METABOLOMICS

Coral Barbas
Danuta Dudzik
Mª Fernanda Rey-Stolle
Francisco J. Rupérez
Antonia Garcia

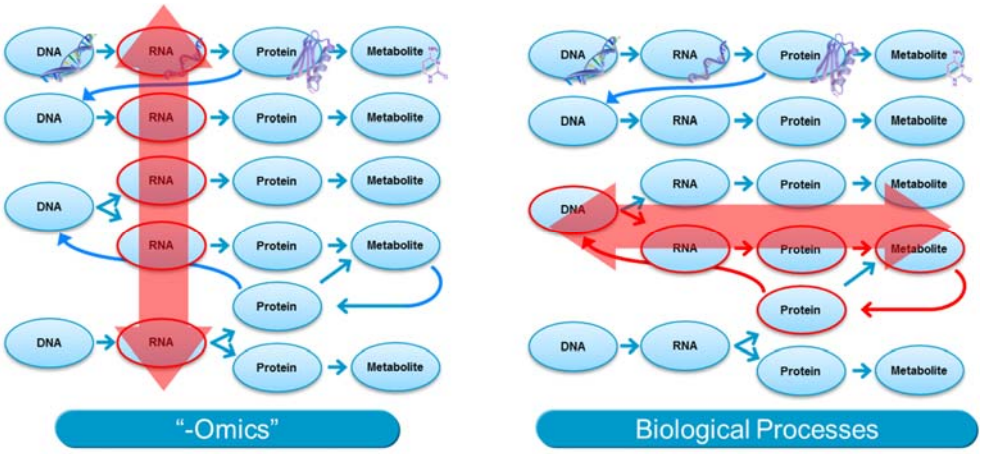
SUMMARY

1. Introduction to metabolomics
2. Analytical approaches in metabolomics
   - Workflow of the metabolomics study
   - Quality Control and Quality Assurance Procedure in Metabolomics
3. Data processing and identification of metabolites
   - Data processing pipeline
   - Non-targeted metabolomics data treatment
   - Metabolite identification
   - Statistical analysis
4. Data analysis
   - From data identification to pathways
   - Biomarker validation
5. Practical sessions
   - Targeted and non-targeted metabolomics
   - Metabolomics with free online tools

Metabolomics

New emerging field of "omics" research (which includes genomics, proteomics and metabolomics) concerned with comprehensive characterization of the small molecule metabolites present in biological systems.

Omics & Systems Biology

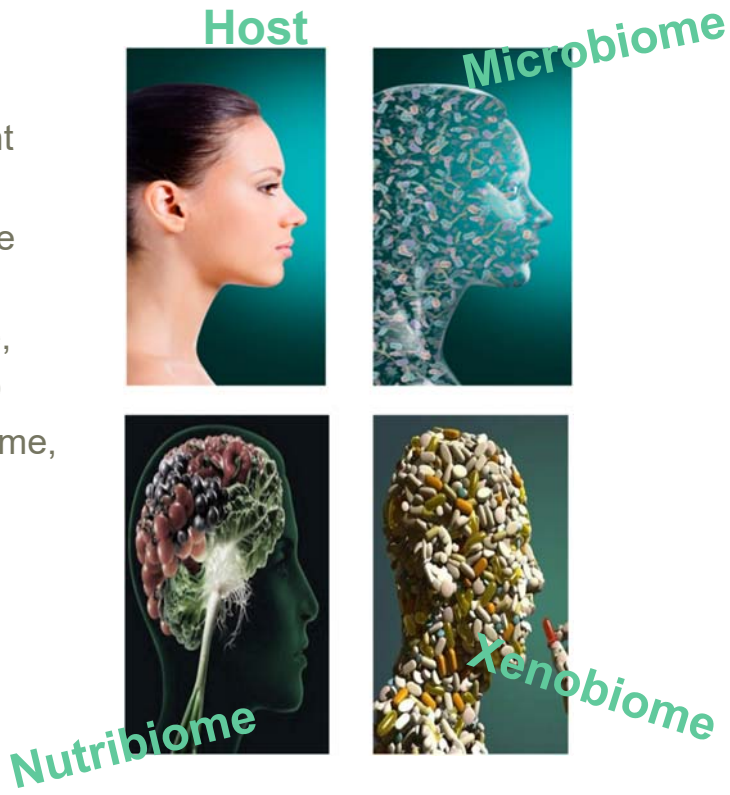


Addendum: Definition of Metabonomics

- Measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification (Nicholson, 1999)
  - quantitative measurement of the time-related "total" metabolic response to pathophysiological (nutritional, xenobiotic, surgical or toxic) stimuli
- MetaboLomics - the picture, MetaboNomics – the movie
- Nowadays, everything is Metabolomics

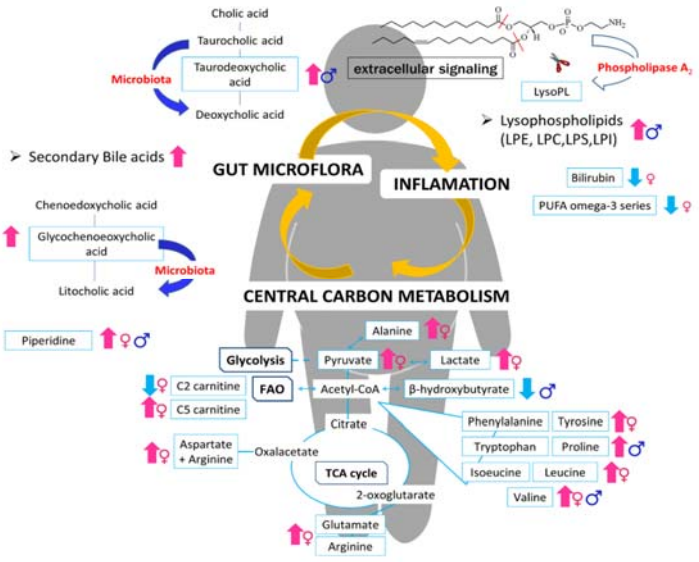
Definition of Metabolome

- "...the **complete set of metabolites**/low-molecular-weight intermediates, which are context dependent, varying according to the physiology, developmental or pathological state of the cell, tissue, organ or organism…" (Oliver 2002)
- Origin: Endometabolome, Microbiome, Xenobiome, Nutribiome…
- Nature: Glycome, lipidome, sphingolipidome, peptidome…
- Metabolome ↔ Phenotype



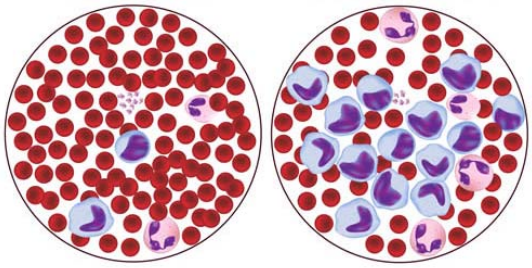
What metabolomics can provide (I)

- Overview of the metabolic status and global biochemical events associated with a cellular or biological system.
  - Pathological situations without known mechanism, i.e. relationship between obesity and insulin resistance



What metabolomics can provide (II)

- Identification (proposal) of new biomarkers, important in the process of new drug discovery or as in vitro diagnostics tools.
  - For instance, new diagnostic biomarkers for aggressiveness in chronic lymphatic leukemia



Utility of validated metabolites as biomarkers of aggressive state of CLL

Metabolite	AUC	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Acetylcarnitine	0.695	43.2	93.0	86.1	62.1
Butyrylcarnitine	0.548	10.8	98.0	84.4	52.4
Hexanoylcarnitine	0.690	27.0	96.0	87.1	56.8
Octanoylcarnitine	0.651	29.7	95.0	85.6	57.5
Decanoylcarnitine	0.662	27.0	94.0	81.8	56.3
Palmitoylcarnitine	0.719	40.5	94.0	87.1	61.2
Dodecanamide	0.497	8.1	100.0	100.0	52.1
Hexadecanamide	0.516	5.4	100.0	100.0	51.4
Oleamide	0.600	18.9	96.0	82.5	54.2
Linoleamide	0.672	16.2	98.0	89.0	53.9
Acylcarnitines[a]	0.743	32.4	95.0	86.6	58.4
FAA[b]	0.662	13.9	96.0	77.6	52.7
Acylcarnitines and FAA	0.750	54.0	89.0	83.1	65.9

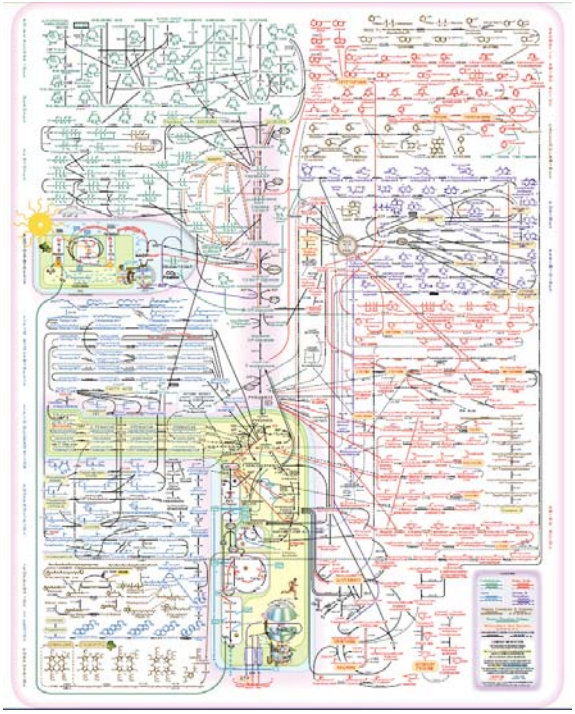
What metabolomics good for…..

- searching for metabolic differences between groups of samples (case vs control; before vs after treatment; One condition vs another)
- identifying compounds that are significant and proposing the mechanisms
- finding out information about the phenotype
- observing the effects of a treatment
- finding new drug targets

What is metabolomics NOT….

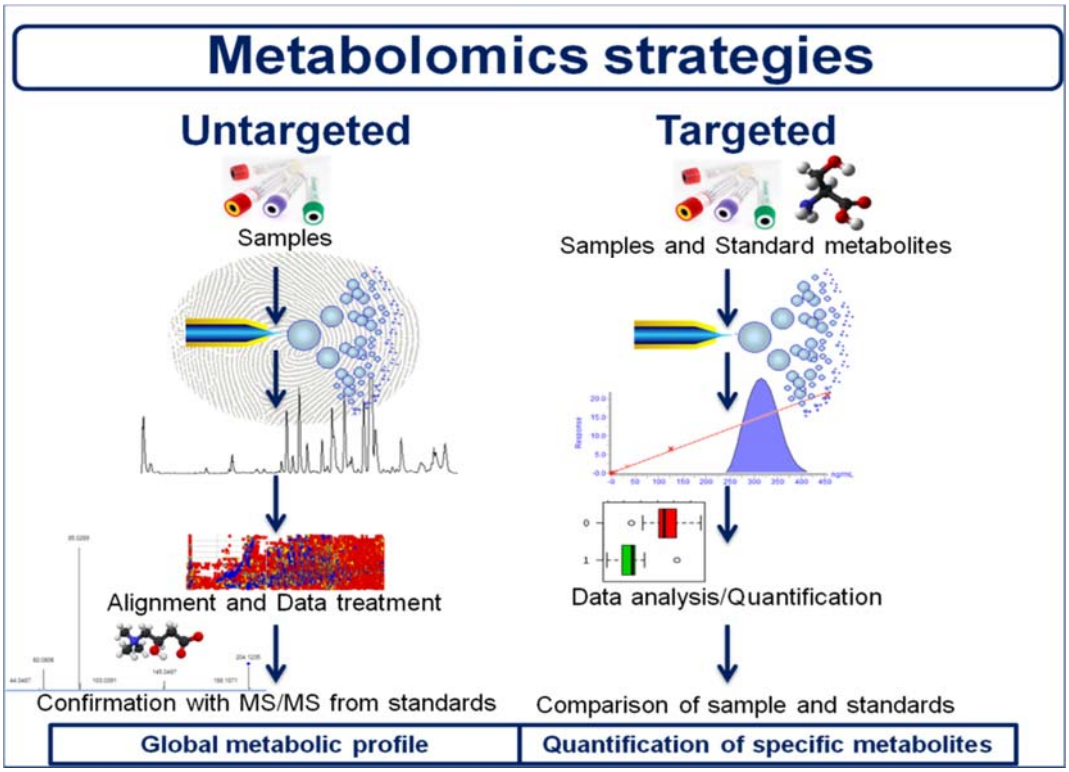
- a method to reveal the fate of a metabolite or drug
- a method for quantification
- the use of a simple kit to quantify a grouop of metabolites (it requires NMR, MS…)
- Possible without simultaneous comparison of samples

Definition of Metabolism

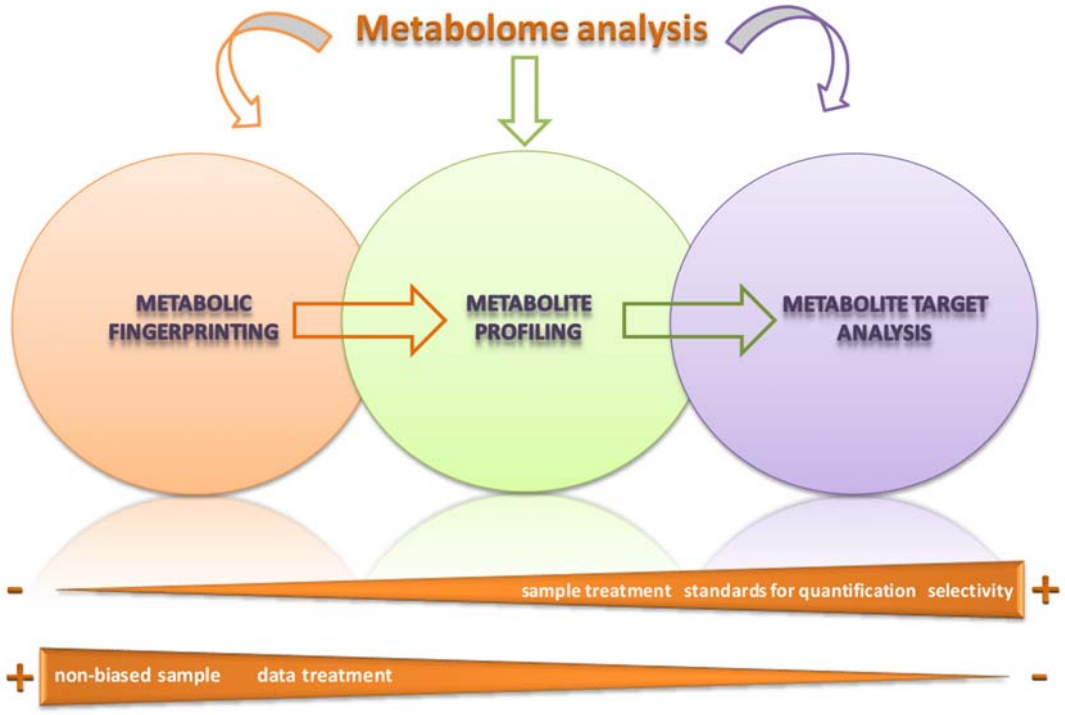


The complete group of (bio)chemical processes within an organelle, cell, tissue, organ or organism, essential for life

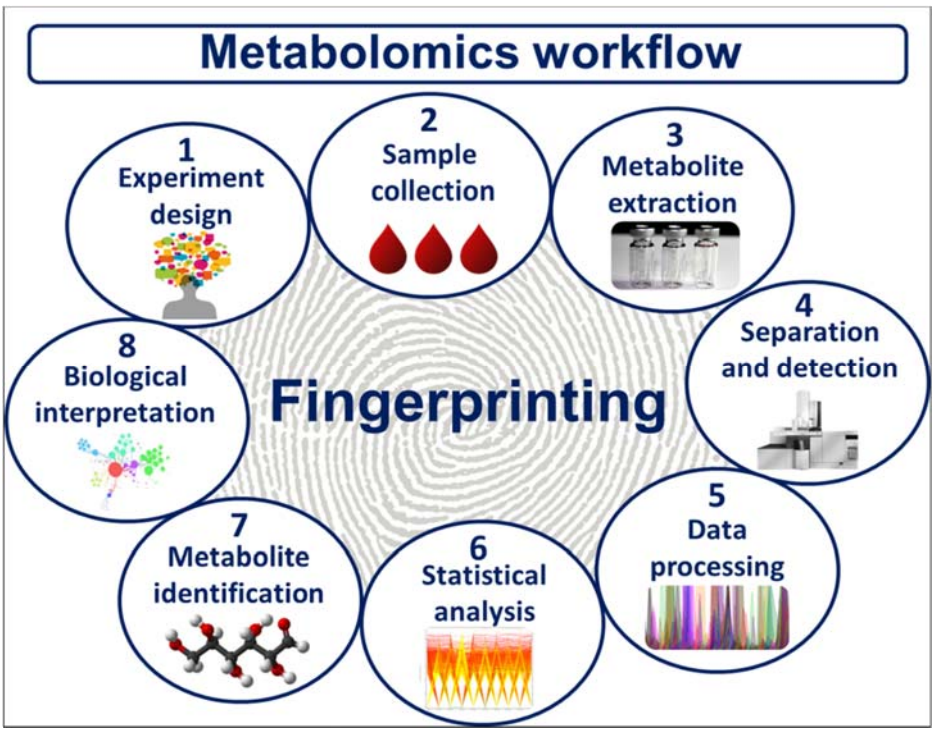
Analytical approaches in metabolomics



Three ways to do metabolomics



WORKFLOW



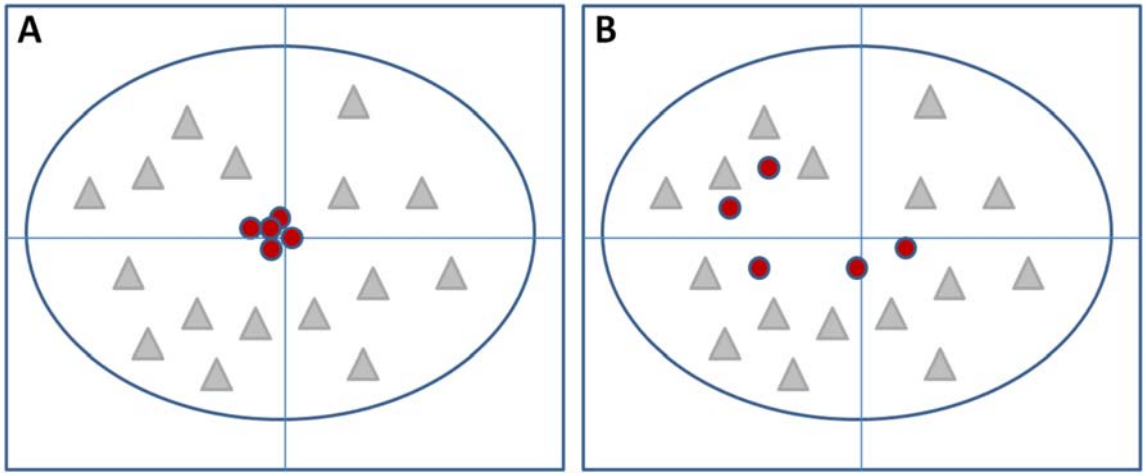
ANALYTICAL TECHNIQUES

- GC/MS: Small polar compounds
  - Mainly water soluble (some hydrophobic)
  - Sample treatment: Derivatization
  - Fragmentation reproducible - databases
- NMR
  - Water-soluble
  - Virtually no sample treatment
  - High LOD
- LC/MS
  - from small to large (<1500 Da) medium to non-polar metabolites
- CE/MS: Small-medium polar compounds
  - Amino acids, acylcarnitines, polyamines, etc.
  - No derivatization

MS Based analytical platforms in metabolomics

Analysis Technique	Application	Advantages	Disadvantages
GC-MS	Separation, identification, and quantification of volatile and thermally stable less polar metabolites.	High chromatographic resolution, availability of large spectrum libraries for metabolites identification.	Inability to analyze thermo-labile and high molecular weight metabolites, the requirement of derivatization for non-volatile metabolites.
LC-MS	Separation, identification, and quantification of very broad groups of metabolites, depending on the type of column and mobile phase.	High sensitivity, large sample capacity, derivatization not required, ability to analyze thermo-labile compounds.	Limited availability of commercial libraries, restriction on LC eluents, matrix effect, limited potential in identification unless an MS-MS technique is used.
CE-MS	Separation, identification, and quantification of polar and ionized metabolites, using reduced sample volumes.	High resolution and rapid analysis, utility for complex biological samples, even if in a small volume.	Limited availability of commercial libraries. Buffer incompatibility, detection limits. Limited potential for identification unless an MS-MS technique is used.

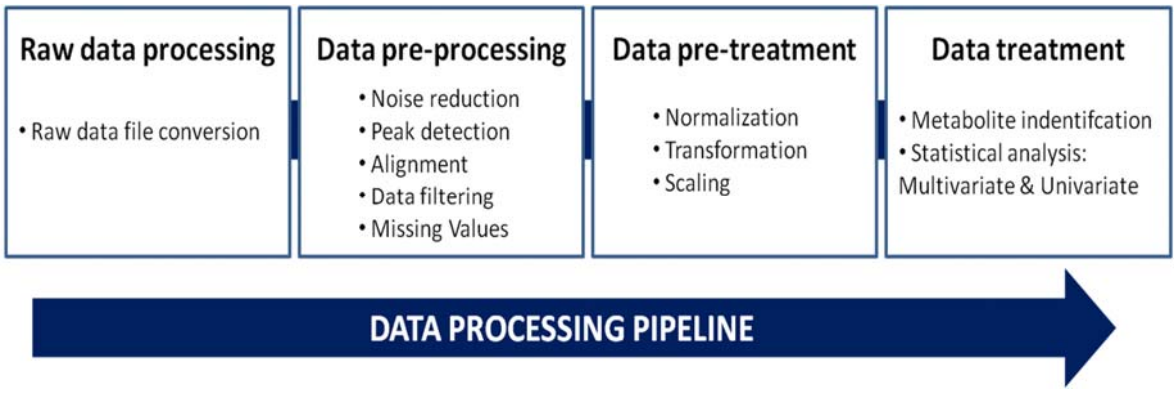
Quality Control and Quality Assurance Procedure in Metabolomics



A: QCs (red dots) clustered together

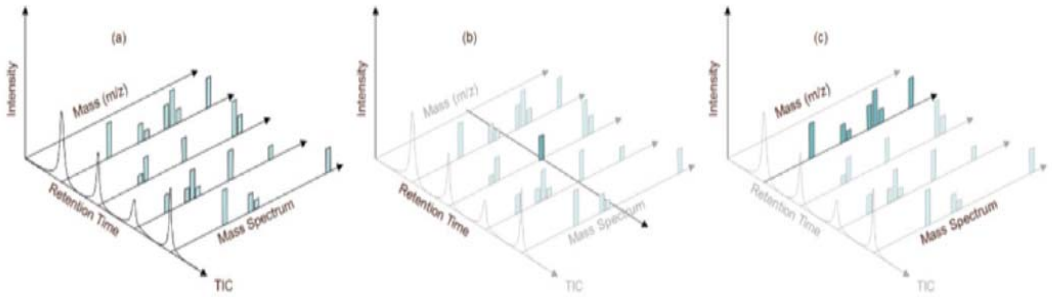
B: QCs spreaded

DATA TREATMENT IN METABOLOMICS: Signal Processing



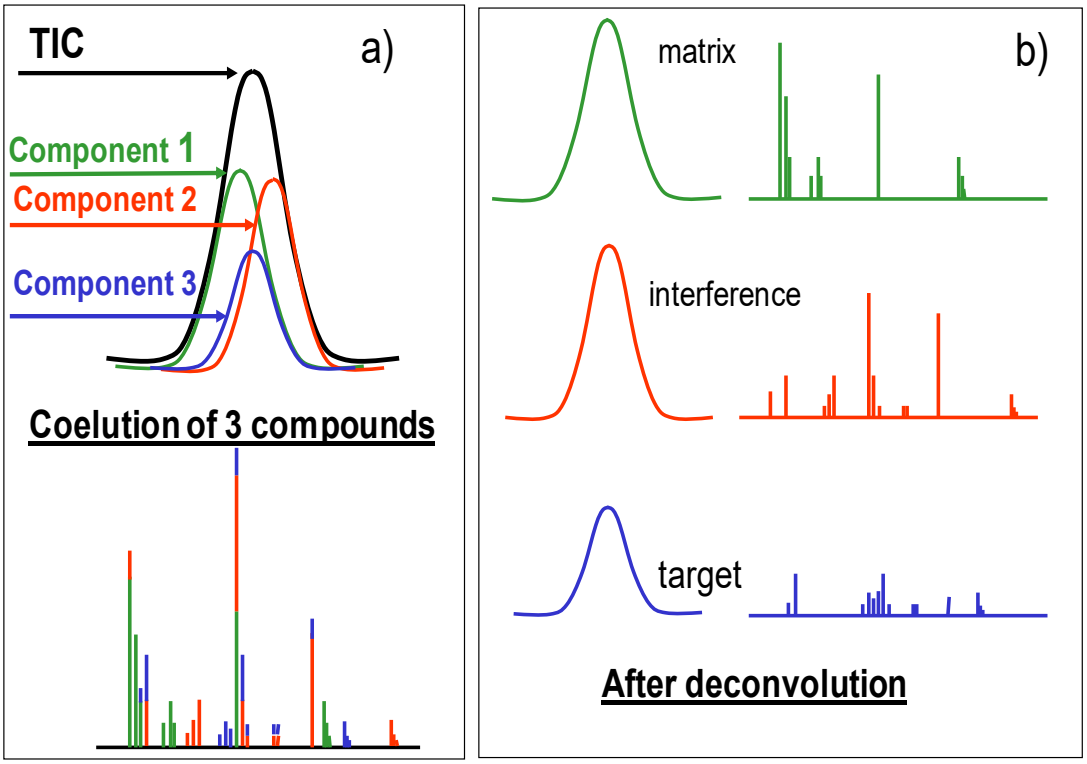
ANALYTICAL TECHNIQUE: GC-MS

- Gas chromatography coupled to mass spectrometry
- Gold standard
  - Highly sensitive and reproducible
  - Information: Quality and Quantity
  - Spectrum libraries for identification purposes
  - 10-20% of the known compounds can be analyzed by GC
    - High metabolic relevance



(a) 3D Data of GC/MS, (b) Extracted Ion chromatogram for the selected ion
(c) A single data point in time gives a single mass spectrum
adapted from Chromatography today

Deconvolution

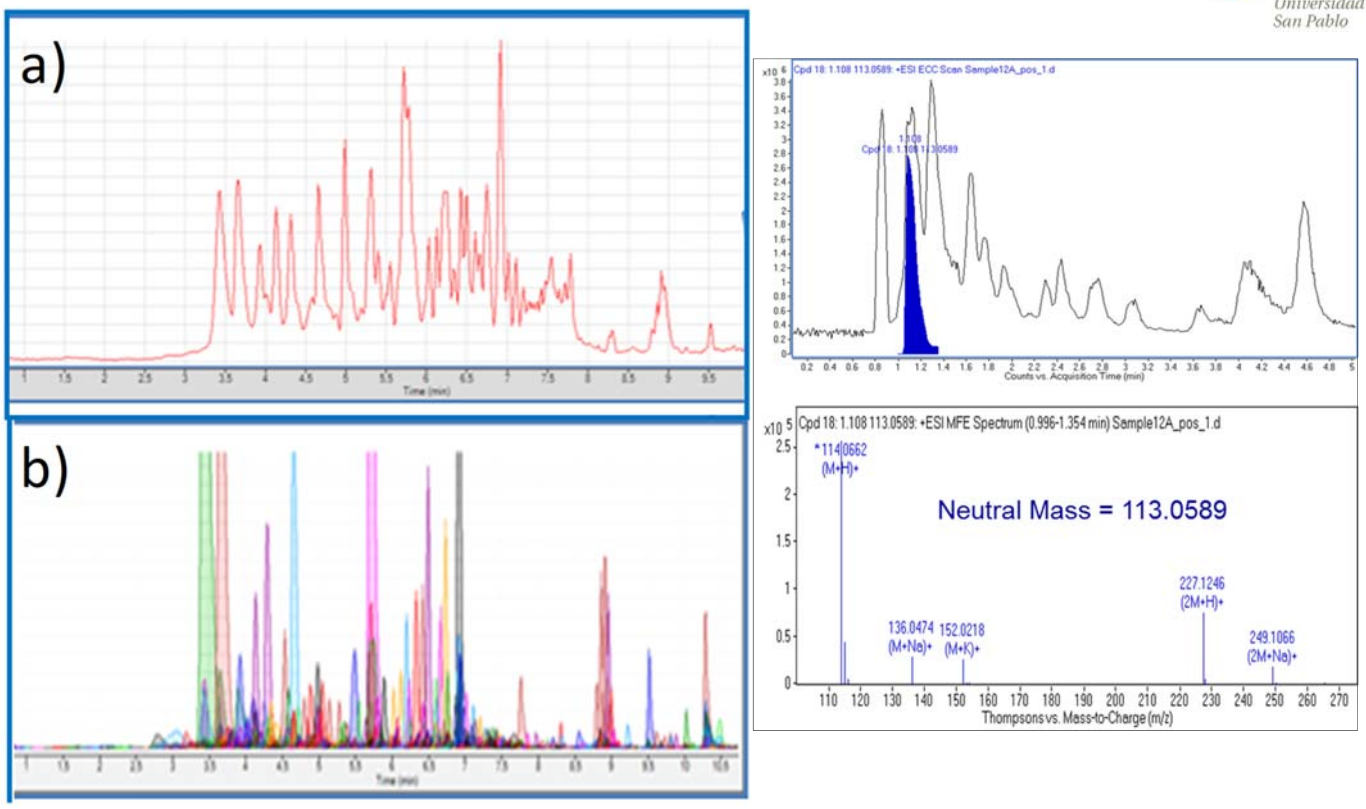


a) Before and b) After the deconvolution process
adapted from https://www.agilent.com/cs/library/Support/Documents/f05017.pdf

Deconvolution in LC-ESI-MS and CE-ESI-MS

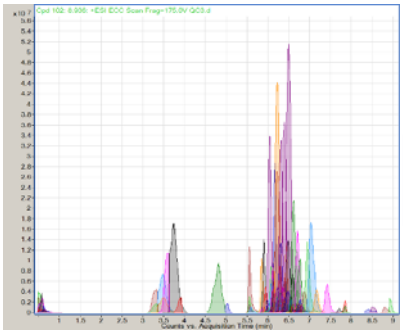
- Peak-based methods
- Molecular Feature Extractor (Agilent) considers the accuracy of the mass measurements to group related ions by charge-state envelope, isotopic distribution, and possible chemical relationships when determining whether different ions are from the same metabolic feature.
- It can consider also related ions like adducts: proton, sodium, potassium and ammonia adducts in positive ionization or loss of a proton, adducts with formate, etc. in negative ionization mode.

After Deconvolution



a) Total Ion Chromatogram
b) Chromatograms from every single compound obtained after deconvolution

Chromatogram or features list?



Automatically Show Columns

Show/Hide	Saturated	RT	m/z	Mass	Polarity	Ions	Height	Area	Vol	Quality Score
☑	S	0.514	280.0923	279.085	Positive	5	2740215		14234373	100
☑	S	0.517	203.0526	202.0454	Positive	6	4045612		20286412	80
☑		0.526	140.0682	139.0609	Positive	3	1530642		7519633	100
☑		0.529	136.0482	135.041	Positive	2	1187260		6016909	100
☑	S	0.57	162.1126	161.1053	Positive	7	2758926		19465836	100
☑	S	0.57	304.2998	303.2926	Positive	4	3021606		14360408	100
☑		0.58	114.0664	113.0591	Positive	4	549599		6985240	80
☑		0.614	175.1192	174.1119	Positive	3	760396		3860452	91.6
☑		0.625	156.0768	155.0696	Positive	7	1055485		6005360	100
☑		0.646	170.0927	169.0854	Positive	4	604901		3355886	100
☑	S	3.29	520.3389	519.3316	Positive	7	3243984		36560884	80.2
☑		3.298	544.3388	543.3315	Positive	4	1339590		12508261	87
☑	S	3.483	520.3389	519.3316	Positive	11	5382675		86088232	85.2
☑	S	3.5	544.3389	543.3316	Positive	7	2216628		36501788	100
☑	S	3.573	496.3389	495.3316	Positive	6	8281163		80624664	86.1
☑	S	3.746	496.3388	495.3316	Positive	13	11742118		200075408	80
☑	S	3.89	522.3545	521.3472	Positive	8	2204035		17763280	87
☑	S	4.801	524.3702	523.3629	Positive	12	6862920		111854136	100
☑		5.027	524.3702	523.3629	Positive	6	1308318		8748089	87
☑		5.559	163.0393	162.032	Positive	4	1349239		5379156	100
☑	S	5.561	391.2839	390.2767	Positive	15	7367381		51796644	100
☑	S	5.561	149.0235	148.0162	Positive	4	2915086		12886771	100
☑		5.597	338.3418	337.3345	Positive	9	787867		4058574	100
☑	S	5.849	675.5425	674.5353	Positive	7	6821314		39496860	100
☑		5.857	627.5339	626.5266	Positive	3	712763		3673685	100
☑	S	5.904	701.5583	700.551	Positive	8	8615328		59468044	100
☑	S	5.941	689.558	688.5507	Positive	8	3178377		18867332	100
☑		5.991	754.5372	753.5299	Positive	4	737180		4127601	100
☑		5.995	730.5369	729.5297	Positive	7	1506513		8484420	100

Chromatogram Results | MS Spectrum

Data preprocesing

- Alignment
  - Peak shifts are observed across the RT axis
  - Two groups:
    - data are aligned before peak detection
    - peak-based alignment methods: detected spectral peaks are aligned across samples.
    - softwares:
      - MetaboAnalyst (metaboanalyst.ca)
      - mzmine and mzmine2 (http://mzmine.sourceforge.net/)
      - metAlign
      - BinBase (fiehnlab.ucdavis.edu)
      - xcms and xcms2 (Scripps)
      - metaXCMS (Scripps)
      - XCMS Online (Scripps)
- Missing values
  - Problems in further analysis
  - Different strategies
    - Replace by the half of the minimum, by mean/median, k-nearest neighbour (KNN), probabilistic PCA (PPCA), Bayesian PCA (BPCA) method, Singular Value Decomposition (SVD) …
- Filtering
  - Variables of very small values - detected using mean or median
  - Variables that are near-constant - detected using standard deviation (SD)
  - Variables that show low repeatability - measured using QC sample

Data pretreatment

- Normalization
  - **Sample-specific** normalization (i.e. weight, volume)
  - Normalization by **sum** or **median**
  - Normalization by reference **sample**
  - Normalization by a **pooled** sample from group control
  - Normalization by reference **feature**
  - **Quantile** normalization
- Data transformation
  - **Log** transformation
  - **Cube root** transformation
- Data scaling
  - **Mean** centering
  - **Auto scaling** (mean-centered and divided by the standard deviation of each variable)
  - **Pareto scaling** (mean-centered and divided by the square root of standard deviation of each variable)
  - **Range scaling** (mean-centered and divided by the range of each variable)

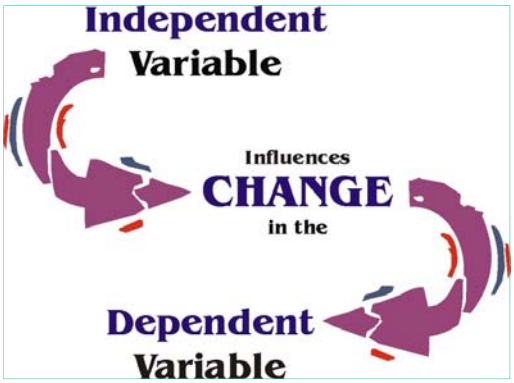
Statistics for Metabolomics

AIMS to:

- detect differences between sample groups at the chemical level
- rank compounds by relative importance for sample differentiation

VARIABLES

- **dependent variable**: represents the output or effect, or is tested to see if there is effect, e.g.: abundance of metabolite
- **independent variable**: represents the inputs or causes, or are tested to see if they are the causes, e.g.: treatment conditions within the experiment

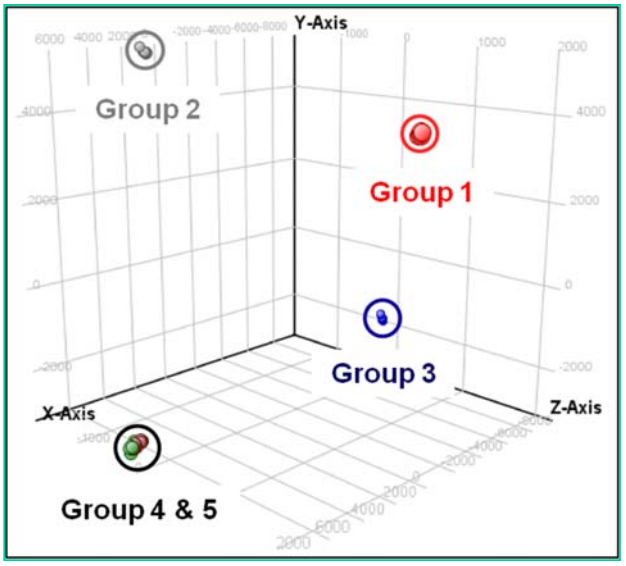


TYPES

- **Univariate analysis UVA:**
  - Normal distribution: Student's *t*-Test, ANOVA,
  - Non-normal distribution: Mann-Whitney U-Test, Kruskal-Wallis
- **Multivariate analysis MVA:** PCA, PLS-DA, OPLSDA

PCA

- used as a tool in exploratory data analysis
- each dot graphically represents each sample measured
- the algorithm has no knowledge of the group associations of the samples – *unsupervised* analysis
- first principal component explains most of the variance
- compound loadings indicate the impact of that compound on the analysis
- each dot is the sum of the compound loadings for a sample
- the tightness of the clustering reflects the variance of the samples



Class prediction

an algorithm using past data to predict the results of future observations

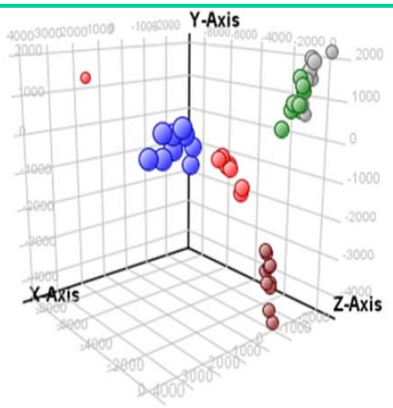
- the algorithm has knowledge of the group associations of the samples – *supervised* analysis
- common algorithms
  - **Partial Least Squares Discriminate Analysis (PLS-DA)**
  - Support Vector Machine
  - Decision Tree
  - Naïve Bayes
  - Neural Network

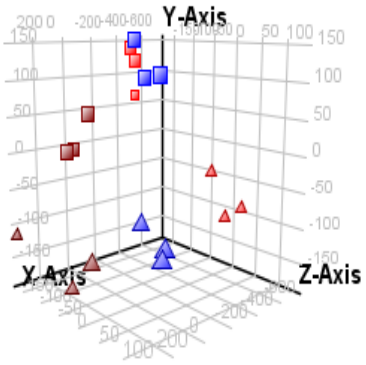
Class prediction: PLS-DA

Partial Least Square - Discriminant Analysis
Projection on Latent Structures - Discriminant Analysis

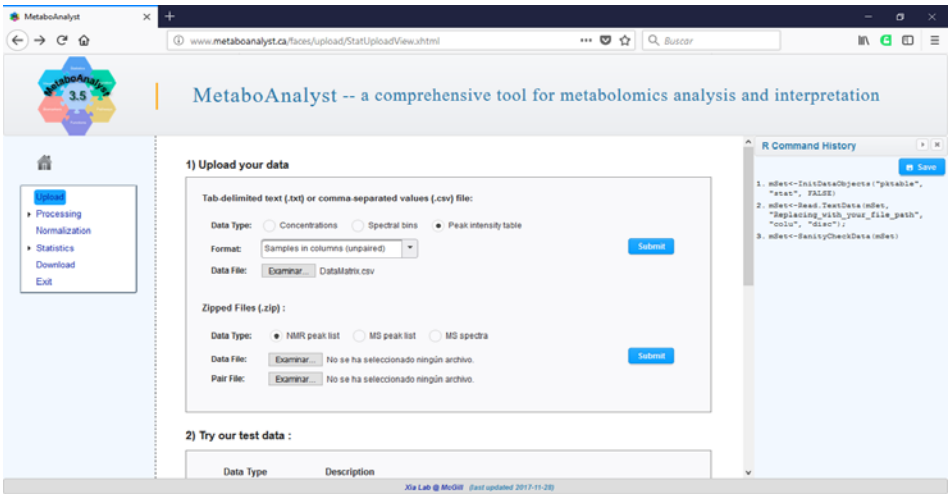
a statistical method that bears some relation to principal components analysis (PCA) but is a *supervised* analysis

- creates a linear regression model by projecting the predicted and observable variables to a new space
- well suited when there are more predictors (compounds) than observations (samples)
- each compound has a t-score that represents its impact on the prediction
- a prediction confidence value is assigned when the model is run





Univariate and Multivariate Statistical Analysis



Class prediction: Validate the model

assesses accuracy of prediction rule that is built and provides an indication of over-fitting models:

leave one out

- all samples in the training set except one is used to build the prediction rule
- using this rule, the class of sample that was left out is predicted
- the sample is returned to the training set while a different sample is left out and the prediction rule is built with remaining samples
- this process is repeated until each sample in training set has been predicted exactly once
- the number of correct and incorrect predictions is then tallied to determine the success rate

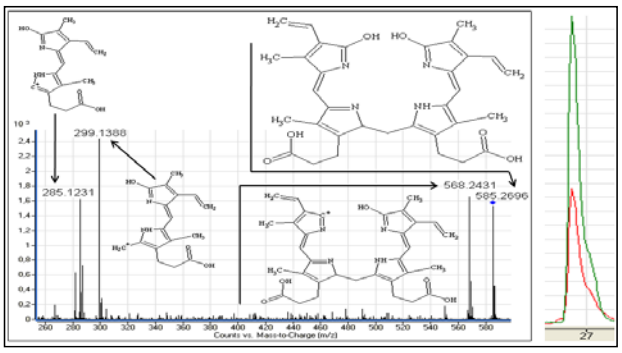
N - fold

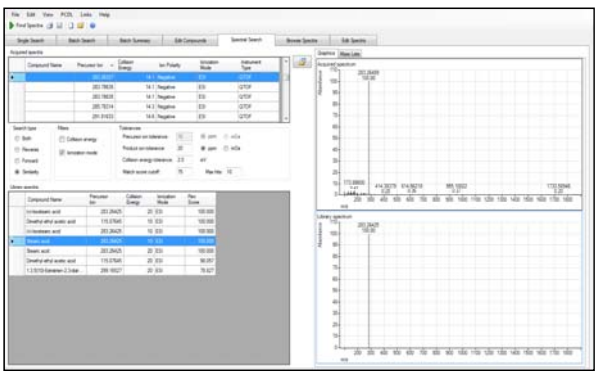
1. samples in the training set are randomly divided into *N* equals subsets, maintaining relative classes frequency
2. *N*-1 subsets are then combined for training and the remaining set is used for testing
3. repeat step 2 step with each group left out in turn
4. repeat step 1, 2, 3 *M* times
5. each sample gets predicted *M* times and majority class predicted over these *M* times is reported in validation results

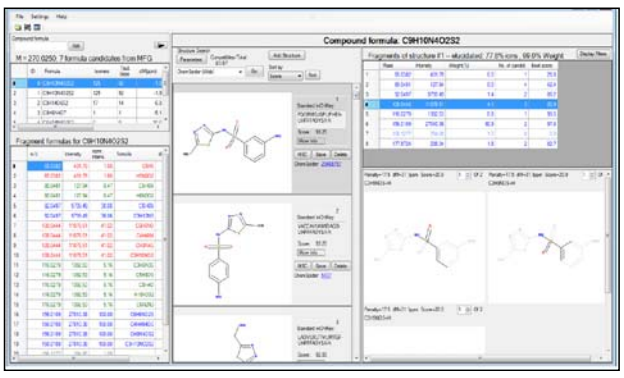
Identification

1. Database matching using accurate mass measurement
2. Database matching with isotope pattern matching
3. Database matching with isotope pattern matching and retention time
4. MS/MS library matching
5. MS/MS library and retention time matching

Confidence







DATABASE CLASIFICATIONS

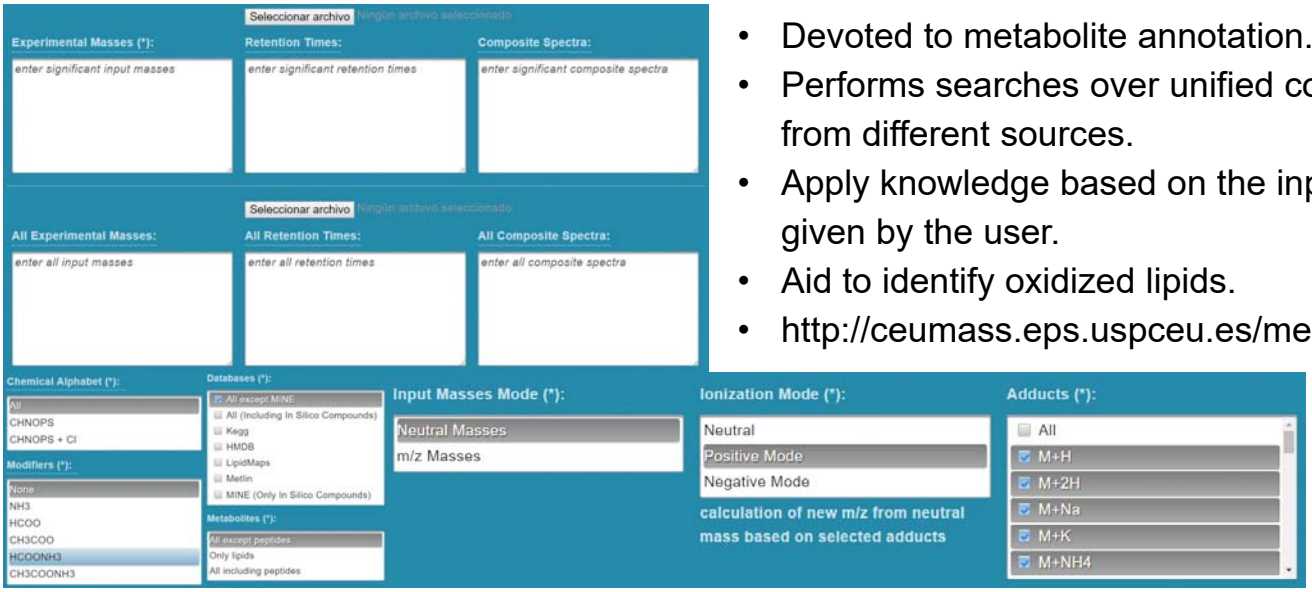
- **Based on Spectral input**
  - Mainly small molecules and not only metabolites
  - NMR
  - MS or MS/MS
- **Based on compound information**
  - Compound name, structures, physical properties, identification
- **Based on Metabolic pathway database**
  - Metabolites, xenobiotics, proteins, signal pathways
- **Complete Metabolomic database**
  - A combination of the previous ones

Database List in 2018

Name	URL	Name	URL
ARALIP	http://aralip.plantbiology.msu.edu/pathways/pathways	KEGG	http://prime.psc.riken.jp/?action=metabolites_index
AtIPD	http://www.atipd.ethz.ch/	KEGG Glycan	http://www.genome.jp/kegg/glycan/
BiGG	http://bigg.ucsd.edu/	KNApSAcK	http://prime.psc.riken.jp/?action=metabolites_index
BioCyc	http://biocyc.org/	LipidMaps	http://www.lipidmaps.org/
BioNumbers	http://bionumbers.hms.harvard.edu/	MarkerDB	http://www.markerdb.ca/users/sign_in
BML-NMR	http://www.bml-nmr.org/	MassBank	http://www.massbank.jp/
BioMagResBank	http://www.bmrb.wisc.edu/metabolomics/	MetaboAnalyst	http://www.metaboanalyst.ca/MetaboAnalyst/
BMDB	http://www.cowmetdb.ca/cgi-bin/browse.cgi	MetaboLights	http://www.ebi.ac.uk/metabolights/index
ChEBI	http://www.ebi.ac.uk/chebi/	MetaCrop	http://metacrop.ipk-gatersleben.de/apex/f?p=269:111:
ChEMBL	https://www.ebi.ac.uk/chembl/about#	MetaCyc	http://metacyc.org/
ChEBI	http://www.ebi.ac.uk/chebi/	METAGENE	http://www.metagene.de/program/a.prg
ChemMine	http://chemminedb.ucr.edu/	METLIN	https://metlin.scripps.edu/index.php
ChemSpider	http://www.chemspider.com/	MMCD	http://mmcd.nmrfam.wisc.edu/
CCD	http://ccd.chemnetbase.com/intro/index.jsp#about	mzCloud	https://mzcloud.org/
CSF Metabolome Database	http://www.csfmetabolome.ca/	OMIM	http://www.ncbi.nlm.nih.gov/omim/
CyberCell Database	http://ccdb.wishartlab.com/CCDB/	OMMBID	http://ommbid.mhmedical.com/
DrugBank	http://www.drugbank.ca/	Oryzabase	http://www.shigen.nig.ac.jp/rice/oryzabase/
ECMDB	http://www.ecmdb.ca/	PepBank	http://pepbank.mgh.harvard.edu/
ExPaSy Pathways	http://web.expasy.org/pathways/	PharmGKB	http://www.pharmgkb.org/
Fiehn GC-MS Database	http://fiehnlab.ucdavis.edu/Metabolite-Library-2007/	PMN	http://www.plantcyc.org/
FooDB	http://www.foodb.ca	PubChem	http://pubchem.ncbi.nlm.nih.gov/
GMDB	http://gmd.mpimp-golm.mpg.de/	Reactome	http://www.reactome.org/
HMDB	http://metabolomics.pharm.uconn.edu/iimdb/	RiceCyc	http://pathway.gramene.org/gramene/ricecyc.shtml
HumanCyc	http://www.genome.jp/kegg/	Serum Metabolome Database	http://www.serummetabolome.ca/
IIDMB	http://www.genome.jp/kegg/glycan/	SetupX & BinBase	http://fiehnlab.ucdavis.edu/projects/binbase_setupx

AACLifeSci

CEU Universidad San Pablo

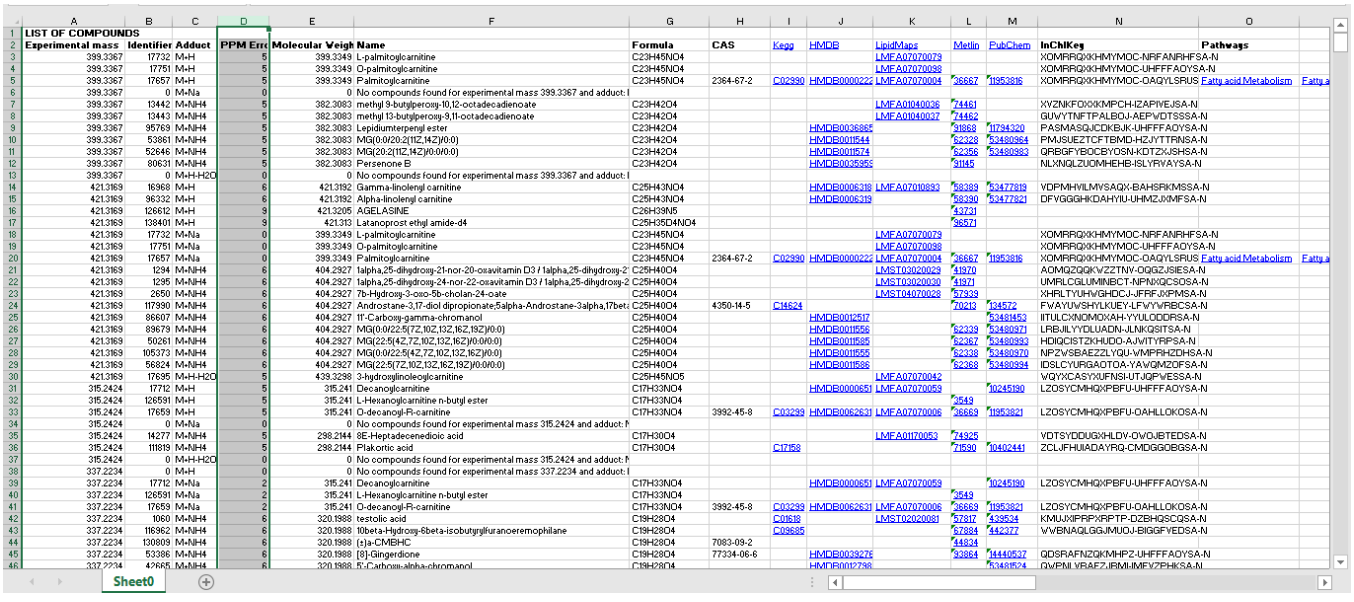


- Devoted to metabolite annotation.
- Performs searches over unified compounds from different sources.
- Apply knowledge based on the input data given by the user.
- Aid to identify oxidized lipids.
- http://ceumass.eps.uspceu.es/mediator

AACLifeSci

CEU Universidad San Pablo

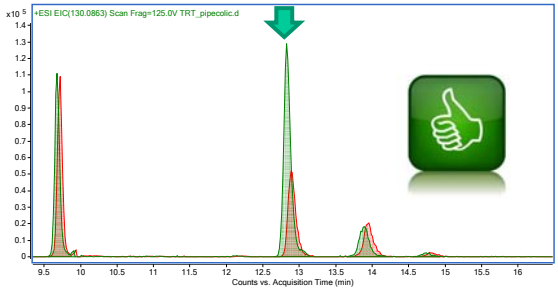
CEU
MASS MEDIATOR



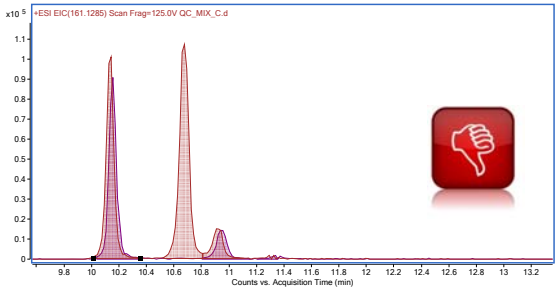
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	LIST OF COMPOUNDS														
2	Experimental mass	Identifier	Adduct	PPM Erro	Molecular Weigh	Name	Formula	CAS	Kegg	HMDB	LipidMaps	Metlin	PubChem	InChIKey	Pathways
3	399.3367	17732	M+H	5	399.3349	L-palmitoylcarnitine	C23H45NO4				LMFA07070079			XOMRRQXKHMYMOC-NRFANRHFSA-N	
4	399.3367	17751	M+H	5	399.3349	O-palmitoylcarnitine	C23H45NO4				LMFA07070098			XOMRRQXKHMYMOC-UHFFFAOYSA-N	
5	399.3367	17657	M+H	5	399.3349	Palmitoylcarnitine	C23H45NO4	2364-67-2	C02990	HMDB0000222	LMFA07070004	36667	11953816	XOMRRQXKHMYMOC-OAQYLSRUS	Fatty acid Metabolism
6	399.3367	0	M+Na	0	0	No compounds found for eseprimental mass 399.3367 and adduct:									
7	399.3367	13442	M+NH4	5	382.3083	methyl 9-butylperoxy-10,12-octadecadienoate	C23H42O4				LMFA01040036	74461		XVZNKFOXXKMPCH-IZAPIVEJSA-N	
8	399.3367	13443	M+NH4	5	382.3083	methyl 13-butylperoxy-9,11-octadecadienoate	C23H42O4				LMFA01040037	74462		GUVYTNFTPALBOJ-AEPVDTSSSA-N	
9	399.3367	95769	M+NH4	5	382.3083	Lepidiumterpengl ester	C23H42O4			HMDB0036865		91868	11794320	PASMASQJCDKBJK-UHFFFAOYSA-N	
10	399.3367	53861	M+NH4	5	382.3083	MG(0:0/20:2(11Z,14Z)/0:0)	C23H42O4			HMDB0011544		52328	53480964	PMJSUEZTCFTBMD-HZJYTTRNSA-N	
11	399.3367	52646	M+NH4	5	382.3083	MG(20:2(11Z,14Z)/0:0/0:0)	C23H42O4			HMDB0011574		52356	53480863	QRBSIFYBOCBYQSN-KDTZXJSHSA-N	
12	399.3367	80631	M+NH4	5	382.3083	Persenone B	C23H42O4			HMDB0025955		91145		NLXNQLZUOMHEHB-ISLYRYAYSA-N	
13	399.3367	0	M+H-H2O	0	0	No compounds found for eseprimental mass 399.3367 and adduct:									
14	421.3169	16968	M+H	6	421.3192	Gamma-linolenyl carnitine	C25H43NO4			HMDB0006318	LMFA07010893	68389	53477819	VDPMHVILMVSAQX-BAHSRKMSSA-N	
15	421.3169	96332	M+H	6	421.3192	Alpha-linolenyl carnitine	C25H43NO4			HMDB0006319		68390	53477821	DFVGGGHKDAHYIU-UHMZJXMFSA-N	
16	421.3169	126612	M+H	9	421.3205	AGELASINE	C26H39N5					43721			
17	421.3169	138401	M+H	9	421.313	Latanoprost ethyl amide-d4	C25H35D4NO4					96571			
18	421.3169	17732	M+Na	0	399.3349	L-palmitoylcarnitine	C23H45NO4				LMFA07070079			XOMRRQXKHMYMOC-NRFANRHFSA-N	
19	421.3169	17751	M+Na	0	399.3349	O-palmitoylcarnitine	C23H45NO4				LMFA07070098			XOMRRQXKHMYMOC-UHFFFAOYSA-N	
20	421.3169	17657	M+Na	0	399.3349	Palmitoylcarnitine	C23H45NO4	2364-67-2	C02990	HMDB0000222	LMFA07070004	36667	11953816	XOMRRQXKHMYMOC-OAQYLSRUS	Fatty acid Metabolism
21	421.3169	1294	M+NH4	6	404.2927	1alpha,25-dihydroxy-21-nor-20-oxavitamin D3 / 1alpha,25-dihydroxy-2	C25H40O4				LMST03020029	41970		AOMQZQQKVZZTNV-OQGZJSIESA-N	
22	421.3169	1295	M+NH4	6	404.2927	1alpha,25-dihydroxy-24-nor-22-oxavitamin D3 / 1alpha,25-dihydroxy-2	C25H40O4				LMST03020030	41971		UMRLCGLUMINBCT-NPNXQCSOSA-N	
23	421.3169	2650	M+NH4	6	404.2927	7b-Hydroxy-3-oxo-5b-cholan-24-oate	C25H40O4				LMST04070028	57939		XHRLTYUHVGHDCJ-JFRFJXPMSA-N	
24	421.3169	117990	M+NH4	6	404.2927	Androstane-3,17-diol dipropionate;5alpha-Androstane-3alpha,17beta	C25H40O4	4350-14-5	C14624			70213	134572	FVAYUVSHYLUEY-LFWYYRBCSA-N	
25	421.3169	86607	M+NH4	6	404.2927	11'-Carboxy-gamma-chromanol	C25H40O4			HMDB0012517			53481453	ITULCXNOMOXAH-YYULODDRSA-N	
26	421.3169	89679	M+NH4	6	404.2927	MG(0:0/22:5(7Z,10Z,13Z,16Z,19Z)/0:0)	C25H40O4			HMDB0011556		52339	53480971	LRBJILYYDLUADN-JLNKQSITSA-N	
27	421.3169	50261	M+NH4	6	404.2927	MG(22:5(4Z,7Z,10Z,13Z,16Z)/0:0/0:0)	C25H40O4			HMDB0011585		52367	53480993	HDIQCISTZKHUDO-AJVITYRPSA-N	
28	421.3169	105373	M+NH4	6	404.2927	MG(0:0/22:5(4Z,7Z,10Z,13Z,16Z)/0:0)	C25H40O4			HMDB0011555		52338	53480970	NPZVSBAEZZLYQU-VMPRHZDHSA-N	
29	421.3169	56824	M+NH4	6	404.2927	MG(22:5(7Z,10Z,13Z,16Z,19Z)/0:0/0:0)	C25H40O4			HMDB0011586		52368	53480994	IDSLCYURGAOTOA-YAWQMZOFSA-N	
30	421.3169	17695	M+H-H2O	5	439.3298	3-hydroxylinoleoylcarnitine	C25H45NO5				LMFA07070042			WQYXCASYXUFNSI-UTJGPVESSA-N	
31	315.2424	17712	M+H	5	315.241	Decanoylcarnitine	C17H33NO4			HMDB0000651	LMFA07070059		10245190	LZOSYCMHQXPBFU-UHFFFAOYSA-N	
32	315.2424	126591	M+H	5	315.241	L-Hexanoylcarnitine n-butyl ester	C17H33NO4					3549			
33	315.2424	17659	M+H	5	315.241	O-decanoyl-R-carnitine	C17H33NO4	3992-45-8	C03299	HMDB0062631	LMFA07070006	36669	11953821	LZOSYCMHQXPBFU-OAHLLOKOSA-N	
34	315.2424	0	M+Na	0	0	No compounds found for eseprimental mass 315.2424 and adduct:									
35	315.2424	14277	M+NH4	5	298.2144	8E-Heptadecenedioic acid	C17H30O4				LMFA01170053	74925		VDTSYDDUGXHLDV-OWOJBTEDSA-N	
36	315.2424	111819	M+NH4	5	298.2144	Plakortic acid	C17H30O4		C17158			71590	10402441	ZCLJFHUIADAYRQ-CMDGGOBGSA-N	
37	315.2424	0	M+H-H2O	0	0	No compounds found for eseprimental mass 315.2424 and adduct:									
38	337.2234	0	M+H	0	0	No compounds found for eseprimental mass 337.2234 and adduct:									
39	337.2234	17712	M+Na	2	315.241	Decanoylcarnitine	C17H33NO4			HMDB0000651	LMFA07070059		10245190	LZOSYCMHQXPBFU-UHFFFAOYSA-N	
40	337.2234	126591	M+Na	2	315.241	L-Hexanoylcarnitine n-butyl ester	C17H33NO4					3549			
41	337.2234	17659	M+Na	2	315.241	O-decanoyl-R-carnitine	C17H33NO4	3992-45-8	C03299	HMDB0062631	LMFA07070006	36669	11953821	LZOSYCMHQXPBFU-OAHLLOKOSA-N	
42	337.2234	1060	M+NH4	6	320.1988	testolic acid	C19H28O4		C09618		LMST02020081	57817	439534	KMUJXIPRPXRPTP-DZBHQSCQSA-N	
43	337.2234	116962	M+NH4	6	320.1988	10beta-Hydroxy-6beta-isobutyrylfuranoeremophilane	C19H28O4		C09685			57884	442377	WWBNAQLGGJMUOJ-BIGGFVEDSA-N	
44	337.2234	130809	M+NH4	6	320.1988	(±)a-CMBHC	C19H28O4	7083-09-2				44824			
45	337.2234	53386	M+NH4	6	320.1988	[8]-Gingerdione	C19H28O4	77334-06-6		HMDB0039276		33864	14440537	QDSRAFNZQKMHPZ-UHFFFAOYSA-N	
46	337.2234	42665	M+NH4	6	320.1988	5'-Carboxy-alpha-chromanol	C19H28O4			HMDB0012798			53481524	QWPNLYBAEZJBMH-MEYZPHKSA-N	

Sheet0

Confirmation by Standard addition



Pipecolic acid

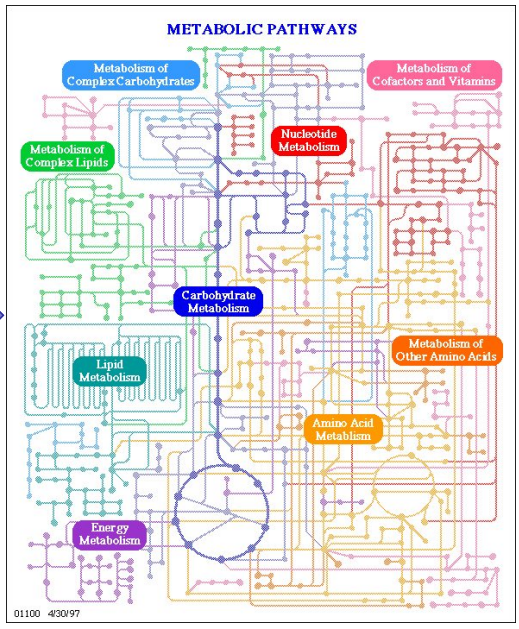


Methyl-lysine

From Lists to Pathways

metabolomics

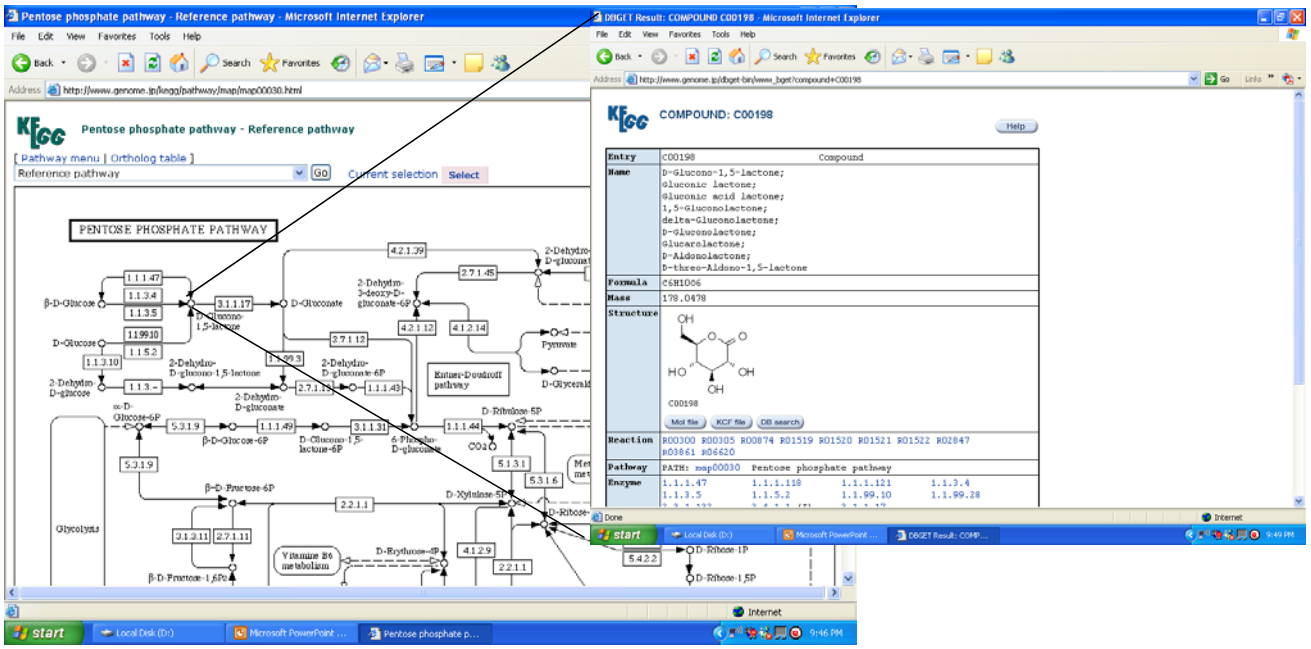
Compound	Retention Time (min)	Conc. in Urine (µM)	Compound	Retention Time (min)	Conc. in Urine (µM)
Dns-o-phospho -L-serine	0.92	<D.L. *	**Dns-Ile**	6.35	25
Dns-o-phospho -L-tyrosine	0.95	<D.L.	**Dns-3-aminosalicylic acid**	6.44	0.5
Dns-adnosine monophosphate	0.99	<D.L.	**Dns-pipecolic acid**	6.50	0.5
Dns-o-phosphoethanolamine	1.06	16	**Dns-Leu**	6.54	54
Dns-glucosamine	1.06	22	**Dns-cystathionine**	6.54	0.3
Dns-o-phospho -L-threonine	1.09	<D.L.	**Dns-Leu-Pro**	6.60	0.4
Dns-6-dimet hylamine purine	1.20	<D.L.	**Dns-5-hydroxylysine**	6.65	1.6
Dns-3-methyl -histidine	1.22	80	**Dns-Cystine**	6.73	160
Dns-taurine	1.25	834	**Dns-N-norleucine**	6.81	0.1
Dns-carnosine	1.34	28	Dns-5-hydroxydopamine	7.17	<D.L.
Dns-Arg	1.53	36	**Dns-dimethylamine**	7.33	293
Dns-Asn	1.55	133	**Dns-5-HIAA**	7.46	18
Dns-hypotaurine	1.58	10	**Dns-umbelliferone**	7.47	1.9
Dns-homocarnosine	1.61	3.9	Dns-2,3 -diaminopropionic acid	7.63	<D.L.
Dns-guanidine	1.62	<D.L.	**Dns-L-ornithine**	7.70	15
Dns-Gln	1.72	633	**Dns-4-acetyamidophenol**	7.73	51
Dns-allantoin	1.83	3.8	**Dns-procaine**	7.73	8.9
Dns-L-citrulline	1.87	2.9	**Dns-homocystine**	7.76	3.3
Dns-1 (or 3 -)-methylhistamine	1.94	1.9	**Dns-acetaminophen**	7.97	82
Dns-adenosine	2.06	2.6	**Dns-Phe-Phe**	8.03	0.4
Dns-methylguanidine	2.20	<D.L.	**Dns-5-methyo xysalicylic acid**	8.04	2.1
Dns-Ser	2.24	511	**Dns-Lys**	8.16	184
Dns-aspartic acid amide	2.44	26	Dns-aniline	8.17	<D.L.
Dns-4-hydroxy -proline	2.56	2.3	**Dns-leu-Phe**	8.22	0.3
Dns-Glu	2.57	21	**Dns-His**	8.35	1550
Dns-Asp	2.60	90	Dns-4-thialysine	8.37	<D.L.
Dns-Thr	3.03	157	Dns-benzylamine	8.38	<D.L.
Dns-epinephrine	3.05	<D.L.	**Dns-1-ephedrine**	8.50	0.6
Dns-ethanolamine	3.11	471	**Dns-tryptamine**	8.63	0.4
Dns-aminoadipic acid	3.17	70	Dns-pyridoxamine	8.94	<D.L.
Dns-Gly	3.43	2510	Dns-2-methyl -benzylamine	9.24	<D.L.
Dns-Ala	3.88	593	**Dns-5-hydroxytrptophan**	9.25	0.12
Dns-aminolevulinic acid	3.97	30	**Dns-1,3 -diaminopropane**	9.44	0.23
Dns-r-amino -butyric acid	3.98	4.6	**Dns-putrescine**	9.60	0.5
Dns-p-amino -hippuric acid	3.98	2.9	**Dns-1,2 -diaminopropane**	9.66	0.1
Dns-5-hydro xymethyluricil	4.58	1.9	**Dns-tyrosinamide**	9.79	29
Dns-tryptophanamide	4.70	5.5	**Dns-dopamine**	10.08	140
Dns-isoguanine	4.75	<D.L.	**Dns-cadaverine**	10.08	0.08
Dns-5-aminopentanoic acid	4.79	1.6	**Dns-histamine**	10.19	0.4
Dns-sarcosine	4.81	7.2	**Dns-3-methoxy -tyramine**	10.19	9.2
Dns-3-amino -isobutyrate	4.81	85	**Dns-Tyr**	10.28	321
Dns-2-aminobutyric acid	4.91	17	Dns-cysteamine	10.44	<D.L.



Pathway Databases

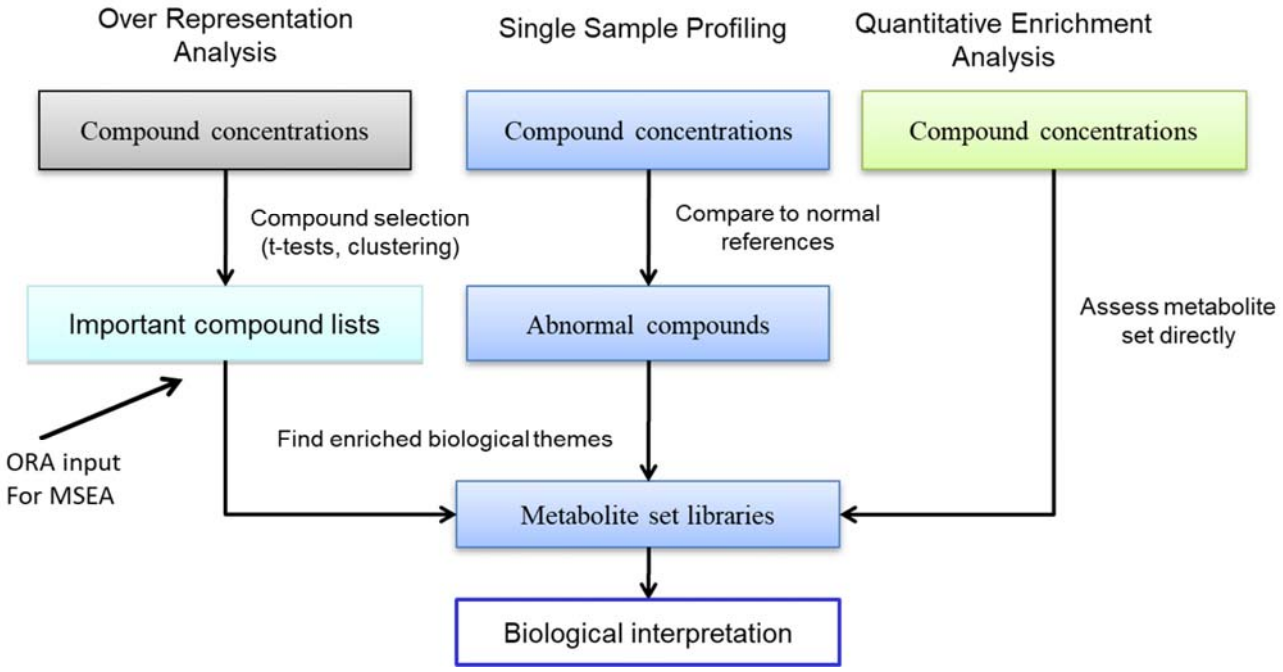
- Rich source of biological data that relates metabolites to genes, proteins, diseases, signaling events and processes
- Provide various tools to permit visualization and gene/metabolite mapping
- Often cover multiple species
- KEGG (www.genome.jp/kegg/), BioCyc/MetaCyc (https://biocyc.org/), SMPDB (www.smpdb.ca), Reactome (www.reactome.org), WikiPathways (http://www.wikipathways.org)...
- "*Strictly speaking, one could argue that pathways don't exist... there are only networks*." (WikiPathways.org)

KEGG – Kyoto Encyclopedia of Genes and Genomes

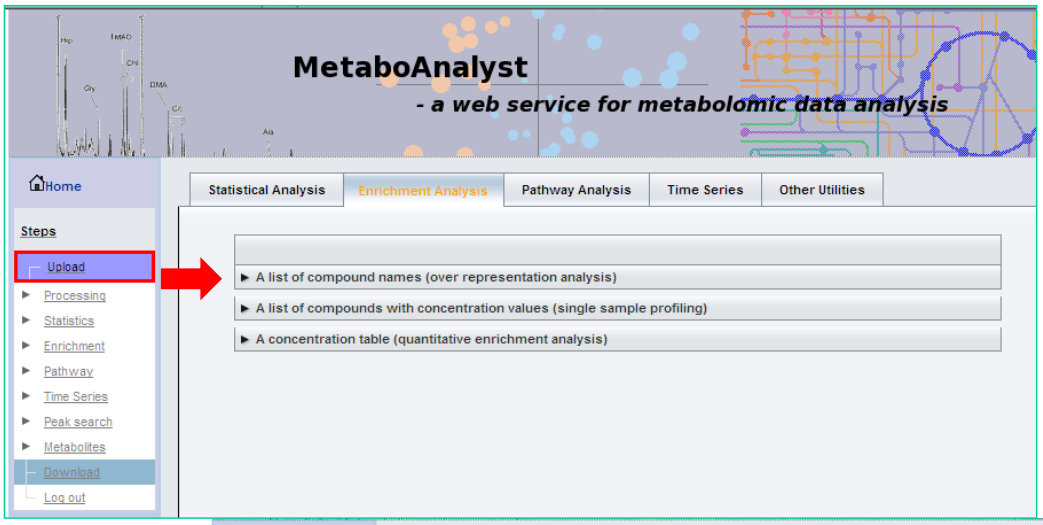


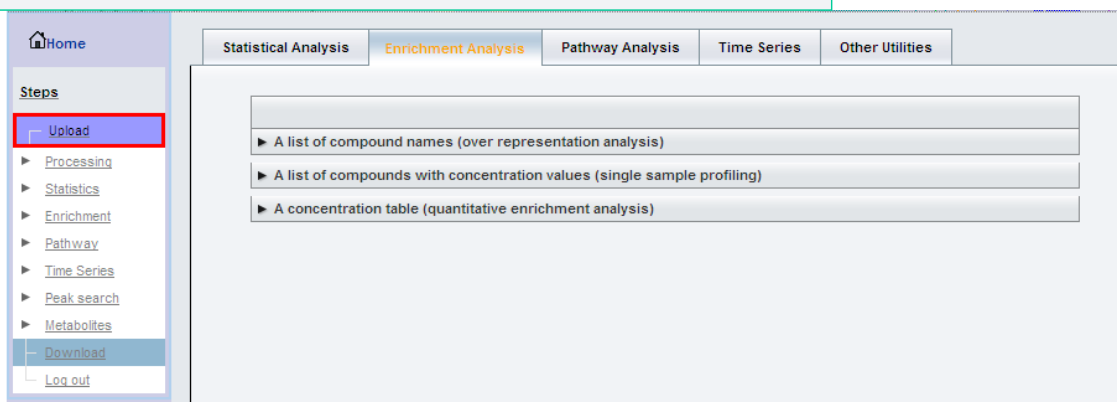
http://www.genome.jp/kegg/

The Metabolite Set Enrichment Analysis MSEA approach

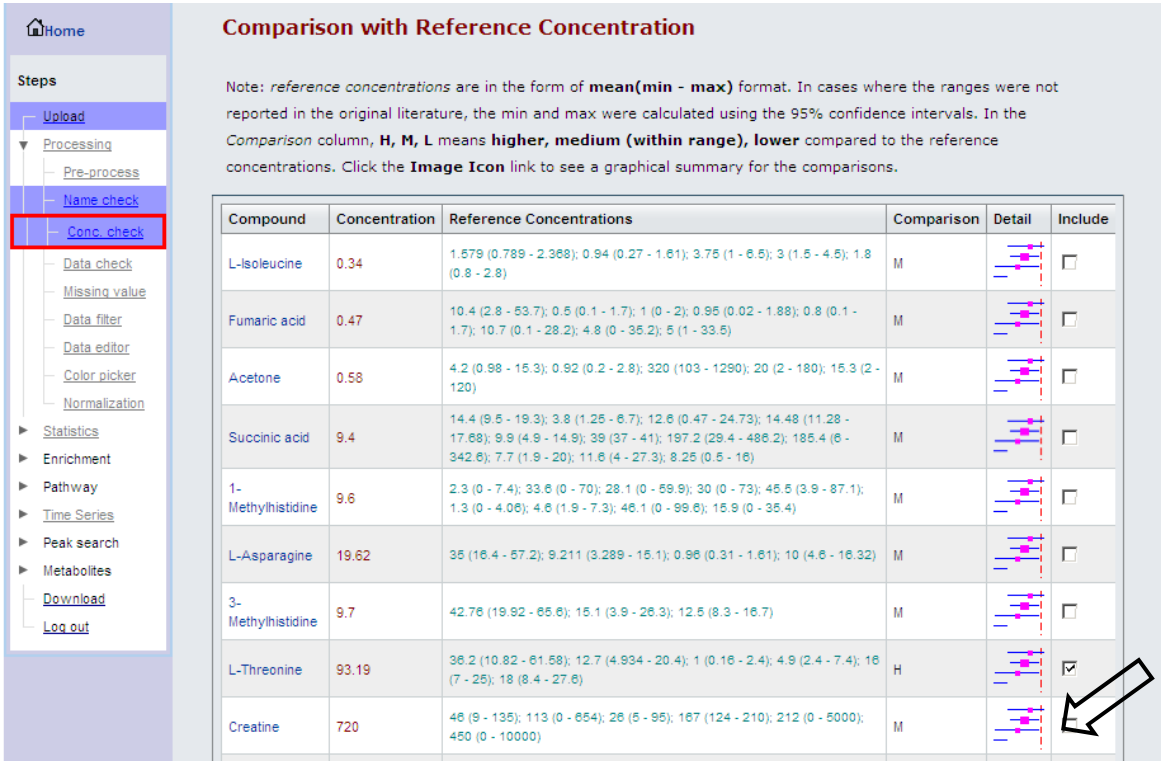


Start with a compound List

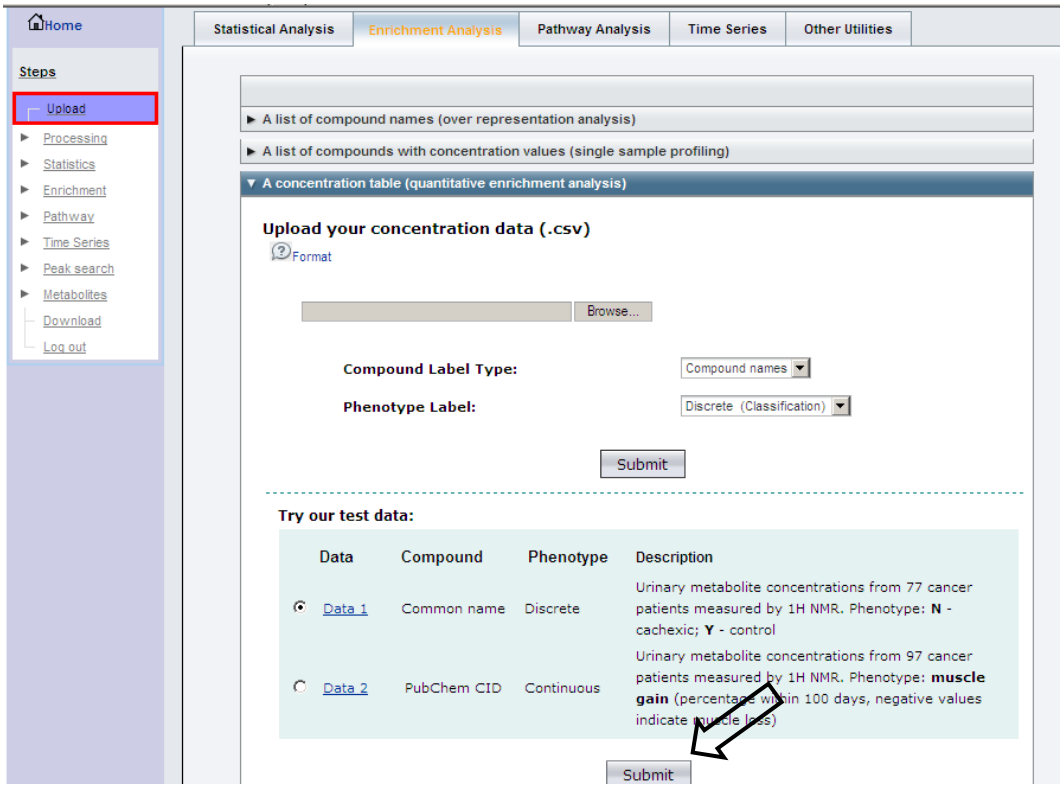




Concentration Comparison



Quantitative Enrichment Analysis

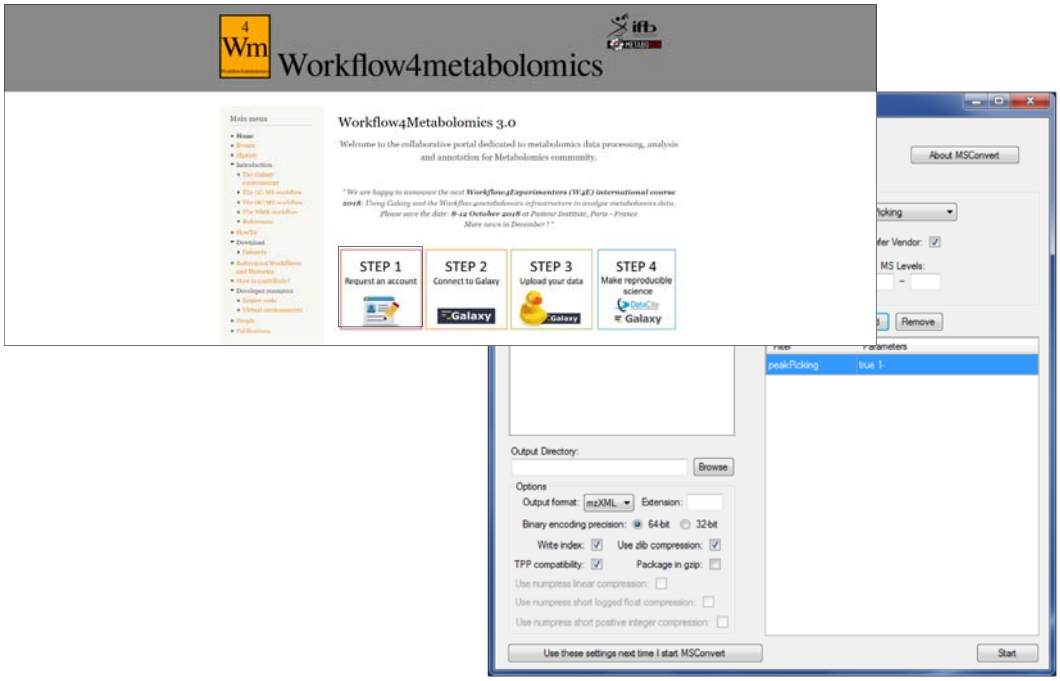


RESULT

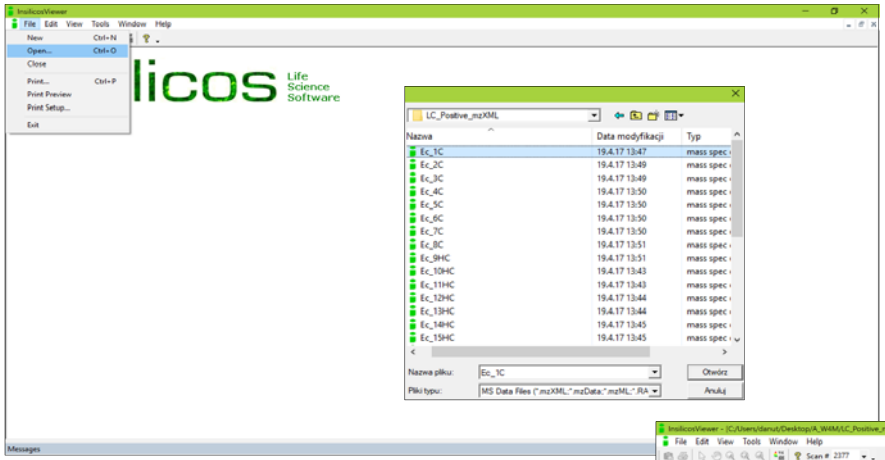
Metaboanalyst
Metabolic Pathway Analysis (MetPA)

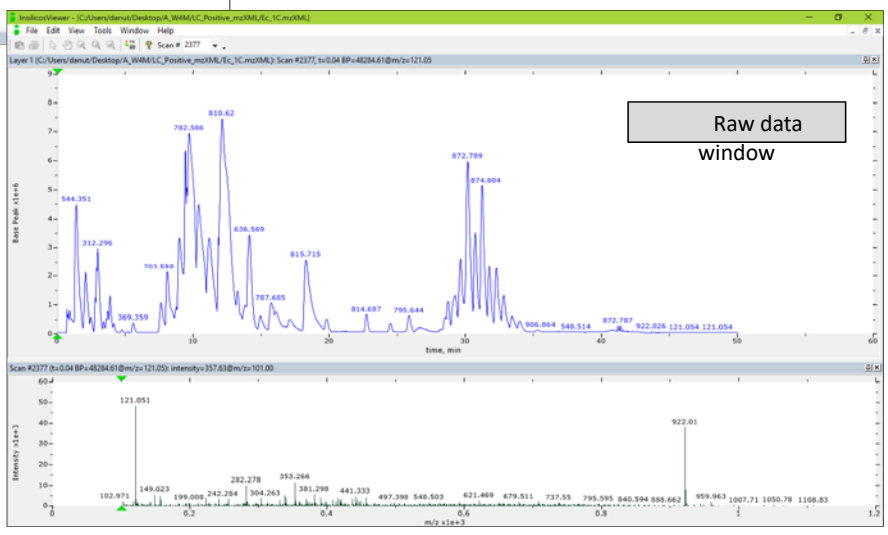
- Purpose: to extend and enhance metabolite set enrichment analysis for pathways by
  - Considering the **structures of pathway**
  - Dynamic pathway visualization
- Currently supports ~1500 pathways covering 17 organisms (based on KEGG)

PRACTICAL SESSION. VISUALS_1

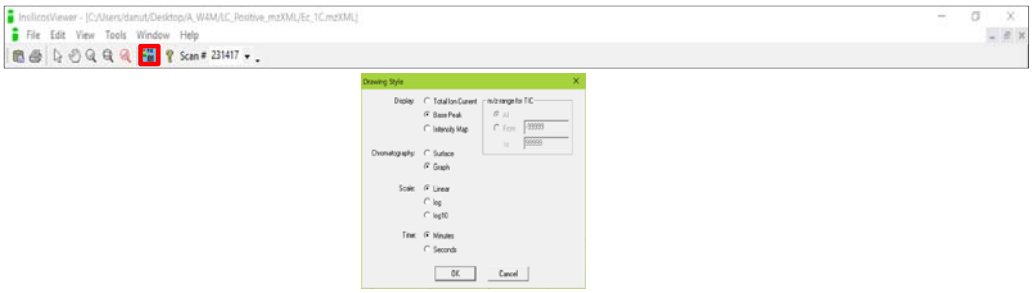


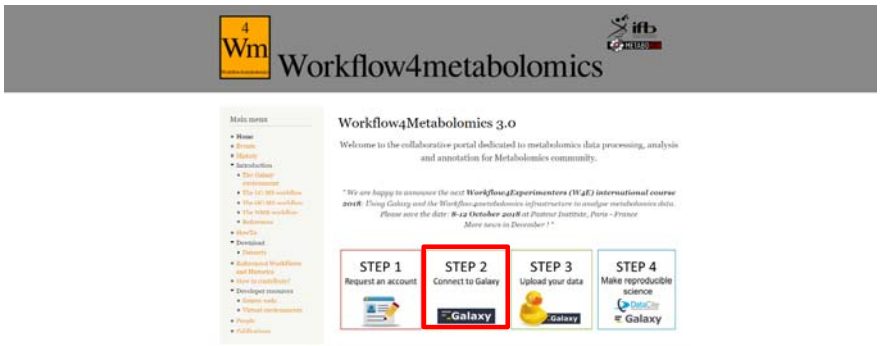
PRACTICAL SESSION. VISUALS_2



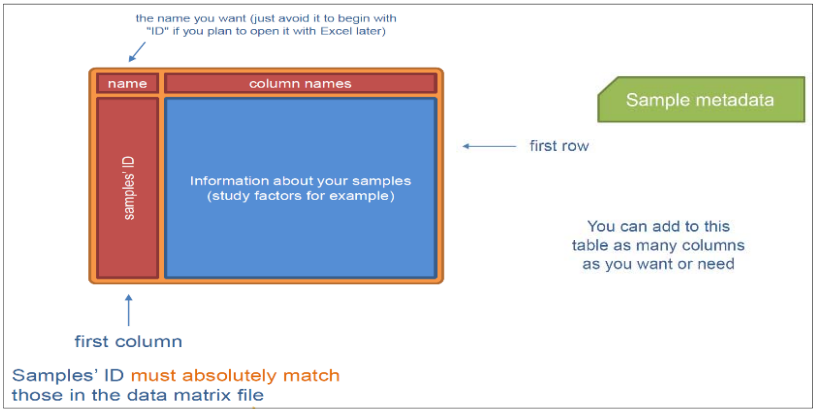


PRACTICAL SESSION. VISUALS_3





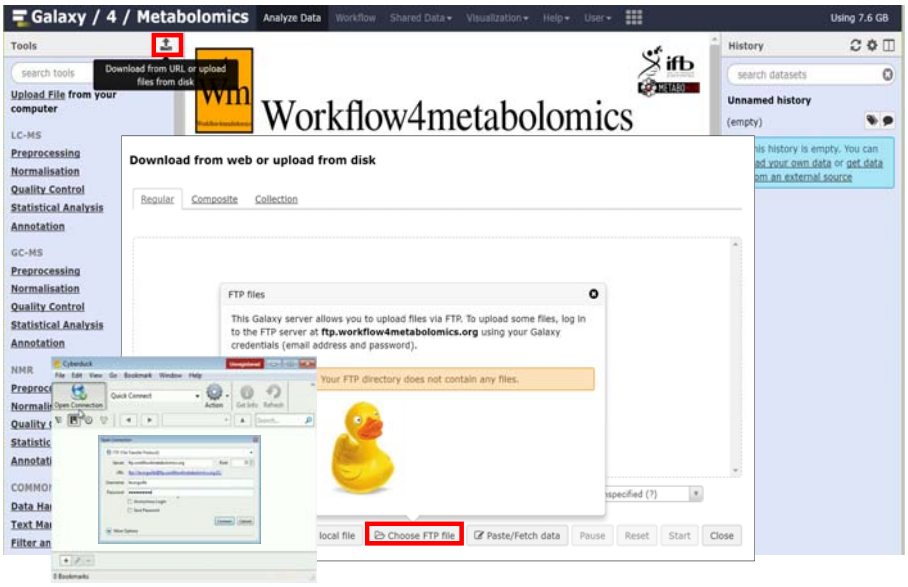
PRACTICAL SESSION. VISUALS_4

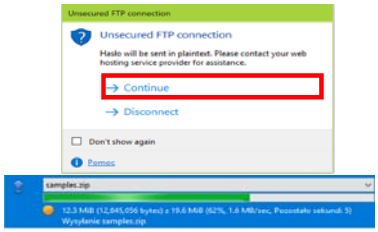


PRACTICAL SESSION. VISUALS_5

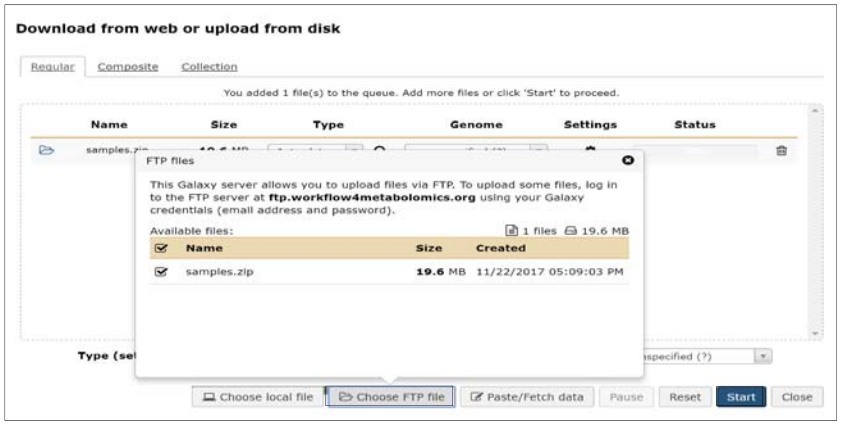
sampleName	class	polarity	sampleType	batch	injectionOrder	diet
QC	one	positive	pool	B1	1	NA
C1	one	positive	sample	B1	7	C
HC3	one	positive	sample	B1	10	HC
BL	one	positive	blank	B1	12	NA
…	…	…	…	…	…	…

PRACTICAL SESSION. VISUALS_6

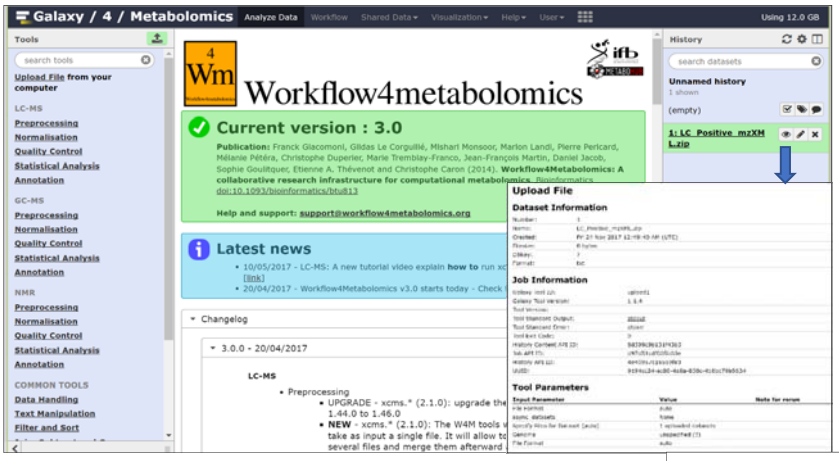


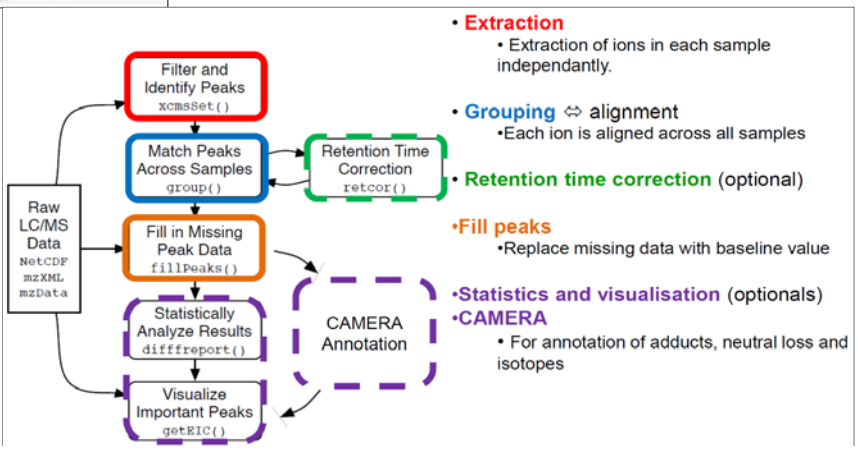


PRACTICAL SESSION. VISUALS_7

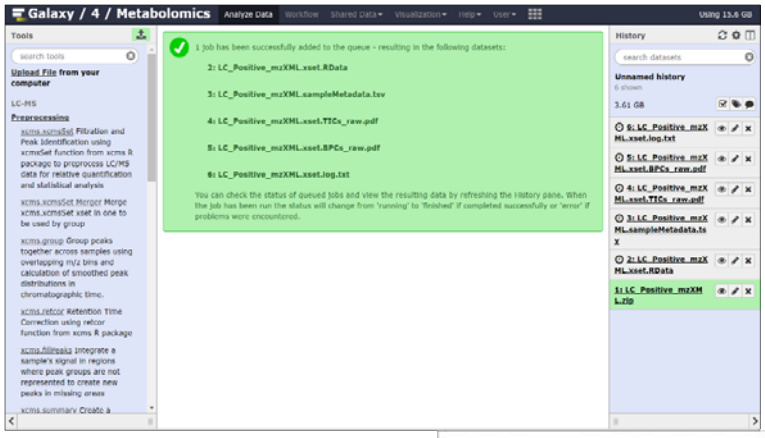


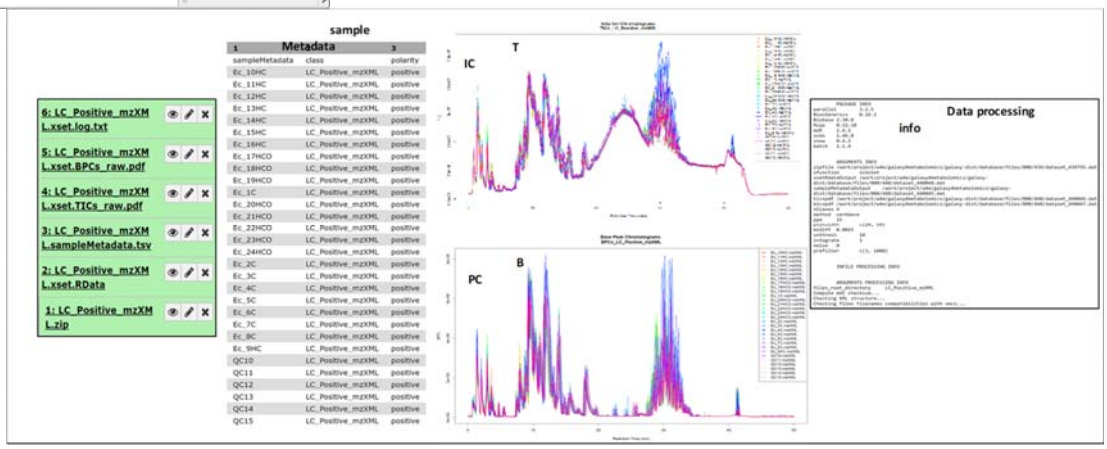
PRACTICAL SESSION. VISUALS_8





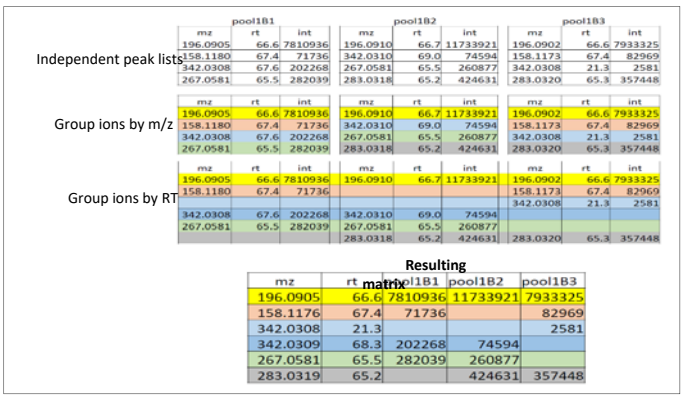
PRACTICAL SESSION. VISUALS_9





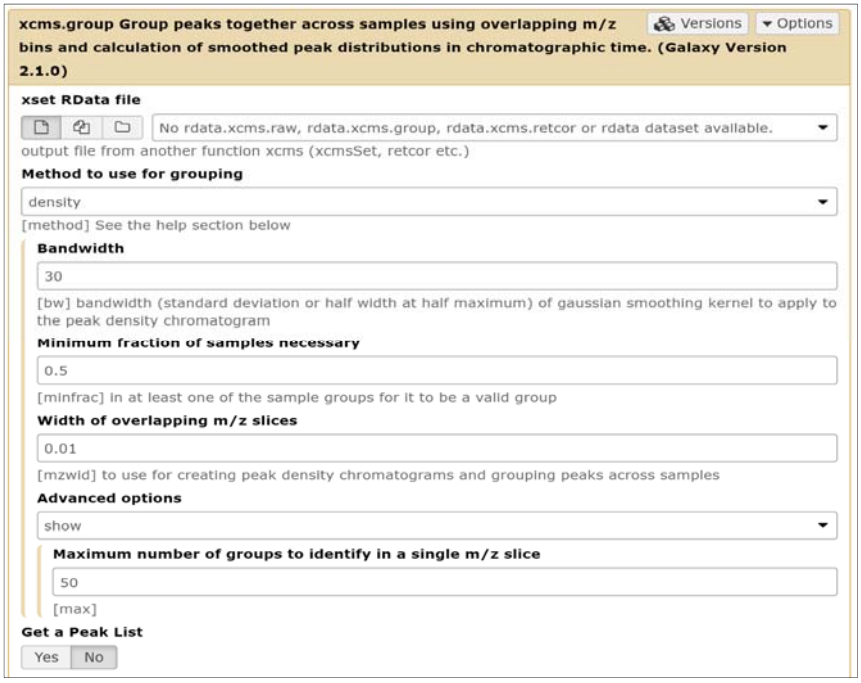
PRACTICAL SESSION. VISUALS_10

xcms.group Group peaks together across samples using overlapping m/z bins and calculation of smoothed peak distributions in chromatographic time.



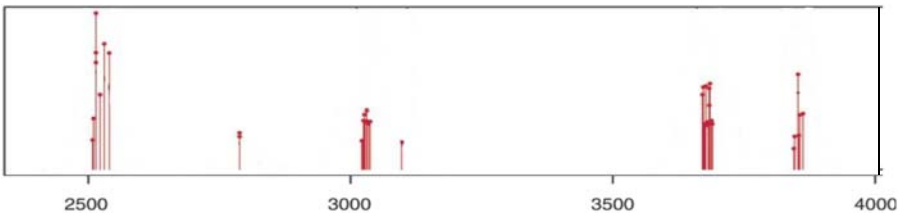
Parameter : num + label	Format
Or : RData file | rdata.xcms.raw	
Or : RData file | rdata.xcms.retcor	

PRACTICAL SESSION. VISUALS_11

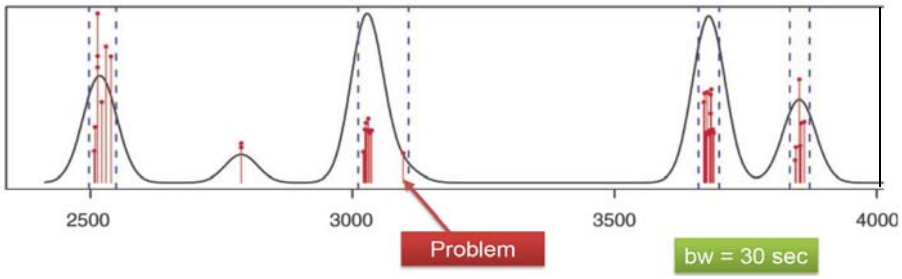


PRACTICAL SESSION. VISUALS_12

Grouping peaks in mass bin: 337.975 – 338.225 m/z (mzwid)

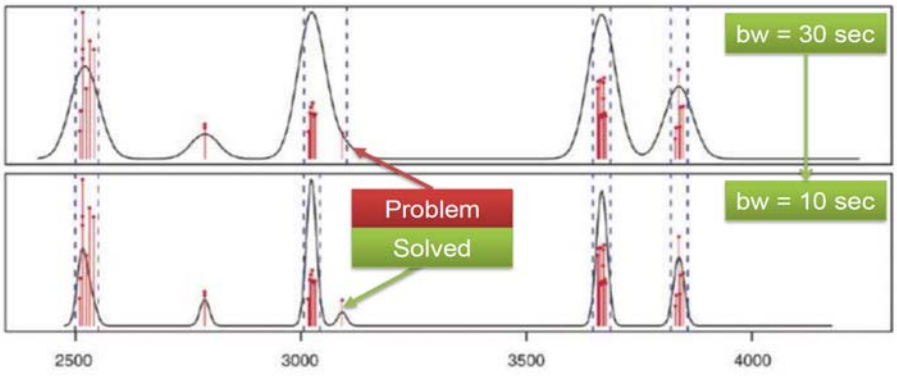


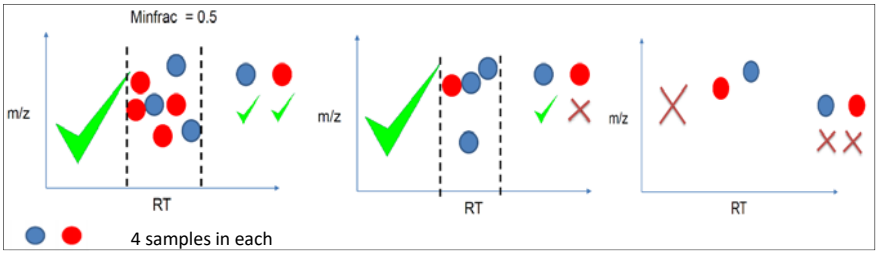
Grouping peaks in mass bin: 337.975 – 338.225 m/z (mzwid)



PRACTICAL SESSION. VISUALS_13

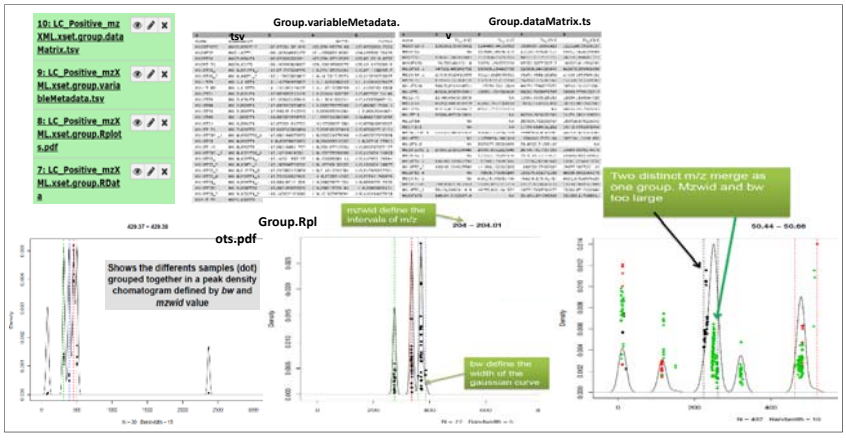
Grouping peaks in mass bin: 337.975 – 338.225 m/z (mzwid)





4 samples in each group

PRACTICAL SESSION. VISUALS_14



PRACTICAL SESSION. VISUALS_15

xcms.retcor Retention Time Correction using retcor function from xcms R package

Parameter : num + label	Format
1 : RData file	rdata.xcms.group

xcms.retcor Retention Time Correction using retcor function from xcms R package (Galaxy Version 2.1.0) Versions Options

xset RData file

No rdata.xcms.raw, rdata.xcms.group or rdata dataset available.

output file from another function xcms (xcmsSet, retcor etc.)

Method to use for retention time correction

peakgroups

[method] See the help section below

Smooth method

loess

[smooth] either 'loess' for non-linear alignment or 'linear' for linear alignment

Number of extra peaks to allow in retention time correction correction groups

1

[extra]

Number of missing samples to allow in retention time correction groups

1

Number of admitted missing well behaved peak in a group.

[missing]

Advanced options

hide

Resubmit your raw dataset or your zip file

Execute

PRACTICAL SESSION. VISUALS_16

Advanced options

show

Degree of smoothing for local polynomial regression fitting

0.2

[span]

Family

gaussian

[family] if gaussian fitting is by least-squares with no outlier removal, and if symmetric a re descending M estimator is used with Tukey's biweight function, allowing outlier removal

plottype

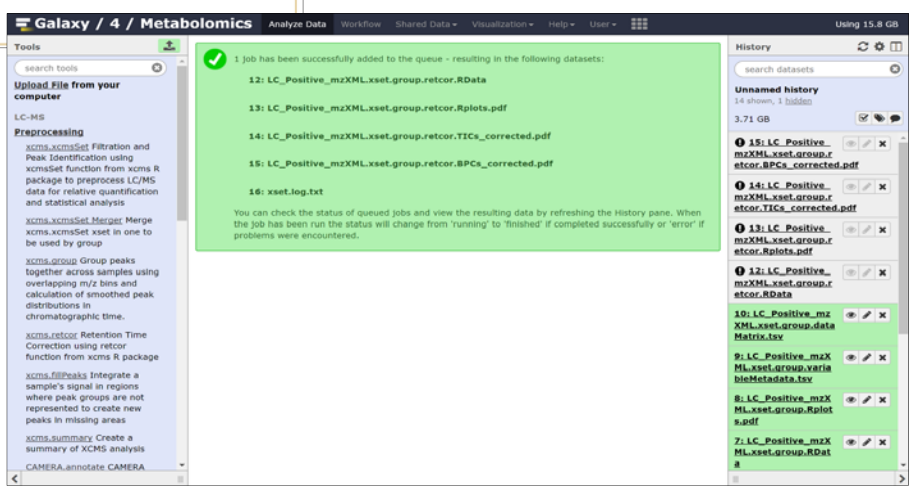
deviation

Plot to visualize the result of the retention time correction.

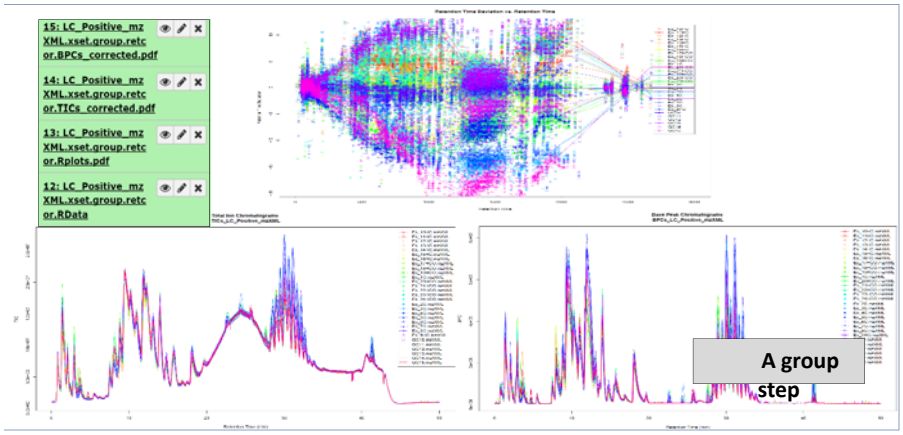
[plottype] if deviation plot retention time deviation points and regression fit, and if mdevden also plot peak overall peak density and retention time correction peak density

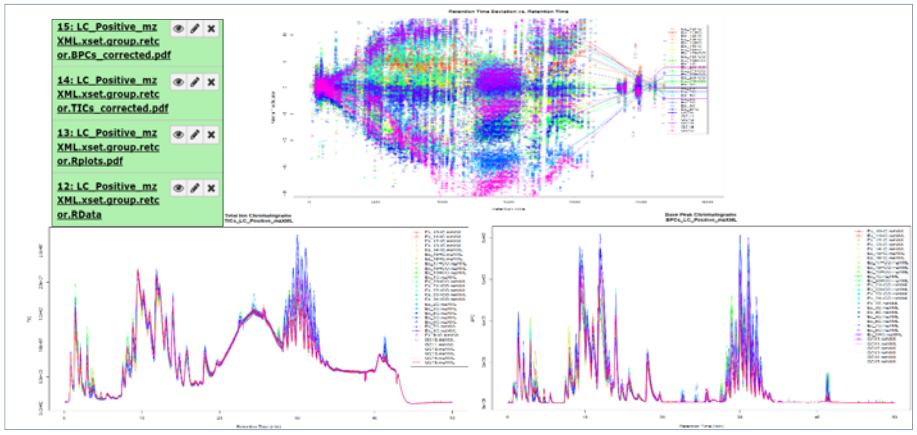
Resubmit your raw dataset or your zip file

Execute

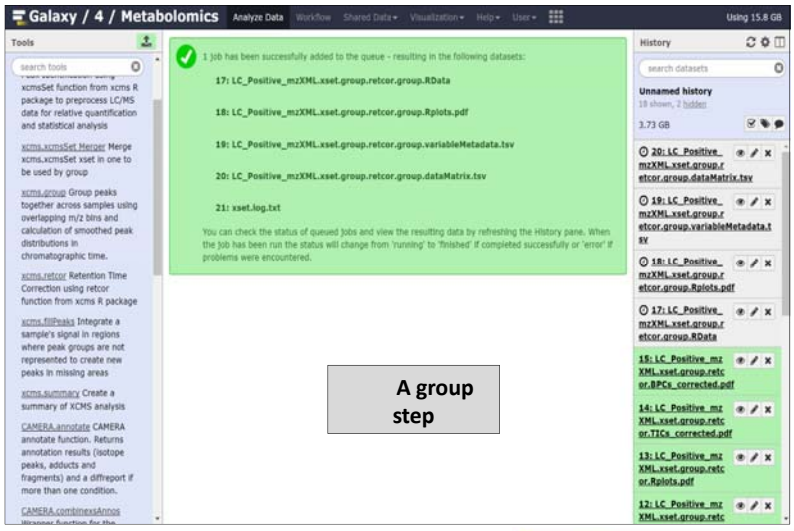


PRACTICAL SESSION VISUALS_17



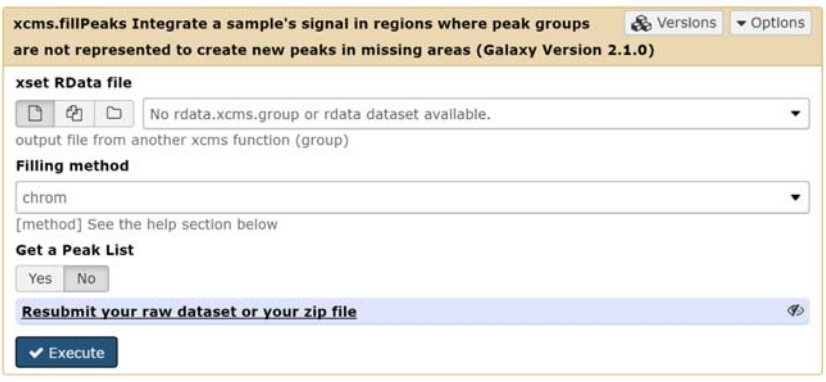


PRACTICAL SESSION. VISUALS_18

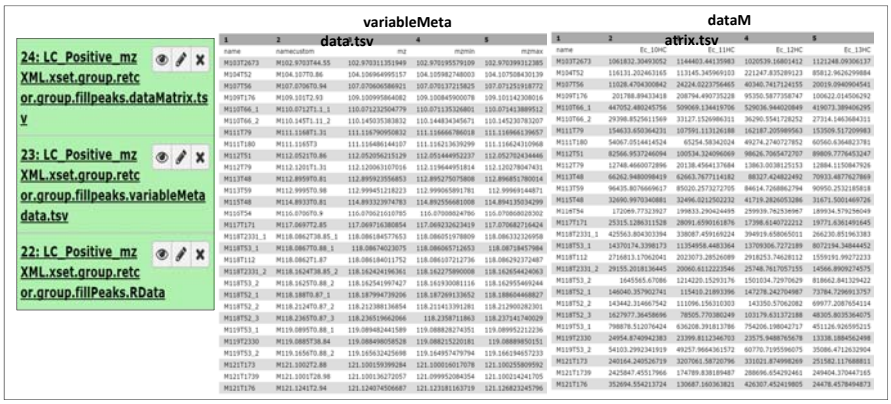


xcms.fillPeaks Integrate a sample's signal in regions where peak groups are not represented to create new peaks in missing areas

Parameter : num + label	Format
1 : RData file	rdata.xcms.group

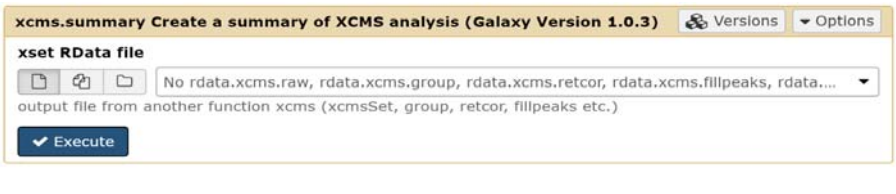


PRACTICAL SESSION. VISUALS_19



PRACTICAL SESSION. VISUALS_20

xcms.summary Create a summary of XCMS analysis



This project has been funded with support from the European Commission.

This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein