Metabolomics data sciences for health

Etienne Thévenot

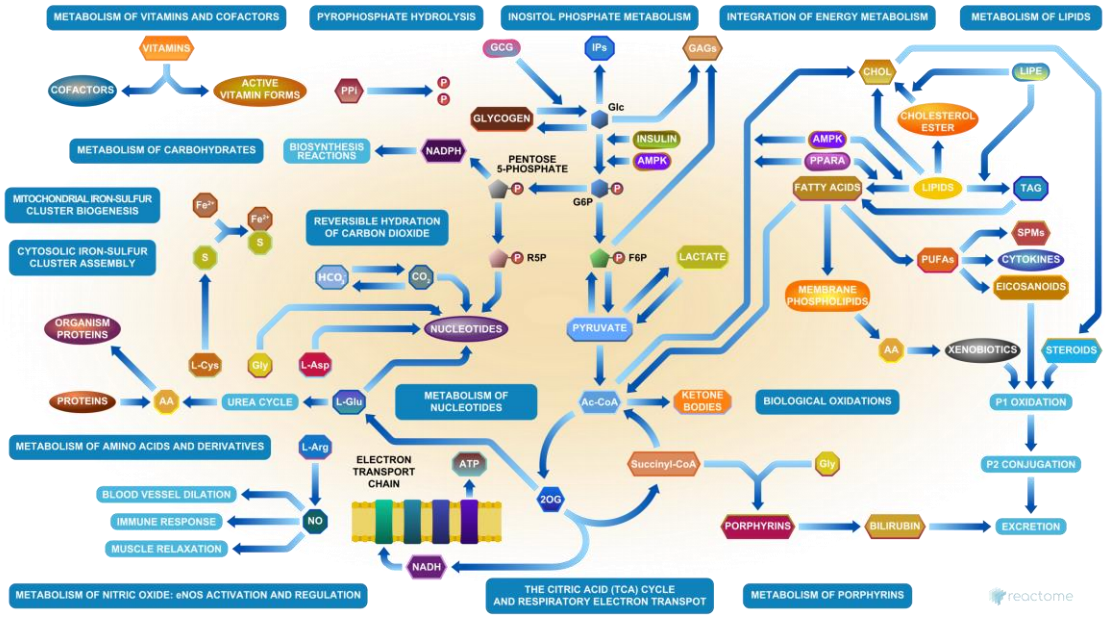
etienne.thevenot@cea.fr

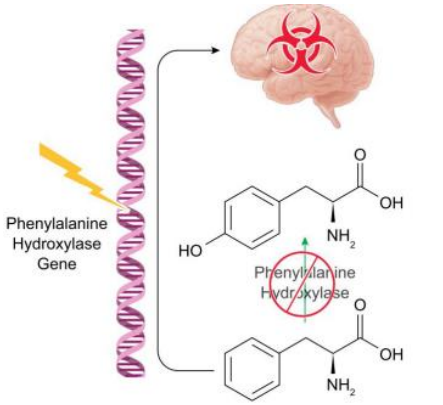
Equipe sciences des données métabolomiques (Odiscé)
Laboratoire Innovations en spectrométrie de Masse pour la Santé (LI-MS)
& Laboratoire d'Immuno-Allergie Alimentaire (LIAA)
Service de Pharmacologie et Immunoanalyse (SPI)
CEA, Centre de Saclay, F-91191 Gif sur Yvette, France

Introduction

Metabolism and metabolites

- **metabolism**: set of chemical reactions in the cell that allow the organism
  - to maintain itself alive, to reproduce, to develop, to respond to environmental stimuli
- **metabolites**
  - elementary organic compounds involved in these reactions
    - substrates, products, cofactors
  - multitude of functions
    - production and storage of energy, elementary bricks of macromolecules, role in signalling, immunity, sensors of the environment
- **metabolic diseases**
  - genetic (inborn errors of metabolism): e.g. accumulation of certain metabolites
  - chronic pathologies: obesity, diabetes, cardiovascular diseases
  - cancer, immunology neurological diseases (cell transformation and expansion)

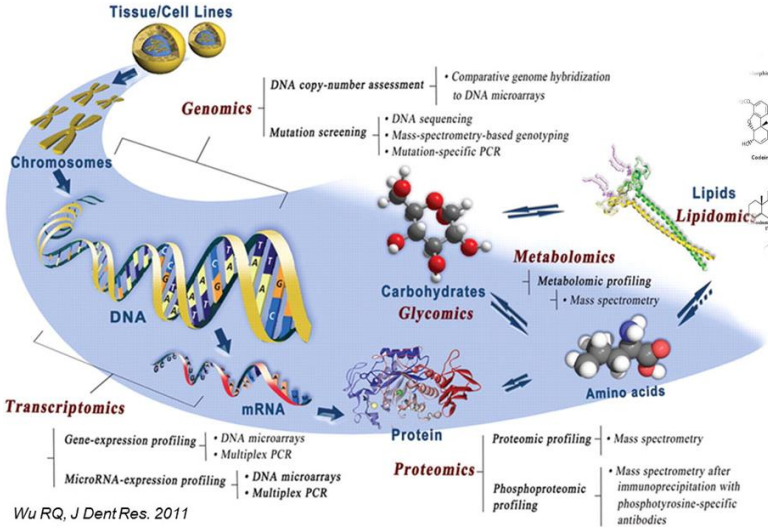


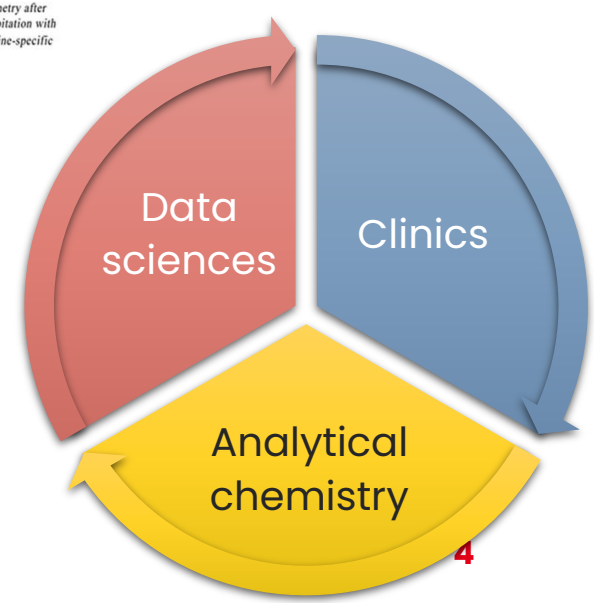


Phenylalanine

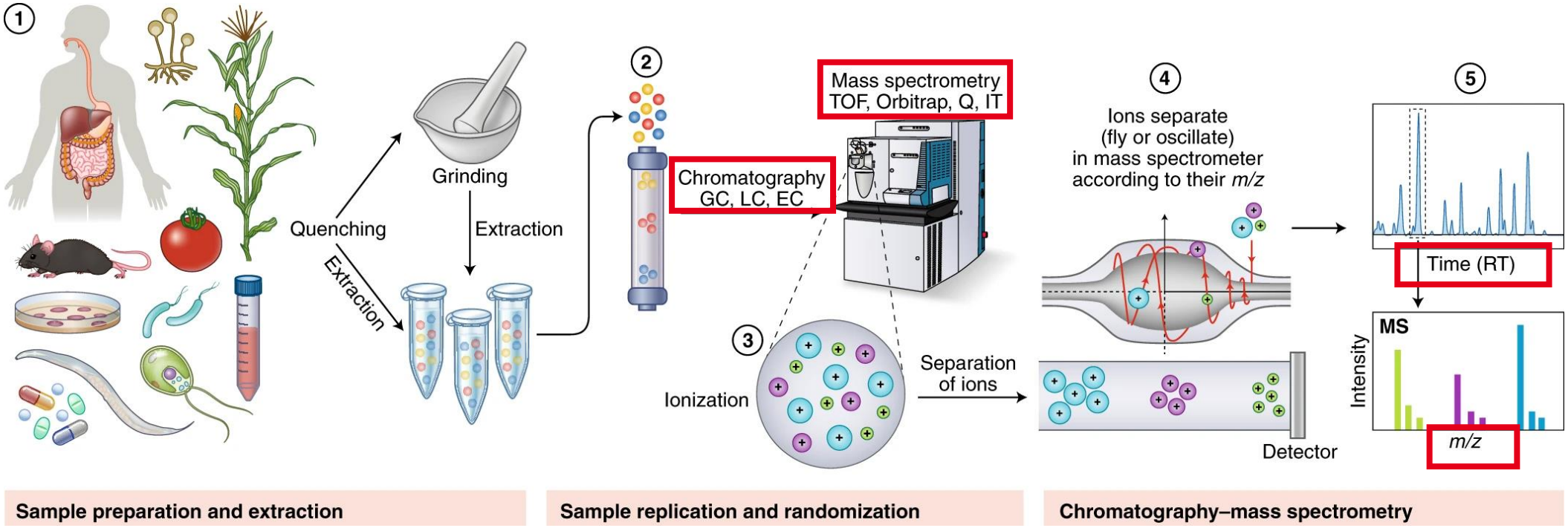
Metabolomics

- characterizing all small molecules (< 1.5 kDa) present in the body and the variations in their concentrations
  - endogenous metabolites (from the body, gut microbiota, diet)
  - exogenous metabolites (drugs, contaminants)
- based on large-scale analytical chemistry coupled with mathematical methods of signal processing and statistical analysis
- size of human metabolome > 10^5 compounds
- structural diversity (> 3,000 chemical classes)
- closest to the phenotype: integrates all upstream genetic and environmental perturbations





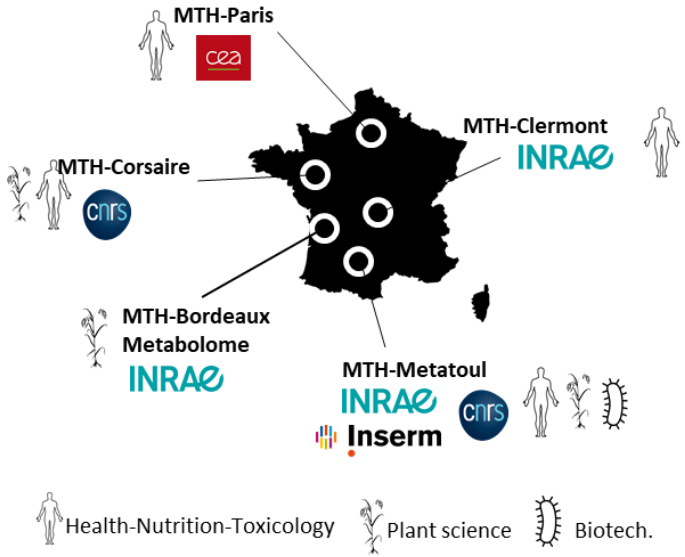
High-Resolution Mass Spectrometry coupled to Liquid Chromatography (LC-HRMS)



Alseekh,S. *et al.* (2021) Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nat Methods*, **18**, 747–756.

MetaboHUB: the French infrastructure for metabolomics and fluxomics

MetaboHUB aims at **providing cutting-edge analytical and bioinformatic services**, **training** and **support** to **academic research teams and industrial partners** in the fields of **nutrition, health, agriculture and biotechnology**



17.6 M€ National PIA funding since 2013

>130 Scientists

5 Cutting edge regional infrastructures

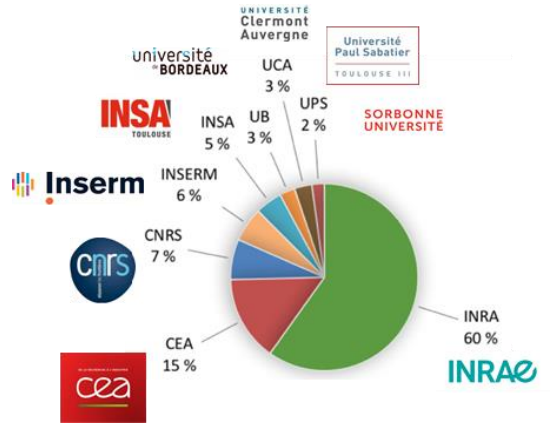
>470 Registred users

12 NMR spectrometers

51 (+15) Mass Spectrometers

>100 /y Publications

>100 EU and National grants



1. provide **high throughput quantitative analytical and digital technologies** for phenotyping from **cells to populations**,
2. **annotate metabolomes** through implementation and maintenance of **reference spectra databases**,
3. develop broadband **flux measurements**,
4. provide access to **high value-added services** to national scientific community and industry players,
5. **attract and train a new generation** of scientists and users

https://www.metabohub.fr/

Equipe sciences des données métabolomiques pour la santé

- Maria Teresa Avella
- Oneeb Nasir
- Diana Karaki
- Dylan Saunier
- Sylvain Dechaumet
- Eric Venot
- Etienne Thévenot

https://odisce.github.io/

- Alumni: Alma Toscano, Romuald Ait Bachir, Ghita Tajeddine, Ahmed Mohamedi, Krystyna Biletska, Camille Roquencourt, Camilo Broc, Alyssa Imbert, Pierrick Roger, Alexis Delabrière, Philippe Rinaudo, Natacha Lenuzza

Missions

- Major
  - Develop innovative methods and tools in phenotyping data sciences for precision medicine
  - Manage FAIR bioinformatics resources at LI-MS for high-throughput analysis of cohorts
- Minor
  - Computational metabolomics node of CEA and MetaboHUB Paris
  - Trainings in data sciences

Odisce

https://odisce.github.io/

Laboratoire Innovations en spectrométrie de Masse pour la Santé (LI-MS, CEA)
François Fenaille
Plateforme SMArt-Omics (Florence Castelli)

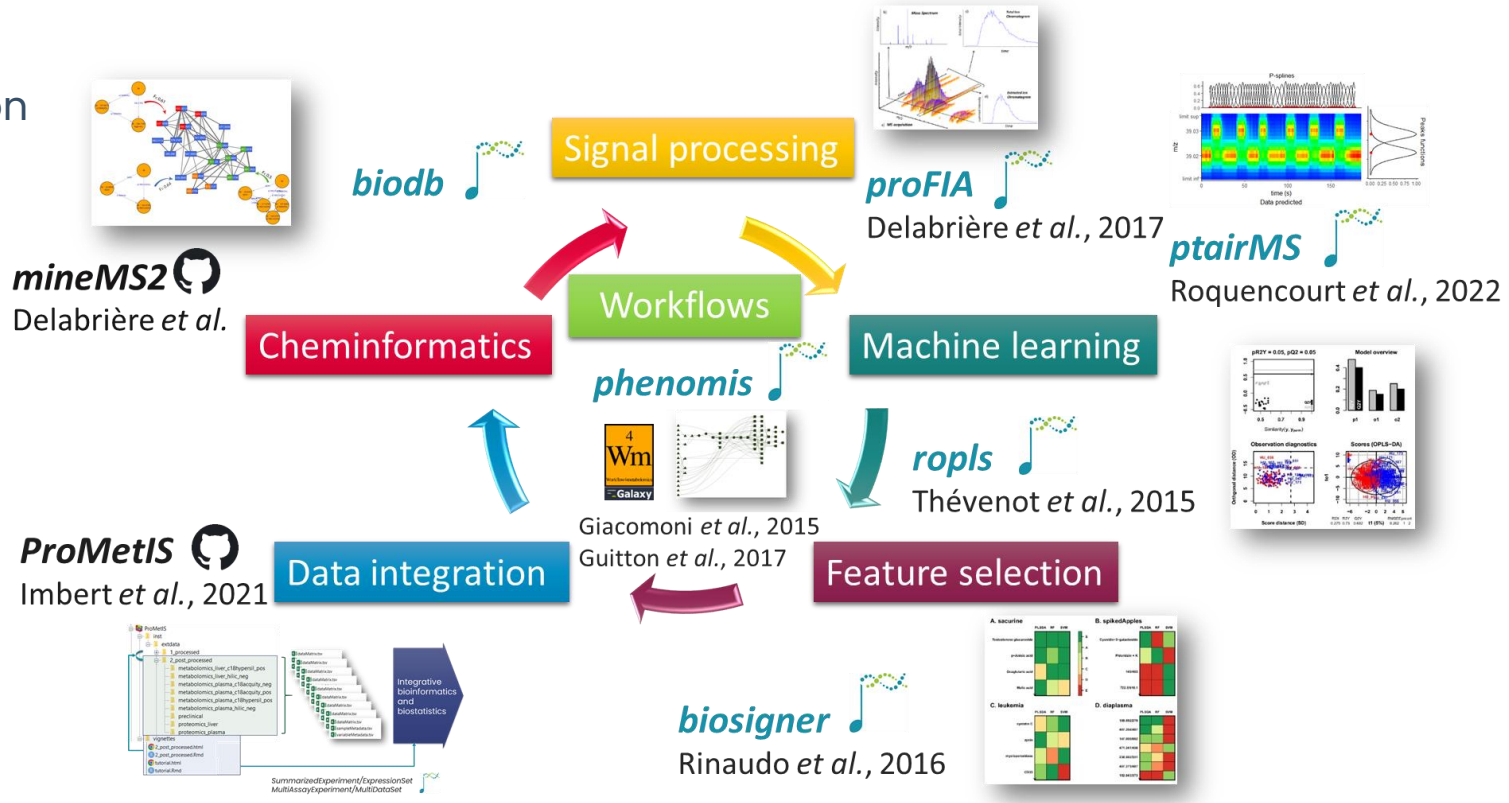
Laboratoire d'Immuno-Allergie Alimentaire (LIAA, INRAE)
Karine Adel-Patient

Service de Pharmacologie et d'Immunoanalyse (SPI)
Stéphanie Simon

Département Médicaments et Technologies pour la Santé (DMTS)
Christophe Junot

Computational metabolomics workflow

- Experimental design
- Preprocessing
  - peak detection and quantification (within samples)
  - peak grouping (between samples)
- Annotation
- Post-processing
  - signal drift and batch effect correction
  - log transformation
  - quality control
  - imputation
- Statistical analysis
  - exploratory analysis
  - univariate hypothesis testing
  - machine learning
    - feature selection
    - data integration
- Biological interpretation
  - metabolic pathway/network analysis

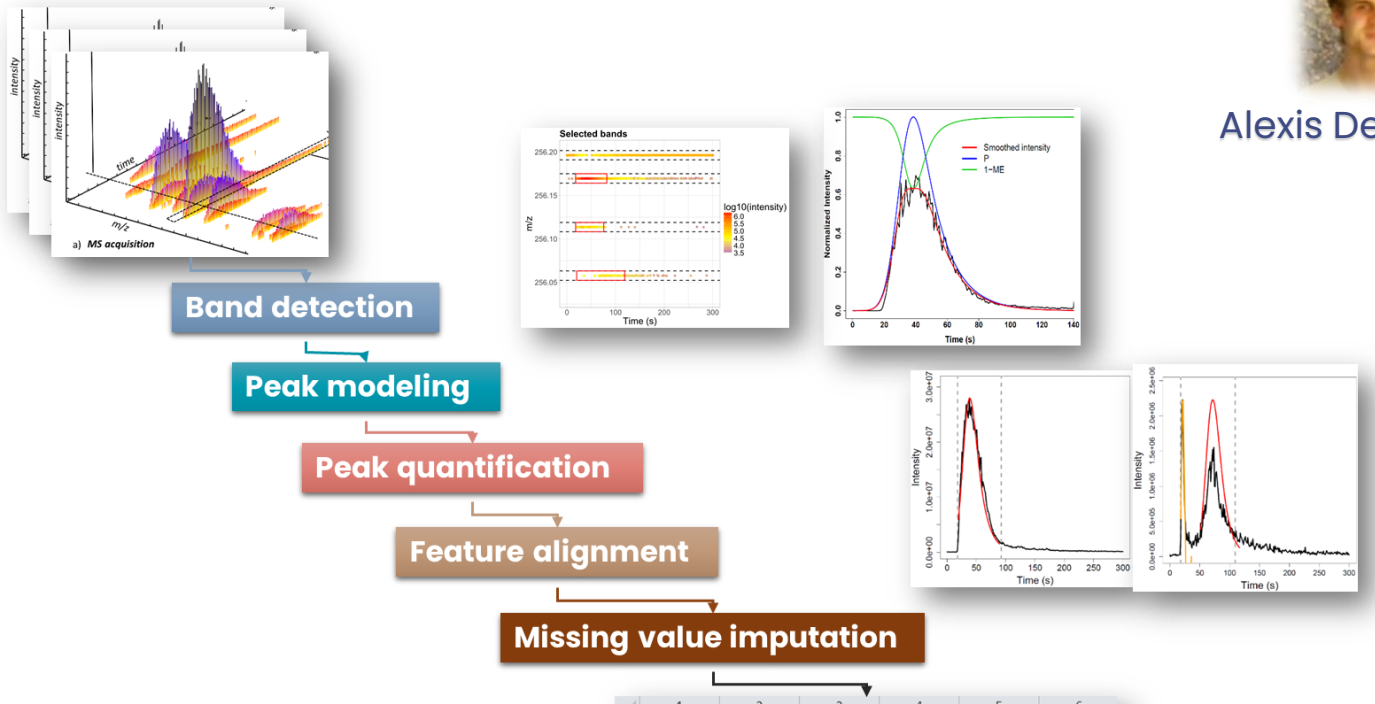


Data processing

proFIA: processing of FIA-HRMS data

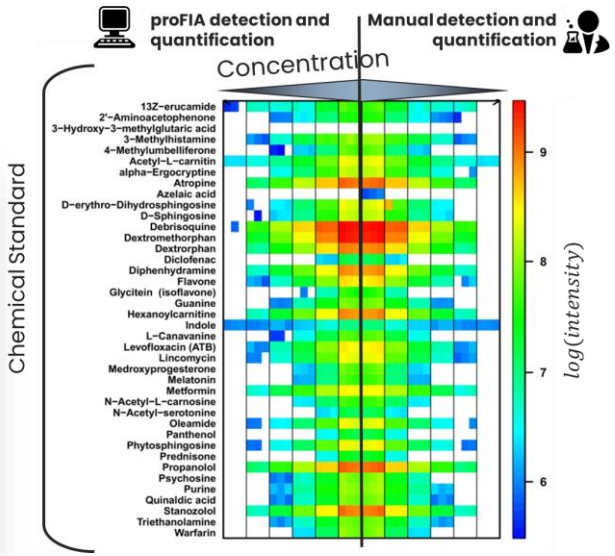
Alexis Delabrière

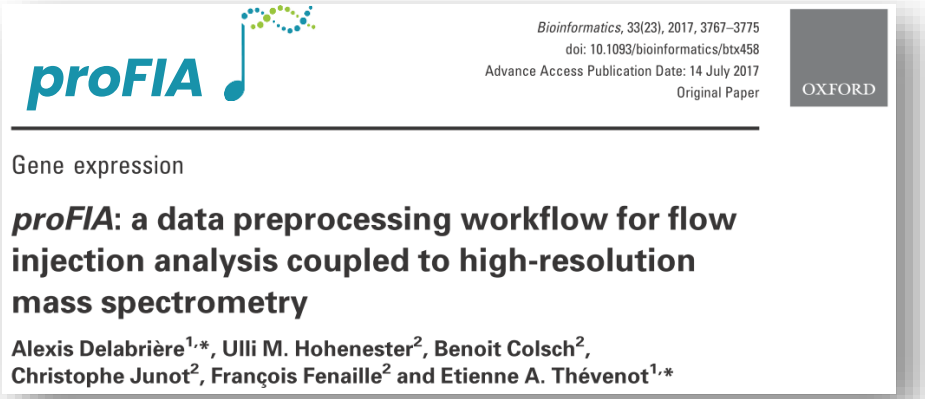
- Flow injection is of high-interest for high-throughput analysis
- Specific algorithms are required for these data
- *proFIA* achives high performance compared to manual integration:
  - precision of 0.96
  - recall of 0.98
  - mean intensity error < 5%



	1	2	3	4	5	6
						UIH344_12_
1	mzMed	mzMin	mzMax	meanSolvent	corMean	A
2	163.02777	163.02767	163.02782	845186.744	0.42770334	2105187
3	163.03893	163.03889	163.039	0	0.71292516	10484
4	163.11577	163.11562	163.11586	0	0.54386442	183310
5	164.02923	164.02919	164.0294	0	0.6646757	119575

'variables by samples' table of peak intensities

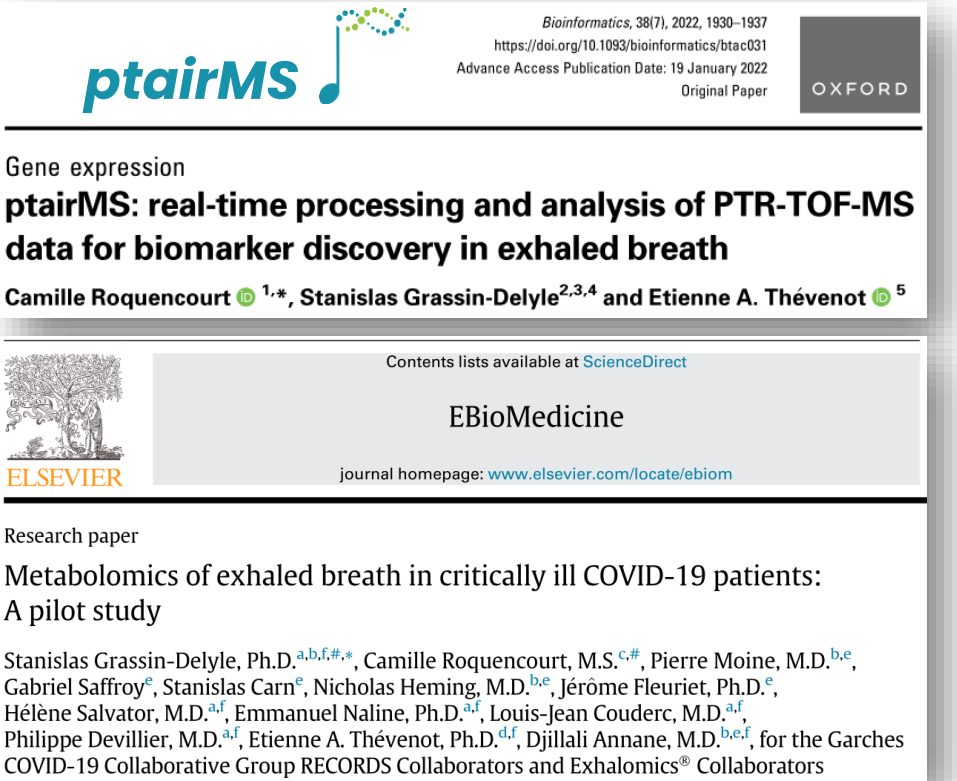




ptairMS: processing of PTR-TOF-MS data for volatolomics; application to COVID-19 diagnosis

- Full processing of PTR-TOF-MS data for biomarker discovery in exhaled air
- Four patented markers of severe COVID-19 infection
- Used in routine in 2 hospitals

Camille Roquencourt



ptairMS

Bioinformatics, 38(7), 2022, 1930–1937
https://doi.org/10.1093/bioinformatics/btac031
Advance Access Publication Date: 19 January 2022
Original Paper

OXFORD

Gene expression

ptairMS: real-time processing and analysis of PTR-TOF-MS data for biomarker discovery in exhaled breath

Camille Roquencourt [1,*], Stanislas Grassin-Delyle[2,3,4] and Etienne A. Thévenot [5]

ELSEVIER

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.elsevier.com/locate/ebiom

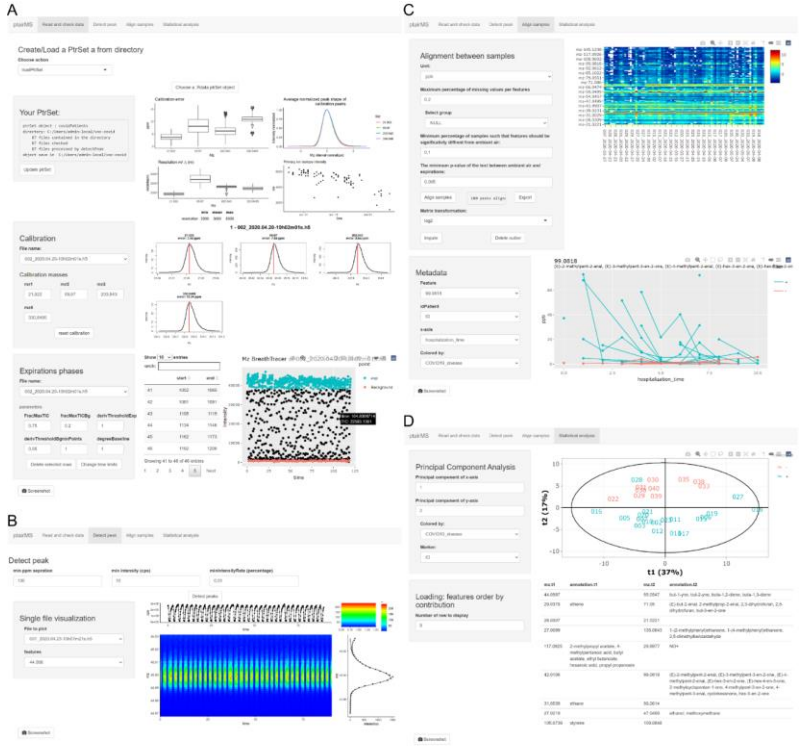
Research paper

Metabolomics of exhaled breath in critically ill COVID-19 patients: A pilot study

Stanislas Grassin-Delyle, Ph.D.[a,b,f,#,*], Camille Roquencourt, M.S.[c,#], Pierre Moine, M.D.[b,e], Gabriel Saffroy[e], Stanislas Carn[e], Nicholas Heming, M.D.[b,e], Jérôme Fleuriet, Ph.D.[e], Hélène Salvator, M.D.[a,f], Emmanuel Naline, Ph.D.[a,f], Louis-Jean Couderc, M.D.[a,f], Philippe Devillier, M.D.[a,f], Etienne A. Thévenot, Ph.D.[d,f], Djillali Annane, M.D.[b,e,f], for the Garches COVID-19 Collaborative Group RECORDS Collaborators and Exhalomics® Collaborators

HOPITAL FOCH

PTR-TOF-MS Ionicon



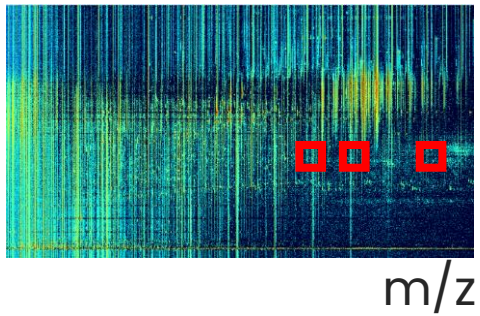
DeepMass: processing raw data by deep learning

Romuald Ait Bachir

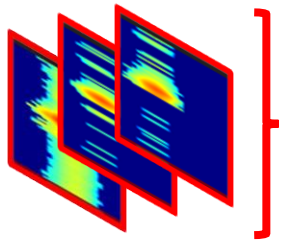
Biological sample

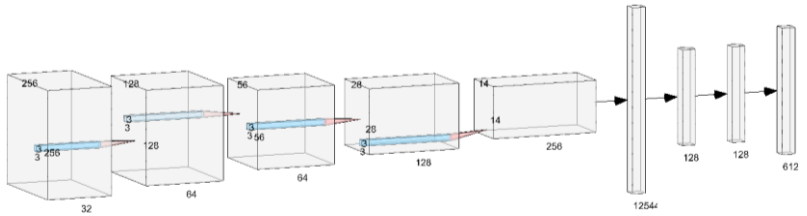
LC-MS analysis

Acquisition file



- For each acquisition file (i.e. for each sample)
  - For each metabolite of interest
    - Extract images from 3 reference ions (256x256 pixel; 50ppm x 60s)
    - Feed the CNN to predict the dilution factor and the molecule name





Molecule

Dilution factor

For each of the targeted metabolites

Extraction of the images from 3 reference ions

Processing by the CNN

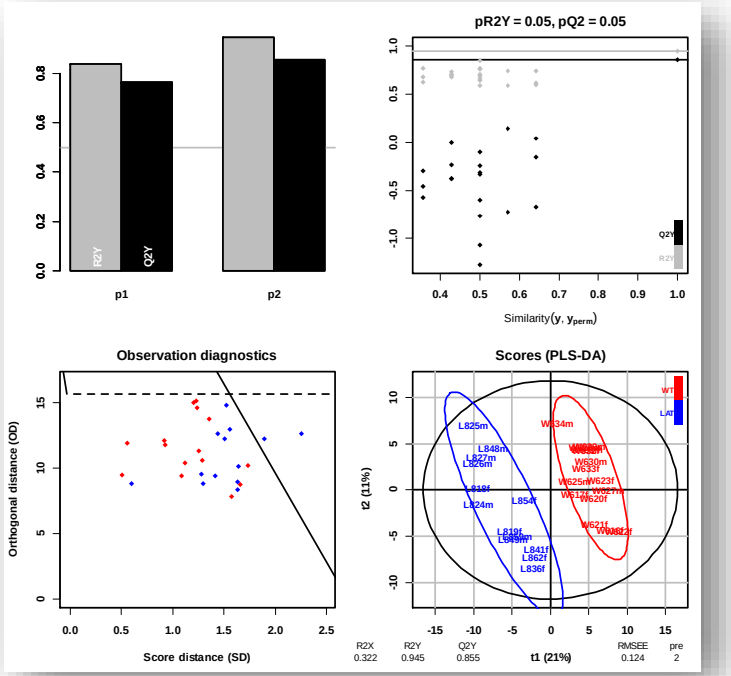
Prediction

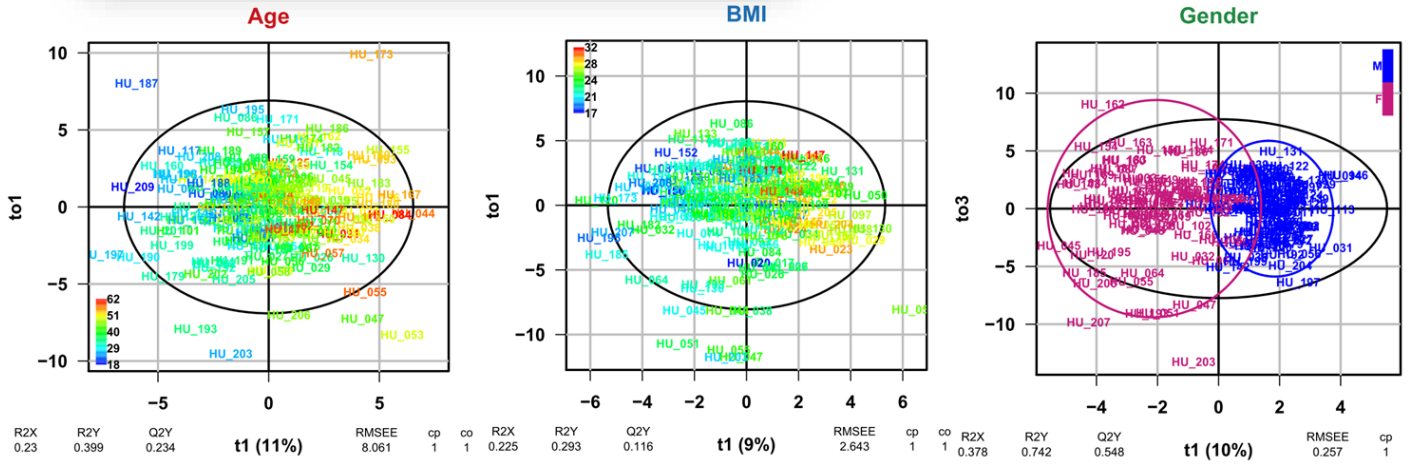
Machine learning

ropls: open-source implementation of (O)PLS(-DA) application to variations of the urine metabolome

Etienne Thévenot

- First open-source implementation of the Orthogonal Partial Least-Squares from Trygg and Wold (2002).
- Includes many numerical and graphical diagnostics and results
- Within the top 10% of the most Bioconductor downloaded packages
- Applied to the study of the physiological variations of the urine metabolome (*sacurine* dataset)





Journal of proteome research

ropls

Article

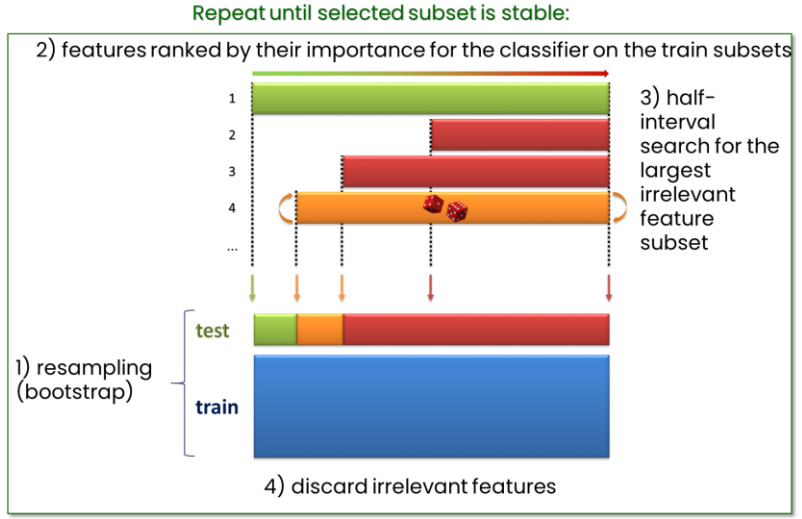
pubs.acs.org/jpr

Analysis of the Human Adult Urinary Metabolome Variations with Age, Body Mass Index, and Gender by Implementing a Comprehensive Workflow for Univariate and OPLS Statistical Analyses

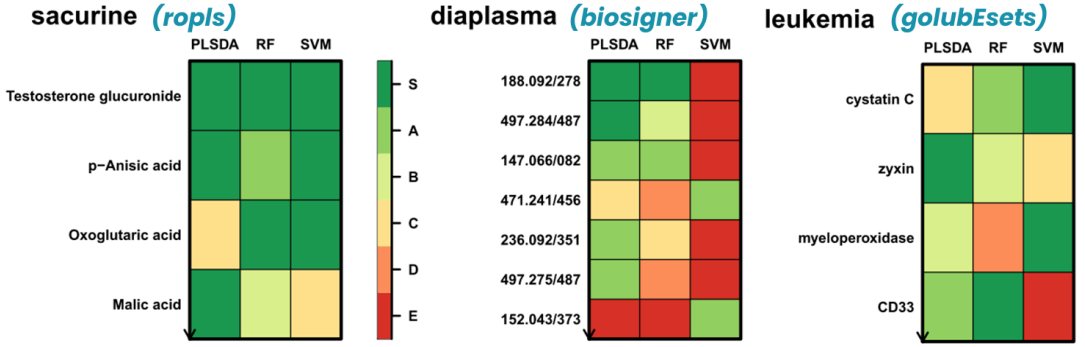
Etienne A. Thévenot,*,†,¶ Aurélie Roux,‡,¶ Ying Xu,‡ Eric Ezan,‡ and Christophe Junot*,‡

biosigner: selecting the features that *significantly* contributes to the prediction

- Recursive approach which selects feature which significantly contributes to the prediction performance
- Available for 3 classifiers (PLS-DA, Random Forest, Support Vector Machines)
- Application to (metabol)omics datasets shows that small and stable signatures are selected



Philippe Rinaudo



frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 21 June 2016
doi: 10.3389/fmolb.2016.00026

biosigner: A New Method for the Discovery of Significant Molecular Signatures from Omics Data

*Philippe Rinaudo[1], Samia Boudah[2], Christophe Junot[2] and Etienne A. Thévenot[1]**

		sacurine	diaplasma	leukemia
factor		gender	diabetic type	ALL/AML
samples		183	69	72
features		109	5,501	7,129
signatures		[2-3]	[0-2]	[1-2]
performances (full -> restricted)	PLS-DA	87% -> **89%**	83% -> **91%**	95% -> **87%**
	Random Forest	86% -> **86%**	81% -> **81%**	92% -> **92%**
	SVM	88% -> **89%**	83% -> **na**	93% -> **95%**

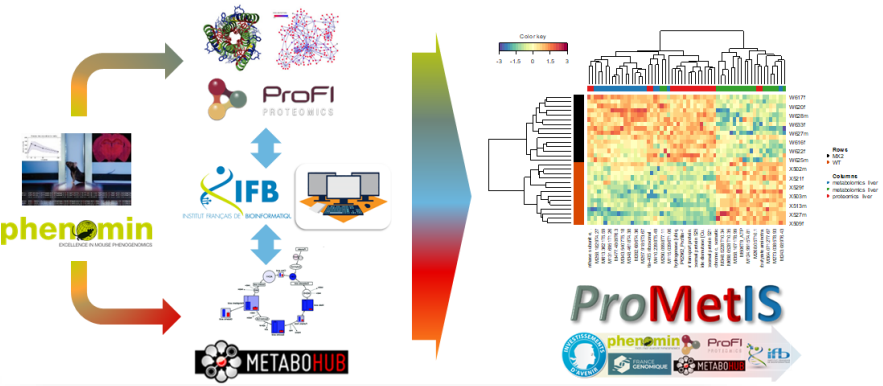
Data integration

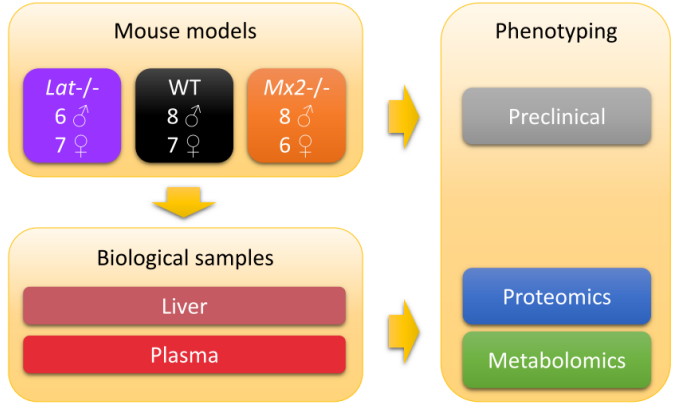
ProMetIS: deep phenotyping by proteomics and metabolomics

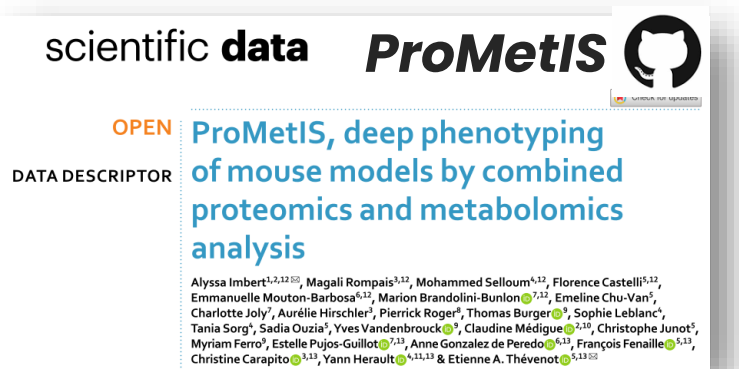
Alyssa Imbert

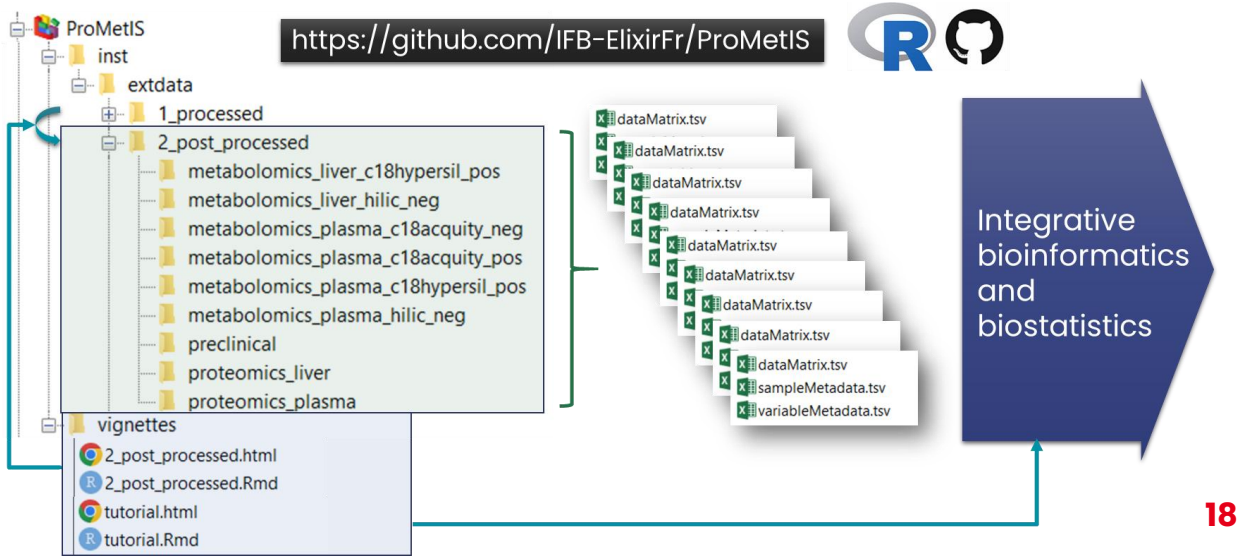
Oneeb Nasir

- Preclinical, proteomics and metabolomics analysis of liver and plasma from LAT and Mx2 KO mice
  - 9 datasets publicly available
- Single- and multi-omics analysis (*in progress*)
  - strong and complementary molecular changes







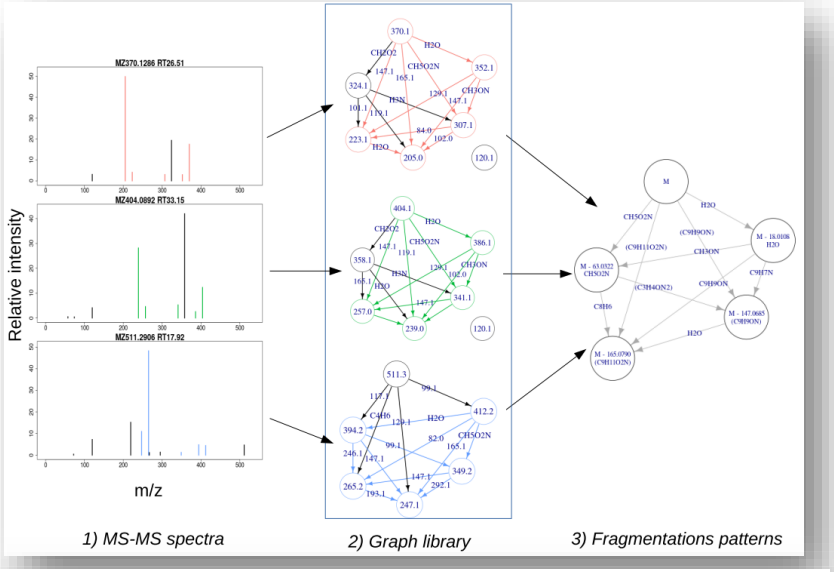


Identification

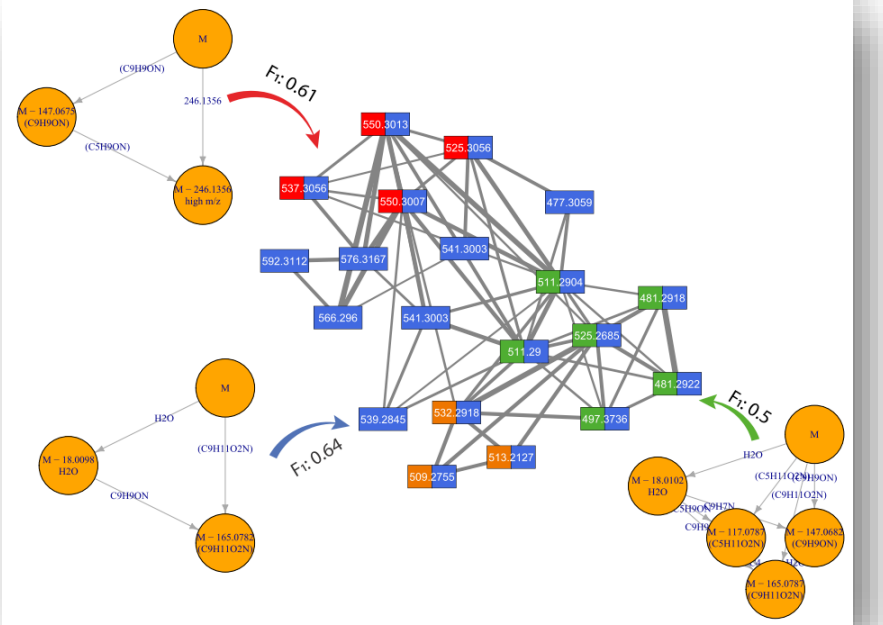
mineMS2: Annotation of spectral libraries with exact fragmentation patterns

Alexis Delabrière

- new method for automatically determining fragmentation patterns within MS/MS spectrum libraries



- application to the interpretation of molecular network components

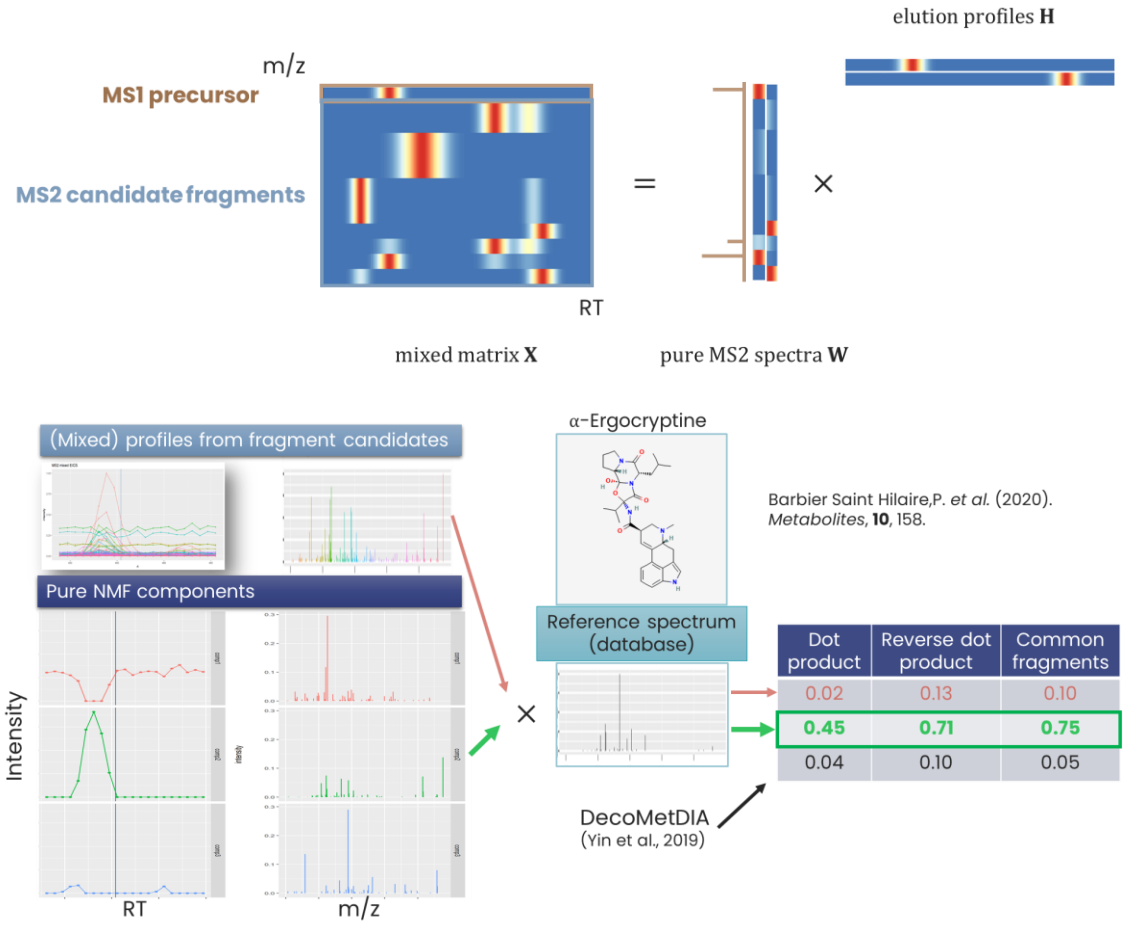


mineMS2

MS2Markers: Unmixing high-throughput DIA MS/MS data for comprehensive annotation

Diana Karaki

- Recent analytical approaches (SWATH DIA) enable to acquire MS/MS data for all ions
- The resulting MS/MS spectra are mixed resulting in poor matching scores with spectral databases
- New unmixing approaches are required
- MS/MS data processing workflow
  - MS1 peak detection
  - **for each MS1 precursor**
    - building the MS2 « mixed » matrix
    - unmixing of the MS2 matrix by NMF
    - matching the resulting « pure » MS2 spectrum to a spectral database



biodb: Searching chemical and biological databases

Pierrick Roger

biodb

- Query by IDs, MS, and MS/MS spectra
- Pathway mapping
- Unified, extensible framework

NCBI	CCDS, Gene, PubChem Compound, PubChem Substance	National Center for Biotechnology Information
miRBase	Mature	The microRNA database
UniProt		Universal Protein Resource
ExPASy	ENZYME	SIB Bioinformatics Resource Portal
HMDB	Metabolites	Human Metabolome Database
PeakForest	Compound, Mass	MetaboHUB spectral database
ChEBI		Chemical Entities of Biological Interest
LIPID MAPS	Structure	Lipidomics Gateway
KEGG	Compounds, Enzyme, Genes, Module, Pathway, Reaction	Kyoto Encyclopedy of Genes and Genomes

FAIR and scalable workflows

Building a comprehensive suite of software libraries application to single and multi-omics studies

Data integration

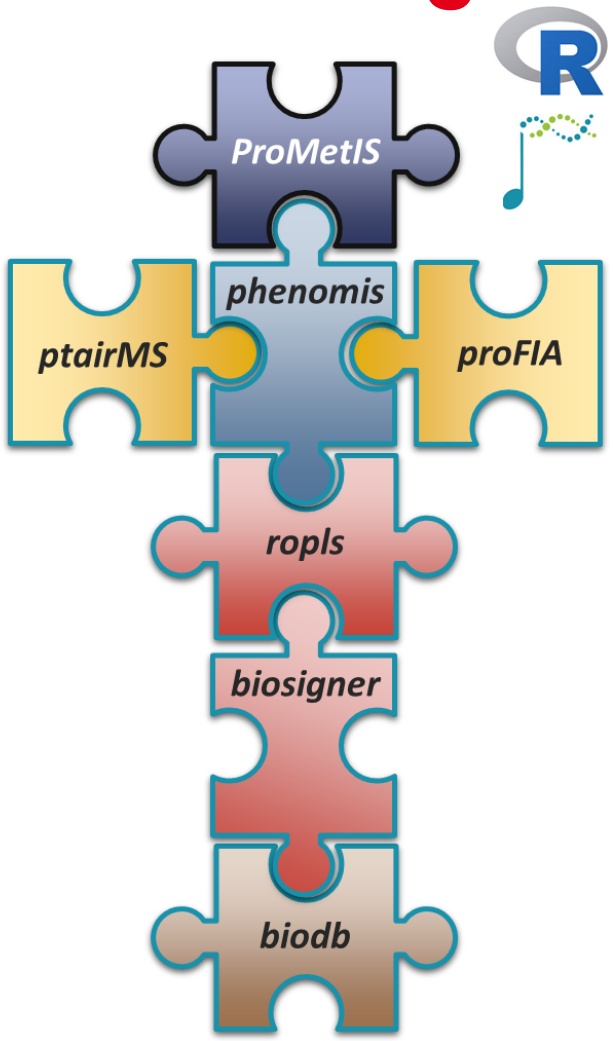
Preprocessing

Post-processing
Univariate analysis

Multivariate analysis

Feature selection

Annotation



Importing

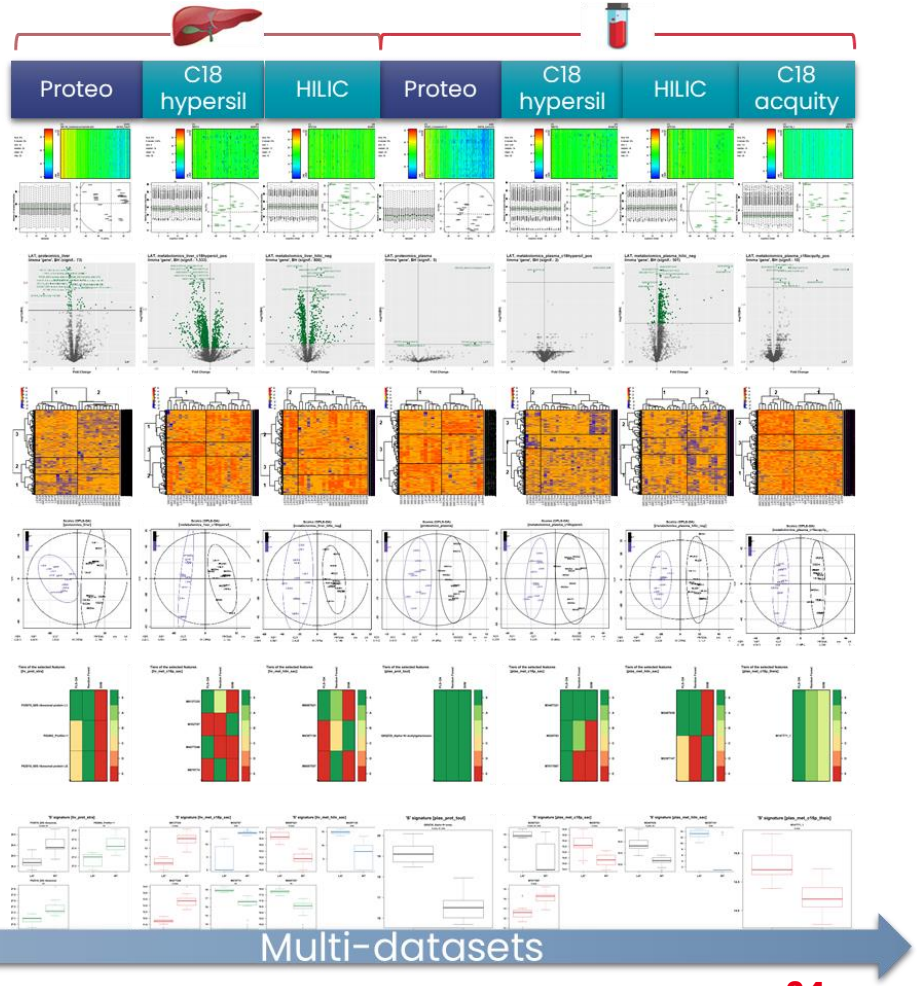
Post-processing
- Quality control
- Signal drift and batch effect correction
- Transformation

Statistics
- Univariate hypothesis testing
- PCA
- Heatmap
- (O)PLS(-DA)
- Feature selection

Annotation
- Chemical annotation

Exporting

Multi-steps

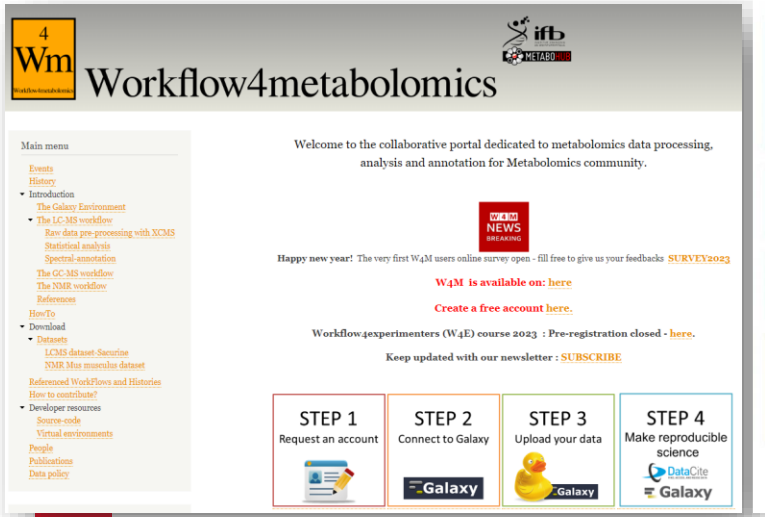


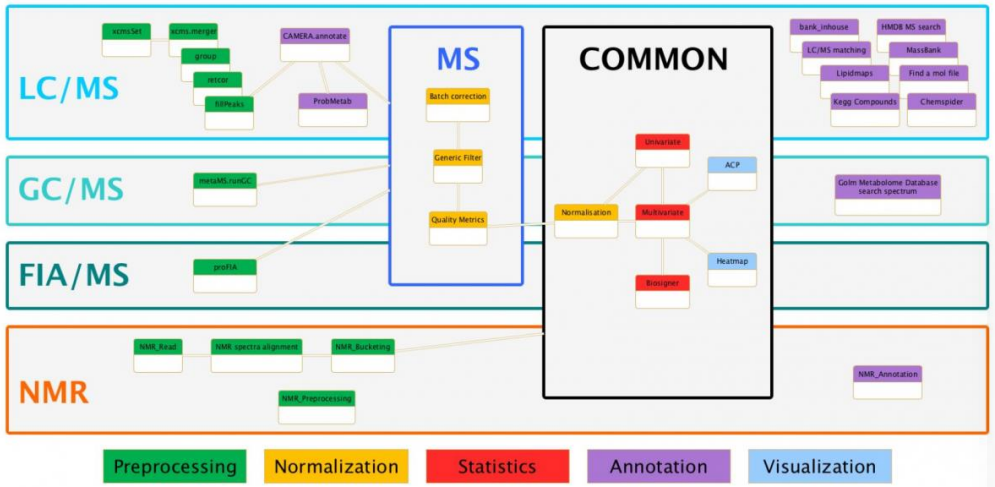
➢ Suite of 6 R packages available on Bioconductor for comprehensive statistical analyis

Workflow4Metabolomics

The W4M core team

- online public Galaxy instance dedicated to metabolomics
- > 100 modules
- reference workflows and histories (DOIs)
- annual training courses with hands-on on the participants data since 2014
- jointly supported by the French infrastructures for bioinformatics (IFB) and metabolomics (MetaboHUB)





Bioinformatics, 31(9), 2015, 1493–1495
doi: 10.1093/bioinformatics/btu813
Advance Access Publication Date: 19 December 2014
Applications Note

Gene expression

Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics

Franck Giacomoni[1,†], Gildas Le Corguillé[2,†], Misharl Monsoor[2], Marion Landi[1], Pierre Pericard[2], Mélanie Pétéra[1], Christophe Duperier[1], Marie Tremblay-Franco[3], Jean-François Martin[3], Daniel Jacob[4], Sophie Goulitquer[2], Etienne A. Thévenot[5,*] and Christophe Caron[2,*]

Contents lists available at ScienceDirect

International Journal of Biochemistry and Cell Biology

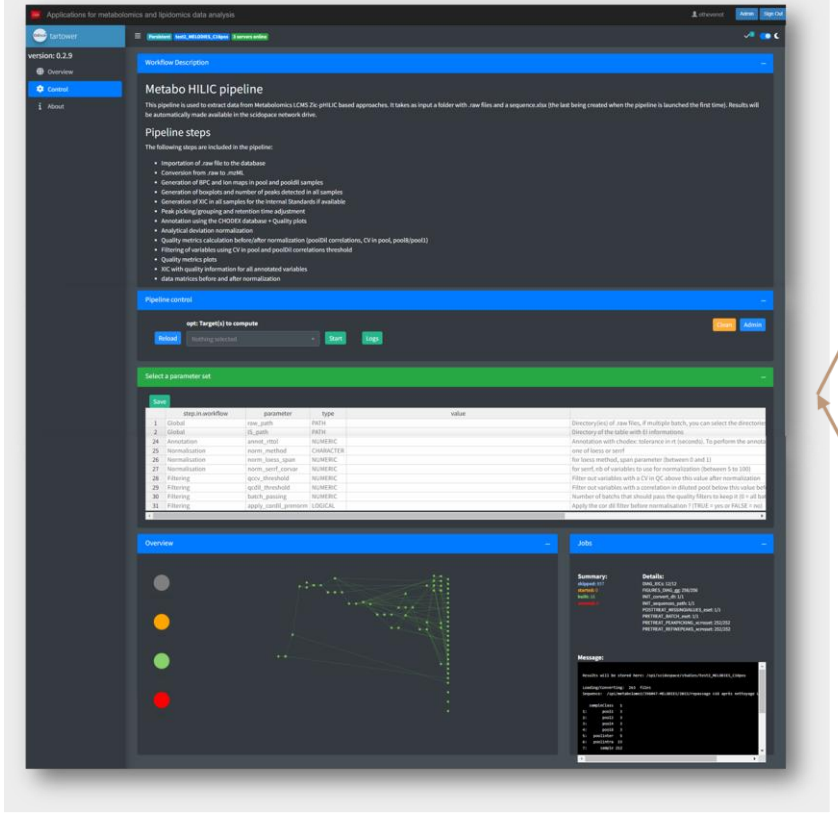
journal homepage: www.elsevier.com/locate/biocel

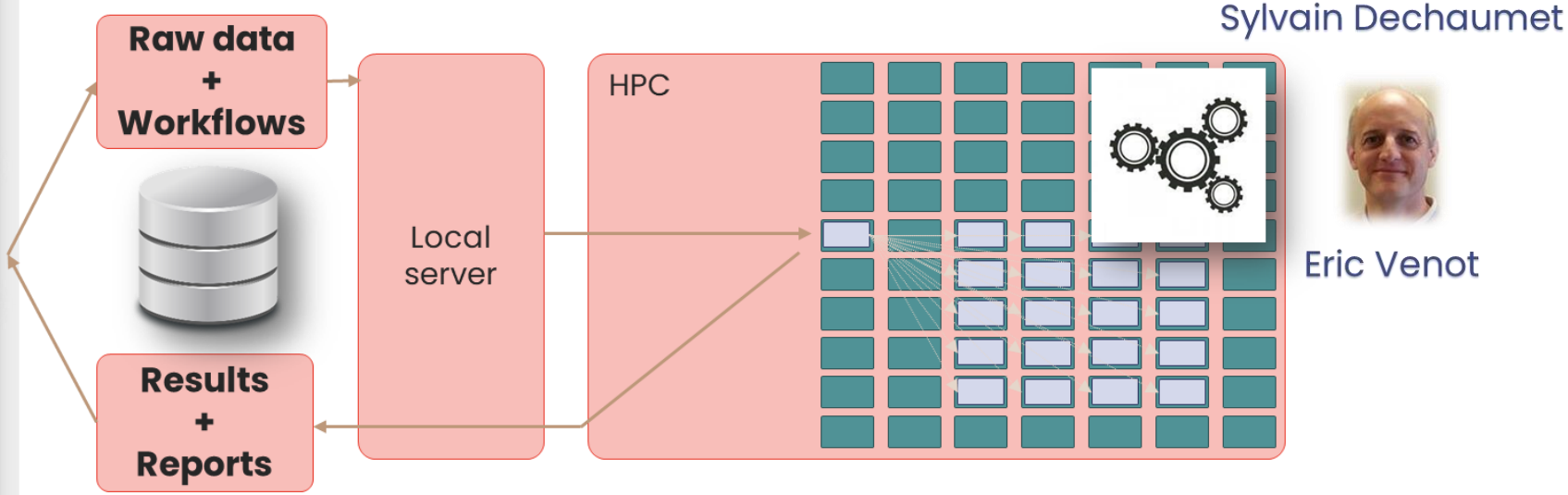
Create, run, share, publish, and reference your LC–MS, FIA–MS, GC–MS, and NMR data analysis workflows with the Workflow4Metabolomics 3.0 Galaxy online infrastructure for metabolomics

Yann Guitton[a,1], Marie Tremblay-Franco[b,1], Gildas Le Corguillé[c], Jean-François Martin[b], Mélanie Pétéra[d], Pierrick Roger-Mele[e], Alexis Delabrière[e], Sophie Goulitquer[f], Misharl Monsoor[c], Christophe Duperier[d], Cécile Canlet[b], Rémi Servien[b], Patrick Tardivel[b], Christophe Caron[g], Franck Giacomoni[d,*,2], Etienne A. Thévenot[e,*,2]

R/singularity based internal infrastructure for high performance computing & workflow management

Sylvain Dechaumet





Eric Venot

- Core infrastructure for computational metabolomics at SPI
  - launched in March 2023
- Running on the cluster from the Joliot Institute (as Singularity images)
- Shiny interface for workflow management
- More than 150,000 jobs already (>16,000 raw files)

- Maria Teresa Avella
- Oneeb Nasir
- Diana Karaki
- Dylan Saunier
- Sylvain Dechaumet
- Eric Venot
- Etienne Thévenot

Thank you!

https://odisce.github.io/

- François Becher
- Florence Castelli
- Céline Chollet
- Emeline Chu-Van
- Benoit Colsch
- Annelaure Damont
- Eric Ezan
- Blanche Guillon
- Arnaud Martel
- Jean-Claude Tabet
- Karine Adel-Patient
- François Fenaille
- Stéphanie Simon
- Christophe Junot
- Antoine Souloumiac
- Vincent Frouin
- Edouard Duchesnay
- Stanislas Grassin-Delyle
- Camille Roquencourt
- Guillaume Marti
- Marie Tremblay-Franco
- Marion Brandolini-Bunlon
- Estelle Pujos-Guillot
- Thomas Burger
- Christine Carapito
- Yves Vandenbrouck
- Myriam Ferro
- Claudine Médigue
- Jacques van Helden
- Yann Herault