Mass Spectrometry and Proteomics/Metabolomics using R and Bioconductor

Laurent Gatto
lg390@cam.ac.uk

Cambridge Centre For Proteomics
University of Cambridge

European Bioinformatics Institute (EBI)

18^{th} November 2010

Plan

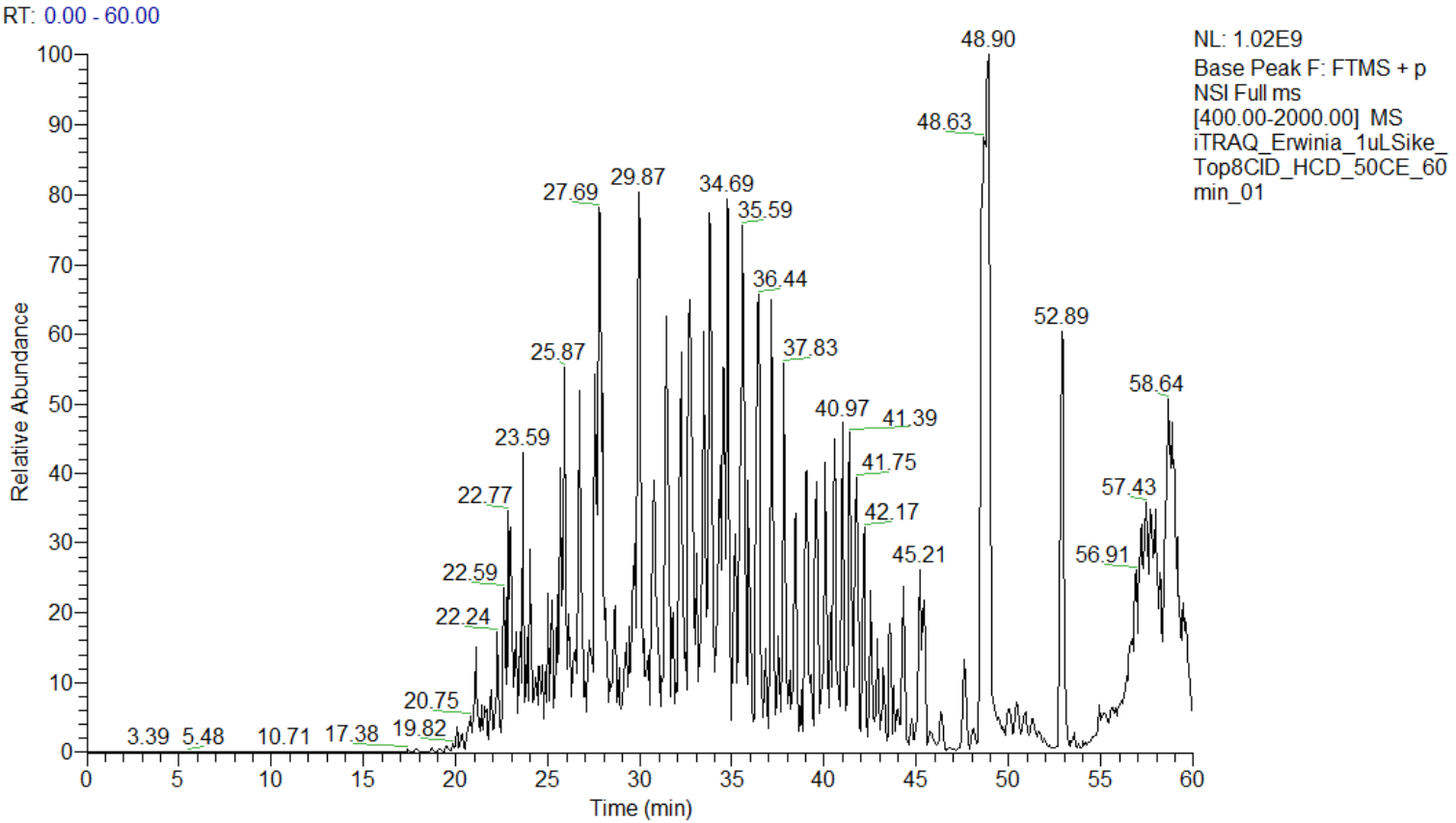
Plan

Short definition

- Mass spectrometry (MS) is an analytical technique that measures the mass-to-charge ratio of charged particles.
- Used to study various chemical compounds – peptides (as surrogates for proteins) or metabolites (see `xcms` talk).
- Allows **identification** and **quantification**.

Separation based on analyte physical properties

- HP[L|G]C (online) → **chromatogram**
- 1D and 2D Gels using Cy-dyes labelled proteins (`digeR` talk)



Separation

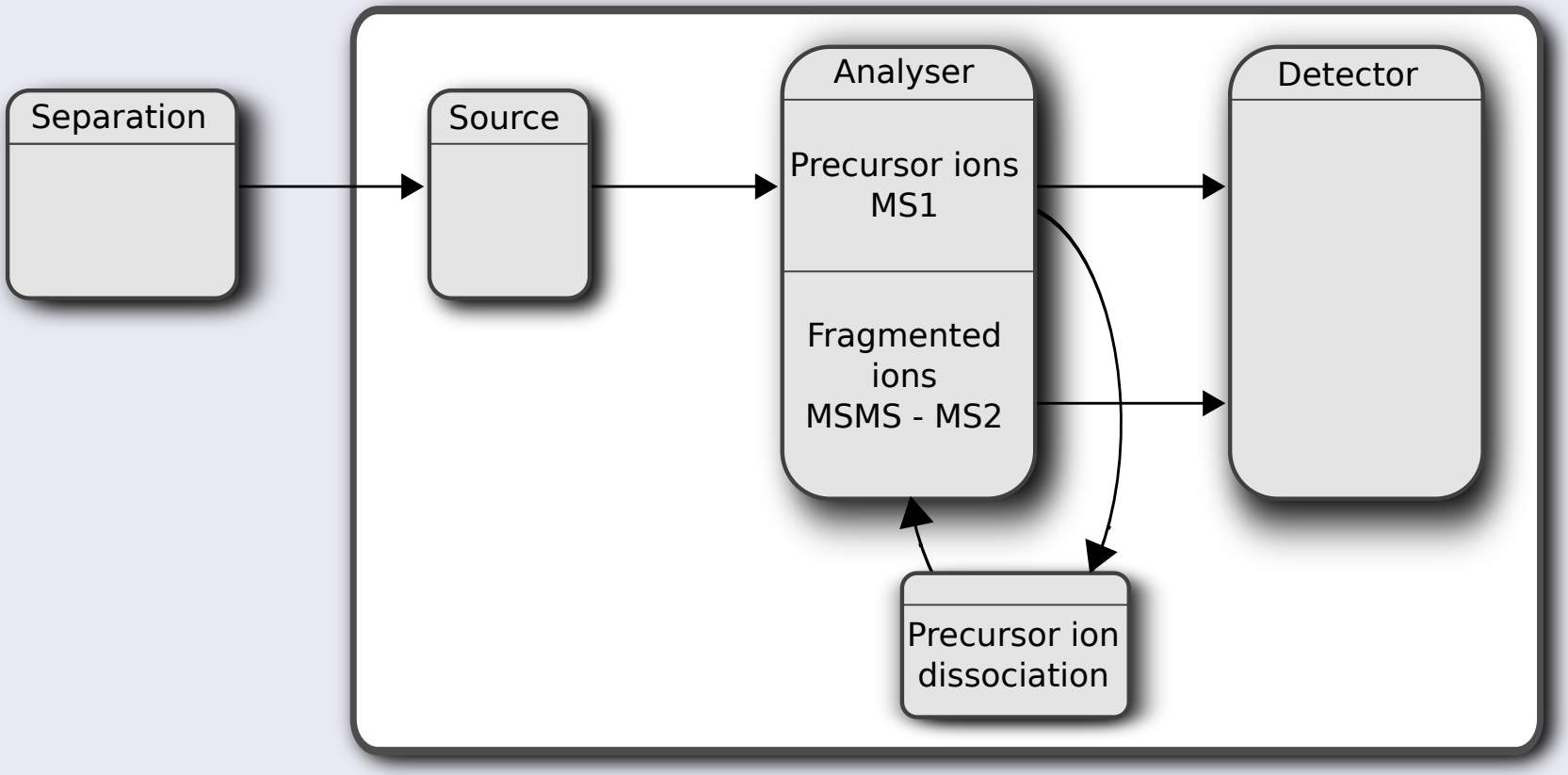
MS stages

Across retention time...

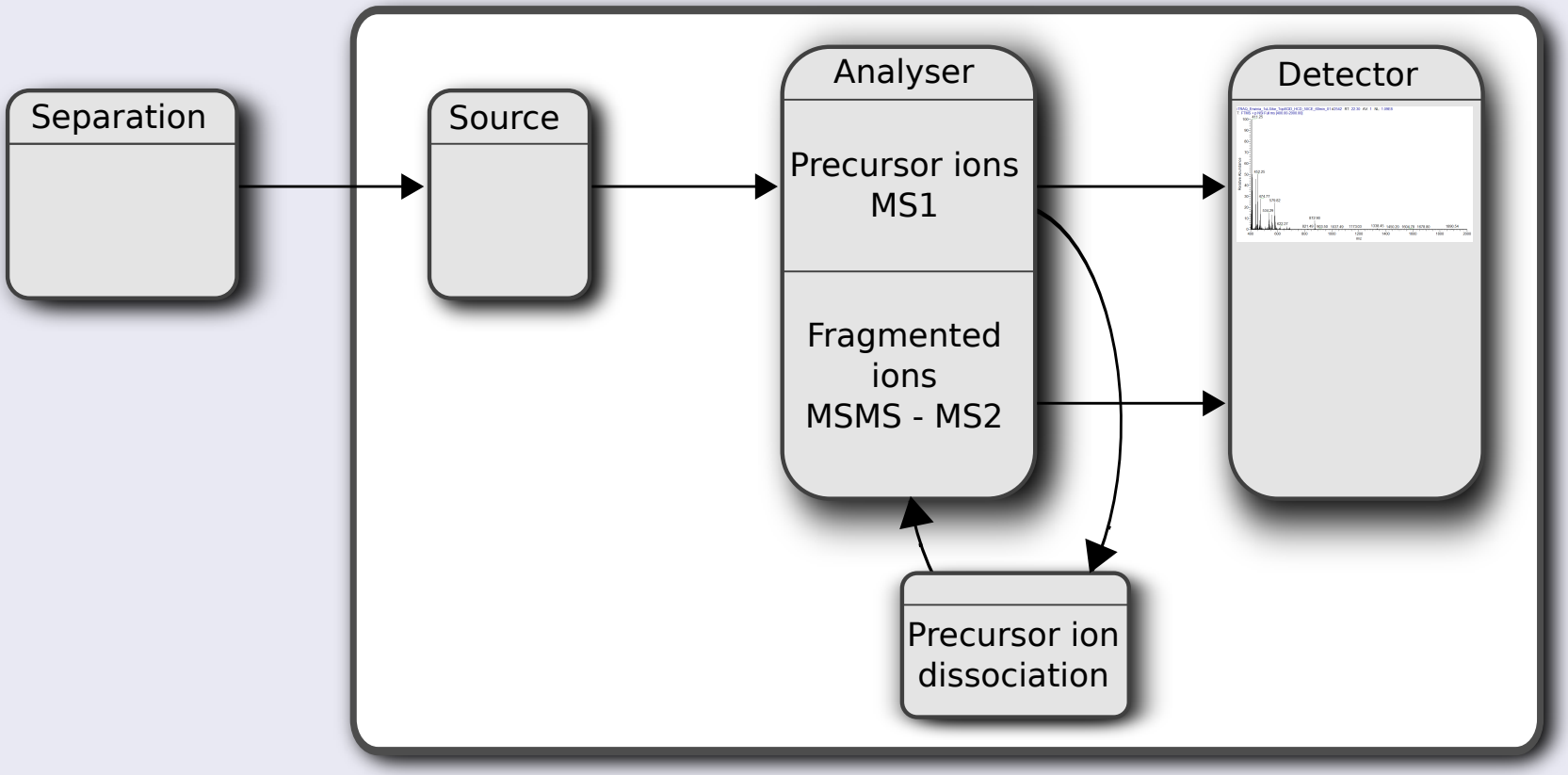
- Ionisation source – formation of gas-phase ions (analytes): ESI, MALDI
- Mass analyzer – separation of the ions according to their mass (M) to charge (Z) ratio.
- Detector – ion current monitoring and amplification (ion counts)

$\rightarrow$ mass **spectrum** (intensity vs. M/Z)

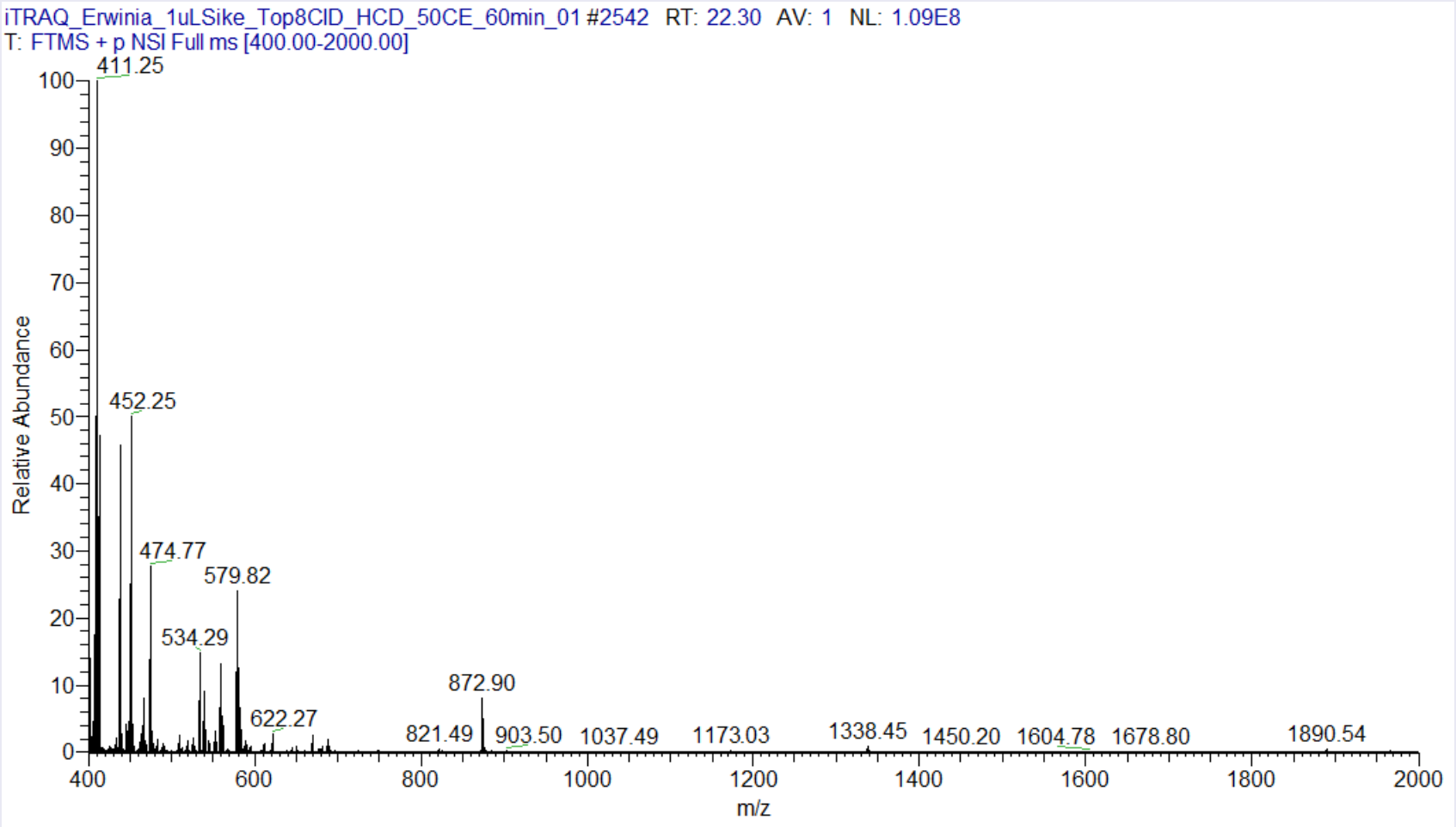
Schematic MS(MS) worflow



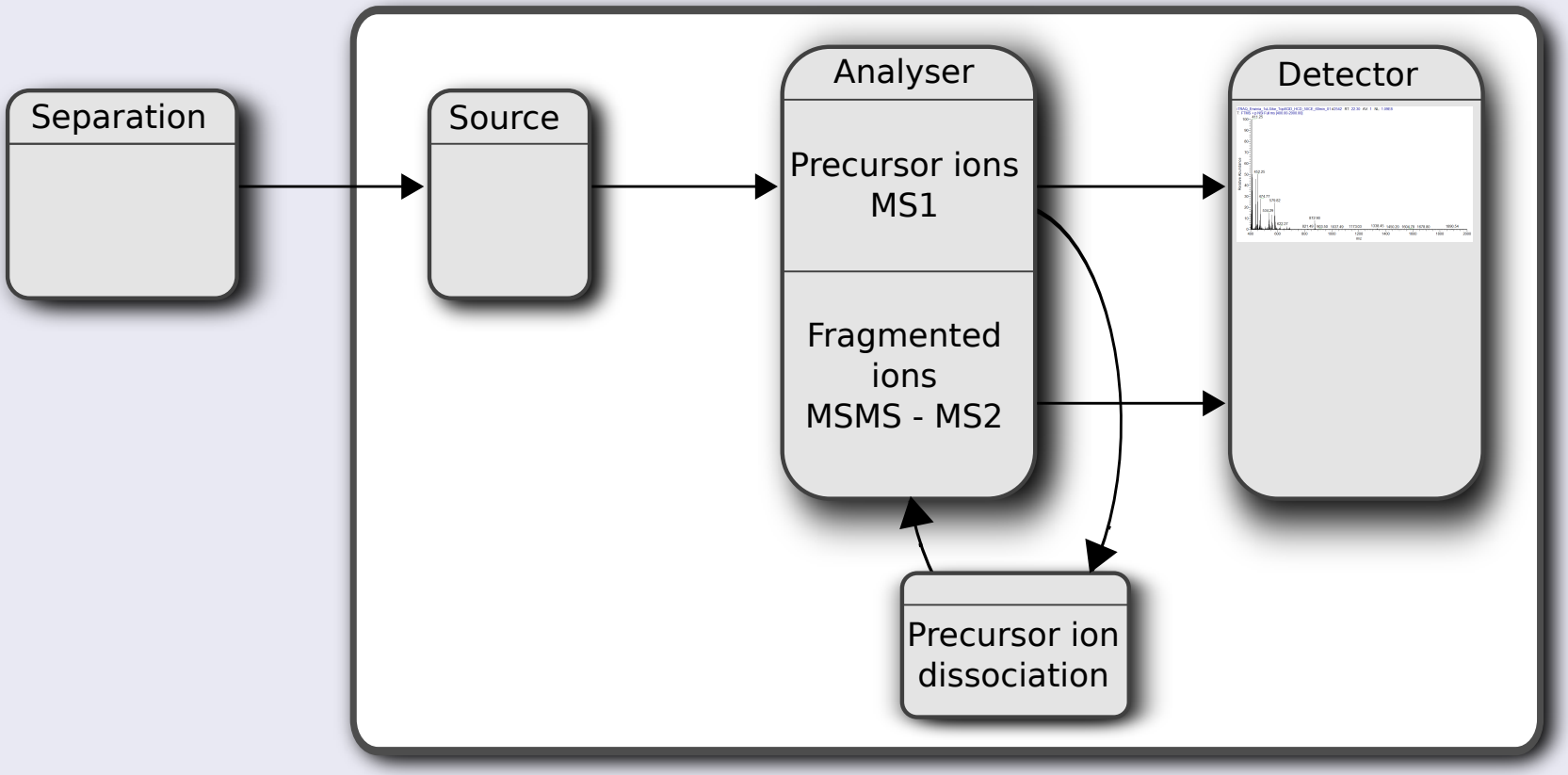
Schematic MS(MS) worflow



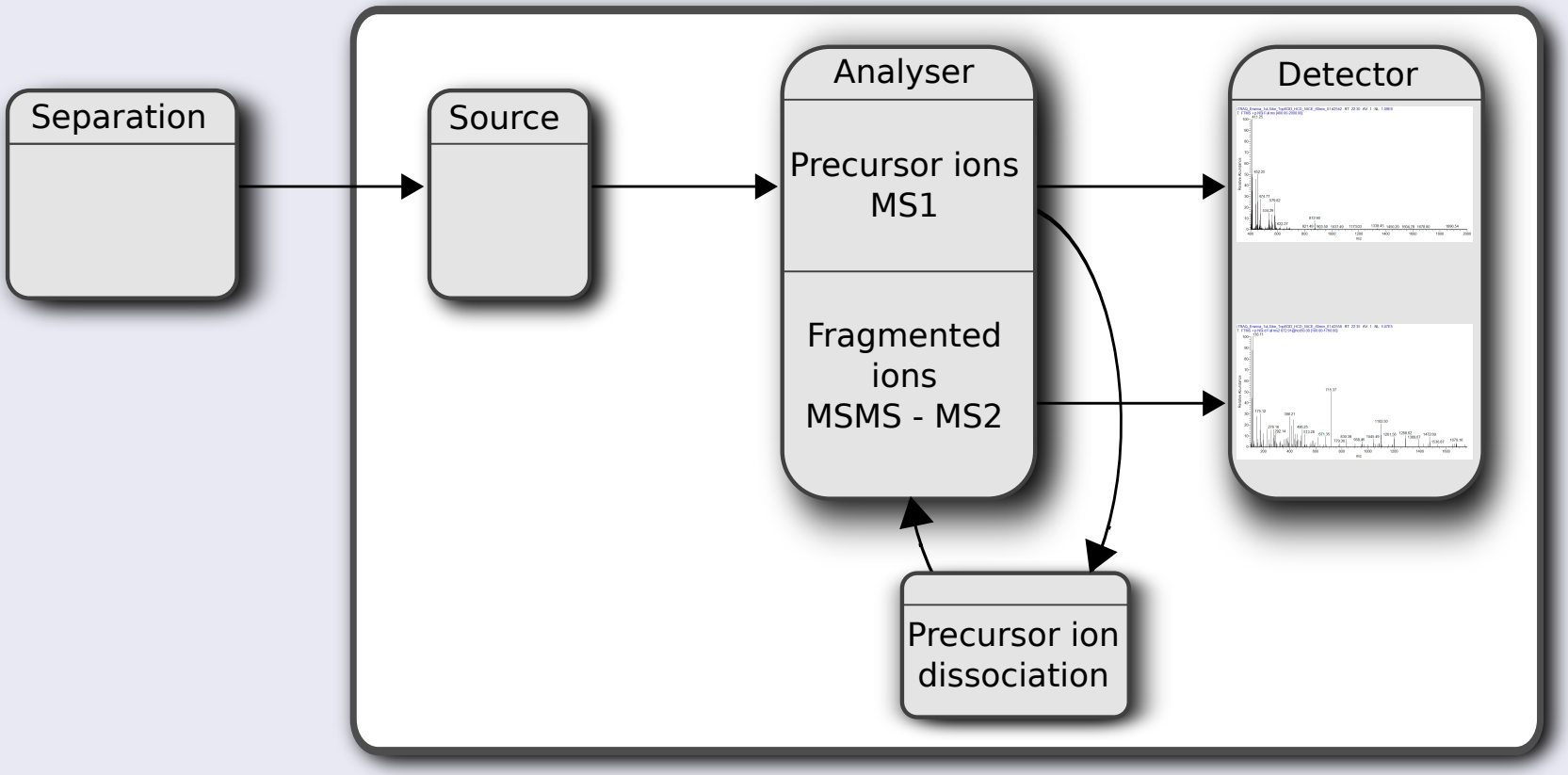
MS1 scan


iTRAQ_Erwinia_1uLSike_Top8CID_HCD_50CE_60min_01 #2542 RT: 22.30 AV: 1 NL: 1.09E8
T: FTMS + p NSI Full ms [400.00-2000.00]
Relative Abundance
m/z
411.25
452.25
474.77
579.82
534.29
622.27
872.90
821.49
903.50
1037.49
1173.03
1338.45
1450.20
1604.78
1678.80
1890.54

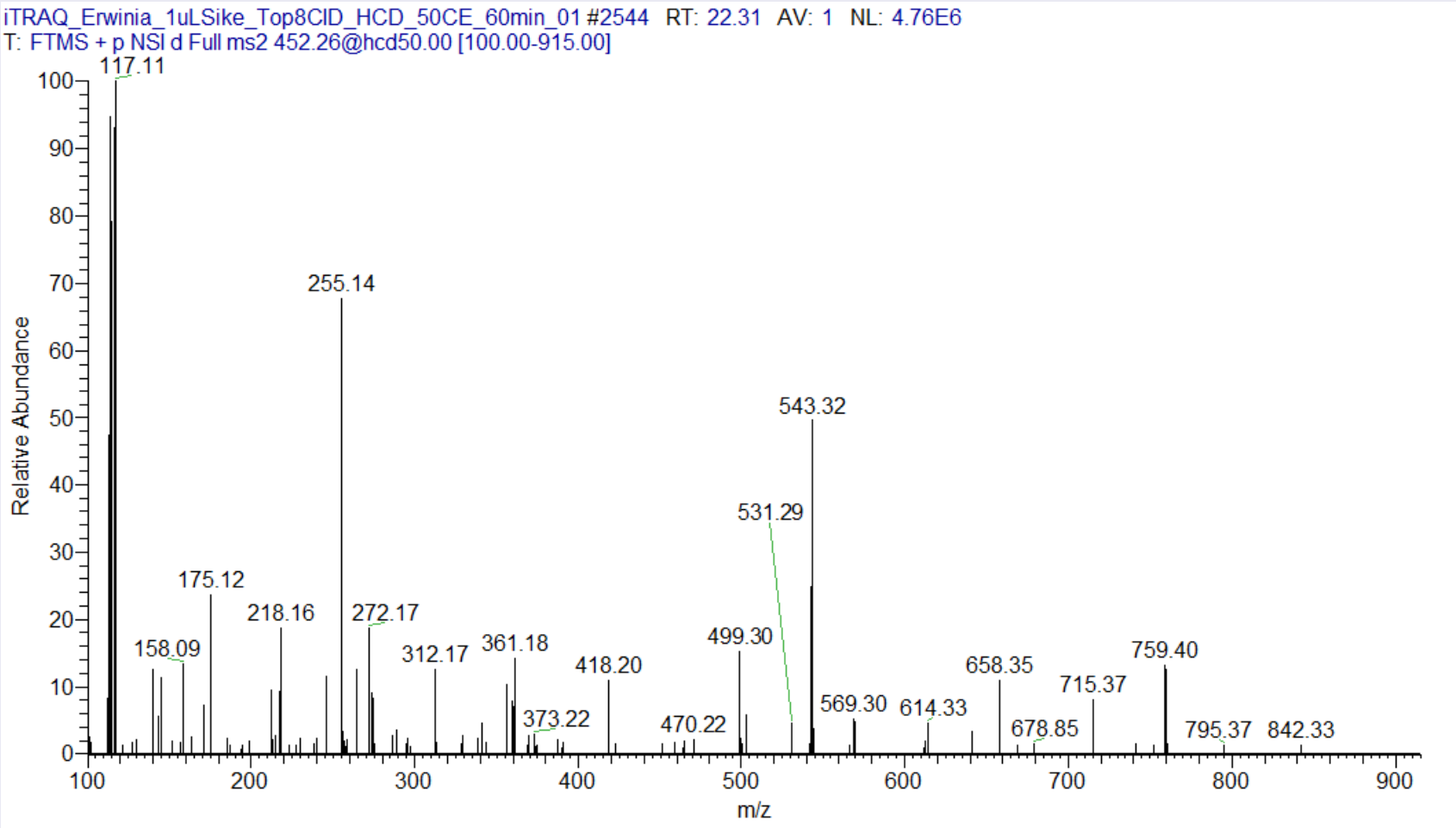
Schematic MS(MS) worflow



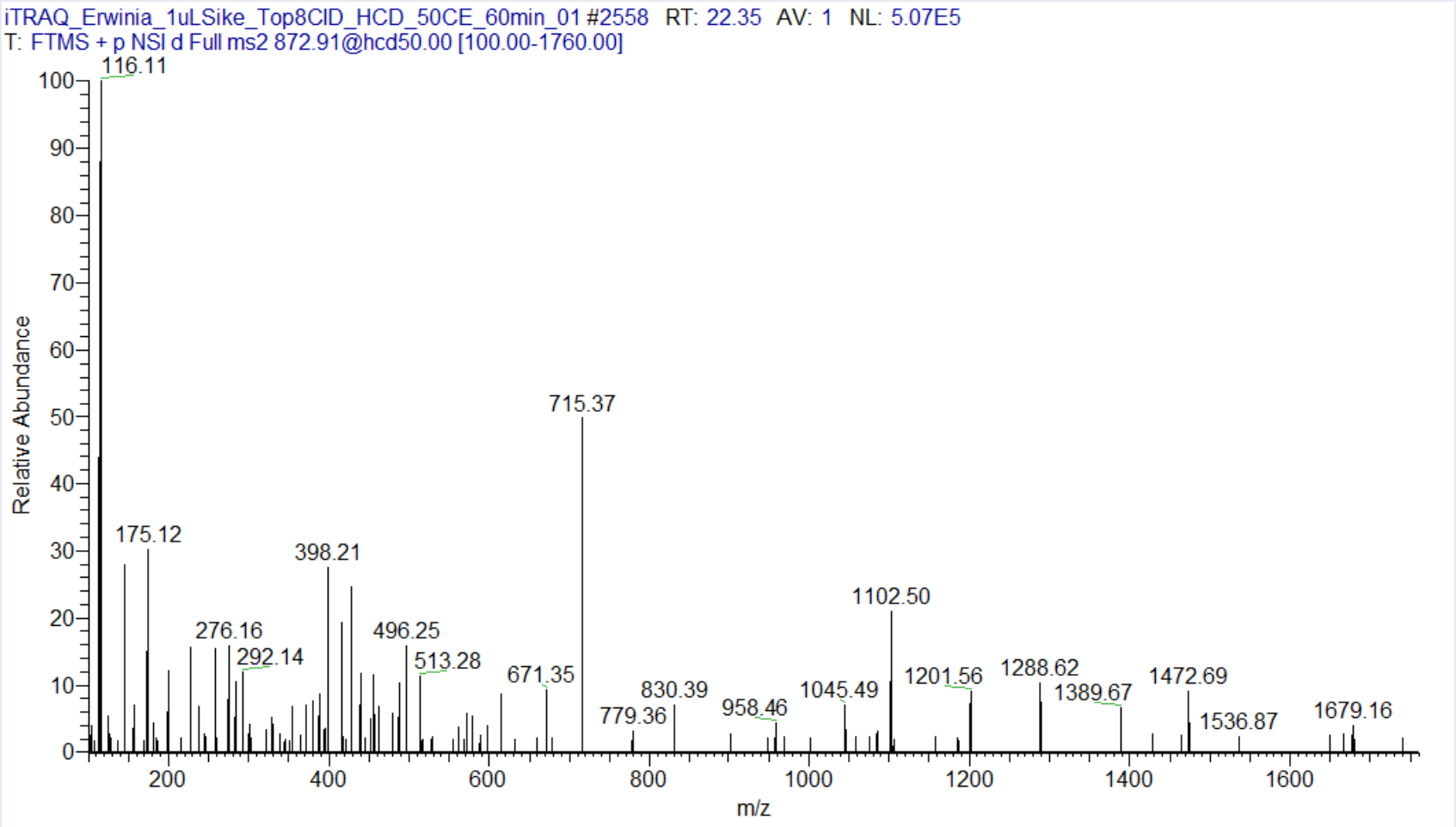
Schematic MS(MS) worflow



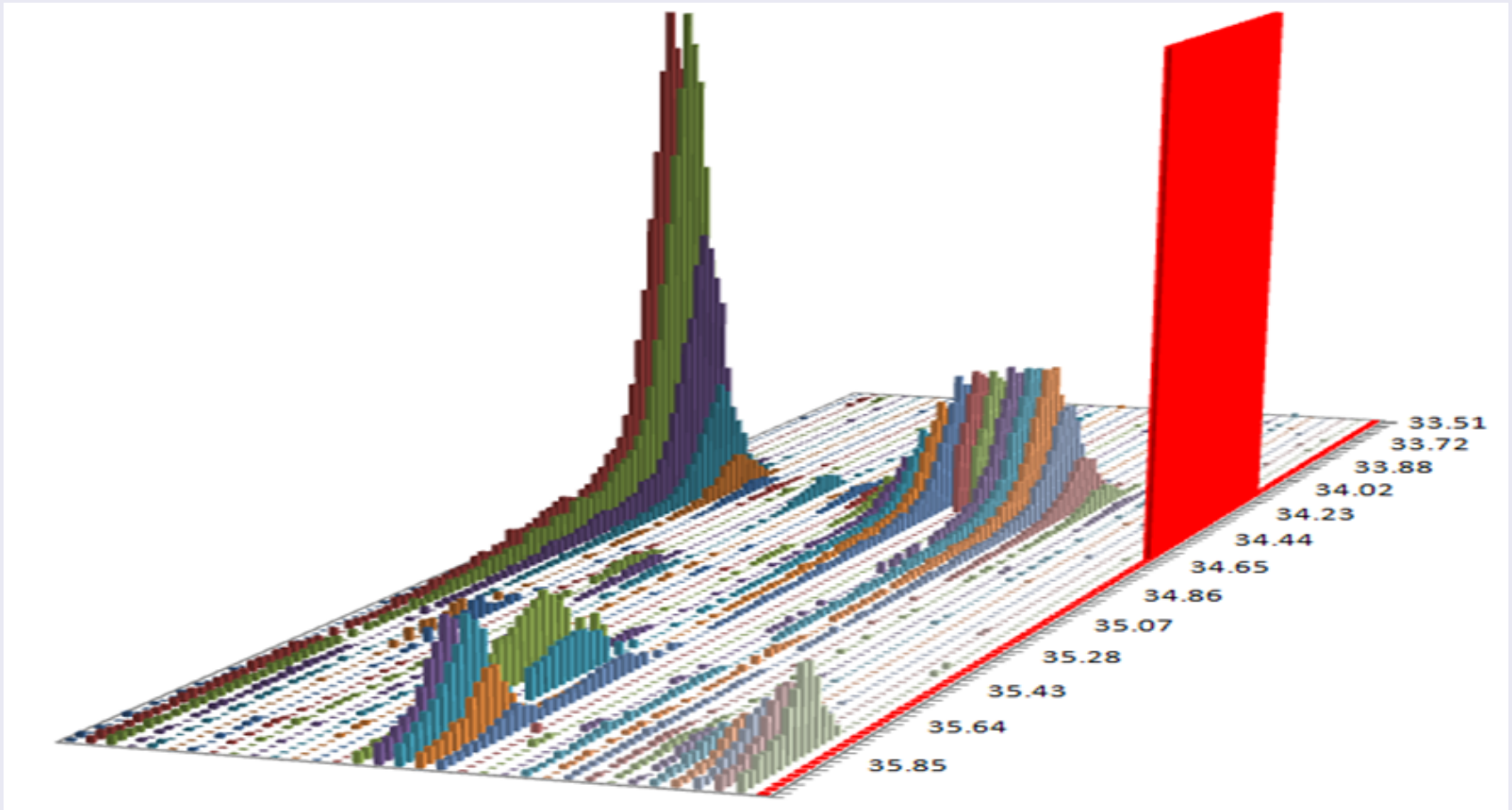
MSMS scans


iTRAQ_Erwinia_1uLSike_Top8CID_HCD_50CE_60min_01 #2544 RT: 22.31 AV: 1 NL: 4.76E6
T: FTMS + p NSI d Full ms2 452.26@hcd50.00 [100.00-915.00]
Relative Abundance
m/z
117.11
255.14
543.32
531.29
175.12
218.16
272.17
158.09
312.17
361.18
418.20
499.30
658.35
759.40
715.37
569.30
614.33
373.22
470.22
678.85
795.37
842.33

MSMS scans

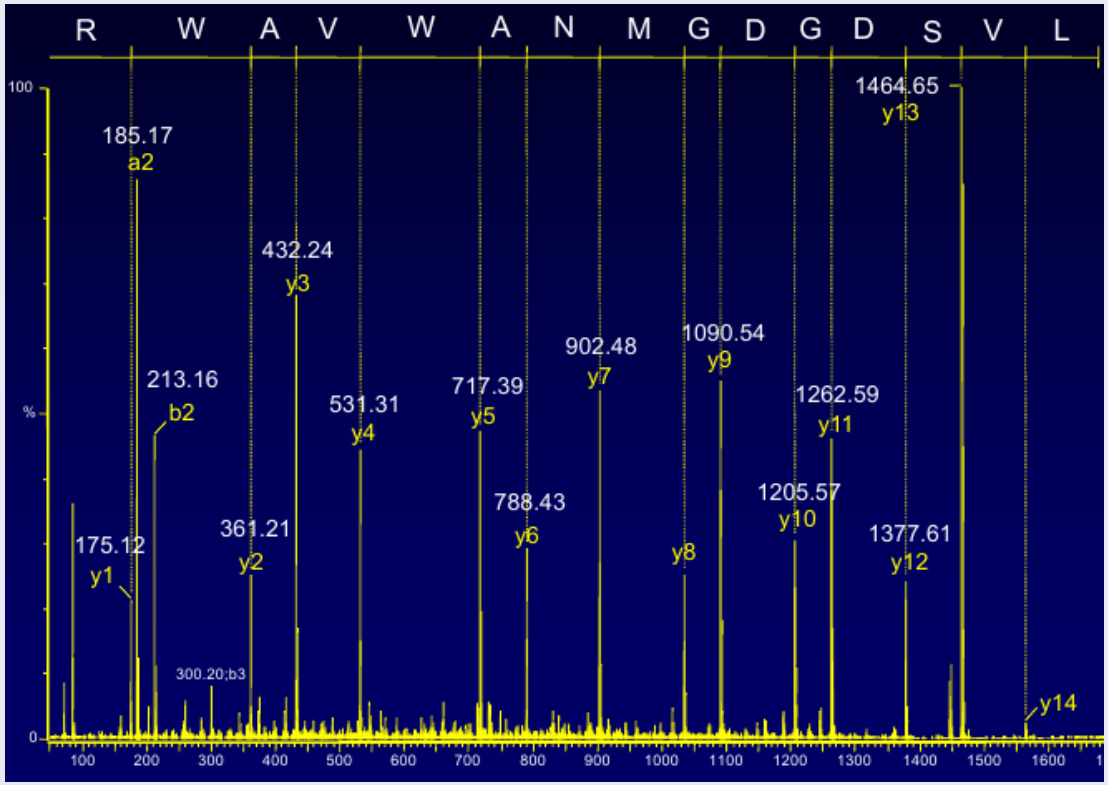

iTRAQ_Erwinia_1uLSike_Top8CID_HCD_50CE_60min_01 #2558 RT: 22.35 AV: 1 NL: 5.07E5
T: FTMS + p NSI d Full ms2 872.91@hcd50.00 [100.00-1760.00]
Relative Abundance
m/z
116.11
175.12
276.16
292.14
398.21
496.25
513.28
671.35
715.37
779.36
830.39
958.46
1045.49
1102.50
1201.56
1288.62
1389.67
1472.69
1536.87
1679.16

MZ vs. time vs int.



Identification

- MS1 M/Z for the *usual suspects* in metabolomics.
- MS2 spectra are matched against theoretical spectra databases.
- De novo peptide sequencing.

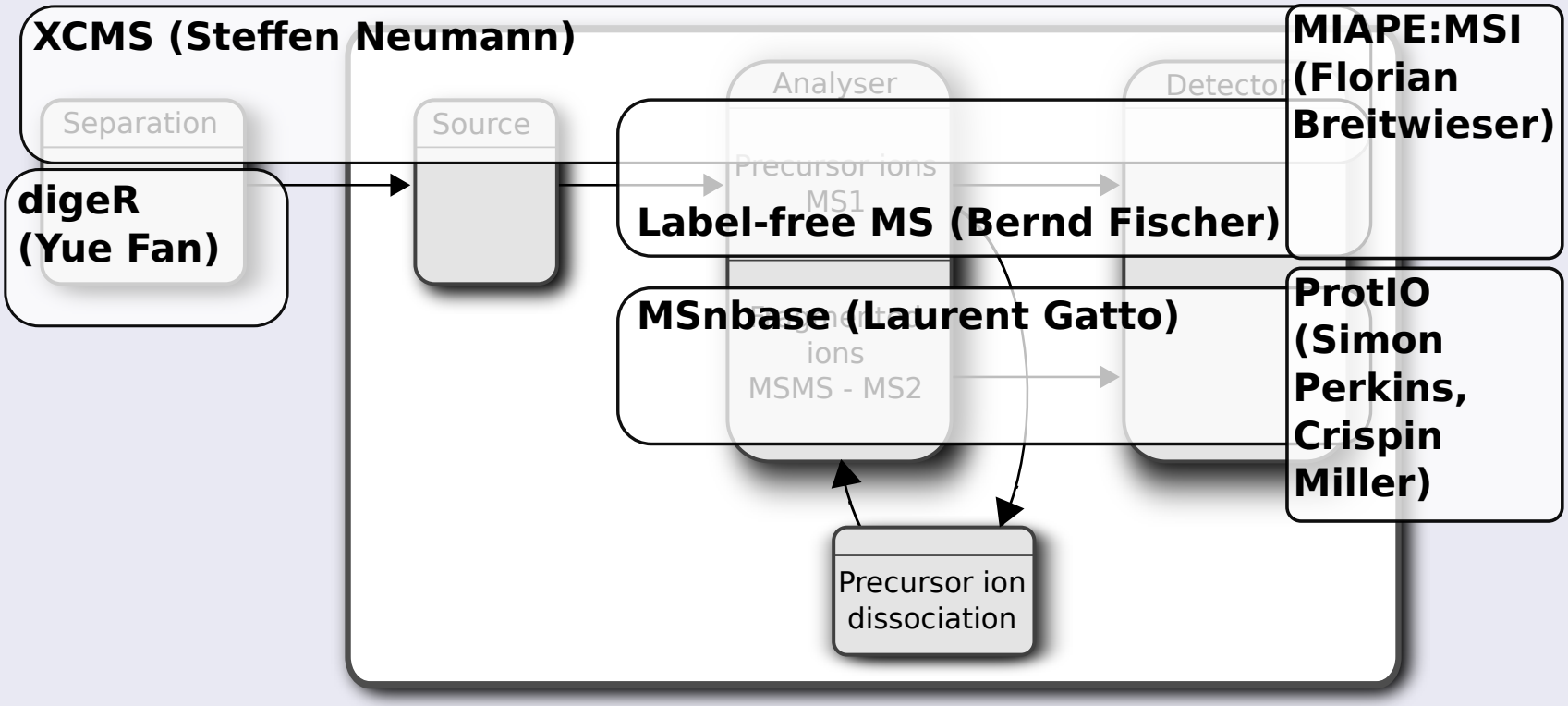


Quantification

- Using MS1 data over retention time (Bernd's talk).
- In MS2 using **spectral counting**, assessing abundance based on **protein coverage** or using **reporter ions** (see `MSnbase`).

Plan

R/Bioc packages



Plan

Applications

- Identify peptides/proteins/metabolites.
- Relative and absolute quantification (`xcms`, `MSnbase` and `digeR` talks).
- Post-translational modifications (PTM).
- Interaction partners (PPI, P-metabolite)
- Sub-cellular localisation of proteins/peptides (see `pRoloc` talk).

Some challenges

- Dynamic range (6 orders of magnitude in human cells and >10 in serum. . . and no PCR)
- Chemical complexity (membrane vs. soluble proteins)
- Only some, and preferentially most abundant once, analytes are (1) randomly sampled for MS1 and (2) selected for MS2 $\rightarrow$ **missing data.**
- Identification is entirely dependent on the quality of the annotation at hand.
- Peptide surrogacy: non-unique peptides, how many peptides per protein, mapping multiple modifications to proteins.

Session progamme

- Metabolomics using `xcms` (Steffen Neumann)
- Label-free differential quantification for proteomics (Bernd Fischer)
- MSMS data with `MSnbase` (Laurent Gatto)
- Protein localisation with `pRoloc` (Laurent Gatto)
- Interfacing proteomics data and R/Bioc with `ProtIO` (Crispin Miller and Simon Perkins)
- DIGE gels with `digeR` (Yue Fan)
- MIAPE:MSI and pep. $\rightarrow$ prot. (Florian Breitwieser)
- Discussion