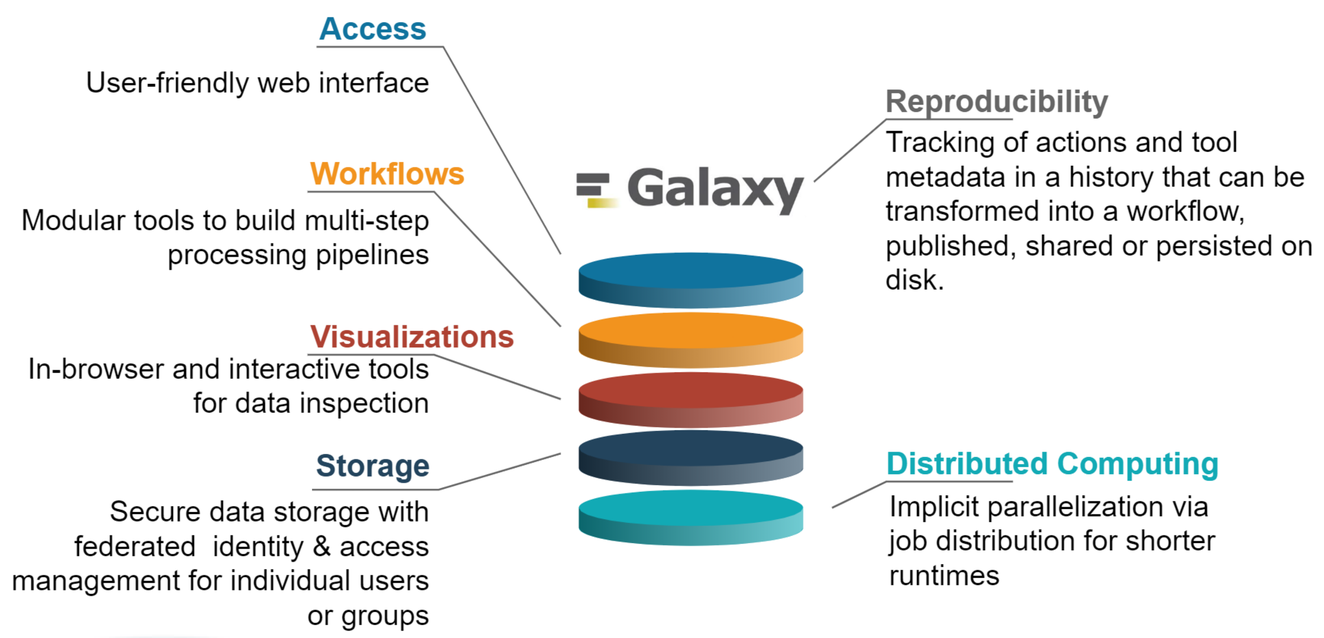
Developing Modular Galaxy Workflows for High-Resolution Mass Spectrometry Exposomics Data Processing

Helge Hecht[1*], Jiří Novotný[2*], Karolína Trachtová[1,2,3], Martin Čech[1,4], Maksym Skoryk[2], Ondřej Melichar[1,2], Aleš Křenek[2+], Jana Klánová [1], Elliott Price [1+]

1 RECETOX Centre, Masaryk University, Brno, Czech Republic; 2 Institute of Computer Science, Masaryk University, Brno, Czech Republic; 3 Central European Institute of Technology, Masaryk University, Brno, Czech Republic; 4 Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Prague, Czech Republic; * Authors contributed equally; + elliott.james.price@gmail.com / ljocha@ics.muni.cz

Introduction

Galaxy is an **open-source web based** platform making fully reproducible bioinformatic analysis simpler for scientists [1, 2]. Current **Galaxy tools** to process mass spectrometry of small molecules data are limited [3, 4]. An **Exposome toolbox** for Galaxy is being developed to ease **analysis of high-resolution mass spectrometry (HRMS)** datasets, tailored towards measures of environmentally-derived chemicals.



Methods

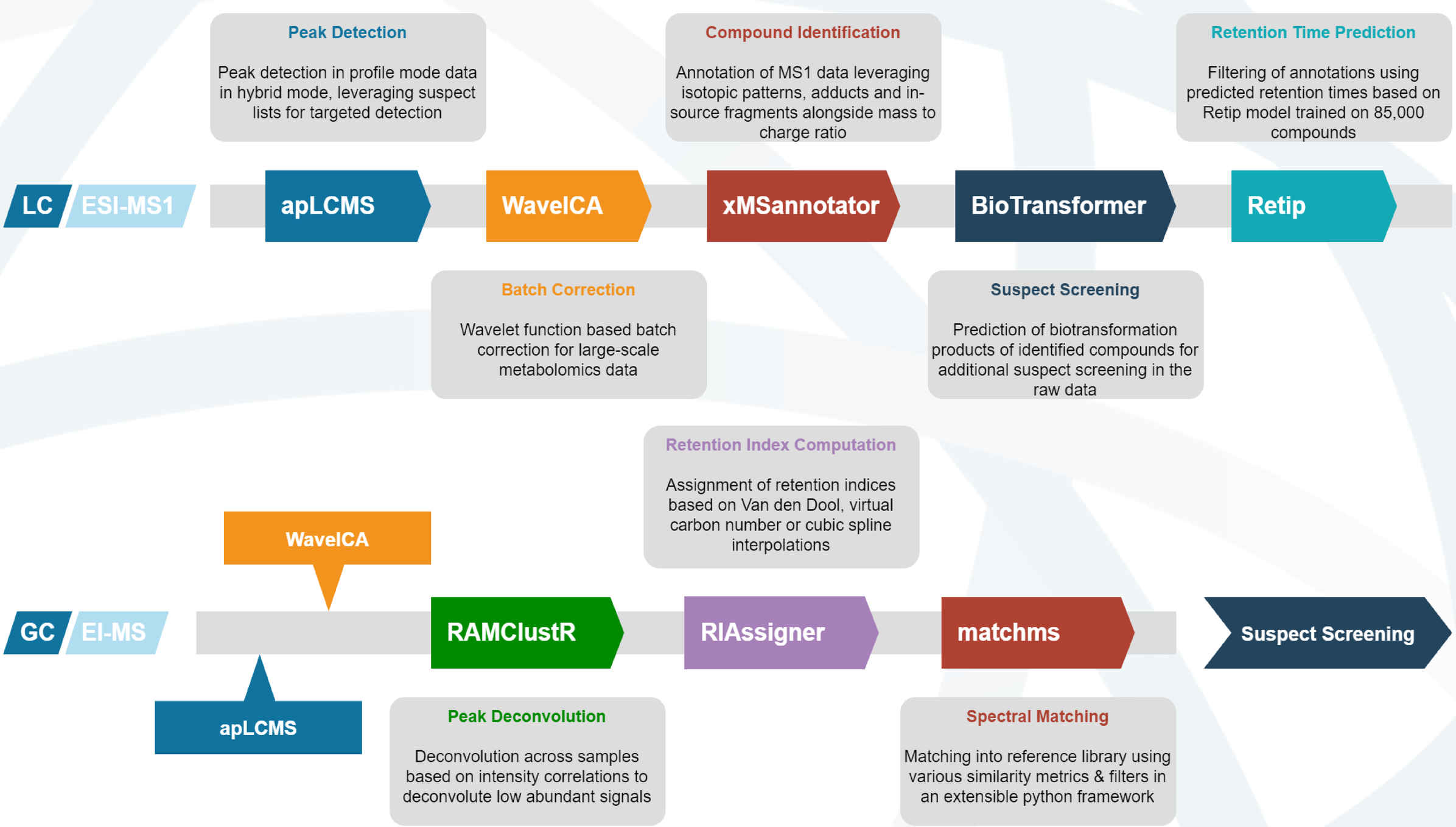
Pipelines were implemented as **modular workflows in Galaxy**. Tools were made available as **Docker containers** or **conda packages** following Galaxy standards and ensuring reproducibility [5]. **Code quality** and maintainability was **improved** with packages partially re-written, **versioned** and **committed to repositories** [6]. Reporting follows the mzTab-M standard [7].

Results

We expanded the available Galaxy resources by several tools and **two novel workflows** for HRMS data analysis. All modifications were carried out following **open-source** and **FAIR data principle**s, in consensus with the original authors. Packages are available on **https://github.com/RECETOX/galaxytools** and tools hosted at **https://umsa.cerit-sc.cz/**.

Future Work

The selected tools form the foundation for future additions and are **complementary to existing Galaxy resources** (e.g Galaxy Europe and Workflow4Metabolomics [8]).
We welcome **contributions** and **collaboration** for ongoing and future developments.



Acknowledgements

1. RECETOX research infrastructure (the Czech Ministry of Education, Youth and Sports: LM2018121)
2. Computational resources were provided by the project e-INFRA LM2018140
3. CETOCOEN EXCELLENCE Teaming 2 project supported by Horizon2020 (857560) and the Czech ministry of Education, Youth and Sports (02.1.01/0.0/0.0/18_046/0015975).

References

1. Afgan, E. et al. Galaxy: A Gateway to Tools in e-Science. in Guide to e-Science (eds. Yang, X., Wang, L. & Jie, W.) 145–177 (Springer, 2011). doi:10.1007/978-0-85729-439-5_6.
2. 2. Jalili, V. et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. Nucleic Acids Res. 48, W395–W402 (2020). doi: 10.1093/nar/gkaa434
3. Weber, R. J. M. et al. Computational tools and workflows in metabolomics: An international survey highlights the opportunity for harmonisation through Galaxy. Metabolomics 13, 12 (2017). doi: 10.1007/s11306-016-1147-x
4. David, A. et al. Towards a comprehensive characterisation of the human internal chemical exposome: Challenges and perspectives. Environ. Int. 156, 106630 (2021). doi: 10.1016/j.envint.2021.106630
5. Grüning, B. et al. Practical Computational Reproducibility in the Life Sciences. Cell Syst. 6, 631–635 (2018). doi: 10.1016/j.cels.2018.03.014
6. Chang, H.-Y. et al. A Practical Guide to Metabolomics Software Development. Anal. Chem. 93, 1912–1923 (2021). doi: 10.1021/acs.analchem.0c03581
7. Hoffmann, N. et al. mzTab-M: A Data Standard for Sharing Quantitative Results in Mass Spectrometry Metabolomics. Anal. Chem. 91, 3302–3310 (2019). doi: 10.1021/acs.analchem.8b04310
8. 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. Int. J. Biochem. Cell Biol. 93, 89–101 (2017). doi: 10.1016/j.biocel.2017.07.002