Metabolomics for nutrition and toxicology, *in vivo*, *in vitro* and *in silico* studies. Overview of the French metabolomics community

Fabien JOURDAN

INRA Toulouse

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

Outline

- Metabolomics to study food contaminants
- Identifying biomarkers with high resolution mass spectrometry
- Metabolic network analysis
- Metabolic network reconstruction for cell lines
- Metabolomics in France

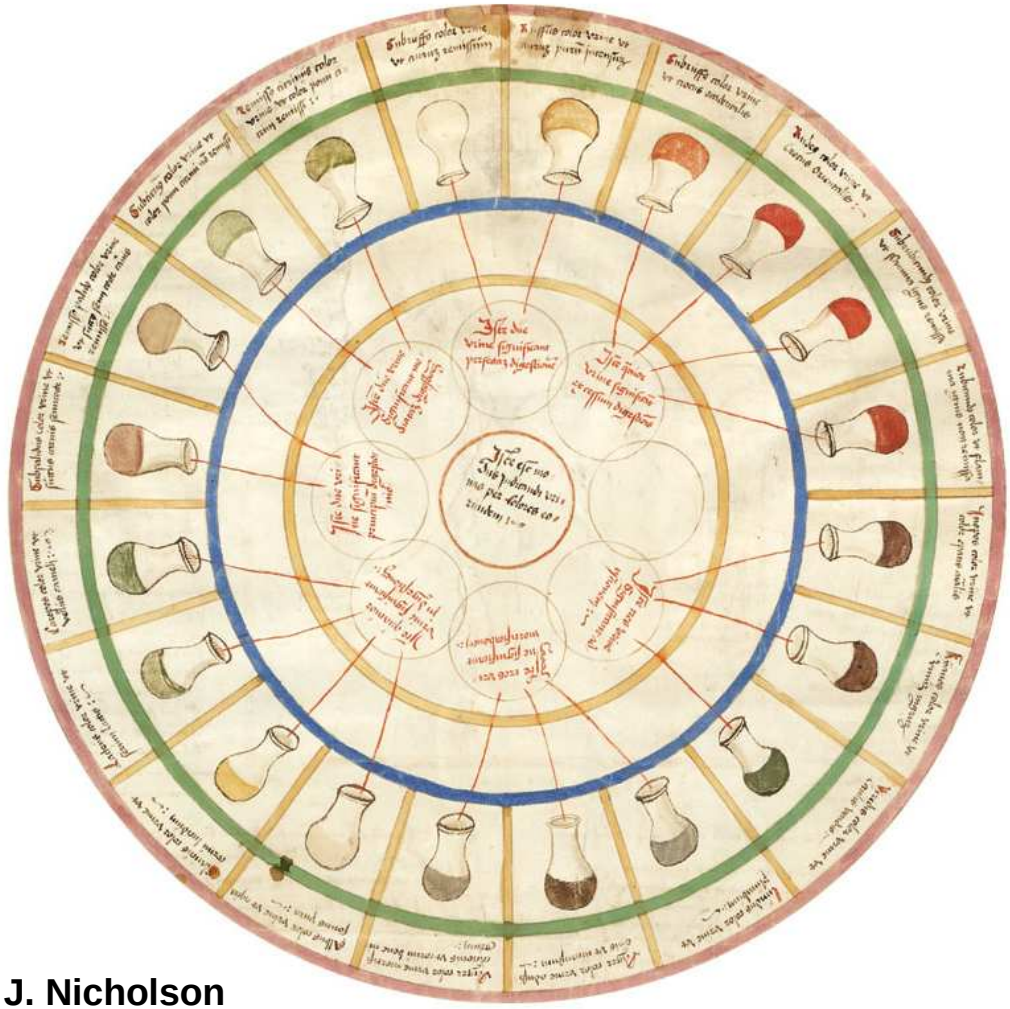



Metabolomics to study food contaminants

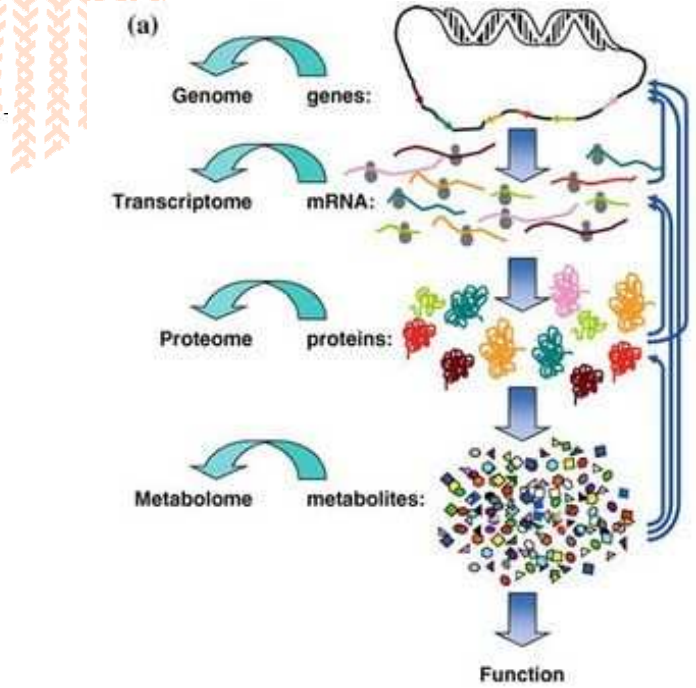



ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

From genes to metabolites



J. Nicholson



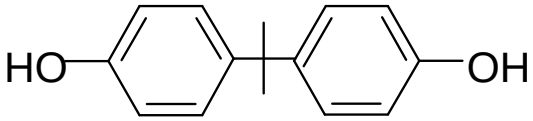
Using the metabolome shift as a phenotypic marker

Food contaminants

MeX team (D Zalko),
INRA Toulouse

Develop novel and efficient strategies to address modern toxicological challenges linked with food contaminants

- Chronic & low dose exposures issues
- Exposure during critical periods of the development



Bisphenol A
Prod. *ca.* 3 Million T/year

Polycarbonates

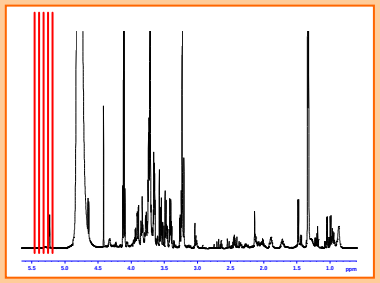
Epoxy Resins

Free monomer = model Xeno-estrogen

Effects / human health ?

Fingerprinting

Data acquisition



Spectrum bucketing

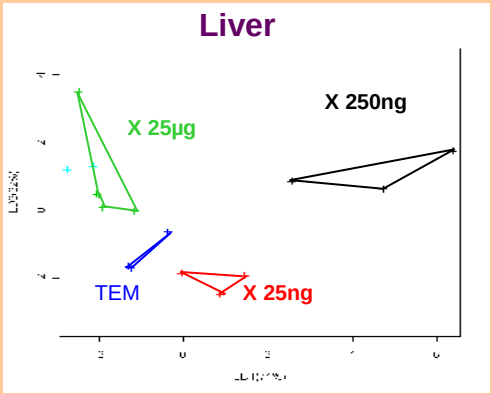
Numerical data

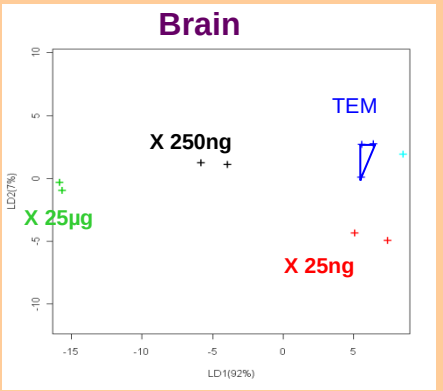
Variable 1 v1,v2…

Variable 2 v1,v2…

…

Multivariate statistics





Discrimiation of adult mice exposed *in utero* to low doses of xenobiotics

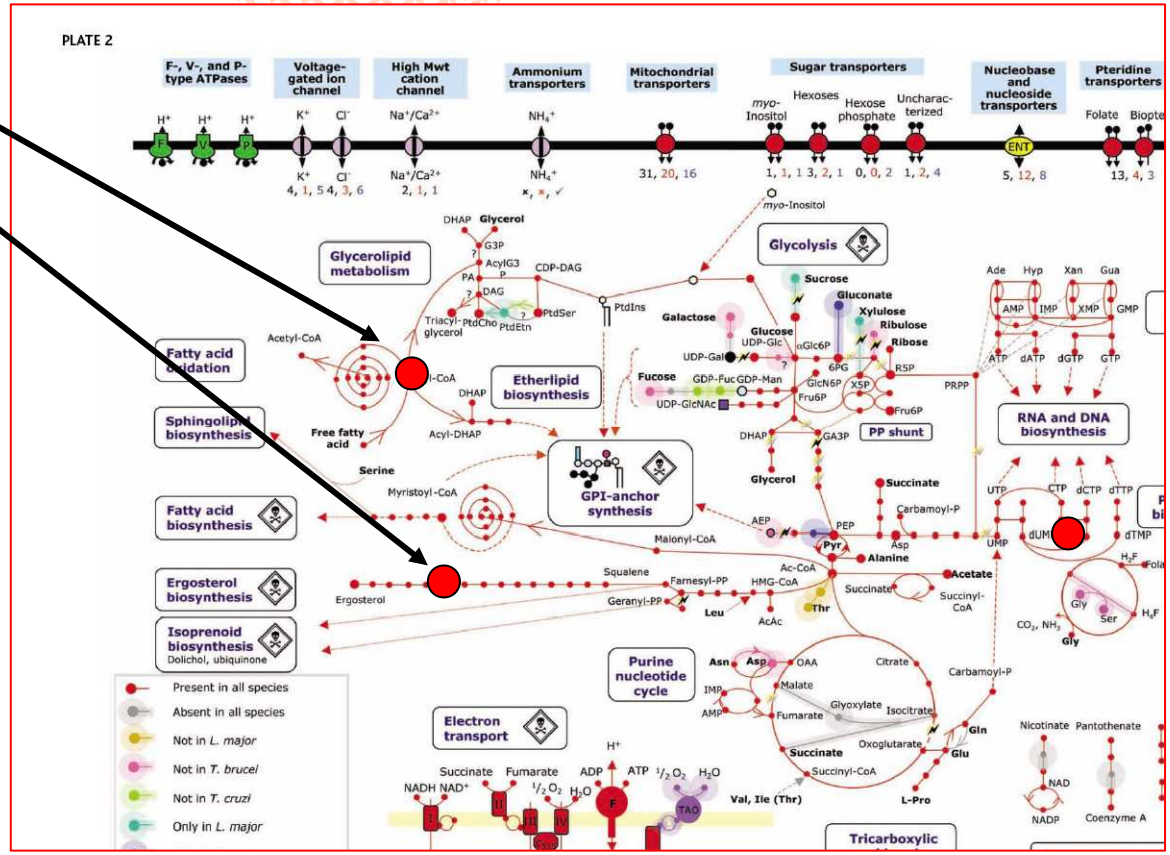
Biomarkers

hippurate/tryptophane
valine
lactate
lysine
cétoglutarate/succinate
glucose
TMAO/phénylalanine

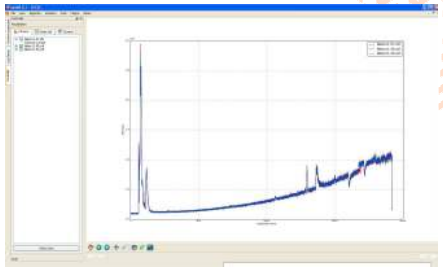
From Biomarkers to System Biology

hippurate/tryptophane
valine
lactate
lysine
cétoglutarate/succinate
glucose
TMAO/phénylalanine

Given a list of biomarkers…



…find the processes involved

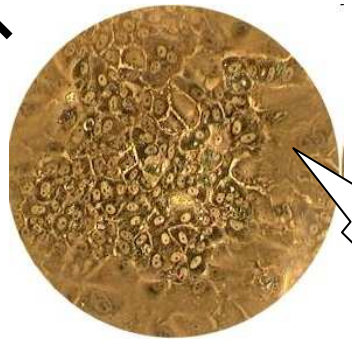


MS raw data

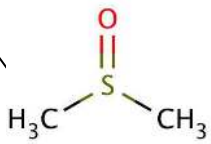
Compound Name	Identifier	Mass	Formula	Pathways	sample
(R)-3-hydroxybutanoate	CPD-835	104.047862	C4H8O3	0 pathway	
3-hydroxy-isobutyrate	3-HYDROXY-ISOBUTYRATE	104.047862	C4H8O3	1 pathway Display pathways	
L-isoleucine	ILE	131.094805	C6H13NO2	7 pathways Display pathways	
L-leucine	LEU	131.094805	C6H13NO2	4 pathways Display pathways	
S-adenosyl-L-homocysteine	ADENOSYL-HOMO-CYS	384.120629	C14H20N6O5S1	10 pathways Display pathways	
8-hydroxyquinoline	CPD-9017	136.03985	C9H7NO1	0 pathway	

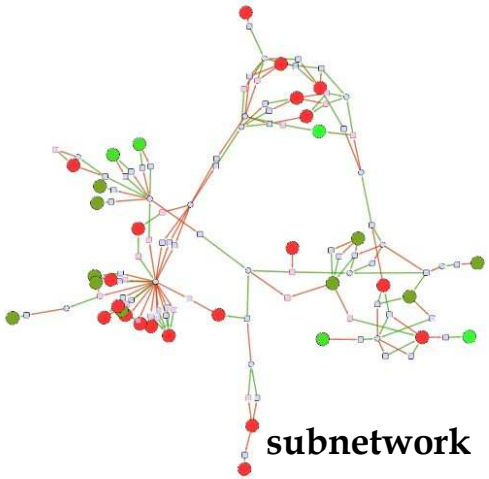
Metabolite list

LC-HRMS

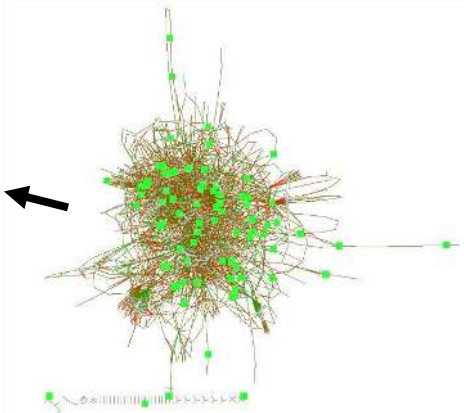


HepaRG





subnetwork



Data in human network

Identifying biomarkers with high resolution mass spectrometry




High Resolution Mass Spectrometry

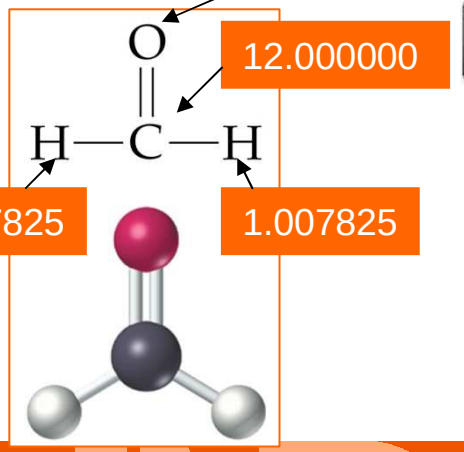
NMR

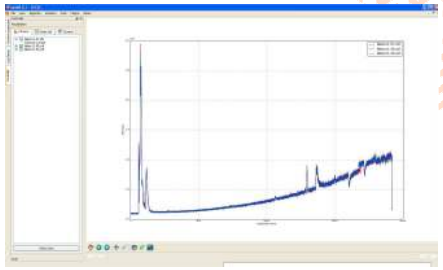
Mass spectrometry
- LC-MS
- GC-MS
- CE- MS

High resolution MS :
- FT ICR
- Orbitrap

30

30.049160	30.006100	30.065484	30.010565
CH4N	NO	C2H4	CH2O





MS raw data

Compound Name	Identifier	Mass	Formula	Pathways	sample
(R)-3-hydroxybutanoate	CPD-836	104.047862	C4H8O3	0 pathway	
3-hydroxy-isobutyrate	3-HYDROXY-ISOBUTYRATE	104.047862	C4H8O3	1 pathway Display pathways	
L-isoleucine	ILE	131.094805	C6H13NO2	7 pathways Display pathways	
L-leucine	LEU	131.094805	C6H13NO2	4 pathways Display pathways	
S-adenosyl-L-homocysteine	ADENOSYL-HOMO-CYS	384.120629	C14H20N6O5S1	10 pathways Display pathways	
8-hydroxyquinoline	CPD-9017	136.03805	C9H7NO	0 pathway	

Metabolite list

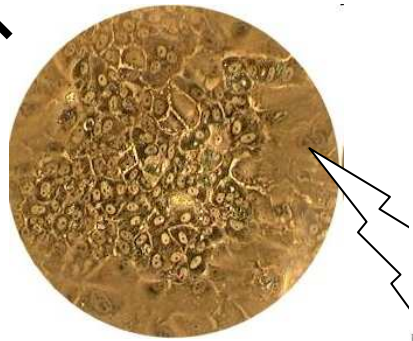
LC-HRMS

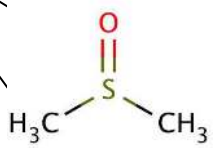
AXIOM

genotoul metatoul

Going from raw data to a metabolite list

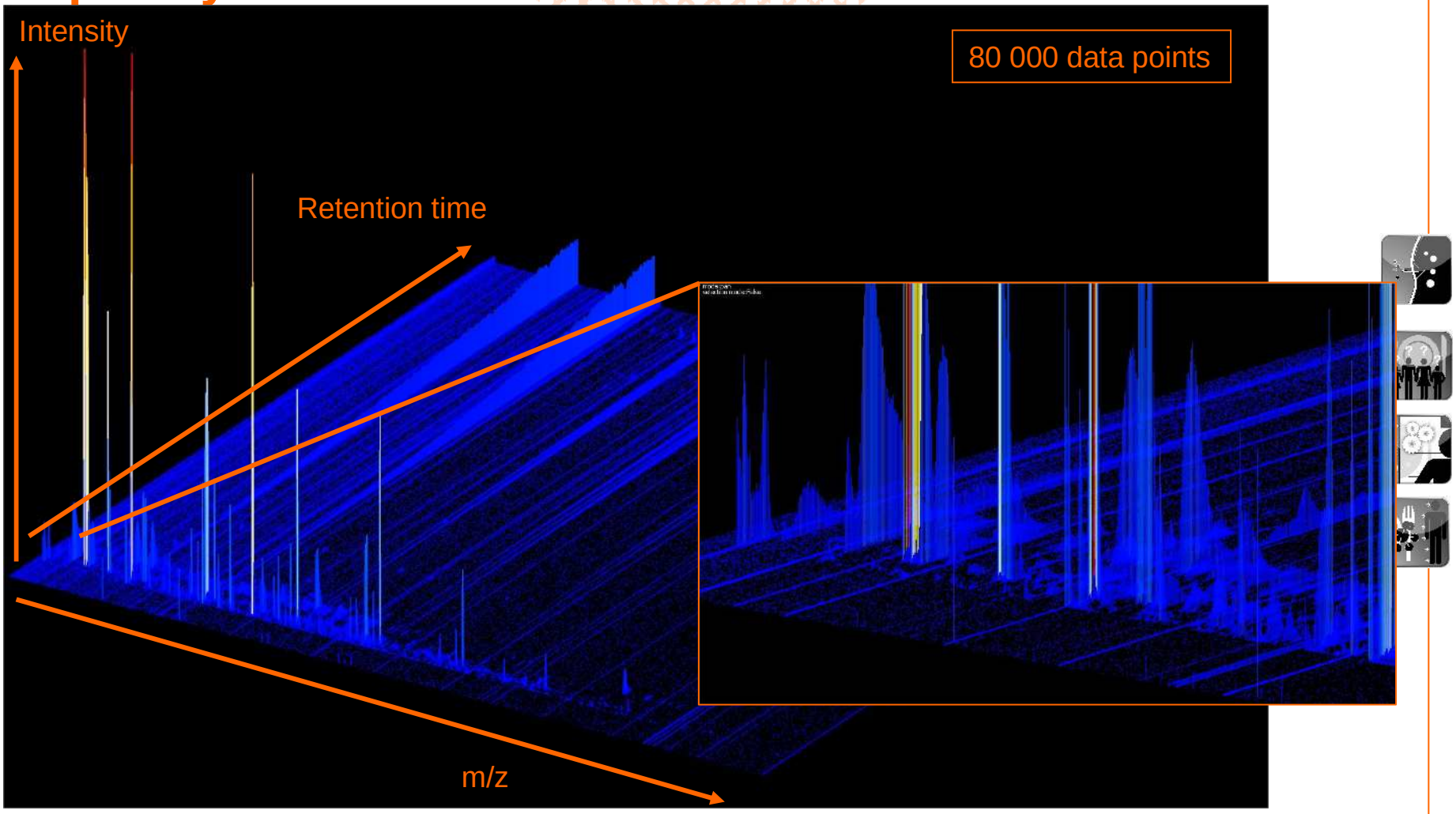
Challenge : tenth of thousands of peaks among which only few hundreds are relevant



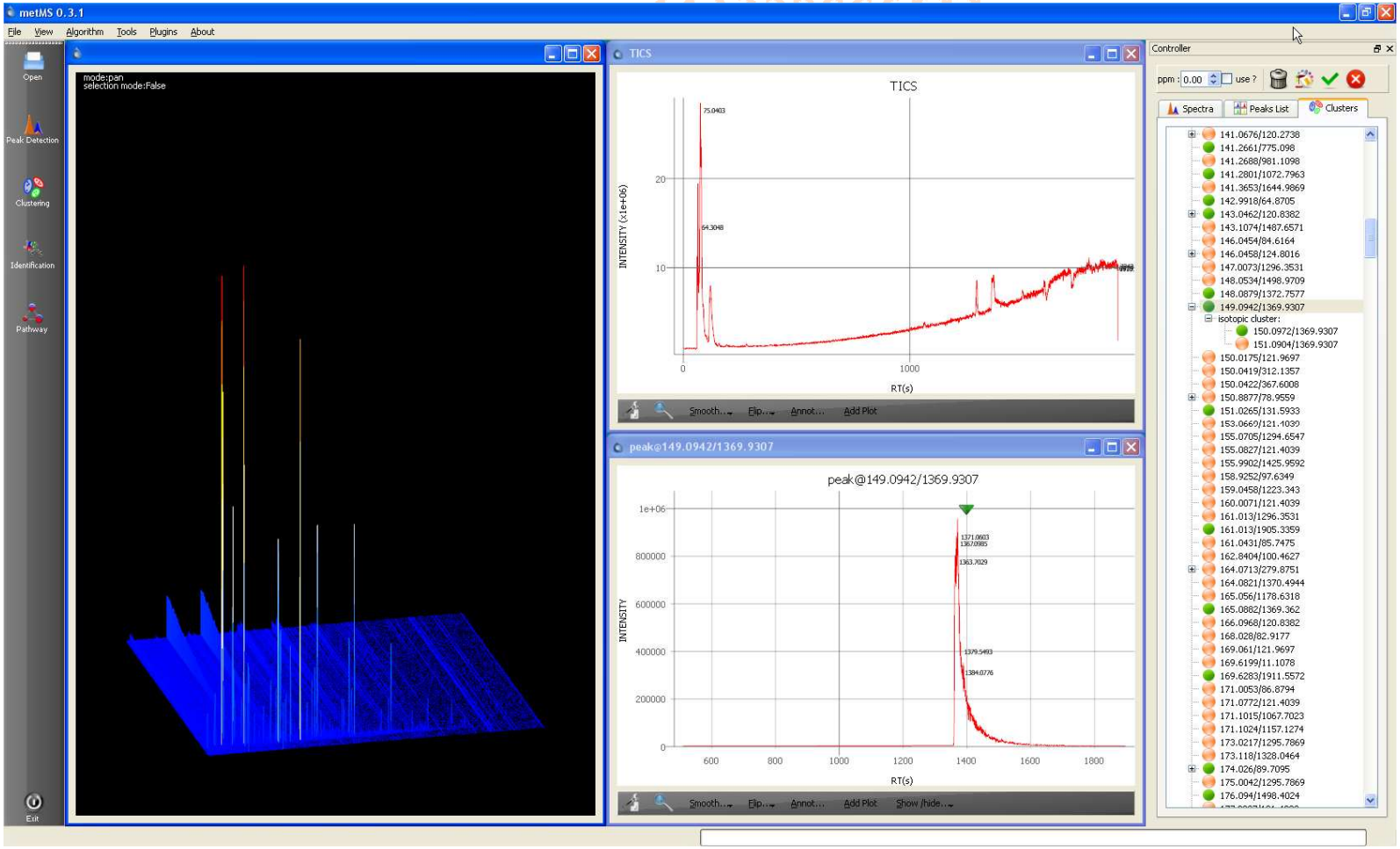


HepaRG

Complexity of LC-HRMS data



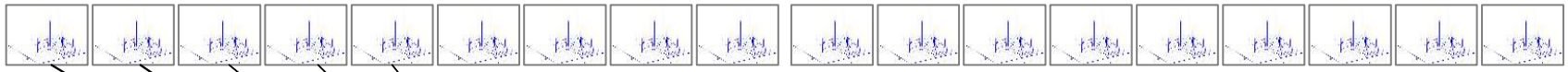
MetMS: implementation of LC-HRMS data treatment






Marc Dubois

Computational Challenge



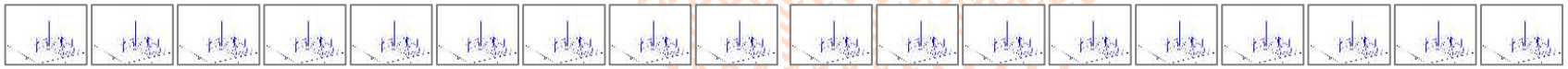
18 samples for a condition

Scripps Center For Metabolomics
METLIN: Metabolite and Tandem MS Database

XCMS

	A	B
1	**exact mass**	**retention time**
2	307.063427809286	1014.87168102335
3	403.138561743954	994.002293383756
4	447.068375627694	171.512766016008
5	299.005338547854	713.542363217923
6	305.022719248425	1643.99557163300
7	283.027781167477	1359.31933067860
8	396.105844158039	1194.21002632329
9	217.002878600446	1450.42819933016
10	201.024981090449	1532.99014233025
11	283.027663292645	531.752199938246
12	171.027015142314	1289.31181975929
13	283.027770816323	1177.84505625444
14	283.027830387926	1443.68769151619
15	308.086501554988	1193.79131977766
16	171.026887987536	158.235226597383
17	60.0172695993117	1540.27482811595
18	299.005298963986	650.500998390242
19	305.022469730536	1087.42609034510
20	288.075700357267	107.424259304909
21	463.011248047309	1241.38126586809
22	463.01118039837	1116.92830294752

Applying treatments on each sample…but it can take more than a day !!



18 samples for a condition

Computation is performed on a computer cluster and takes less than an hour

	A	B
1	**exact mass**	**retention time**
2	307.063427809286	1014.87168102335
3	403.138561743954	994.002293383756
4	447.068375627694	171.512766016008
5	299.005338547854	713.542363217923
6	305.022719248425	1643.99557163300
7	283.027781167477	1359.31933067860
8	396.105844158039	1194.21002632329
9	217.002878600446	1450.42819933016
10	201.024981090449	1532.99014233025
11	283.027663292645	531.752199938246
12	171.027015142314	1289.31181975929
13	283.027770816323	1177.84505625444
14	283.027830387926	1443.68769151619
15	308.086501554988	1193.79131977766
16	171.026887987536	158.235226597383
17	60.0172695993117	1540.27482811595
18	299.005298963986	650.500998390242
19	305.022469730536	1087.42609034510
20	288.075700357267	107.424259304909
21	463.011248047309	1241.38126586809
22	463.011180398...	1116.928309475...

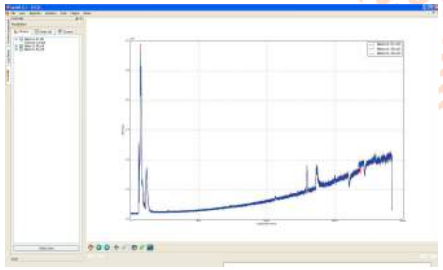
Data treatment is a challenge in HRMS metabolomics

Output of raw data treatment

Compound Name	Identifier	Mass	Formula	Pathways	sample
(R)-3-hydroxybutanoate	CPD-335	104.047362	C4H8O3	0 pathway	1
3-hydroxy-isobutyrate	3-HYDROXY-ISOBUTYRATE	104.047362	C4H8O3	1 pathway Display pathways	1
L-isoleucine	ILE	131.094606	C6H13N1O2	7 pathways Display pathways	1
L-leucine	LEU	131.094606	C6H13N1O2	4 pathways Display pathways	1
S-adenosyl-L-homocysteine	ADENOSYL-HOMO-CYS	384.120629	C14H20N6O5S1	10 pathways Display pathways	1
8-hydroxypurine	CPD-9017	136.03805	C5H4N4O1	0 pathway	1

Metabolic Network Analysis



MS raw data

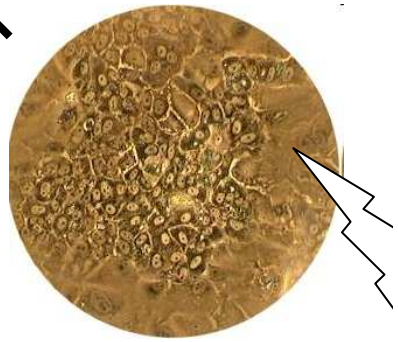
Compound Name	Identifier	Mass	Formula	Pathways	sample
(R)-3-hydroxybutanoate	CPD-836	104.047862	C4H8O3	0 pathway	
3-hydroxy-isobutyrate	3-HYDROXY-ISOBUTYRATE	104.047862	C4H8O3	1 pathway Display pathways	
L-isoleucine	ILE	131.094805	C6H13NO2	7 pathways Display pathways	
L-leucine	LEU	131.094805	C6H13NO2	4 pathways Display pathways	
S-adenosyl-L-homocysteine	ADENOSYL-HOMO-CYS	384.120629	C14H20N6O5S1	10 pathways Display pathways	
8-hydroxyquinoline	CPD-9017	136.03985	C9H7NO1	0 pathway	

Metabolite list

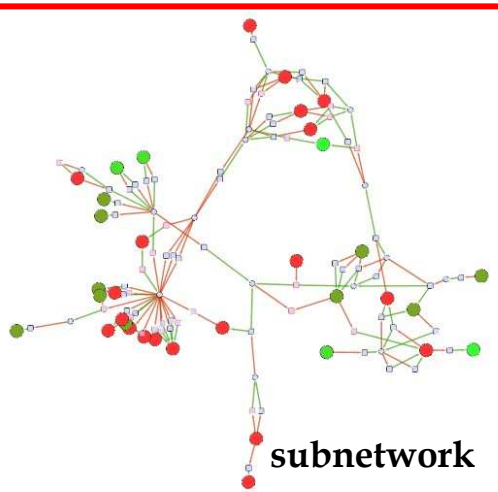
LC-HRMS

AXIOM

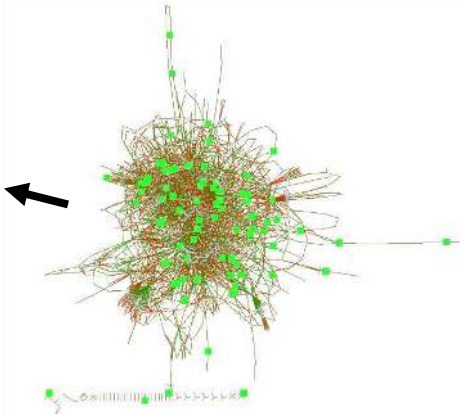
geno toul metatoul



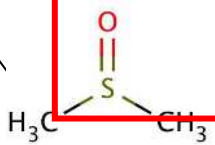
HepaRG



subnetwork

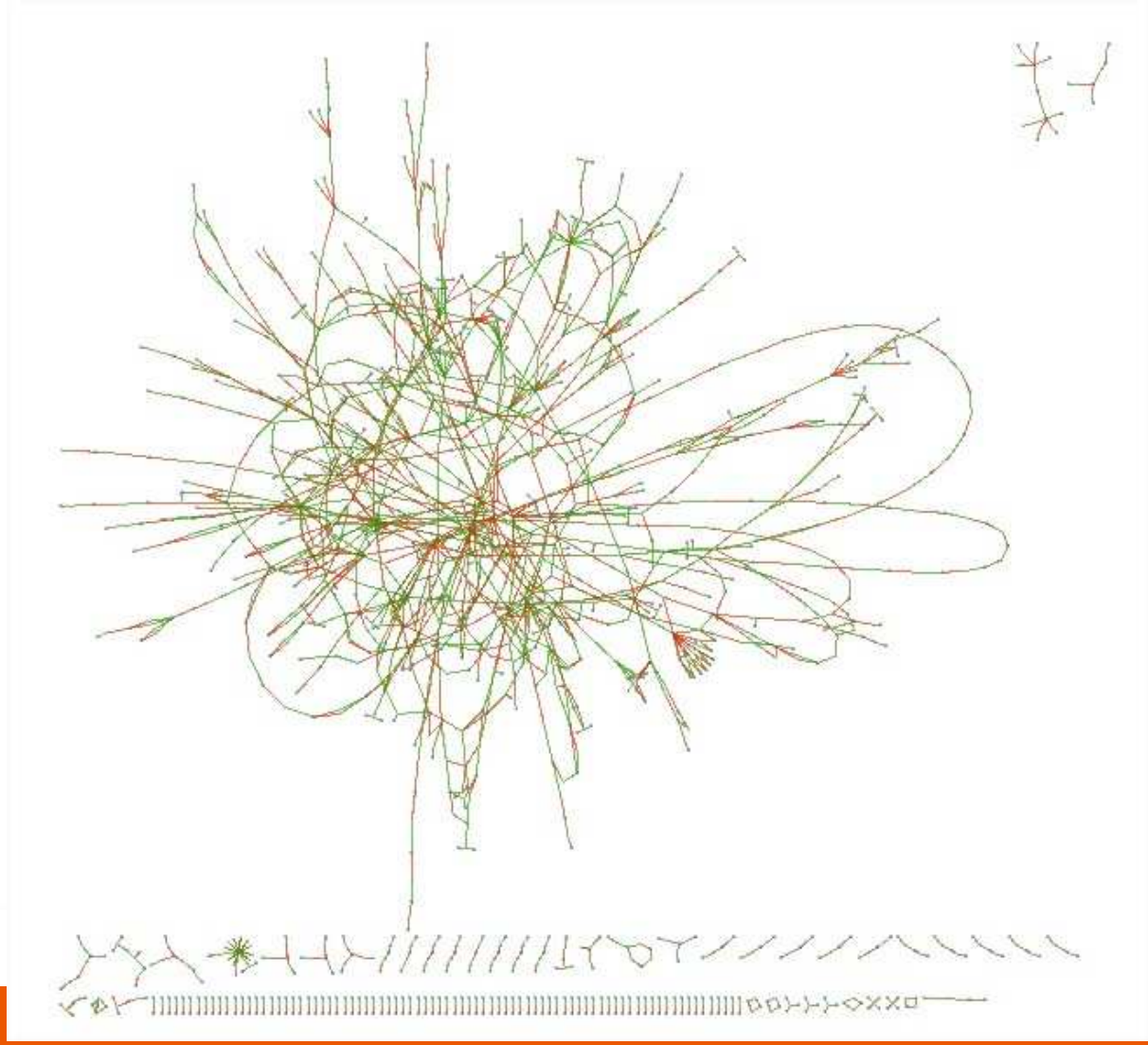


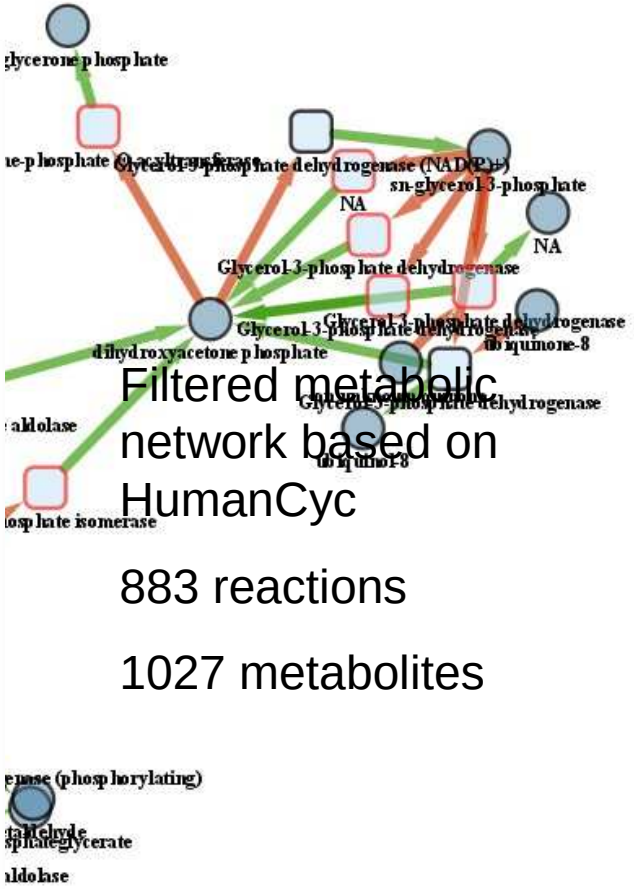
Data in human network



ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA

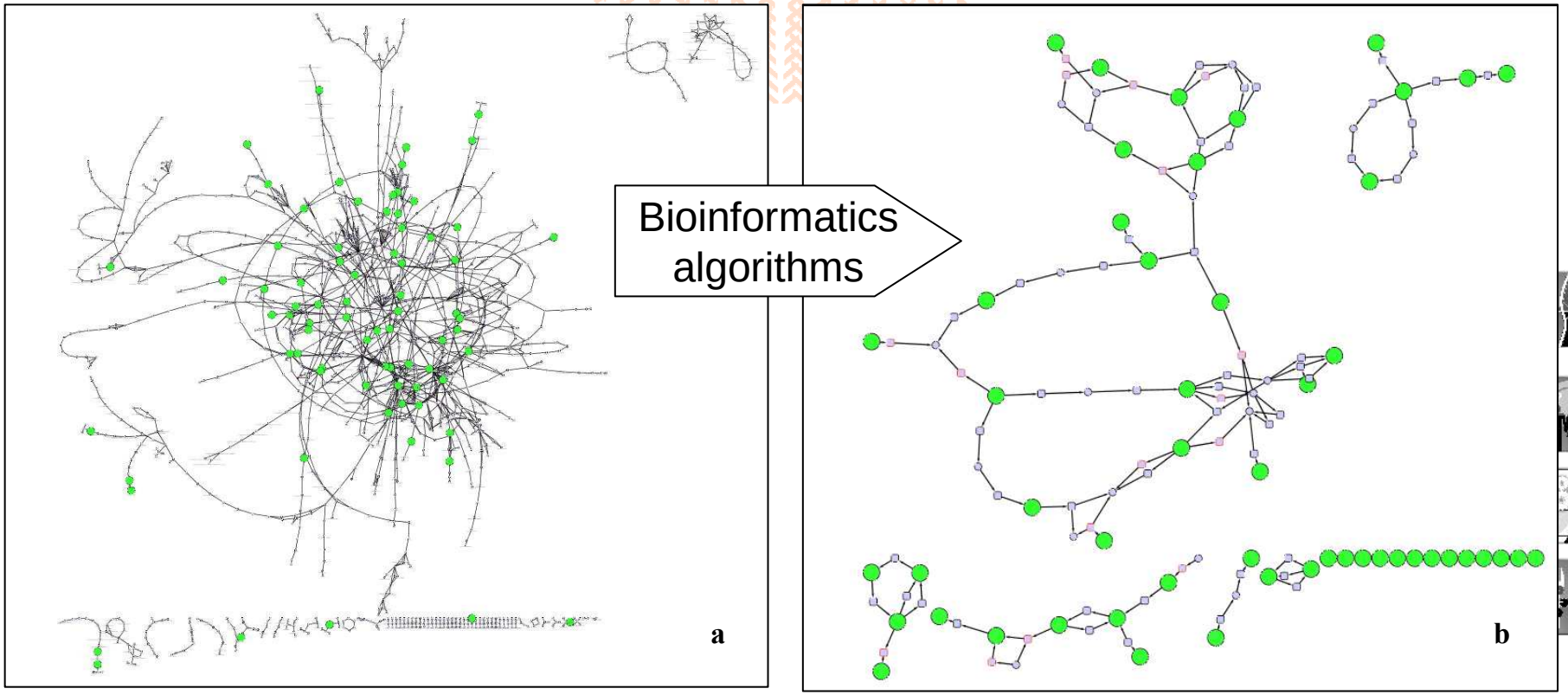




Filtered metabolic network based on HumanCyc

883 reactions

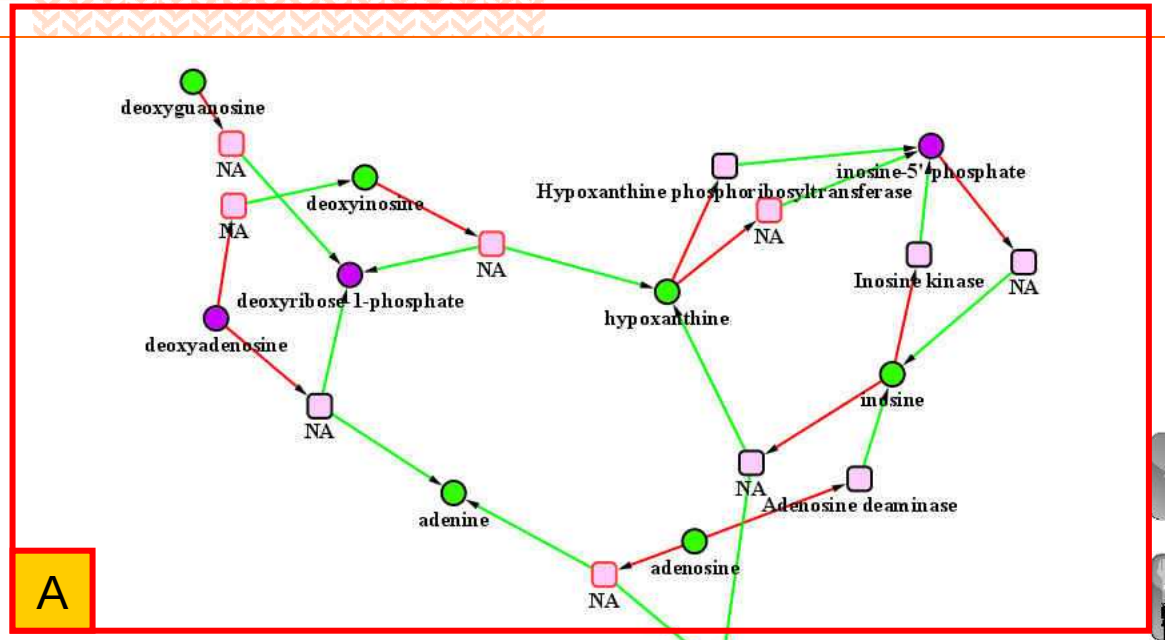
1027 metabolites



Aim : given a set of metabolites, extracting a “sparser” sub-network containing all the identified compounds.

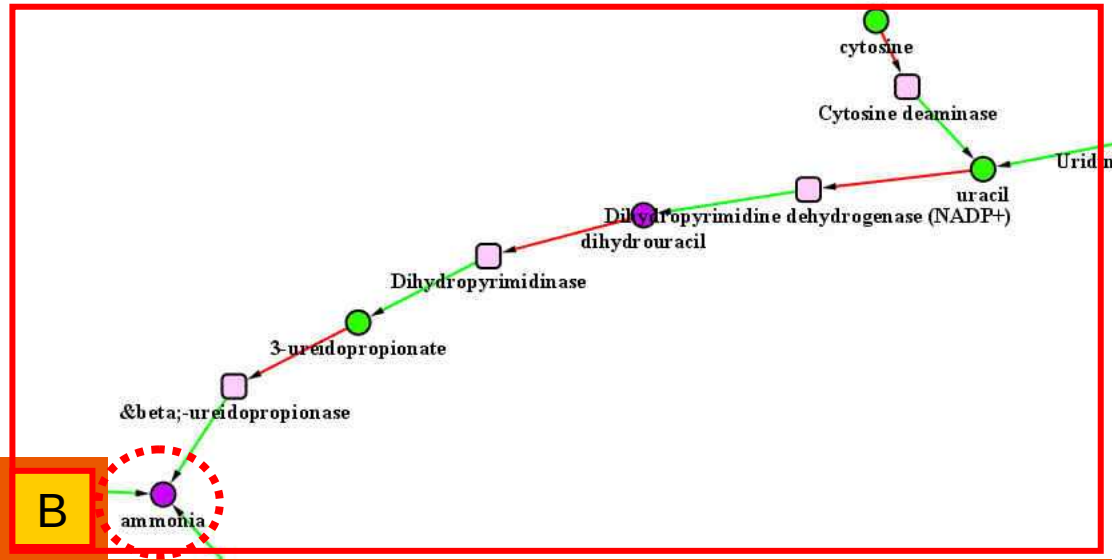
Not all the metabolites are identified, it requires to fill the gaps

Jourdan *et al.* *Metabolomics*, **2010**, *6*, 312-321



Purine Pathway

ribose-1-phosphate



Pyrimidine Pathway

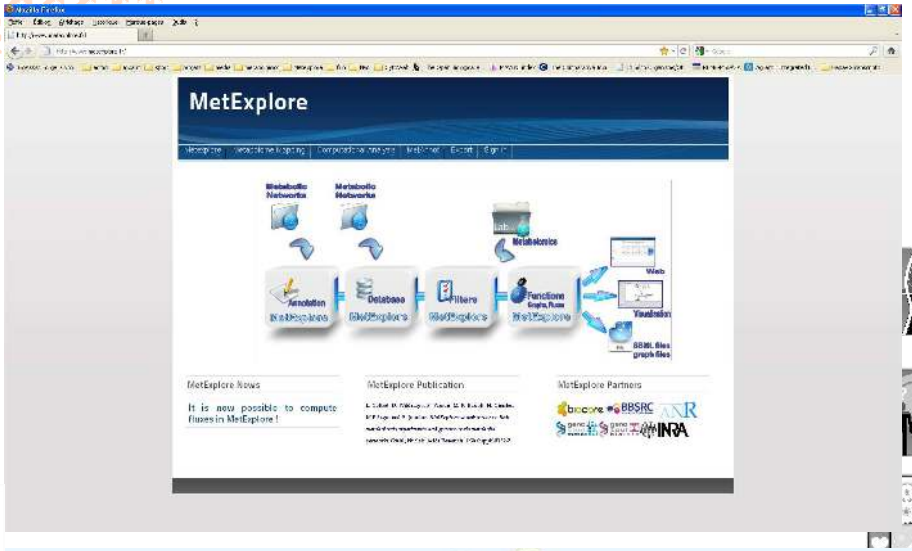
Overview MetExplore

URL: www.metexplore.fr

Registered users:52

Metabolic networks:174

Reactions:19927



W132–W137 *Nucleic Acids Research, 2010, Vol. 38, Web Server issue* *Published online 5 May 2010*
doi:10.1093/nar/gkq312

MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks

Ludovic Cottret[1,*], David Wildridge[2], Florence Vinson[1], Michael P. Barrett[2], Hubert Charles[3,4], Marie-France Sagot[3,5] and Fabien Jourdan[1]

[1]INRA, UMR1089, Xénobiotiques, F-31000 Toulouse, France, [2]Division of Infection and Immunity, Glasgow Biomedical Research Centre, University of Glasgow, Glasgow, UK, [3]Bamboo Team, INRIA Grenoble-Rhône-Alpes, 38330 Montbonnot Saint-Martin, [4]UMR203 Biologie Fonctionnelle Insectes et Interactions (BF2I), INRA, INSA-Lyon, Université de Lyon, F-69621 Villeurbanne and [5]Université de Lyon, F-69000, Lyon; Université Lyon 1; CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622, Villeurbanne, France

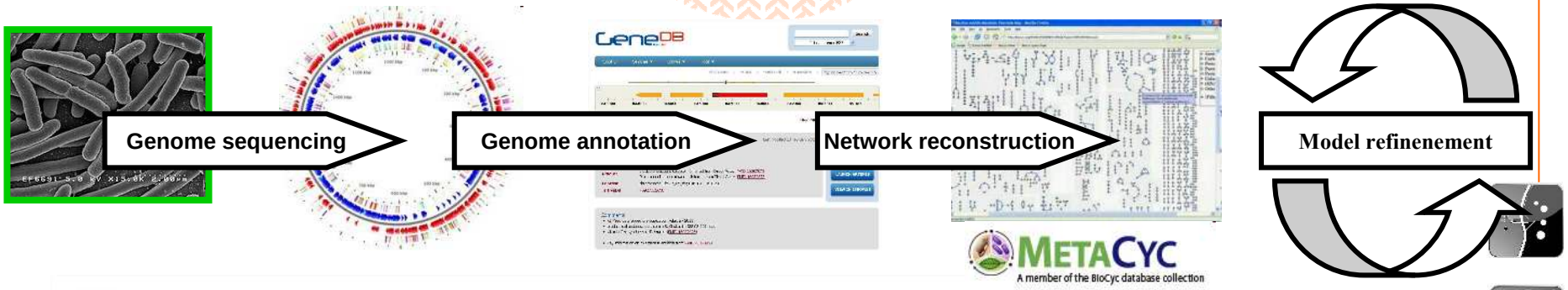
Received January 29, 2010; Revised March 30, 2010; Accepted April 17, 2010

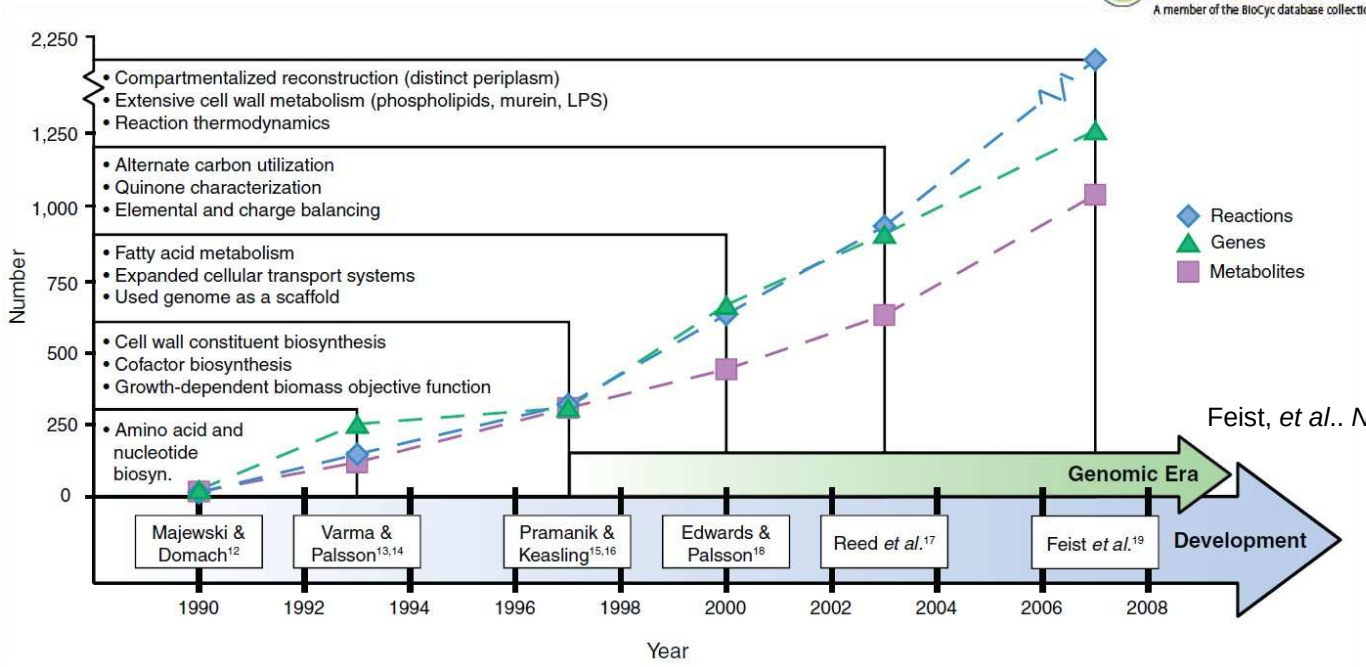
2,437 visits came from 304 cities

Metabolic network reconstruction for cell lines

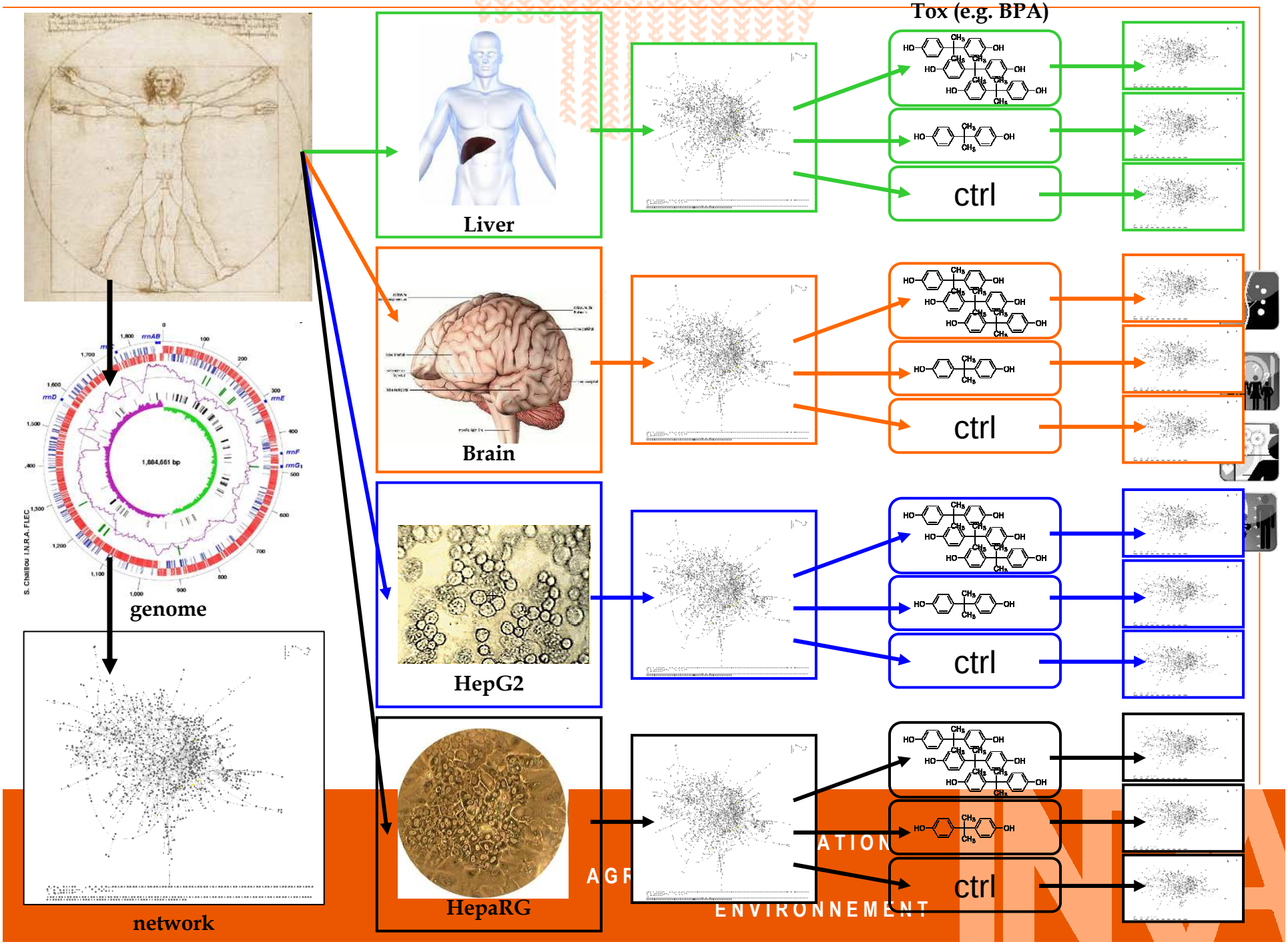



From genome to metabolic networks





Feist, *et al.*. *Nat Biotechnol,* **2008**, *26*, 659-667


Tox (e.g. BPA)
Liver
Brain
HepG2
HepaRG
ctrl
ctrl
ctrl
ctrl
genome
network
ENVIRONNEMENT

Network-based prediction of human tissue-specific metabolism

Tomer Shlomi[1,4], Moran N Cabili[1,4], Markus J Herrgård[2], Bernhard Ø Palsson[2] & Eytan Ruppin[1,3]

Direct *in vivo* investigation of mammalian metabolism is complicated by the distinct metabolic functions of different tissues. We present a computational method that successfully describes the tissue specificity of human metabolism on a large scale. By integrating tissue-specific gene- and protein-expression data with an existing comprehensive reconstruction of the global human metabolic network, we predict tissue-specific metabolic activity in ten human tissues. This reveals a central role for post-transcriptional regulation in shaping tissue-specific metabolic activity profiles. The predicted tissue specificity of genes responsible for metabolic diseases and tissue-specific differences in metabolite exchange with biofluids extend markedly beyond tissue-specific differences manifest in enzyme-expression data, and are validated by large-scale mining of tissue-specificity data. Our results establish a computational basis for the genome-wide study of normal and abnormal human metabolism in a tissue-specific manner.

Metabolic network modeling of biological systems involves the analysis and prediction of metabolic flux distributions under diverse physiological and genetic conditions. Traditional modeling techniques are based on mathematical approaches that require detailed information on kinetics and on enzyme and metabolite concentrations[1,2]. However, a lack of accurate information of kinetic constants and enzyme and metabolite intracellular concentrations limits the current applicability of such methods to small-scale systems. Constraint-based modeling bypasses this hurdle by analyzing the function of large-scale metabolic networks through relying solely on simple physical-chemical constraints[3]. In recent years, constraint-based modeling has been frequently used to successfully predict various phenotypes of microorganisms, such as their growth rates, rates of nutrient uptake, by-product secretion and the lethality of gene knockouts (see ref. 4 for review).

Despite this progress in applying constraint-based modeling to studying the metabolism of microorganisms, large-scale modeling of human metabolism is still in its infancy. Nonetheless, the emergence of metabolic diseases such as diabetes and obesity as major sources of morbidity and mortality[5,6] has stimulated research into human metabolism and its regulation. Metabolic enzymes and their regulators are increasingly considered viable drug targets for these and other conditions[7,8]. However, in reconstructing human metabolic networks, most of the previous work has focused on characterizing distinct metabolic pathways[9,10]. Until recently, reconstructions of large-scale human metabolic networks had been performed only for specific cell types and organelles[11–13]. Although fundamental steps forward, reconstructions of the global human metabolic network based on an extensive evaluation of genomic and bibliomic data (that is, comprehensive assessment of the literature)[14,15] are not tissue specific. In adapting constraint-based modeling methods from the realm of microorganisms to that of multicellular organisms, one encounters two main hurdles. The first is that different tissues have different metabolic objectives that are not well characterized and remain largely unknown. This is in contrast to modeling microorganisms where a simple objective function (such as maximizing the biomass production rate) can be used together with flux balance analysis[4] to predict biologically plausible flux distributions. The second major obstacle is the lack of information on tissue-specific metabolite uptake and secretion, which is essential for employing flux balance analysis.

We present a new constraint-based computational method for systematically predicting human tissue-specific metabolic behavior by integrating a genome-scale metabolic network with tissue-specific gene- and protein-expression data. Changes in gene- and protein-expression levels play a major role in controlling tissue-specific metabolic functions[16–18], and a strong correlation between gene expression and measured[19,20] and predicted[21–24] metabolic fluxes is reported for microorganisms. To account for metabolic flux activity that is not reflected in the expression data (that is, post-transcriptional regulatory effects), we treat tissue-specific variations in enzyme-expression levels not as the final determinants of enzyme activity, but as cues for the likelihood that the enzyme in question supports metabolic flux in its associated reaction(s). Network integration is then used to accumulate these cues into a global, consistent metabolic behavior, which reflects the outcome of putative post-transcriptional regulatory effects. Our method's reliance on enzyme-expression data to infer tissue-specific metabolic flux eliminates the need for a priori knowledge of tissue-specific objective functions and metabolites exchanged by the tissue with biofluids. Instead, the method provides predictions regarding tissue-specific metabolite uptake and secretion.

To examine our method's ability to correctly predict metabolic behavior based on gene-expression data, we first apply it to predicting the metabolic state of the yeast *Saccharomyces cerevisiae* under

[1]School of Computer Science, Tel-Aviv University, Tel-Aviv 69978, Israel. [2]Department of Bioengineering, University of California, San Diego, La Jolla, California 92093-0412, USA. [3]School of Medicine, Tel-Aviv University, Tel-Aviv 69978, Israel. [4]These authors contributed equally to this work. Correspondence should be addressed to T.S. (shlomito@post.tau.ac.il) or E.R. (ruppin@post.tau.ac.il).

Published online 17 August 2008; doi:10.1038/nbt.1487

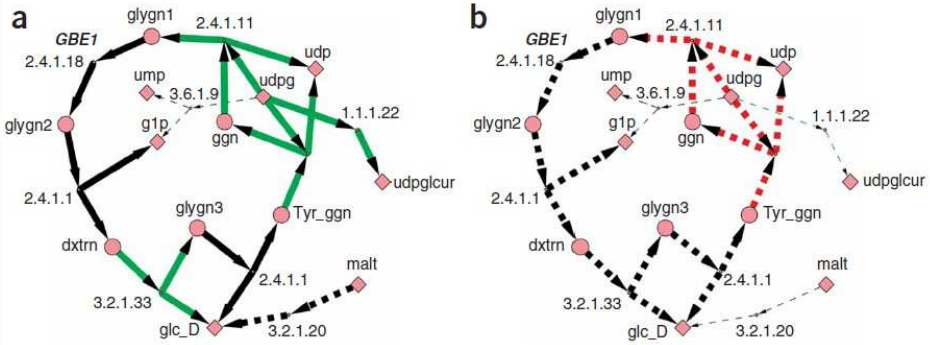
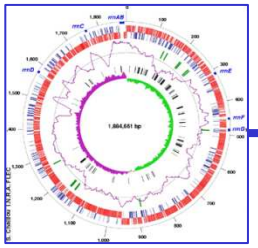


Figure 4 An example of a sub-network of glycogen metabolism. (**a**,**b**) The predicted tissue-specific activity of *GBE1* (1,4-alpha-glucan branching enzyme) in the liver (**a**) and its inactivity in the spleen (**b**) are illustrated. Nodes

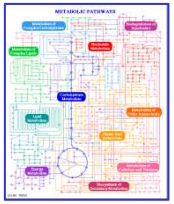
Cell lines annotation pipeline

in silico reconstruction



genome

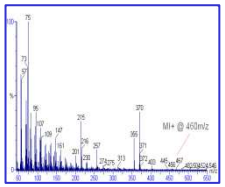
Reconstruction softwares



KEGG

BioCyc Database Collection

Metabolomics data

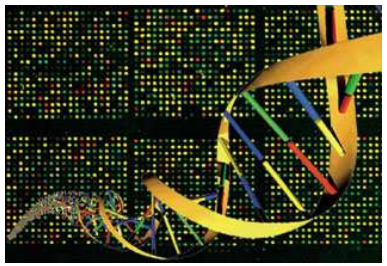


HRMS

MetMS

metabolome

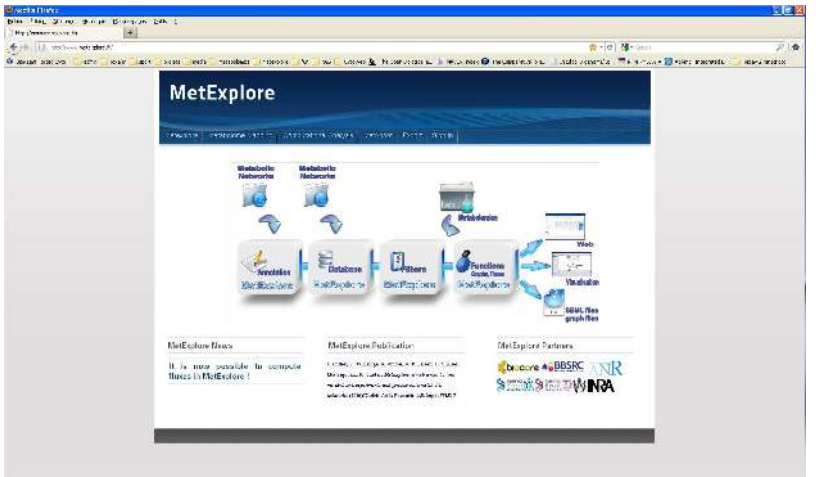
Transcriptomics data



Nistec, Kismet

ANR

Data integration



MetExplore: web server the storage, manual annotation and analysis of metabolic networks

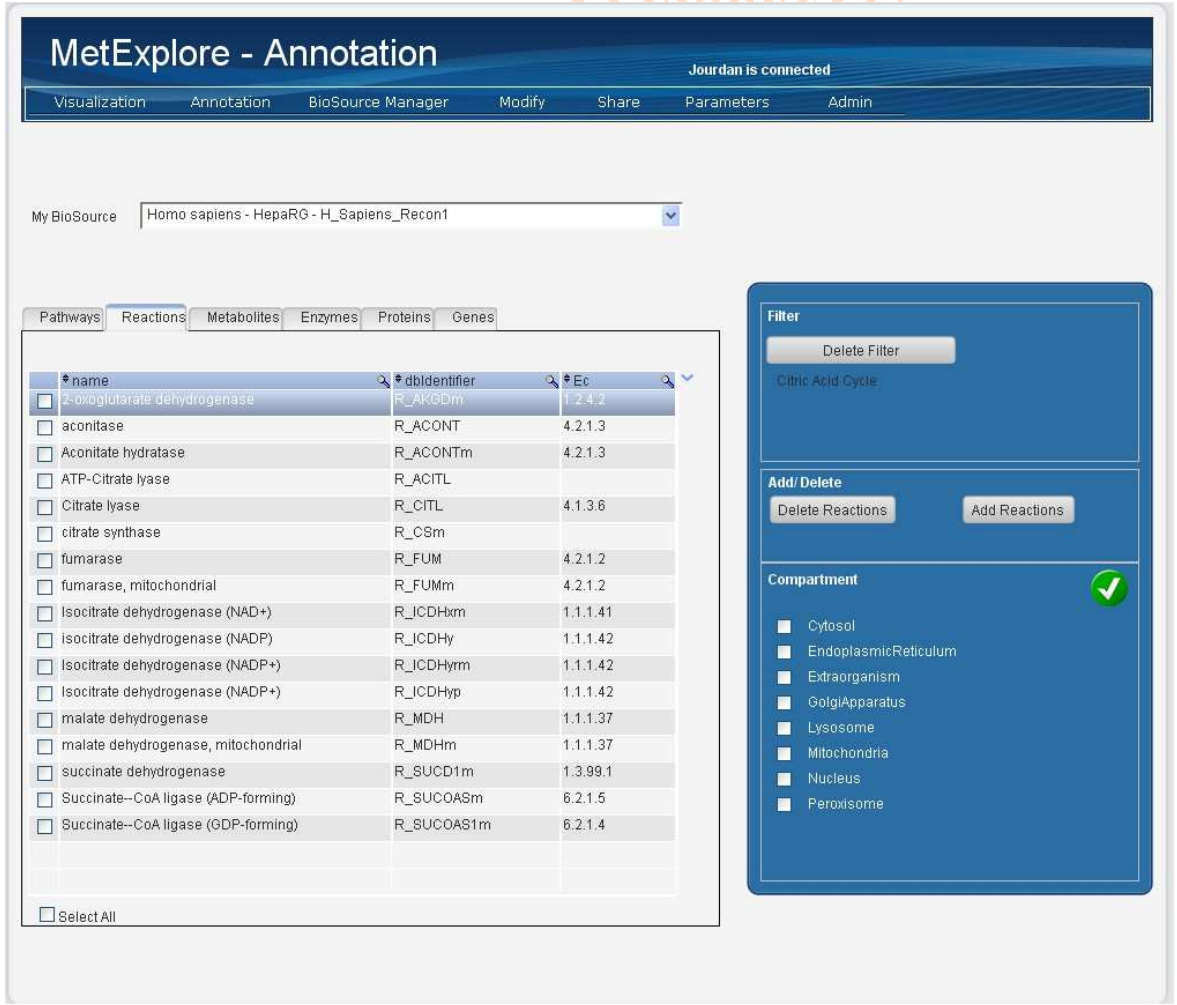
http://www.metexplore.fr

Cottret, L.; Wildridge, D.; Vinson, F.; Barrett, M. P.; Charles, H.; Sagot, M.-F. & Jourdan, F. MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks. *Nucleic Acids Res*, **2010**, *38*, W132-W137

ALIMENTATION
GRICULTURE
ENVIRONNEMENT

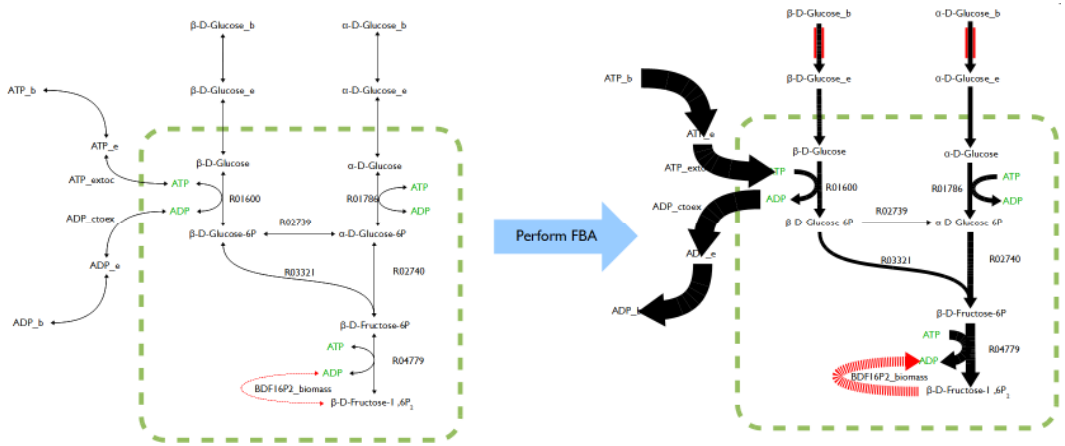
INRA

Annotation Project, manual curation



Florence Vinson

Next step: Metabolic Flux Plasticity



FBA: [2] Fell DA, Small JR: Fat synthesis in adipose tissue.
An examination of stoichiometric constraints**.** *Biochem J* 1986, **238**:781-786
FBA: [3] Orth JD, Thiele I, Palsson B:
What is flux balance analysis? *Nat Biotechnol* 2010, **28**:245-248.

Xenobiotics act on regulation of enzymes leading to changes in flux distribution.

In the case of cancer cells, these fluxes are different from healthy cell lines (Rupin *et al.*)

What are the flux distribution shifts when cells are exposed to xenobiotics?

What are the long time effects?




Metabolomics in France




ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA

French Metabolomics and Fluxomics network

BOARD:

D ROLIN (president, Bordeaux)

C JUNOT (Vice president, Saclay)

C DEBORDE (Secretary, Bordeaux)

JC MARTIN (Treasurer, Marseille)

C CANLET (Toulouse)

AM DELORS (Clermont-Ferrand)

F JOURDAN (Toulouse)

A PARIS (Paris)

JC PORTAIS (Toulouse)

E PUJOS-Guillot (Clermont-Ferrand)

RFMF
Réseau Français de Métabolomique et Fluxomique

Aims: animation, formation, networking, funding young researchers to attend conferences




National Conference

Date	Attendance	Labs	Invited Speakers
2005	86	27	
2006	107	39	
2008	109	44	
2010	101	56	J. Nicholson
2011	136	71	R. Breitling, M. Oresic

If you want to communicate with us:
Mailing list:
rfmf@*listes.inra.fr*

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

MetaboHub: creating a french metabolomics infrastructure

Developping and providing metabolomics analysis for large scale studies

Three scientific work packages

- Metabolomics (HT, standardisation, target metabolomics)
- Fluxomics (network, synthetic biology)
- Bioinformatics (data management)

Four facilities

- Paris (CEA)
- Bordeaux (INRA)
- Toulouse (INRA)
- Clermont Ferrand (INRA)




Conclusion

- Metabolomics bioinformatics pipeline in HRMS is a challenging computational problem
- Metabolic networks can be used for system biology interpretation of metabolomics data
- Genome based reconstructions have to be improved to provide better models
- Flux analysis will help to move towards dynamic studies




France Metabolomics

INRA Toulouse: Florence Vinson, Marc Dubois, M Dussart, D Zalko (MeX team)
CEA Paris: C Junot
INSA Toulouse: Ludovic Cottret, JC Portais (Metasys team)

International Metabolomics

University of Glasgow: M.P Barrett, D Wildridge
Sanger Institute: F Logan, M Beriman

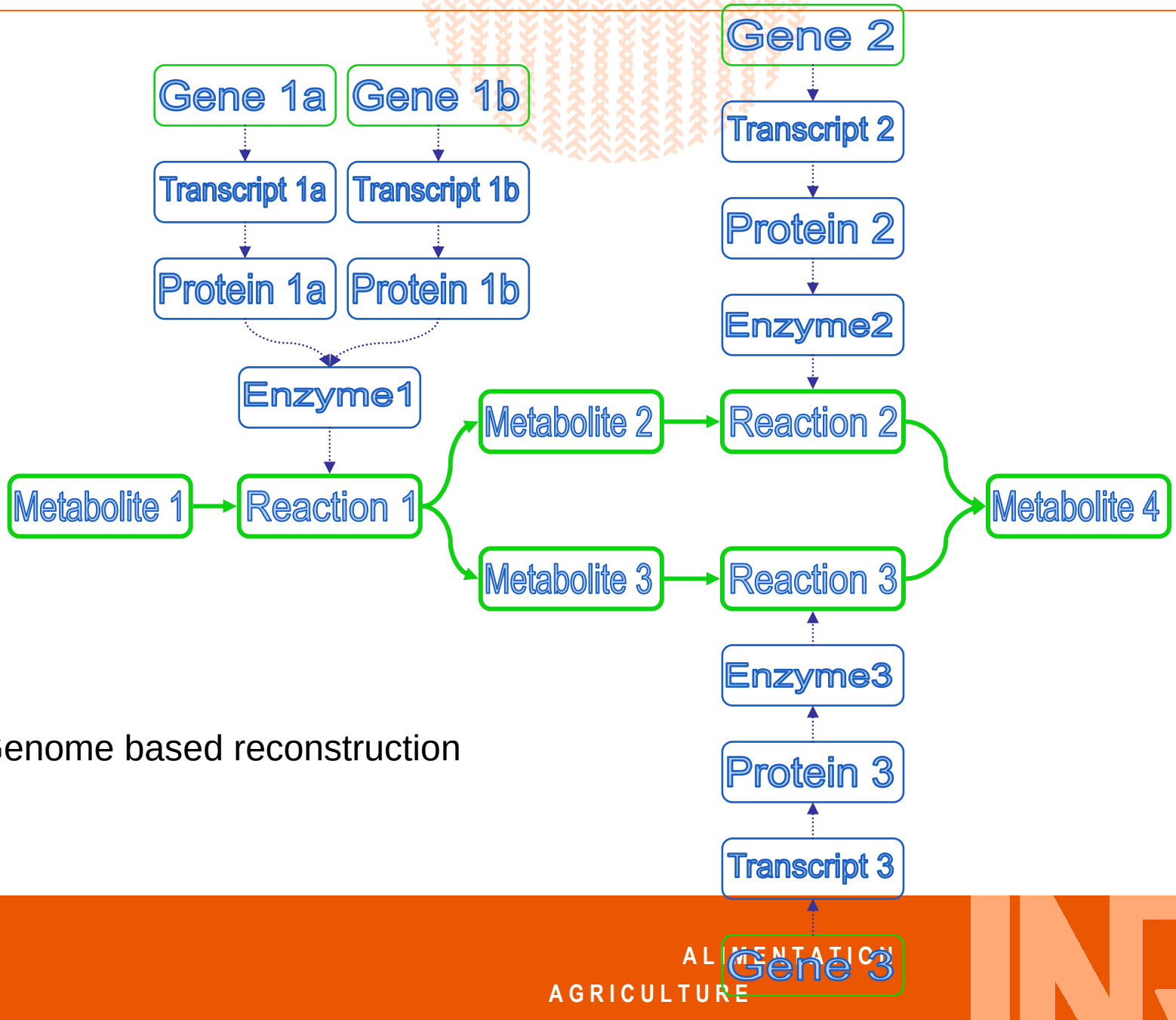
France Bioinformatics

INRIA Lyon: MF Sagot, V Lacroix, V Acuña, P Milreu, C Klein
University of Évry: E Birmelé

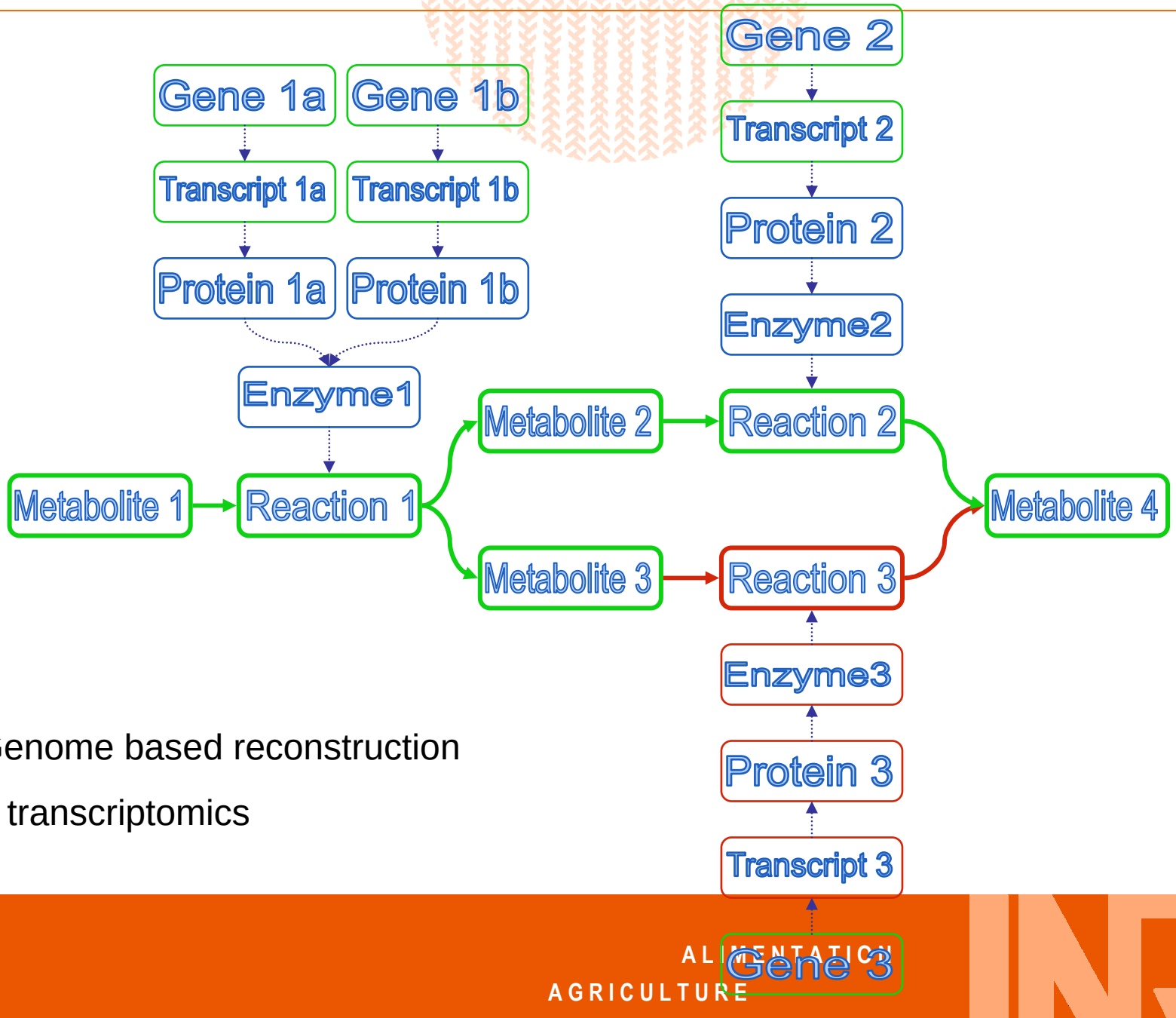
International Bioinformatics

Università degli Studi di Firenze: P Crescenzi, A Marino
Univ. La Sapienza, Rome: A Marchetti-Spaccamela
Free Univ. Amsterdam & CWI: L Stougie

And thanks for your attention !

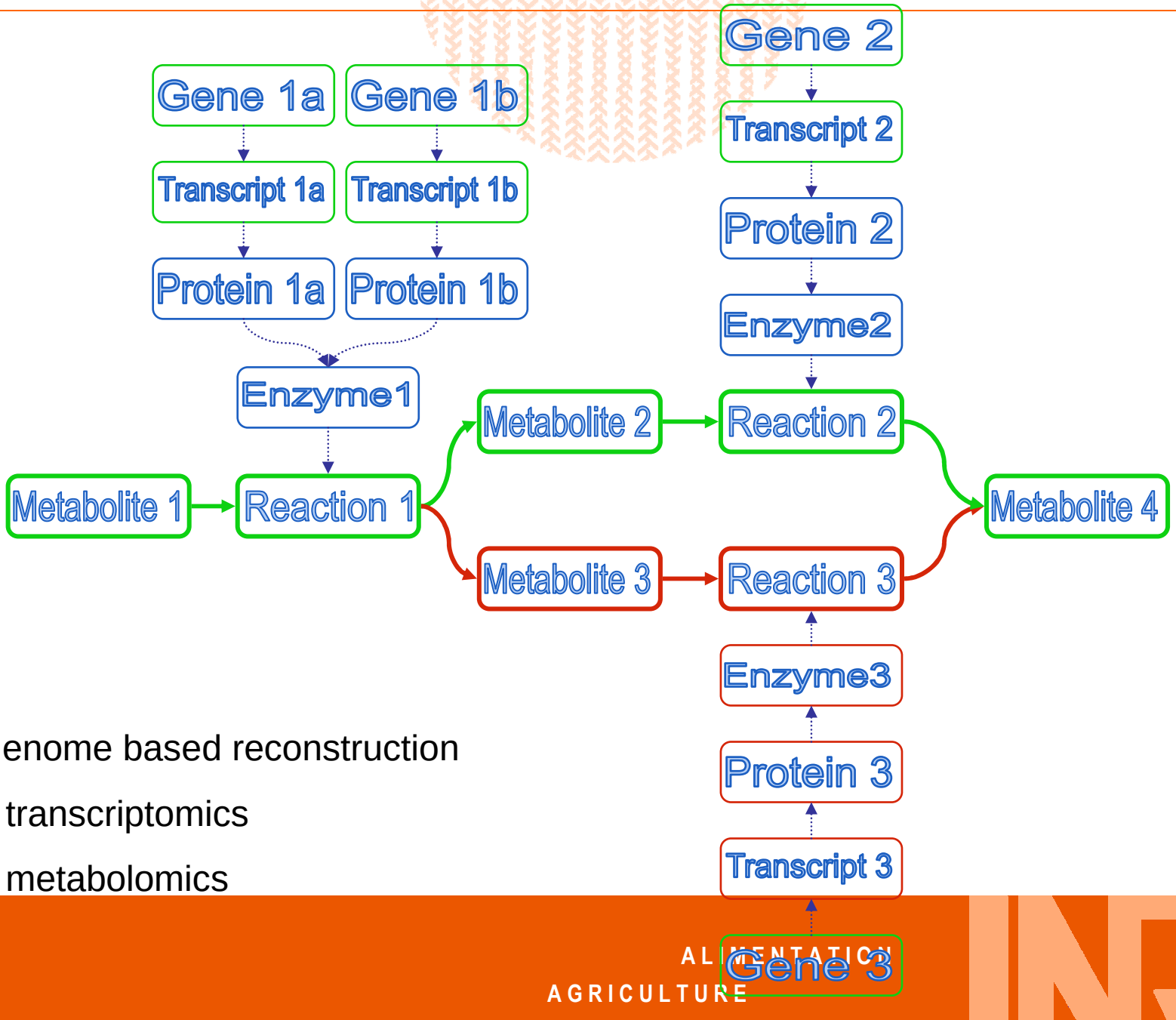


Genome based reconstruction



Genome based reconstruction

+ transcriptomics



Genome based reconstruction

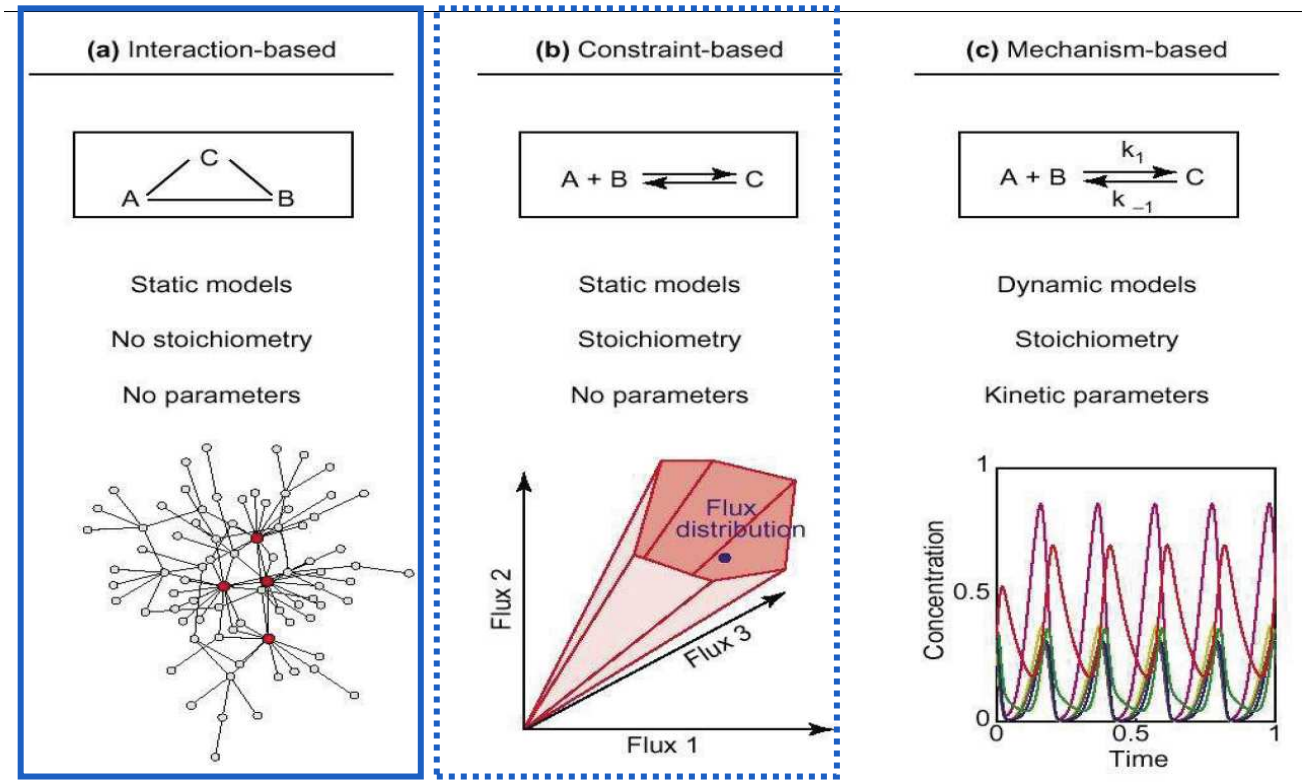
+ transcriptomics

+ metabolomics

Where we stand ?

Global

Local

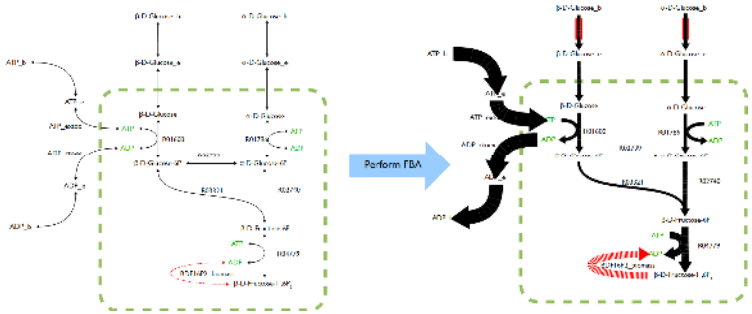


Qualitative

Quantitative

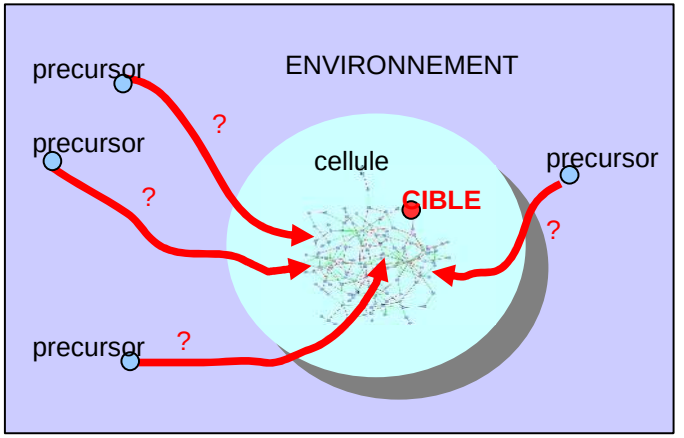
Use of metabolic network models

Synthetic biology



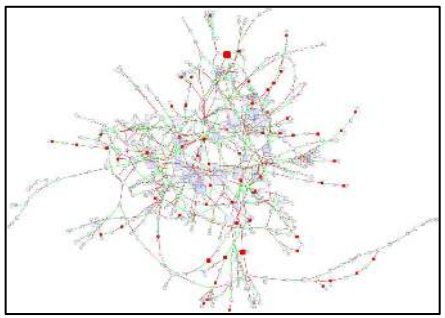
FBA: [2] Fell DA, Small JR: Fat synthesis in adipose tissue.
An examination of stoichiometric constraints**.** *Biochem J* 1986, **238**:781-786
FBA: [3] Orth JD, Thiele I, Palsson B:
What is flux balance analysis? *Nat Biotechnol* 2010, **28**:245-248.

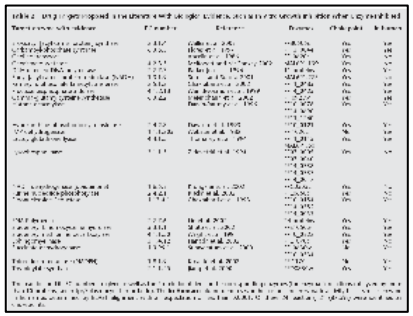
Symbiosis



Precursors : [4] Cottret et al. *WABI 08 Proceedings*, 2008, 233-244

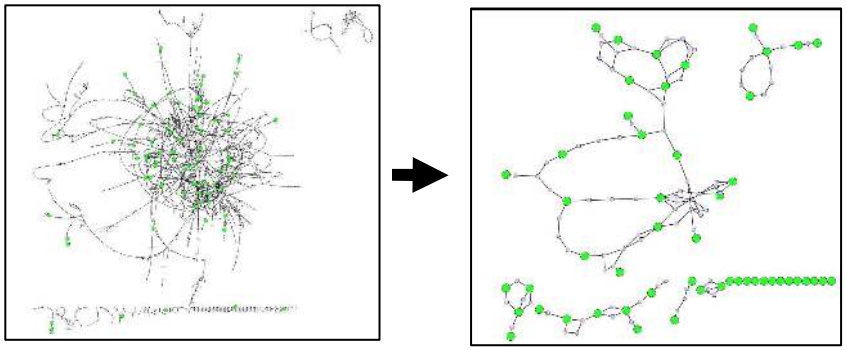
Drug target discovery



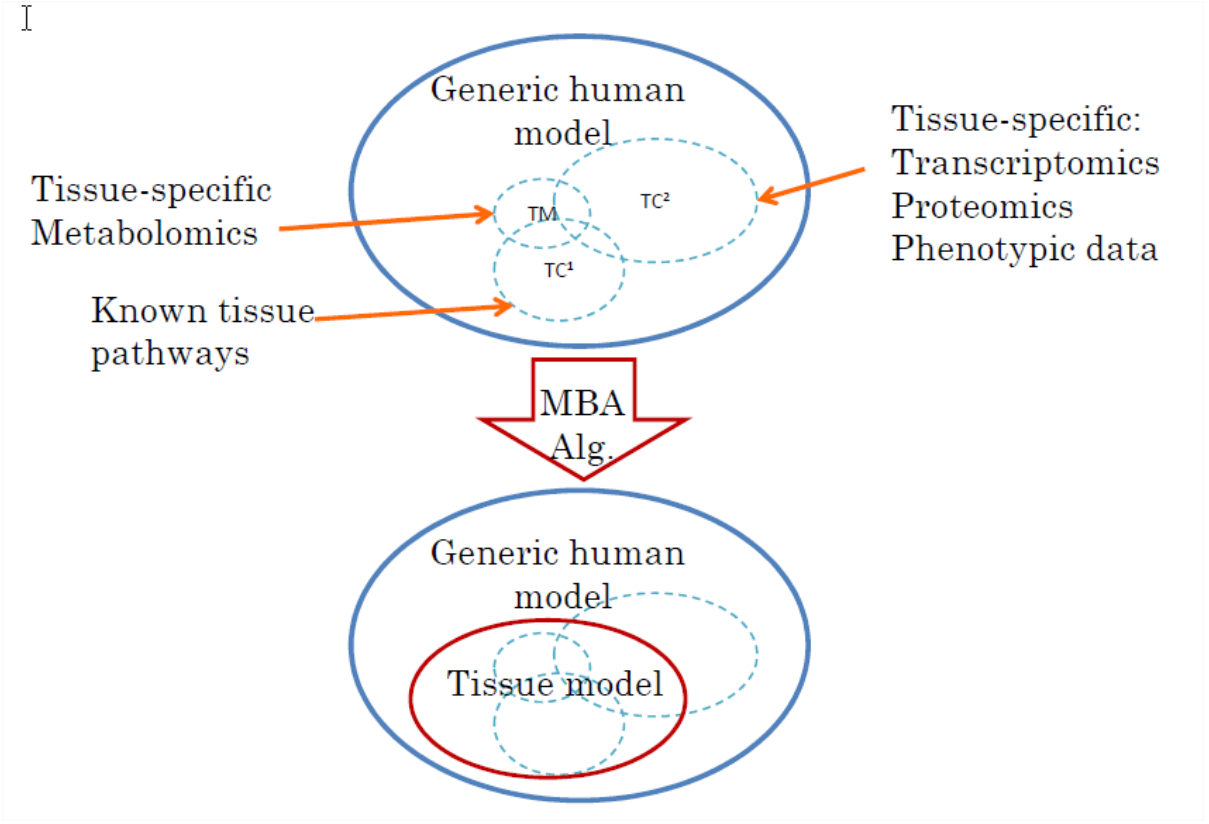


Chocke points : [5]Yeh et al. *Genome Res,* **2004***, 14*, 917-924

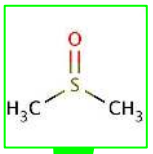
Metabolic effects in toxicology

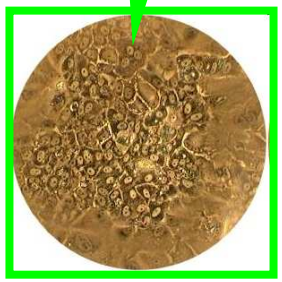


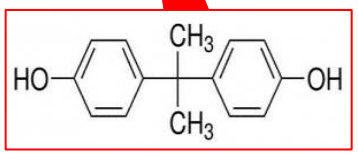
Sub-processes : [6] Antonov et al. *FEBS J,* **2009***, 276*, 2084-2094
Sub-processes : [7] Jourdan et al. *Metabolomics*, **2010**, 6(2): 312-321


Generic human model
Tissue-specific Metabolomics
Known tissue pathways
TM
TC²
TC¹
Tissue-specific:
Transcriptomics
Proteomics
Phenotypic data
MBA Alg.
Generic human model
Tissue model

DMSO

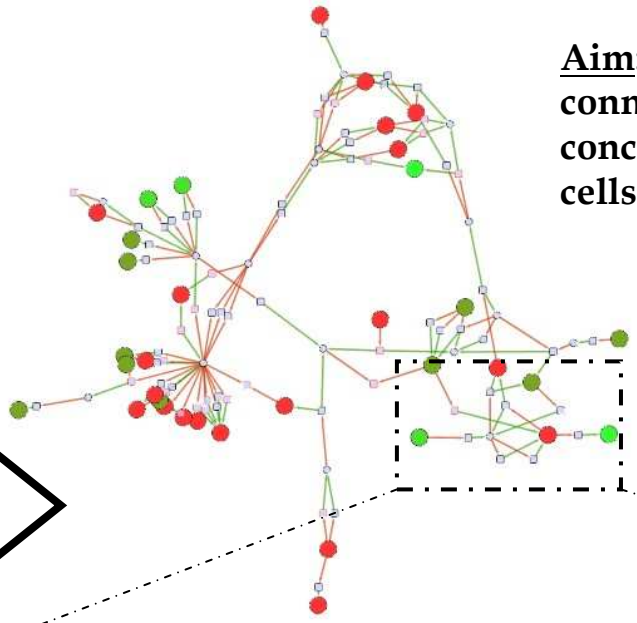






BPA

Network analysis



<u>Aim</u>: understanding wich process connect the metabolites which concentrations are affected when cells are exposed to BPA