MS2Snoop: an R package for automatic extraction and curation of MS/MS spectra in metabolomics

Kevin Wagner, Lain Pavot, Franck Giacomoni, Guillaume Marti, Yann Guitton, Laurent Debrauwer, Emilien L. Jamin, Jean-François Martin

MS2Snoop: a R package for automatic extraction and curation of MS/MS spectra in metabolomics

Kévin Wagner[1,2], Lain Pavot[3], Franck Giacomoni[3], Guillaume Marti[2,4], Yann Guitton[5], Laurent Debrauwer[1,2], Emilien Jamin[1,2], Jean-François Martin[1,2]

[1] Toxalim (Research Center in Food Toxicology), Toulouse University, INRAE, ENVT, INP-Purpan, Toulouse, France
[2] MetaboHUB-MetaToul, National Infrastructure of Metabolomics and Fluxomics, Toulouse, France
[3] Université Clermont Auvergne, INRAE, UNH, Plateforme d'Exploration du Métabolisme, MetaboHUB Clermont, Clermont-Ferrand, France
[4] Metatoul-AgromiX Platform, LRSV, Université de Toulouse, CNRS, UT3, INP, Toulouse, France
[5] Oniris, INRAE, LABERCA, Nantes, France

Introduction

Omics approaches including metabolomics generate huge datasets requiring powerful processing tools. In the field of non-targeted mass spectrometry metabolomics, HRMS data mining has been the subject of much work, with the development of powerful interactive tools, including Workflow4Metabolomics (W4M) [1-2]. One of the main bottlenecks of current workflows lies in the complexity of the interpretation of results, in particular a lack of reliable MS/MS spectral reference databases (DBs) to be used for confident identification (level 1 and 2 according to the metabolomics standard initiative). Creating reliable and well curated spectral DBs remains a laborious and challenging task, but necessary to ensure the quality of the reference spectra. In this work, we present a tool inspired by expert manual curation allowing the export of MS/MS data from both targeted and data dependent (DDA) fragmentation methods to laboratory spectral DBs, named "MS2Snoop".

MS2snoop

- **Workflow (R):**

XCMS

Peak picking

- From mzML file(s) containing MS & MS/MS
- Use of xcmsSet & group functions of XCMS
- Generation of an « xcmsSet » object (.tsv)

msPurity

Precursors quality

- From same mzML file(s)
- Use of purityA function to evaluate the purity of precursors [3]
- Generation of a « purityA » object (.tsv)

frag4feature

Datasets merging

- Association of fragment ions with precursor by merging « purityA » and « xcmsSet » objects
- Generation of .tsv files for fragment & precursor ions

FilterFrag

Fragment ions Filtration

- Definition of detection level of fragment ions (intensity, S/N ratio)
- Generation of .tsv files for fragment & precursor ions

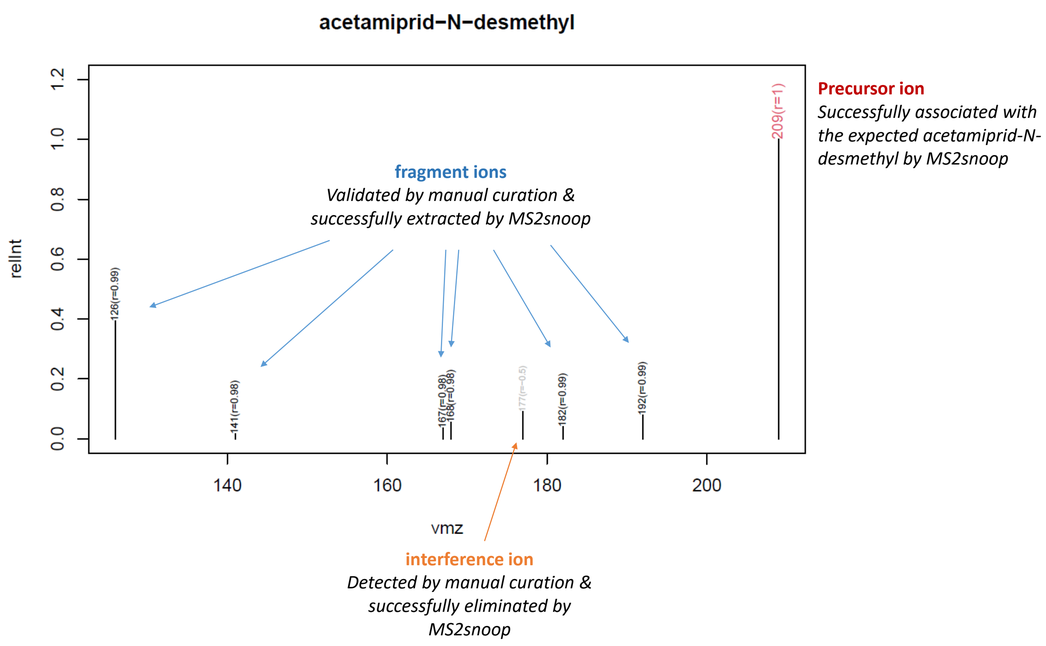
MS2snoop

MS/MS spectra reconstruction

- Merging of fragment & precursor ions with list of expected compounds
- Correlation of fragment intensity with precursor intensity

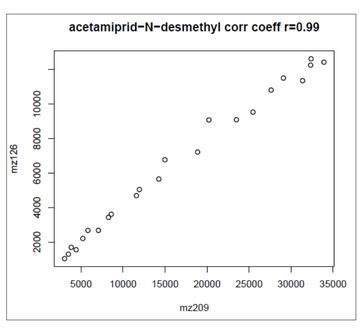
- **Examples of applications**

Targeted MS/MS by resonant CID achieved on a LTQ-Orbitrap XL using ESI in the positive mode

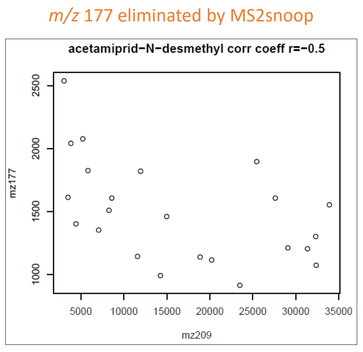


Curation by correlation between precursor ion intensity & fragment ion intensities

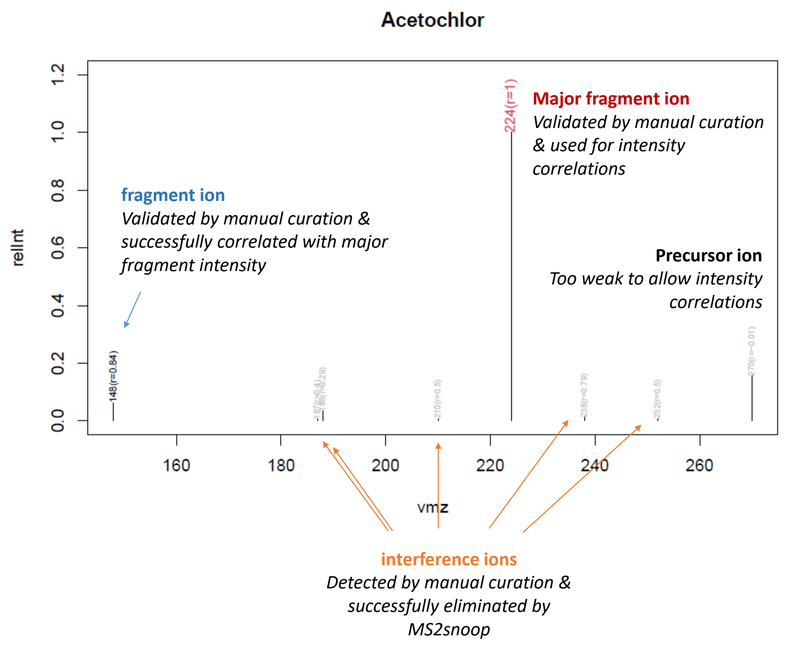
m/z 126 kept by MS2snoop



m/z 177 eliminated by MS2snoop



Successful application to the particular case of acetochlor with a very weak precursor ion intensity



Conclusions

MS2Snoop already allows extracting and cleaning MS/MS datasets using commonly used metabolomics data processing steps as a starting point, then combined with inputs of a list of expected compounds and MS/MS raw data deconvoluted by MSPurity. Firstly developed and validated with targeted MS/MS spectra from about 30 compounds, MS2Snoop is compatible with DDA spectra, although limited by the number of MS/MS scans acquired. This package will allow high throughput extraction of MS/MS spectra in order to feed the PeakForest project [4]. Preliminary results obtained with MS2Snoop allowed considering it as a valuable tool for confident annotation of metabolites on chromatographic methods coupled to non-targeted mass spectrometry.

A beta version of this original package is already publicly available on W4M (https://workflow4metabolomics.org/) and recent update allows fragments ions formula annotation with Sirius.

References

[1] Giacomoni F. et al. Bioinformatics, (2015) 31, 1493
[2] Guitton Y. et al, Int. J. Biochem. Cell Biol., (2017) 93, 89
[3] Lawson T. N. et al., Anal. Chem., (2017) 89, 1432
[4] Paulhe N. et al., Metabolomics, (2022)18, 40

Poster presented at Analytics2022, Nantes, France

Acknowledgments

HRMS analyses were achieved in the frame of the French national infrastructure MetaboHUB (MetaboHUB-ANR-11-INBS-0010). Special thanks to JF Martin for his long career and the good times !