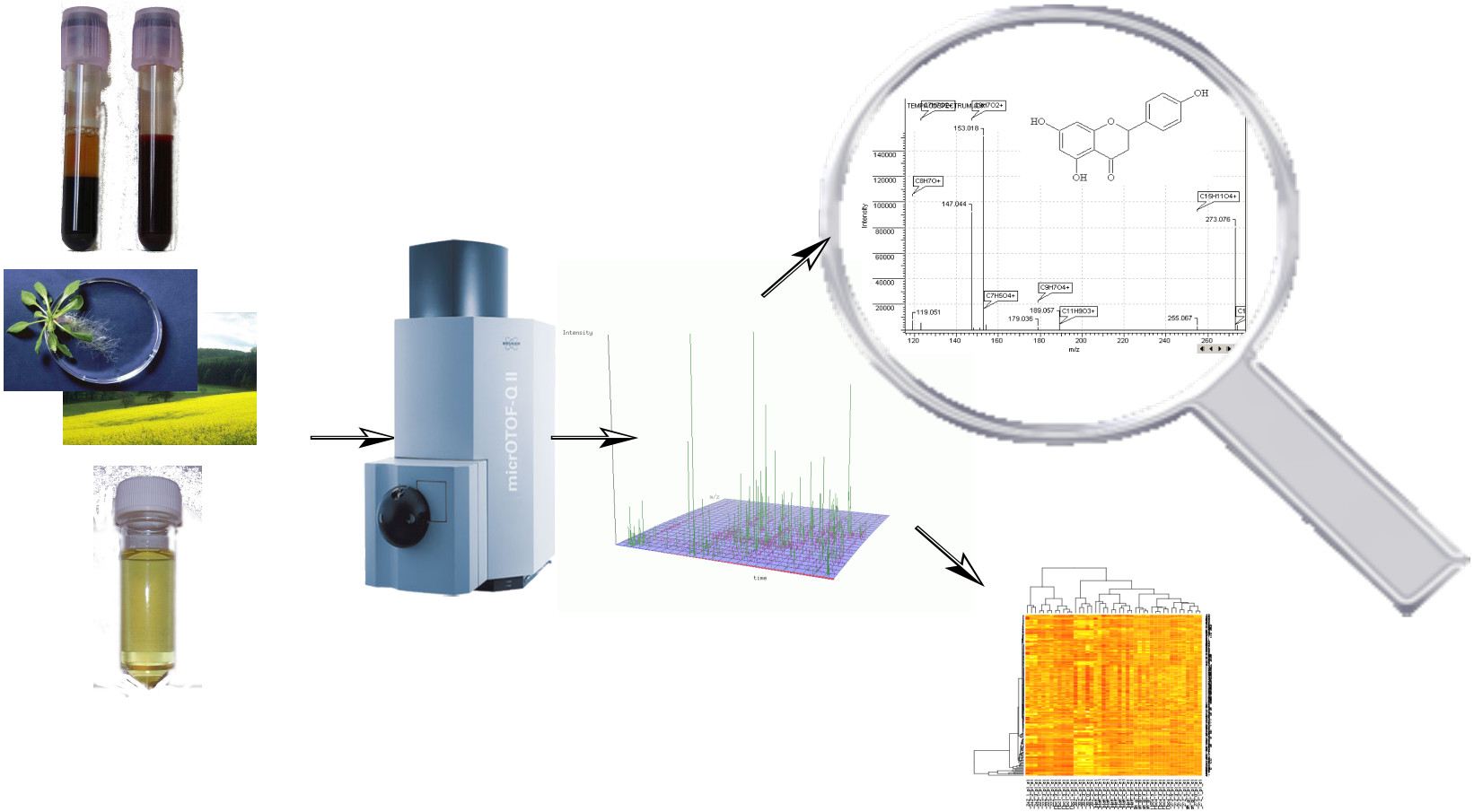
RmassBank Decelopments: XCMS+CAMERA=MetShot

Steffen Neumann, Erik Müller

Leibniz Institute of Plant Biochemistry

ThOB 13.6.2013

Metabolomics – The Pipeline



Outline

Outline

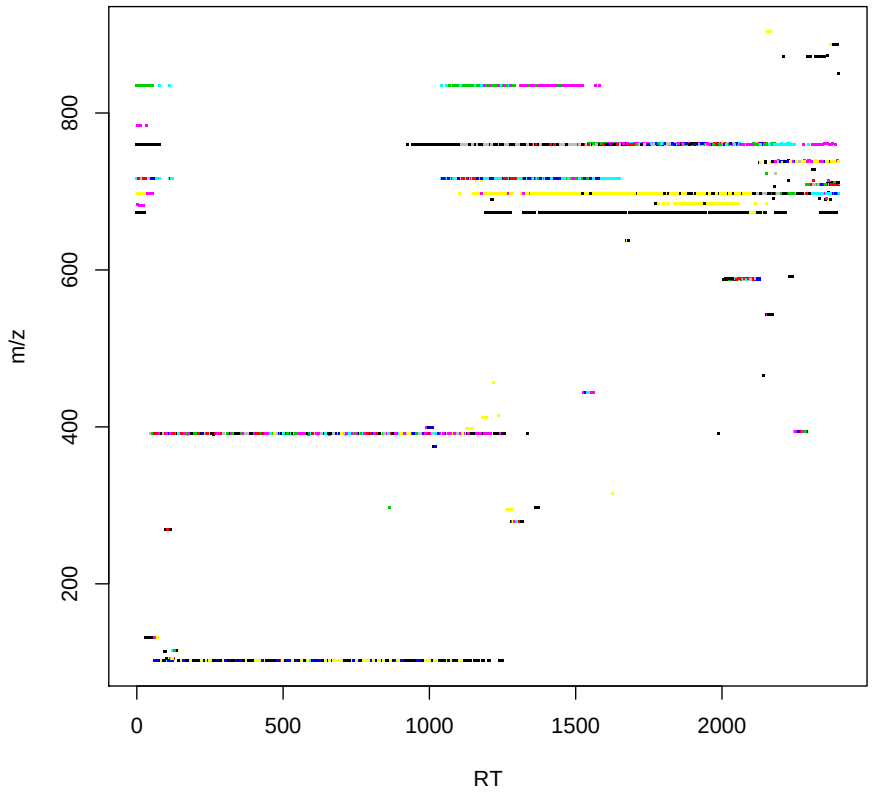
Instrument Data-dependent MS/MS

Instrument Auto-MS/MS

- Full scan first
- Then: MS/MS of most intense peak(s)
- **Online**

Problems:

- Boring peaks: plasticiser, solvent, . . .
- Molecular ion / Biologically relevant peaks may not be the most intense
- Noise spectra / Too few / too many peaks



precursor selection
in 8 LC-MS/MS runs

DI

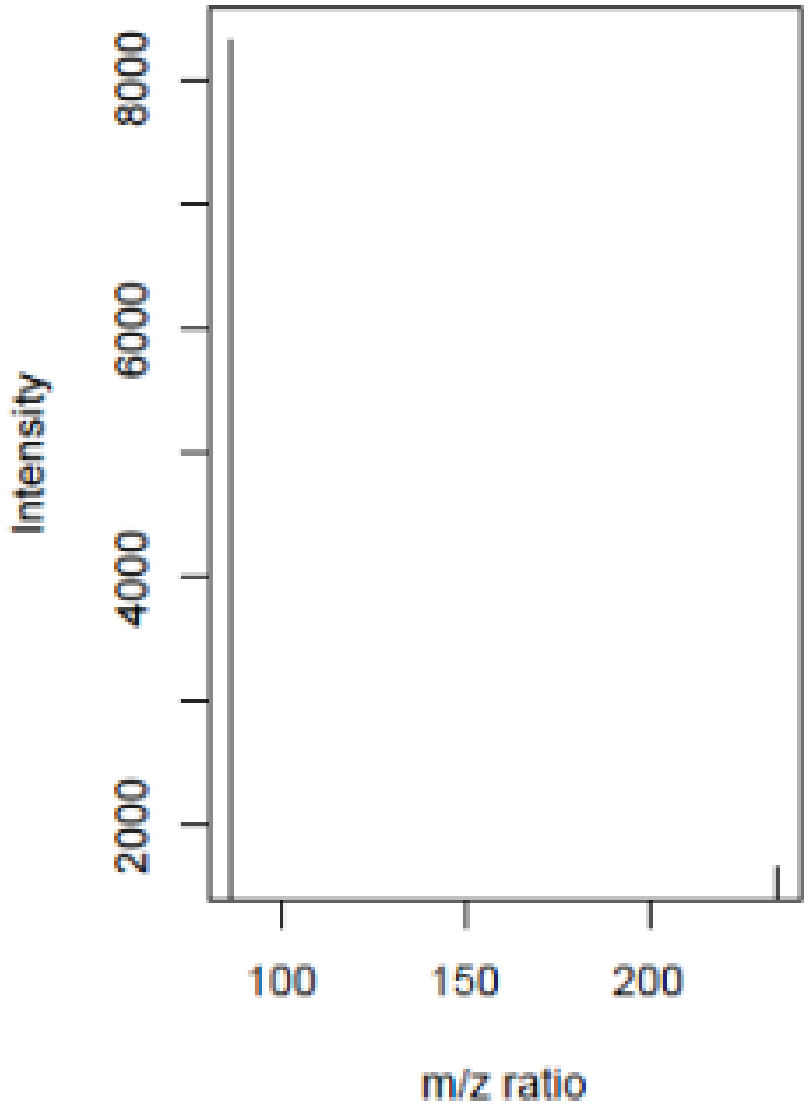
Instrument Data-dependent MS/MS

Instrument Auto-MS/MS

- Full scan first
- Then: MS/MS of most intense peak(s)
- **Online**

Problems:

- Boring peaks: plasticiser, solvent, . . .
- Molecular ion / Biologically relevant peaks may not be the most intense
- Noise spectra / Too few / too many peaks



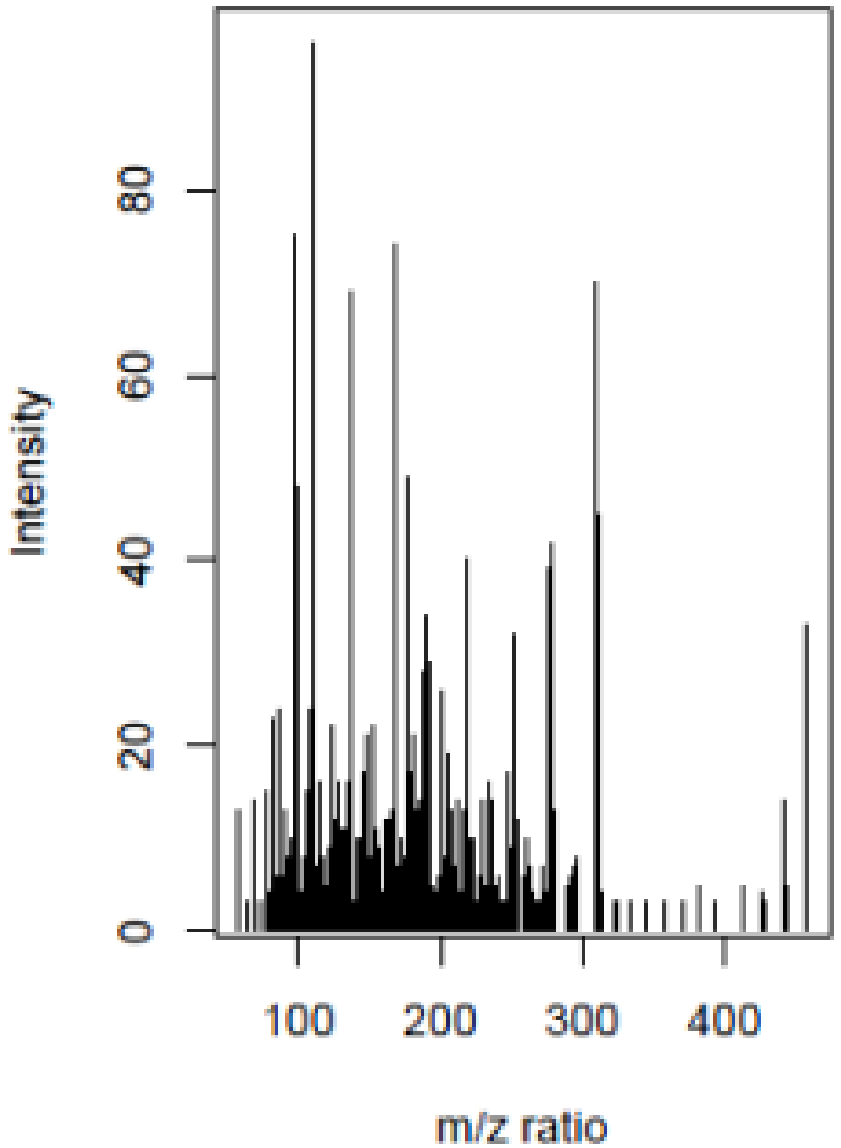
Instrument Data-dependent MS/MS

Instrument Auto-MS/MS

- Full scan first
- Then: MS/MS of most intense peak(s)
- **Online**

Problems:

- Boring peaks: plasticiser, solvent, . . .
- Molecular ion / Biologically relevant peaks may not be the most intense
- Noise spectra / Too few / too many peaks



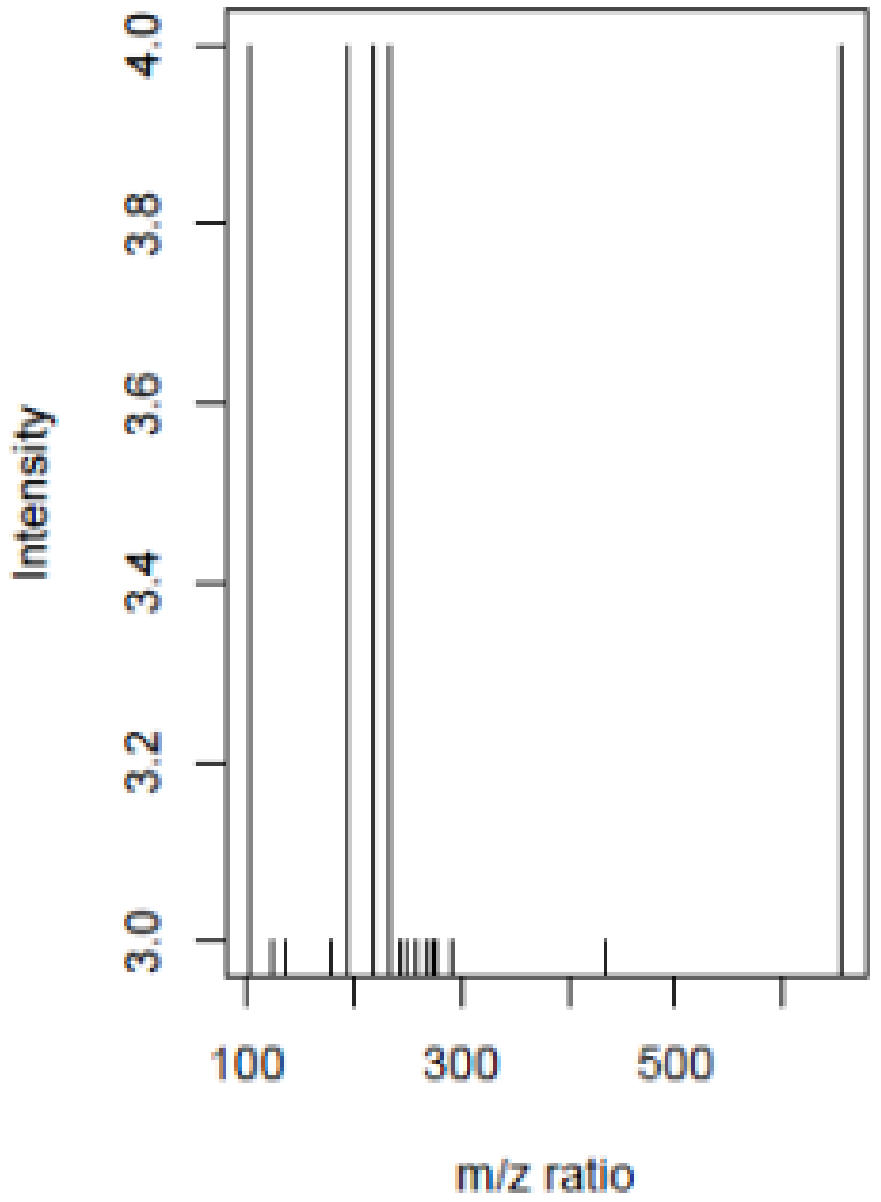
Instrument Data-dependent MS/MS

Instrument Auto-MS/MS

- Full scan first
- Then: MS/MS of most intense peak(s)
- **Online**

Problems:

- Boring peaks: plasticiser, solvent, . . .
- Molecular ion / Biologically relevant peaks may not be the most intense
- Noise spectra / Too few / too many peaks



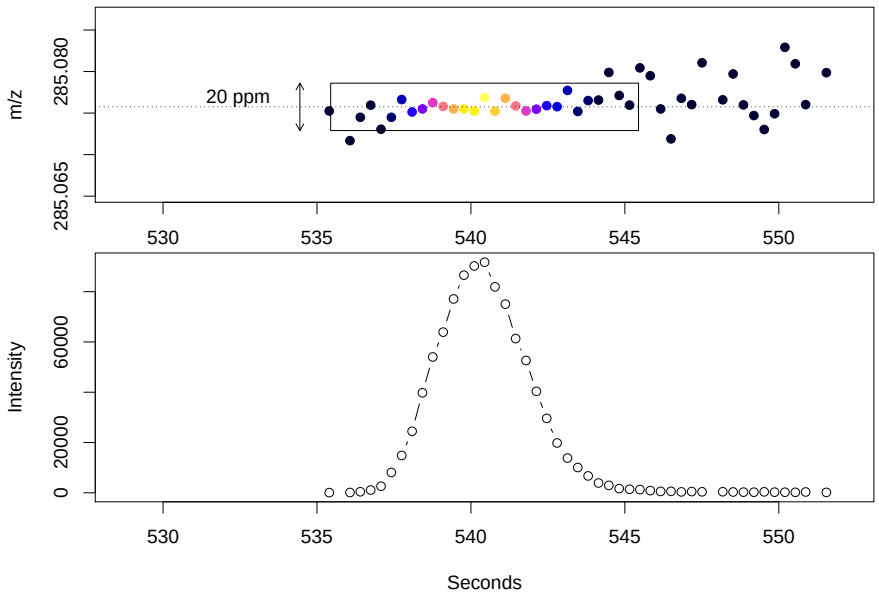
Single spectrum extraction

m/z accuracy:

- Low intensity signals
  $\rightarrow$ low m/z accuracy
- Spectra merging
  $\rightarrow$ improves m/z accuracy

Co-elution / co-isolation:

- MS/MS spectrum contains fragment peaks from more than one metabolite
- Poor identification



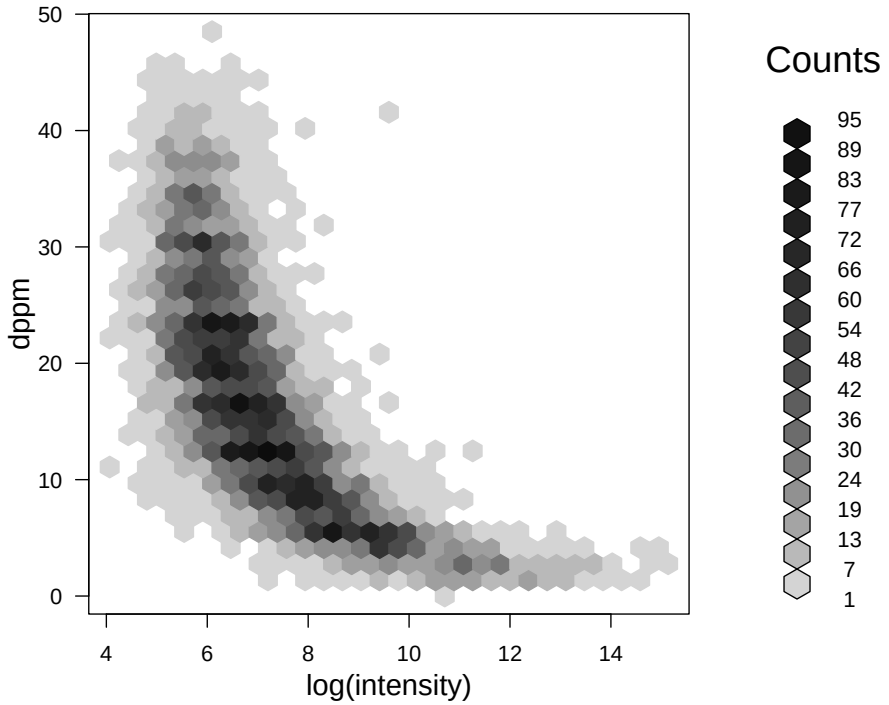
Single spectrum extraction

m/z accuracy:

- Low intensity signals
  $\rightarrow$ low m/z accuracy
- Spectra merging
  $\rightarrow$ improves m/z accuracy

Co-elution / co-isolation:

- MS/MS spectrum contains fragment peaks from more than one metabolite
- Poor identification



Acquisition of LC-MS/MS

Manual LC-MS/MS acquisition:

- Choose m/z – RT window(s)
  - → build instrument method(s)
- 2nd run/injection for MS/MS
- Manual extraction of MS/MS spectrum
- Controlled quality data
- **Offline**, low throughput

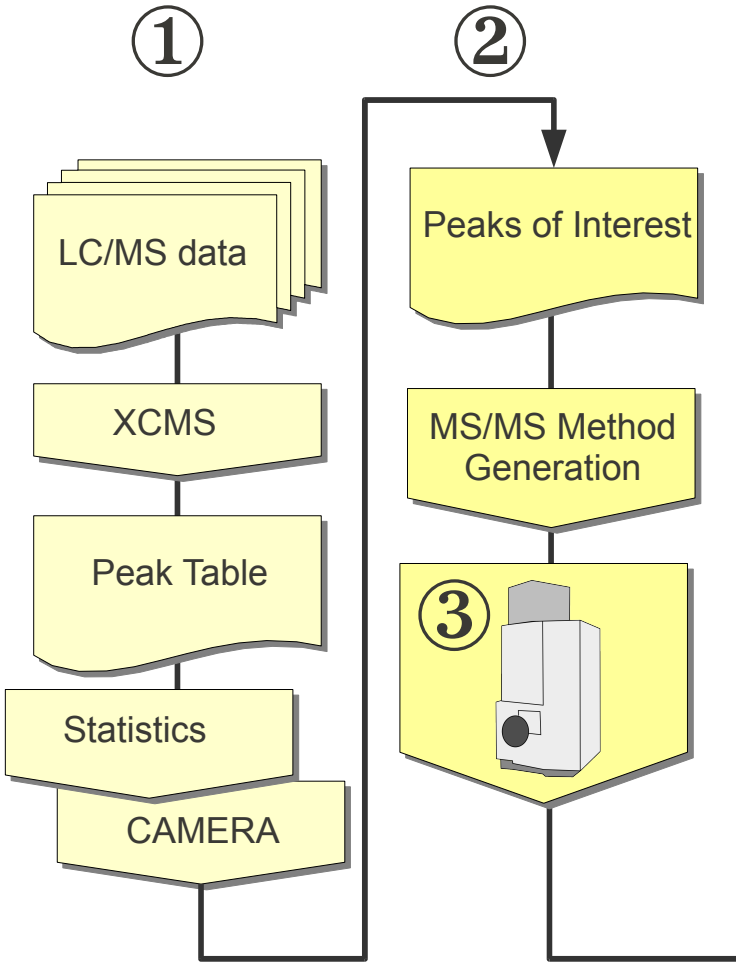
Data-dependent MS/MS:

- Selects most intense peak(s)
- Low quality spectra:
  - "wrong" precursor(s)
  - m/z accuracy
  - effective MS1 scan rate reduced
- **Online**, little control

Choice of two approaches with different drawbacks

MetShot is a hybrid approach, where the instrument methods are created automatically after the profiling and statistics

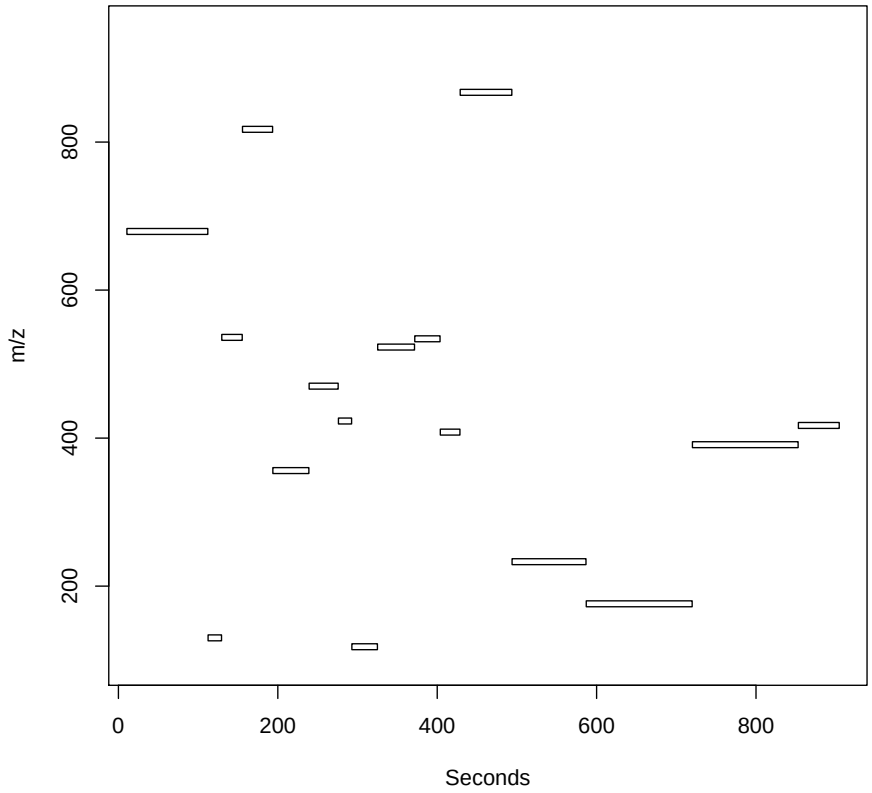
MetShot: *nearline* data-dependent tandem MS



Steffen Neumann, Andrea Thum and Christoph Böttcher: Nearline acquisition and processing of liquid chromatography-tandem mass spectrometry data. Metabolomics (2012)

Scheduling of MS/MS spectra

- "Features of Interest" list from LC-MS profiling
- Optimal packing of LC-MS/MS:
  - most important first
  - without RT overlap
  - method files for Bruker, Waters, Agilent, TraML XML
- Ordered by p-Value (biological *relevance*)

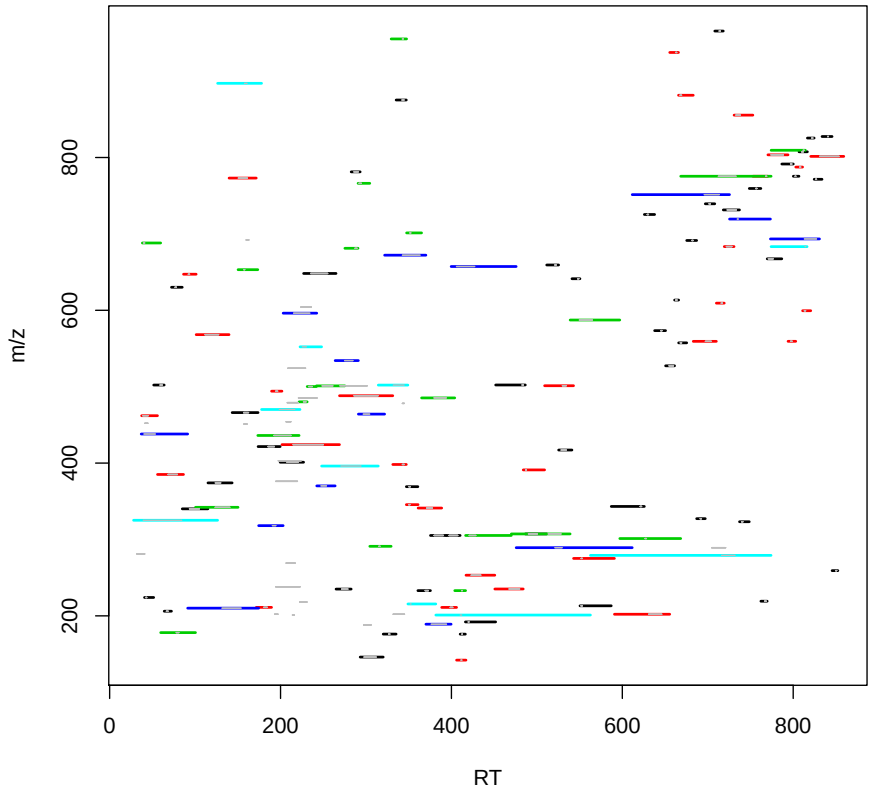


⇒ Only few LC-MS/MS runs to cover many peaks
R package: `http://msbi.ipb-halle.de/msbi/MetShot`

Steffen Neumann, Andrea Thum and Christoph Böttcher: Nearline acquisition and processing of liquid chromatography-tandem mass spectrometry data. Metabolomics (2012)

Scheduling of MS/MS spectra

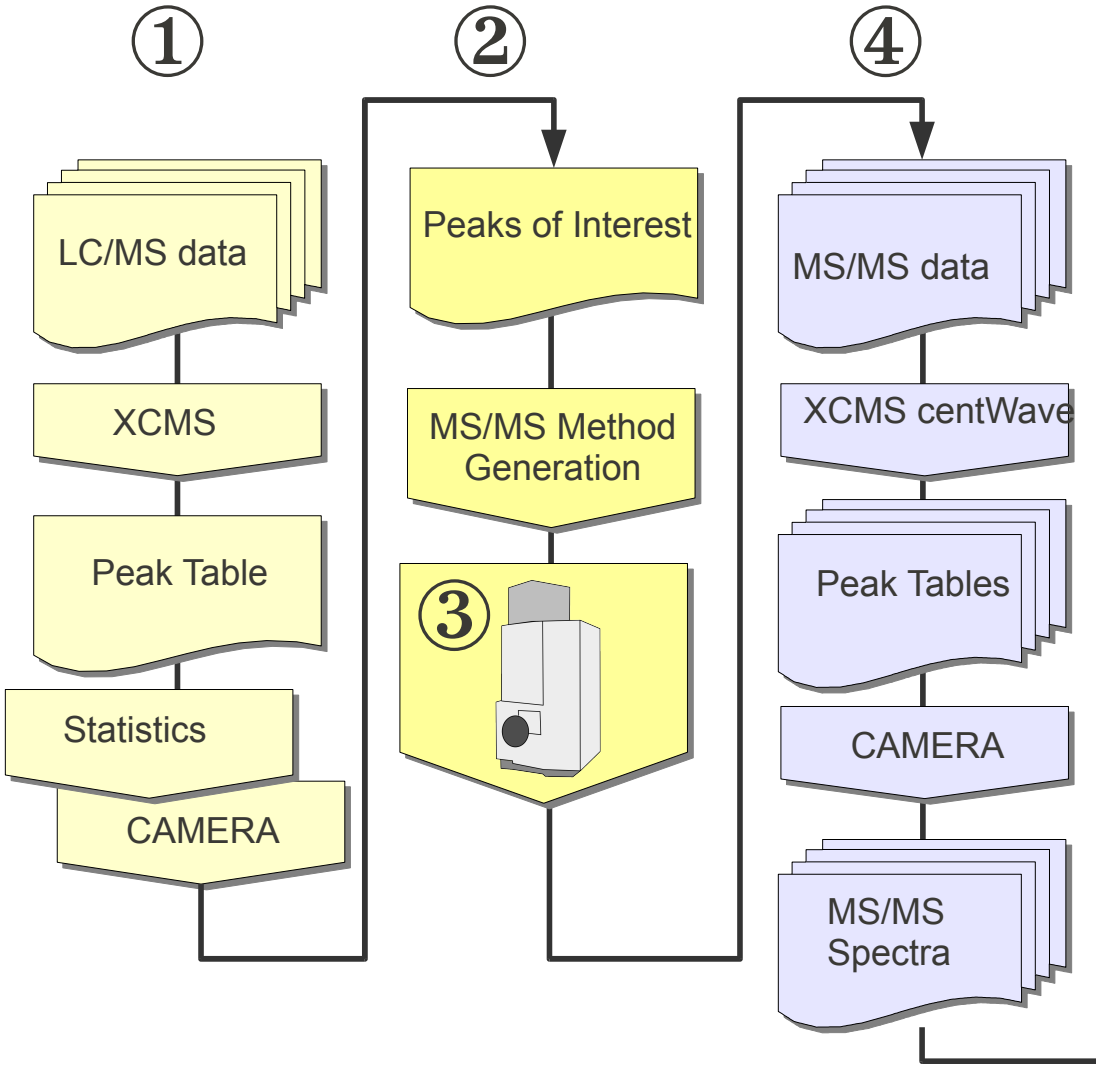
- "Features of Interest" list from LC-MS profiling
- Optimal packing of LC-MS/MS:
  - most important first
  - without RT overlap
  - method files for Bruker, Waters, Agilent, TraML XML
- Ordered by p-Value (biological *relevance*)



⇒ Only few LC-MS/MS runs to cover many peaks
R package: `http://msbi.ipb-halle.de/msbi/MetShot`

Steffen Neumann, Andrea Thum and Christoph Böttcher: Nearline acquisition and processing of liquid chromatography-tandem mass spectrometry data. Metabolomics (2012)

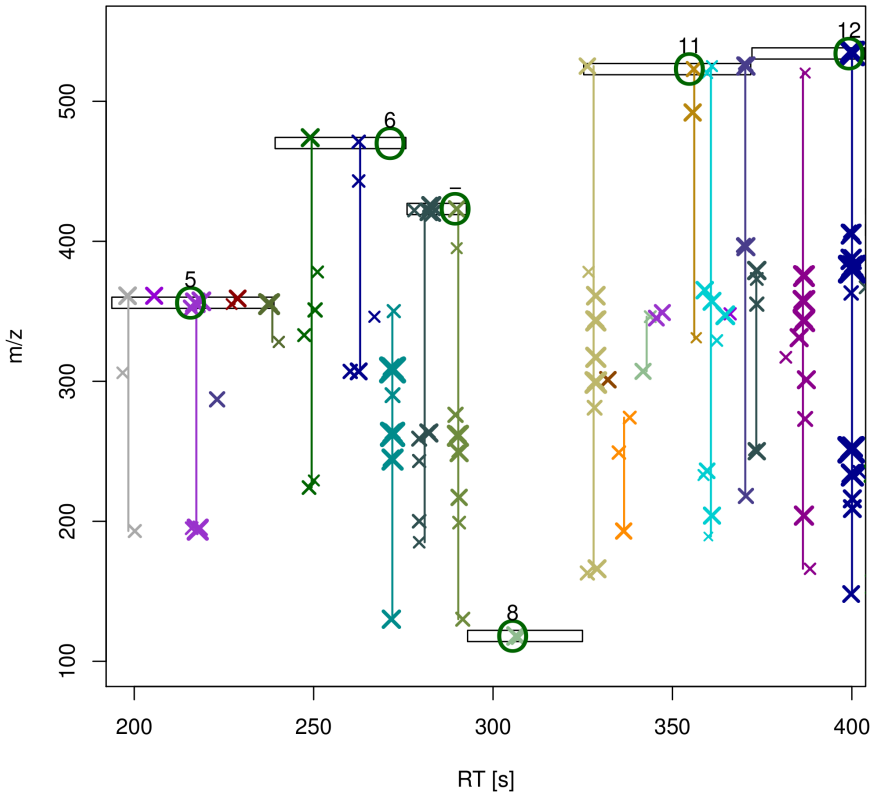
MetShot: *nearline* data-dependent tandem MS



Steffen Neumann, Andrea Thum and Christoph Böttcher: Nearline acquisition and processing of liquid chromatography-tandem mass spectrometry data. Metabolomics (2012)

Collect and Assemble LC-MS and LC-MS/MS

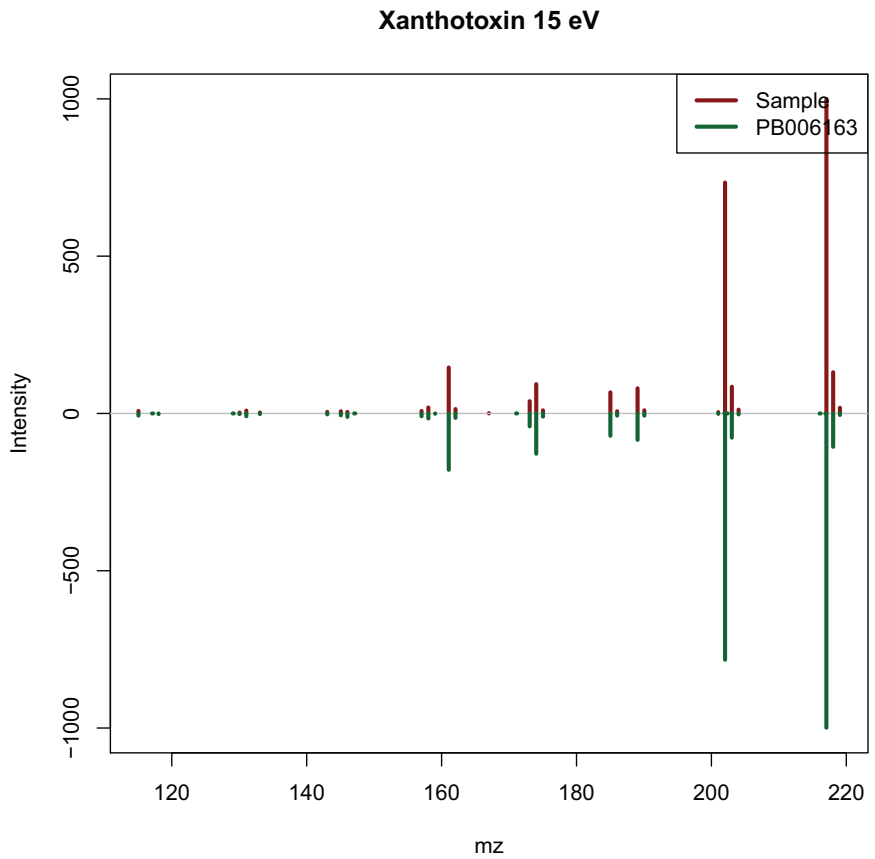
- Feature detection with XCMS in LC-MS/MS raw data
- Collect compound spectra with CAMERA
- Assign spectra to nearest precursor peak of interest



Steffen Neumann, Andrea Thum and Christoph Böttcher: Nearline acquisition and processing of liquid chromatography-tandem mass spectrometry data. Metabolomics (2012)

Evaluation of extracted spectra

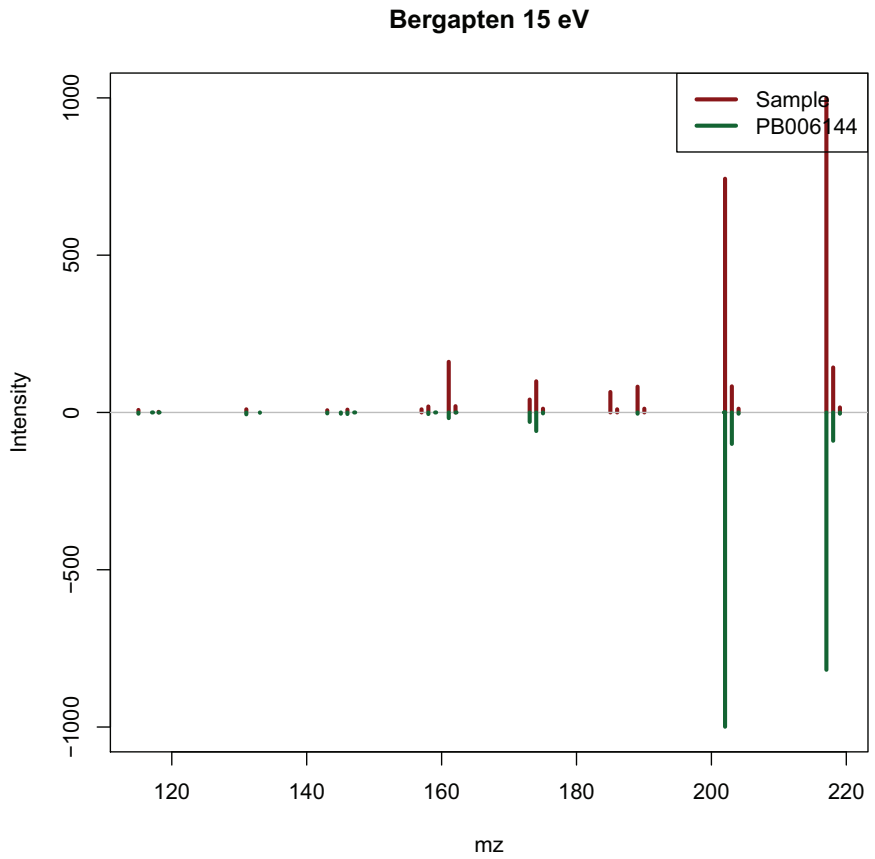
- Mixture of 27 standards
- Manual spectra extraction with vendor software, available in MassBank
- MassBank spectral score automatic vs. manual
- Score >0.95 for 25 of 27



Steffen Neumann, Andrea Thum and Christoph Böttcher: Nearline acquisition and processing of liquid chromatography-tandem mass spectrometry data. Metabolomics (2012)

Evaluation of extracted spectra

- Mixture of 27 standards
- Manual spectra extraction with vendor software, available in MassBank
- MassBank spectral score automatic vs. manual
- Score >0.95 for 25 of 27



Steffen Neumann, Andrea Thum and Christoph Böttcher: Nearline acquisition and processing of liquid chromatography-tandem mass spectrometry data. Metabolomics (2012)

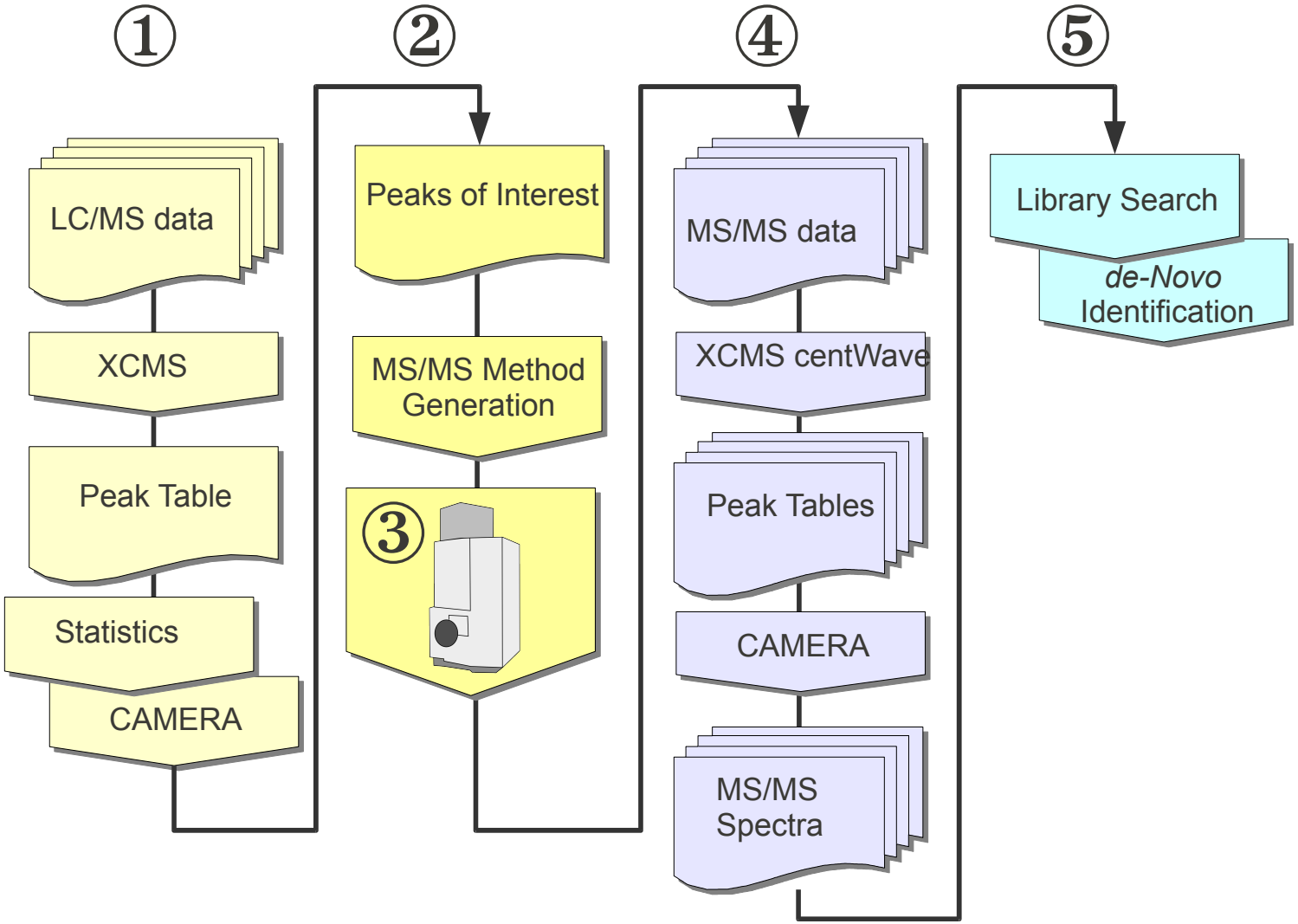
Evaluation of extracted spectra

- Mixture of 27 standards
- Manual spectra extraction with vendor software, available in MassBank
- MassBank spectral score automatic vs. manual
- Score >0.95 for 25 of 27

Compound Name	MassBank	Score
Anisic acid	PB006062	0.9981
Xanthotoxin	PB006163	0.9966
Genistein	PB005713	0.9953
Amentoflavone	PB006304	0.9934
Xanthohumol	PB005728	0.9919
Biochanin A	PB005708	0.9906
Cinchonine	PB005842	0.9902
Kaempferol	PB005703	0.9889
4-Methylumbelliferyl glucoronide	PB006265	0.9888
Bicuculline	PB005942	0.9882
Chlorogenic Acid	PB006181	0.9872
Laudanosine	PB005881	0.9864
(+/-) Salsolinol	PB006041	0.9861
Rotenone	PB005742	0.9849
Naringin	PB005723	0.9832
Indole-3-acetyl-L-valine	PB005789	0.9831
Rutin	PB006202	0.9828
Indole-3-carboxylic acid	PB005786	0.9803
Reserpine	PB005761	0.9789
Hesperidin methyl chalcone	PB006242	0.9771
Chelidonine	PB005902	0.9740
Erythromycin	PB005802	0.9692
3,4,5-Trimethoxycinnamic acid	PB005822	0.9679
Quercetin	PB006205	0.9654
Bergapten	PB006144	0.9527
Indole-3-acetonitrile	PB006045	0.9440
Emetine	PB005910	0.8247

Steffen Neumann, Andrea Thum and Christoph Böttcher: Nearline acquisition and processing of liquid chromatography-tandem mass spectrometry data. Metabolomics (2012)

MetShot: *nearline* data-dependent tandem MS



Extracting MS2 spectra

Other additions: Visualisation

- RMassBank focussed on record generation
- Added summaries for different processing stages
- Added Visualisations

Parser and Record validation

- Parser for MassBank v1 and v2 records
  - Useful for validation, conversion etc.
- Rules for validation:
  - Consistent instrument names
  - Mass of SMILES consistent with isolation window
  - No ions above isolation window
  - Empty peak lists
- Create a HTML report table
- Can be used on MassBank servers in general
  $\Rightarrow$ curation of MassBank data

Developments under the hood

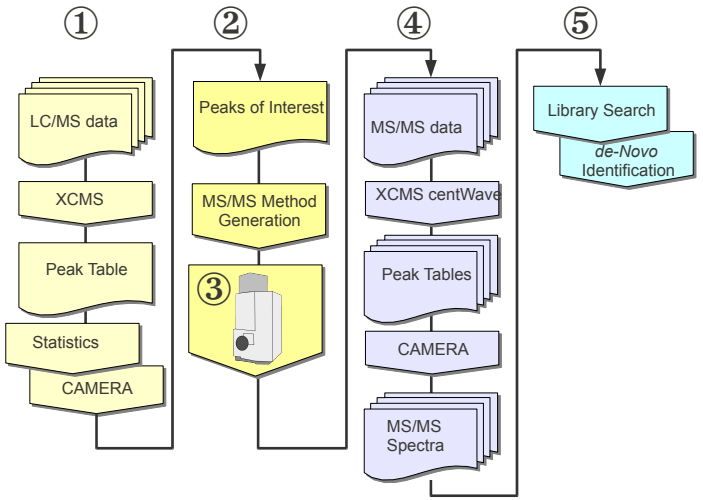
- RMassBank GUI is great for casual use
- RMassBank workflows are great for repeated analysis
- But flexible MS/MS acquisition strategies require flexible data analysis
- Simplification and Modularisation planned

Outline

Summary

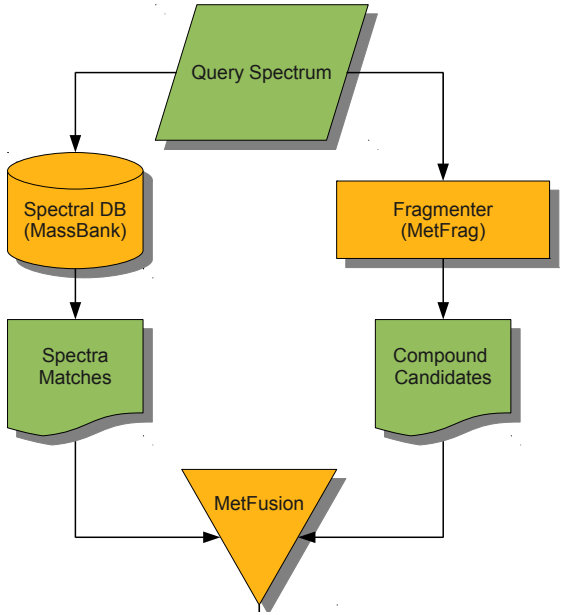
MetShot aproach

- High-quality MS/MS spectra
- Biologically relevant features
- Signal processing with XCMS and CAMERA



Metabolite identification

- With reference spectra (MassBank)
- With *in silico* tools (MetFrag)
- With a combination thereof (MetFusion)



Member of the Leibniz Association

Thanks to . . .

MSBI @ IPB Halle:

- Christoph Ruttkies
- Michael Gerlich
- D. Trutschel, D. Schober

Acknowledgements:

- Profs. Nishioka, Arita and Horai (MassBank Japan)
- Ralf Tautenhahn (SCRIPPS, Metlin Fragments)
- Emma Schymanski (eawag, Zürich)
- Funding: DFG grant “ChemFrag”, EU FP7 “COSMOS”

Member of the Leibniz Association

Only supplemental slides beyond this point