P-Mart – Interactive Analysis of Ion Abundance Global Proteomics Data

*Lisa M. Bramer[1], Kelly G. Stratton[1], Amanda M. White[1], Ameila H. Bleeker[1], Markus A. Kobold[1], Katrina M. Waters[2], Thomas O. Metz[2], Karin D. Rodland[2], Bobbie-Jo M. Webb-Robertson[1]**

[1]Computing & Analytics Division, Pacific Northwest National Laboratory, 902 Battelle Blvd, Richland, WA USA

[2] Biological Sciences Division, Pacific Northwest National Laboratory, 902 Battelle Blvd, Richland, WA USA

Corresponding Author: bj@pnnl.gov

ABSTRACT

The use of mass spectrometry-based techniques for global protein profiling of biomedical or environmental experiments has become a major focus in research centered on biomarker discovery. However, one of the most important issues recently highlighted in the new era of omics data generation is the ability to perform analyses in a robust and reproducible manner. This has been hypothesized to be one of the issues hindering the ability of clinical proteomics to successfully identify clinical diagnostic and prognostic biomarkers of disease. P-Mart (https://pmart.labworks.org) is a new interactive web-based software environment that enables domain scientists to perform quality control processing, statistics, and exploration of large-complex proteomics datasets without requiring statistical programming. P-Mart is developed in a manner that allows researchers to perform analyses via a series of modules, explore the results using interactive visualization, and finalize the analyses with a collection of output files documenting all stages of the analysis and a report to allow reproduction of the analysis.

KEYWORDS: software, proteomics, web-service, reproducibility, statistics, exploratory data analysis, visualization

1. INTRODUCTION

Global mass spectrometry (MS)-based proteomics is a high-throughput technology that allows hundreds of thousands of peptides to be mapped to tens of thousands of proteins from many types of samples (e.g., blood, urine, stool, etc.). This offers an incredible opportunity to understand biological functions at the protein level in relationship to phenotypes of interest[1-4]. However, as with many complicated high-throughput technologies the data is complex and plagued with

multiple sources of variability, such as sample preparation, ionization and peptide identification, which consequently generate a challenging analysis process.

Global proteomics data is generated by measuring spectra that are then matched, typically using a similarity algorithm[5-8], to peptides. These peptides can then be quantified using metrics such as the ion abundance or intensity. After this task is complete, downstream analysis is targeted at generating high-quality statistics at the peptide or protein level from complex quantified peptide datasets. This downstream analysis is often performed using functions available through statistical packages, such as R[9-12]. The benefit of these packages to the community is immense, offering a large number of statistical and exploratory analysis tools. There have also been multiple tools provided to the community to simplify the downstream processing of proteomics data, such as Galaxy and Taverna[13-15]. These tools are extremely powerful for creating and deploying workflows either within the R computing environment or across services. In terms of stand-alone software the most robust is Perseus[16], which offers a lot of downstream processing options and is a great solution for many proteomics data processing tasks. The gaps that P-Mart fills in respect to the existing suite of capabilities for downstream processing of proteomics data are the ability to analyze data in a web-service environment, unique quality control and data analysis processing capabilities, and statistical analyses, including exploratory data analysis, that do not require imputation of the missing data. In addition, P-Mart offers unique trellis visualizations that allow researchers to look at individual proteins and move through the large amount of data based on statistics and data characteristics.

We present a new approach, P-Mart, to both simplify these analyses and increase reproducibility for computational and biological researchers. P-Mart takes several new and existing R functions and provides a holistic approach to data analysis, allowing all steps of analysis from quality control processing through pattern discovery, to be performed in a workflow-based manner that ends in a detailed documentation of the methods that were employed, as well as multiple export files that capture the output of each stage that modifies the data. The web-development process for P-Mart integrates R functions in a straight-forward manner, which will allow developers to easily add new capabilities to P-Mart beyond the current set included in the initial release. P-Mart's overarching capabilities in the context of analyzing existing Clinical Proteomics Tumor Analysis Consortium (CPTAC) data[17] has been previously described[18]. Herein we describe P-Mart as a capability for use with user-uploaded data to allow discovery and exploration of potential biomarker candidates for many user-defined global peak-intensity proteomic datasets, Figure 1.

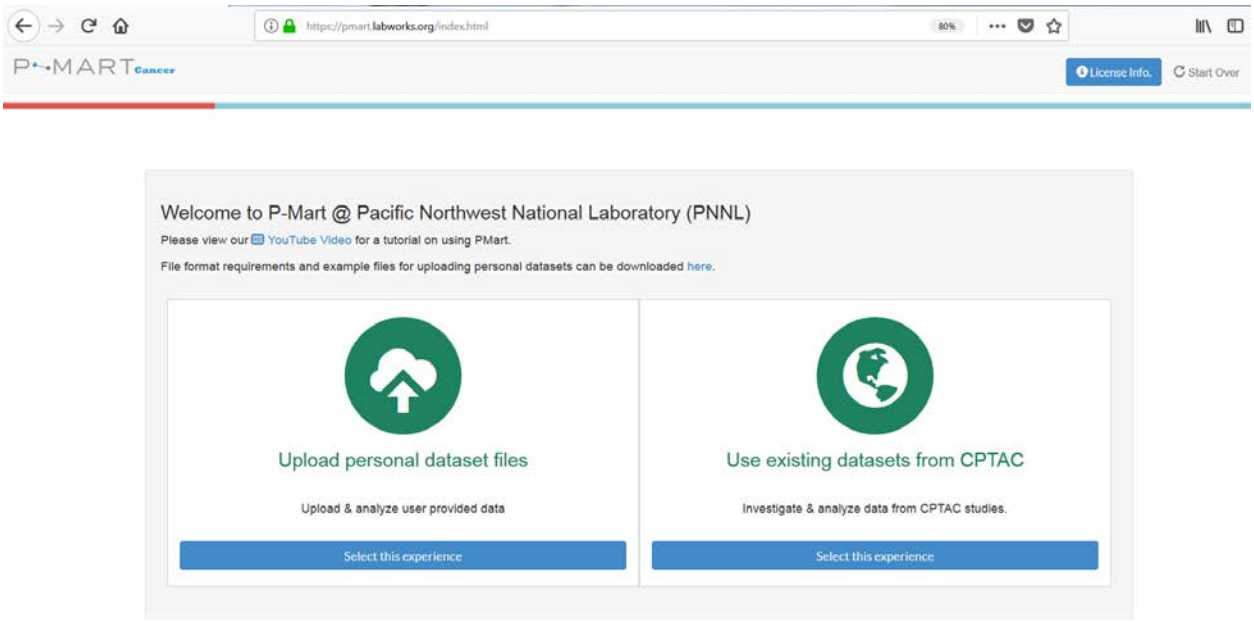

Figure 1. The opening page of P-Mart, which offers analyses of personal data or existing CPTAC data. Video tutorials are available from this page at http://bit.ly/PMartPNNL to train users on multiple aspects of the software.

2. SOFTWARE METHODS

P-Mart consists of two connected, yet distinct, capabilities. First is the underlying statistical functionality developed in R and Rcpp and the second is the user web-interface developed in JavaScript. These two components are linked through a tool, Rserve[19], to make one seamless experience for the user of the web service. A user can add desired statistical or data processing functionality through modification of existing R functions or the addition of a custom R function added to the available R package. If the new functionality is added to an existing module, the user only needs to modify the web interface JavaScript code to include the new function/process as an option, or a user can simply create a new module to add to the existing workflow options. A user may include custom functions to their local version of P-Mart through the addition of a new R

function to the package[20] and exposing the function in the web-service through the addition of JavaScript code to the Spring Framework[21].

A subset of the full R functions, as well as example data, are available for R programmers on GitHub to allow for customized processes to be developed (https://github.com/pmartR/). This includes pmartR and pmartRdata packages. The functions available as part of the web-service are described in Supplemental Table 1. The pmartR package includes the functions for data transformation, filtering, quality control analysis of samples, normalization and statistical tests. The pmartRdata package describes and gives example files for using the functions in pmartR.

The web-interface was developed with JavaScript and utilizes High-Charts[22] for the interactive visualizations. The web-service is wrapped up as a Docker Container, so it can be used in various computing environments (https://hub.docker.com/r/pnnl/pmart-web/). The web-service offers the same files as in the pmartRdata package in the format needed for upload into the web application, as well as access to multiple cancer proteomics datasets and their clinical data generated through the CPTAC preloaded for exploration[17-18].

3. RESULTS AND DISCUSSION

As seen in Figure 1, P-Mart offers a broad suite of capabilities. The statistics and protein quantification capabilities will be described briefly, but additional detail can be found in the P-MartCancer description[18]. A concise and simple format for user uploads has been developed that can be generated from most peptide or protein abundance files. This format is easily generated from common formats output by MaxQuant and mzMine[23-25]. Instructions for the file format and

conversion tutorials are available for download from the data upload page. In addition, a description of our interactive exploratory visualization capability that interacts with statistical parameters to quickly highlight key potential biomarkers will be given, as well as a detailed view of the export and documentation capabilities to enable reproducibility. Finally, we will give a brief use-case based on the example data available on the website.

3.1. Overall P-Mart Capabilities

P-Mart offers access to analyze user data through the main web page (Figure 1). P-Mart capabilities are demonstrated via video tutorials which are related to user uploads as well as on using the module-based workflow (http://bit.ly/PMartPNNL). Below we give an overview of the capabilities but point readers to the videos for details of user uploads and software manipulation. One of the novel components of P-Mart compared to existing tools is that all tasks below are completed without imputation, which has been demonstrated to have a significant impact and potential to introduce significant bias on downstream analyses[26-27].

3.1.1. *User Uploads*

The user-upload functionality requires an upload of two or three .csv files as well as the specification of certain information about the data (e.g., is it peptide, protein, or gene level data; has the data been transformed and/or normalized already; which columns correspond to which biomolecules; what variables are of clinical interest). The upload screens for the use-case are shown in Figure 2. The first required file contains the quantified data for each biomolecule (rows) and sample (columns) and the second required file contains sample information including sample names and variables of interest (e.g., treatment groups or demographics). The third file is optional, and contains metadata on the biomolecules, such as mappings from peptides to proteins or genes.

In order to quantify from peptides to proteins or genes, this metadata file must be provided and include columns for peptide and protein or gene.

3.1.2. *Statistical Pre-processing and Analyses*

The statistical functions of P-Mart are focused on quality control processing, normalization, and basic statistical tests. The quality control functions include filtering of low-quality peptide or protein information, as well as identification of samples with outlier behavior and normalization guidance[9, 28-31]. This allows for the generation of a dataset with improved properties in the context of variance and missing data for the purposes of statistical analysis. Finally, statistics allow for comparison of two or more groups via analysis of variance-based approaches for quantitative differences and a G-test for qualitative differences, as well as options for multiple test corrections.

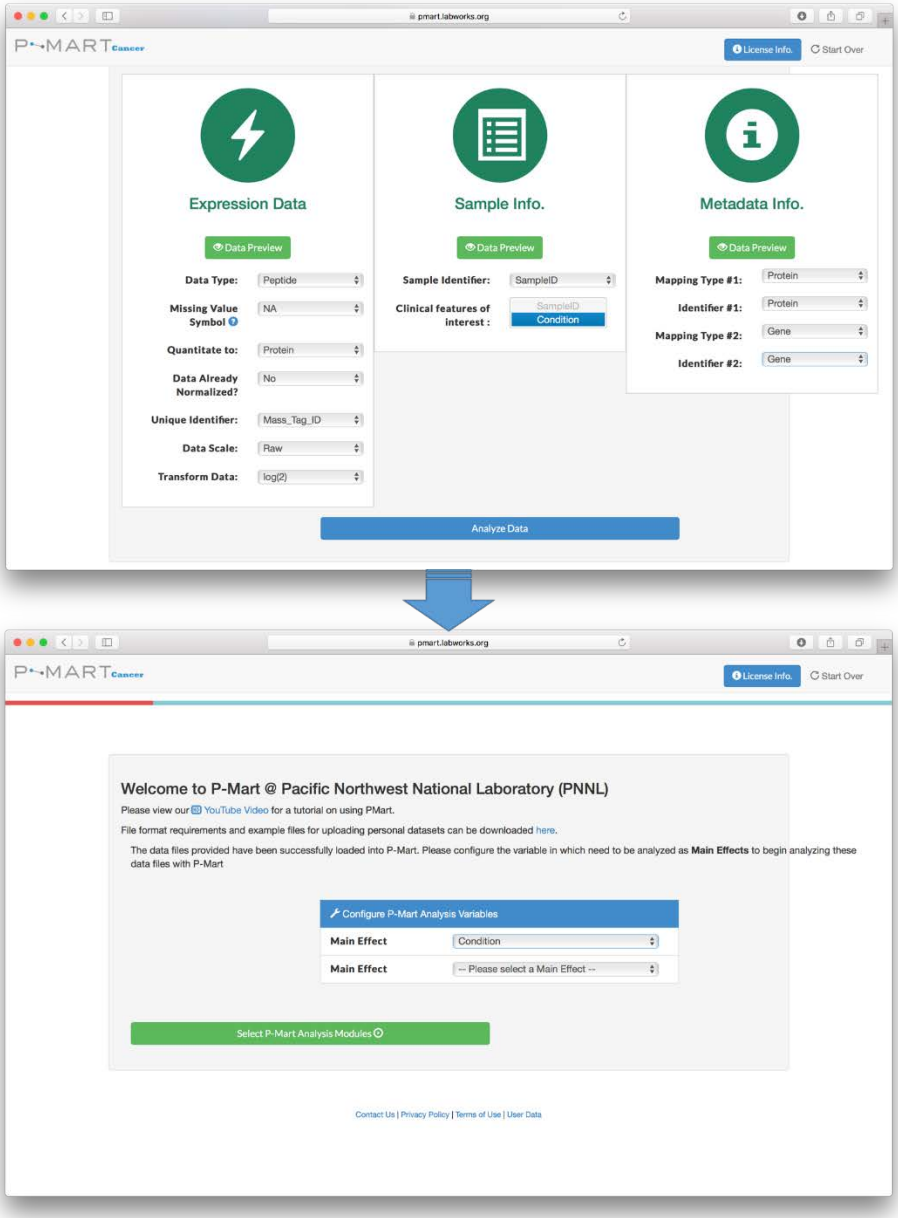

Figure 2. P-Mart user interface showing the selections made for the example MERS-Co dataset upload.

3.1.3. *Exploratory Data Analyses*

P-Mart offers standard pattern discovery tools, such as Principal Component Analysis (PCA) to evaluate visual separations of groups in the data, which is commonly performed in computational biology[26, 32-33]. A core component of P-Mart is that this task is completed on the data without requiring imputation[34].

P-Mart also consists of a unique capability for proteomics, an interactive visualization tool Trelliscope[35], which provides a flexible and scalable way to divide datasets and analysis results into meaningful subsets, apply a plot method to each subset, and then arrange those plots in a grid and interactively sort, filter, and query panels of the display based on metrics of interest[36]. This allows a user to interactively explore their data and statistical results to make comparisons, evaluate statistical results and models, and uncover the structure of data even when the structure is quite complex. We have used Trelliscope to allow specific queries of interest across peptides, proteins or genes for various levels of information (e.g., coverage, statistical significance, fold-change) and to display the results in a fashion that allows interactive discovery in lieu of sorting through spreadsheets.

3.1.4. *Exports and Reporting*

One of the most important features of P-Mart is the reporting and export functionality because it offers all the necessary information to reproduce an analysis exactly and all data files throughout the process where the underlying data has been modified (e.g., quantified to protein, normalized) in addition to the final statistics. This focus on reproducibility is a key issue currently in the field of biology[37-39] and P-Mart is one methodology pursing strategies to ensure reproducible analyses.

3.2. Case Study – Virology Proteomics Analysis

We present a case-study of real data to demonstrate the utility and novel features of P-Mart. We utilize a proteomics dataset from a study on Middle Eastern Respiratory Syndrome Coronavirus (MERS-CoV), where the objective was to analyze the response to MERS-CoV at 18 hours post infection in human Calu-3 cells[40-41]. The dataset includes 3 uninfected samples and 9 infected samples and is available for download from the P-Mart website. Figure 2 shows P-Mart at the data upload stage where we have input the three files via the user uploads page. The feature of interest in the MERS-CoV dataset is "Condition", which refers to whether the sample is infected or not, so this variable is selected in the user-interface as the main effect. The user can custom-select a workflow or allow P-Mart to make a suggestion based on the data provided. We selected workflow modules for quality assessment on peptides and samples, peptide-level statistics, protein quantification, protein-level statistics, pattern discovery and interactive Trelliscope displays (Figure 3).

Figure 3. Module selection in P-Mart allows the user to add and removed existing capabilities easily, as well as new capabilities in R to be easily added to P-Mart.

The pre-processing of the data included a sample outlier filter[29], with a significance threshold of 0.001, which removes one sample, and a peptide coverage filter[31] removes 7,346 peptides (Figure 4). Thus, at the end of the quality control processing there are 11 samples and 10,081 peptides ready for statistical analysis and further processing. The 7,346 peptides were removed because the low occurrence indicates that no meaningful statistical comparison can be made. These datasets were then normalized by global median centering[28].

Differential peptide statistics compare the infected samples to the uninfected mock control samples using both quantitative (t-test) and qualitative (G-test) tests[31]. Next, a standard reference-based median quantification is used to obtain data at the protein level[42-43]. Protein evaluation continues with statistical tests analogous to the peptide-level data. The results of the statistical analysis are bar graphs showing the number of peptides or proteins that are significant at a user defined threshold (e.g., p-value < 0.05). In this example dataset there are over 600 significant proteins. These proteins are moved forward for further analysis with methods such as PCA or query-based evaluation in Trelliscope.

We focus the remainder of the use-case on identification of interesting candidate markers using the Trelliscope capability. Trelliscope yields both peptide and protein level displays from which the user can choose. For the protein data, each graph represents a single protein, resulting in 630 displays for this dataset. We use the Panel Labels options to customize the information shown for each protein, ultimately selecting the p-value from the t-test, fold change, and protein name (Figure 5).

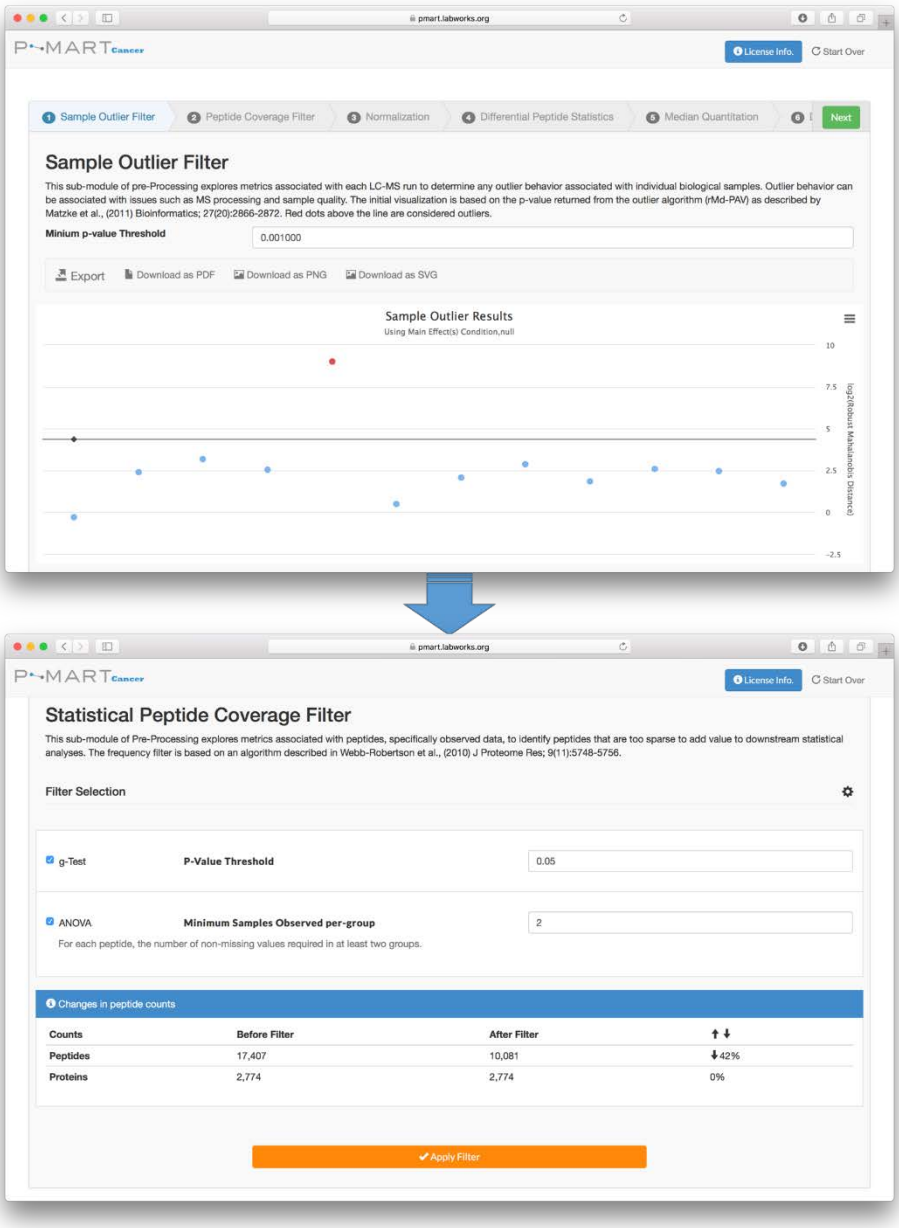

Figure 4. P-Mart quality assessment tools to evaluate the quality of samples (top) and adequate statistical coverage of peptides (bottom). Each module allows the user to modify the parameters to be more or less strict in the selection criteria.

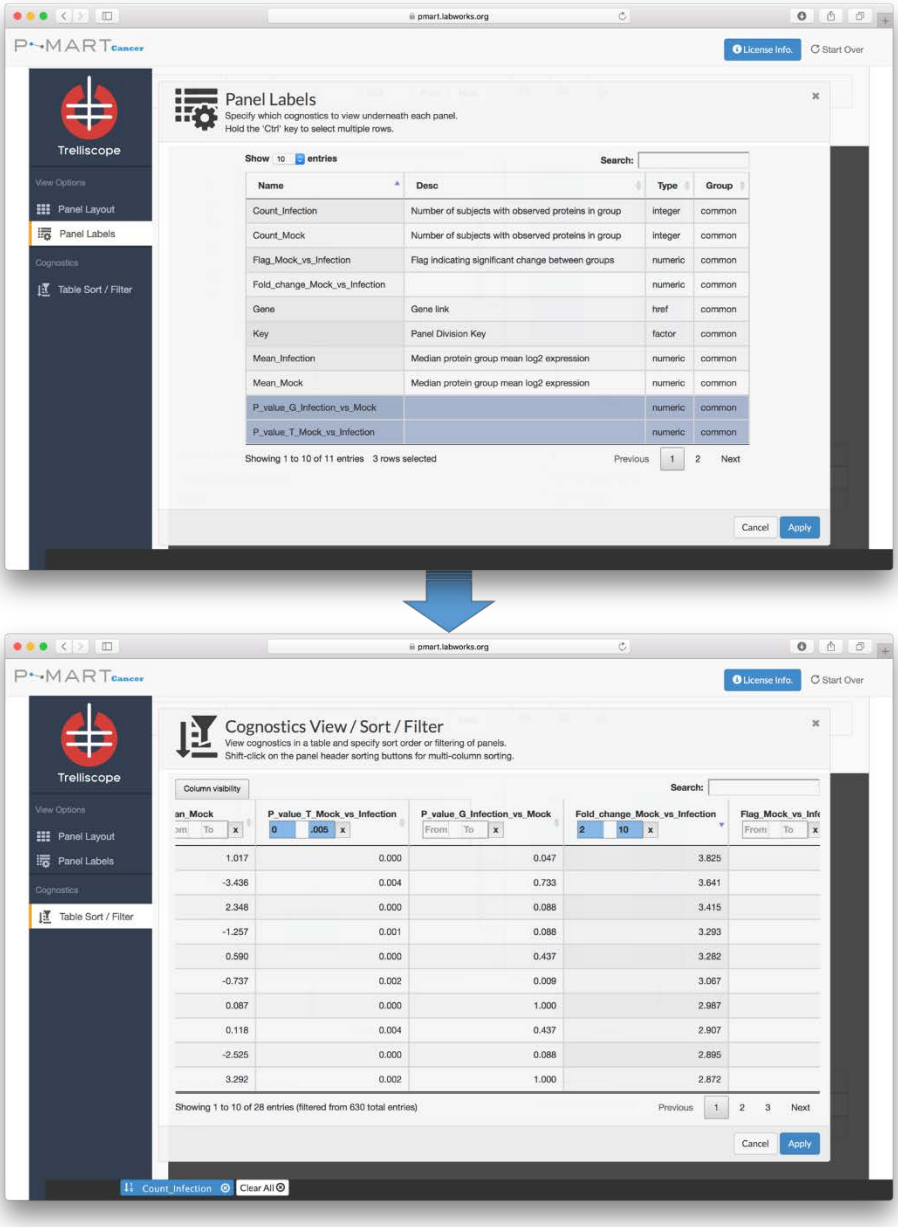

Figure 5. P-Mart Trelliscope features allow customization of the metrics that will be displayed with each plot (top) and the capability to sort, filter and search based on parameters of interest.

Trelliscope allows a user to sort and filter these in order to focus on the proteins most interesting to the user. To sort and filter the proteins we use the Table Sort/Filter options and filter down to those proteins with p-value < 0.005 and log2 fold change between 2 and 10 (corresponding to proteins displaying down regulation in the infected group), reducing the displays to a subset of 28 proteins to examine (bottom Figure 5). We then sort by fold change. Among the subset of proteins are NGAL_HUMAN, DRG1_HUMAN, GANAB_HUMAN, and BLVRB_HUMAN, which are involved with kidneys and lung function, which MERS-CoV is known to affect these organs. Through Trelliscope, we were also able to identify several proteins linked to genes associated with antigen presentation, which displayed behavior consistent with prior findings in this study [41]. For example, TBA4B (involved in antigen presentation) and B2MG (involved in immunodeficiency) were in the subset of 28 proteins. Trelliscope provides a link to the Uniprot page[44] for each protein, allowing verification of the types of genes and processes related to the proteins of interest. Figure 6 gives an example of this visualization, NGAL_HUMAN protein level data displayed as a boxplot and when the link is selected the Uniprot page gives information on the protein selected. It is comforting to identify proteins that are associated with functions expected to be observed in the experiment, however a key benefit of the rapid visual exploration in Trelliscope are the considerable number of proteins with unknown function or no obvious connection to explore either computationally or experimentally.

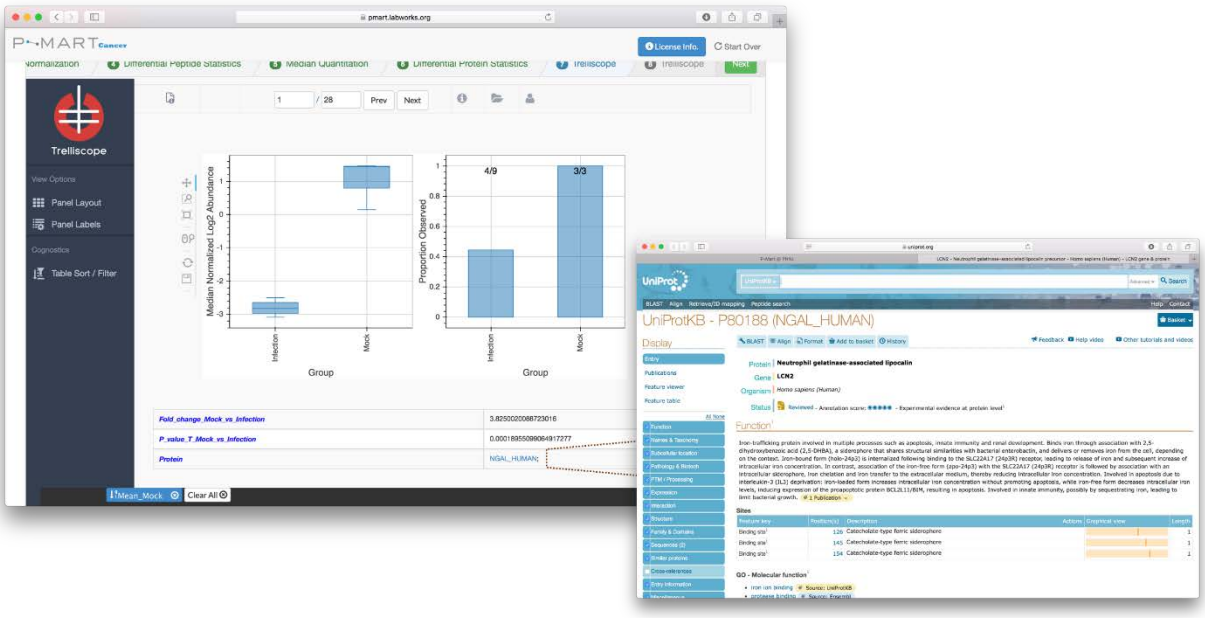

Figure 6. P-Mart Trelliscope gives a visual display of each protein that was selected based on user interaction (Figure 5) and allows interactive evaluation of the proteins through links to public resources, such as Uniprot.

One of the most important features of P-Mart is the reporting and file downloads. Figure 7 shows the report that is produced at the end of the analysis. Scrolling through this report a user will find all the steps that were performed on the data with enough detail to reproduce the analysis. The bottom of the report offers references and a link to download all files associated with the analysis. As seen in Figure 8 all files that are created throughout the analysis, including intermediate files at steps such as normalization, are provided to the user, as well as the statistical analyses results.

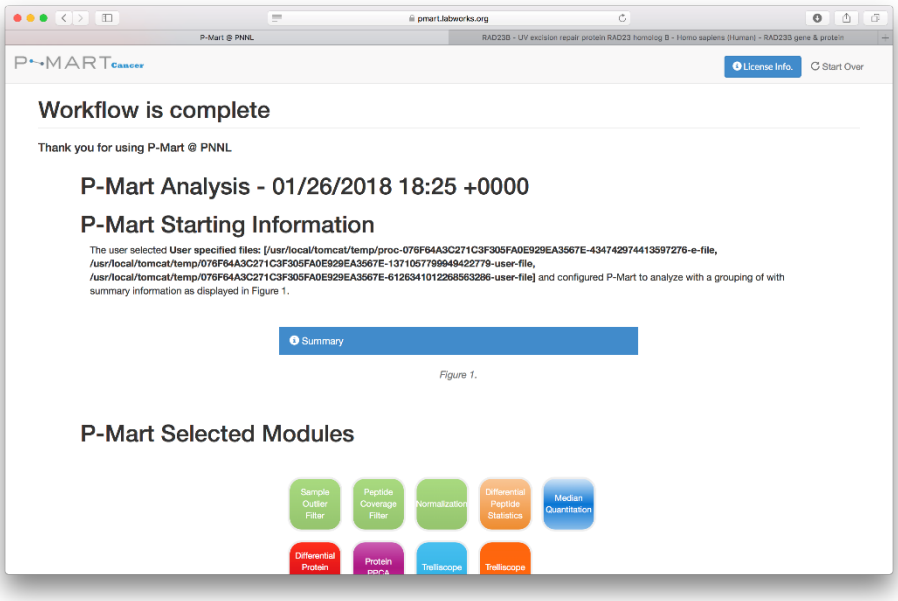

Figure 7. Example of the report yielded at the end of the P-Mart analysis.

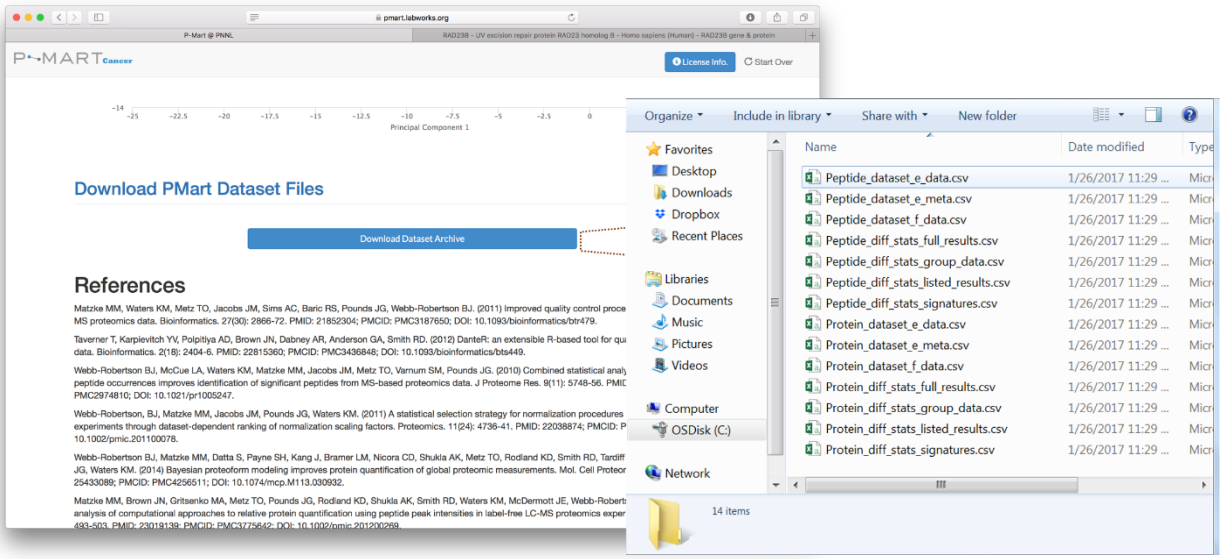

Figure 8. P-Mart download is available at the end of the report. One the right is the collection of data that is available to the user, including normalized data, protein-level data and all statistics that were performed.

4. CONCLUSIONS

P-Mart is a new online software tool to enable statistical processing and interactive discovery from peak-intensity global proteomics data in a reproducible fashion. The user can select a workflow from a series of modules and can easily progress through the workflow with visualization and tabular report capabilities integrated to allow the user to observe the changes in the data throughout the workflow. Novice users can use default parameters and expert users can tailor the tools to their specific requirements. P-Mart is a software, but aspires to meet the FAIR (Findable, Accessible, Interoperable, and Reusable) principles[45-46]. The underlying code base is available at various levels (https://github.com/pmartR/) and (https://hub.docker.com/r/pnnl/pmart-web/) and accessible through maintained online resources. As a web-service P-Mart can interoperable with other work-flow capabilities[13-15]. In addition, the R-code can be used directly to operate with other R packages[9, 12] allowing easy reuse of the code. The user is offered a report at the end of the analysis that allows the processing of a dataset to be reproduced and all data and statistics files, including intermediate analyses are available to the user.

SUPPORTING INFORMATION: The following supporting information is available free of charge at ACS website http://pubs.acs.org

Supplemental Table S1: P-Mart functions

AUTHOR INFORMATION

Corresponding Author

*Bobbie-Jo Webb-Robertson (bj@pnnl.gov).

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding Sources

This work has been supported by NCI grant U01-1CA184783 to B.J.M. Webb-Robertson. The example datasets were developed with support by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award Number U19A106772.

ACKNOWLEDGMENT

P-Mart was developed at Pacific Northwest National Laboratory, a multi-program national laboratory operated by Battelle for the U.S. Department of Energy under contract DE-AC06-76RL01830.

REFERENCES

1. Gajadhar, A. S.; Johnson, H.; Slebos, R. J.; Shaddox, K.; Wiles, K.; Washington, M. K.; Herline, A. J.; Levine, D. A.; Liebler, D. C.; White, F. M.; Clinical Proteomic Tumor Analysis, C., Phosphotyrosine signaling analysis in human tumors is confounded by systemic ischemia-driven artifacts and intra-specimen heterogeneity. *Cancer Res* **2015,** *75* (7), 1495-503.

2. Mertins, P.; Mani, D. R.; Ruggles, K. V.; Gillette, M. A.; Clauser, K. R.; Wang, P.; Wang, X.; Qiao, J. W.; Cao, S.; Petralia, F.; Kawaler, E.; Mundt, F.; Krug, K.; Tu, Z.; Lei, J. T.; Gatza, M. L.; Wilkerson, M.; Perou, C. M.; Yellapantula, V.; Huang, K. L.; Lin, C.; McLellan, M. D.; Yan, P.; Davies, S. R.; Townsend, R. R.; Skates, S. J.; Wang, J.; Zhang, B.; Kinsinger, C. R.; Mesri, M.; Rodriguez, H.; Ding, L.; Paulovich, A. G.; Fenyo, D.; Ellis, M. J.; Carr, S. A.; Nci, C., Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **2016,** *534* (7605), 55-62.

3. Slebos, R. J.; Wang, X.; Wang, X.; Zhang, B.; Tabb, D. L.; Liebler, D. C., Proteomic analysis of colon and rectal carcinoma using standard and customized databases. *Sci Data* **2015,** *2*, 150022.

4. Zhang, H.; Liu, T.; Zhang, Z.; Payne, S. H.; Zhang, B.; McDermott, J. E.; Zhou, J. Y.; Petyuk, V. A.; Chen, L.; Ray, D.; Sun, S.; Yang, F.; Chen, L.; Wang, J.; Shah, P.; Cha, S. W.; Aiyetan, P.; Woo, S.; Tian, Y.; Gritsenko, M. A.; Clauss, T. R.; Choi, C.; Monroe, M. E.; Thomas, S.; Nie, S.; Wu, C.; Moore, R. J.; Yu, K. H.; Tabb, D. L.; Fenyo, D.; Bafna, V.; Wang, Y.; Rodriguez, H.; Boja, E. S.; Hiltke, T.; Rivers, R. C.; Sokoll, L.; Zhu, H.; Shih Ie, M.; Cope, L.; Pandey, A.; Zhang, B.; Snyder, M. P.; Levine, D. A.; Smith, R. D.; Chan, D. W.; Rodland, K. D.; Investigators, C., Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell* **2016,** *166* (3), 755-65.

5. Duncan, D. T.; Craig, R.; Link, A. J., Parallel tandem: a program for parallel processing of tandem mass spectra using PVM or MPI and X!Tandem. *J Proteome Res* **2005,** *4* (5), 1842-7.

6. Tabb, D. L., The SEQUEST family tree. *J Am Soc Mass Spectrom* **2015,** *26* (11), 1814-9.

7. Xu, T.; Park, S. K.; Venable, J. D.; Wohlschlegel, J. A.; Diedrich, J. K.; Cociorva, D.; Lu, B.; Liao, L.; Hewel, J.; Han, X.; Wong, C. C. L.; Fonslow, B.; Delahunty, C.; Gao, Y.; Shah, H.; Yates, J. R., 3rd, ProLuCID: An improved SEQUEST-like algorithm with enhanced sensitivity and specificity. *J Proteomics* **2015,** *129*, 16-24.

8. Yates, J. R., 3rd, Pivotal role of computers and software in mass spectrometry - SEQUEST and 20 years of tandem MS database searching. *J Am Soc Mass Spectrom* **2015,** *26* (11), 1804-13.

9. Choi, M.; Chang, C. Y.; Clough, T.; Broudy, D.; Killeen, T.; MacLean, B.; Vitek, O., MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* **2014,** *30* (17), 2524-6.

10. Mueller, L. N.; Brusniak, M. Y.; Mani, D. R.; Aebersold, R., An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. *J Proteome Res* **2008,** *7* (1), 51-61.

11. Pendarvis, K.; Kumar, R.; Burgess, S. C.; Nanduri, B., An automated proteomic data analysis workflow for mass spectrometry. *BMC Bioinformatics* **2009,** *10 Suppl 11*, S17.

12. Zauber, H.; Schulze, W. X., Proteomics wants cRacker: automated standardized data analysis of LC-MS derived proteomic data. *J Proteome Res* **2012,** *11* (11), 5548-55.

13. Abouelhoda, M.; Issa, S. A.; Ghanem, M., Tavaxy: integrating Taverna and Galaxy workflows with cloud computing support. *BMC Bioinformatics* **2012,** *13*, 77.

14. Giardine, B.; Riemer, C.; Hardison, R. C.; Burhans, R.; Elnitski, L.; Shah, P.; Zhang, Y.; Blankenberg, D.; Albert, I.; Taylor, J.; Miller, W.; Kent, W. J.; Nekrutenko, A., Galaxy: a platform for interactive large-scale genome analysis. *Genome Res* **2005,** *15* (10), 1451-5.

15. Oinn, T.; Addis, M.; Ferris, J.; Marvin, D.; Senger, M.; Greenwood, M.; Carver, T.; Glover, K.; Pocock, M. R.; Wipat, A.; Li, P., Taverna: a tool for the composition and enactment of bioinformatics workflows. *Bioinformatics* **2004,** *20* (17), 3045-54.

16. Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M. Y.; Geiger, T.; Mann, M.; Cox, J., The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **2016,** *13* (9), 731-40.

17. NCI, National Cancer Institute Office of Cancer Clinical Proteomics Research, https://cptac-data-portal.georgetown.edu/cptacPublic/. **2018**.

18. Webb-Robertson, B. M.; Bramer, L. M.; Jensen, J. L.; Kobold, M. A.; Stratton, K. G.; White, A. M.; Rodland, K. D., P-MartCancer-Interactive Online Software to Enable Analysis of Shotgun Cancer Proteomic Datasets. *Cancer Res* **2017,** *77* (21), e47-e50.

19. Rserve Binary R server, https://www.rforge.net/Rserve/ https://www.rforge.net/Rserve/.

20. Wickham, H., *R packages: organize, test, document, and share your code*. O'Reilly Media: 2015; p 202.

21. Kayal, D., *Pro Java EE Spring Patterns: Best Practices and Design Strategies Implementing Java EE Patterns with the Spring Framework*. 1 ed.; Apress: 2008; p 344.

22. HighCharts https://www.highcharts.com/

23. Katajamaa, M.; Miettinen, J.; Oresic, M., MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics* **2006,** *22* (5), 634-6.

24. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M., MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **2010,** *11*, 395.

25. Tyanova, S.; Temu, T.; Cox, J., The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* **2016,** *11* (12), 2301-2319.

26. Watkins, D. M.; Sego, L. H.; Holmes, A. E.; Webb-Robertson, B. J.; White, A. M.; Wunschel, D. S.; Kreuzer, H.; Corley, C. D.; Tardiff, M. R., Assessing performance and tradeoffs of bioforensic signature systems. In *EEE International Conference on Technologies for Homeland Security*, IEEE: 2013; pp 304-309.

27. Webb-Robertson, B. J.; Wiberg, H. K.; Matzke, M. M.; Brown, J. N.; Wang, J.; McDermott, J. E.; Smith, R. D.; Rodland, K. D.; Metz, T. O.; Pounds, J. G.; Waters, K. M.,

Review, evaluation, and discussion of the challenges of missing value imputation for mass spectrometry-based label-free global proteomics. *J Proteome Res* **2015,** *14* (5), 1993-2001.

28. Callister, S. J.; Barry, R. C.; Adkins, J. N.; Johnson, E. T.; Qian, W. J.; Webb-Robertson, B. J.; Smith, R. D.; Lipton, M. S., Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J Proteome Res* **2006,** *5* (2), 277-86.

29. Matzke, M. M.; Waters, K. M.; Metz, T. O.; Jacobs, J. M.; Sims, A. C.; Baric, R. S.; Pounds, J. G.; Webb-Robertson, B. J., Improved quality control processing of peptide-centric LC-MS proteomics data. *Bioinformatics* **2011,** *27* (20), 2866-72.

30. Webb-Robertson, B. J.; Matzke, M. M.; Jacobs, J. M.; Pounds, J. G.; Waters, K. M., A statistical selection strategy for normalization procedures in LC-MS proteomics experiments through dataset-dependent ranking of normalization scaling factors. *Proteomics* **2011,** *11* (24), 4736-41.

31. Webb-Robertson, B. J.; McCue, L. A.; Waters, K. M.; Matzke, M. M.; Jacobs, J. M.; Metz, T. O.; Varnum, S. M.; Pounds, J. G., Combined statistical analyses of peptide intensities and peptide occurrences improves identification of significant peptides from MS-based proteomics data. *J Proteome Res* **2010,** *9* (11), 5748-56.

32. Alonso-Gutierrez, J.; Kim, E. M.; Batth, T. S.; Cho, N.; Hu, Q.; Chan, L. J.; Petzold, C. J.; Hillson, N. J.; Adams, P. D.; Keasling, J. D.; Garcia Martin, H.; Lee, T. S., Principal component analysis of proteomics (PCAP) as a tool to direct metabolic engineering. *Metab Eng* **2015,** *28*, 123-33.

33. Verhoeckx, K. C.; Bijlsma, S.; de Groene, E. M.; Witkamp, R. F.; van der Greef, J.; Rodenburg, R. J., A combination of proteomics, principal component analysis and

transcriptomics is a powerful tool for the identification of biomarkers for macrophage maturation in the U937 cell line. *Proteomics* **2004,** *4* (4), 1014-28.

34. Yu, L.; Snapp, R. R.; Ruiz, T.; Radermacher, M., Probabilistic principal component analysis with expectation maximization (PPCA-EM) facilitates volume classification and estimates the missing data. *J Struct Biol* **2010,** *171* (1), 18-30.

35. Hafen, R.; Rodland, K. D.; Gosink, L.; Kleese-Van Dam, K.; McDermott, J. E.; Cleveland, W. S., Trelliscope: A System for Detailed Visualization in the Deep Analysis of Large Complex Data. In *Large Data Analysis and Visualization (LDAV)*, 2013.

36. Trelliscope http://deltarho.org/trelliscope-0-9-7.

37. Begley, C. G.; Ioannidis, J. P., Reproducibility in science: improving the standard for basic and preclinical research. *Circ Res* **2015,** *116* (1), 116-26.

38. Gaudart, J.; Huiart, L.; Milligan, P. J.; Thiebaut, R.; Giorgi, R., Reproducibility issues in science, is P value really the only answer? *Proc Natl Acad Sci U S A* **2014,** *111* (19), E1934.

39. Gonzalez Martin-Moro, J., The science reproducibility crisis and the necessity to publish negative results. *Arch Soc Esp Oftalmol* **2017,** *92* (12), e75-e77.

40. Menachery, V. D.; Mitchell, H. D.; Cockrell, A. S.; Gralinski, L. E.; Yount, B. L., Jr.; Graham, R. L.; McAnarney, E. T.; Douglas, M. G.; Scobey, T.; Beall, A.; Dinnon, K., 3rd; Kocher, J. F.; Hale, A. E.; Stratton, K. G.; Waters, K. M.; Baric, R. S., MERS-CoV Accessory ORFs Play Key Role for Infection and Pathogenesis. *MBio* **2017,** *8* (4).

41. Menachery, V. D.; Schafer, A.; Burnum-Johnson, K. E.; Mitchell, H. D.; Eisfeld, A. J.; Walters, K. B.; Nicora, C. D.; Purvine, S. O.; Casey, C. P.; Monroe, M. E.; Weitz, K. K.; Stratton, K. G.; Webb-Robertson, B. M.; Gralinski, L. E.; Metz, T. O.; Smith, R. D.; Waters, K.

M.; Sims, A. C.; Kawaoka, Y.; Baric, R. S., MERS-CoV and H5N1 influenza virus antagonize antigen presentation by altering the epigenetic landscape. *Proc Natl Acad Sci U S A* **2018**.

42. Matzke, M. M.; Brown, J. N.; Gritsenko, M. A.; Metz, T. O.; Pounds, J. G.; Rodland, K. D.; Shukla, A. K.; Smith, R. D.; Waters, K. M.; McDermott, J. E.; Webb-Robertson, B. J., A comparative analysis of computational approaches to relative protein quantification using peptide peak intensities in label-free LC-MS proteomics experiments. *Proteomics* **2013,** *13* (3-4), 493-503.

43. Taverner, T.; Karpievitch, Y. V.; Polpitiya, A. D.; Brown, J. N.; Dabney, A. R.; Anderson, G. A.; Smith, R. D., DanteR: an extensible R-based tool for quantitative analysis of -omics data. *Bioinformatics* **2012,** *28* (18), 2404-6.

44. The UniProt, C., UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **2017,** *45* (D1), D158-D169.

45. Boeckhout, M.; Zielhuis, G. A.; Bredenoord, A. L., The FAIR guiding principles for data stewardship: fair enough? *Eur J Hum Genet* **2018,** *26* (7), 931-936.

46. Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J. W.; da Silva Santos, L. B.; Bourne, P. E.; Bouwman, J.; Brookes, A. J.; Clark, T.; Crosas, M.; Dillo, I.; Dumon, O.; Edmunds, S.; Evelo, C. T.; Finkers, R.; Gonzalez-Beltran, A.; Gray, A. J.; Groth, P.; Goble, C.; Grethe, J. S.; Heringa, J.; t Hoen, P. A.; Hooft, R.; Kuhn, T.; Kok, R.; Kok, J.; Lusher, S. J.; Martone, M. E.; Mons, A.; Packer, A. L.; Persson, B.; Rocca-Serra, P.; Roos, M.; van Schaik, R.; Sansone, S. A.; Schultes, E.; Sengstag, T.; Slater, T.; Strawn, G.; Swertz, M. A.; Thompson, M.; van der Lei, J.; van Mulligen, E.; Velterop, J.; Waagmeester, A.; Wittenburg, P.; Wolstencroft, K.; Zhao, J.; Mons, B., The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **2016,** *3*, 160018.

TOC IMAGE


Quality control
Normalization
Protein quantification
Statistics
Pattern discovery
Interactive displays
Reporting and results
P-mart
Panomics marketplace:
analyses via
reproducible
techniques