MAPPING METABOLITES IN METABOLIC NETWORKS: THE IDENTIFIER CHALLENGE

METABOLIC NETWORKS FOR MULTI-OMICS STUDIES

Florence VINSON
Fabien JOURDAN
Senior Researcher
French Agronomic Research Institute (INRA) Toulouse, FRANCE
@MetExplore
Fabien.Jourdan@inra.fr

Diseases

Food contaminants

MetExplore Web server development

Cottret et al. **NAR.** 2018
Chazalviel et al. **Bioinformatics** 2018

Cells

Tissues

Biofluids

transcriptomics

metabolomics

global mathematical modeling of metabolism

Suggest metabolic profile interpretation

Del Mar Amador et al. **J. Inherit. Metab. Dis**. 2018
Frainay et al. **Bioinformatics** 2018
Frainay and Jourdan **Brief. In Bioinfo**. 2017

understand the metabolic characteristics of cells/tissues

Stuani et al. **Nature Cancer**, 2019 Under review
Poupin et al**. J. Prot. Research** 2018

predict specific biomarkers of expositon/disease

Poupin et al. **Sci. Reports** 2019

Predictive toxicology of Metabolic Disrupting Compounds

Mesnage et al. **Archives of Tox.** 2018

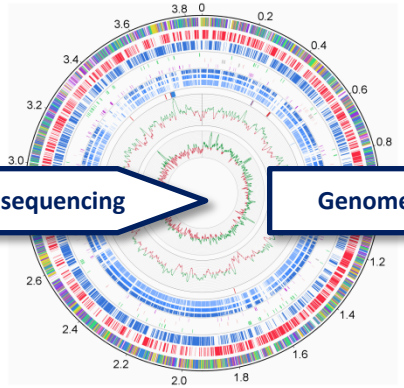
Apports du CATI pour …

- L'axe transversal du département Alimentation Humaine
  - Animation des bio-info-math-stat du département
  - Eviter les doublons dans les projets
  - **Lieu pour réaliser des projets communs et transversaux aux questions de recherche AlimH**
- MetaboHub
  - Travailler sur des méthodes d'intérêt pour les plateformes analytiques
  - Assurer l'interopérabilité entre les systèmes d'information (PeakForest, W4M, MetExplore)
  - **Identifier les futurs défis et fronts technologiques**
- Vous
  - Avoir un lieu d'échange méthodologique prenant en compte les attentes de l'institut … tout en se faisant plaisir!!!
  - **Suggestion: favoriser les hackathons, les partages techniques/méthodologiques**

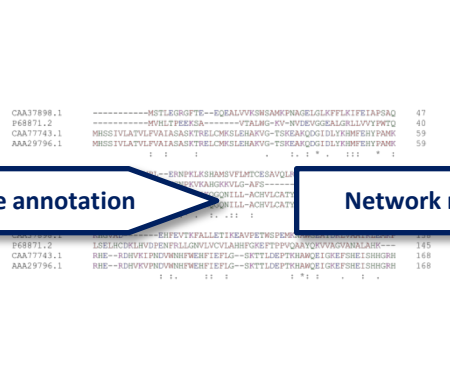
INTRODUCTION

Gathering metabolic knowledge – genome scale models

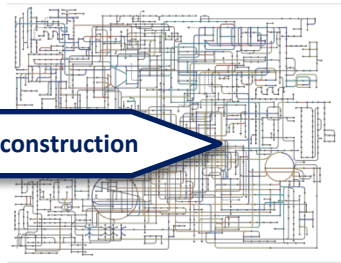
Genome sequencing



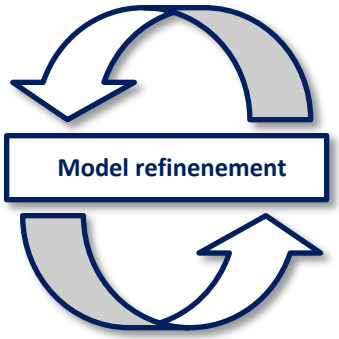
Genome annotation



Network reconstruction



Model refinenement



stoichiometric associations

Thiele I, Palsson BØ. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat. Protoc. 2010; 5:93–121

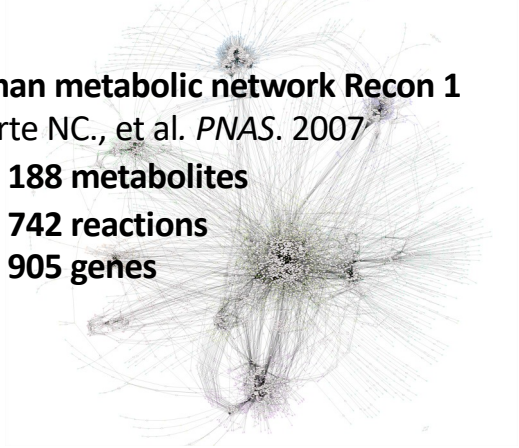
list of metabolites
$gluc_e$
$gluc_c$
Na_e
Na_c
ATP_c
ADP_c
$g6p_c$
$f6p_c$
$6pgl_c$
$nadp_c$
$nadph_c$
$gluc_r$

list of reactions		list of genes
GLCt4	$Na_e + gluc_e \leftrightarrow Na_c + gluc_c$	(6524) or (6526)
HEX1	$gluc_c + ATP_c \rightarrow ADP_c + g6p_c$	(3098)
PGI	$g6p_c \rightarrow f6p_c$	(2821)
G6PDH2r	$g6p_c + nadp_c \rightarrow 6pgl_c + nadph_c$	(2539)
GLCter	$gluc_c \leftrightarrow gluc_r$	

"Essentially, all models are wrong, but some are useful."

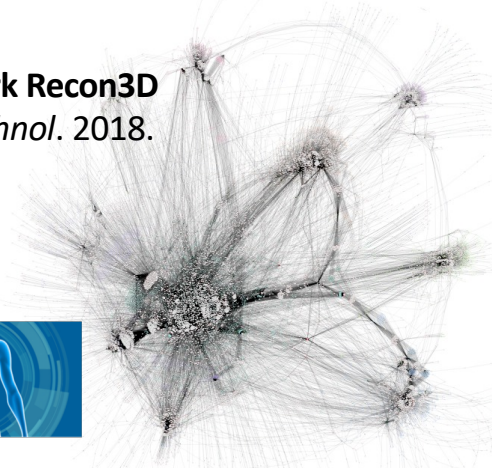
George E. P. Box, Norman R. Draper (1987) Empirical Model-Building and Response Surfaces.

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

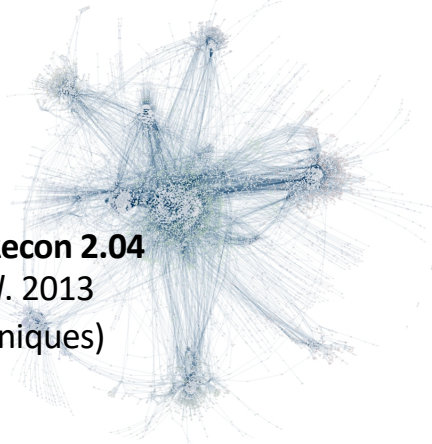
Human metabolic network Recon3D
Brunk E., et al. *Nat Biotechnol*. 2018.
8 399 metabolites
13 543 reactions
3 697 genes



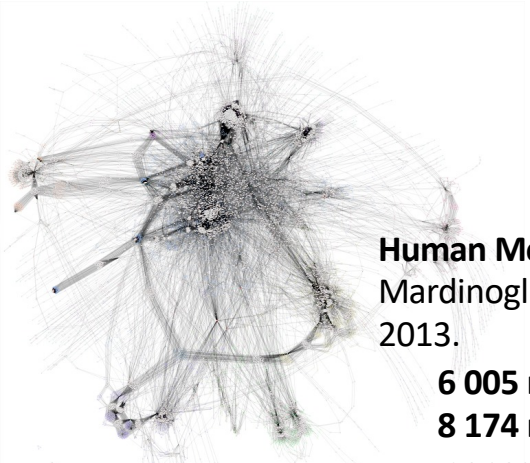
2007

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

MAPPING METABOLITES/LIPIDS IN NETWORKS

Metabolic model and lipidomics dataset used as example

Genome-scale metabolic network

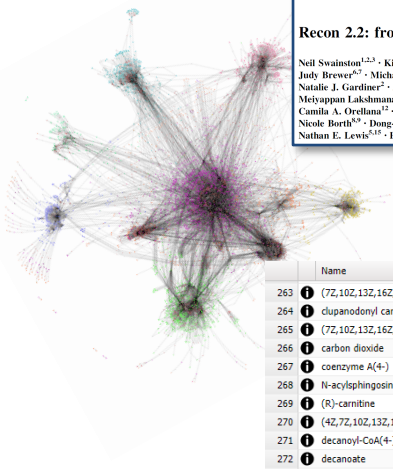
Human metabolic network Recon 2.2

Swainston et al. *Metabolomics* 2016

5324 metabolites (2652 uniques)

7785 reactions

1675 genes



Metabolomics (2016)12:109
DOI 10.1007/s11306-016-1051-4

SHORT COMMUNICATION

Recon 2.2: from reconstruction to model of human metabolism

Neil Swainston[1,2,3] · Kieran Smallbone[3] · Hooman Hefzi[4,5] · Paul D. Dobson[3] · Judy Brewer[6,7] · Michael Hanscho[8,9] · Daniel C. Zielinski[4] · Kok Siong Ang[10,11] · Natalie J. Gardiner[2] · Jahir M. Gutierrez[4,5] · Sarantos Kyriakopoulos[11] · Meiyappan Lakshmanan[11] · Shangzhong Li[4,5] · Joanne K. Liu[12] · Veronica S. Martínez[12] · Camila A. Orellana[12] · Lake-Ee Quek[12] · Alex Thomas[5,13] · Juergen Zanghellini[9] · Nicole Borth[8,9] · Dong-Yup Lee[10,11] · Lars K. Nielsen[12] · Douglas B. Kell[1,14] · Nathan E. Lewis[5,15] · Pedro Mendes[1,3,16]

		Name	Identifier
263		(7Z,10Z,13Z,16Z,19Z)-docosapentaenoyl-CoA(4-)	M_clpndcoa_c
264		clupanodonyl carnitine	M_clpndcrn_c
265		(7Z,10Z,13Z,16Z,19Z)-docosapentaenoic acid	M_clpnd_c
266		carbon dioxide	M_co2_c
267		coenzyme A(4-)	M_coa_c
268		N-acylsphingosine	M_crm_hs_c
269		(R)-carnitine	M_crn_c
270		(4Z,7Z,10Z,13Z,16Z,19Z)-docosahexaenoate	M_crvnc_c
271		decanoyl-CoA(4-)	M_dcacoa_c
272		decanoate	M_dca_c
273		docosa-4,7,10,13,16-pentaenoyl coenzyme A	M_dcsptn1coa_c
274		docosa-4,7,10,13,16-pentaenoyl carnitine	M_dcsptn1crn_c
275		(4Z,7Z,10Z,13Z,16Z)-docosa-4,7,10,13,16-pentaenoic acid	M_dcsptn1_c

#4311 in MetExplore

Lipidomics dataset

SCIENTIFIC REPORTS

OPEN

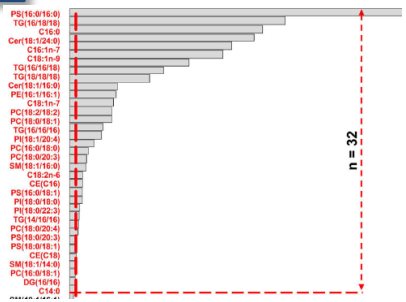
Metabolism dysregulation induces a specific lipid signature of nonalcoholic steatohepatitis in patients

Received: 15 November 2016
Accepted: 28 March 2017
Published: 24 April 2017

Franck Chiappini[1,2,3], Audrey Coilly[1,2,3,4,*], Hanane Kadar[5,*], Philippe Gual[6,7,8], Albert Tran[6,7,8], Christophe Desterke[9,10], Didier Samuel[1,2,3,4], Jean-Charles Duclos-Vallée[1,2,3,4], David Touboul[5], Justine Bertrand-Michel[11], Alain Brunelle[5], Catherine Guettier[1,2,3,12] & François Le Naour[1,2,3,9,10]

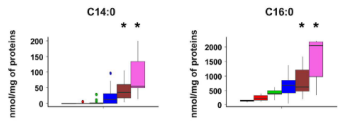
- Lipidomics (GC & LC-MS) analyses
- Human liver biopsies
- Healthy, 3 grades of NAFL, NASH

→ signature of **32 discriminating lipids = "NASH signature"**



Metabolite names

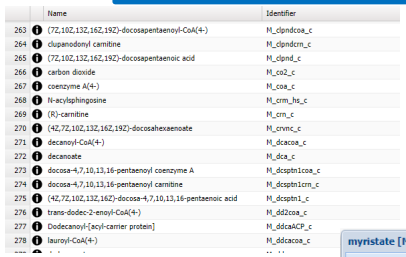
- C14:0
- C18:1n-9
- CE(C16)
- DG(16/16)
- TG(14/16/16)
- Cer(d18:1/16:0)
- PC(16:0/18:0)
- PE(16:1/16:1)
- PI(18:0/18:0)
- PS(16:0/16:0)
- SM(18:1/14:0)
- …



The metabolite identifier issue

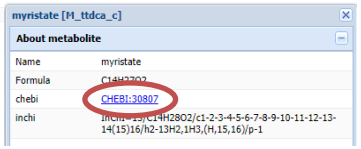
e.g. **C14:0 ↔ myristic acid**

Genome-scale metabolic network



List of **names** and **network-specific idenfiers** (**M_*met*_*cpt***)

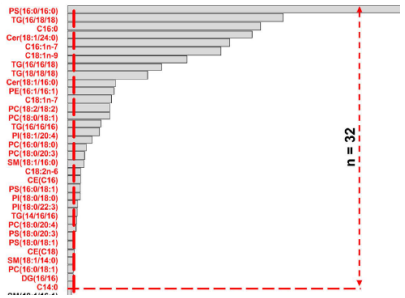
+ additional identifiers from public databases (LipidMAPS, ChEBI, HMDB …)



Lipidomics dataset

Metabolite names

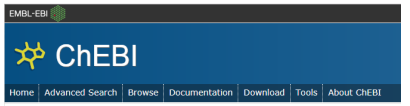
C14:0
C16:1n-7
CE(C16)
DG(16/16)
TG(14/16/16)
Cer(d18:1/16:0)
PC(16:0/18:0)
PE(16:1/16:1)
PI(18:0/18:0)
PS(16:0/16:0)
SM(18:1/14:0)
…



name	Identifier	CHEBI	LipidMaps
myristate	M_ttdca_c	30807	-

?

name
C14:0



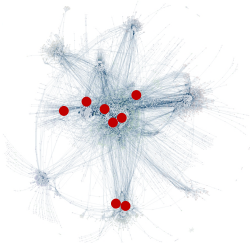
① finding relevant database identifiers (CHEBI)

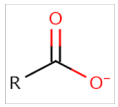
CHEBI	name
30807	C14:0

② matching with network metabolites

Network Identifier	Dataset name
M_ttdca_c	C14:0

③ mapping





ChEBI
Name: Fatty acid anion
ChEBI ID: 28868

is conjugate base of

ChEBI
Name: Fatty acid
ChEBI ID: 35366

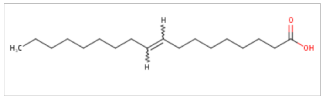
is a

ChEBI
Name: Long-chain fatty acid
ChEBI ID: 15904

is a

ChEBI
Name: Octadecenoic acid
ChEBI ID: 25634

is a



ChEBI
Name: Octadec-9-enoic acid
ChEBI ID: 36021

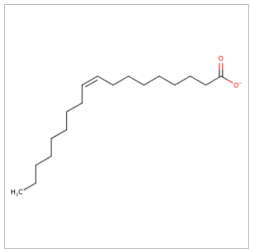
is a is a

ChEBI
Name: Oleic acid
ChEBI ID: 16196

ChEBI
Name: Elaidic acid
ChEBI ID: 27997

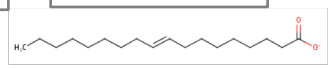
is conjugate base of

is conjugate base of



ChEBI
Name: Oleate
ChEBI ID: 30823

ChEBI
Name: Elaidate
ChEBI ID: 30825



Recon 2.2
Name: Fatty acid anion
Recon ID: M_Rtotal
ChEBI: 28868

Distance: 3.1

Lipidomics
Name: C18:1n-9
ChEBI ID: 36021

No exact network match

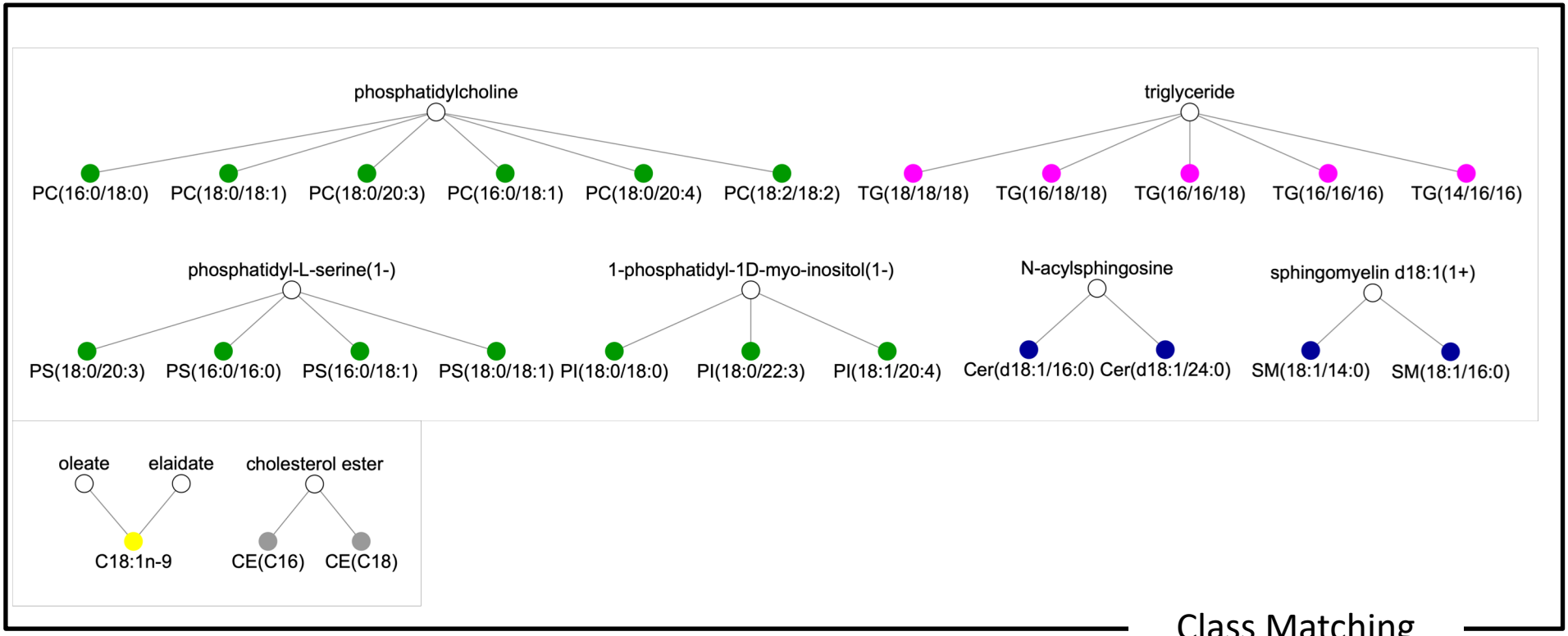
Network match
Distance: -1.1

Network match
Distance: -1.1

Recon 2.2
Name: Oleate
Recon ID: M_ocdcea
ChEBI: 30823

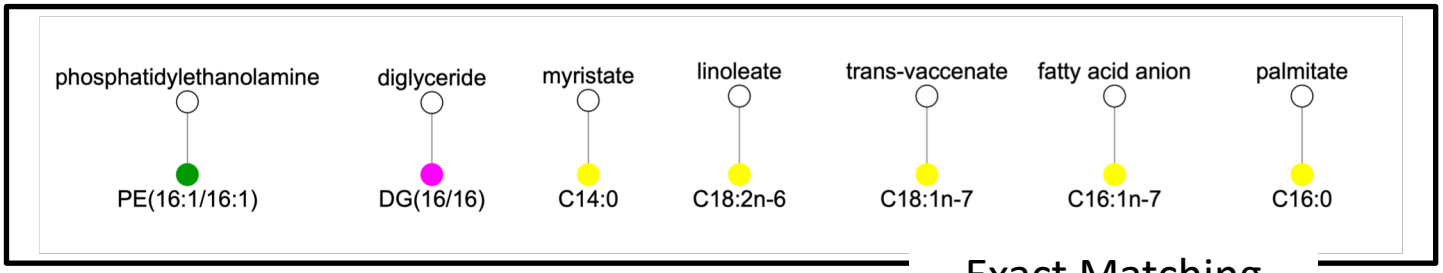
Recon 2.2
Name: Elaidate
Recon ID: M_elaid
ChEBI: 30825

Class network matching



Class Matching

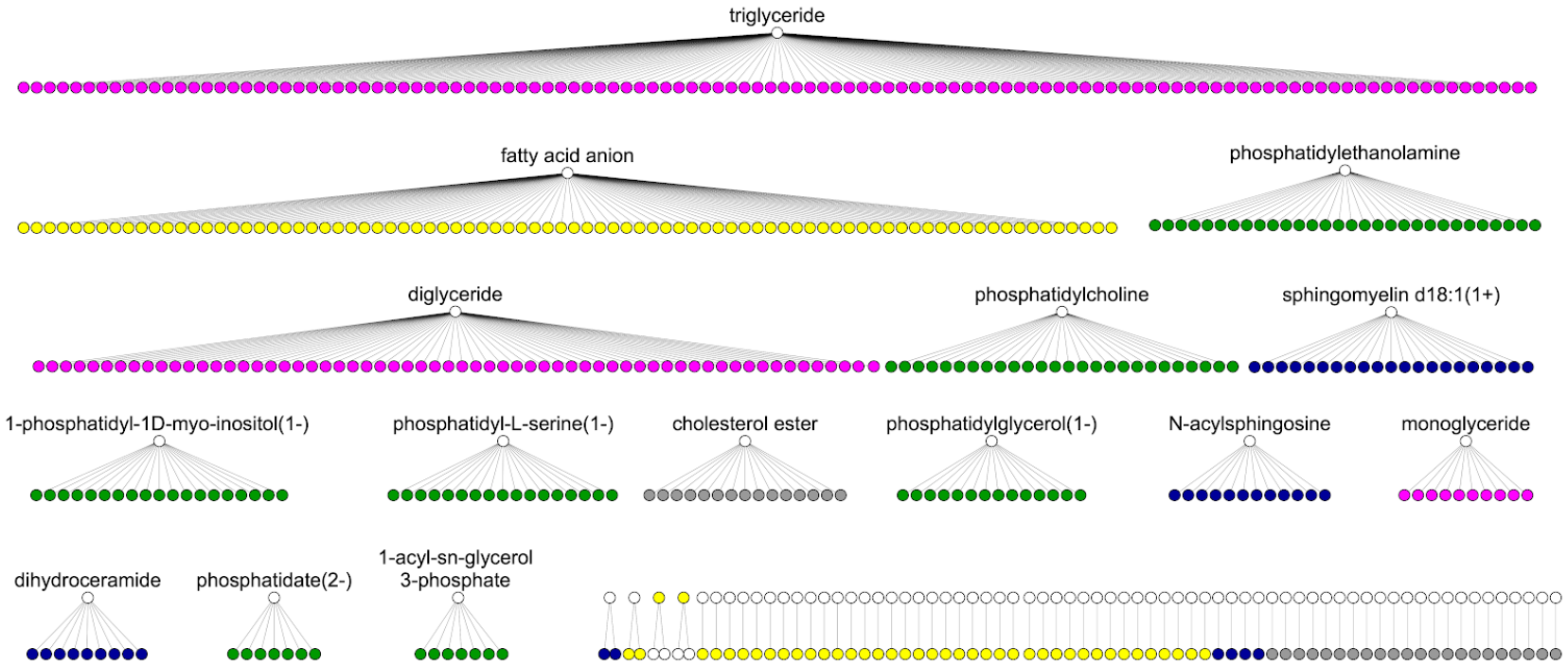
- network
- Fatty acids
- Glycerophospholipids
- Glycerolipids
- Sterol lipids
- Sphyngolipids



Exact Matching

Available in MetExplore

Lipidomic library mapping



Mapping of all lipids which can be measured by MetaboHub lipidomics facilities

METABOHUB

MetExplore : omics data analysis in genome scale networks

10 YEARS

INRA SCIENCE & IMPACT

METABOHUB

medDay PHARMACEUTICALS

PhenoMeNal

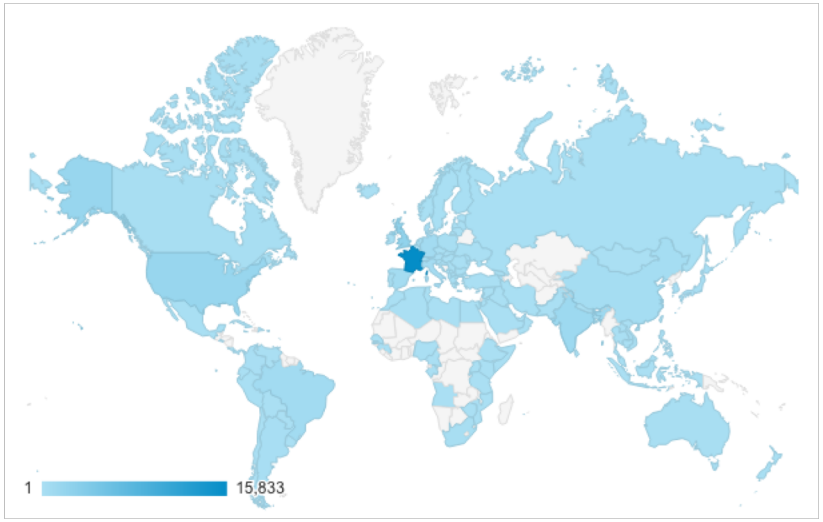
The project

- *Publications* :
  - Cottret et al (2018). Nucleic Acids Research
  - Chazalviel et al (2017). Bioinformatics
  - Frainay et al (2018). Bioinformatics
- *Number of citations*: >140
- *Metrics*:
  - > **800** registered users,
  - **> 1400** networks
  - **> 600** persons trained
- Involed in several national and EU grants
- 1 industrial partner (MedDay pharma)
- part of ELIXIR-FR Service Delivery Plan

elixir

Website

http://www.metexplore.fr/



INRA SCIENCE & IMPACT

MeX

genotoul metatoul

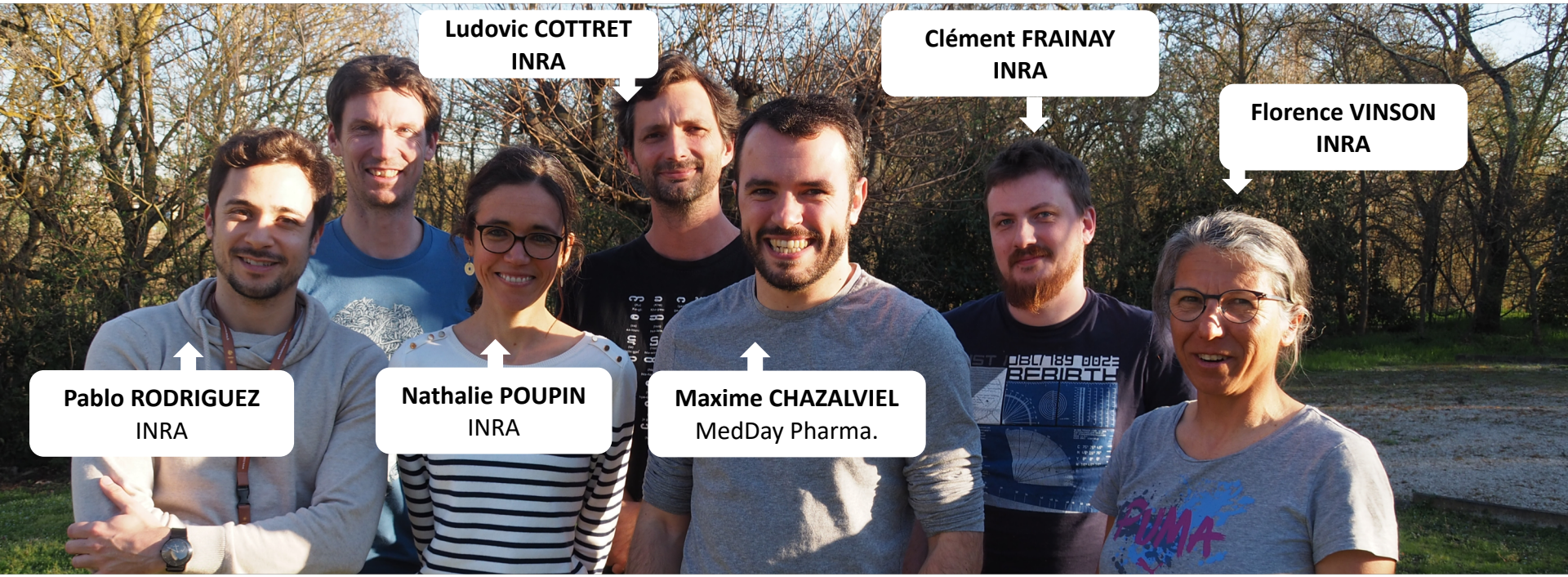
Work in progress

- Python library to perform ontology mapping (done)
- Implementation in MetExplore (done)
- Publication (on its way)
- Future directions: allows same thing on other ontologies (e.g. Lipidmaps)

Conclusion

- In most networks we lack standardised identifiers
- Metabolomics data often come with no identifier
- Suggestion:
  - Have on both sides two kind of identifiers:
    - Chemistry based identifiers (SMILES, INCHI)
    - Ontology based identifiers (ChEBI)
- Not only usefull for mapping, we also need identifiers for computations (e.g. atom mapping in reactions)

- Expected role of the CATI on this aspect:
  - Ensure this inter-operability upward by providing relevant identifiers


Ludovic COTTRET
INRA
Clément FRAINAY
INRA
Florence VINSON
INRA
Pablo RODRIGUEZ
INRA
Nathalie POUPIN
INRA
Maxime CHAZALVIEL
MedDay Pharma.