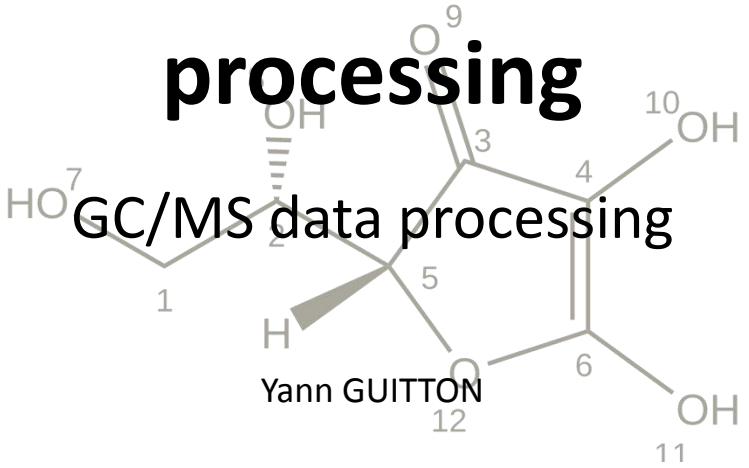
MS data processing

GC/MS data processing

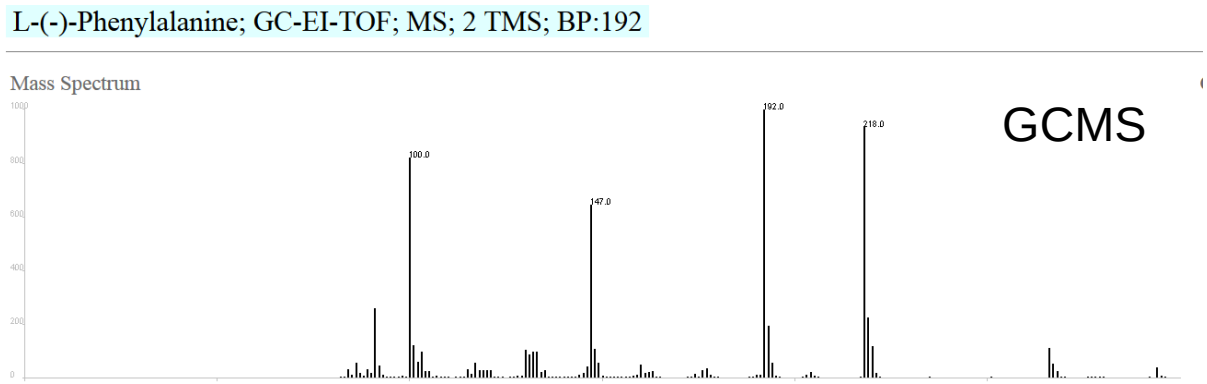


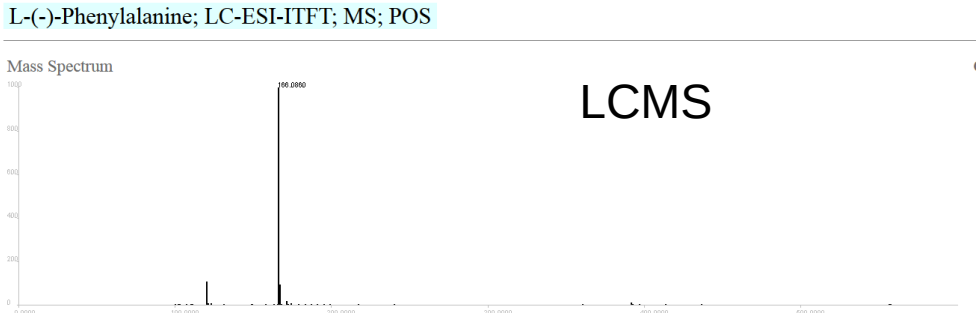
Yann GUITTON

v 1.0.0

Processing GC-MS data in Galaxy

- Why using a specific tool for processing GC-MS data ?
  - GC-MS data have their own specificity
    - First one is higer amount of ions in a mass spectra with EI compared to ESI





Processing GC-MS data in Galaxy

Galaxy / 4 / Metabolomics

Analyze Data | Workflow | Shared Data | Visualization

Tools

search tools

Upload File from your computer

Export Data

LC-MS

Format Conversion

Preprocessing

Normalisation

Quality Control

Statistical Analysis

Annotation

GC-MS

Preprocessing

metaMS.runGC GC-MS data preprocessing using metaMS package

Normalisation

Quality Control

Statistical Analysis

Annotation

dbsearch.golm Search GCMS spectra from msp files into Golm Metabolome database

NMR

Preprocessing

Normalisation

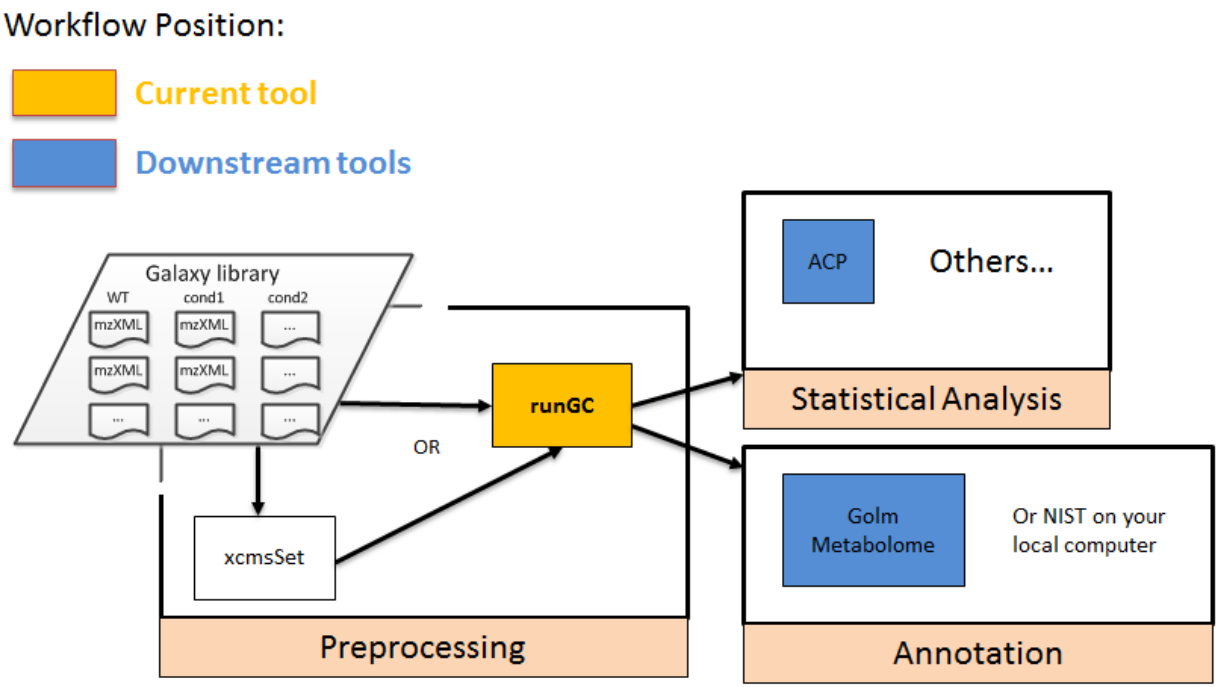
Quality Control

Statistical Analysis

Downstream tools

Name	Output file	Format
Determine Vdk or Lowess	dataMatrix.tsv	Tabular
Normalization Vdk/Lowess	dataMatrix.tsv	Tabular
Anova	dataMatrix.tsv	Tabular
PCA	dataMatrix.tsv	Tabular
Hierarchical Clustering	dataMatrix.tsv	Tabular
Golm Metabolome Search	peakspectra.msp	Text

General schema of the metabolomic workflow for GCMS



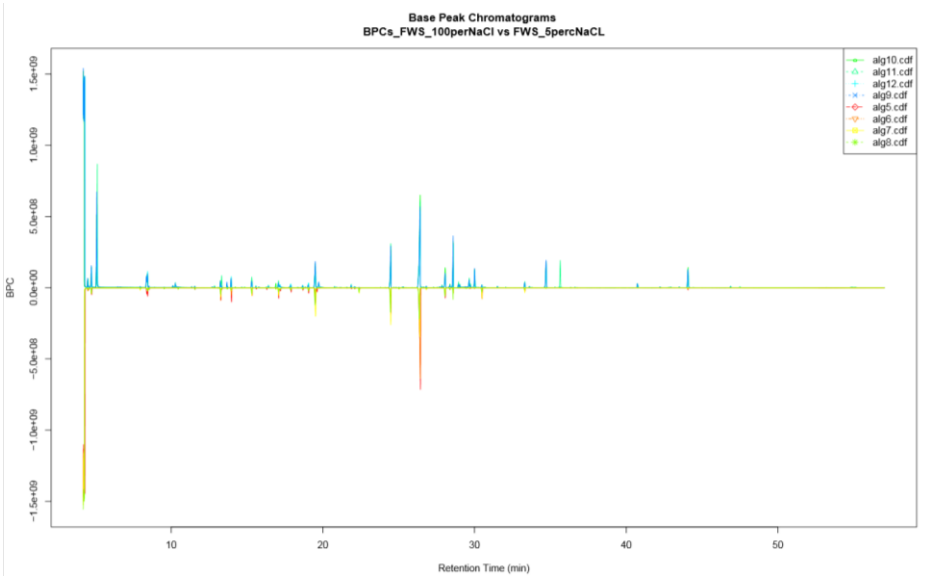
Processing GC-MS data in Galaxy

- *In fine* that tool use xcms/CAMERA under metaMS R package from R. Wehrens

metaMS : Wehrens, R.; Weingart, G.; Mattivi, F. Journal of Chromatography B.

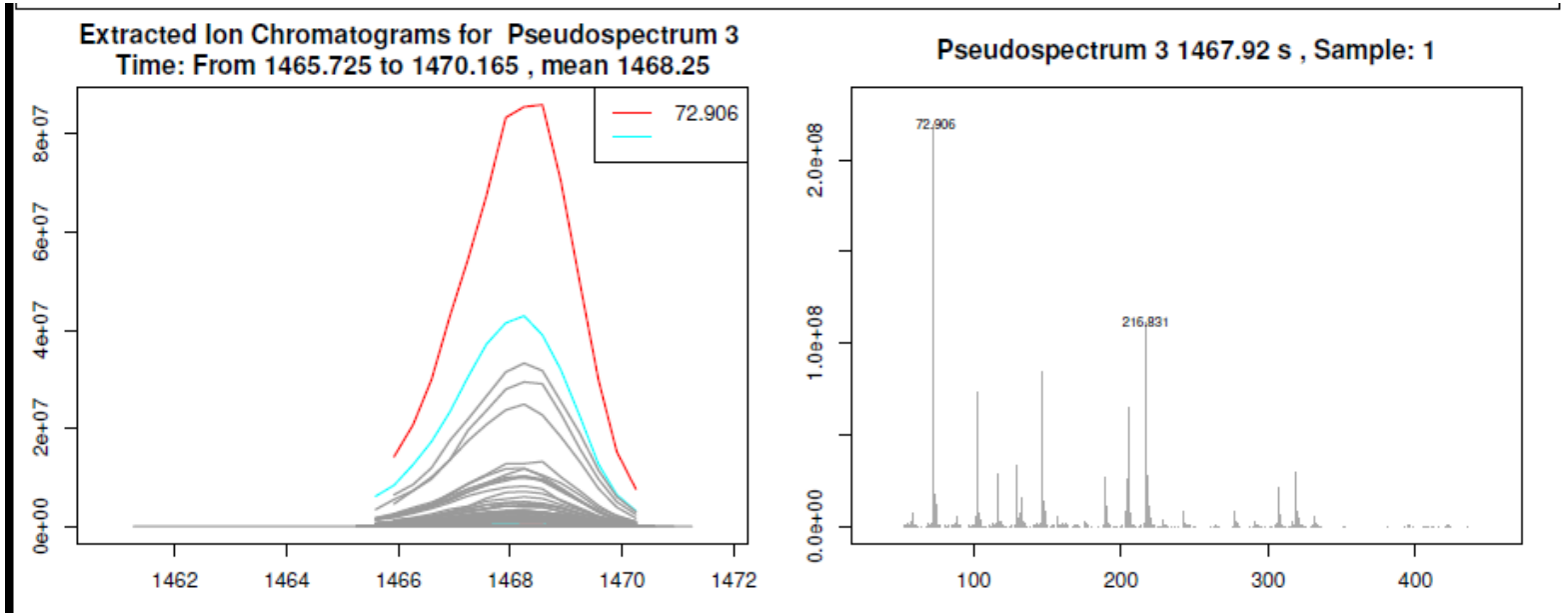
- with adapted parameters and dedicated outputs

1: GCMS_Idealq_FWS_SWS.zip
2: peaktable.tsv
3: sampleMetadata.tsv
4: variableMetadata.tsv
5: dataMatrix.tsv
6: peakspectra.msp
7: TICs_raw.pdf
8: BPCs_raw.pdf
9: GCMS_EIC.pdf
10: rungc.RData
11: rungc.log.txt
12: rungc.zip
20: GolmResult.html



Processing GC-MS data in Galaxy

9: GCMS_EIC.pdf

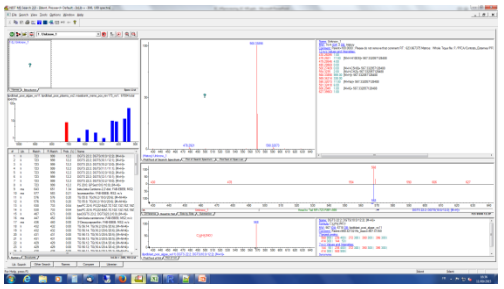


Peakpicking/grouping quality control

Processing GC-MS data in Galaxy

6: peakspectra.msp

A MSP export of pseudospectras readable by NIST or Golm Metabolome



NIST MSsearch

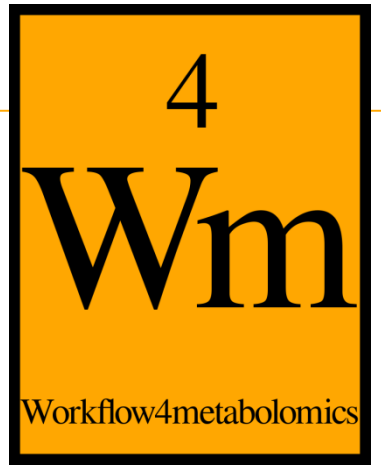
20: GolmResult.html

cname	analyteName	ri	DotproductDistance	EuclideanDistance	HammingDistance	JaccardDistance	s12GowerLegendr
Unknown 1	Citric acid (4TMS)	1803.92065	0.01	0.01	15	0.28	0.4
	Isocitric acid (4TMS)	1805.39966	0.04	0.01	109	0.74	0.71
	Gluconic acid (6TMS)	1984.52283	0.11	0.02	85	0.82	0.82
	Malic acid, 3-oxalo- (1MEOX) (4TMS) BP	2200	0.12	0.02	868	0.96	0.89
	Galactaric acid (6TMS)	2030.46814	0.12	0.02	461	0.92	0.85
	Arabinoheptulosonic acid enol, 3-deoxy- (5TMS) MP	1926.63843	0.11	0.02	499	0.92	0.85
	Malic acid, 3-oxalo- (1MEOX) (4TMS) MP	2190.4812	0.12	0.02	478	0.93	0.87
	myo-Inositol-1-phosphate (7TMS)	2414.291	0.12	0.02	403	0.92	0.85

http://web11.sb-roscoff.fr/download/w4m/howto/w4m_HowToUseNIST_V01.pdf

Processing GC-MS data in Galaxy

- Then go to other databases with the .MSP file
- Or go to statistical analysis with



Processing GC-MS data in Galaxy

HOW TO

How to : Processing GC-MS data in Galaxy

Galaxy / 4 / Metabolomics

Analyze Data | Workflow | Shared Data | Visualization

Tools

search tools

Upload File from your computer

Export Data

LC-MS

Format Conversion

Preprocessing

Normalisation

Quality Control

Statistical Analysis

Annotation

GC-MS

Preprocessing

metaMS.runGC GC-MS data preprocessing using metaMS package

Normalisation

Quality Control

Statistical Analysis

Annotation

dbsearch.golm Search GCMS spectra from msp files into Golm Metabolome database

NMR

Preprocessing

Normalisation

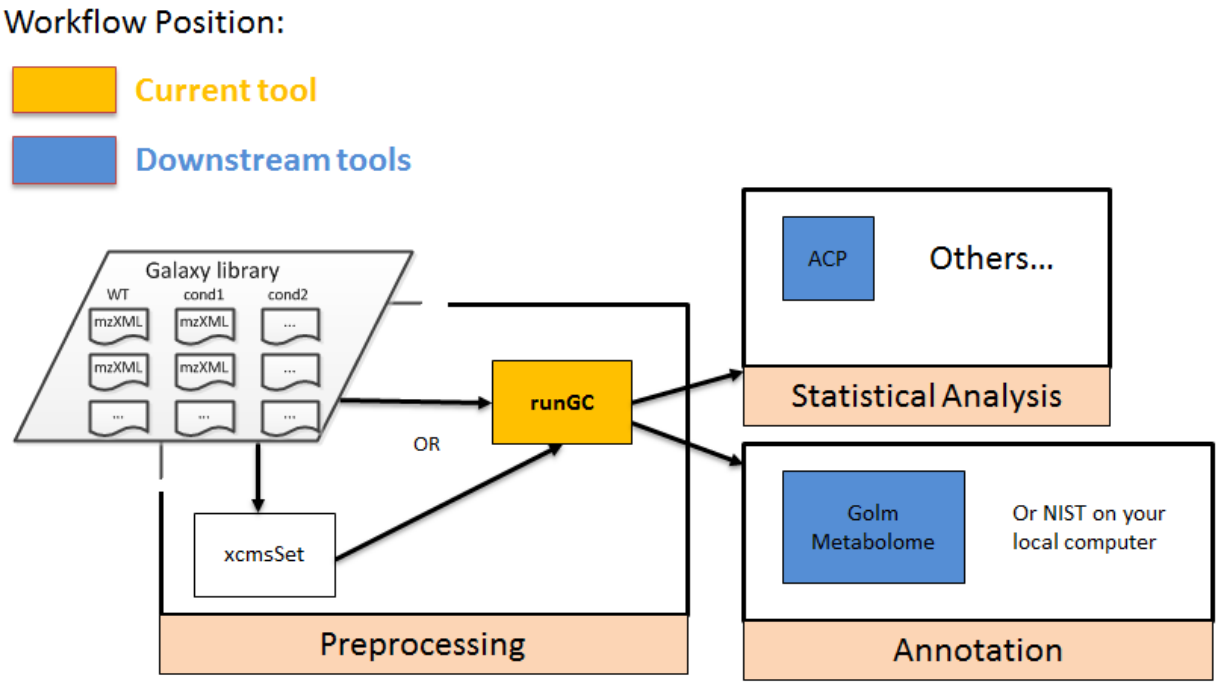
Quality Control

Statistical Analysis

Downstream tools

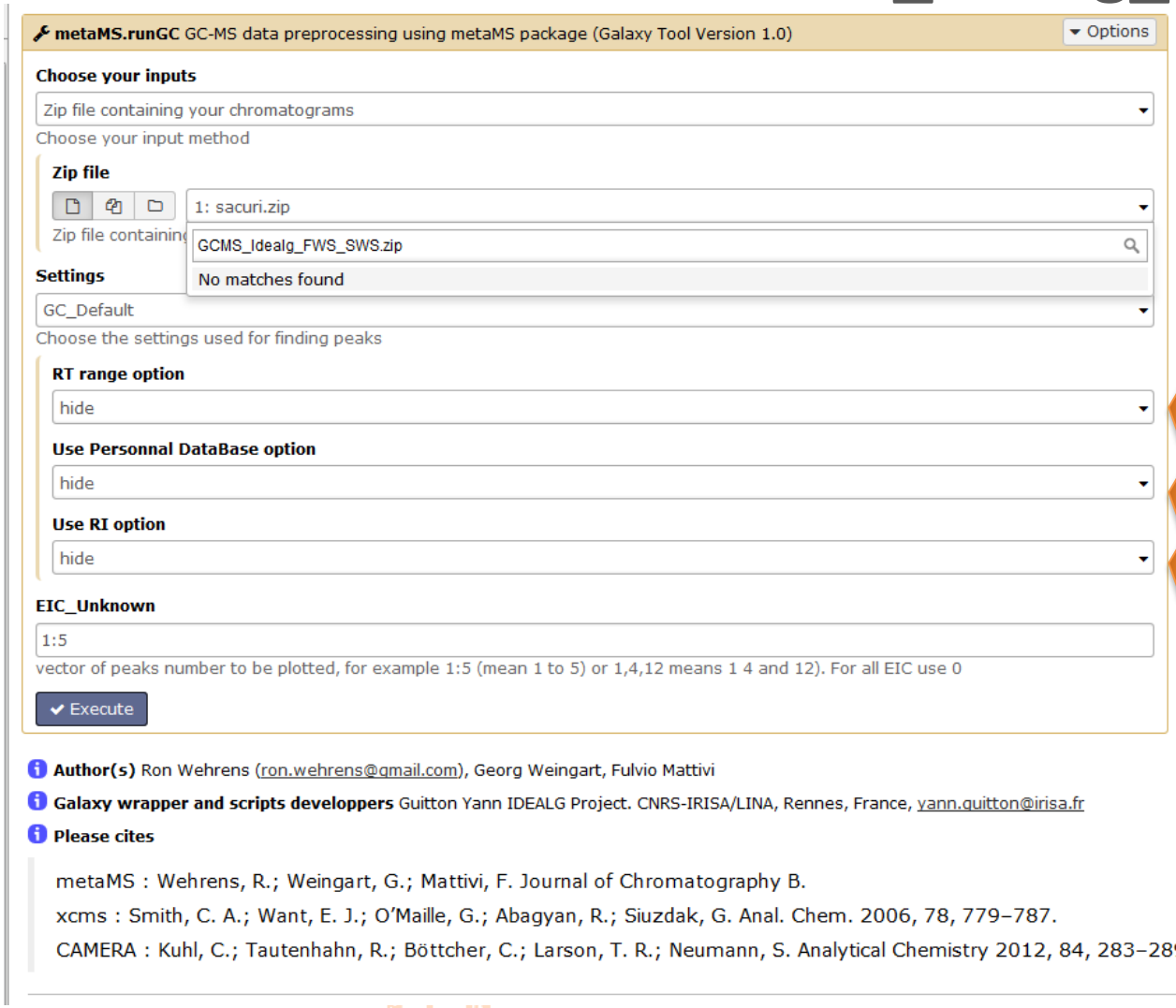
Name	Output file	Format
Determine Vdk or Lowess	dataMatrix.tsv	Tabular
Normalization Vdk/Lowess	dataMatrix.tsv	Tabular
Anova	dataMatrix.tsv	Tabular
PCA	dataMatrix.tsv	Tabular
Hierarchical Clustering	dataMatrix.tsv	Tabular
Golm Metabolome Search	peakspectra.msp	Text

General schema of the metabolomic workflow for GCMS



How to : Processing GC-MS data in Galaxy

- Can be tested with GCMS_Idealg_FWS_SWS.zip



Input your data as zip, or as library (data should be in CDF, mzXML or mzData) and organised in sub folders

GC default settings, give it a try

RT range: cut a part of your chromatogramme

Can import a DB in MSP format

Calculate Ris need a file with alcane RT

Only the five first EIC by default use 0 to draw all

How to : Processing GC-MS data in Galaxy

Settings

User_Defined

Choose the settings used for finding peaks

FWHM

5

The FWHM of your peaks (matchedFilter method is used)

RT range option

hide

RT_Diff

0.05

The allowed RT shift between same molecule in different sample

Min_Features

5

The minimum number of ion in a mass spectra to consider it a molecule

similarity_threshold

0.7

The minimum similarity allowed between peaks mass spectra to be considered as equal

min.class.fract

0.5

The fraction of samples in which a pseudospectrum is present before it is regarded as an unknown

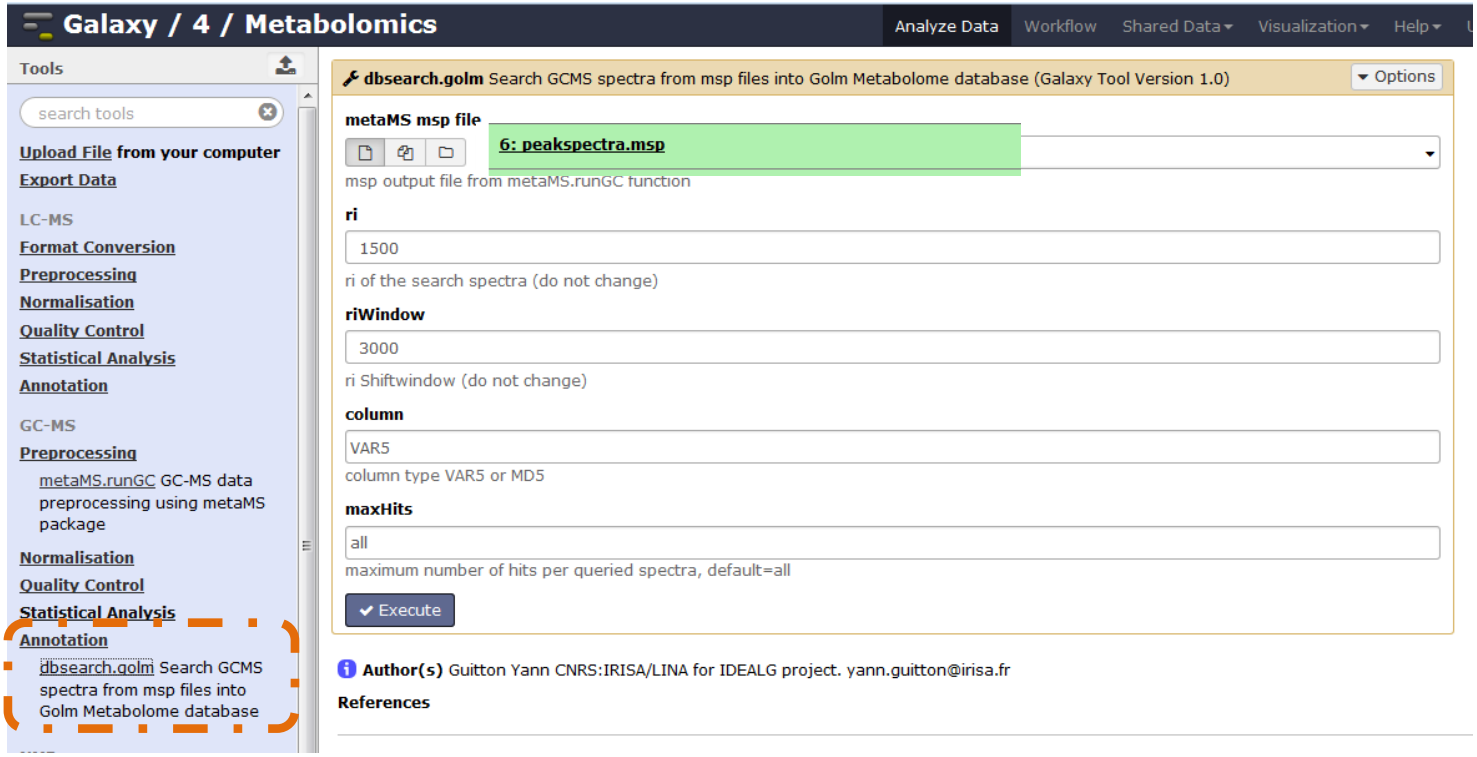
min.class.size

3

The absolute number of samples in which a pseudospectrum is present before it is regarded as an unknown

In fact all are xcms/CAMERA parameters

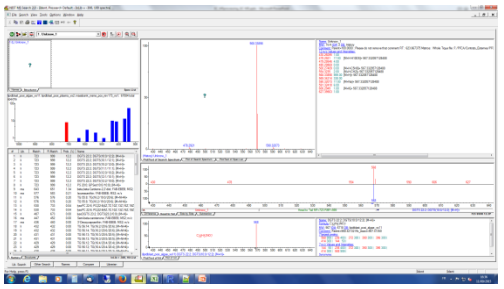
How to : Processing GC-MS data in Galaxy



How to: Processing GC-MS data in Galaxy

6: peakspectra.msp

A MSP export of pseudospectras readable by NIST or Golm Metabolome



NIST MSsearch

20: GolmResult.html

cname	analyteName	ri	DotproductDistance	EuclideanDistance	HammingDistance	JaccardDistance	s12GowerLegendr
Unknown 1	Citric acid (4TMS)	1803.92065	0.01	0.01	15	0.28	0.4
	Isocitric acid (4TMS)	1805.39966	0.04	0.01	109	0.74	0.71
	Gluconic acid (6TMS)	1984.52283	0.11	0.02	85	0.82	0.82
	Malic acid, 3-oxalo- (1MEOX) (4TMS) BP	2200	0.12	0.02	868	0.96	0.89
	Galactaric acid (6TMS)	2030.46814	0.12	0.02	461	0.92	0.85
	Arabinoheptulosonic acid enol, 3-deoxy- (5TMS) MP	1926.63843	0.11	0.02	499	0.92	0.85
	Malic acid, 3-oxalo- (1MEOX) (4TMS) MP	2190.4812	0.12	0.02	478	0.93	0.87
	myo-Inositol-1-phosphate (7TMS)	2414.291	0.12	0.02	403	0.92	0.85

http://web11.sb-roscoff.fr/download/w4m/howto/w4m_HowToUseNIST_V01.pdf

Thanks!