Technological developments and opportunities with Workflow4Metabolomics

Mélanie Pétéra, Yann Guitton, Charlotte Joly, Florence Souard, Etienne A Thévenot, Céline Dalle, Ralf J.M. Weber, Sylvain Chéreau, Binta Diémé, Cédric Delporte, et al.

Technological developments and opportunities with Workflow4Metabolomics

Mélanie PÉTÉRA[1], Yann GUITTON[2], Charlotte JOLY[1], Florence SOUARD[3], Etienne THEVENOT[4], Céline DALLE[5], Ralf J. M. WEBER[6,7], Sylvain CHÉREAU, Binta DIÉMÉ[1], Cédric DELPORTE[8], Gildas LE CORGUILLÉ[9], Helge HECHT[10], W4M Core Team[1,2,3,4,5,6,7,8,9,11,12,13,14]

[1] *Université Clermont Auvergne, INRAE, UNH, Plateforme d'Exploration du Métabolisme, MetaboHUB Clermont, F-63000 Clermont-Ferrand, France*
[2] *Oniris, INRAE, LABERCA, 44300, Nantes, France*
[3] *DPP Department - Unit of Pharmacology, Pharmacotherapy and Pharmaceutical care, Faculty of Pharmacy, Université libre de Bruxelles, Brussels, Belgium,*
[4] *CEA, INRAE, Département Médicaments et Technologies pour la Santé (DMTS), MetaboHUB, Université Paris-Saclay, 91191 Gif-sur-Yvette, France*
[5] *French Armed Forces Biomedical Research Institute (IRBA), Analytical Development and Bioanalysis Unit, F-91220 Brétigny-sur-Orge, France*
[6] *School of Biosciences, University of Birmingham, Birmingham, B15 2TT, UK*
[7] *Phenome Centre Birmingham, University of Birmingham, Birmingham, B15 2TT, UK*
[8] *RD3 Department-Unit of Pharmacognosy, Bioanalysis and Drug Discovery, Faculty of Pharmacy, Université libre de Bruxelles, Brussels, Belgium*
[9] *Sorbonne Université, CNRS, FR2424, ABiMS, Station Biologique, 29680, Roscoff, France*
[10] *RECETOX, Faculty of Science, Masaryk University, Kotlářská 2, Brno, Czech Republic*
[11] *ABiMS, FR2424 CNRS-UPMC, Station Biologique, Place Georges Teissier, 29680, Roscoff, France*
[12] *Toxalim – Research Center in Food Toxicology, Toulouse University, INRAE UMR 1331, ENVT, INP-Purpan, Paul Sabatier University, F-31027 Toulouse, France*
[13] *Metatoul-AXIOM platform, National Infrastructure for Metabolomics and Fluxomics, MetaboHUB, Toxalim, INRAE UMR 1331, F-31027 Toulouse, France*
[14] *INRAE, UMR IGEPP, 35653 Le Rheu, France*

Introduction

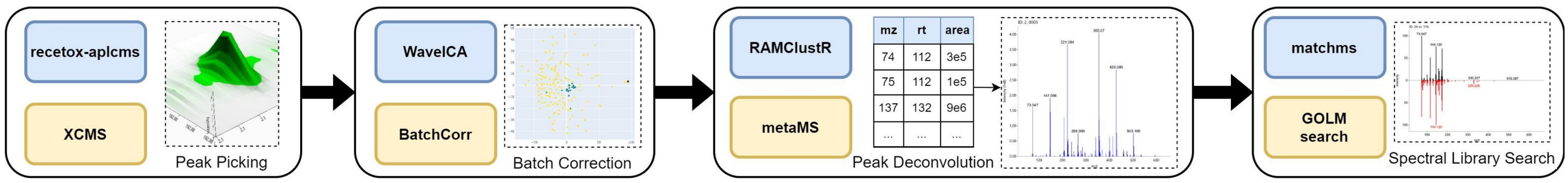
Metabolomics data analysis is a complex and multistep process, which is constantly evolving with the development of new analytical technologies, mathematical methods, bioinformatics tools and databases. By a common effort from several institutional structures as MetaboHUB (French national infrastructure) and the IFB (French ELIXIR node), **Workflow4Metabolomics**[1] (W4M) endeavours to break through the barriers that are obstructing data analysis practices in this field. W4M aims to promote open science in Metabolomics and facilitate knowledge dissemination by providing community resources. In particular, W4M provides Galaxy[2] tools to the community, in addition to the upkeep of a dedicated subdomain at usegalaxy.fr.

New W4M LC-MS, LC-MSMS, GC-MS and NMR pipelines

The W4M core team makes continuous upgrades including pipeline enhancement, tool updates and new component development. The latest version of the W4M infrastructure integrates improvements in several parts of the current tool suite, including statistics (*e.g.* new features for the 'mixmodel: ANOVA for repeated measures statistics' tool) and annotation (*e.g.* a new version of the '2D NMR Annotation' tool for complex mixture bidimensional NMR spectra). New tools were also added to the catalogue, as the 'MS2snoop' module to investigate spectra of standards further using the results of the 'MSPurity[3]' MS tool suite. A suite of tools (BankInHouse for GCMS and LCMS, NMR peak-matching) should be proposed to connect with the PeakForest project (peakforest.org[4]), a multi-platform digital infrastructure for interoperable metabolite spectral data and metadata management. W4M's recent works also cover the development of Galaxy Interactive Tools, for example two tools for dynamic data visualisation and annotation: XSeeker - a modified version of HaloSeeker[5] to visualize and annotate MS spectra; NMRPro[6] - for interactive processing and visualization of NMR spectra.

W4M also recently added to its tool panel a new set of tools for GC-MS data processing (blue nodes depicted beneath) provided by RECETOX, Masaryk University (Czech EIRENE node)[7] to complement the workflow from W4M (yellow nodes). Recent developments focus on interoperabilty and scalability of tools. The newly added tools enable processing of profile mode HRMS data and searching in-house libraries using multiple spectral similarity metrics.

W4M NEWS BREAKING



Open Science - sharing - training: contributing to the community

W4M is involved in the Open science movement. In addition to providing Digital Object Identifiers (DOI) for user data history references, we invest in training materials based on the Galaxy Training Network (**GTN**)[8] concept and format.

Get to grips with Metabolomics data analysis at your own pace using Galaxy Training Materials!

What is the GTN?

Online, open-access resources about data analysis

Interactive learning via hands-on tutorials built around a 'research story'

One website, aggregating training material covering many current research topics

Developed and maintained by the community in a collaborative way using GitHub

GTN materials for Metabolomics

At workflow4metabolomics.usegalaxy.fr, you will be able to test a variety of GTN materials, including the following for Metabolomics:

- Galaxy Training! Mass spectrometry: LC-MS analysis
- Galaxy Training! Mass spectrometry: LC-MS preprocessing with XCMS
- Galaxy Training! Mass spectrometry: LC-MS data processing
- Galaxy Training! Mass spectrometry: GC-MS data processing (with XCMS, RAMClustR, RIAssigner, and matchms)
- Galaxy Training! Nuclear Magnetic Resonance: data preprocessing (coming soon)

Latest materials

W4M contributes to and promotes the use of metabolomics software and open-source software package production.
Contributors share new metabolomics data analysis tools compliant with the galaxy framework on the GitHub **"tools-metabolomics"** repository.

The **W4M** infrastructure is hosted on the **usegalaxy.fr** facility that now provides **t.i.a.s.s** which is a secured space opened on the cluster during your training session

W4M provides one-week-long on-site training once a year through its 'bring-your-own-data' international school **Workflow4Experimenters**.
This year's school took place in Paris in March, 2023. Keep posted for next editions !

Join us on workflow4metabolomics.org!

W4M web portal **W4M pipelines** **Events** **Twitter** **Mailing list** **Support forum**

Acknowledgments This work is supported by the French Ministry of Research and National Research Agency as part of the French metabolomics and fluxomics infrastructure (MetaboHUB-ANR-INBS-0010).

Funding: La Région Auvergne-Rhône-Alpes, Région Bretagne, chopin, France 2030, European Union European Structural and Investment Funds Operational Programme Research, Development and Education

[1] Guitton Y. et al. "Create, run, share, publish, and reference your LC-MS, FIA-MS, GC-MS, and NMR data analysis workflows with the Workflow4Metabolomics 3.0 Galaxy online infrastructure for metabolomics." The international journal of biochemistry & cell biology vol. 93 (2017): 89-101.
[2] Afgan E. et al. "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update." Nucleic acids research vol. 46,W1 (2018): W537-W544. doi:10.1093/nar/gky379
[3] Lawson T. N. et al. "msPurity: Automated Evaluation of Precursor Ion Purity for Mass Spectrometry-Based Fragmentation in Metabolomics." Analytical chemistry vol. 89,4 (2017): 2432-2439. doi:10.1021/acs.analchem.6b04358
[4] Paulhe N. et al. "PeakForest: a multi-platform digital infrastructure for interoperable metabolite spectral data and metadata management." Metabolomics : Official journal of the Metabolomic Society vol. 18,6 40. 14 Jun. 2022, doi:10.1007/s11306-022-01899-3
[5] Léon A. et al. "HaloSeeker 1.0: A User-Friendly Software to Highlight Halogenated Chemicals in Nontargeted High-Resolution Mass Spectrometry Data Sets." Analytical chemistry vol. 91,5 (2019): 3500-3507. doi:10.1021/acs.analchem.8b05103
[6] Mohamed A. et al. "NMRPro: an integrated web component for interactive processing and visualization of NMR spectra." Bioinformatics (Oxford, England) vol. 32,13 (2016): 2067-8. doi:10.1093/bioinformatics/btw102
[7] doi:10.5821/zenodo.8083373
[8] Batut B. et al. "Community-Driven Data Analysis Training for Biology." Cell systems vol. 6,6 (2018): 752-758.e1. doi:10.1016/j.cels.2018.05.012