MOLAR: a streamlined data analysis pipeline for metabolomics data processing and interpretation

Bioinformatics

Christine Isaguirre (christine.isaguirre@vai.org), Van Andel Research Institute, **Abigail Ellis**, Van Andel Research Institute, **Carlos Castello**, Van Andel Research Institute, **Russell Jones**, Van Andel Research Institute, **Brejnev Muhire**, Van Andel Research Institute, **Ryan Sheldon**, Van Andel Institute

Metabolomics cores often return data to users as a raw peak area table of metabolites in biological samples. This forces end-users with varying metabolomics experience to properly process data and draw biological conclusions Objective: To increase user-accessibility, the Van Andel Institute Metabolomics and Bioenergetics Core developed an R-based program: MetabOlomics Analytical Results (MOLAR), which accepts a raw peak area table, performs statistical analyses, and outputs a biologically interpretable, user-friendly report.

Methods: MOLAR was built by compiling open-access metabolomics and statistical tools (MetaboAnalyst, StatTarget) into one R-program which synchronizes analysis between packages. MOLAR accepts a raw peak area table generated by MassHunter Quantitative Analysis (Quant; Agilent) after completion of peak picking and integration. The Quant method used is generated from QQQ MRM metabolite profiling and includes approximately 250 metabolites. The raw peak area table is combined with two additional user generated input files: one containing sample group metadata and the other identifying compounds that pass visual inspection of peak picking and integration. The MOLAR script was equipped with a graphical user interface (GUI) made in R-shiny for ease of use. The GUI contains adjustable settings to allow the user to tailor data processing for the unique needs of each analysis.

Results: MOLAR both minimizes input data formatting, completes pre-processing (blank filtration, normalization, and signal drift correction from pooled quality control samples) and post-processing (statistical analyses, data visualization, and reporting). The output is a PDF report with relevant statistical figures including principal component analysis (PCA) plots, volcano plots, heatmaps, and pathway analysis (generated from the Kegg pathway database). MOLAR also generates CSV files with numerical t-test, volcano plot, fold change, and PCA results to allow end-users to generate their own plots as needed.

Conclusions: MOLAR is a data analysis pipeline which streamlines statistical analyses and data visualization. MOLAR generates data reports, which improve data accessibility and ability to draw biological inference.