Published in final edited form as:

J Proteome Res. 2019 March 01; 18(3): 1426–1432. doi:10.1021/acs.jproteome.8b00840.

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

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

[†] Computing & Analytics Division, Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, Washington 99352, United States

[‡] Biological Sciences Division, Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, Washington 99352, United States

[§] Department of Cell, Developmental, and Cancer Biology, Oregon Health & Science University, Portland, Oregon 97221, United States

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 data sets 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.

Graphical Abstract

*Corresponding Author: 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.

The authors declare no competing financial interest.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.jproteome.8b00840](https://doi.org/10.1021/acs.jproteome.8b00840).

Supplemental Table S1: P-Mart functions (PDF)

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 are 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 data sets. 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 many downstream processing options and is a great solution for many proteomics data-processing tasks. The gaps that P-Mart fills with 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 the 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 straightforward 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) data¹⁷ has been previously described.¹⁸ Herein we describe P-Mart as a capability for use with user-uploaded data to allow the discovery and exploration of potential biomarker candidates for many user-defined global peak-intensity proteomics data sets.

2. SOFTWARE METHODS

P-Mart consists of two connected, yet distinct capabilities. The 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 the 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, then the user only needs to modify the web interface JavaScript code to include the new function/process as an option, or a user can 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 package²⁰ 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-Charts²² 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 those in the pmartRdata package in the format needed for upload into the web application as well as access to multiple cancer proteomics data sets and their clinical data generated through the CPTAC preloaded for exploration.^{17,18}

3. RESULTS AND DISCUSSION

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 Web site.

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 that are related to user uploads as well as to 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 with 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; have 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. 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 Preprocessing 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 the identification of samples with outlier behavior and normalization guidance.^{9,28–31} This allows for the generation of a data set with improved properties in the context of variance and missing data for the purposes of statistical analysis. Finally, statistics allow for

the comparison of two or more groups via the analysis of variance-based approaches for quantitative differences and a G-test for qualitative differences as well as options for multiple test corrections.

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 data sets 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 spread-sheets.

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 of the necessary information to reproduce an analysis exactly and all data files throughout the process where the underlying data have 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 pursuing 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 data set from a study on Middle Eastern respiratory syndrome coronavirus (MERS-CoV), where the objective was to analyze the response to MERS-CoV at 18 h post-infection in human Calu-3 cells.^{40,41} The data set includes three uninfected samples and nine infected samples and is available for download from the P-Mart Web site. 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 data set is “Condition”, which refers to whether the sample is infected, 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).

The preprocessing of the data included a sample outlier filter,²⁹ with a significance threshold of 0.001, which removes one sample, and a peptide coverage filter³¹ removes 7346 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 7346 peptides were

removed because the low occurrence indicates that no meaningful statistical comparison can be made. These data sets 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 data set 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 the 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 data set. 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).

Trelliscope allows a user to sort and filter these 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 downregulation in the infected group), reducing the displays to a subset of 28 proteins to examine (Figure 5, bottom). 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, and 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 the 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 page⁴⁴ 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 are 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 is the considerable number of proteins with unknown function or no obvious connection to explore either computationally or experimentally.

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 of 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.

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 is interoperable with other workflow 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 data set to be reproduced, and all data and statistics files, including intermediate analyses, are available to the user.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

P-Mart was developed at Pacific Northwest National Laboratory, a multiprogram national laboratory operated by Battelle for the U.S. Department of Energy under contract DE-AC06-76RL01830. This work has been supported by NCI grant U01-1CA184783 to B.-J.M.W.-R. The example data sets were developed with support by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number U19A106772, and proteomics data were collected in the Environmental Molecular Science Laboratory, a U.S. Department of Energy (DOE) national scientific user facility at Pacific Northwest National Laboratory in Richland, WA.

REFERENCES

- (1). Gajadhar AS; Johnson H; Slebos RJ; Shaddox K; Wiles K; Washington MK; Herline AJ; Levine DA; Liebler DC; White FM 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. [PubMed: 25670172]
- (2). Mertins P; Mani DR; Ruggles KV; Gillette MA; Clauser KR; Wang P; Wang X; Qiao JW; Cao S; Petralia F; Kawaler E; Mundt F; Krug K; Tu Z; Lei JT; Gatza ML; Wilkerson M; Perou CM; Yellapantula V; Huang KL; Lin C; McLellan MD; Yan P; Davies SR; Townsend RR; Skates SJ; Wang J; Zhang B; Kinsinger CR; Mesri M; Rodriguez H; Ding L; Paulovich AG; Fenyo D; Ellis MJ; Carr SA Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* 2016, 534 (7605), 55–62. [PubMed: 27251275]
- (3). Slebos RJ; Wang X; Wang X; Zhang B; Tabb DL; Liebler DC Proteomic analysis of colon and rectal carcinoma using standard and customized databases. *Sci. Data* 2015, 2, 150022. [PubMed: 26110064]
- (4). Zhang H; Liu T; Zhang Z; Payne SH; Zhang B; McDermott JE; Zhou JY; Petyuk VA; Chen L; Ray D; et al. Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell* 2016, 166 (3), 755–65. [PubMed: 27372738]
- (5). Duncan DT; Craig R; Link AJ 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. [PubMed: 16212440]

- (6). Tabb DL The SEQUEST family tree. *J. Am. Soc. Mass Spectrom* 2015, 26 (11), 1814–9. [PubMed: 26122518]
- (7). Xu T; Park SK; Venable JD; Wohlschlegel JA; Diedrich JK; Cociorva D; Lu B; Liao L; Hewel J; Han X; Wong CCL; Fonslow B; Delahunty C; Gao Y; Shah H; Yates JR 3rd ProLuCID: An improved SEQUEST-like algorithm with enhanced sensitivity and specificity. *J. Proteomics* 2015, 129, 16–24. [PubMed: 26171723]
- (8). Yates JR 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. [PubMed: 26286455]
- (9). Choi M; Chang CY; 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. [PubMed: 24794931]
- (10). Mueller LN; Brusniak MY; Mani DR; 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. [PubMed: 18173218]
- (11). Pendarvis K; Kumar R; Burgess SC; Nanduri B An automated proteomic data analysis workflow for mass spectrometry. *BMC Bioinf.* 2009, 10 (Suppl 11), S17.
- (12). Zauber H; Schulze WX Proteomics wants cRacker: automated standardized data analysis of LC-MS derived proteomic data. *J. Proteome Res.* 2012, 11 (11), 5548–55. [PubMed: 22978295]
- (13). Abouelhoda M; Issa SA; Ghanem M Tavaxy: integrating Taverna and Galaxy workflows with cloud computing support. *BMC Bioinf.* 2012, 13, 77.
- (14). 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 Res.* 2005, 15 (10), 1451–5. [PubMed: 16169926]
- (15). Oinn T; Addis M; Ferris J; Marvin D; Senger M; Greenwood M; Carver T; Glover K; Pocock MR; Wipat A; Li P Taverna: a tool for the composition and enactment of bioinformatics workflows. *Bioinformatics* 2004, 20 (17), 3045–54. [PubMed: 15201187]
- (16). Tyanova S; Temu T; Sinitcyn P; Carlson A; Hein MY; Geiger T; Mann M; Cox J The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* 2016, 13 (9), 731–40. [PubMed: 27348712]
- (17). NCI, National Cancer Institute Office of Cancer Clinical Proteomics Research. Data Portal. <https://cptac-data-portal.georgetown.edu/cptacPublic/> (accessed 2018).
- (18). Webb-Robertson BM; Bramer LM; Jensen JL; Kobold MA; Stratton KG; White AM; Rodland KD P-MartCancer-Interactive Online Software to Enable Analysis of Shotgun Cancer Proteomic Datasets. *Cancer Res.* 2017, 77 (21), e47–e50. [PubMed: 29092938]
- (19). Rserve Binary R server. <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, 1st 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. [PubMed: 16403790]
- (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 Bioinf.* 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. [PubMed: 27809316]
- (26). Watkins DM; Sego LH; Holmes AE; Webb-Robertson BJ; White AM; Wunschel DS; Kreuzer H; Corley CD; Tardiff MR Assessing Performance and Tradeoffs of Bioforensic Signature Systems. In *IEEE International Conference on Technologies for Homeland Security*; IEEE, 2013; pp 304–309.

- (27). Webb-Robertson BJ; Wiberg HK; Matzke MM; Brown JN; Wang J; McDermott JE; Smith RD; Rodland KD; Metz TO; Pounds JG; Waters KM 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. [PubMed: 25855118]
- (28). Callister SJ; Barry RC; Adkins JN; Johnson ET; Qian WJ; Webb-Robertson BJ; Smith RD; Lipton MS Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J. Proteome Res.* 2006, 5 (2), 277–86. [PubMed: 16457593]
- (29). Matzke MM; Waters KM; Metz TO; Jacobs JM; Sims AC; Baric RS; Pounds JG; Webb-Robertson BJ Improved quality control processing of peptide-centric LC-MS proteomics data. *Bioinformatics* 2011, 27 (20), 2866–72. [PubMed: 21852304]
- (30). Webb-Robertson BJ; Matzke MM; Jacobs JM; Pounds JG; Waters KM A statistical selection strategy for normalization procedures in LC-MS proteomics experiments through datasetdependent ranking of normalization scaling factors. *Proteomics* 2011, 11 (24), 4736–41. [PubMed: 22038874]
- (31). Webb-Robertson BJ; McCue LA; Waters KM; Matzke MM; Jacobs JM; Metz TO; Varnum SM; Pounds JG 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. [PubMed: 20831241]
- (32). Alonso-Gutierrez J; Kim EM; Bath TS; Cho N; Hu Q; Chan LJ; Petzold CJ; Hillson NJ; Adams PD; Keasling JD; Garcia Martin H; Lee TS Principal component analysis of proteomics (PCAP) as a tool to direct metabolic engineering. *Metab. Eng.* 2015, 28, 123–33. [PubMed: 25554074]
- (33). Verhoeckx KC; Bijlsma S; de Groene EM; Witkamp RF; van der Greef J; Rodenburg RJ 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. [PubMed: 15048983]
- (34). Yu L; Snapp RR; Ruiz T; Radermacher M Probabilistic principal component analysis with expectation maximization (PPCAEM) facilitates volume classification and estimates the missing data. *J. Struct. Biol.* 2010, 171 (1), 18–30. [PubMed: 20385241]
- (35). Hafen R; Rodland KD; Gosink L; Kleese-Van Dam K; McDermott JE; Cleveland WS Trelliscope: A System for Detailed Visualization in the Deep Analysis of Large Complex Data. 2013 IEEE Symposium on Large-Scale Data Analysis and Visualization (LDAV) 2013, 105.
- (36). Trelliscope. <http://deltarho.org/trelliscope-0-9-7>.
- (37). Begley CG; Ioannidis JP Reproducibility in science: improving the standard for basic and preclinical research. *Circ. Res.* 2015, 116 (1), 116–26. [PubMed: 25552691]
- (38). Gaudart J; Huiart L; Milligan PJ; 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. [PubMed: 24760820]
- (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. [PubMed: 28890235]
- (40). Menachery VD; Mitchell HD; Cockrell AS; Gralinski LE; Yount BL; Graham RL; McAnarney ET; Douglas MG; Scobey T; Beall A; Dinno K; Kocher JF; Hale AE; Stratton KG; Waters KM; Baric RS MERS-CoV Accessory ORFs Play Key Role for Infection and Pathogenesis. *mBio* 2017, 8 (4), No. e00665–17. [PubMed: 28830941]
- (41). Menachery VD; Schafer A; Burnum-Johnson KE; Mitchell HD; Eisfeld AJ; Walters KB; Nicora CD; Purvine SO; Casey CP; Monroe ME; Weitz KK; Stratton KG; Webb-Robertson BM; Gralinski LE; Metz TO; Smith RD; Waters KM; Sims AC; Kawaoka Y; Baric RS MERS-CoV and H5N1 influenza virus antagonize antigen presentation by altering the epigenetic landscape. *Proc. Natl. Acad. Sci. U. S. A* 2018, 115, E1012. [PubMed: 29339515]
- (42). Matzke MM; Brown JN; Gritsenko MA; Metz TO; Pounds JG; Rodland KD; Shukla AK; Smith RD; Waters KM; McDermott JE; Webb-Robertson BJ 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. [PubMed: 23019139]

- (43). Taverner T; Karpievitch YV; Polpitiya AD; Brown JN; Dabney AR; Anderson GA; Smith RD DanteR: an extensible Rbased tool for quantitative analysis of -omics data. *Bioinformatics* 2012, 28 (18), 2404–6. [PubMed: 22815360]
- (44). The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 2017, 45 (D1), D158–D169. [PubMed: 27899622]
- (45). Boeckhout M; Zielhuis GA; Bredenoord AL The FAIR guiding principles for data stewardship: fair enough? *Eur. J. Hum. Genet.* 2018, 26 (7), 931–936. [PubMed: 29777206]
- (46). Wilkinson MD; Dumontier M; Aalbersberg IJ; Appleton G; Axton M; Baak A; Blomberg N; Boiten JW; da Silva Santos LB; Bourne PE; Bouwman J; Brookes AJ; Clark T; Crosas M; Dillo I; Dumon O; Edmunds S; Evelo CT; Finkers R; Gonzalez-Beltran A; Gray AJ; Groth P; Goble C; Grethe JS; Heringa J; 't Hoen PA; Hooft R; Kuhn T; Kok R; Kok J; Lusher SJ; Martone ME; Mons A; Packer AL; Persson B; Rocca-Serra P; Roos M; van Schaik R; Sansone SA; Schultes E; Sengstag T; Slater T; Strawn G; Swertz MA; 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. [PubMed: 26978244]

Figure 1. 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.

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

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

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.

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

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. On the right is the collection of data that are available to the user, including normalized data, protein-level data, and all statistics that were performed.