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/PLS) and annotation tools. In addition, some future complementary developments are in progress: machine learning, pathway analysis (ProbMetab), banks combiner, spectra matching ...