workflow4metabolomics.org

Galaxy and the metabolomics analysis Universe

Misharl MONSOOR [1], Gildas LE CORGUILLE [1], Marion LANDI [2], Mélanie PETERA [2], Pierre PERICARD [1], Christophe DUPERIER [2], Marie TREMBLAY-FRANCO [3], Jean-François MARTIN [3], Sophie GOULITQUER [1], Etienne THEVENOT [4], Franck GIACOMONI [2], Christophe CARON [1]

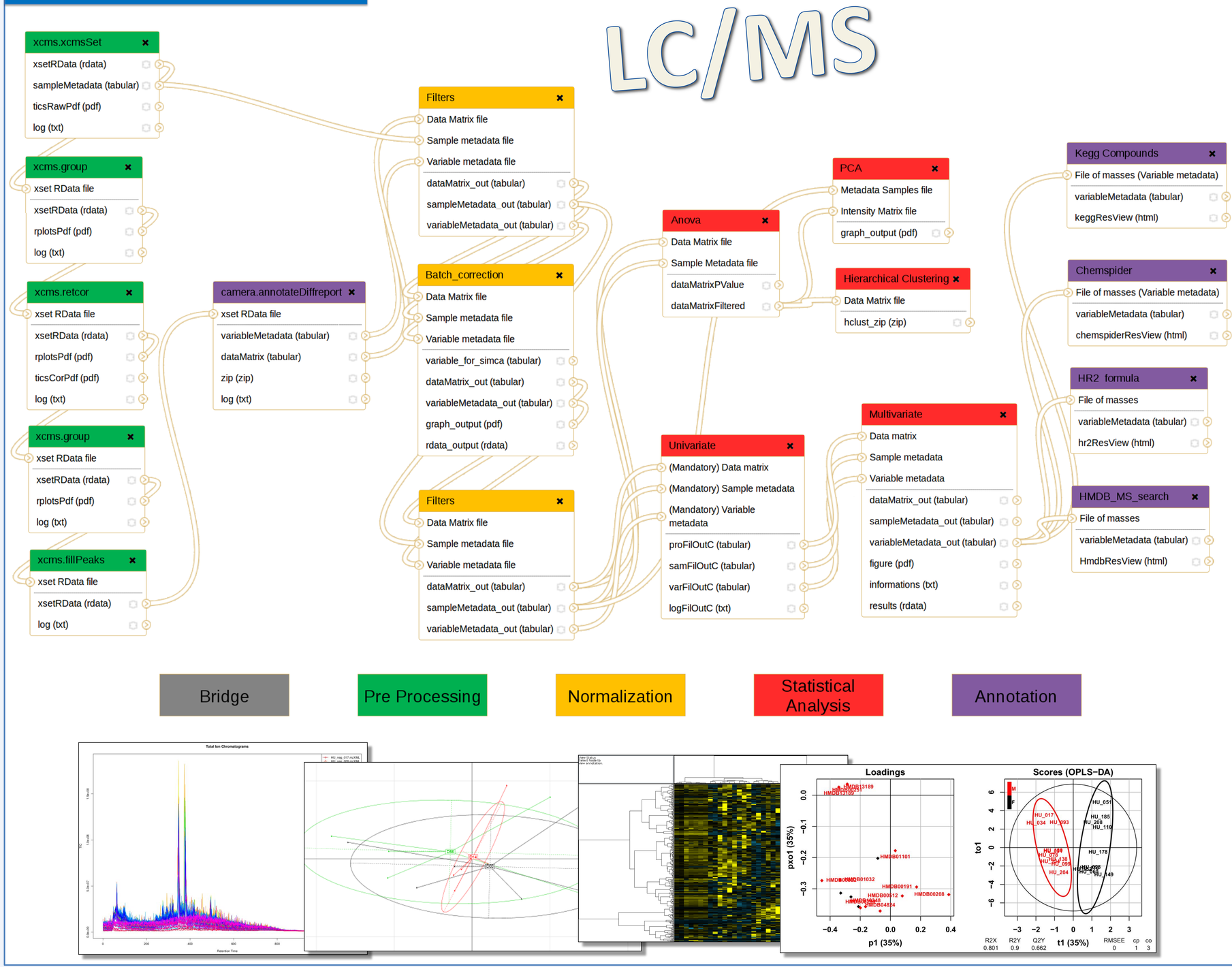
[1] ABiMS, FR2424 CNRS-UPMC, Station Biologique, Place Georges Teissier, 29680, Roscoff, France, [2] PFEM, UMR1019 INRA, Centre Clermont-Ferrand-Theix, 63122, Saint Genes Champanelle, France, [3] PF MetaToul-AXIOM, UMR 1331 Toxalim INRA, 180 chemin de Tournefeuille, F-31027, Toulouse, France, [4] CEA, LIST, Saclay Centre, F-91191, Gif-sur-Yvette, France

Abstract

Facing the emergence of new technologies in the field of metabolomics, treatment solutions adopted so far (XCMS, R scripts, etc.) clearly show their ergonomic limits for the experimenters point of views. Bottlenecks affect unified access to core applications as well as computing infrastructure and storage.
In the context of collaboration between metabolomics (MetaboHUB French infrastructure) and bioinformatics platforms (IFB: Institut Français de Bioinformatique), we have developed a full pipeline using Galaxy framework for data analysis including preprocessing, normalization, quality control, statistical analysis and annotation steps.
This modular and extensible workflow is composed with existing components (XCMS and CAMERA functions, etc.) but also a whole suite of complementary statistical tools. This implementation is accessible through a web interface, which guarantees the parameters completeness. The advanced features of Galaxy have made possible the integration of components from different sources and of different types. An extensible environment is offered to metabolomics communities (platforms, end users, etc.), and enables preconfigured workflows sharing for new users, but also experts in the field.
Finally, a real strategy of metabolite annotation improvement is proposed, combining powerful statistics (Univariate, Multivariate PLS/OPLS) and annotation tools. In addition, some future complementary developments are in progress : machine learning, pathway analysis (ProbMetab), banks combiner, spectra matching …

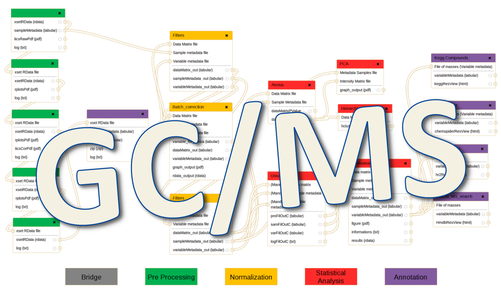
The workflow

LC/MS

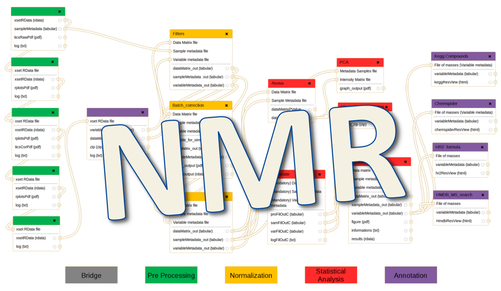


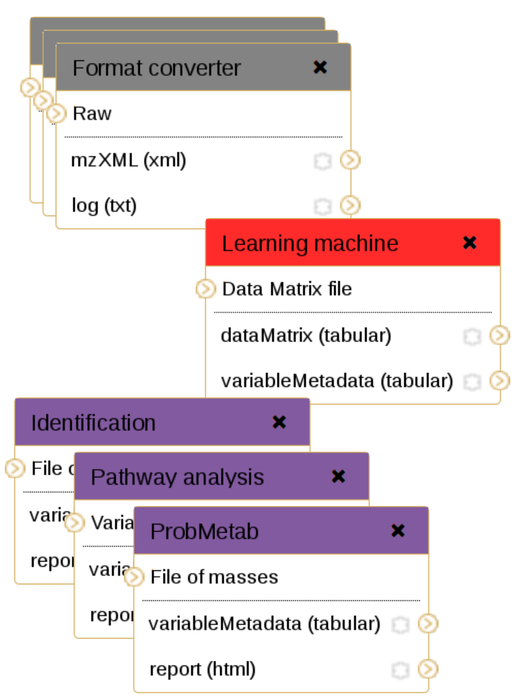
Next

GC/MS

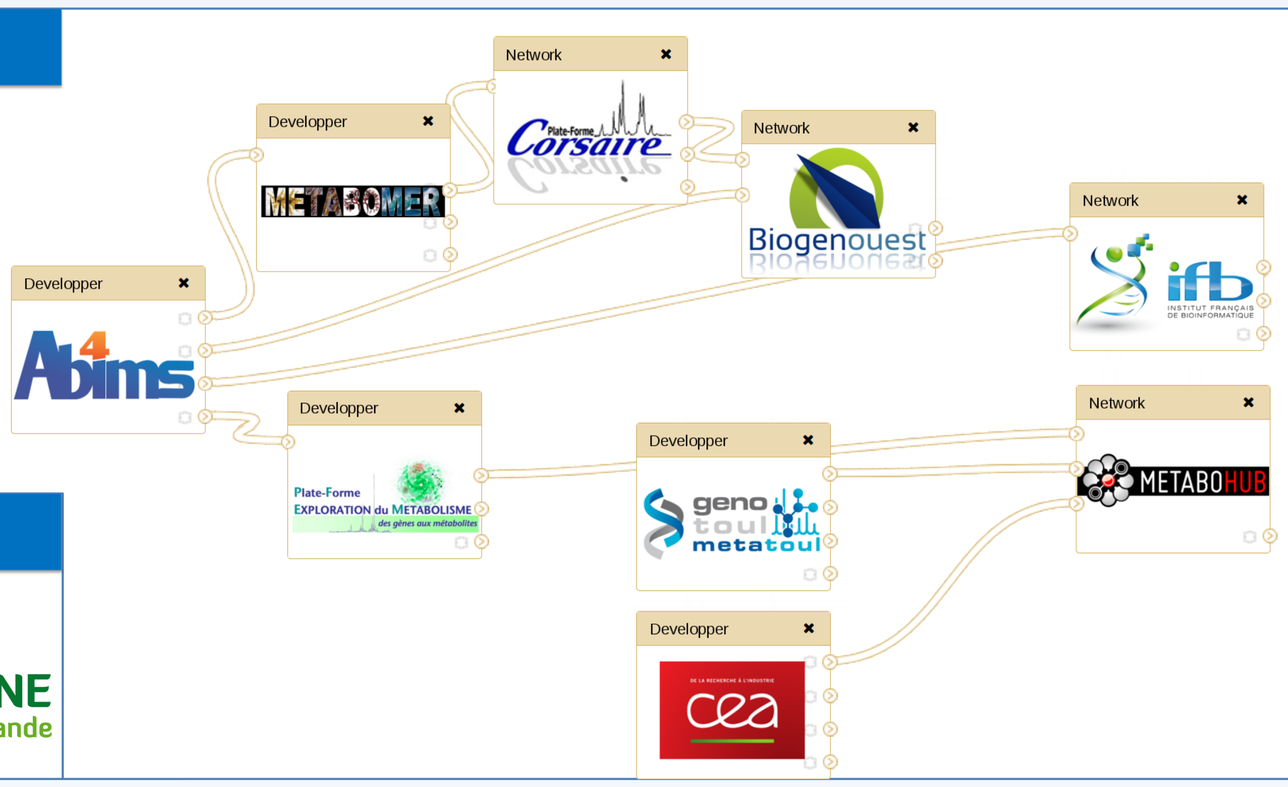


NMR





Community



Fundings

Conclusion

- A functional pipeline accessible through a web interface right now.
- Future developments in progress:
  - **Preconfigured workflows** for MS specific instruments (High and Low res, …)
  - **New tools** for bridge, statistical analysis, annotation categories.
  - Extension to **new resources**: GC/MS, NMR

contact@workflow4metabolomics.org

- Goecks, J, Nekrutenko, A, Taylor, J and The Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. 2010 Aug 25;11(8):R86.
- Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J. "Galaxy: a web-based genome analysis tool for experimentalists". Current Protocols in Molecular Biology. 2010 Jan; Chapter 19:Unit 19.10.1-21.
- Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A. "Galaxy: a platform for interactive large-scale genome analysis." Genome Research. 2005 Oct; 15(10):1451-5.
- Smith, C.A., Want, E.J., O'Maille, G., Abagyan,R., Siuzdak and G. (2006). "XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching and identification." Analytical Chemistry, 78, pp. 779–787.
- Kuhl C, Tautenhahn R, Boettcher C, Larson TR and Neumann S (2012). "CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets." Analytical Chemistry, 84, pp. 283–289.
- R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/