ProteomeExpert: A Docker image based web-server for exploring, modeling, visualizing, and mining quantitative proteomic data sets

Tiansheng Zhu1,2,3, Hao Chen1,2, Xishan Yan4, Zhicheng Wu1,2,3, Xiaoxu Zhou1,2, Qi Xiao1,2, Weigang Ge1,2, Qiushi Zhang1,2, Chao Xu4, Luang Xu1,2, Guan Ruan1,2, Zhangzhi Xue1,2, Chunhui Yuan1,2,5, Guo-Bo Chen5,6 and Tiannan Guo1,2,7,*

1 Zhejiang Provincial Laboratory of Life Sciences and Biomedicine, Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, 18 Shilongshan Road, Hangzhou, Zhejiang, China. 2 Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, 18 Shilongshan Road, Hangzhou, Zhejiang Province, China. 3 Shanghai Key Lab of Intelligent Information Processing, and School of Computer Science, Fudan University, China. 4 College of Mathematics and Informatics, Digital Fujian Institute of Big Data Security Technology, Fujian Normal University, China. 5 Clinical Research Institute, Zhejiang Provincial People’s Hospital, People’s Hospital of Hangzhou Medical College, Hangzhou, Zhejiang, China.

*To whom correspondence should be addressed.

Abstract

The rapid progresses of high throughput sequencing technology-based omics and mass spectrometry (MS)-based proteomics such as data-independent acquisition (DIA) and its penetration to clinical studies have generated increasing number of proteomic data sets containing 100s-1000s samples. To analyze these quantitative proteomic data sets and other -omics data sets more efficiently and conveniently, we present a web server-based software tool ProteomeExpert implemented in Docker, which offers various analysis tools for experimental design, data mining, interpretation, and visualization of quantitative proteomic data sets. ProteomeExpert can be deployed on an operating system with Docker installed or with R language environment.

Availability and implementation: The Docker image of ProteomeExpert is freely available from https://hub.docker.com/r/lifeinfo/proteomeexpert. The source code of ProteomeExpert is also openly accessible at http://www.github.com/lifeinfo/ProteomeExpert. In addition, a demo server is provided at https://proteomic.shinyapps.io/peserver.

Contact: guotianm@westlake.edu.cn; chen.guobo@foxmail.com; yuanchunhui@westlake.edu.cn

SUPPLEMENTARY INFORMATION: SUPPLEMENTARY DATA ARE AVAILABLE AT BIOINFORMATICS ONLINE.

1 Introduction

Recent advances in liquid chromatographic mass spectrometry (LC-MS) based proteomics technology permit acquisition of hundreds to thousands of proteomics data sets in a relatively short time, especially using data-independent acquisition (DIA) MS strategy (Gillet, et al., 2012; Guo, et al., 2015; Yue, et al., 2020; Zhang, et al., 2020). The substantial increase of proteomics data during the last few years necessitates effective algorithms and software tools for automatic interpretation of the resultant quantitative data sets to obtain valuable biological insights or assist clinical diagnosis. Currently existing tools require programming skills, such as SWATH2stats (Blattmann, et al., 2016) and MSstats (Choi, et al., 2014). Other tools like mapDIA (Teo, et al., 2015) is a statistical analysis package for differentially expressed protein using DIA fragment-level intensities. PANDA-view (Chang, et al., 2018) and Perseus (Tyanova, et al., 2016), both depending on other packages, do not include functionalities in feature selection, peptide to protein inference, and experimental design. Therefore, proteomics analysis tools supporting various functions across tasks still needs replenishment.

Here we provide an easy-to-use web server-based comprehensive data analysis platform for quantitative proteomics data and other -omics data that covers experimental design, data pre-processing, data quality control, protein inference, statistics, feature selection, unsupervised learning, supervised learning and visualization. In order to facilitate installing and deployment, we release this platform as a Docker image except for GitHub, which is easy to install and share on operating systems such as Linux, Mac or Windows.

2 Methods

© The Author(s) 2021. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
In summary, we have developed ProteomeExpert to meet the requirement for processing large-scale quantitative proteomics data sets. Most, if not all, quantitative proteomic data sets can be fed into ProteomeExpert, including but not limited to DIA-MS with or without ion mobility, label-free or stable isotope labeling-based data-dependent acquisition (DDA) MS, and parallel reaction monitoring (PRM) MS, multiple reaction monitoring (SRM) MS. Transcriptomic and metabolomic data sets can also be processed by this tool. ProteomeExpert is compatible with other -omics tools by uploading their results in tab-delimited or comma-separated text file format and excel file. Moreover, ProteomeExpert includes comprehensive methods for data preprocessing, visualization, statistics and machine learning. It can be hosted within R shiny environment under Windows, Linux and Mac system or deploy in docker available as a web server.

Funding
This work is supported by grants from the National Key R&D Program of China (No. 2020YFE002200), Zhejiang Provincial Natural Science Foundation for Distinguished Young Scholars (LR19C050001), Hangzhou Agriculture and Society Advancement Program (20190101A04), National Natural Science Foundation of China (81972492) and National Science Fund for Young Scholars (21904107).

References
Blattmann, P., Heusel, M. and Aebersold, R. SWATH2stats: an r/bioconductor package to process and convert quantitative SWATH-MS proteomics data for downstream analysis tools. *PLoS One* 2016;11(4):e0153160.
Chang, C., et al. PANDA-view: an easy-to-use tool for statistical analysis and visualization of quantitative proteomics data. *Bioinformatics* 2018;34(20):3594-3596.
Choi, M., et al. MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* 2014;30(17):2524-2526.
Gillet, L.C., et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* 2012;11(6):O111 016717.
Guo, T., et al. Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. *Nat Med* 2015;21(4):407-413.
Teo, G., et al. mapDIA: preprocessing and statistical analysis of quantitative proteomics data from data independent acquisition mass spectrometry. *J Proteomics* 2015;129:108-120.
Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 1996;58(1):267-288.
Tyanova, S., et al. The perseus computational platform for comprehensive analysis of (pro)teomics data. *Nat Methods* 2016;13(9):731-740.
Yue, L., et al. Generating proteomic big data for precision medicine. *Proteomics* 2020:e1900358.
Zhang, F., et al. Data-Independent acquisition mass spectrometry-based proteomics and software tools: a glimpse in 2020. *Proteomics* 2020:e1900276.