2-16650003	5.076	5.251
t3-16650005	5.853	5.535
...		

→ Expression values for
23 567 identified transcripts



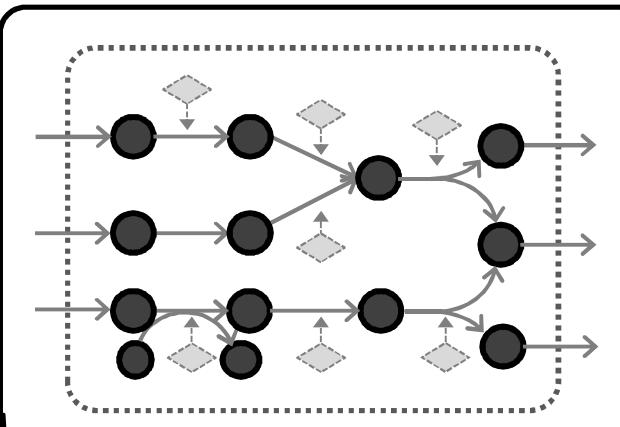
Published March 7, 2016
JEM

Brief Definitive Report

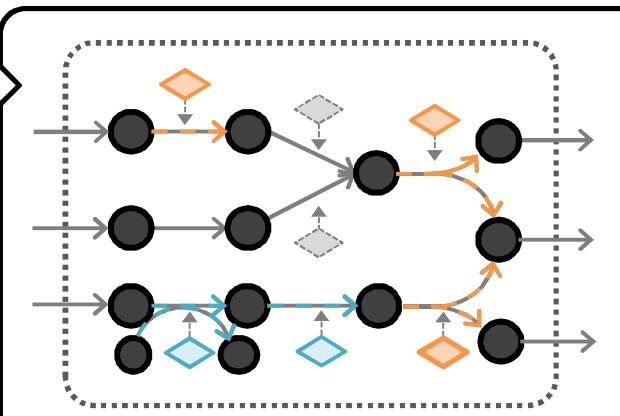
Isocitrate dehydrogenase 1 mutations prime the all-trans
retinoic acid myeloid differentiation pathway in acute
myeloid leukemia

Boutzen et al. (2016). J Exp Med.

GENERIC HUMAN METABOLIC NETWORK



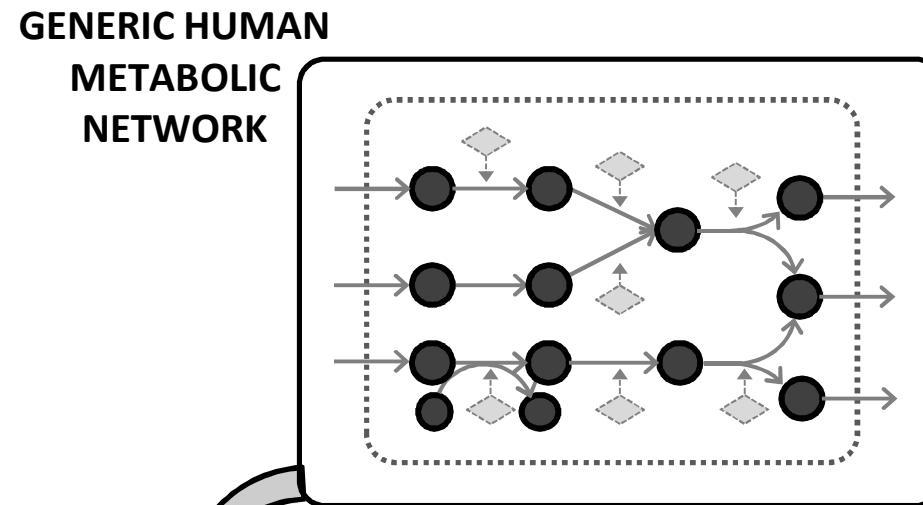
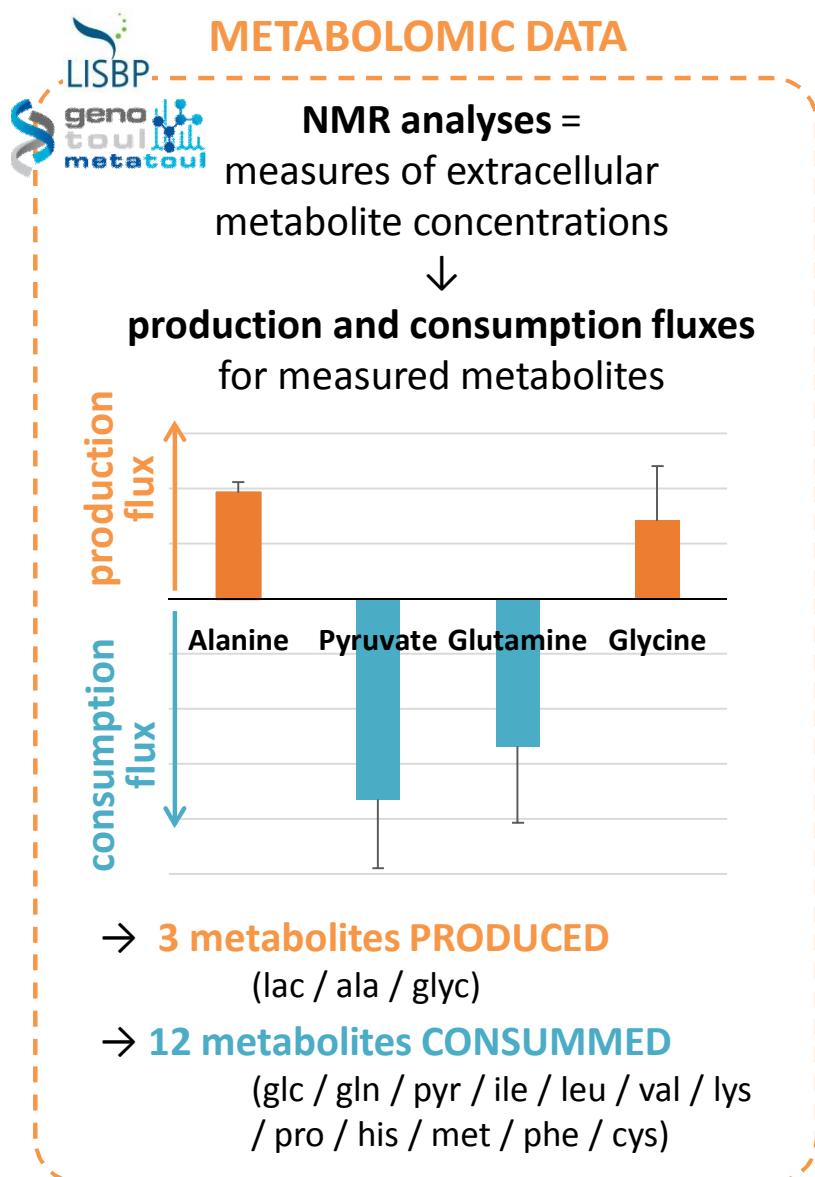
mapping on model



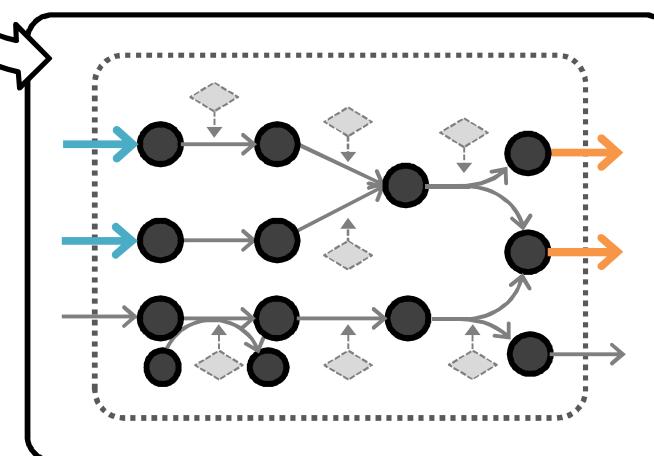
6-7% reactions associated with highly expressed genes
3-5% reactions associated with not expressed genes



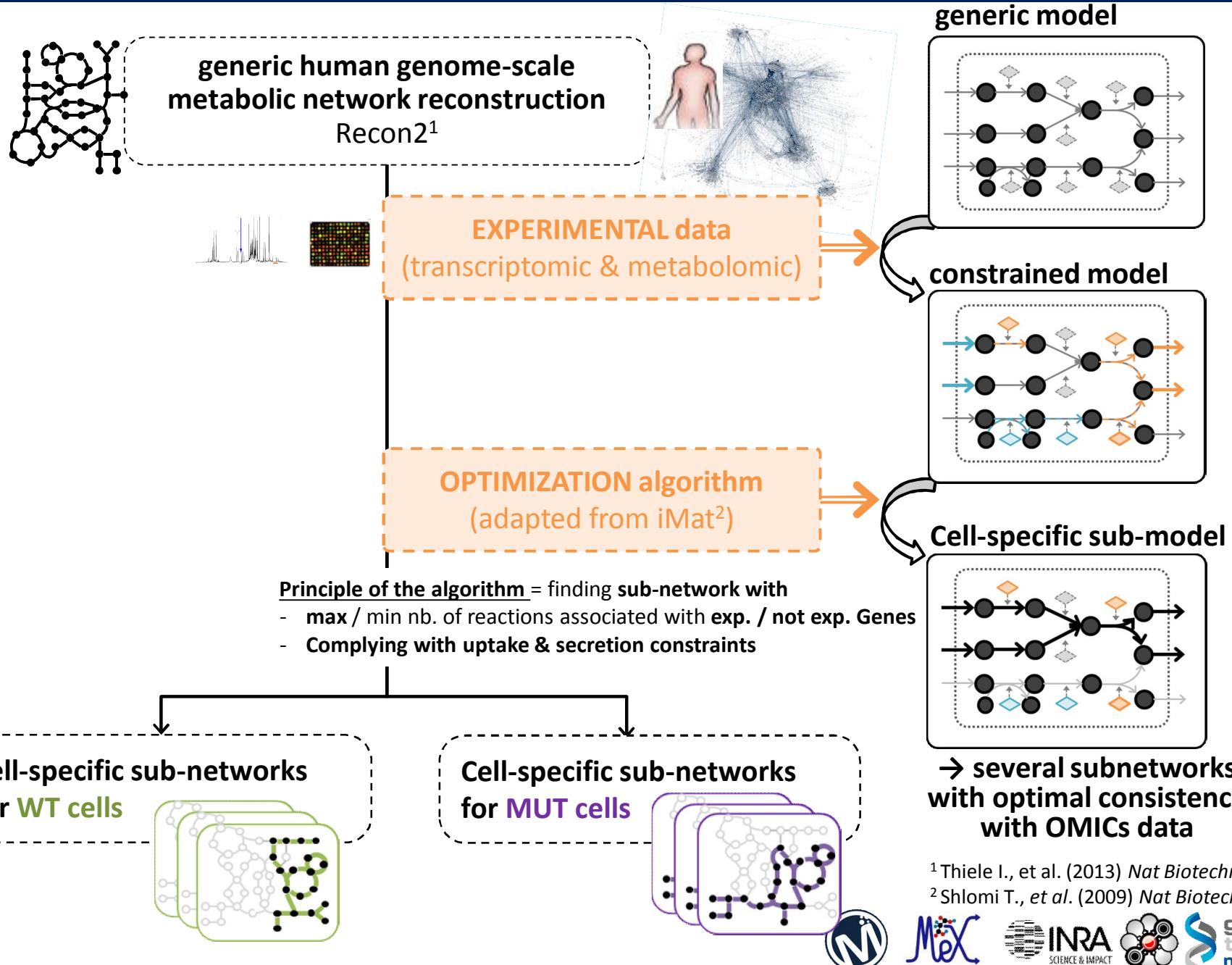
Constraints from METABOLOMIC data



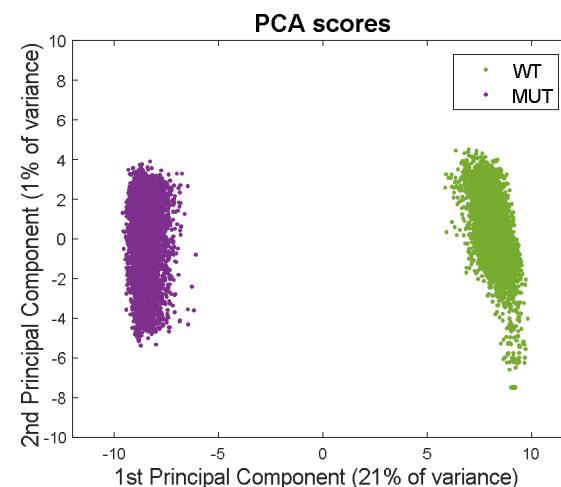
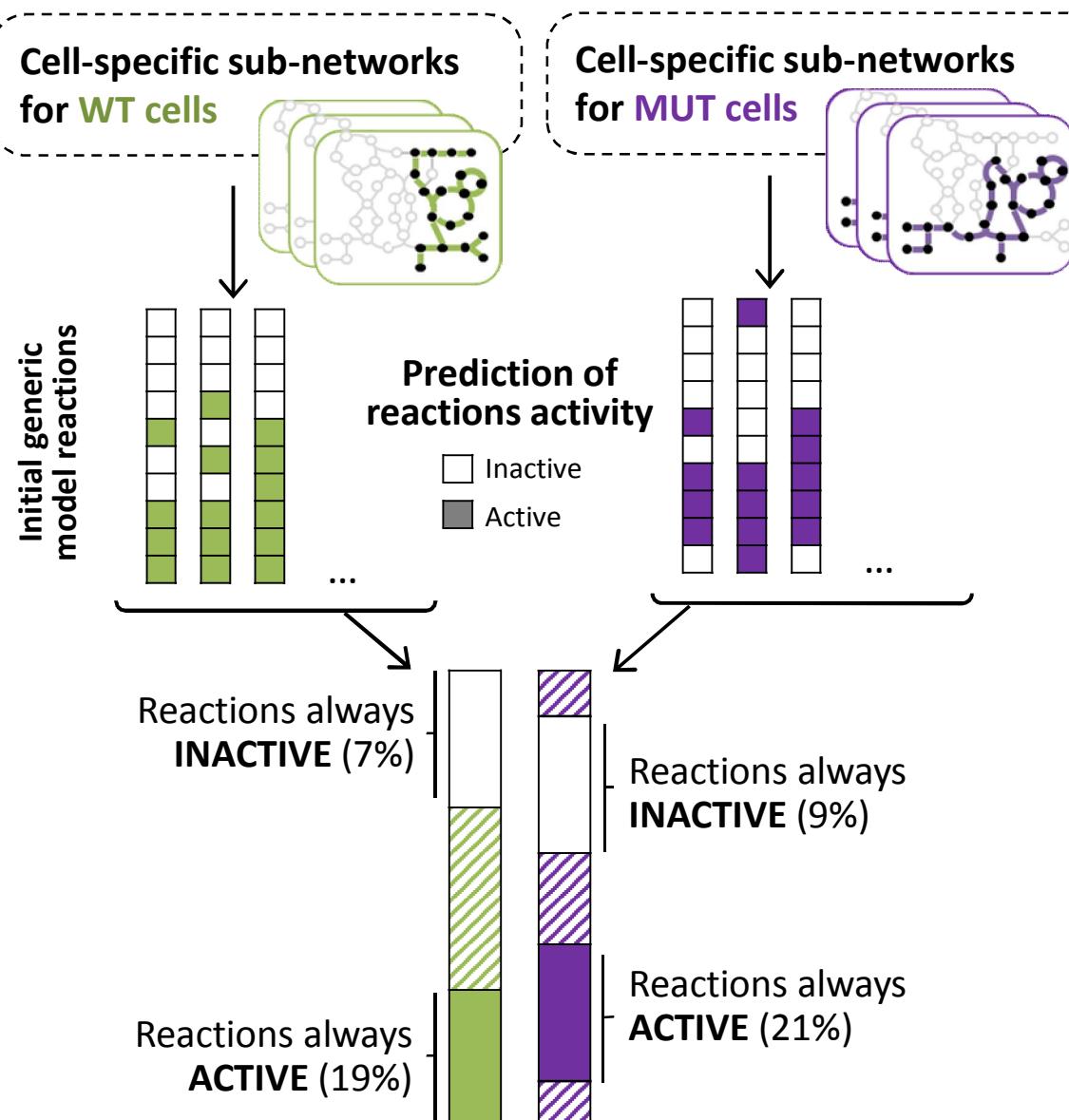
→ **Constraints on model**
= uptake or secretion of metabolites



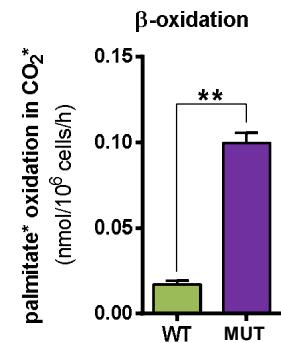
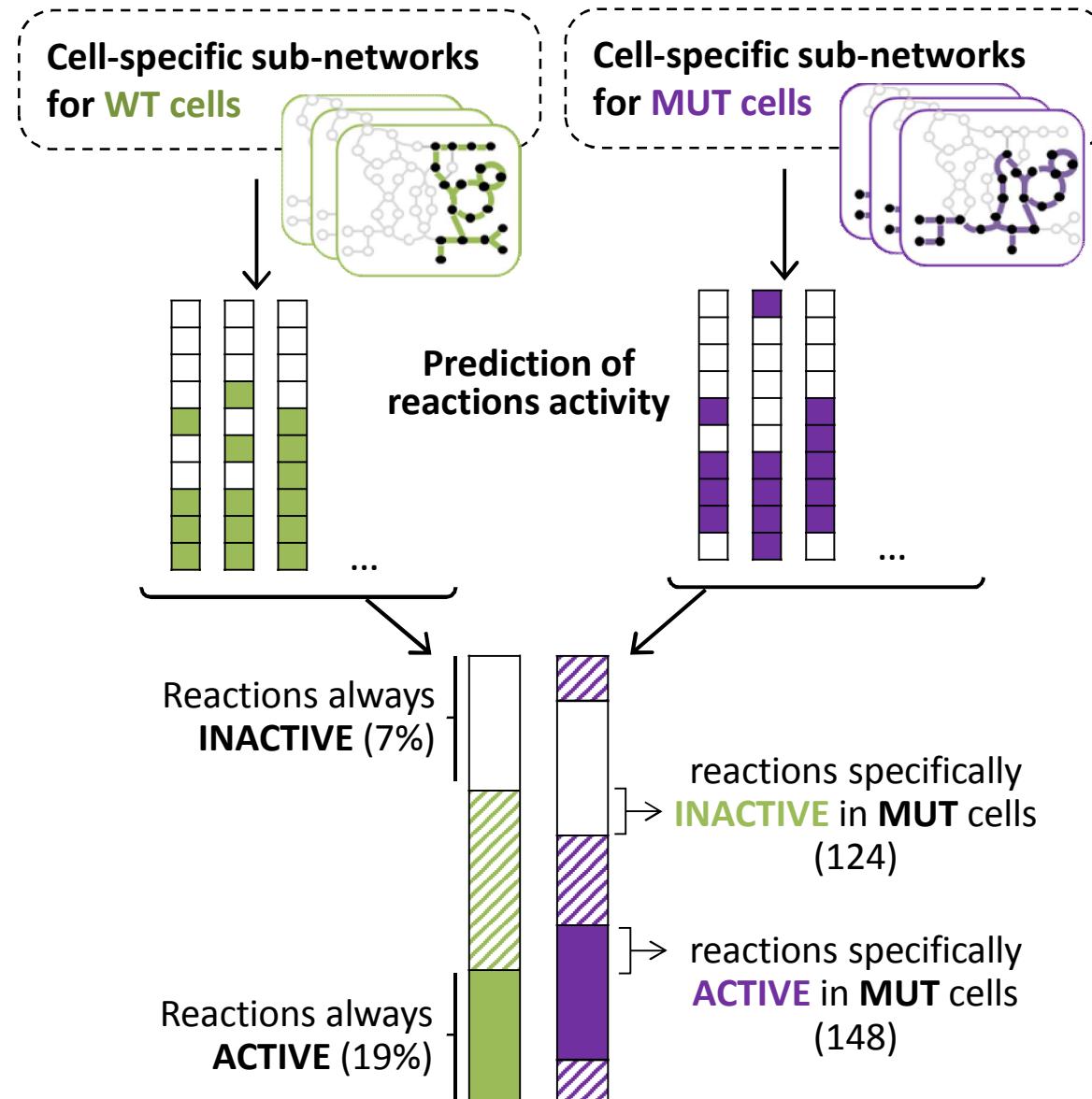
Building cell-specific models for WT and MUT cells



Comparison of cell-specific subnetworks



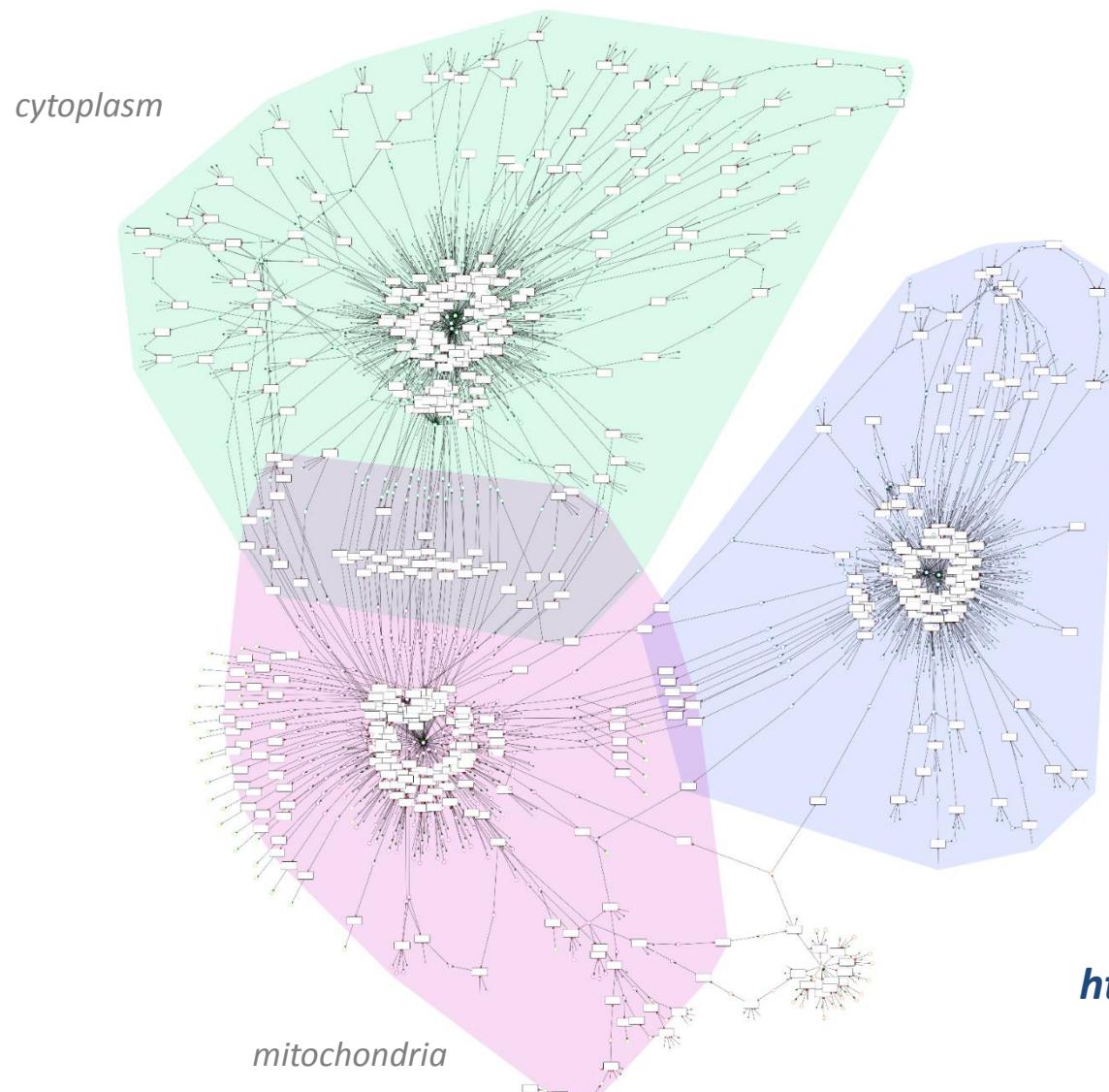
Comparison of cell-specific subnetworks



Pathway enrichment

Fatty Acid oxidation	***
Steroid metabolism	**
Glyoxylate & dicarboxylate metab.	*
Fatty Acid oxidation	***
N-glycan synthesis	***
Fructose & mannose metab.	**
Glyoxylate & dicarboxylate metab.	**

Visualization helps in the interpretation of modulated reactions

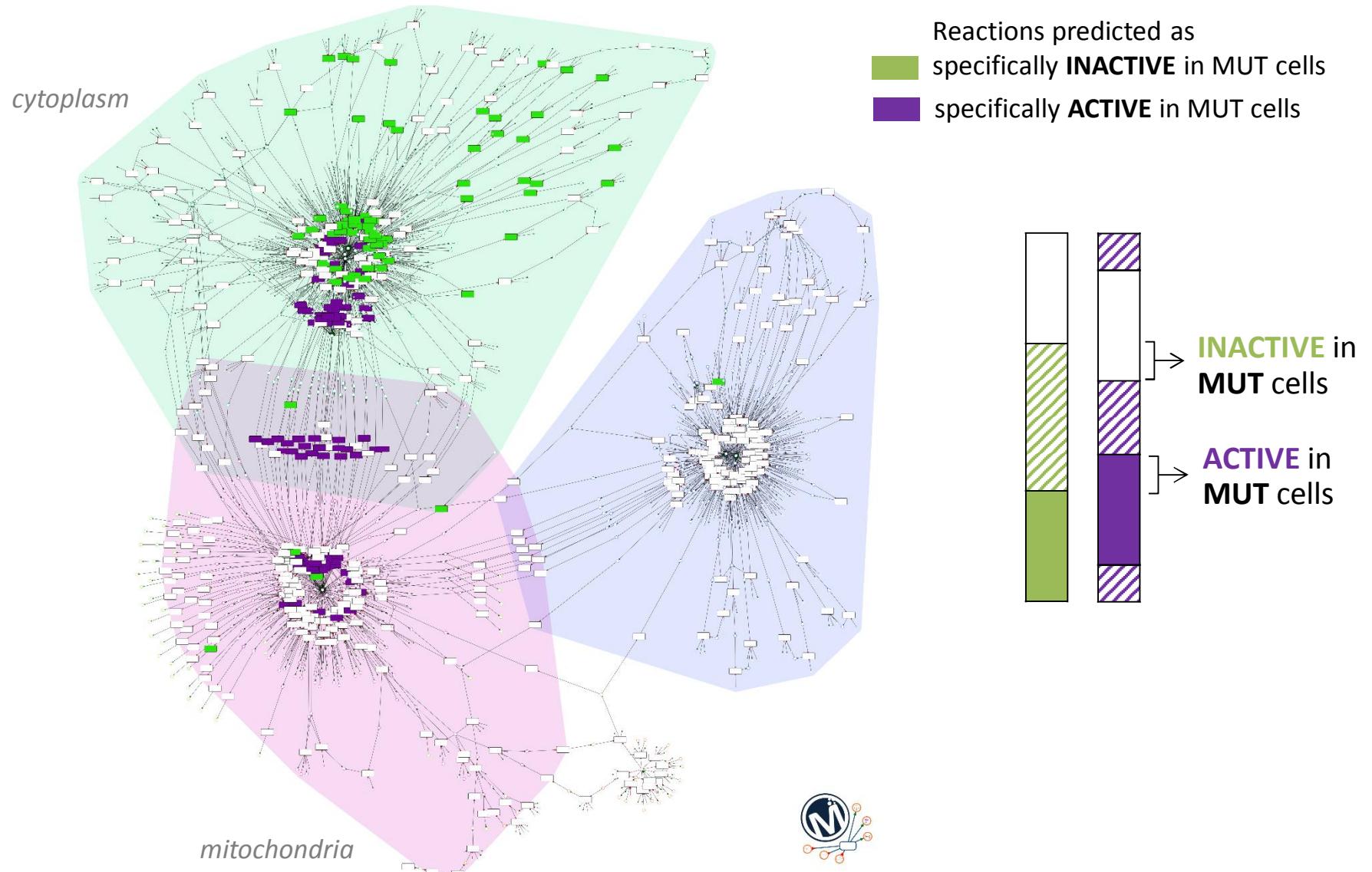


**reactions in the Fatty Acid oxidation pathway
(577 reactions)**

<https://metexplore.toulouse.inra.fr>



Visualization helps in the interpretation of modulated reactions



LETTER

doi:10.1038/nature10363

Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase

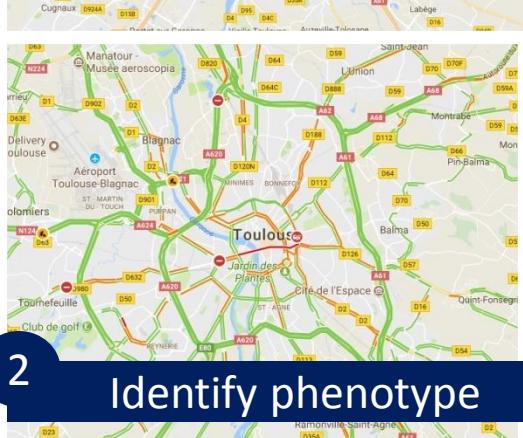
Christian Frezza¹, Liang Zheng¹, Ori Folger², Kartik N. Rajagopalan³, Elaine D. MacKenzie¹, Livnat Jerby², Massimo Micaroni⁴, Barbara Chaneton¹, Julie Adam⁵, Ann Hedley¹, Gabriela Kalna¹, Ian P. M. Tomlinson⁶, Patrick J. Pollard⁵, Dave G. Watson⁷, Ralph J. Debernardinis³, Tomer Shlomi^{8*}, Eytan Ruppin^{2,9*} & Eyal Gottlieb¹

IN SILICO PERTURBATIONS: DRUG TARGET
PREDICTION

3 Perturbate in silico the system and predict changes



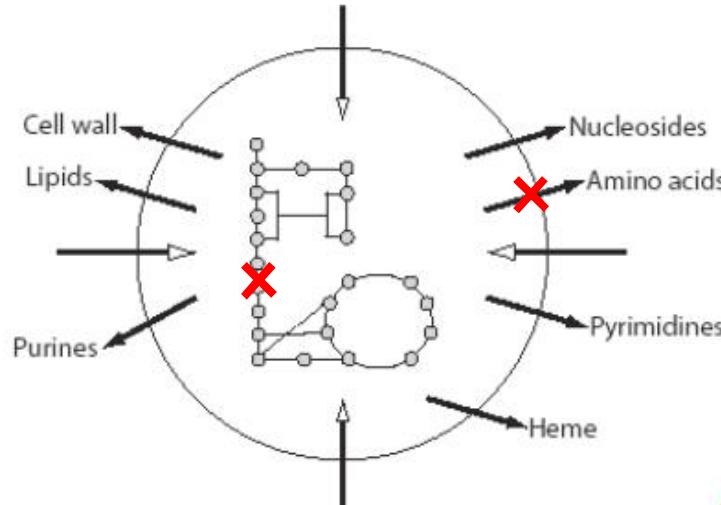
1 Reconstruction



2 Identify phenotype



3 Simulate perturbations



Flux Balance Analysis: predicting if changes in fluxes allow or not to reach a metabolic objective (e.g. producing sufficient biomass for growth)

Allows testing in silico impact of gene KO (or combination of gene KOs) on metabolism.

Assist in drug target discovery.

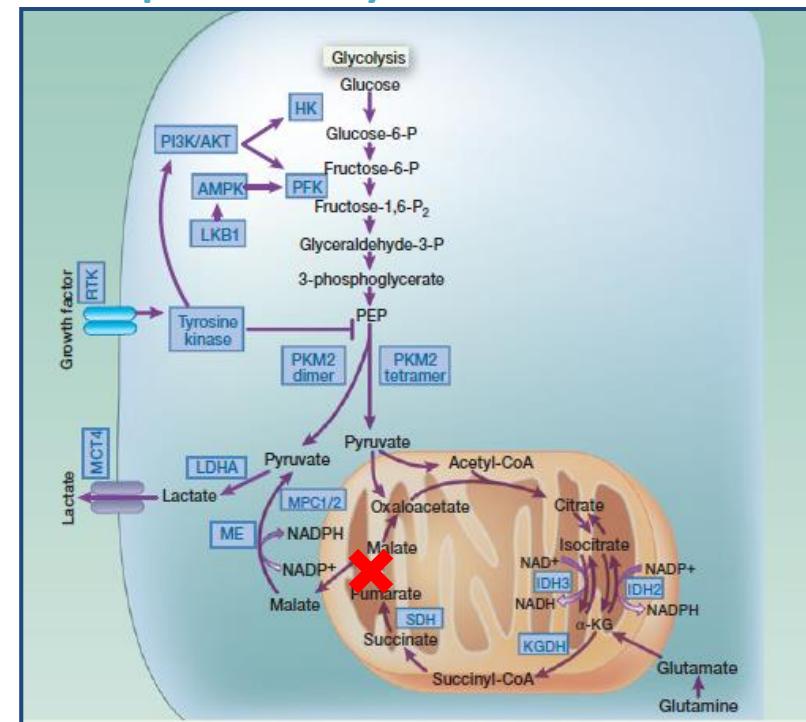
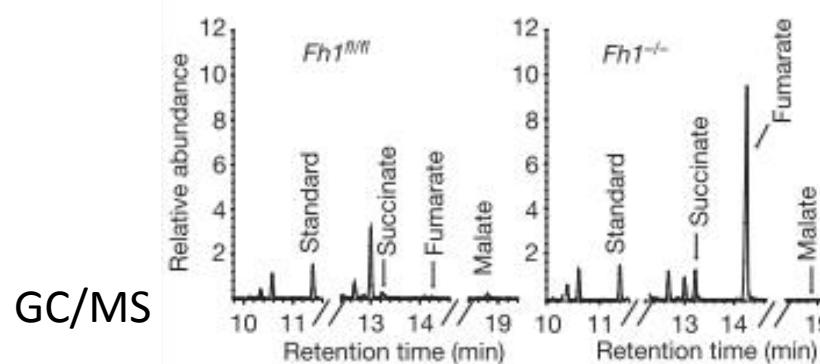


Predict system behaviour: gene deletion analysis

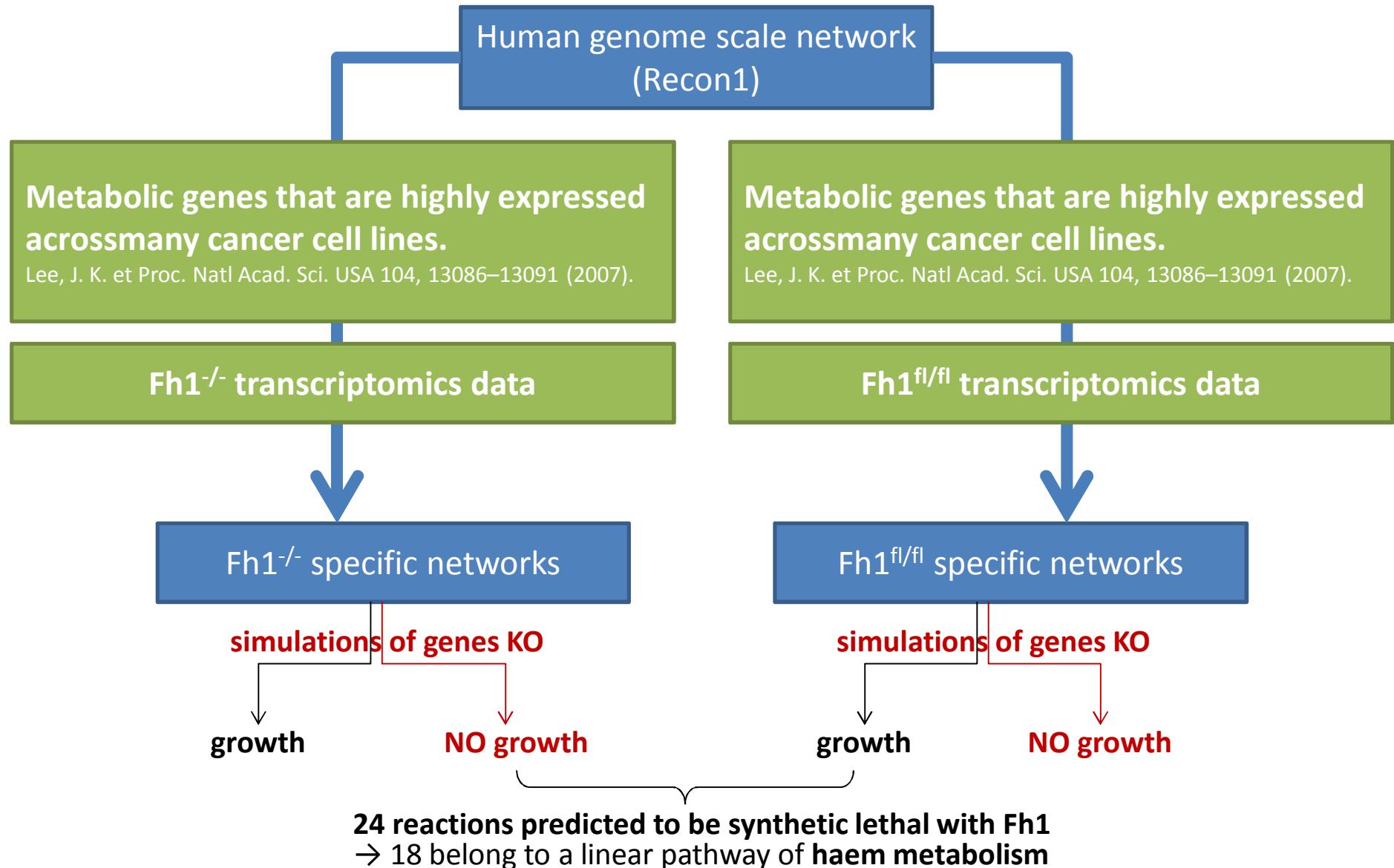
Example: Targeting of synthetic lethal gene for the treatment of HLRCC patients
(Frezza C. et al., Nature, 2011)

Fumarate hydratase (FH) is an **enzyme of the TCA cycle** that catalyses the hydration of **fumarate into malate**. **Renal-cancer cells (HLRCC)** are **deficient for FH** and display a truncated TCA cycle. As TCA is a major source for mitochondrial NADH, its truncature may have severe bioenergetic outcomes. **But cancer cells are able to survive despite a non functional TCA, and no mechanism to explain this particularity had been provided.**

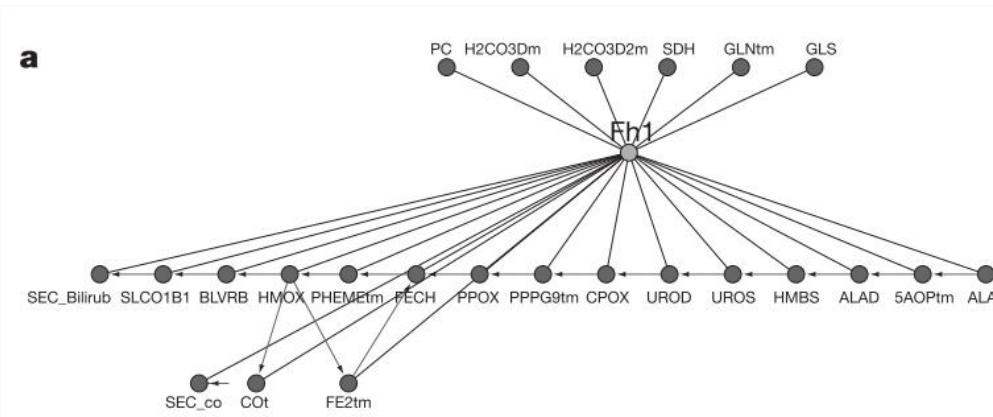
→ ***in silico* modelling approach to identify genes that are synthetic lethal with Fh1**
(predict genes KO, that together with Fh1 mutation, would selectively affect the growth ability of cancer cells without affecting the wild-type cells).



Construction of Fh1^{-/-} and Fh1^{fl/fl} specific networks



Predicting synthetic lethal genes

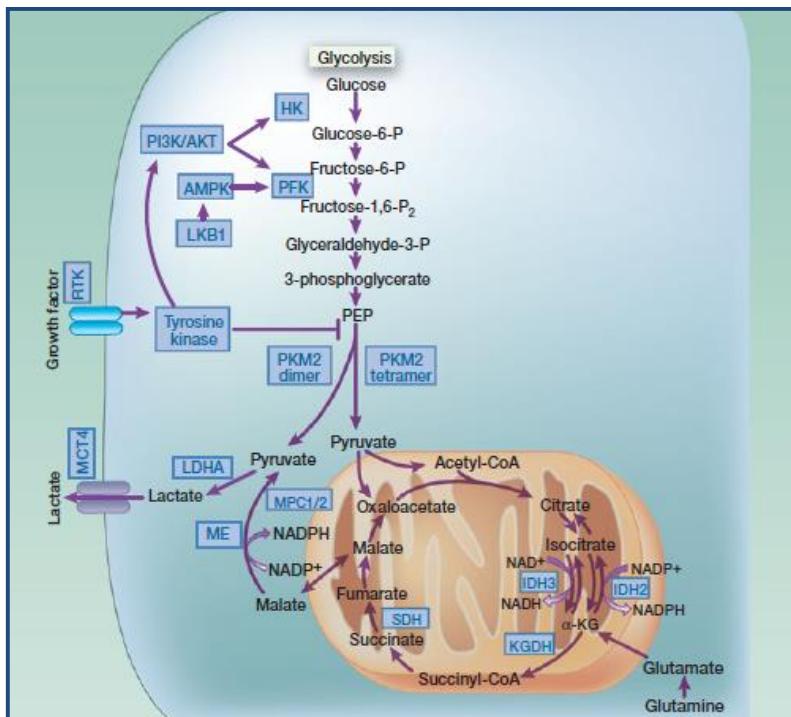


- “ 24 reactions were predicted to be synthetic lethal with Fh1
- “ 18 belong to a linear pathway of haem metabolism
- “ haem metabolism was also predicted by the model to have increased flux in $Fh1^{-/-}$ cells.

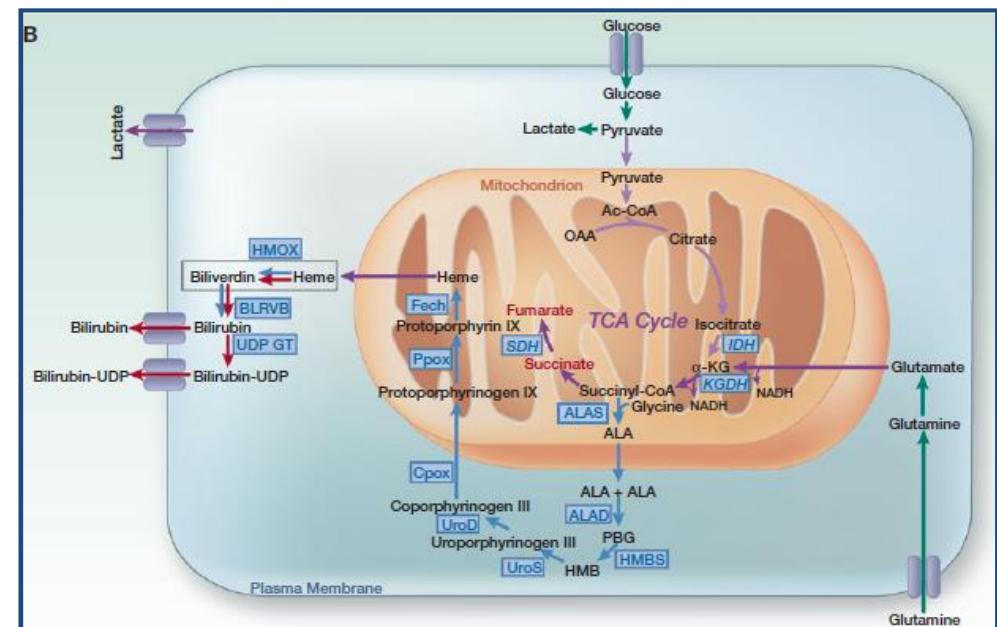
Predict system behaviour: gene deletion analysis

Example: Targeting of synthetic lethal gene for the treatment of HLRCC patients
(Frezza C. et al., Nature, 2011)

The inhibition of the **haem biosynthesis/degradation pathway**, and in particular **Hmox**, is synthetically lethal with Fh1 → valid therapeutic window for the treatment of HLRCC patients.



in wild-type cells



in cancer cells

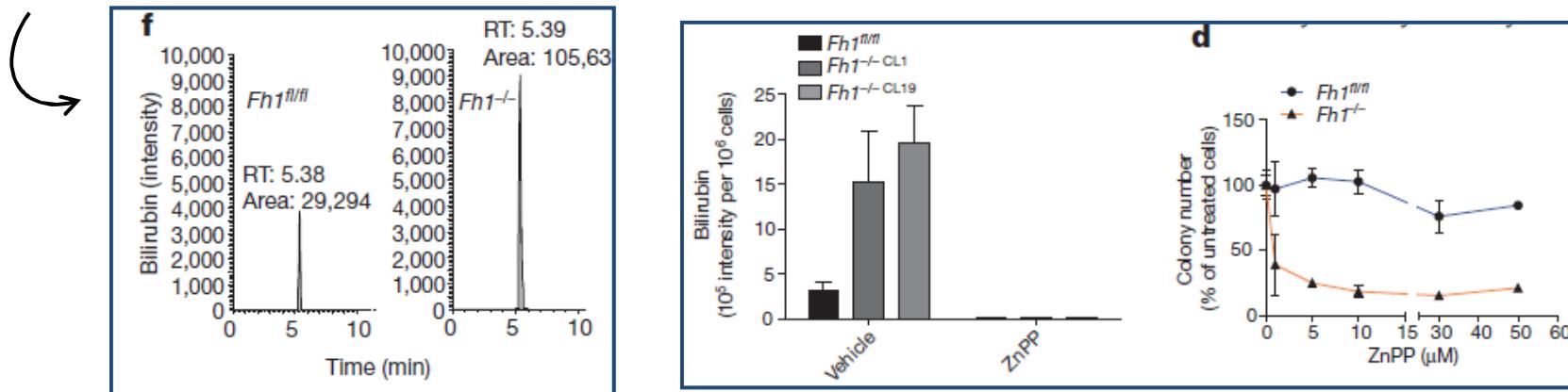


Predict system behaviour: gene deletion analysis

Example: Targeting of synthetic lethal gene for the treatment of HLRCC patients
(Frezza C. et al., Nature, 2011)

Experimental validation for the implication of the Haem metabolic pathway

1. the levels of excreted bilirubin were higher in Fh1-deficient cells.



2. in both wild-type and Fh1-deficient cells bilirubin excretion was completely blocked by ZnPP (**zinc protoporphyrin**), a Hmox inhibitor.

→ acute treatment with ZnPP had no profound effect on wild-type cells BUT it decreased the growth of cancer cells

PREDICTING BIOMARKERS EXAMPLE

Molecular Systems Biology 5; Article number 263; doi:10.1038/msb.2009.22
Citation: *Molecular Systems Biology* 5:263
© 2009 EMBO and Macmillan Publishers Limited All rights reserved 1744-4292/09
www.molecularsystemsbiology.com

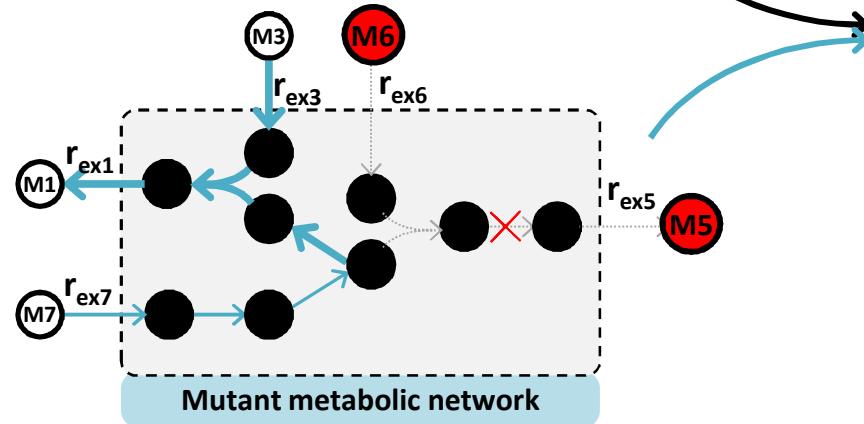
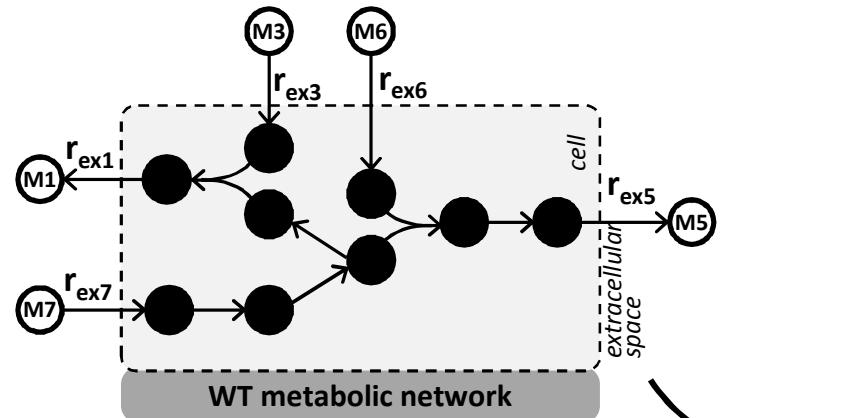


REPORT

Predicting metabolic biomarkers of human inborn errors of metabolism

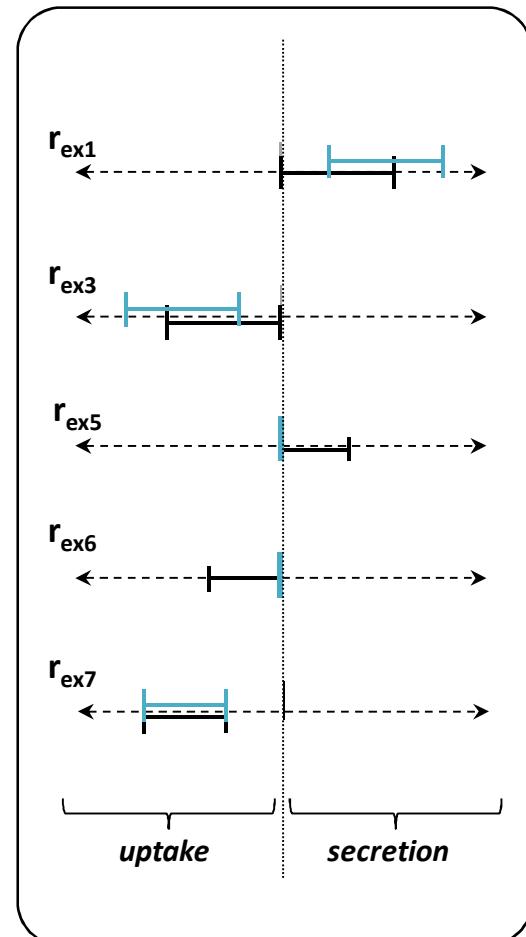
Tomer Shlomi^{1,4,*}, Moran N Cabili^{2,4,*} and Eytan Ruppin^{2,3,*}

Predicting biomarkers



identification of biomarkers
= metabolites with significant changes
in extracellular concentration

Predicted flux ranges
for exchange reactions



Predicting metabolic biomarkers of human inborn errors of metabolism

Shlomi et al. Mol. Syst. Biol. 2009

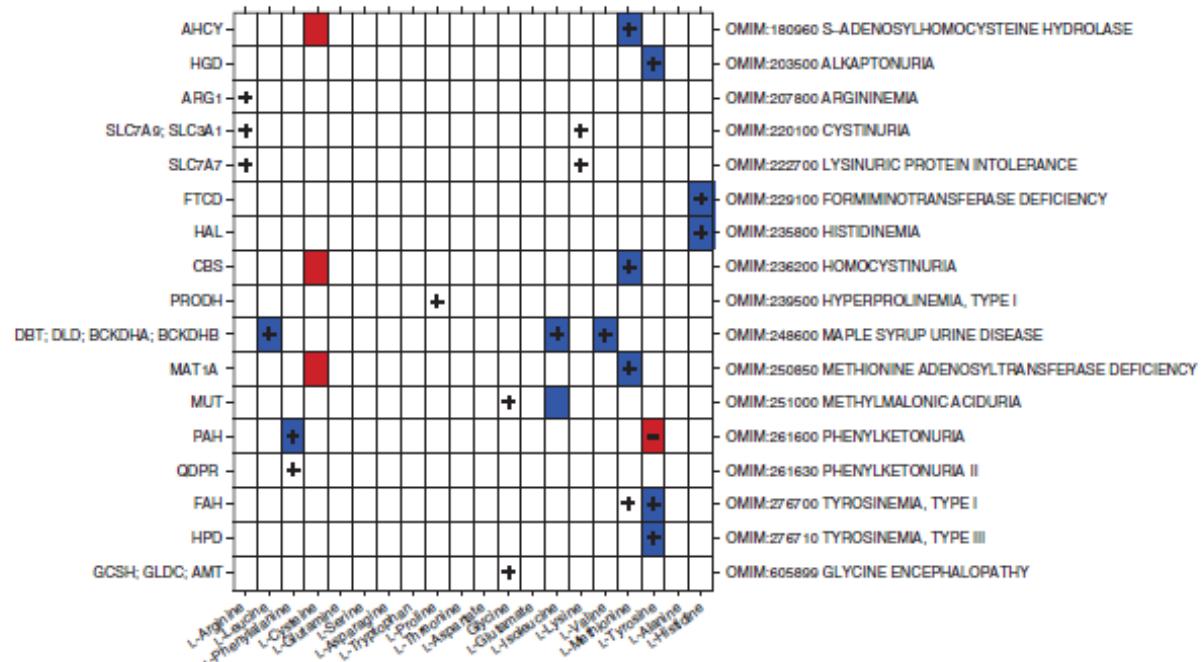
Simulation of enzyme deficiencies in erythrocyte (inborn errors of AA metabolism)



prediction of biomarkers
(metabolites with increased or decreased exchange rate)



comparison with observed changes in metabolite concentrations according to OMIM database



- metabolite predicted to have a reduced concentration
- metabolite predicted to have an increased concentration
- metabolite having a reduced concentration according to OMIM
- + metabolite having a reduced concentration according to OMIM



MetExplore: web server for network analysis of omics data

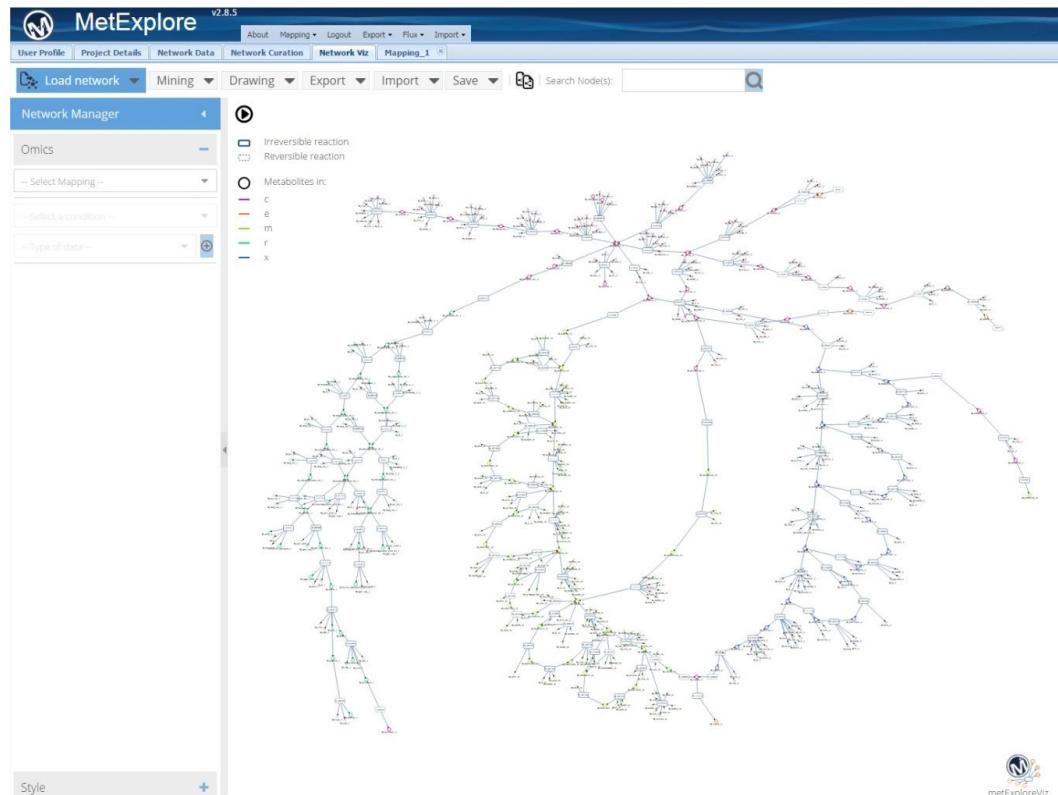
METEXPLORE Computational infrastructure for metabolic network analysis

Funding: ANR MetaboHub, H2020 Phenomenal

“ Long lasting project established in 2009

“ >400 registered users, >350 persons trained, >20 000 visits since 2009

“ Shared platform in international projects



Cottret et al., NAR, 2010

“ Database of metabolic networks

“ Collaborative annotation of metabolic networks

“ Import of omics data

“ Visualization of metabolic networks

“ Sub-network extraction

www.metexplore.fr



MetExplore : omics data analysis in genome scale networks

Nucleic Acids Research

Issues Section browse ▾ Advance articles Submit ▾ Purchase About ▾ All Nucleic Acids Re



Volume 46, Issue W1
2 July 2018

MetExplore: collaborative edition and exploration of metabolic networks

Ludovic Cottret , Clément Frainay, Maxime Chazalviel, Floréal Cabanettes, Yoann Gloaguen, Etienne Camenen, Benjamin Merlet, Stéphanie Heux, Jean-Charles Portais, Nathalie Poupin, Florence Vinson, Fabien Jourdan 

Nucleic Acids Research, Volume 46, Issue W1, 2 July 2018, Pages W495–W502,
<https://doi.org/10.1093/nar/gky301>

Published: 30 April 2018 Article history ▾

The project

” Publications :

” Cottret et al (2018). *Nucleic Acids Research*

” Chazalviel et al (2017). *Bioinformatics*

” Frainay et al (2018). *Bioinformatics*

” Number of citations: >120

” Metrics:

” > 670 registered users,

” > 1300 networks

” > 400 persons trained

” Involved in several national and EU grants

” 1 industrial partner

medDay
PHARMACEUTICALS
• - - - •



Website

<http://www.metexplore.fr/>

Functions

- + Collaborative curation of networks
- + omics data mapping
- + Network visualisation
- + graph algorithms
- + Webservices
- + Expertise and user support



New MetExplore features: having publication ready drawings!

The screenshot shows the MetExplore web application interface. At the top, there's a navigation bar with links for User Profile, Network Data, Network Curation, Network Viz, BioSources, Compartments (1/1), Pathways (88/88), Reactions (1931/1931), Metabolites (1572/1572), Enzymatic Complexes (1455/1455), Gene Products (1455/1455), and Genes (44). The main content area displays a table of metabolic pathways for Homo sapiens, listing their names, identifiers (hsa numbers), and the number of reactions. A specific pathway, "Glycosylphosphatidylinositol (GPI)-anchor biosynthesis - Homo sapiens (human)", is highlighted with a blue background and has a cursor pointing at its identifier. To the right of the table is a sidebar titled "Selected BioSource" which provides details about the public biosource used (Homo sapiens, KEGG Map, Version 24/08/20).

Name	Identifier	Nb Reactions
14 Butanoate metabolism - Homo sapiens (human)	hsa00650	14
15 Caffeine metabolism - Homo sapiens (human)	hsa00232	8
16 Carbon metabolism - Homo sapiens (human)	hsa01200	82
17 Citrate cycle (TCA cycle) - Homo sapiens (human)	hsa00020	23
18 Cysteine and methionine metabolism - Homo sapiens (human)	hsa00270	33
19 D-Arginine and D-ornithine metabolism - Homo sapiens (human)	hsa00472	2
20 D-Glutamine and D-glutamate metabolism - Homo sapiens (human)	hsa00471	4
21 Drug metabolism - cytochrome P450 - Homo sapiens (human)	hsa00982	43
22 Drug metabolism - other enzymes - Homo sapiens (human)	hsa00983	28
23 Ether lipid metabolism - Homo sapiens (human)	hsa00565	19
24 Fatty acid biosynthesis - Homo sapiens (human)	hsa00061	38
25 Fatty acid degradation - Homo sapiens (human)	hsa00071	36
26 Fatty acid elongation - Homo sapiens (human)	hsa00062	34
27 Fatty acid metabolism - Homo sapiens (human)	hsa01212	98
28 Folate biosynthesis - Homo sapiens (human)	hsa00790	27
29 Fructose and mannose metabolism - Homo sapiens (human)	hsa00051	20
30 Galactose metabolism - Homo sapiens (human)	hsa00052	25
31 Glutathione metabolism - Homo sapiens (human)	hsa00480	26
32 Glycerolipid metabolism - Homo sapiens (human)	hsa00561	19
33 Glycerophospholipid metabolism - Homo sapiens (human)	hsa00564	49
34 Glycine, serine and threonine metabolism - Homo sapiens (human)	hsa00260	34
35 Glycolysis / Gluconeogenesis - Homo sapiens (human)	hsa00010	34
36 Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate - Homo sapiens (human)	hsa00532	8
37 Glycosaminoglycan biosynthesis - heparan sulfate / heparin - Homo sapiens (human)	hsa00534	3
38 Glycosaminoglycan degradation - Homo sapiens (human)	hsa00531	20
39 Glycosphingolipid biosynthesis - ganglio series - Homo sapiens (human)	hsa00604	22
40 Glycosphingolipid biosynthesis - globo and isogloblo series - Homo sapiens (human)	hsa00603	16
41 Glycosphingolipid biosynthesis - lacto and neolacto series - Homo sapiens (human)	hsa00601	44
42 Glycosylphosphatidylinositol (GPI)-anchor biosynthesis - Homo sapiens (human)	hsa00563	9
43 Glyoxylate and dicarboxylate metabolism - Homo sapiens (human)	hsa00630	23
44 Histidine metabolism - Homo sapiens (human)	hsa00340	15

Edit your drawing -> save -> overlay different data

Efforts toward visualisation

MetExplore v2.18.19

User Profile Network Data Network Curation Network Viz Mapping

Load network Mining Drawing Export Import Save Exit edit mode Search Node(s)

Network Manager

Omics Pathways

Highlight pathways

Highlight pathways on links

Select/Unselect all

Fatty acid oxidation : Citric acid cycle : Glycerophospholipid metabolism : Aminosugar metabolism : Valine, leucine, and isoleucine metabolism : Alanine and aspartate metabolism : Pyruvate metabolism : Glycine, serine, alanine and threonine metabolism : Unassigned : Methionine and cysteine metabolism : Bile acid synthesis : Phenylalanine metabolism : Glyoxylate and dicarboxylate metabolism :

Compartments

Drawing parameters

Cycle

Reaction Reversible reaction Metabolites

The diagram illustrates a complex metabolic network with numerous nodes (metabolites) and edges (reactions). Key nodes include L-glutamine, L-aspartate(1-), L-aspartate(2-), L-glutamate(1-), 2-oxoglutarate(2-), N-Carbamoyl-L-aspartate, 3-methyl-2-oxopentanoate, Ubiquinone-10, Ubiquinol-10, tiglyl carnitine, propionyl-carnitine, isovaleryl-carnitine, 3-hydroxy-isovaleryl carnitine, L-Carnitine, butyryl carnitine, adipoyl carnitine, 3-hydroxyisovaleryl-CoA, AMP, adipoic acid, succinyl-CoA, glycine, 5-Amino-4-oxopentanoate, D-glycerate, choholyl-CoA(4-), glycocholate, Porphobilinogen, N-acetylneuraminate, N-acetyl-D-glucosamine, (S)-malate(2-), N-acetyl-D-glucosamine 6-phosphate(2-), D-alucosamine 6-phosphate(2-), L-alanine, L-serine, pyruvate, glyoxylate, 3-hydroxypyruvate, 2-oxaloacetate, acetate, citrate, cis-aconitate(3-), N-Acetyl-L-alanine, N-Acetyl-L-methionine, acetylcholine, choline, Butanoyl-CoA, 2-methylbutanoyl-CoA(4-), 2-methylcrotonyl-CoA(4-), Propanoyl-CoA, isovaleroyl-CoA, and phenylacetyl-CoA(4-).



Conclusions & future directions

Conclusions

- “ Genome scale metabolic networks provide a **good context for combined analysis of transcriptomics and metabolomics data.**
- “ **Need to go beyond the description of metabolic pathways** for interpretations
- “ **New information about the cell-specific activity state of reactions** can be gained by using the stoichiometry and the topology of the metabolic network.
- “ **The quality of predictions is highly dependent on the quality of the initial genome-scale metabolic network reconstruction**

Future Directions

- “ **Integration of fluxomic data** will help further constraining the model and getting more accurate predictions for non central metabolism reactions
- “ **Decreasing computation time** to allow large scale phenotyping

“The most exciting phrase to hear in science, the one that heralds new discoveries, is not «Eureka !» but «That’s funny.»”

Isaac Asimov

Bioinformatics: science of recommendation

