



Computational Metabolomics

Kumar Saurabh Singh

Postdoctoral researcher

Plant Microbe Interaction group, Utrecht University, The Netherlands

Bioinformatics group, Wageningen University & Research, The Netherlands



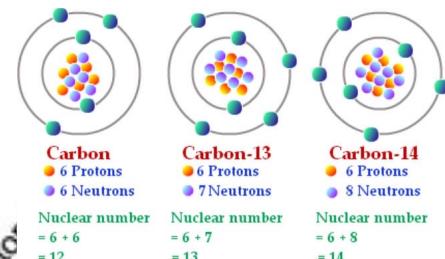
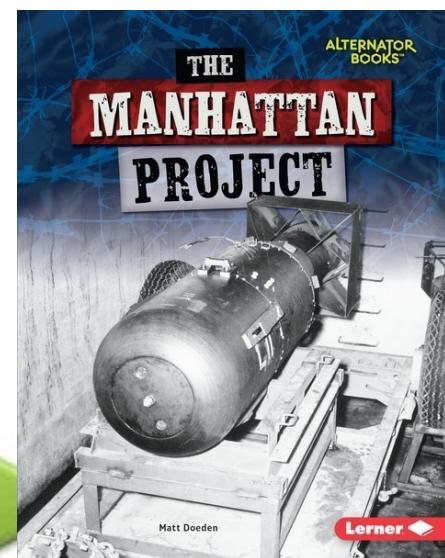
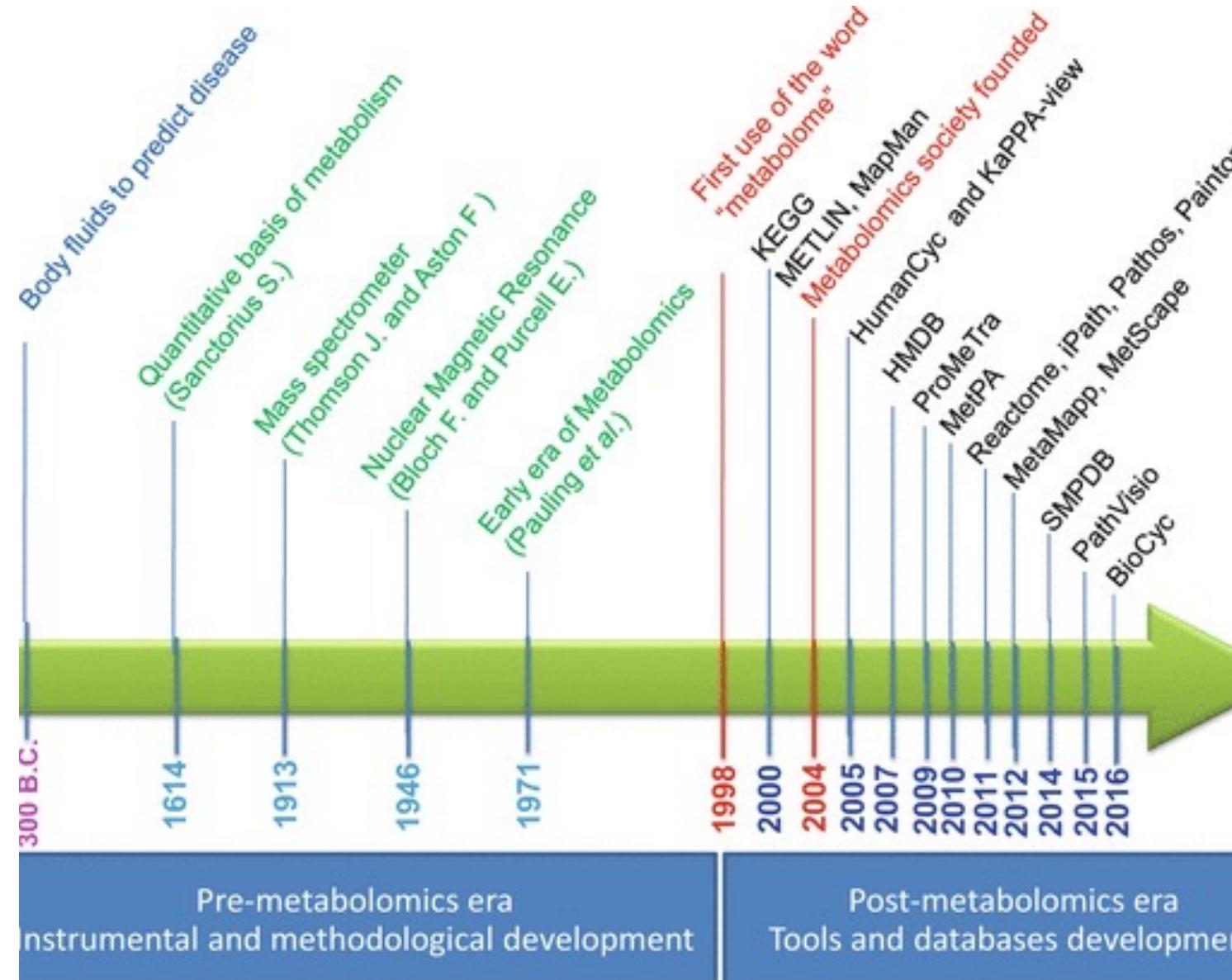
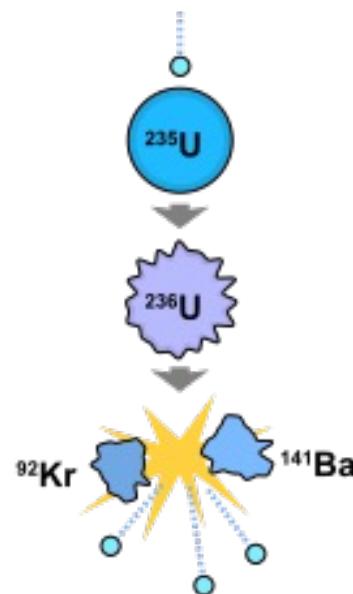
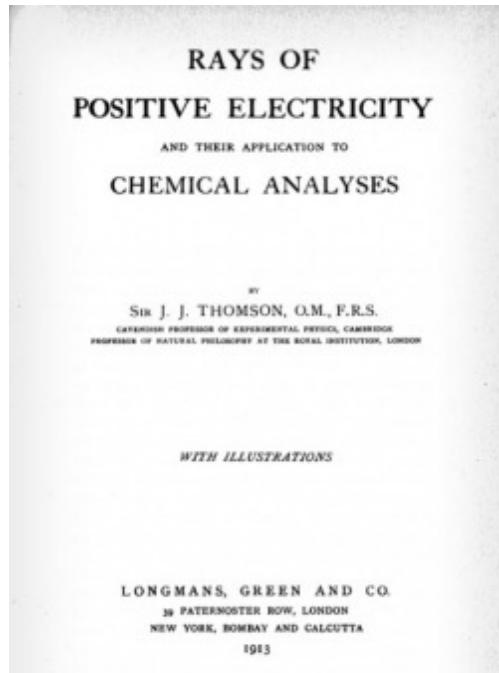
Utrecht University



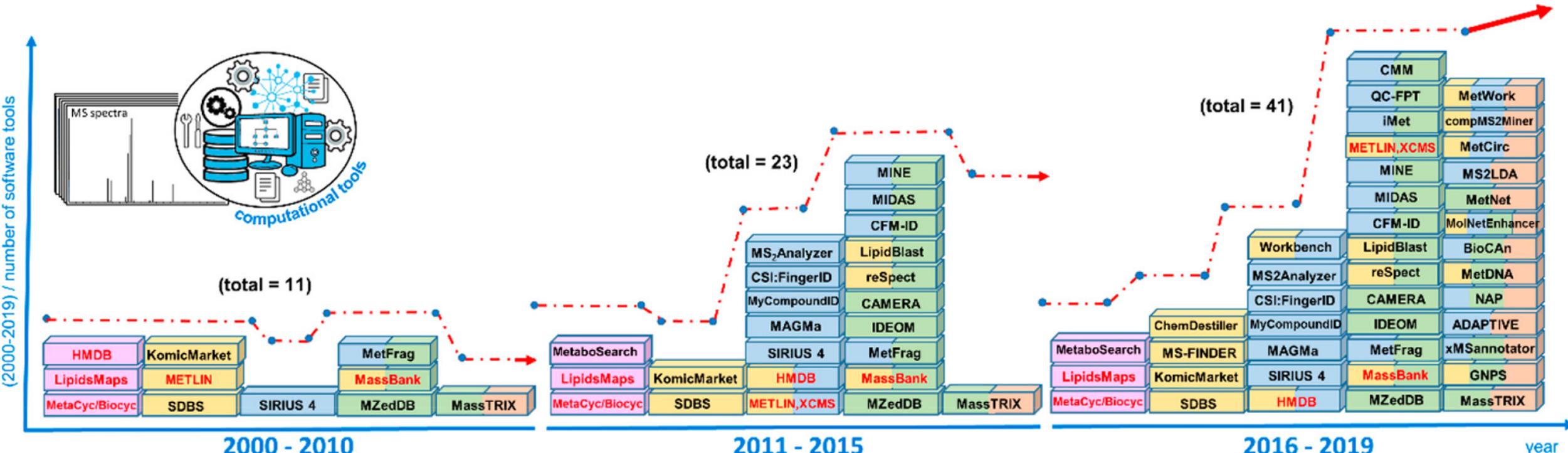
WAGENINGEN
UNIVERSITY & RESEARCH

- 
- (20-30 min) A short primer on Metabolomics, including a brief timeline, technology, data formats and analysis software.
 - (20-30 min) Introduction of MEANtools (Metabolomics and Transcriptomics integration).
 - (5-10) min bio break.
 - (40-50 min) Tutorial on data processing and analysis of LC-MS data using XCMS.

History

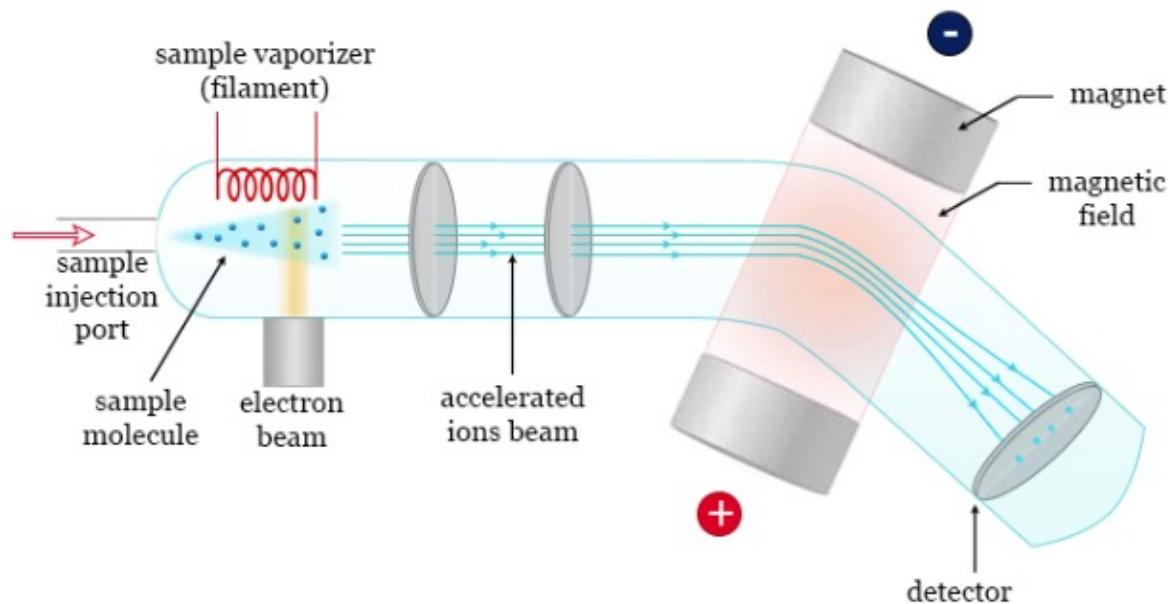


Developments

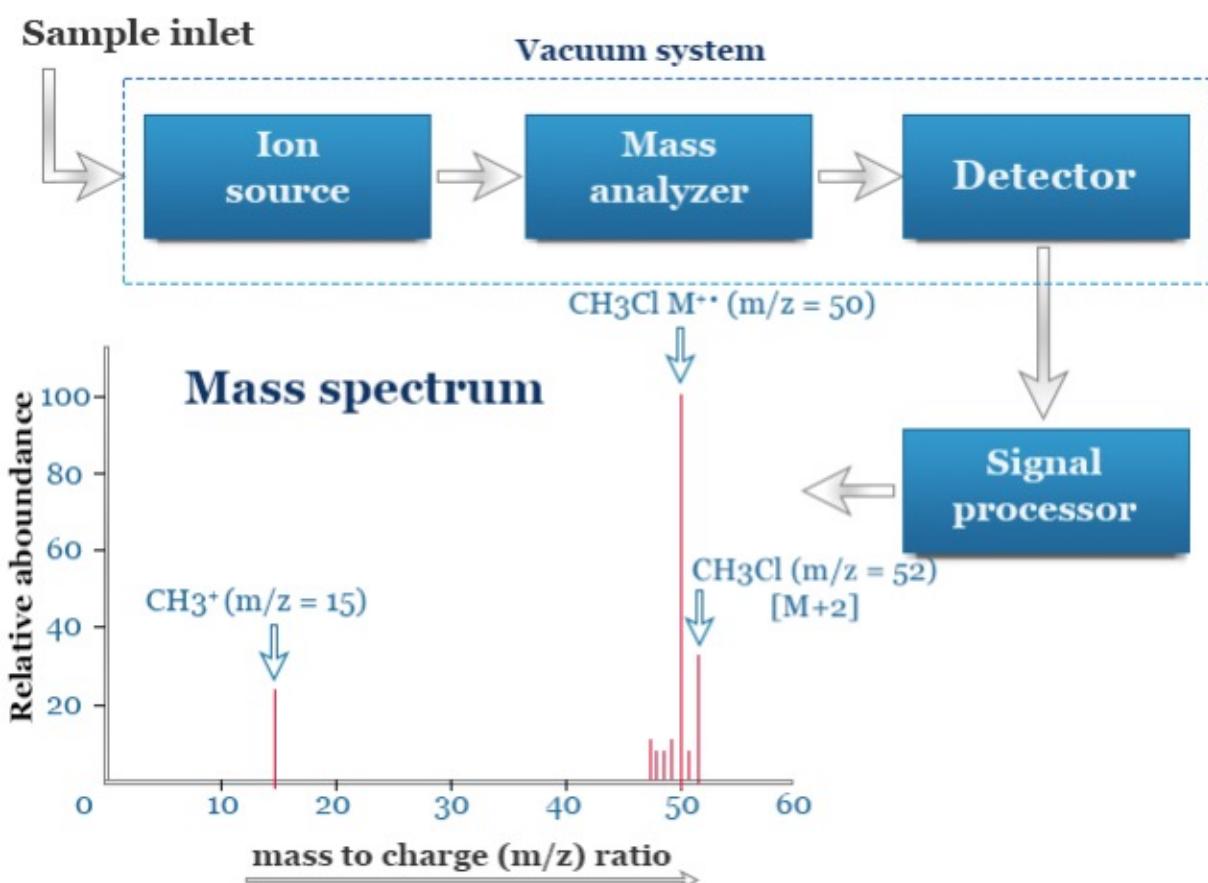


- 0** **Category 0:** Software tools allowing *m/z* matching of the precursor ion or the monoisotopic mass. They may contain more information (physiochemical, biological, ontological, etc.), but the search is performed only through the *m/z* of the precursor ion.
 - 1** **Category 1:** Software tools allowing search through the **MS/MS** or **MSⁿ** spectra against experimental spectra contained in one or more databases.
 - 2** **Category 2:** Software tools allowing search through the **MS/MS** or **MSⁿ** spectra against *in-silico* spectra previously predicted or predicted in run-time based on the putative structures obtained for the *m/z* precursor ion.
 - 3** **Category 3:** Software tools performing metabolite annotation/identification using **orthogonal information** to provide evidence/score about the putative annotations. It includes chromatographic (RT, CE), ontological or biological information, among others. They can use information from MS¹ and/or MSⁿ in addition to the orthogonal information.
 - 4** **Category 4:** Software tools performing metabolite annotation/identification creating **molecular networks** between the putative annotations obtained for the features. They use approaches to put the annotations into a biological context to provide evidence pointing to confirm or refute them. They can use information from MS¹, MSⁿ and/or orthogonal information in addition to the biological context created in the molecular network.

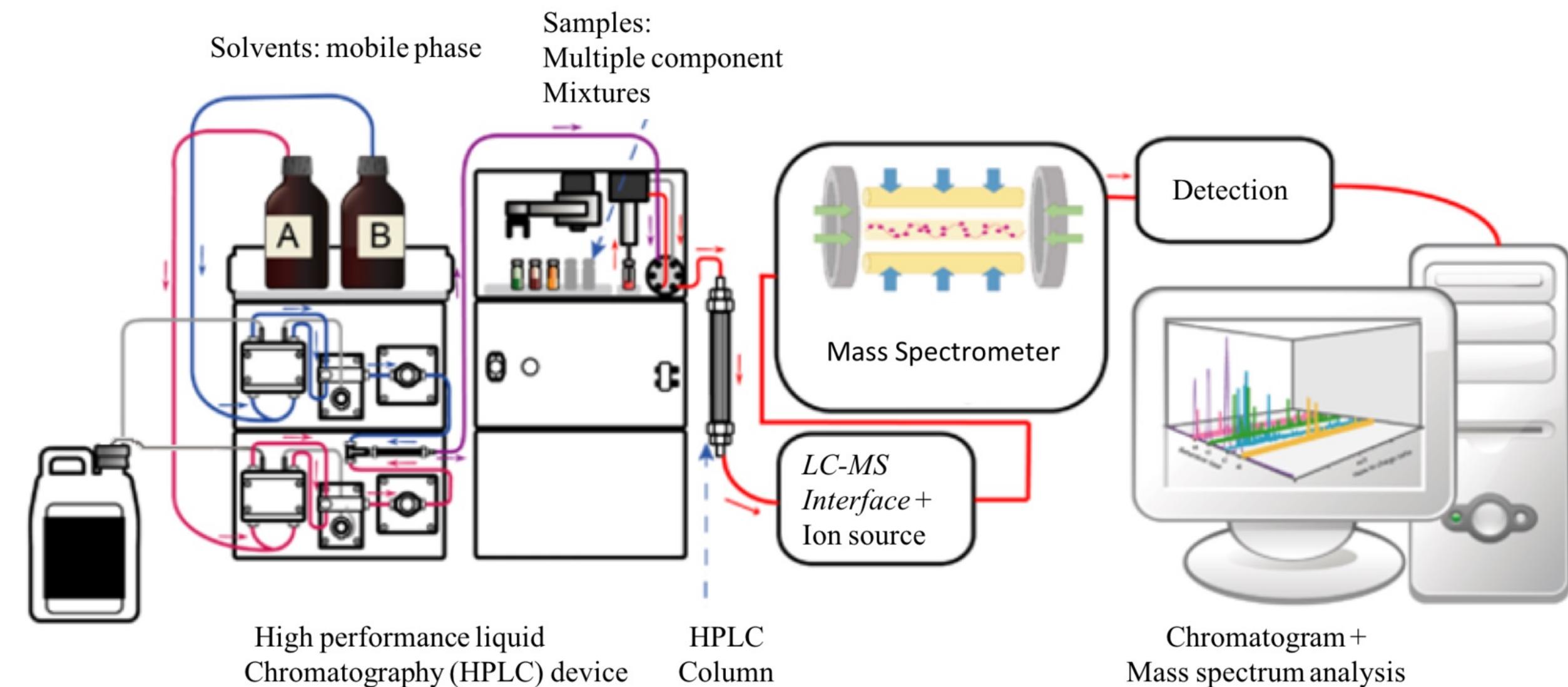
Mass spectrometry



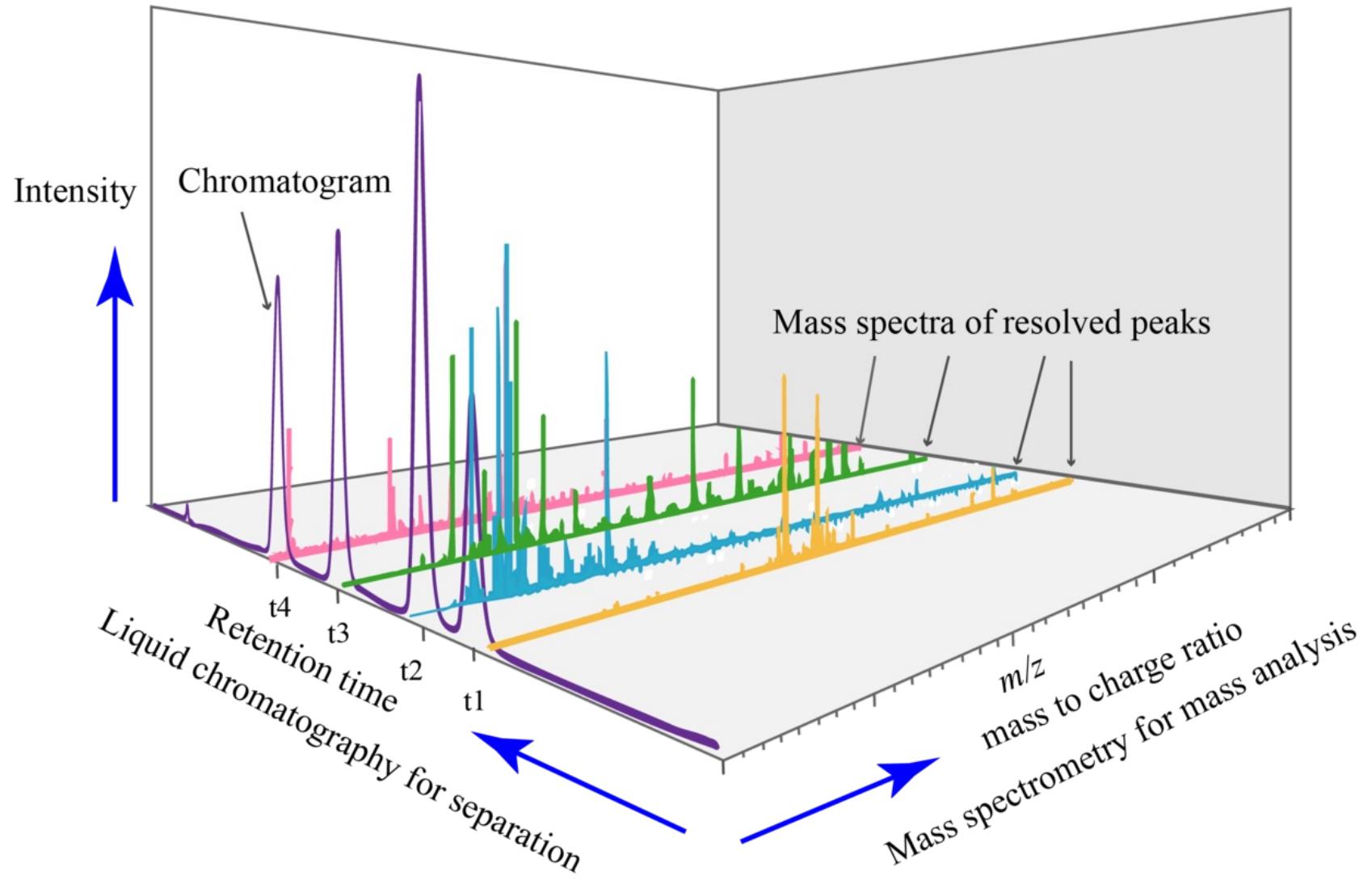
1. Sample Handling system
2. Ionization chamber
3. Ion separator
4. Ion Collector
5. Mass analyzer
6. Detector
7. Signal processor unit



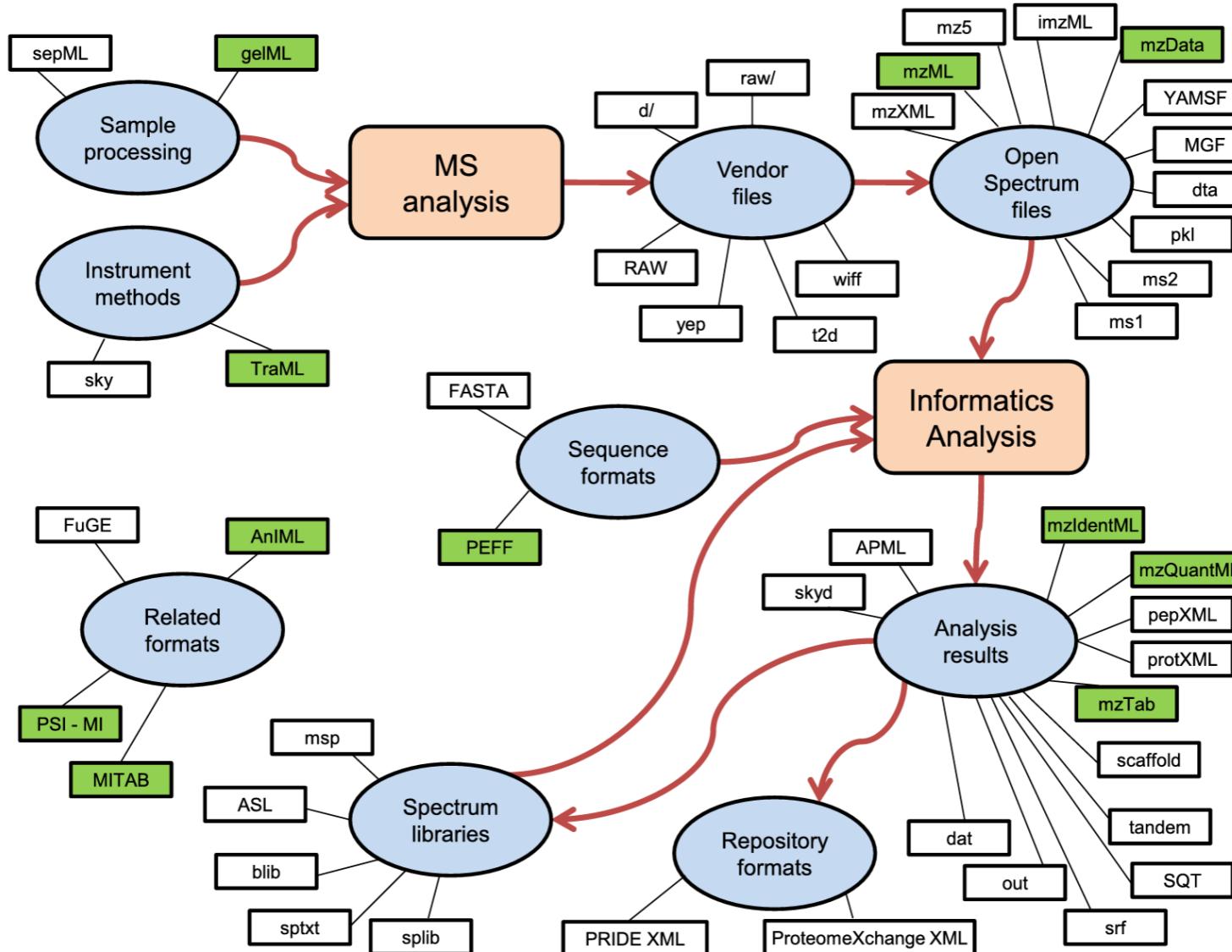
Combining chromatography with mass spectrometry



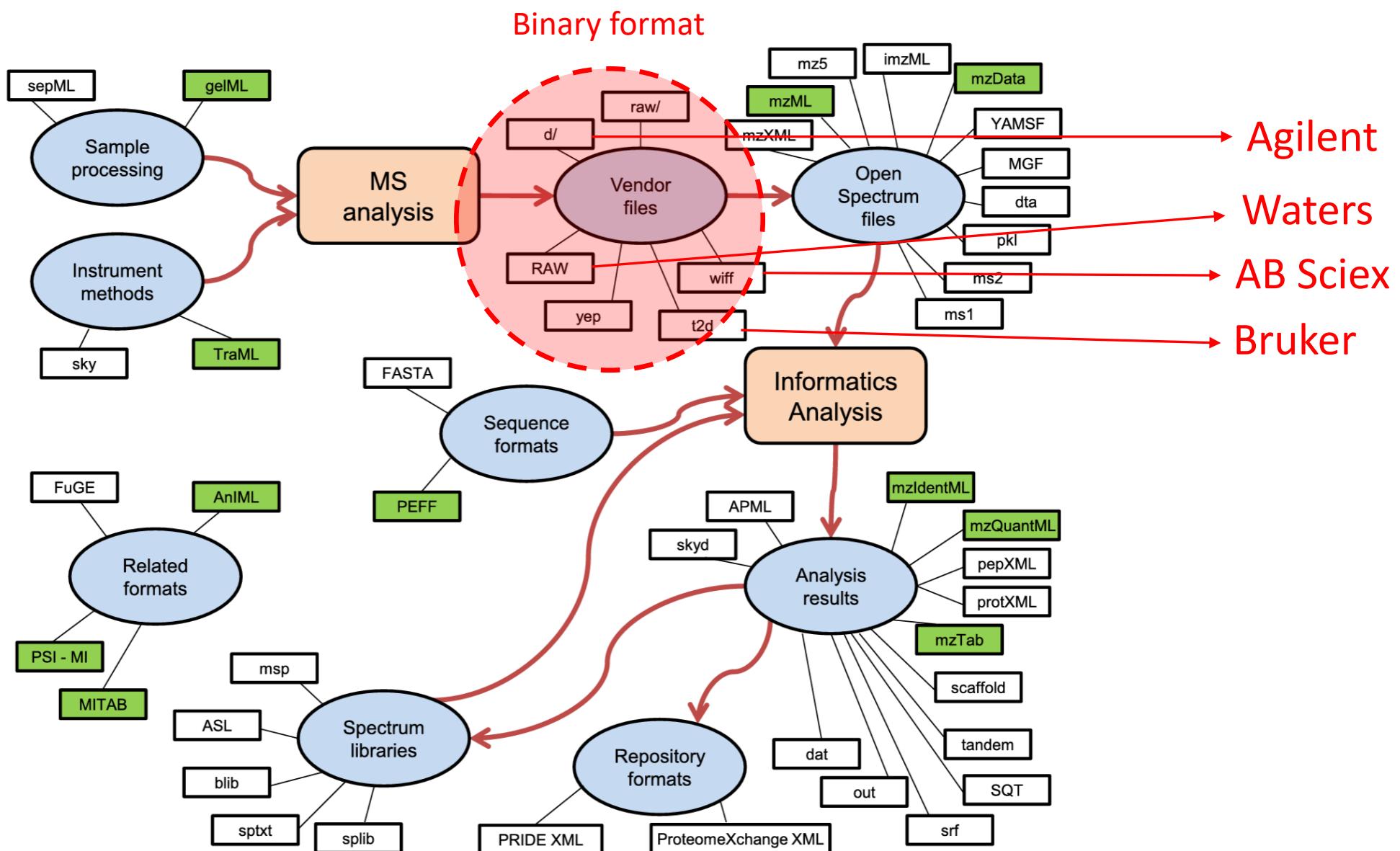
LC-MS Spectrum



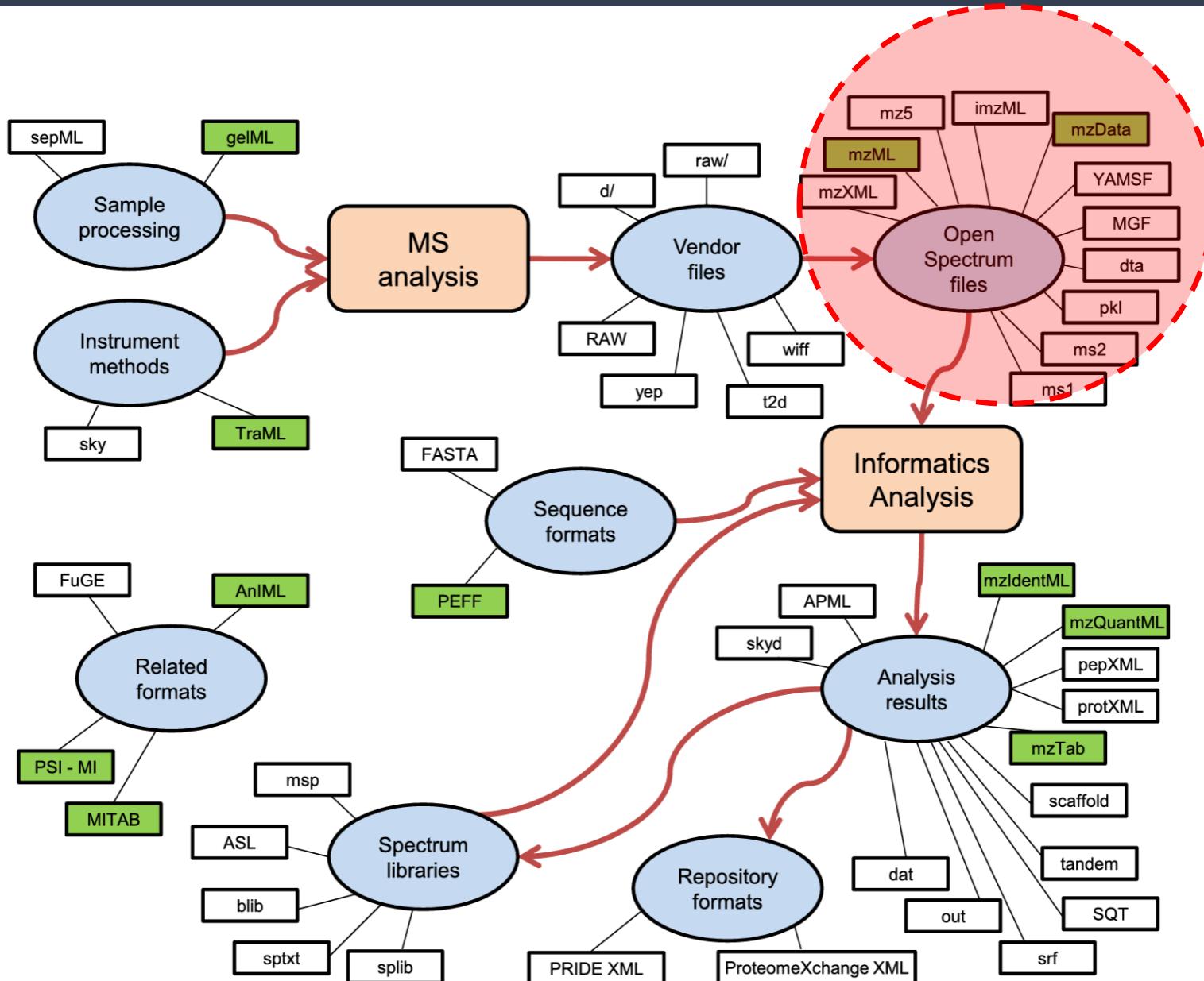
Data formats



Data formats



Data formats



Open Source software can convert vendor specific file format to other more accessible formats
like msConvert

Data formats

UMC Utrecht (WKZ)

Machine	Vendor	Manufacturer software	Default output format	Conversion software	Conversion output formats
UPLC	Waters	Masslynx	.RAW directory	Wolf	*.mzXML, *.cdf
QTOF	Waters	Masslynx	.RAW directory	Wolf	*.mzXML, *.cdf
GC-TOF	Thermo Scientific	XCalibur	.RAW file	XConvert, ReAdW	*.dat, *.ms, *.spa, *.raw, *.cdf, *.txt, *.mzXML
GC-MS	Thermo Scientific	XCalibur	.RAW file	XConvert, ReAdW	*.dat, *.ms, *.spa, *.raw, *.cdf, *.txt, *.mzXML
GC-MS	Agilent/HP	Masshunter	.d directory	Trapper(?), mzStar (now mzWiff, beta)	*.mzXML
LC-MS	Waters	Masslynx	.RAW directory	Wolf	*.mzXML, *.cdf

Wageningen (PRI)

Machine	Vendor	Manufacturer software	Default output format	Conversion software	Conversion output formats
GC-MS (TOF)	Leco	chromaTOF	peg (pegasus)		NetCDF
GC-MS (Quadrupool, SPME)	Interscience	XCalibur	.raw		NetCDF
GC-MS (Quadrupool)	HP (Agilent)	HPchemstation	.D		
LC-MS (TOF, Ultima)	Waters (micromass)	Masslynx V 4.0	Masslynx (.raw files)	MetAlign	netcdf

Data analysis using XCMS

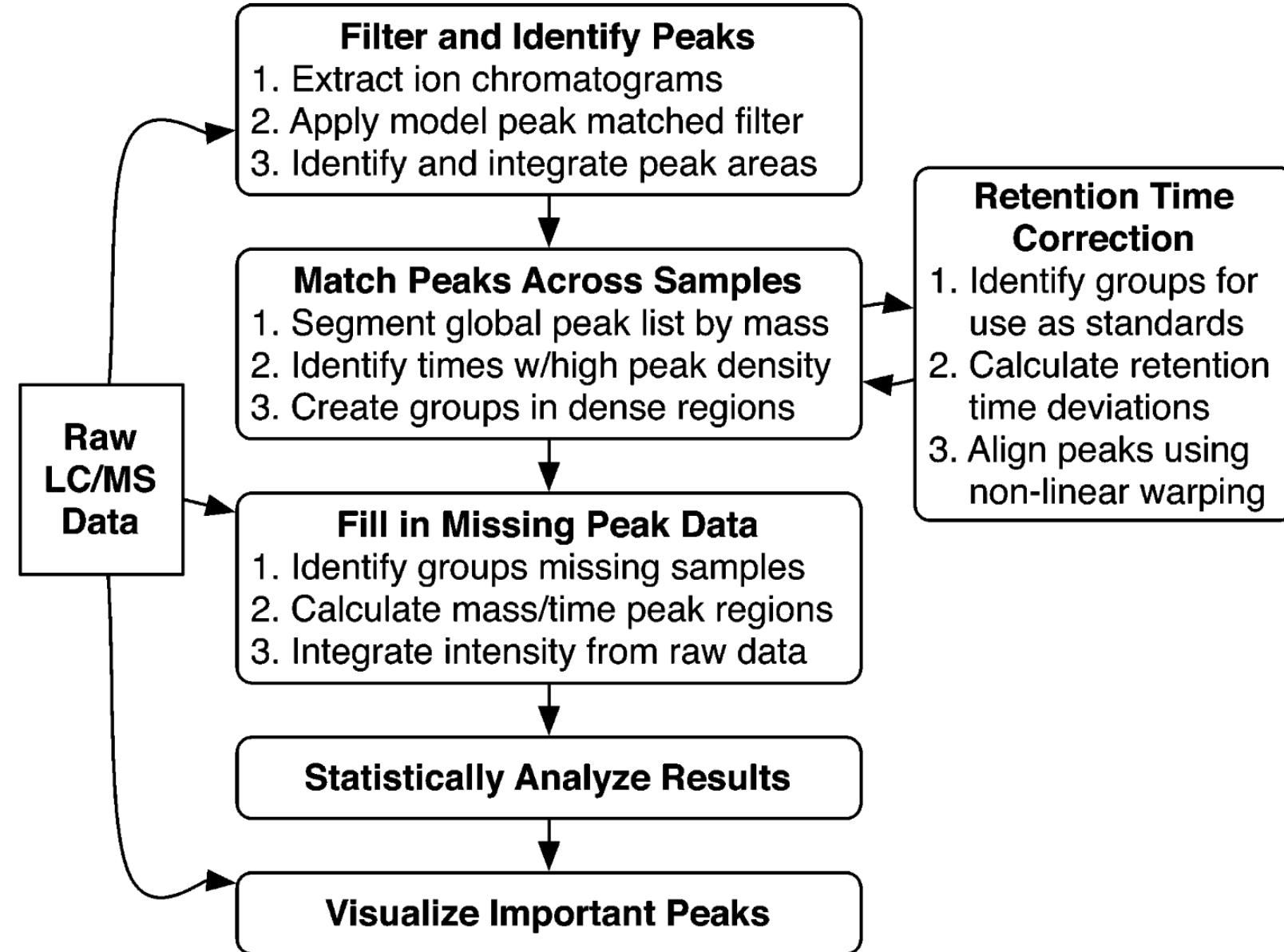
So every metabolite ...

- produces a peak in the ion "current" of all the ions produced during its ionization
- every metabolites produces peaks in more than one ionic trace
- in many cases the same ion could be produced in the ionization of different metabolites

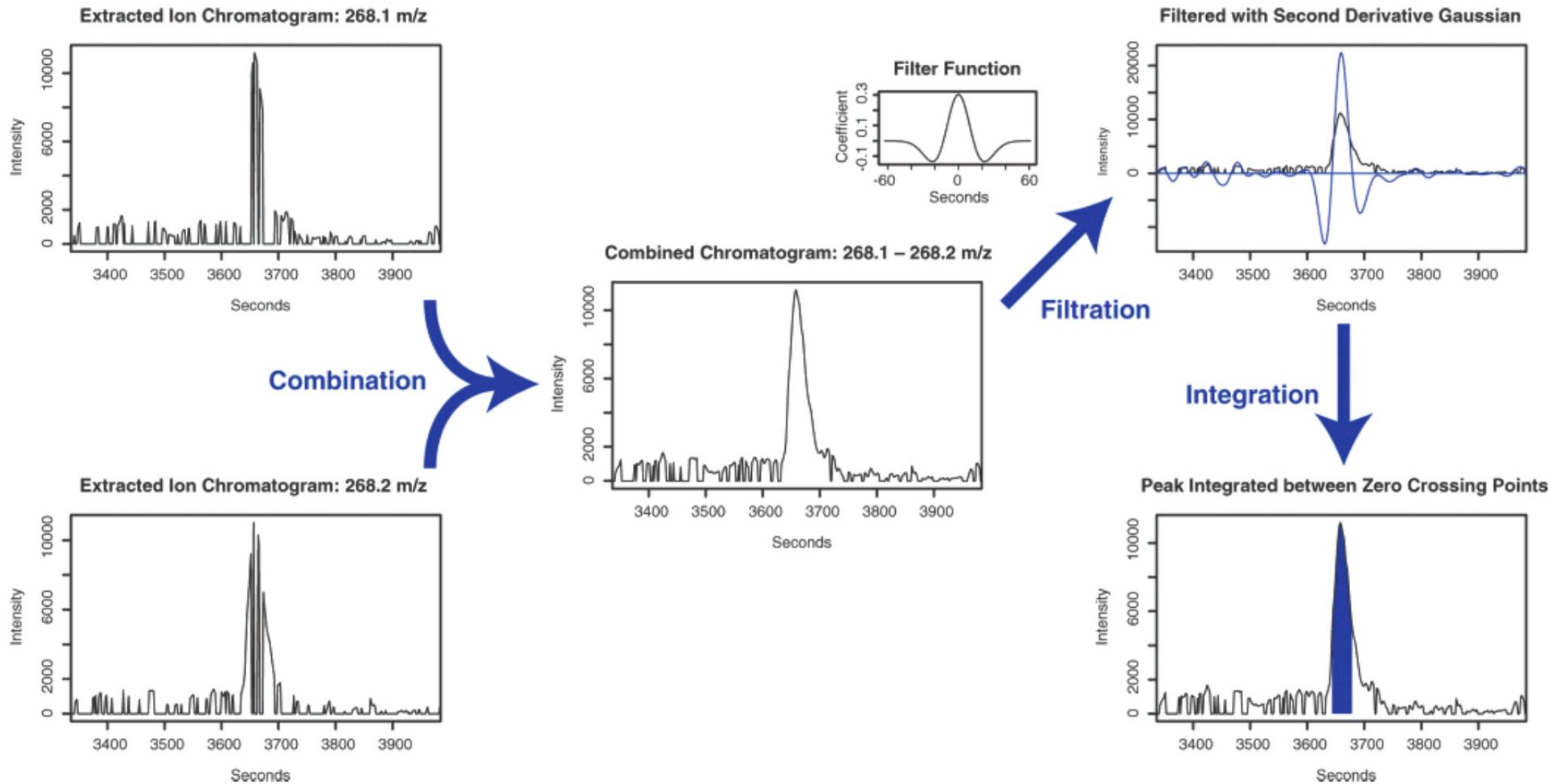
So ...

- to find metabolites we have to automatically look for peaks in the $m/z/rt$ plane
- we will have much more peaks than metabolites... and this will make the analysis of our data matrix extremely challenging

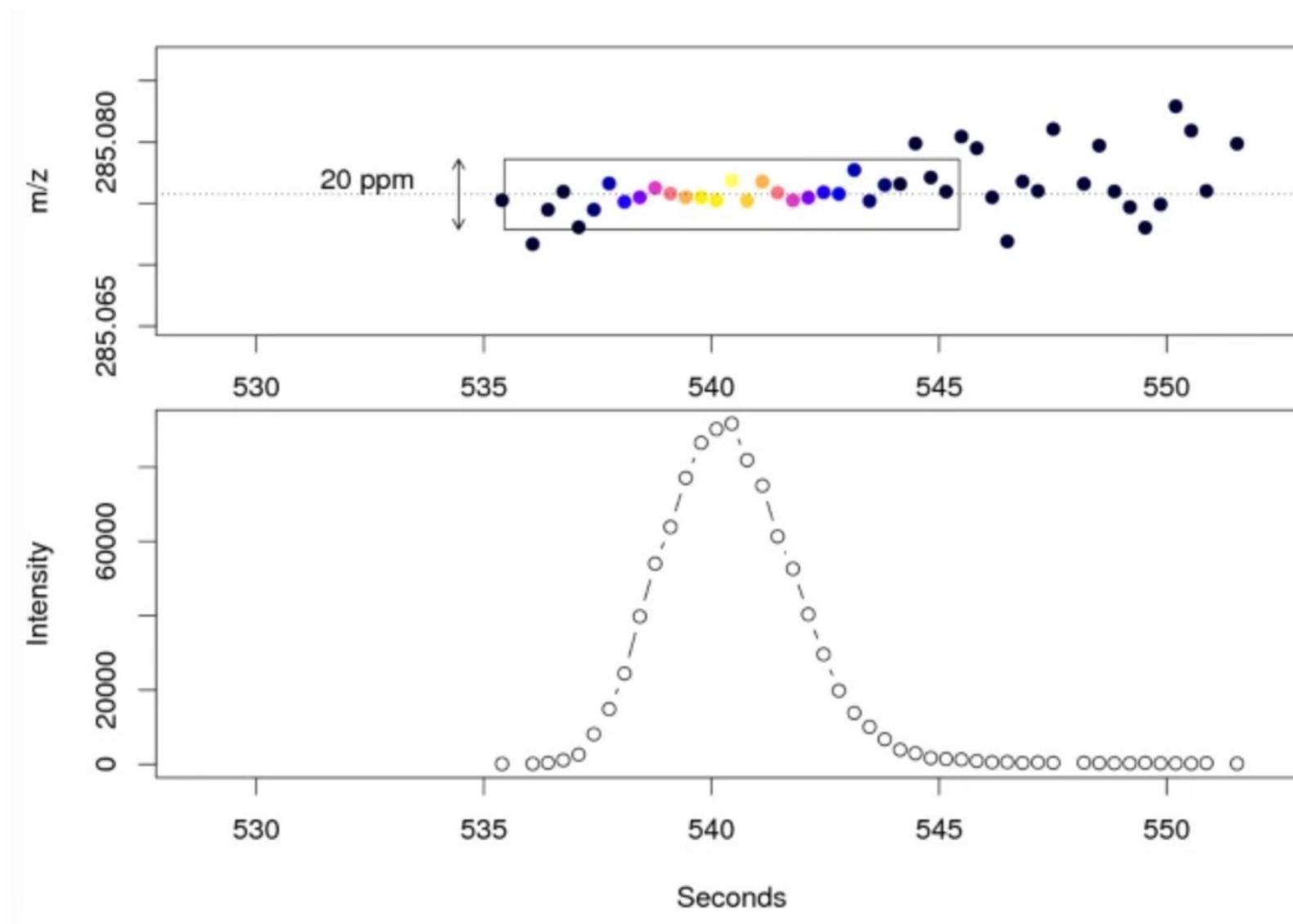
Data analysis using XCMS



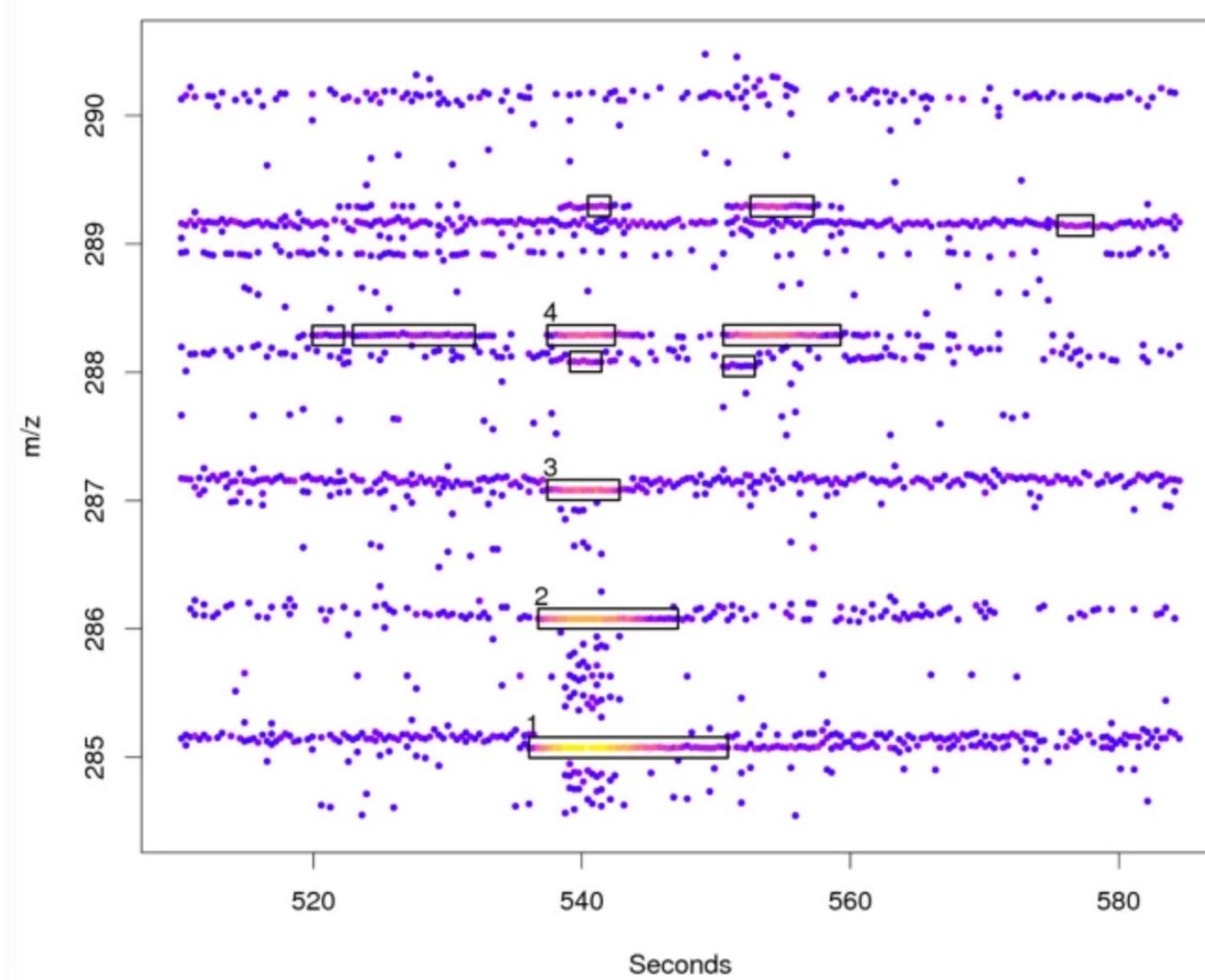
Data analysis using XCMS : Peak picking algorithms; Match filter



Data analysis using XCMS : Peak picking algorithms; CentWave



Data analysis using XCMS : Peak picking algorithms; CentWave



Multiple Algorithms

- Many different algorithms can be used to perform peak picking.
- No best solution. You can only devise best strategies.
- Each algorithm has its own parameters.
- To find the optimal values it is necessary to know the characteristics of the data.
- The “analytical” knowledge is the base of a reasonable educated guess.

Data analysis using XCMS : Peak alignment

Sample

Data analysis using XCMS : Peak alignment

Sample

⋮

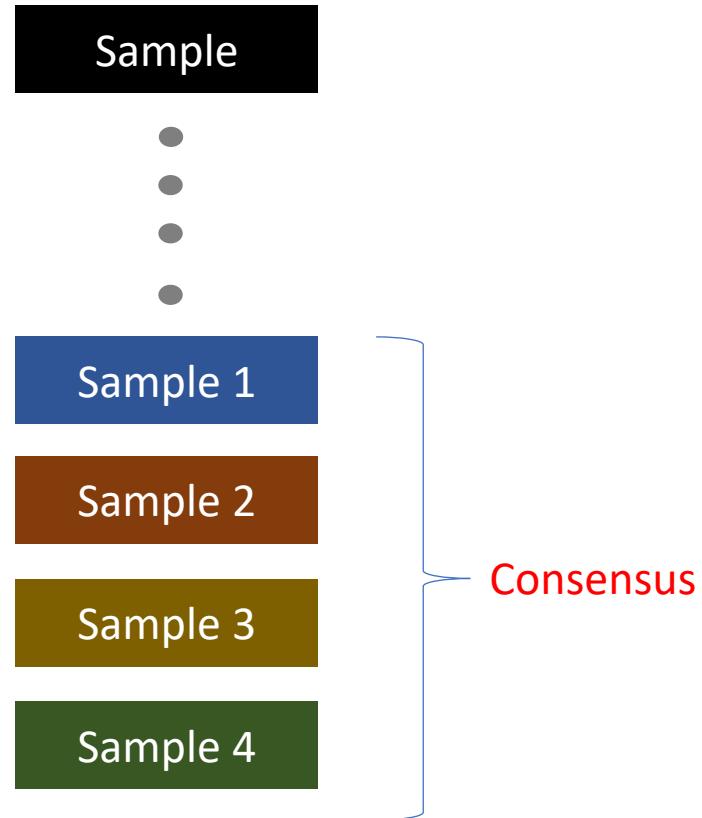
Sample 1

Sample 2

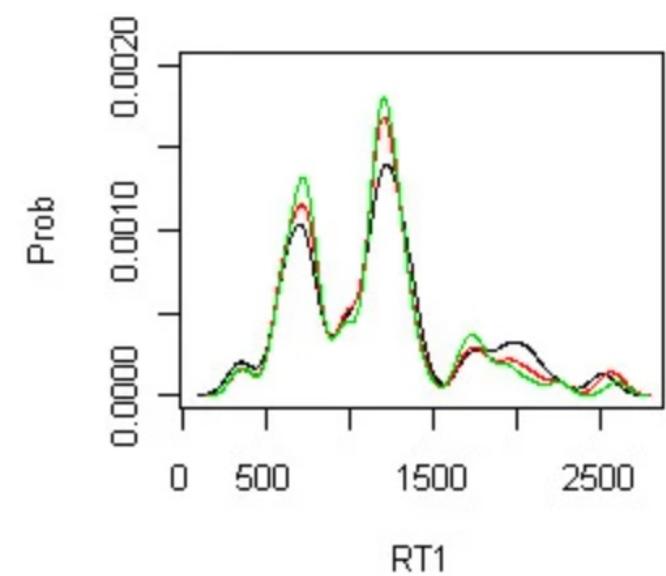
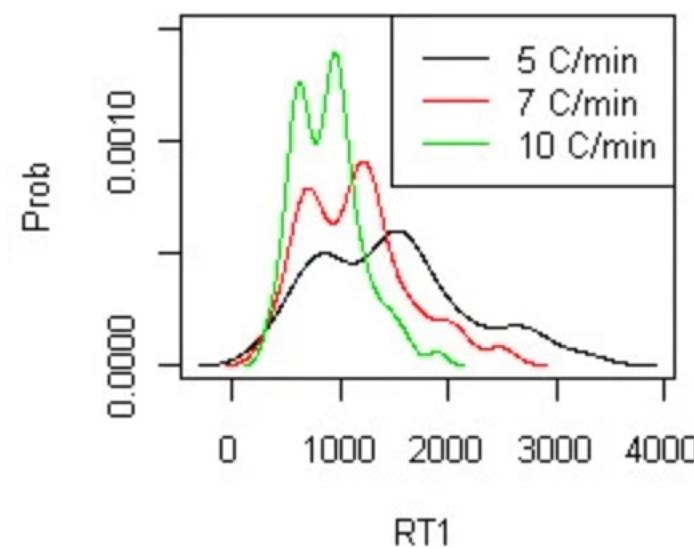
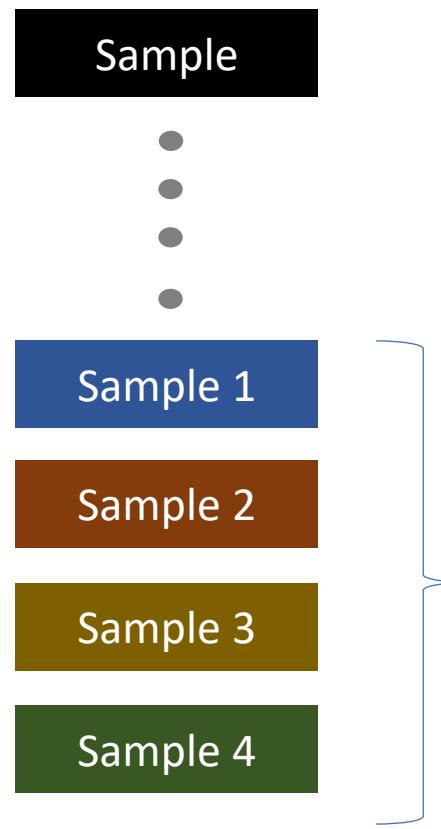
Sample 3

Sample 4

Data analysis using XCMS : Peak alignment

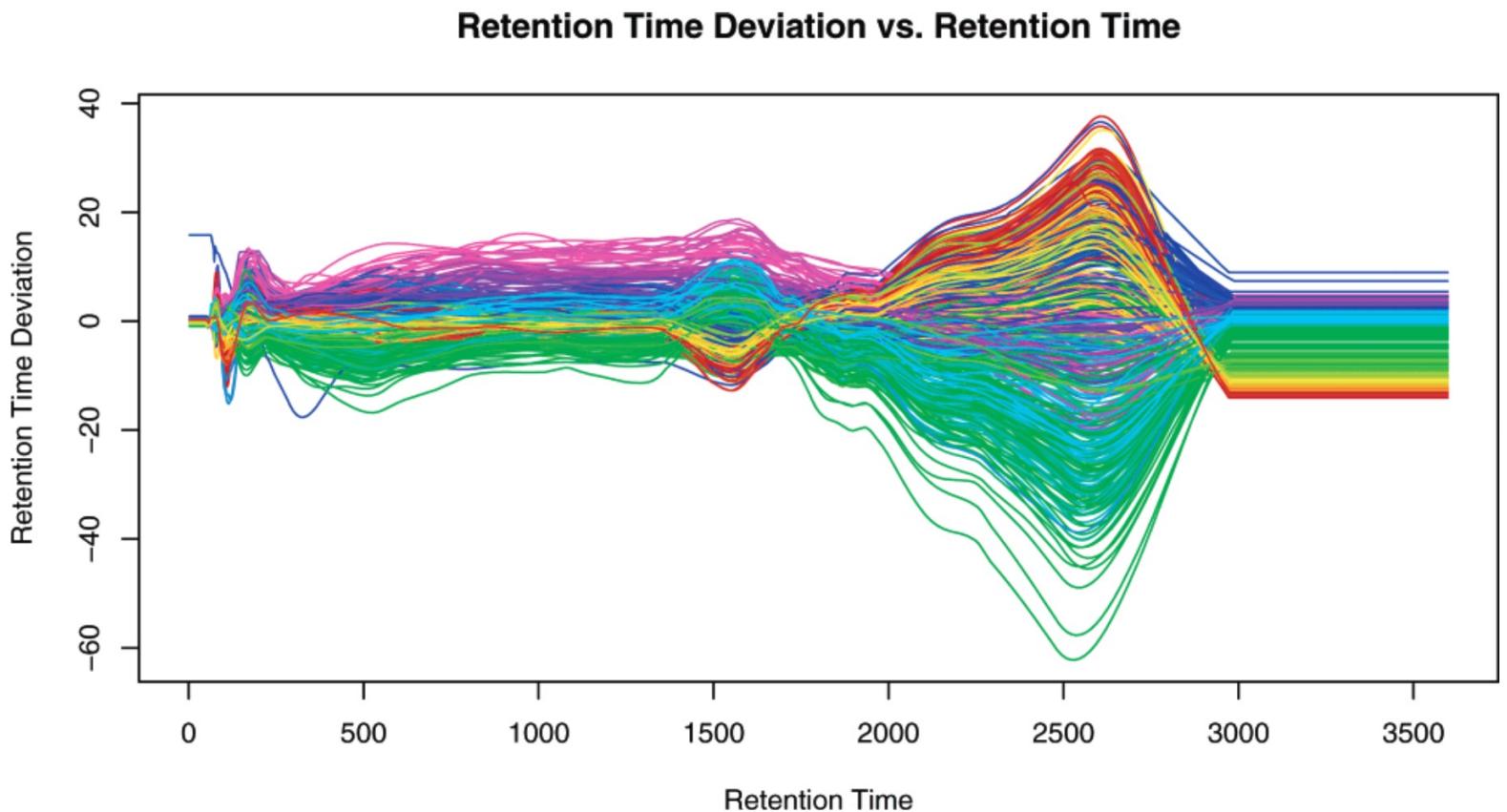
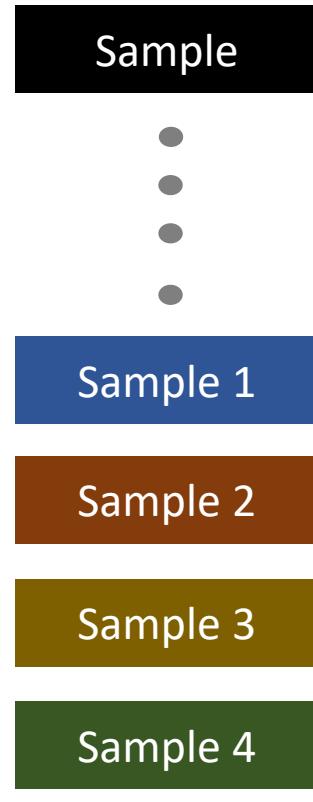


Data analysis using XCMS : Peak alignment

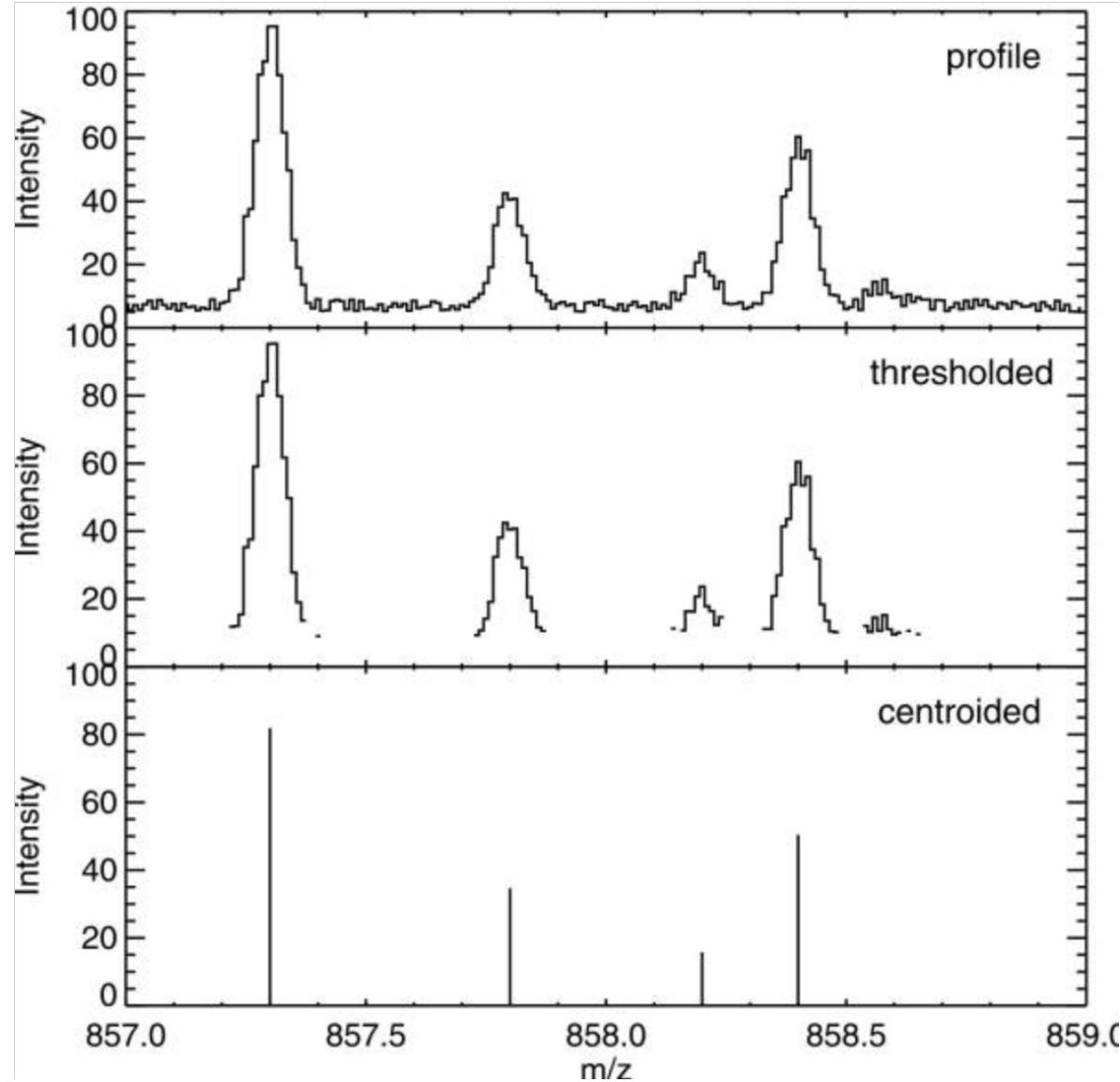


Methods : Binning based approach; Densioty based approach, Time warping, etc.

Data analysis using XCMS : Peak alignment



Data analysis using XCMS : Centroided data

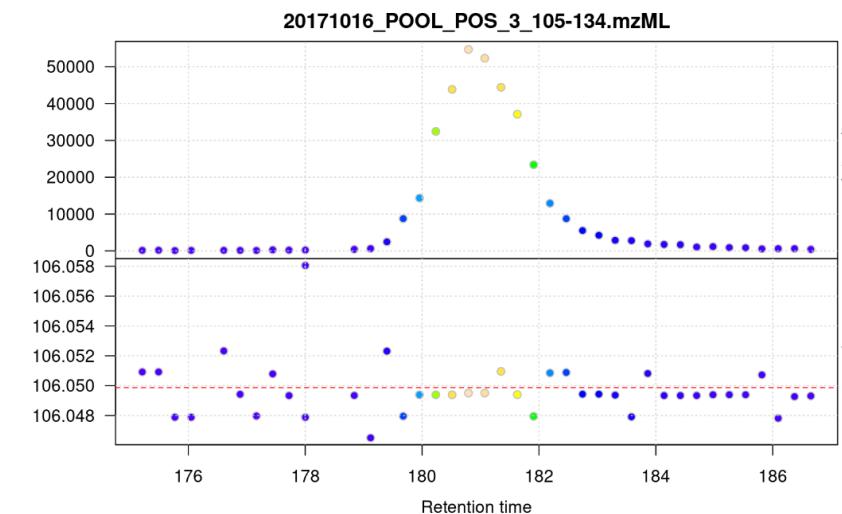
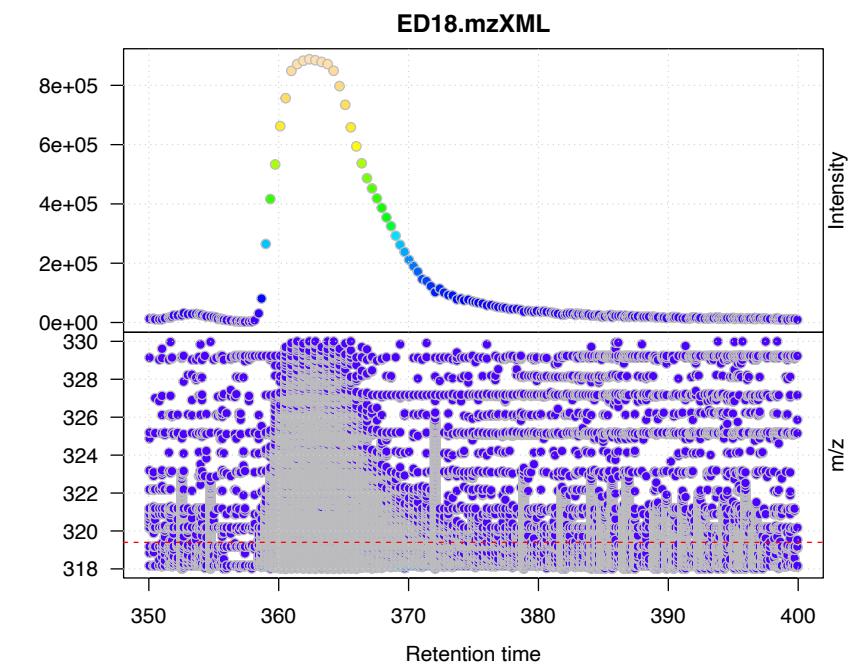
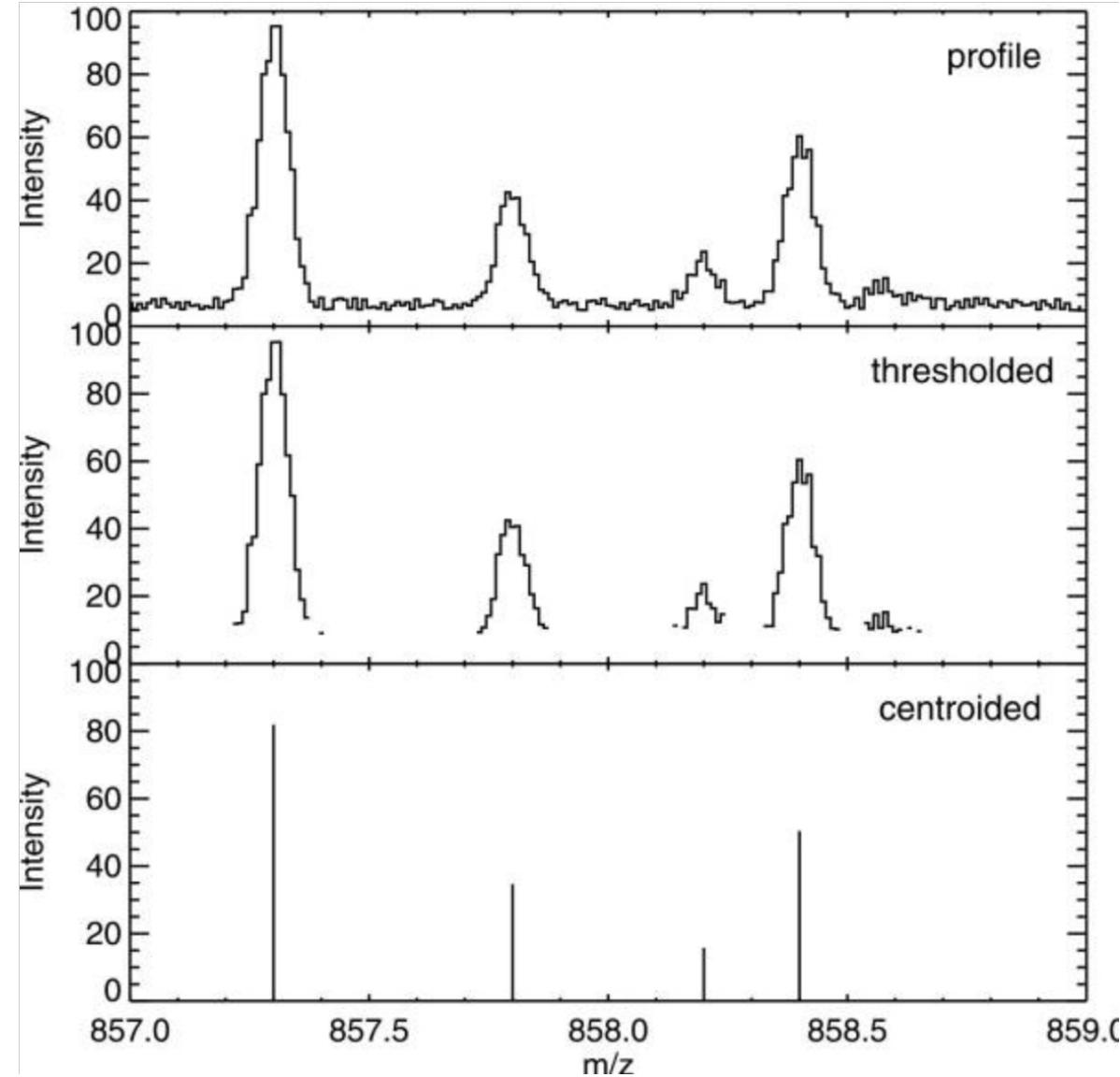


Centroiding aims to reduce the signal distribution to a single representative intensity, a single data point, for the ion in a spectrum. The simplest approach selects the largest intensity for each mass peak and report its intensity and m/z value.

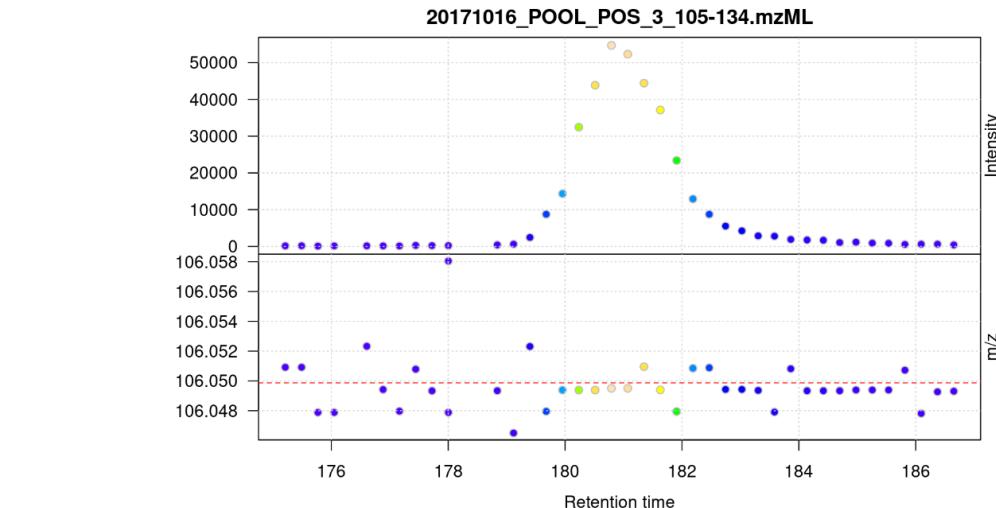
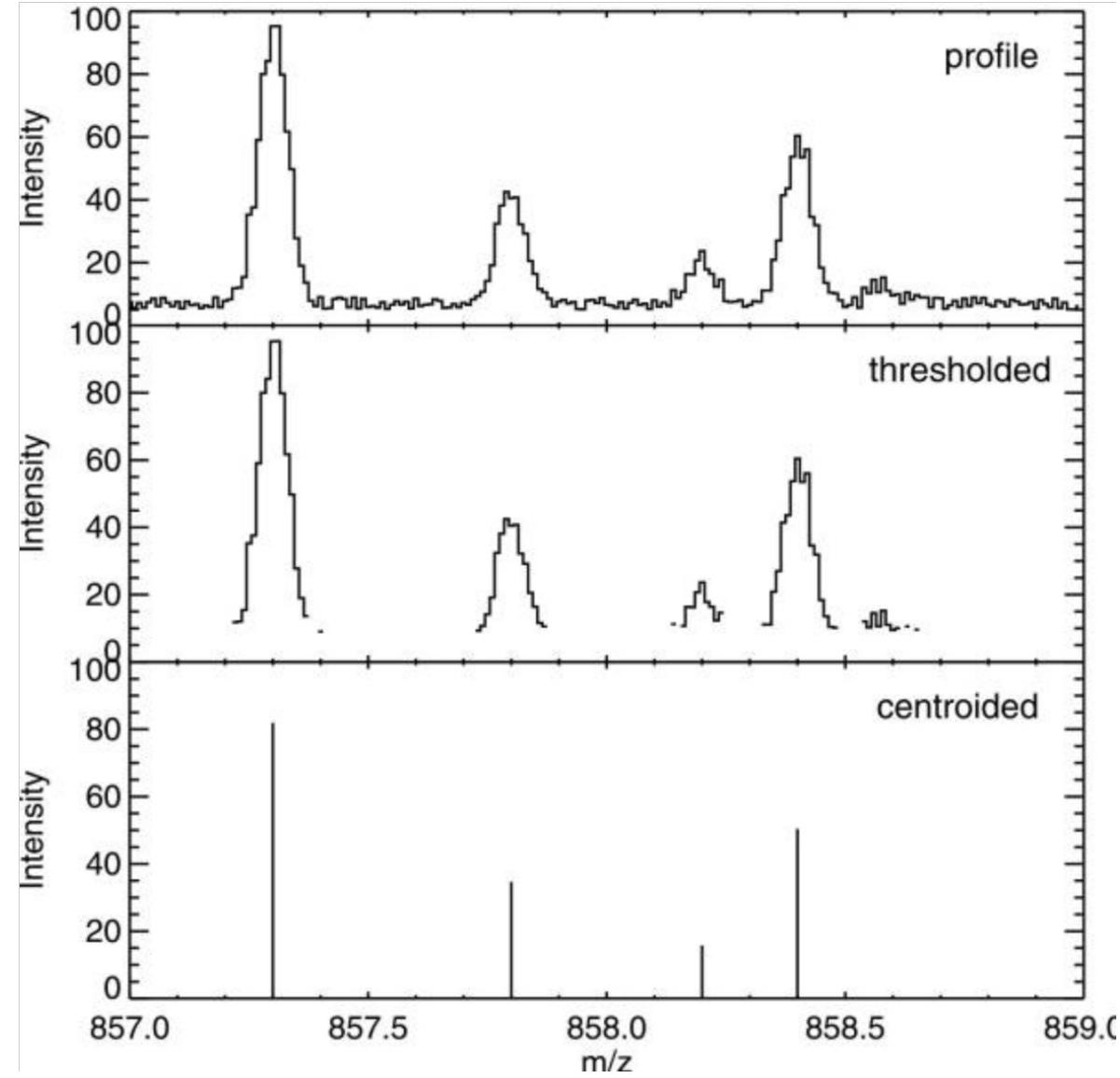
msCovert - To convert profile data to centroid data

XCMS - `pickPeaks()`

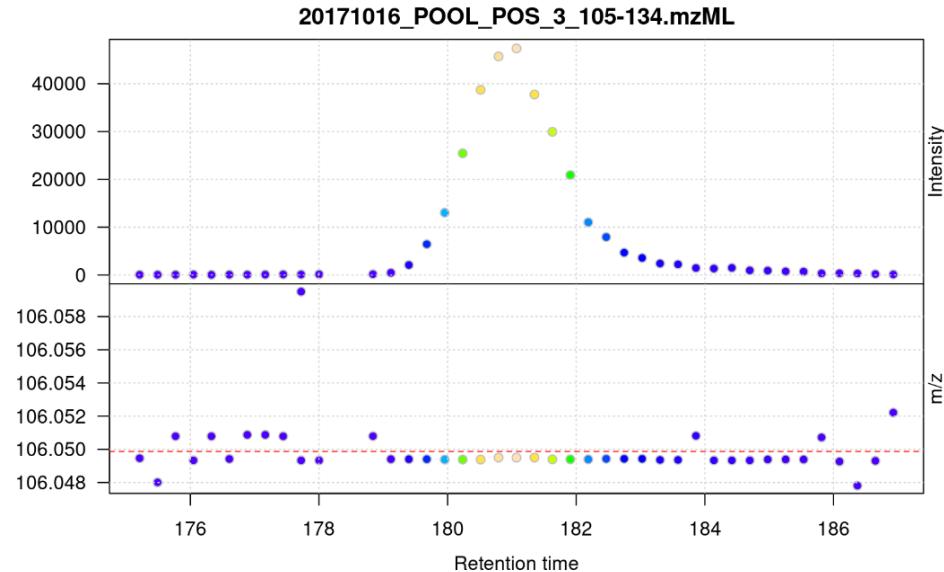
Data analysis using XCMS : Centroided data



Data analysis using XCMS : Centroided data



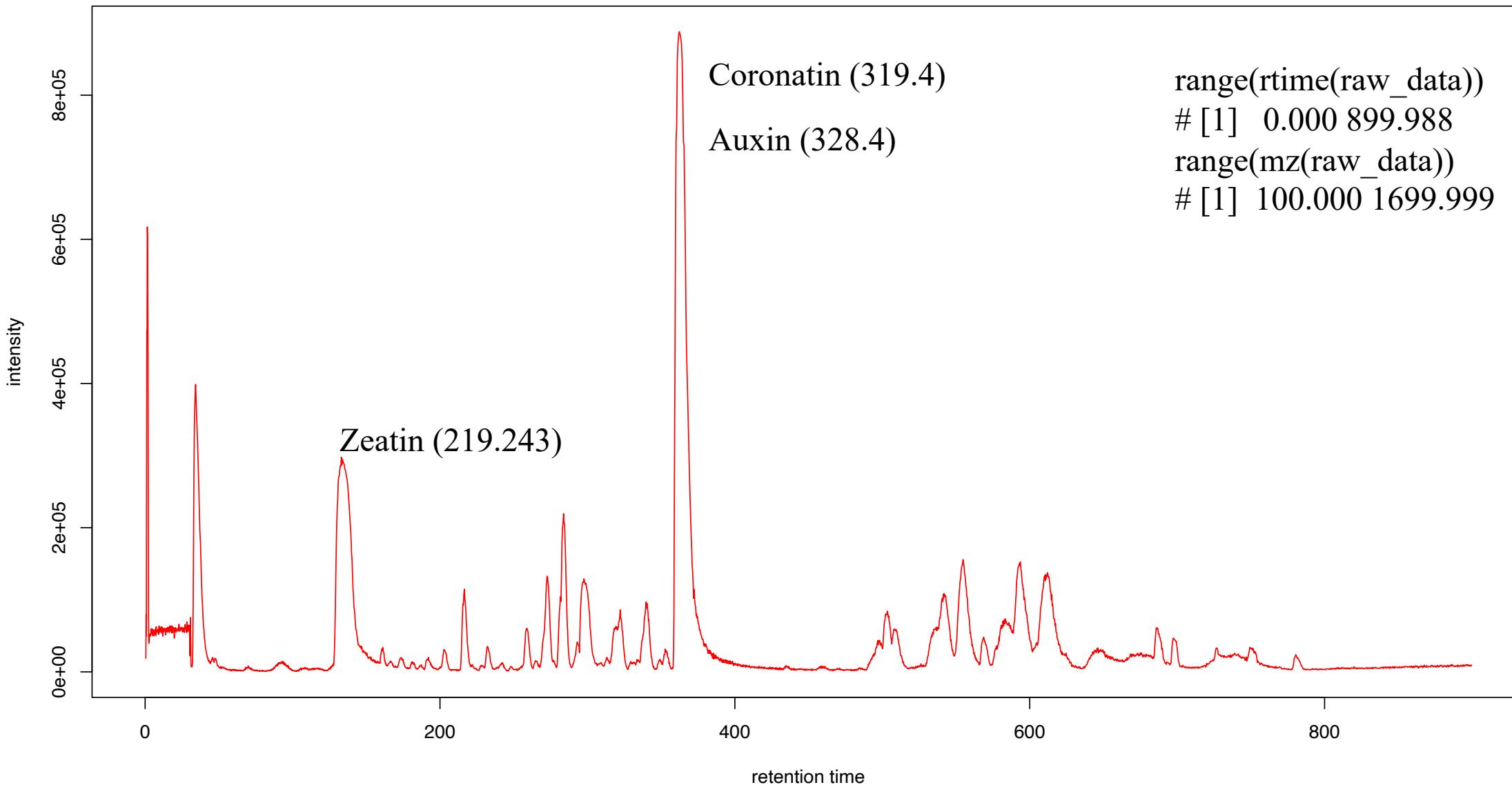
`smooth(method = "SavitzkyGolay", halfWindowSize = 4L)`



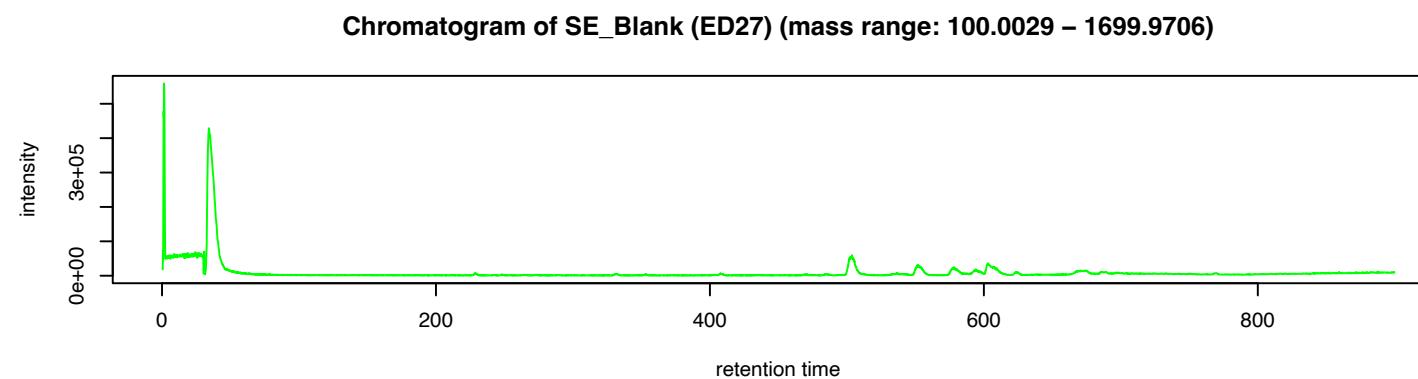
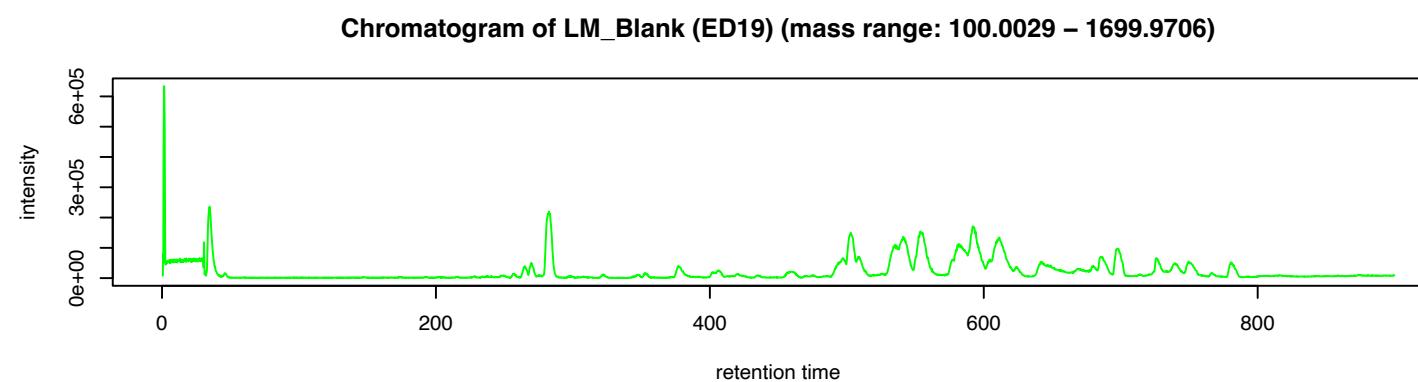
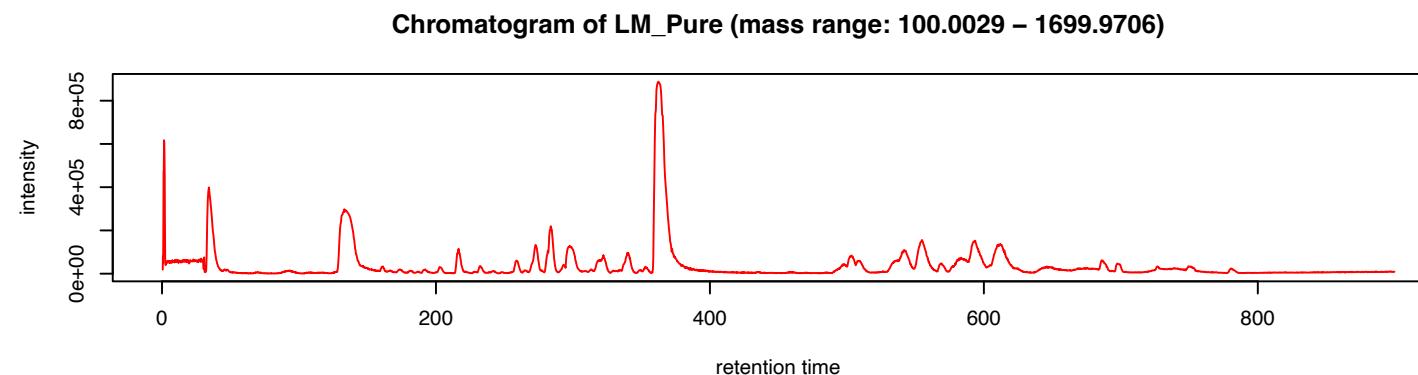
Use of known compounds as samples

Sample 18 – Known compounds

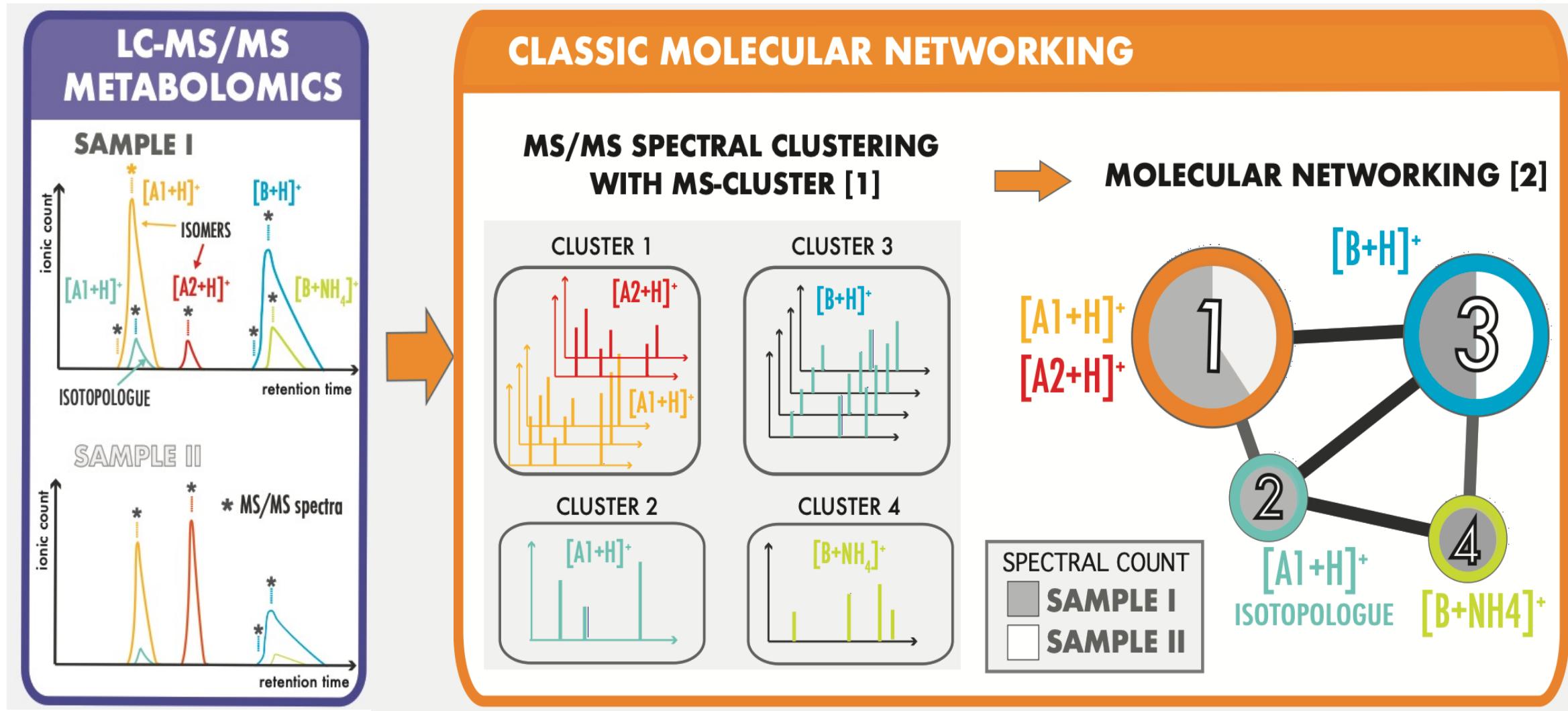
Chromatogram of LM_Pure (mass range: 100.0029 – 1699.9706)



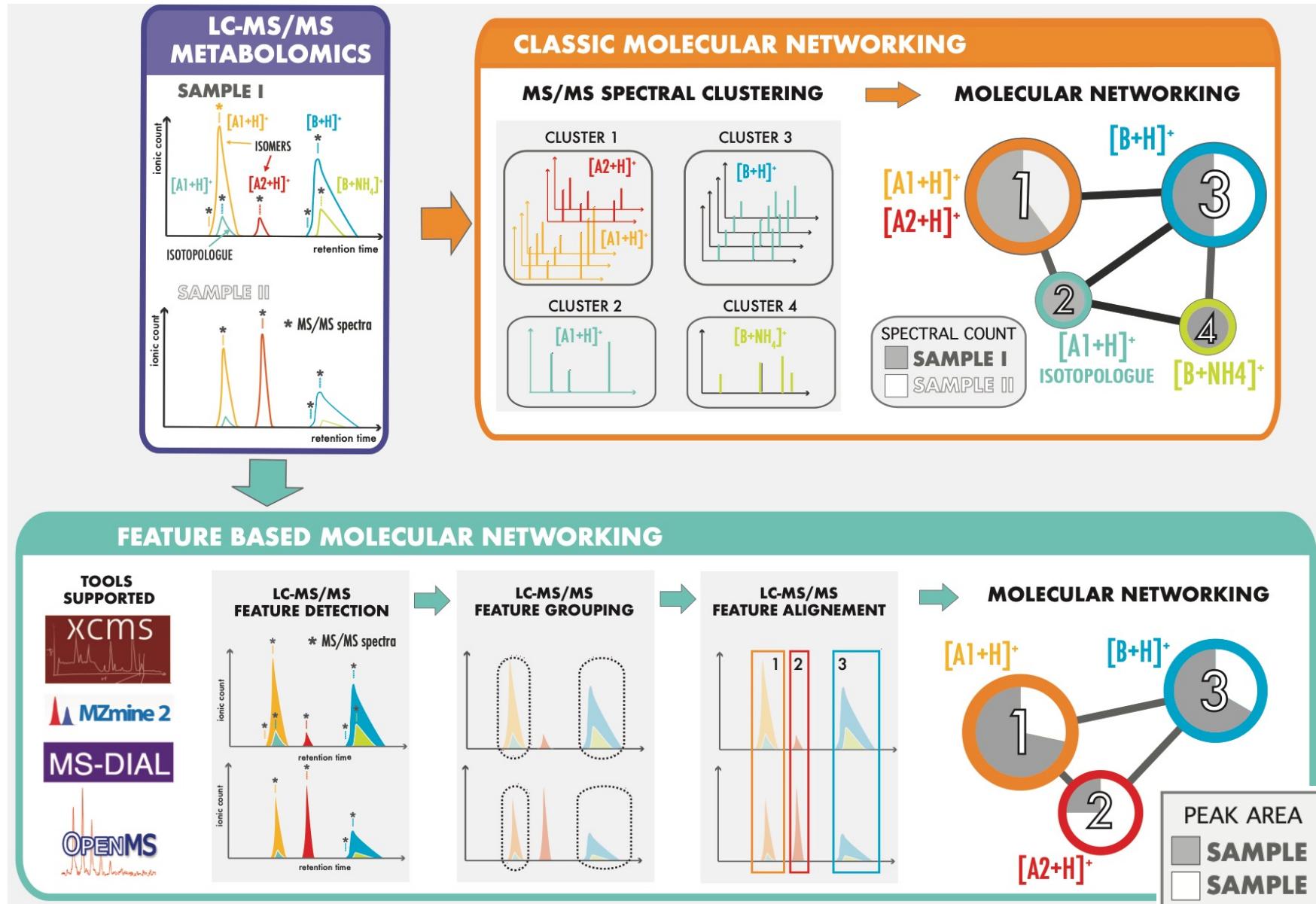
Use of known compounds and blanks



GNPS based classical molecular networking (MN)

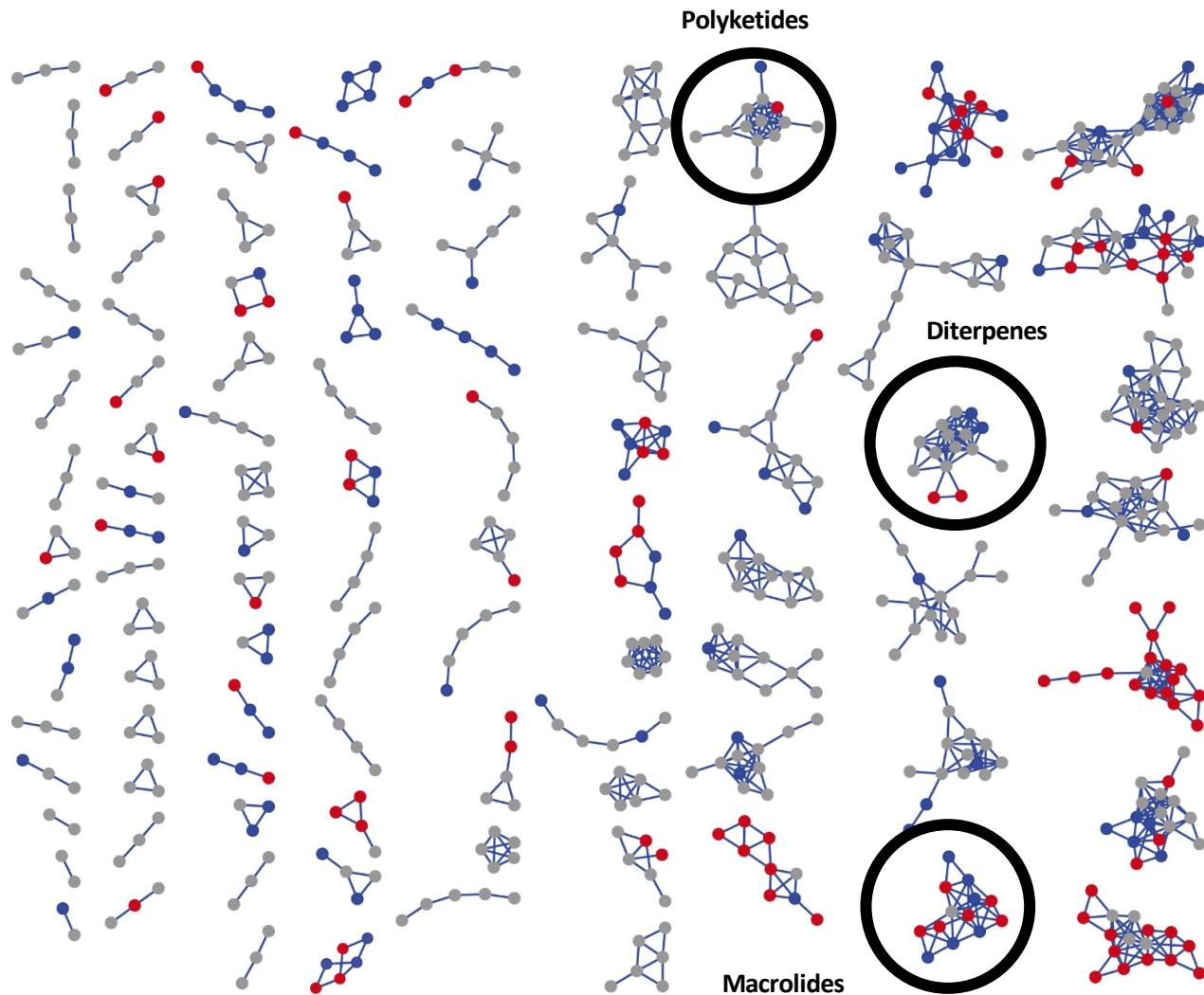


Two methods of molecular networking (MN)



GNPS Tutorial Module 5 –
Feature Based Molecular Networking
Slide by Louis Felix Nothias, GNPS

Molecular Networking (MN)



Wang et al. *Nat. Biotech.* 2016; Slide with D. Petras.

GNPS Tutorial Module 5 - Feature Based Molecular Networking
Slide by Louis Felix Nothias, GNPS

Thank you for your attention!



For discussion / collaboration



k.s.singh@uu.nl
kumar.singh@wur.nl



@Kumar_S_Singh



www.kumarsaurabhSingh.com



+31 (0)657871628



Hugo R. Kruyt building
Padualaan 8, 3584 CH Utrecht
The Netherlands



Utrecht University

NWO | Netherlands Organisation
for Scientific Research

WAGENINGEN
UNIVERSITY & RESEARCH

Mass Spectrometry technologies

Magnetic sector mass spectrometer: Ions with constant kinetic Energy but different mass are soperated by their trajectories in a magnetic field. Kinetic energy is given to the ions via the voltage difference between the acceleration plates in the ion source. The resolving power of a mass spectrometer is limited by the range of kineticenergy of the ions (typically $\sim 0.1\%$). This variation leads to a resolving power of $\sim 10^3$ (resolution of 0.1 at m/z 100). Youc an further increase the resolution by **double focussing mass spectrometers**. An electric sector is added to the ion path prior to the magnetic sector. This way you can select charged ions of a specific KE to enter the magentic field via exit slit. Rest all other ions fail to pass exit slit. This process decreases the variation in KEs of charged ions which further leads to a resolving power of $\sim 10^5$ (resolution of 0.001 at m/z 100).

Transmission Quadrupole MS: Uses additional column of 4 charged poles (+/-+). These charged poles carries DC and AC (Constant voltage and Oscillating coltage). This allows ions to oscillates more and facilitates the selection of specific charged ion (instance $(M+H)^+$). This is very cheap and thus failrly common. It is also able to record 2-8 spectra per second and cover a range up to 4000 m/z units. Sufficient to resolve peaks separated by m/z ~ 0.3 .

NOTE: While MSMS and DFMS act at constant resolving power, quadrupoles operate at constant resolution. That is, in the in the former cateory, separation between peaks decreases as m/z increases. But in the later category, the separation of peaks remains constant as m/z increases.

Time-Of-Flight MS: This uses the time required for ions to travel from the source to the detector as a way of separating ions with different m/z. Assuming all ions have the same KE, then their velocity vary according to $KE = \frac{1}{2} mv^2$. That is lighter ions will be travelling faster than heavier ions. The resolving power here is $10000 - 25000$ (separation of m/z ~ 0.001). The TOF-MS can collect many spectra per second ($10^2 - 10^4$) and can measure very high masses.

3D Quadrupole ion-trap MS: uses a 3D ion trap. Sensitivity is very high ($\sim 1-10$ pg). Resolving power is 1000, 4000 and accuracy of m/z is 0.1.
Linear Quadrupole Ion-Trap MS: Higer sensitiivty the 3DQITMS.

Orbitrap MS: Does require either magentic field or Oscillating current. It uses potential differences to push ions in the center.