

METABOLOMICS

Coral Barbas
Danuta Dudzik
M^a Fernanda Rey-Stolle
Francisco J. Rupérez
Antonia Garcia



SUMMARY



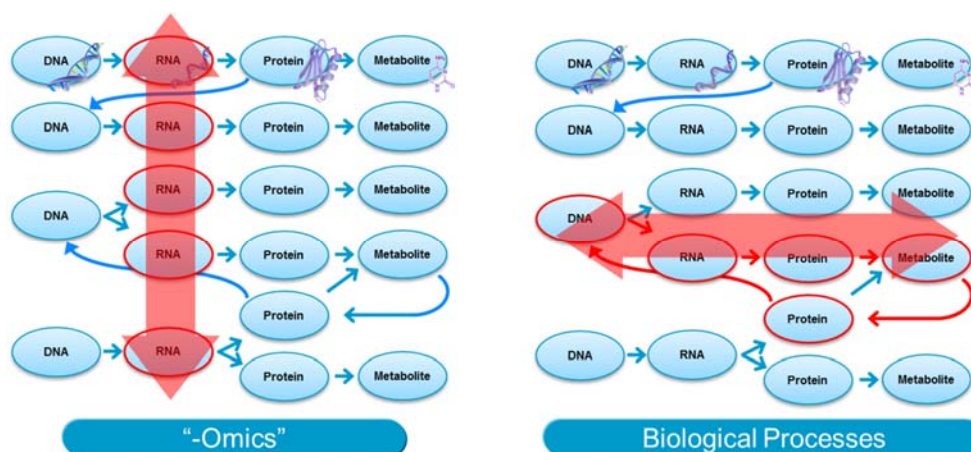
1. Introduction to metabolomics
2. Analytical approaches in metabolomics
 - Workflow of the metabolomics study
 - Quality Control and Quality Assurance Procedure in Metabolomics
3. Data processing and identification of metabolites
 - Data processing pipeline
 - Non-targeted metabolomics data treatment
 - 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

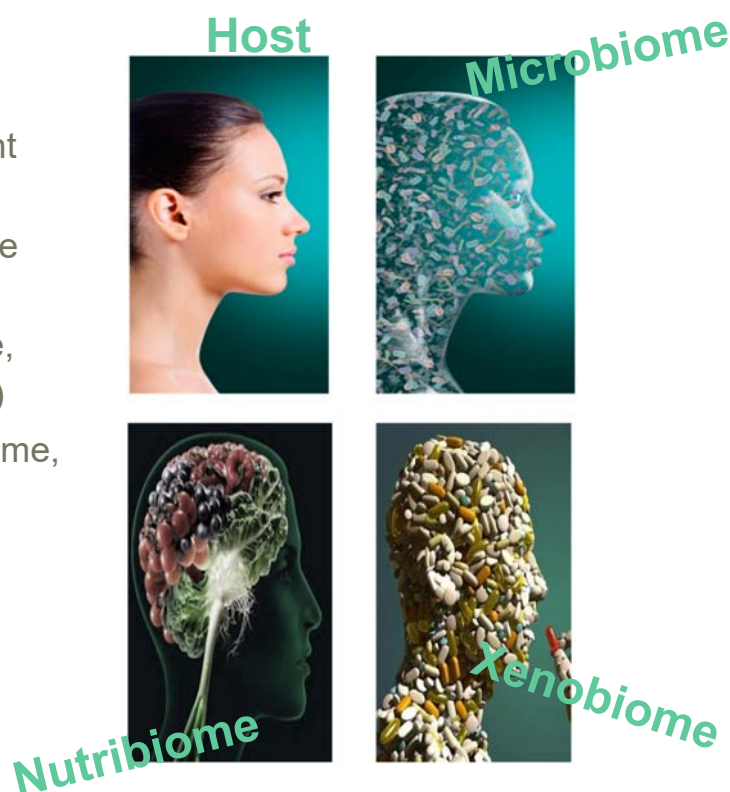


Addendum: Definition of Metabonomics

- 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

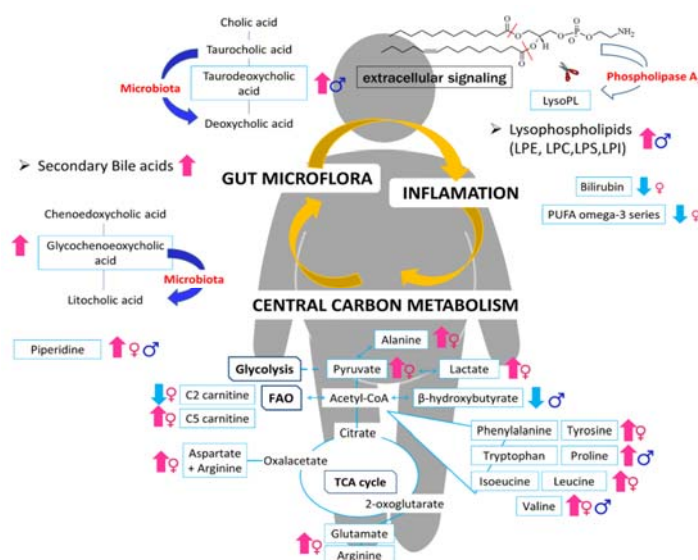
Definition of Metabolome

- “...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, Nutri biome...
- Nature:** Glycome, lipidome, sphingolipidome, peptidome...
- Metabolome ↔ Phenotype



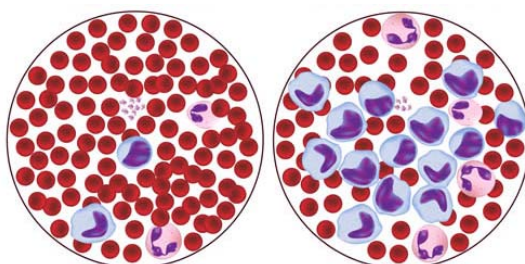
What metabolomics can provide (I)

- 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

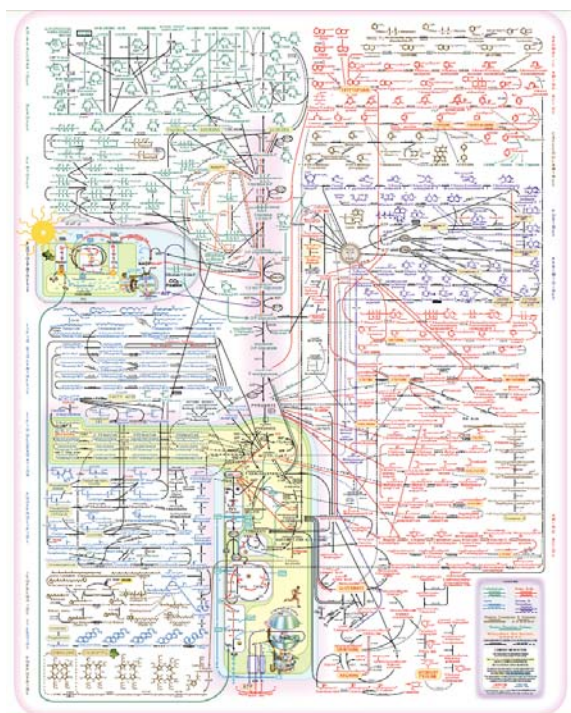
What is metabolomics good for.....

- 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....

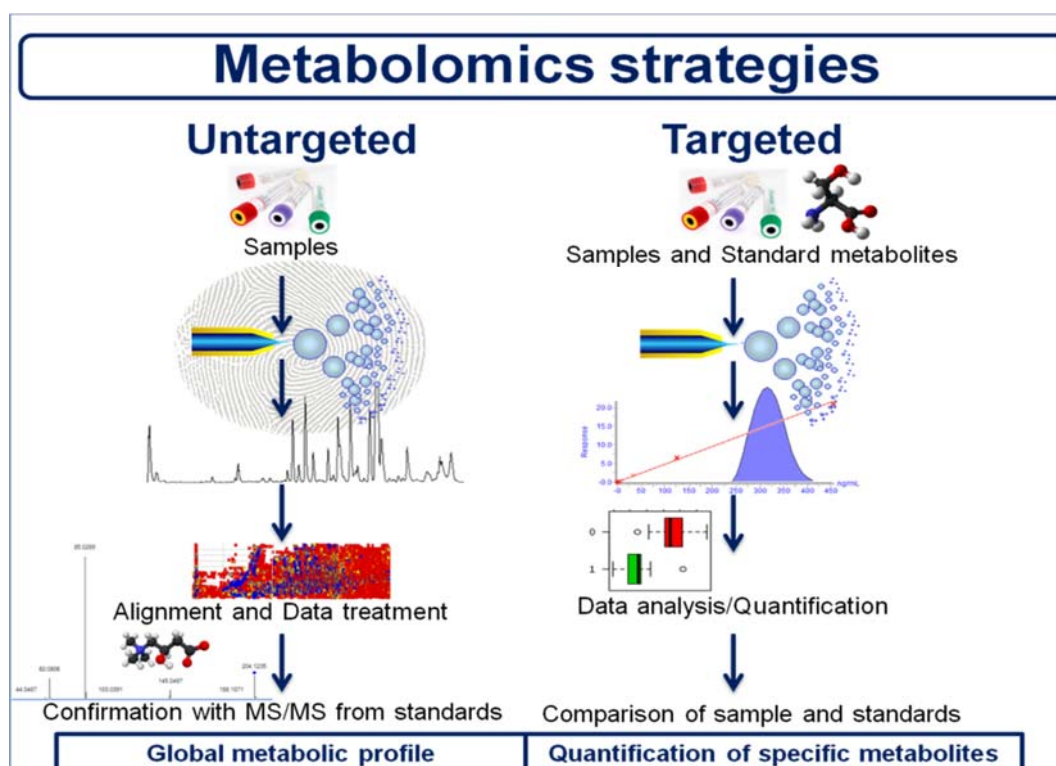
- a method to reveal the fate of a metabolite or drug
- a method for quantification
- the use of a simple kit to quantify a group of metabolites (it requires NMR, MS...)
- Possible without simultaneous comparison of samples

Definition of Metabolism

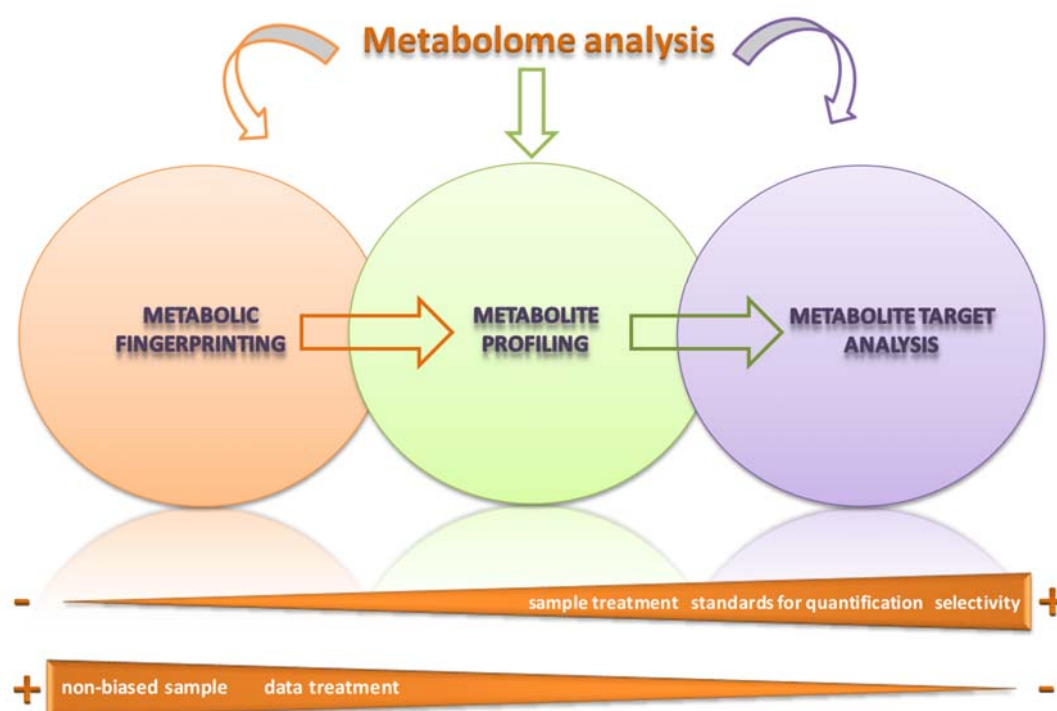


The complete group of (bio)chemical processes within an organelle, cell, tissue, organ or organism, essential for life

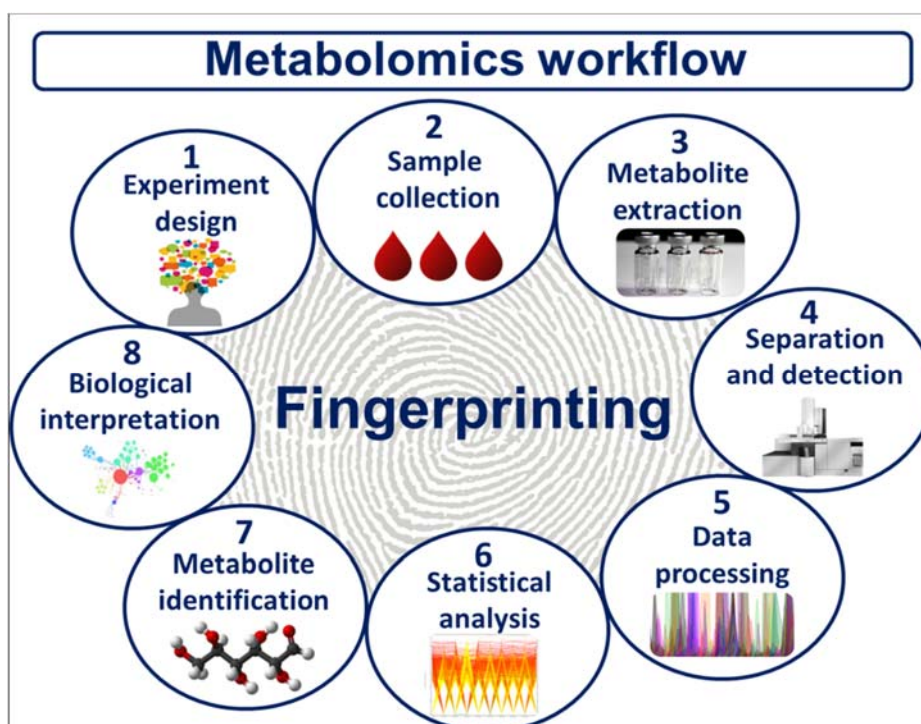
Analytical approaches in metabolomics



Three ways to do metabolomics



WORKFLOW



ANALYTICAL TECHNIQUES

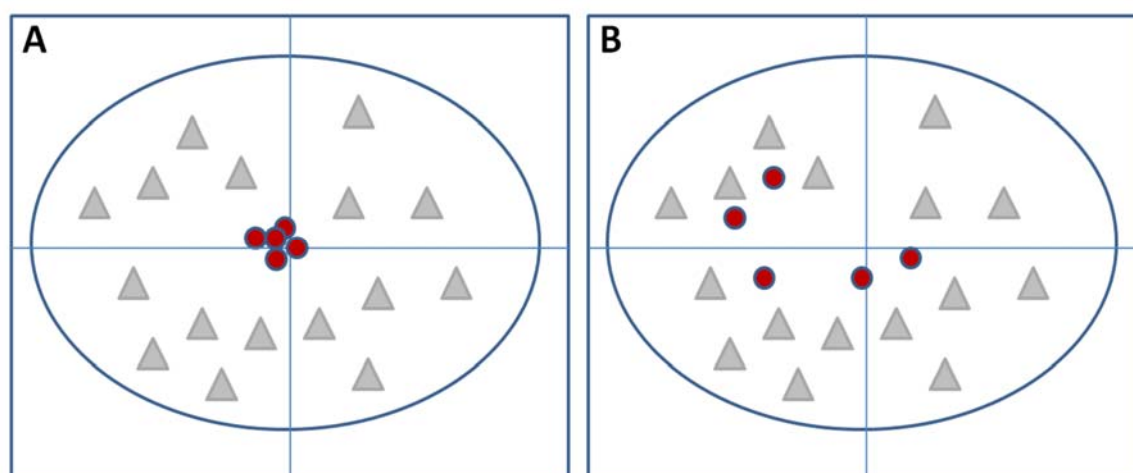
- GC/MS: Small polar compounds
 - Mainly water soluble (some hydrophobic)
 - Sample treatment: Derivatization
 - Fragmentation reproducible - databases
- NMR
 - Water-soluble
 - Virtually no sample treatment
 - High LOD
- LC/MS
 - from small to large (<1500 Da) medium to non-polar metabolites
- CE/MS: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

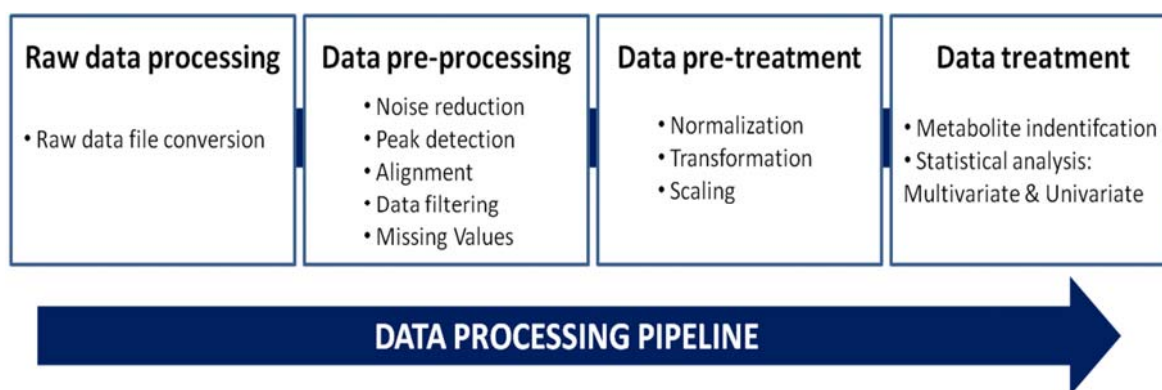
Quality Control and Quality Assurance Procedure in Metabolomics



A: QCs (red dots) clustered together

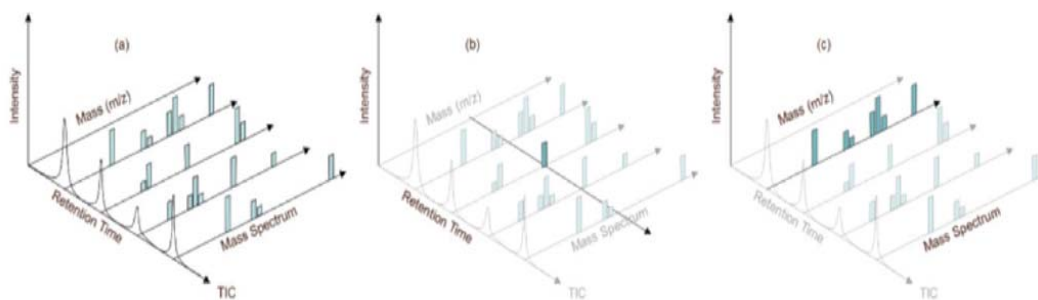
B: QCs spreaded

DATA TREATMENT IN METABOLOMICS: Signal Processing



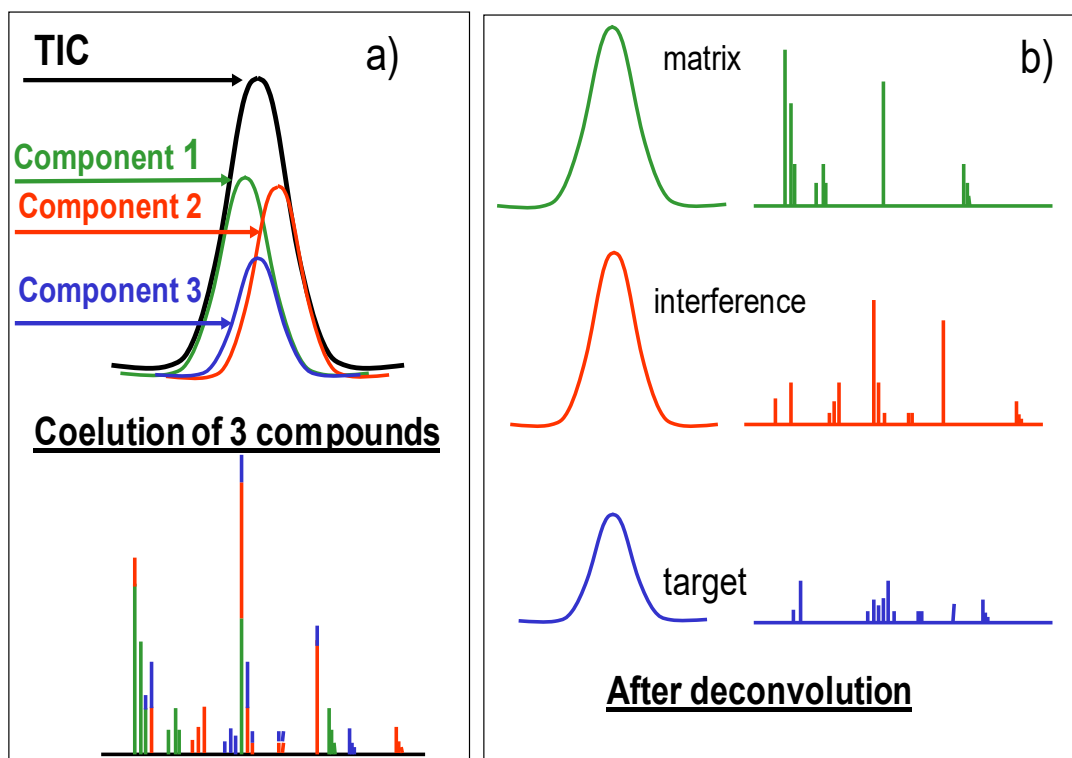
ANALYTICAL TECHNIQUE: GC-MS

- 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

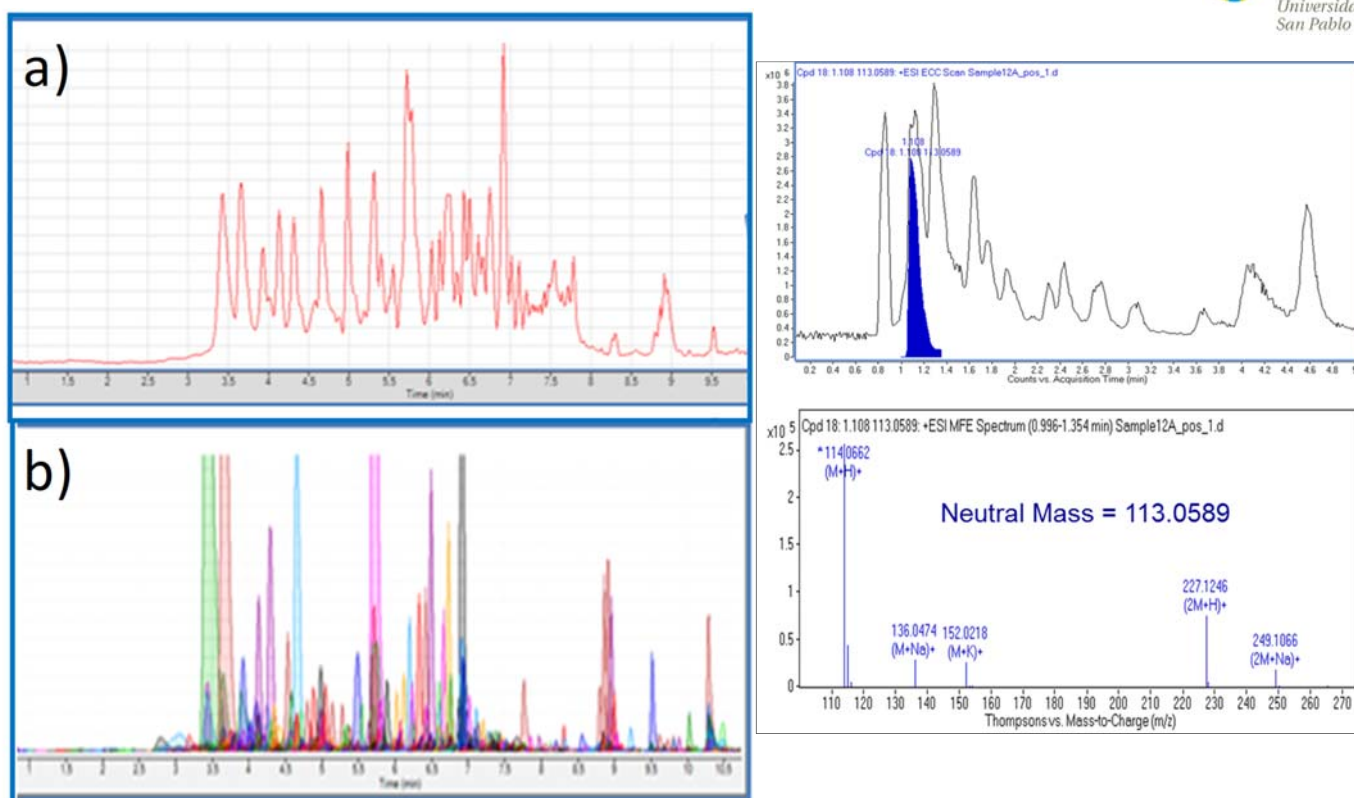


a) Before and b) After the deconvolution process

adapted from <https://www.agilent.com/cs/library/Support/Documents/f05017.pdf>

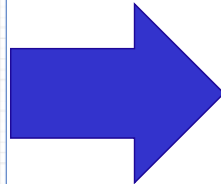
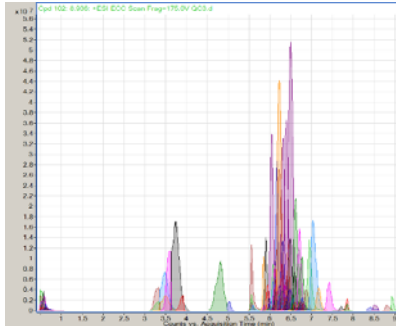
- Peak-based methods
- Molecular Feature Extractor (Agilent) considers the accuracy of the mass measurements to group related ions by charge-state envelope, isotopic distribution, and possible chemical relationships when determining whether different ions are from the same metabolic feature.
- It can 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?



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
☒		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	4059574		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	730.5360	729.5297	Positive	7	1506513	8484440		100

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 preprocessing

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:

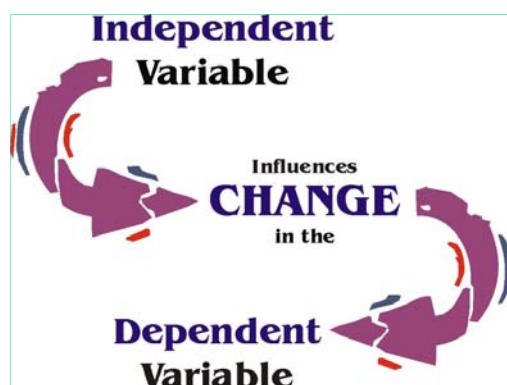
- o detect differences between sample groups at the chemical level
- o rank compounds by relative importance for sample differentiation

VARIABLES

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

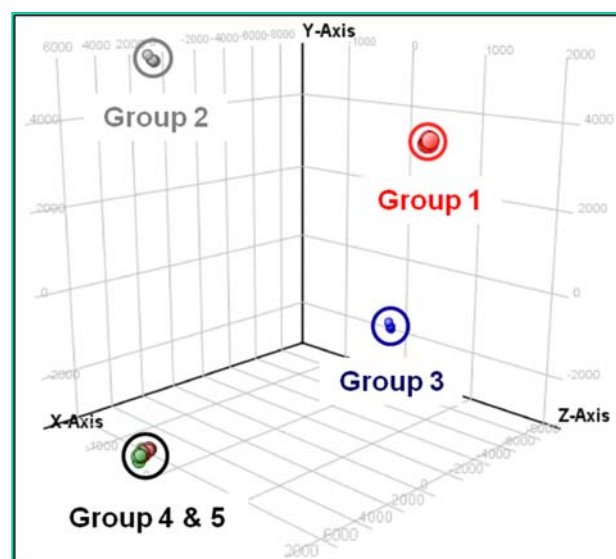
TYPES

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



PCA

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

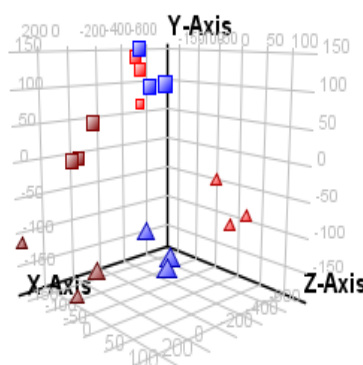
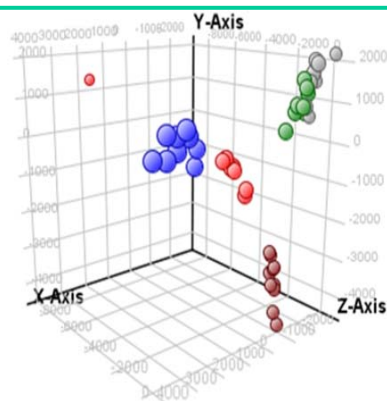


Class prediction: PLS-DA

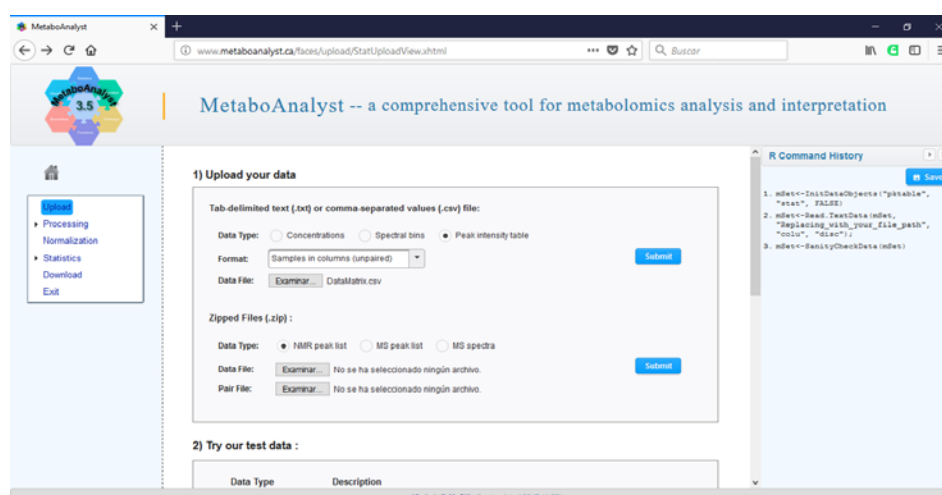
Partial Least Square - Discriminant Analysis Projection on Latent Structures - Discriminant Analysis

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

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

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

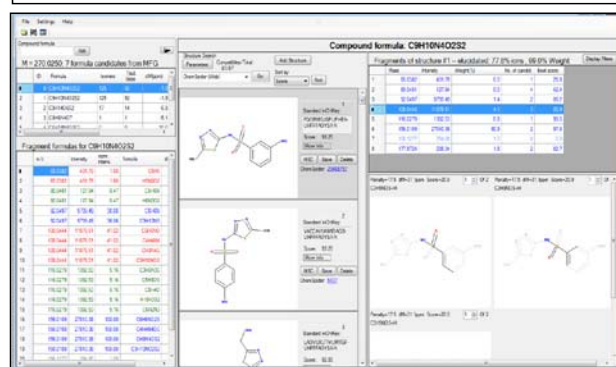
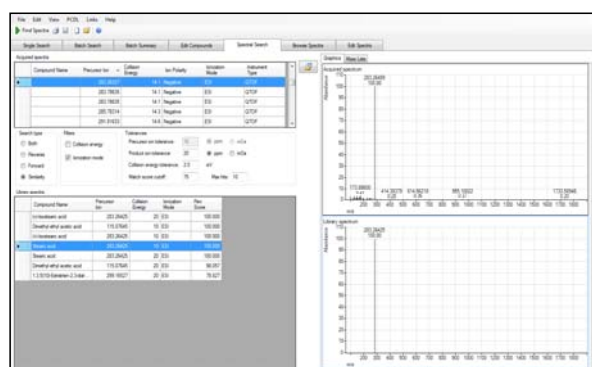
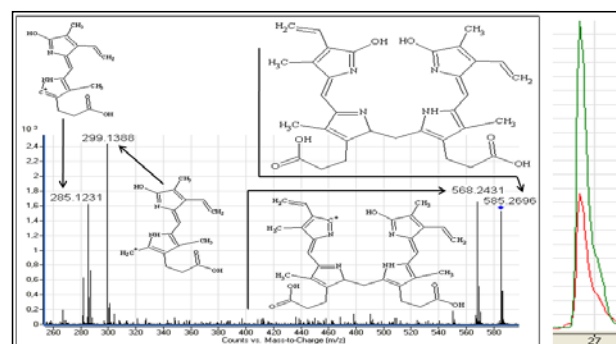
N - fold

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.nmr.fam.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.gamene.org/gamene/ricecyc.shtml
HumanCyc	http://www.genome.jp/kegg/	Serum Metabolome Database	http://www.serummetabolome.ca/
IIMDB	http://www.genome.jp/kegg/glycan/	SetupX & BinBase	http://fiehnlab.ucdavis.edu/projects/binbase_setupx

CEU MASS MEDIATOR

Seleccionar archivo Ningún archivo seleccionado

Experimental Masses (*):
enter significant input masses

Retention Times:
enter significant retention times

Composite Spectra:
enter significant composite spectra

Seleccionar archivo Ningún archivo seleccionado

All Experimental Masses:
enter all input masses

All Retention Times:
enter all retention times

All Composite Spectra:
enter all composite spectra

Chemical Alphabet (*):
All
CHNOPS
CHNOPS + Cl
Modifiers (*):
None
NH3
HCOO
CH3COO
HCOONH3
CH3COONH3

Databases (*):
All except MINE
All (including In Silico Compounds)
Kegg
HMDB
LipidMaps
Metlin
MINE (Only In Silico Compounds)
Metabolites (*):
All except peptides
Only lipids
All including peptides

Input Masses Mode (*):
Neutral Masses
m/z Masses

Ionization Mode (*):
Neutral
Positive Mode
Negative Mode
calculation of new m/z from neutral mass based on selected adducts

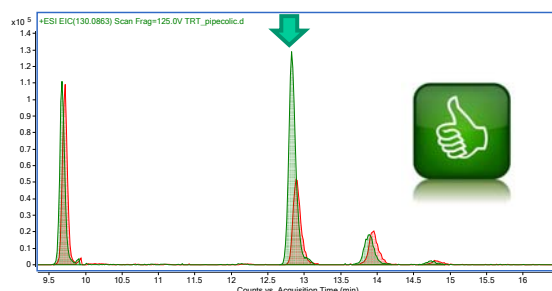
Adducts (*):
All
M+H
M+2H
M+Na
M+K
M+NH4

- 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>

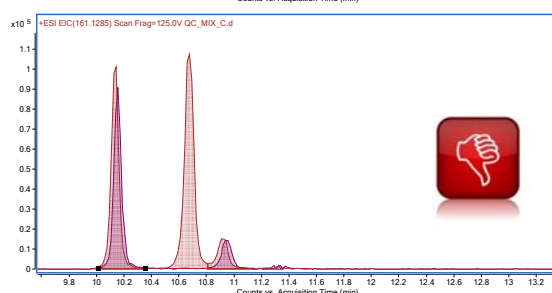
CEU MASS MEDIATOR

	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 Err	Molecular Weight Name	Formula	CAS	Kegg	HMDB	LipidMaps	Metlin	PubChem	InChIKey	Pathways	
3	399.3367	17732	M+H	5	399.3349 L-palmitoleic amine	C20H35NO4							XOMFRQKQHMVMOC-AFFANRHFSA-N		
4	399.3367	17751	M+H	5	399.3349 O-palmitoleic amine	C20H35NO4							XOMFRQKQHMVMOC-UHFFFAOYSA-N		
5	399.3367	17857	M+H	5	399.3349 Palmitoleic amine	C20H35NO4	2364-67-2	C02390	HMDB0000022	LMFA07070004	26867	11953816	XOMFRQKQHMVMOC-OAQVLSRUS	Fatty acid Metabolism	Fatty a
6	399.3367	0	M+Na	0	0 No compounds found for experimental mass 399.3367 and adduct: f										
7	399.3367	13442	M+H	5	382.3083 methyl 9-butylcroton-10,12-octadecadienoate	C23H42O4				LMFA07040026	24461		XVZNFQKQHMPCF-AZPFIYJSA-N		
8	399.3367	13443	M+H	5	382.3083 methyl 13-butylcroton-9,11-octadecadienoate	C23H42O4				LMFA07040027	24462		GUYVTNFPALBJK-AEPVOTSSA-N		
9	399.3367	95769	M+H	5	382.3083 Lipidum tepeperi ester	C23H42O4				HMDB0006865	11734320		PASMASQJCKBKJ-UHFFFAOYSA-N		
10	399.3367	53981	M+H	5	382.3083 Mq(0.6/20.2/12.1/2.0/0.0)	C23H42O4				HMDB0001544	62432	13480364	PMUSUEZTCFTBMD-HZYITRNSA-N		
11	399.3367	52646	M+H	5	382.3083 Mq(20.2/12.1/2.0/0.0/0.0)	C23H42O4				HMDB0001574	62435	13480363	QIBGFIYBCOYOSADT-ZJUSHA-N		
12	399.3367	80631	M+H	5	382.3083 Perseone B	C23H42O4				HMDB0005555	91845		NLQNLZUOMHEB-LSLVRVYSA-N		
13	399.3367	0	M+H+H2O	0	0 No compounds found for experimental mass 399.3367 and adduct: f										
14	421.1969	16360	M+H	6	421.1952 Gamma-linolenic amine	C26H43NO4				HMDB0006319	15472189		YOPMVLMLYSAGK-BAHSRKMSSA-N		
15	421.1969	96332	M+H	6	421.1952 Alpha-linolenic amine	C26H43NO4				HMDB0006319	15471921		DFVGGHKKDAHYU-UHMKZJMFSA-N		
16	421.1969	126612	M+H	9	421.3205 AGELASINE	C26H39N5					43731				
17	421.1969	138401	M+H	9	421.313 Latanoprost ethyl amide-d4	C26H39N4O4					96571				
18	421.1969	17732	M+Na	0	399.3349 L-palmitoleic amine	C20H35NO4									
19	421.1969	17751	M+Na	0	399.3349 O-palmitoleic amine	C20H35NO4									
20	421.1969	17857	M+Na	0	399.3349 Palmitoleic amine	C20H35NO4	2364-67-2	C02390	HMDB0000022	LMFA07070004	26867	11953816	XOMFRQKQHMVMOC-AFFANRHFSA-N		
21	421.1969	1294	M+H	6	404.2327 alpha,25-dihydroxy-21-nor-20-oxoamin D3 / alpha,25-dihydroxy-2	C26H40O4				LMST03020024	41970		ADMQGGQKVVZTNY-OGQZJSESA-N		
22	421.1969	1295	M+H	6	404.2327 alpha,25-dihydroxy-24-nor-22-oxoamin D3 / alpha,25-dihydroxy-2	C26H40O4				LMST03020026	41973		UMRLGLLMBCTAPNKGSCSA-N		
23	421.1969	2650	M+H	6	404.2327 Tri-Hydroxy-2-oxo-5b-cholan-24-ate	C26H40O4				LMST04070028	71939		XHFLTUHVHGDCJ-FFJYJMSA-N		
24	421.1969	117990	M+H	6	404.2327 Androstane-3,17-diol dipropionate, alpha-Androstane-3,17,20-tri	C26H40O4	4350-14-5	C14524			70213	124572	FVAYUVSHYKUEY-LFVYVBCSA-N		
25	421.1969	86607	M+H	6	404.2327 11-Carboxy-gamma-chromanol	C26H40O4				HMDB0001257	2649	13481453	ITULCNDMDKHA-YUUDOPRSA-N		
26	421.1969	89679	M+H	6	404.2327 Mq(0.6/22.5/12.1/2.0/16.2/0.0/0.0)	C26H40O4				HMDB0001556	62433	13480372	IRBULYLYUADN-LIKQSTISA-N		
27	421.1969	50261	M+H	6	404.2327 Mq(22.5/12.1/2.0/16.2/0.0/0.0)	C26H40O4				HMDB0001555	62437	13480363	HOIGCISTZKIDG-AJVTYFSA-N		
28	421.1969	105373	M+H	6	404.2327 Mq(0.6/22.5/12.1/2.0/16.2/0.0/0.0)	C26H40O4				HMDB0001555	62438	13480370	NP2VSBAAZ2LVO-VMPFHZDHA-N		
29	421.1969	56824	M+H	6	404.2327 Mq(22.5/12.1/2.0/16.2/0.0/0.0/0.0)	C26H40O4				HMDB0001558	62438	13480394	DSCYURGADTGA-YAVGMQFSA-N		
30	421.1969	17635	M+H+H2O	5	420.3296 3-hydroxylinoleic amine	C26H43NO5				LMFA07070042	2649		YGVYKASXVFNSTU-IGFVJSSA-N		
31	315.2424	17712	M+Na	2	315.241 Decanolic amine	C17H33NO4				HMDB0006551	10245190		LZOSYCMQKQFBU-UHFFFAOYSA-N		
32	315.2424	126591	M+Na	2	315.241 L-Hexanolic amine n-butyl ester	C17H33NO4					2649				
33	315.2424	17659	M+H	5	315.241 O-Decanoyl-R-carnitine	C17H33NO4	3992-45-8	C02329	HMDB0062631	LMFA07070006	2649	11953821	LZOSYCMQKQFBU-OAHLLOKOSA-N		
34	315.2424	0	M+H	0	0 No compounds found for experimental mass 315.2424 and adduct: f										
35	315.2424	14277	M+H	5	288.2144 8E-Heptadecenedioic acid	C17H30O4				LMFA07070053	74325		VDTSYDDUGHLVO-VOVJTEDSA-N		
36	315.2424	11819	M+H	5	288.2144 Plakotic acid	C17H30O4				C1758			ZCLJFHUADYARQ-CHMDGGBGSA-N		
37	315.2424	0	M+H+H2O	0	0 No compounds found for experimental mass 315.2424 and adduct: f										
38	315.2424	0	M+H	0	0 No compounds found for experimental mass 315.2424 and adduct: f										
39	315.2424	17712	M+Na	2	315.241 Decanolic amine	C17H33NO4				HMDB0006551	10245190		LZOSYCMQKQFBU-UHFFFAOYSA-N		
40	315.2424	126591	M+Na	2	315.241 L-Hexanolic amine n-butyl ester	C17H33NO4					2649				
41	315.2424	17659	M+Na	2	315.241 O-Decanoyl-R-carnitine	C17H33NO4	3992-45-8	C02329	HMDB0062631	LMFA07070006	2649	11953821	LZOSYCMQKQFBU-OAHLLOKOSA-N		
42	315.2424	1860	M+H	6	320.1888 itrolic acid	C19H28O4				LMST02020091	62434	13480374	KMLUFRPXPPT-DBZHQCSA-N		
43	315.2424	18692	M+H	6	320.1888 10beta-Hydroxy-6beta-isobutyrylfranoeremophane	C19H28O4				C08585	62434	1444237	VVBNAQLGGJMUQ-BIGGFYDOSA-N		
44	315.2424	130809	M+H	6	320.1888 (3a-CMEHC	C19H28O4	7083-09-2			HMDB0003923	62434	1444237	QDSRFRN2QKXNP2-UHFFFAOYSA-N		
45	315.2424	53396	M+H	6	320.1888 (3b-CMEHC	C19H28O4	7734-06-6			HMDB0003923	62434	1444237	QDSRFRN2QKXNP2-UHFFFAOYSA-N		
46	315.2424	42968	M+H	6	320.1888 5,7-dihydroxy-alpha-chromanol	C19H28O4				HMDB0003923	62434	1444237	QDSRFRN2QKXNP2-UHFFFAOYSA-N		

Confirmation by Standard addition



Pipecolic acid

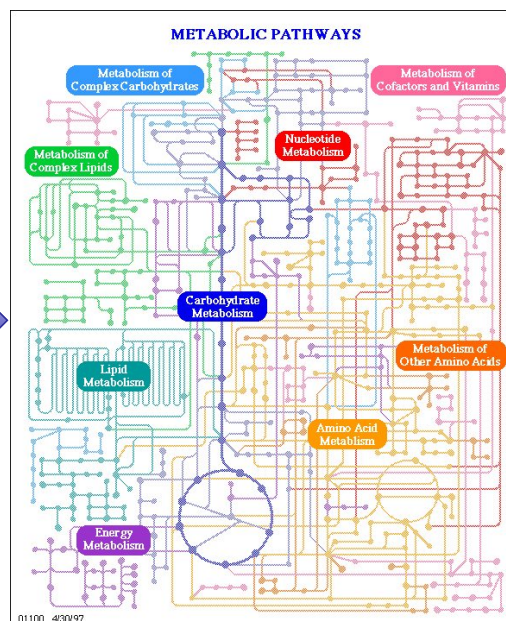


Methyl-lysine

From Lists to Pathways

metabolomics

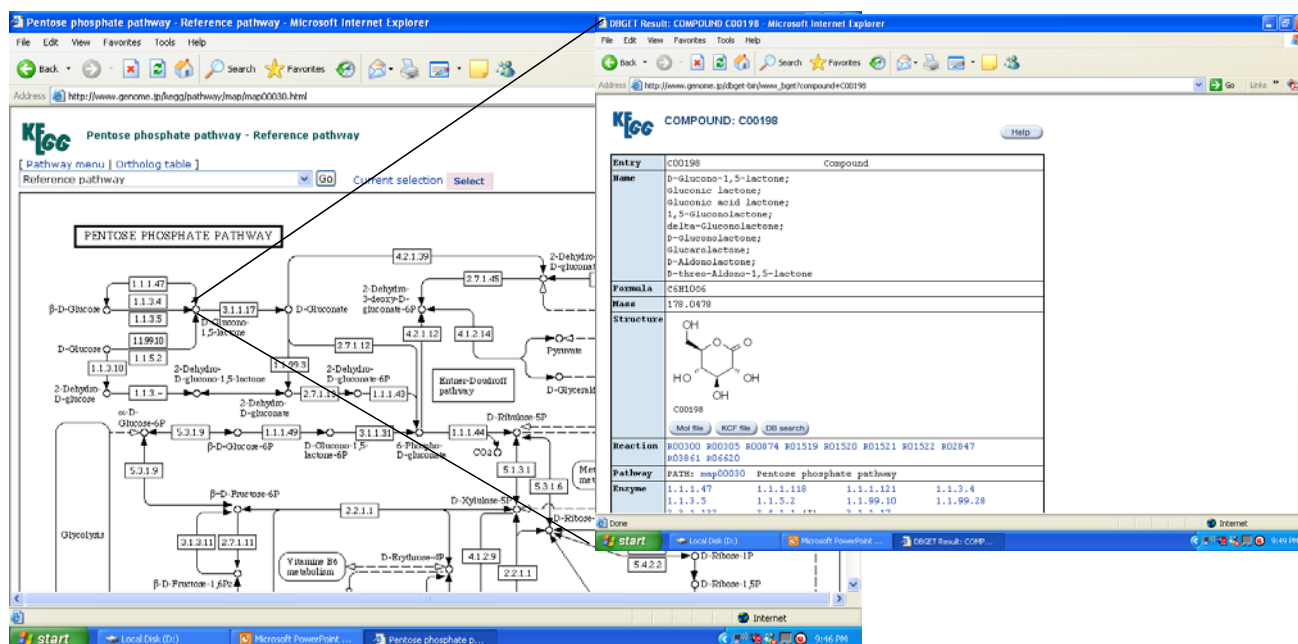
Compound	Retention Time (min)	Conc. in Urine (μM)	Compound	Retention Time (min)	Conc. in Urine (μM)
Dns-o-phospho -L-serine	0.92	<0.1	Dns-Ile	6.35	25
Dns-o-phospho -L-tyrosine	0.95	<0.1	Dns-3-aminosalicylic acid	6.44	0.5
Dns-α-phosphoethanolamine	0.99	<0.1	Dns-pipecolic acid	6.50	0.5
Dns-glucosamine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



Pathway Databases

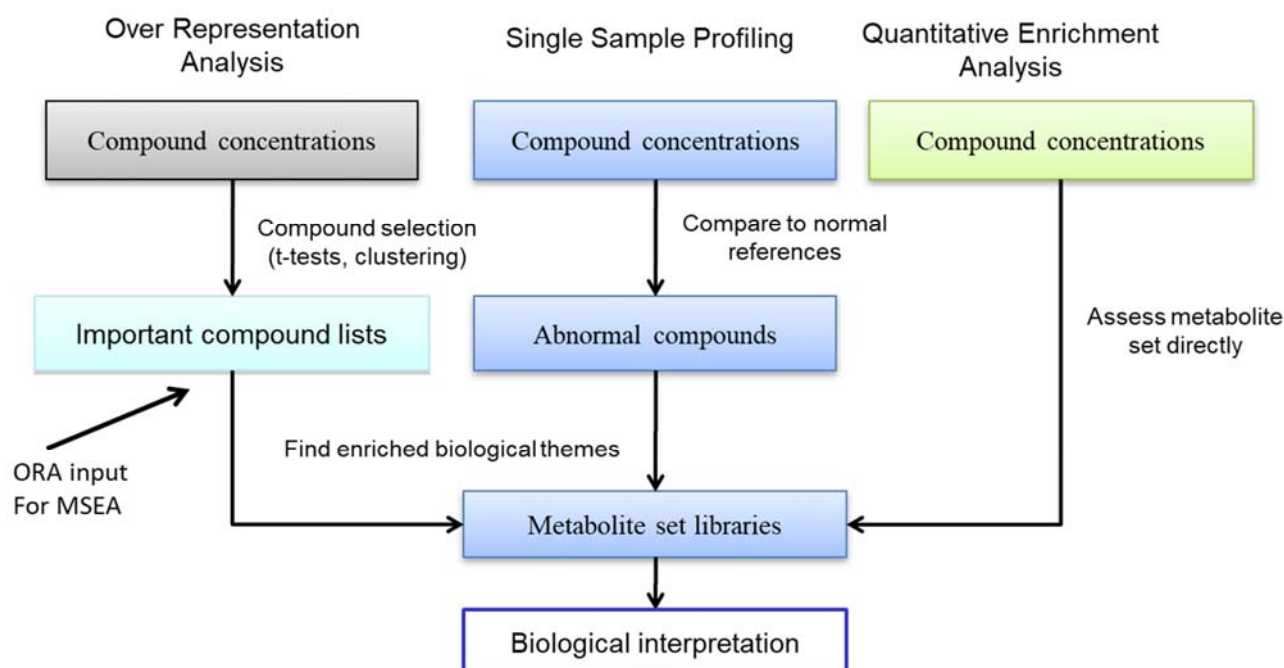
- Rich source of biological data that relates metabolites to genes, proteins, diseases, signaling events and processes
- Provide various tools to permit visualization and gene/metabolite mapping
- Often cover multiple species
- KEGG (www.genome.jp/kegg/), BioCyc/MetaCyc (<https://biocyc.org/>), SMPDB (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

MetaboAnalyst
- a web service for metabolomic data analysis

Home | Statistical Analysis | **Enrichment Analysis** | Pathway Analysis | Time Series | Other Utilities

Steps

- Upload** (highlighted with a red box and arrow)
- Processing
- Statistics
- Enrichment
- Pathway
- Time Series
- Peak search
- Metabolites
- Download
- Log out

Upload options:

- ▶ A list of compound names (over representation analysis)
- ▶ A list of compounds with concentration values (single sample profiling)
- ▶ A concentration table (quantitative enrichment analysis)

Home | Statistical Analysis | **Enrichment Analysis** | Pathway Analysis | Time Series | Other Utilities

Steps

- Upload** (highlighted with a red box)
- Processing
- Statistics
- Enrichment
- Pathway
- Time Series
- Peak search
- Metabolites
- Download
- Log out

Upload options:

- ▶ A list of compound names (over representation analysis)
- ▶ A list of compounds with concentration values (single sample profiling)
- ▶ A concentration table (quantitative enrichment analysis)

Concentration Comparison

[Home](#)

Steps

- [Upload](#)
 - [Processing](#)
 - [Pre-process](#)
 - [Name check](#)
 - [Conc. check](#)
 - [Data check](#)
 - [Missing value](#)
 - [Data filter](#)
 - [Data editor](#)
 - [Color picker](#)
 - [Normalization](#)
 - [Statistics](#)
 - [Enrichment](#)
 - [Pathway](#)
 - [Time Series](#)
 - [Peak search](#)
 - [Metabolites](#)
 - [Download](#)
 - [Log out](#)

Comparison with Reference Concentration

Note: *reference concentrations* are in the form of **mean(min - max)** format. In cases where the ranges were not reported in the original literature, the min and max were calculated using the 95% confidence intervals. In the *Comparison* column, **H, M, L** means **higher, medium (within range), lower** compared to the reference concentrations. Click the **Image Icon** link to see a graphical summary for the comparisons.

Compound	Concentration	Reference Concentrations	Comparison	Detail	Include
L-Isoleucine	0.34	1.579 (0.789 - 2.368); 0.94 (0.27 - 1.61); 3.75 (1 - 6.5); 3 (1.5 - 4.5); 1.8 (0.8 - 2.8)	M		☐
Fumaric acid	0.47	10.4 (2.8 - 53.7); 0.5 (0.1 - 1.7); 1 (0 - 2); 0.95 (0.02 - 1.88); 0.8 (0.1 - 1.7); 10.7 (0.1 - 28.2); 4.8 (0 - 35.2); 5 (1 - 33.5)	M		☐
Acetone	0.58	4.2 (0.98 - 15.3); 0.92 (0.2 - 2.8); 320 (103 - 1290); 20 (2 - 180); 15.3 (2 - 120)	M		☐
Succinic acid	9.4	14.4 (9.5 - 19.3); 3.8 (1.25 - 6.7); 12.6 (0.47 - 24.73); 14.48 (11.28 - 17.68); 9.9 (4.9 - 14.9); 39 (37 - 41); 197.2 (29.4 - 486.2); 185.4 (6 - 342.6); 7.7 (1.9 - 20); 11.6 (4 - 27.3); 8.25 (0.5 - 16)	M		☐
1-Methylhistidine	9.6	2.3 (0 - 7.4); 33.6 (0 - 70); 28.1 (0 - 59.9); 30 (0 - 73); 45.5 (3.9 - 87.1); 1.3 (0 - 4.06); 4.6 (1.9 - 7.3); 46.1 (0 - 99.6); 15.9 (0 - 35.4)	M		☐
L-Asparagine	19.62	35 (16.4 - 57.2); 9.211 (3.289 - 15.1); 0.96 (0.31 - 1.61); 10 (4.6 - 16.32)	M		☐
3-Methylhistidine	9.7	42.76 (19.92 - 65.6); 15.1 (3.9 - 26.3); 12.5 (8.3 - 16.7)	M		☐
L-Threonine	93.19	36.2 (10.82 - 61.58); 12.7 (4.934 - 20.4); 1 (0.16 - 2.4); 4.9 (2.4 - 7.4); 16 (7 - 25); 18 (8.4 - 27.6)	H		☒
Creatine	720	46 (9 - 135); 113 (0 - 654); 26 (5 - 95); 167 (124 - 210); 212 (0 - 5000); 450 (0 - 10000)	M		☐

Quantitative Enrichment Analysis

[Home](#)

Steps

- [Upload](#)
 - [Processing](#)
 - [Statistics](#)
 - [Enrichment](#)
 - [Pathway](#)
 - [Time Series](#)
 - [Peak search](#)
 - [Metabolites](#)
 - [Download](#)
 - [Log out](#)

Statistical Analysis | **Enrichment Analysis** | Pathway Analysis | Time Series | Other Utilities

► A list of compound names (over representation analysis)

► A list of compounds with concentration values (single sample profiling)

▼ A concentration table (quantitative enrichment analysis)

Upload your concentration data (.csv)

[Format](#)

[Browse...](#)

Compound Label Type:

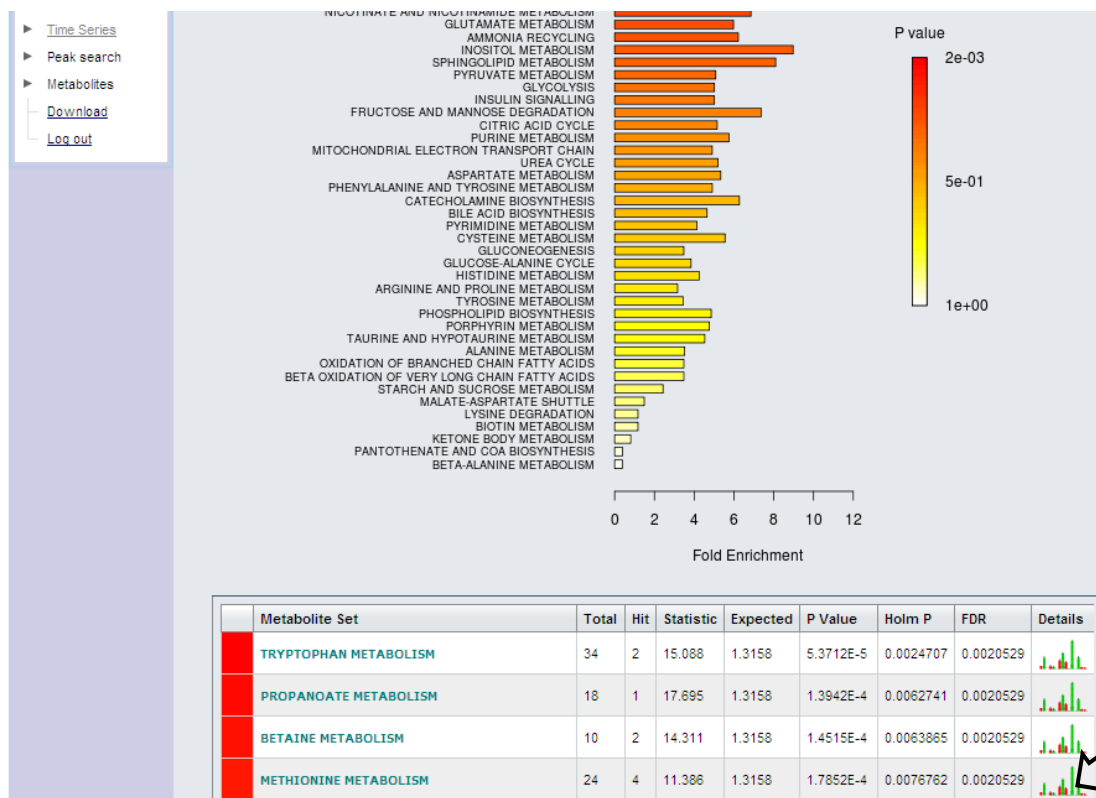
Phenotype Label:

[Submit](#)

Try our test data:

Data	Compound	Phenotype	Description
☒ Data 1	Common name	Discrete	Urinary metabolite concentrations from 77 cancer patients measured by 1H NMR. Phenotype: N - cachexic; Y - control
☐ Data 2	PubChem CID	Continuous	Urinary metabolite concentrations from 97 cancer patients measured by 1H NMR. Phenotype: muscle gain (percentage within 100 days, negative values indicate muscle loss)

[Submit](#)



Metaboanalyst Metabolic Pathway Analysis (MetPA)

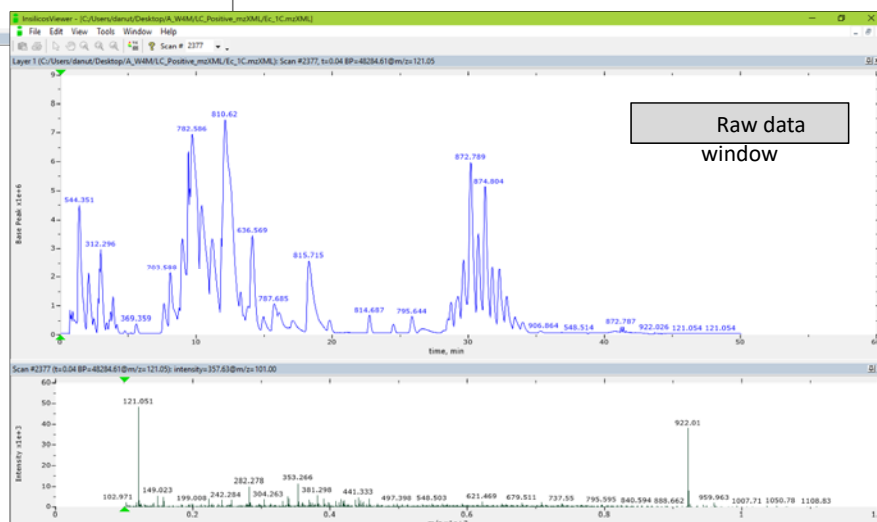
- Purpose: to extend and enhance metabolite set enrichment analysis for pathways by
 - Considering the **structures of pathway**
 - Dynamic pathway visualization
- Currently supports ~1500 pathways covering 17 organisms (based on KEGG)

PRACTICAL SESSION. VISUALS_1

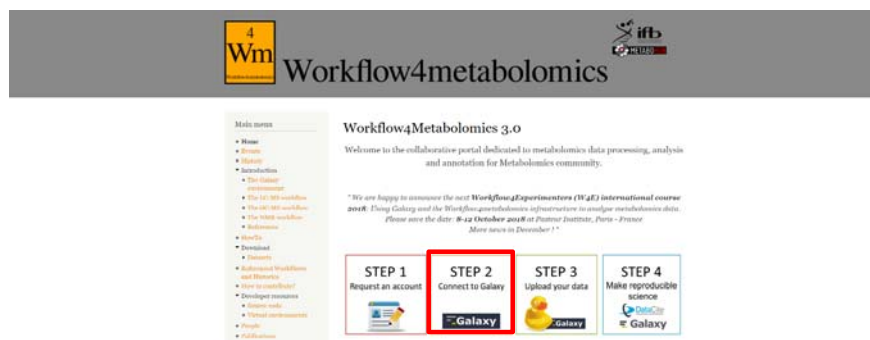
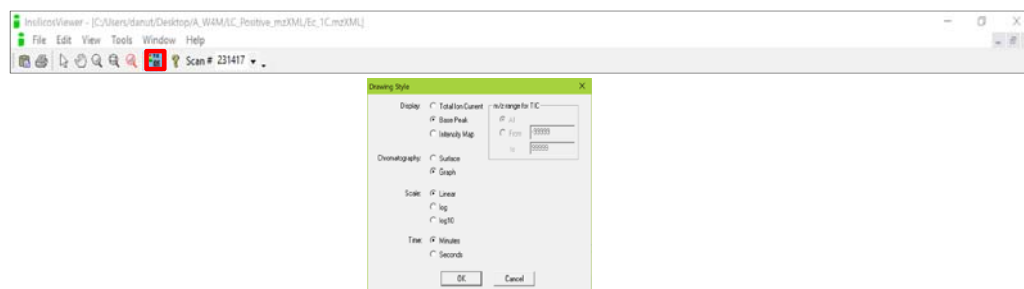
The screenshot shows the Workflow4Metabolomics 3.0 website on the left and the MSConvert application on the right. The website features a 'Main menu' with links to Home, About, Downloads, and more. It also includes a 'Workflow4Metabolomics 3.0' section with a welcome message and a four-step process: STEP 1 Request an account, STEP 2 Connect to Galaxy, STEP 3 Upload your data, and STEP 4 Make reproducible science. The MSConvert application is open, showing the 'About MSConvert' dialog box and the 'Parameters' tab. The 'Parameters' tab includes options for Output format (mzXML), Output extension, Binary encoding precision (64-bit), Write index, Use zlib compression, TPP compatibility, and Package in gap.

PRACTICAL SESSION. VISUALS_2

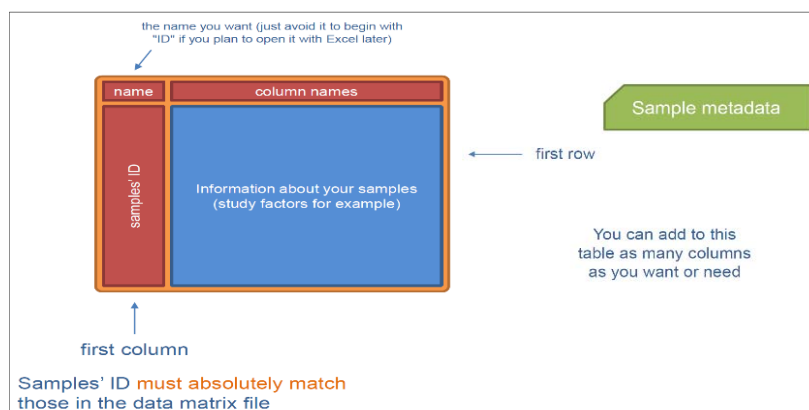
The screenshot shows the Icos Life Science Software interface. The main window displays a list of files in the 'LC_Positive_mzXML' directory. The files are organized by 'Nazwa' (Name), 'Data modyfikacji' (Modification date), and 'Typ' (Type). The files are listed in a table with columns for 'Nazwa', 'Data modyfikacji', and 'Typ'. The 'Nazwa' column shows file names like 'Ec_1C', 'Ec_2C', 'Ec_3C', etc. The 'Data modyfikacji' column shows dates like '19.4.17 13:47', '19.4.17 13:49', etc. The 'Typ' column shows file types like 'mass spec', 'mass spec', etc. The 'Nazwa pliku' (File name) field is set to 'Ec_1C' and the 'Plik typu' (File type) is set to 'MS Data File (*.mzXML;*.mzData;*.mzML;*.RA)'. The 'Otwórz' (Open) button is visible.



PRACTICAL SESSION. VISUALS_3

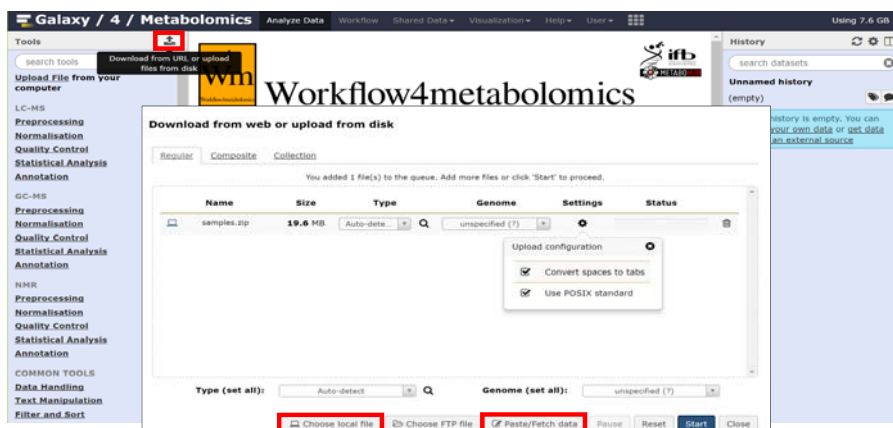


PRACTICAL SESSION. VISUALS_4

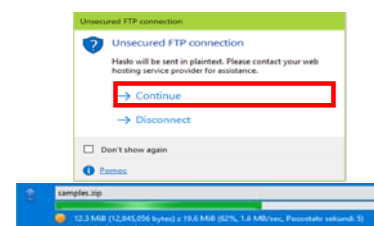
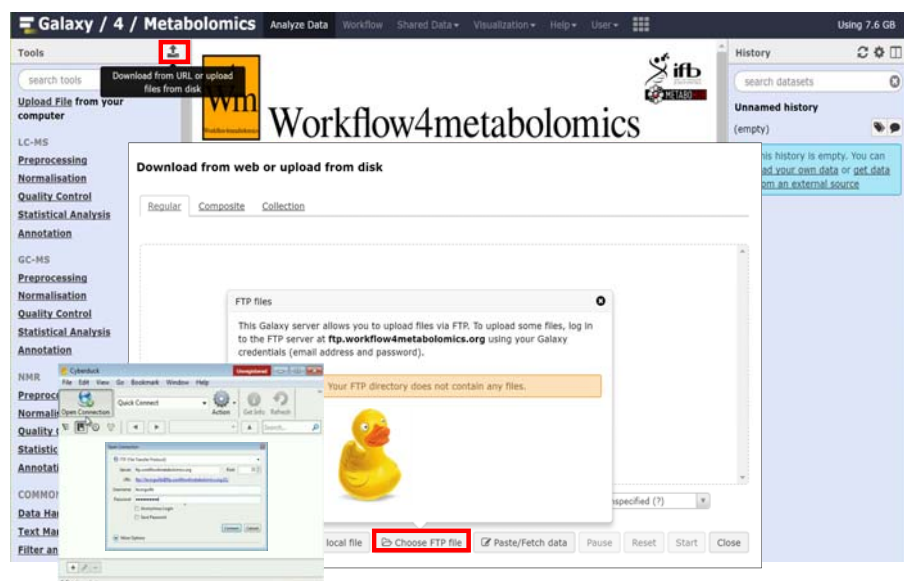


PRACTICAL SESSION. VISUALS_5

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_8

Download from web or upload from disk

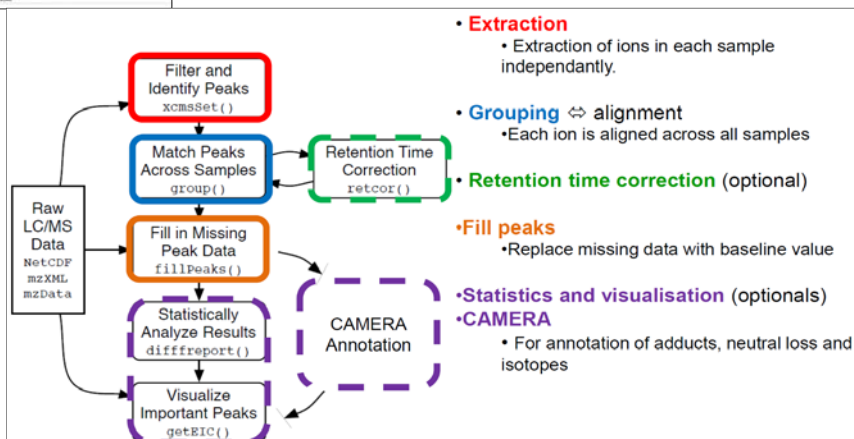
Regular Composite Collection

Name	Size	Type	Genome	Settings	Status
 LC_Positive_mzXML.zip	3.6 GB	Auto-detect  	unspecified (?) 		100% 

Type (set all): 

Genome (set all): 

 Choose local file  Choose FTP file  Paste/Fetch data Pause Reset Start **Close**

[illegible]

PRACTICAL SESSION. VISUALS_9

Galaxy / 4 / Metabolomics Analyze Data | Workflow | Shared Data | Visualization | Help | Tools | Using 15.6 GB

Tools

search tools

Upload File from your computer

LC-MS

Preprocessing

xcms.xcmsSet Filtration and Peak Identification using xcmsSet function from xcms R package to preprocess LC/MS data for relative quantification and statistical analysis

xcms.xcmsSet Merge xcms.xcmsSet in one to be used by group

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

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

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

xcms.summary Create a

1 job has been successfully added to the queue - resulting in the following datasets:

- 2: LC_Positive_mzXML.xset.RData
- 3: LC_Positive_mzXML.sampleMetadata.tsv
- 4: LC_Positive_mzXML.xset.TICs.raw.pdf
- 5: LC_Positive_mzXML.xset.BPCs.raw.pdf
- 6: LC_Positive_mzXML.xset.log.txt

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from "running" to "finished" if completed successfully or "error" if problems were encountered.

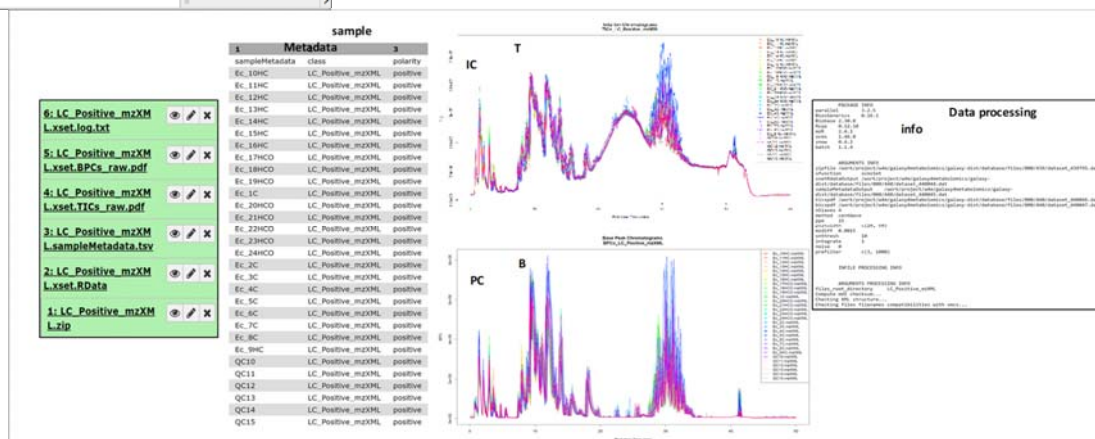
History

search datasets

Unnamed history

3.61 GB

- 6: LC_Positive_mzXML.xset.log.txt
- 5: LC_Positive_mzXML.xset.BPCs.raw.pdf
- 4: LC_Positive_mzXML.xset.TICs.raw.pdf
- 3: LC_Positive_mzXML.sampleMetadata.tsv
- 2: LC_Positive_mzXML.xset.RData
- 1: LC_Positive_mzXML.L.zip



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.

Independent peak list

Group ions by m/z

Group ions by RT

pool1B1				pool1B2				pool1B3			
mz	rt	int		mz	rt	int		mz	rt	int	
196.0905	66.6	7810936		196.0910	66.7	11733921		196.0902	66.6	7933325	
158.1180	67.4	71736		342.0310	69.0	74594		158.1173	67.4	82969	
342.0308	67.6	202268		267.0581	65.5	260877		342.0308	21.3	2581	
267.0581	65.5	282039		283.0318	65.2	424631		283.0320	65.3	357448	

pool1B1				pool1B2				pool1B3			
mz	rt	int		mz	rt	int		mz	rt	int	
196.0905	66.6	7810936		196.0910	66.7	11733921		196.0902	66.6	7933325	
158.1180	67.4	71736		342.0310	69.0	74594		158.1173	67.4	82969	
342.0308	67.6	202268		267.0581	65.5	260877		342.0308	21.3	2581	
267.0581	65.5	282039		283.0318	65.2	424631		283.0320	65.3	357448	

pool1B1				pool1B2				pool1B3			
mz	rt	int		mz	rt	int		mz	rt	int	
196.0905	66.6	7810936		196.0910	66.7	11733921		196.0902	66.6	7933325	
158.1180	67.4	71736		342.0310	69.0	74594		158.1173	67.4	82969	
342.0308	67.6	202268		267.0581	65.5	260877		342.0308	21.3	2581	
267.0581	65.5	282039		283.0318	65.2	424631		283.0320	65.3	357448	

pool1B1				pool1B2				pool1B3			
mz	rt	int		mz	rt	int		mz	rt	int	
196.0905	66.6	7810936		196.0910	66.7	11733921		196.0902	66.6	7933325	
158.1180	67.4	71736		342.0310	69.0	74594		158.1173	67.4	82969	
342.0308	67.6	202268		267.0581	65.5	260877		342.0308	21.3	2581	
267.0581	65.5	282039		283.0318	65.2	424631		283.0320	65.3	357448	

pool1B1				pool1B2				pool1B3			
mz	rt	int		mz	rt	int		mz	rt	int	
196.0905	66.6	7810936		196.0910	66.7	11733921		196.0902	66.6	7933325	
158.1180	67.4	71736		342.0310	69.0	74594		158.1173	67.4	82969	
342.0308	67.6	202268		267.0581	65.5	260877		342.0308	21.3	2581	
267.0581	65.5	282039		283.0318	65.2	424631		283.0320	65.3	357448	

pool1B1				pool1B2				pool1B3			
mz	rt	int		mz	rt	int		mz	rt	int	
196.0905	66.6	7810936		196.0910	66.7	11733921		196.0902	66.6	7933325	
158.1180	67.4	71736		342.0310	69.0	74594		158.1173	67.4	82969	
342.0308	67.6	202268		267.0581	65.5	260877		342.0308	21.3	2581	
267.0581	65.5	282039		283.0318	65.2	424631		283.0320	65.3	357448	

pool1B1				pool1B2				pool1B3			
mz	rt	int		mz	rt	int		mz	rt	int	
196.0905	66.6	7810936		196.0910	66.7	11733921		196.0902	66.6	7933325	
158.1180	67.4	71736		342.0310	69.0	74594		158.1173	67.4	82969	
342.0308	67.6	202268		267.0581	65.5	260877		342.0308	21.3	2581	
267.0581	65.5	282039		283.0318	65.2	424631		283.0320	65.3	357448	

pool1B1				pool1B2				pool1B3			
mz	rt	int		mz	rt	int		mz	rt	int	
196.0905	66.6	7810936		196.0910	66.7	11733921		196.0902	66.6	7933325	
158.1180	67.4	71736		342.0310	69.0	74594		158.1173	67.4	82969	
342.0308	67.6	202268		267.0581	65.5	260877		342.0308	21.3	2581	
267.0581	65.5	282039		283.0318	65.2	424631		283.0320	65.3	357448	

pool1B1				pool1B2				pool1B3			
mz	rt	int		mz	rt	int		mz	rt	int	
196.0905	66.6	7810936		196.0910	66.7	11733921		196.0902	66.6	7933325	
158.1180	67.4	71736		342.0310	69.0	74594		158.1173	67.4	82969	
342.0308	67.6	202268		267.0581	65.5	260877		342.0308	21.3	2581	
267.0581	65.5	282039		283.0318	65.2	424631		283.0320	65.3	357448	

pool1B1				pool1B2				pool1B3			
mz	rt	int		mz	rt	int		mz	rt	int	
196.0905	66.6	7810936		196.0910	66.7	11733921		196.0902	66.6	7933325	
158.1180	67.4	71736		342.0310	69.0	74594		158.1173	67.4	82969	
342.0308	67.6	202268		267.0581	65.5	260877		342.0308	21.3	2581	
267.0581	65.5	282039		283.0318	65.2	424631		283.0320	65.3	357448	

pool1B1				pool1B2				pool1B3			
mz	rt	int		mz	rt	int		mz	rt	int	
196.0905	66.6	7810936		196.0910	66.7	11733921		196.0902	66.6	7933325	
158.1180	67.4	71736		342.0310	69.0	74594		158.1173	67.4	82969	
342.0308	67.6	202268		267.0581	65.5	260877		342.0308	21.3	2581	
267.0581	65.5	282039		283.0318	65.2	424631		283.0320	65.3	357448	

pool1B1				pool1B2				pool1B3			
mz	rt	int		mz	rt	int		mz	rt	int	
196.0905	66.6	7810936		196.0910	66.7	11733921		196.0902	66.6	7933325	
158.1180	67.4	71736		342.0310	69.0	74594		158.1173	67.4	82969	
342.0308	67.6	202268		267.0581	65.5	260877		342.0308	21.3	2581	
267.0581	65.5	282039		283.0318	65.2	424631		283.0320	65.3	357448	

pool1B1				pool1B2				pool1B3			
mz	rt	int		mz	rt	int		mz	rt	int	
196.0905	66.6	7810936		196.0910	66.7	11733921		196.0902	66.6	7933325	
158.1180	67.4	71736		342.0310	69.0	74594		158.1173	67.4	82969	
342.0308	67.6	202268		267.0581	65.5	260877		342.0308	21.3	2581	
267.0581	65.5	282039		283.0318	65.2	424631		283.0320	65.3	357448	

pool1B1				pool1B2				pool1B3			
mz	rt	int		mz	rt	int		mz	rt	int	
196.0905	66.6	7810936		196.0910	66.7	11733921		196.0902	66.6	7933325	
158.1180	67.4	71736		342.0310	69.0	74594		158.1173	67.4	82969	
342.0308	67.6	202268		267.0581	65.5	260877		342.0308	21.3	2581	
267.0581	65.5	282039		283.0318	65.2	424631		283.0320	65.3	357448	

pool1B1				pool1B2				pool1B3			
mz	rt	int		mz	rt	int		mz	rt	int	
196.0905	66.6	7810936		196.0910	66.7	11733921		196.0902	66.6	7933325	
158.1180	67.4	71736		342.0310	69.0	74594		158.1173	67.4	82969	
342.0308	67.6	202268		267.0581	65.5	260877		342.0308	21.3	2581	
267.0581	65.5	282039		283.0318	65.2	424631		283.0320	65.3	357448	

pool1B1				pool1B2				pool1B3			
mz	rt	int		mz	rt	int		mz	rt	int	
196.0905	66.6	7810936		196.0910	66.7	11733921		196.0902	66.6	7933325	
158.1180	67.4	71736		342.0310	69.0	74594		158.1173	67.4	82969	
342.0308	67.6	202268		267.0581	65.5	260877		342.0308	21.3	2581	
267.0581	65.5	282039		283.0318	65.2	424631		283.0320	65.3	357448	

pool1B1				pool1B2				pool1B3			
mz	rt	int		mz	rt	int		mz	rt	int	
196.0905	66.6	7810936		196.0910	66.7	11733921		196.0902	66.6	7933325	
158.1180	67.4	71736		342.0310	69.0	74594		158.1173	67.4	82969	
342.0308	67.6	202268		267.0581	65.5	260877		342.0308	21.3	2581	
267.0581	65.5	282039		283.0318	65.2	424631		283.0320	65.3	357448	

pool1B1				pool1B2				pool1B3			
mz	rt	int		mz	rt	int		mz	rt	int	
196.0905	66.6	7810936		196.0910	66.7	11733921		196.0902	66.6	7933325	
158.1180	67.4	71736		342.0310	69.0	74594		158.1173	67.4	82969	
342.0308	67.6	202268		267.0581	65.5	260877		342.0308	21.3	2581	
267.0581	65.5	282039		283.0318	65.2	424631		283.0320	65.3	357448	

pool1B1				pool1B2				pool1B3			
mz	rt	int		mz	rt	int		mz	rt	int	
196.0905	66.6	7810936		196.0910	66.7	11733921		196.0902	66.6	7933325	
158.1180	67.4	71736		342.0310	69.0	74594		158.1173	67.4	82969	
342.0308	67.6	202268		267.0581	65.5	260877		342.0308	21.3	2581	
267.0581	65.5	282039		283.0318	65.2	424631		283.0320	65.3	357448	

pool1B1				pool1B2				pool1B3			
mz	rt	int		mz	rt	int		mz	rt	int	
196.0905	66.6	7810936		196.0910	66.7	11733921		196.0902	66.6	7933325	
158.1180	67.4	71736		342.0310	69.0	74594		158.1173	67.4	82969	
342.0308	67.6	202268		267.0581	65.5	260877		342.0308	21.3	2581	
267.0581	65.5	282039		283.0318	65.2	424631		283.0320	65.3	357448	

pool1B1				pool1B2				pool1B3			
mz	rt	int		mz	rt	int		mz	rt	int	
196.0905	66.6	7810936		196.0910	66.7	11733921		196.0902	66.6	7933325	
158.1180	67.4	71736		342.0310	69.0	74594		158.1173	67.4	82969	
342.0308	67.6	202268		267.0581	65.5	260877		342.0308	21.3	2581	
267.0581	65.5	282039		283.0318	65.2	424631		283.0320	65.3	357448	

pool1B1				pool1B2				pool1B3			
mz	rt	int		mz	rt	int		mz	rt	int	
196.0905	66.6	7810936		196.0910	66.7	11733921		196.0902	66.6	7933325	
158.1180	67.4	71736		342.0310	69.0	74594		15			

PRACTICAL SESSION. VISUALS_11

xcms.group Group peaks together across samples using overlapping m/z bins and calculation of smoothed peak distributions in chromatographic time. (Galaxy Version 2.1.0)

xset RData file

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

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

Method to use for grouping

density

[method] See the help section below

Bandwidth

30

[bw] bandwidth (standard deviation or half width at half maximum) of gaussian smoothing kernel to apply to the peak density chromatogram

Minimum fraction of samples necessary

0.5

[minfrac] In at least one of the sample groups for it to be a valid group

Width of overlapping m/z slices

0.01

[mzwid] to use for creating peak density chromatograms and grouping peaks across samples

Advanced options

show

Maximum number of groups to identify in a single m/z slice

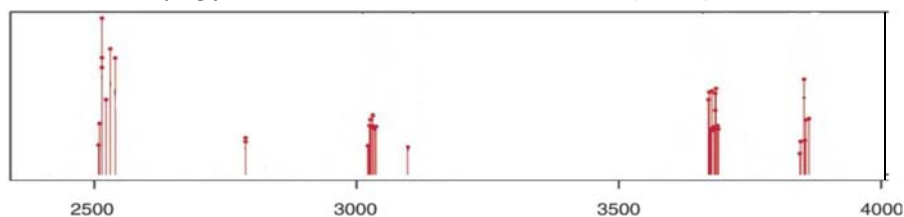
50

[max]

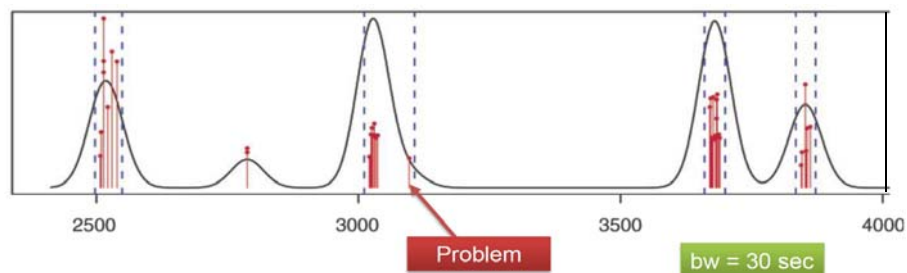
Get a Peak List

PRACTICAL SESSION. VISUALS_12

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

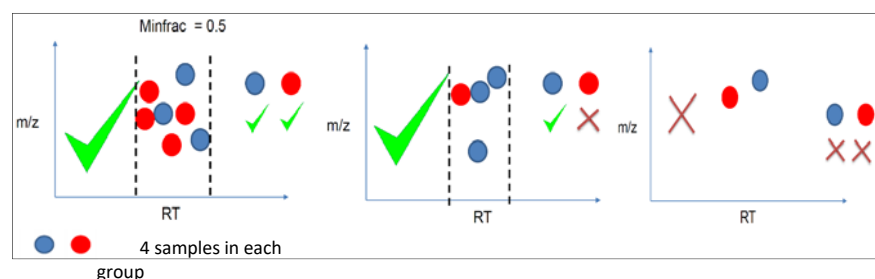
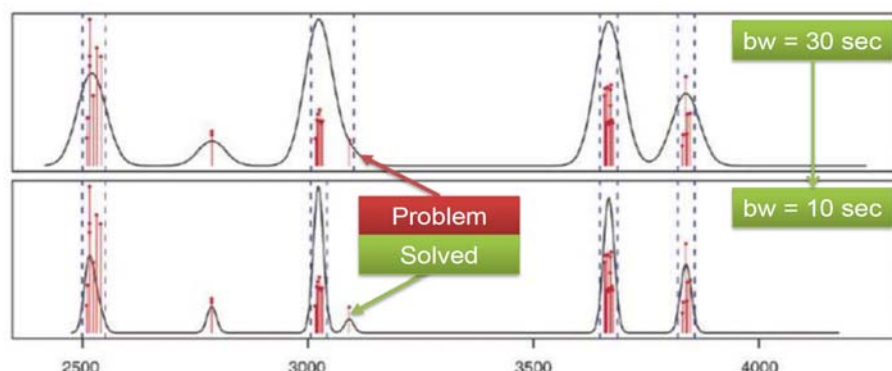


Grouping peaks in mass bin: 337.975 – 338.225 m/z (mzwid)



PRACTICAL SESSION. VISUALS_13

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



PRACTICAL SESSION. VISUALS_14

Galaxy / 4 / Metabolomics Analyze Data Workflow Shared Data Visualization Help User

Tools

Upload File from your computer

LC-MS

Preprocessing

xcms.xcmsSet: Peak Identification and Peak Identification using xcmsSet function from xcms R package to preprocess LC/MS data for relative quantification and statistical analysis

xcms.xcmsSetMerge: Merge xcms.xcmsSet xset in one to be used by group

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

xcms.retain: Retention Time Correction using rtcor function from xcms R package

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

xcms.summary: Create a summary of XCMS analysis

1 job has been successfully added to the queue - resulting in the following datasets:

- 7: LC_Positive_mzXML.xset.group.RData
- 8: LC_Positive_mzXML.xset.group.Rplots.pdf
- 9: LC_Positive_mzXML.xset.group.variableMetadata.tsv
- 10: LC_Positive_mzXML.xset.group.dataMatrix.tsv
- 11: xset.log.txt

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from "running" to "finished" if completed successfully or "error" if problems were encountered.

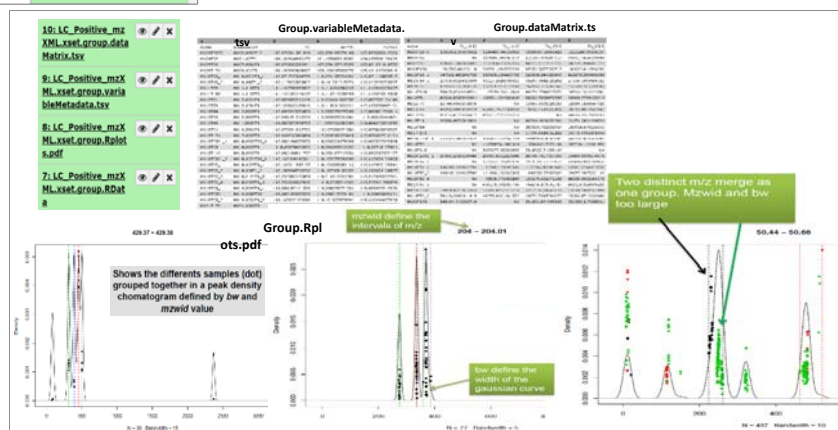
History

search datasets

Unnamed history

3.62 GB

- 10: LC_Positive_mzXML.xset.group.dataMatrix.tsv
- 9: LC_Positive_mzXML.xset.group.variableMetadata.tsv
- 8: LC_Positive_mzXML.xset.group.Rplots.pdf
- 7: LC_Positive_mzXML.xset.group.RData
- 6: LC_Positive_mzXML.xset.log.txt
- 5: LC_Positive_mzXML.xset.RPCL.raw.pdf
- 4: LC_Positive_mzXML.xset.TIC.raw.pdf
- 3: LC_Positive_mzXML.xset.Metadata.tsv



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

[missing]

Advanced options

hide

Resubmit your raw dataset or your zip file

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

Galaxy / 4 / Metabolomics Analyze Data Workflow Shared Data Visualization Help User Using 15.8 GB

Tools

search tools

Upload File from your computer

LC-MS

Preprocessing

xcms.xcmsSet Filtration and Peak Identification using xcmsSet function from xcms R package to preprocess LC/MS data for relative quantification and statistical analysis

xcms.xcmsSet Merger Merge xcms.xcmsSet xset in one to be used by group

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

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

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

xcms.summary Create a summary of XCMS analysis

CAMERA.annotate CAMERA

History

search datasets

Unnamed history

14 shown, 1 hidden

3.71 GB

15: LC_Positive_mzXML.xset.group.retcor.RData

13: LC_Positive_mzXML.xset.group.retcor.Rplots.pdf

14: LC_Positive_mzXML.xset.group.retcor.TICs_corrected.pdf

15: LC_Positive_mzXML.xset.group.retcor.BPCs_corrected.pdf

16: xset.log.txt

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

15: LC_Positive_mzXML.xset.group.retcor.BPCs_corrected.pdf

14: LC_Positive_mzXML.xset.group.retcor.TICs_corrected.pdf

13: LC_Positive_mzXML.xset.group.retcor.Rplots.pdf

12: LC_Positive_mzXML.xset.group.retcor.RData

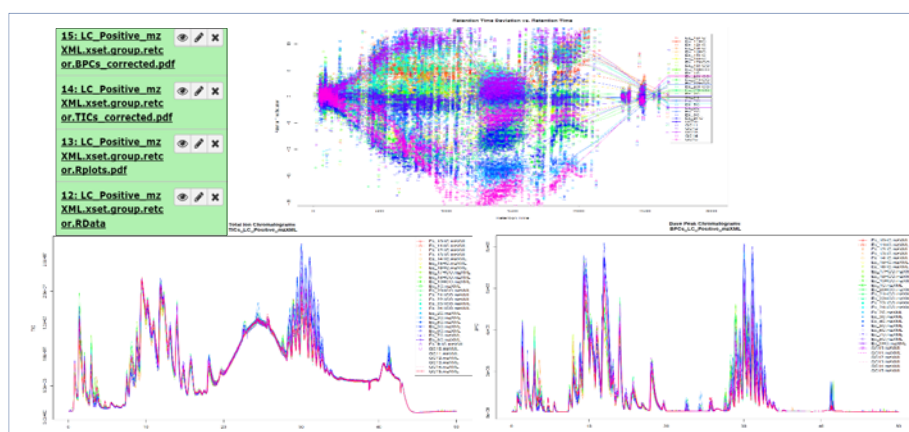
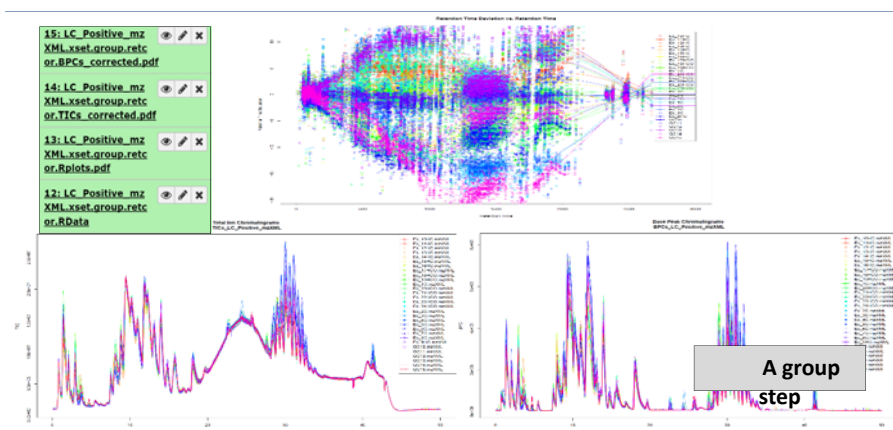
10: LC_Positive_mzXML.xset.group.data.Matix.txt

9: LC_Positive_mzXML.xset.group.variateMetadata.txt

8: LC_Positive_mzXML.xset.group.Rplot.pdf

7: LC_Positive_mzXML.xset.group.RData

PRACTICAL SESSION VISUALS_17



PRACTICAL SESSION. VISUALS_18

Galaxy / 4 / Metabolomics

Tools

1 job has been successfully added to the queue - resulting in the following datasets:

- 17: LC_Positive_mzXML.xset.group.retcor.group.RData
- 18: LC_Positive_mzXML.xset.group.retcor.group.Rplots.pdf
- 19: LC_Positive_mzXML.xset.group.retcor.group.variableMetadata.tsv
- 20: LC_Positive_mzXML.xset.group.retcor.group.dataMatrix.tsv
- 21: xset.log.txt

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History

Unnamed history

18 shown, 2 hidden

3.73 GB

20: LC_Positive_mzXML.xset.group.retcor.group.dataMatrix.tsv

19: LC_Positive_mzXML.xset.group.retcor.group.variableMetadata.tsv

18: LC_Positive_mzXML.xset.group.retcor.group.Rplots.pdf

17: LC_Positive_mzXML.xset.group.retcor.group.RData

16: LC_Positive_mzXML.xset.group.retcor.group.RData

15: LC_Positive_mzXML.xset.group.retcor.group.RData

14: LC_Positive_mzXML.xset.group.retcor.group.RData

13: LC_Positive_mzXML.xset.group.retcor.group.RData

12: LC_Positive_mzXML.xset.group.retcor.group.RData

A group step

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

xcms.fillPeaks Integrate a sample's signal in regions where peak groups are not represented to create new peaks in missing areas (Galaxy Version 2.1.0)

xset RData file

No rdata.xcms.group or rdata dataset available.

output file from another xcms function (group)

Filling method

chrom

[method] See the help section below

Get a Peak List

Resubmit your raw dataset or your zip file

variableMeta	dataM
--------------	-------

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

