A Laboratory Information Management System (LIMS) for High-Throughput LC-MS Metabolomics-Based Biomarker Discovery

Alan M. Smith, Ph.D.[1], Yuerong Zhu, Ph.D. [2], Paul R. West, Ph.D.[1], April M. Weir,M.S.[1], Gabriela G. Cezar, DVM, Ph.D.[1]
[1]Stemina Biomarker Discovery, Inc., 504 S. Rosa Rd., Suite 150, Madison, WI 53719
[2]BioInfoRx, Inc., 1502 Beechwood Cir, Middleton, WI 53562

OVERVIEW

- Developed a LIMS system combining experiment, sample, LC-MS, feature analysis, and interpretation of data into an interconnected web-based user interface
- System designed to handle thousands of LC-MS samples from planning to final interpretation
- Open source framework allows rapid integration of new project specific data analysis tools
- System tested on the Stemina Developmental Toxicity Platform
- Complex analysis procedures were simplified into simple web-based interfaces allowing nearly complete data analysis automation

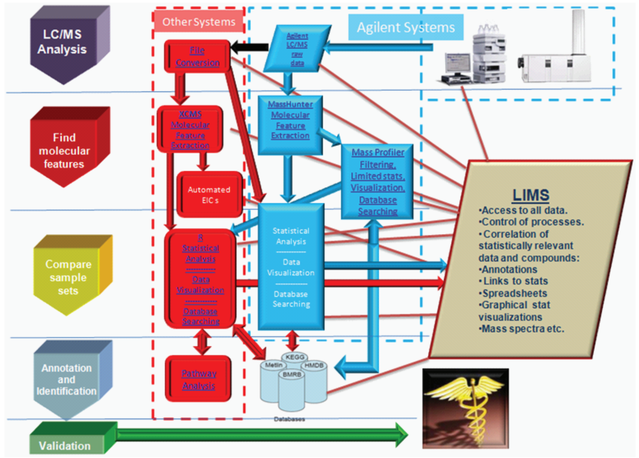

Figure 1. General overview of analysis pipeline combining proprietary and open source software managed by the SteminaLIMS.

INTRODUCTION

Large sample size multifactorial experiments executed under high throughput conditions require laboratory information management systems (LIMS) capable of simplifying an entire project pipeline (Figure 1). SteminaLIMS is a web-based laboratory management tool specifically designed to streamline the processes involved with metabolomics-based biomarker discovery (Figure 2). The SteminaLIMS manages the entire stream of textual, numerical, statistical and graphical data including project information, sample preparation and metadata, mass spectrometry data acquisition and analysis, file conversion, peak picking, statistical analysis, annotation, and small molecule confirmation by MS-MS. The system incorporates convenient user friendly tools that allow collaborative input, tracking, and dissemination of information. The data analysis pipeline uses both proprietary and open source data analysis tools to simplify the trek of small molecule biomarker discovery.

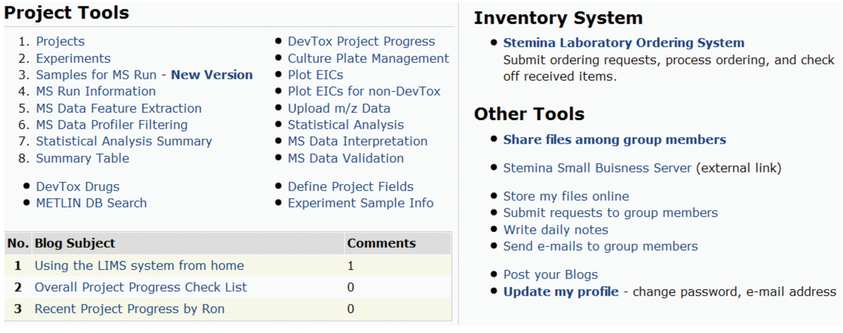

Figure 2. LIMS interface for Developmental Toxicity metabolomics platform. A large number of tools are available for the user to manage and analyze data.

METHODS

The SteminaLIMS is built by BioInfoRx based on its innovative information management technologies. Web programming technologies, such as LAMP (Linux, Apache, MySQL, and PHP) and AJAX (Asynchronous JavaScript and XML), are used to design the backend databases and the web-based interfaces. The system integrates data management with LC-MS analysis processes, providing improved work flow automation. Custom statistical analysis tools were created with open source statistical software R and Bioconductor. These tools allow users to select and preprocess mass feature tables, bin mass features across LC-MS samples, perform differential analysis of feature abundances and create summary tables to identify features of interest. Online mass spectral data analysis tools for EIC and spectra generation are based on the XCMS library in R.

RESULTS

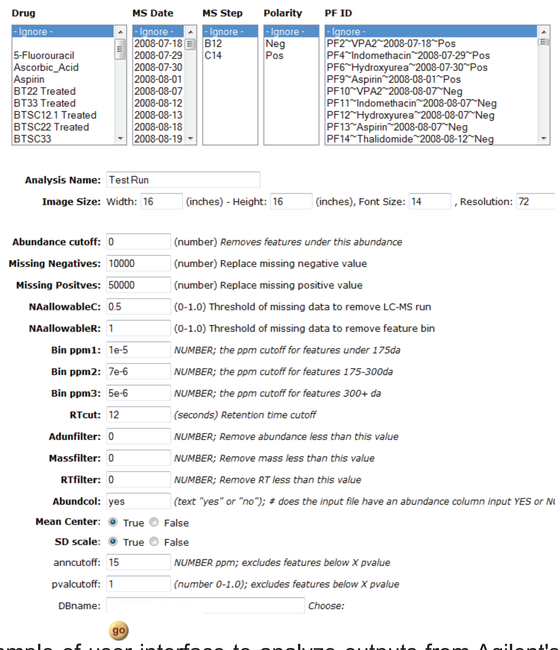

Figure 3. Example of user interface to analyze outputs from Agilent's Mass profiler software. This data analysis procedure reads in Mass Profiler outputs, reorganizes the data, groups mass features across experiments, performs univariate and multivariate statistical analysis, creates project specific graphical outputs of the analysis, annotates the features using several in-house DB's, and generates easy to read data summary documents.

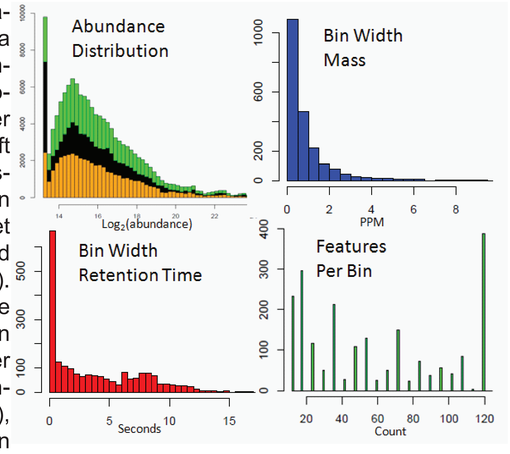

Figure 4. Automated output of feature bin assessment that allows a user to rapidly tune input parameters to optimize the binning process for multiple MassProfiler (Agilent) inputs. The upper left abundance plot examines the distribution of feature abundance in total experimental data set (green), cell cultures (black), and media negative control (orange). The three other plots examine the characteristics of feature bins in terms of length of bin in parts per million mass (upper right), retention time in seconds (lower left), and the number of features per bin (lower right).

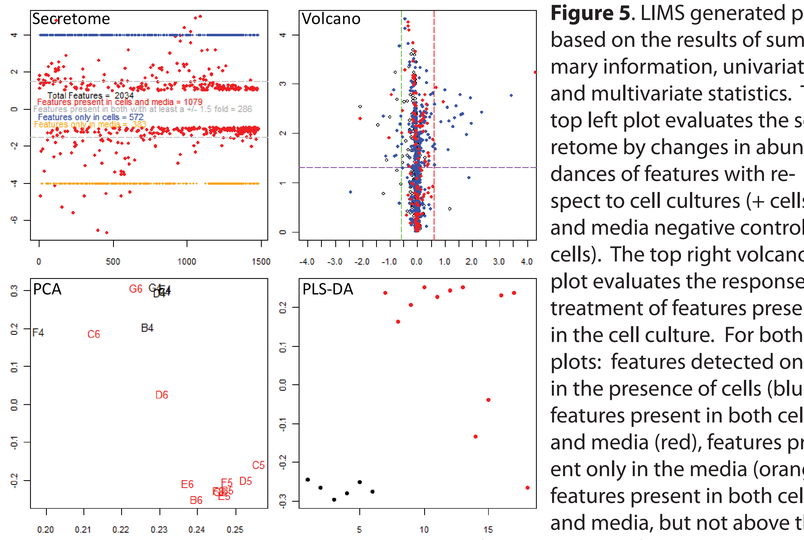

Figure 5. LIMS generated plots based on the results of summary information, univariate, and multivariate statistics. The top left plot evaluates the secretome by changes in abundances of features with respect to cell cultures (+ cells) and media negative controls (- cells). The top right volcano plot evaluates the response to treatment of features present in the cell culture. For both plots: features detected only in the presence of cells (blue), features present in both cells and media (red), features present only in the media (orange), features present in both cells and media, but not above the threshold lines (open black circles). The bottom left loadings plot of the samples contains the first and second principle components from NIPALS PCA (Black = Control, Red=Treated). The bottom right PLS-DA scores plot demonstrates separation of treated (red) and control (black) samples using the SIMPLS method. Information used in each of these plots is also provided in summary spreadsheets such as difference from media, p values, and VIP scores for each feature.

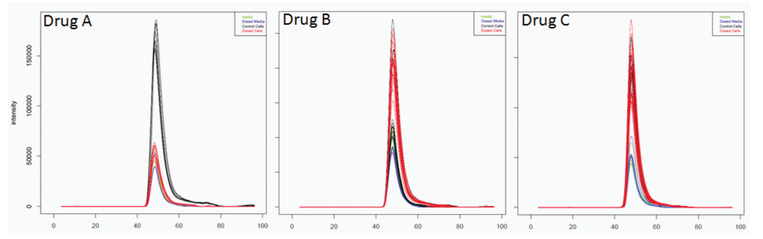

Figure 6. Smoothed extracted ion chromatogram (EIC) for a single mass feature generated by the LIMS (based on features that are secreted by human embryonic stem cells) exhibits statistically significant difference in abundance in response to specific classes of drugs, and were selected based on variable importance from a PLS-DA model.

CONCLUSIONS

- An open-source LIMS creates a working environment that can simplify a complex data analysis pipeline (**Figure 1**) into a simple user interface (**Figure 2**).
- Project specific data analysis procedures can be reduced to simple and easy to use web-based user interfaces with minimal effort (**Figure 3**).
- Tailed outputs for parameter optimization such as combining large numbers of MassProfiler outputs can be easily presented to system users who can evaluate the results (**Figure 4**).
- Project specific outputs and summary statistics are created and presented to users freeing bioinformaticists for other projects (**Figure 5**).
- The final results of statistical analysis can be evaluated by users for features that are acceptable for further confirmation by a mass spectrometrist (**Figure 6**).
- Project data was entered by multiple scientific specialists and passed successfully from experimental design stages to final MS-MS confirmation of differentially secreted small molecules using the SteminaLIMS.

WWW. STEMINA.COM