OXFORD

Briefings in Bioinformatics, 2023, **24(1)**, 1–9
https://doi.org/10.1093/bib/bbac477
Problem Solving Protocol

Label-free proteome quantification and evaluation

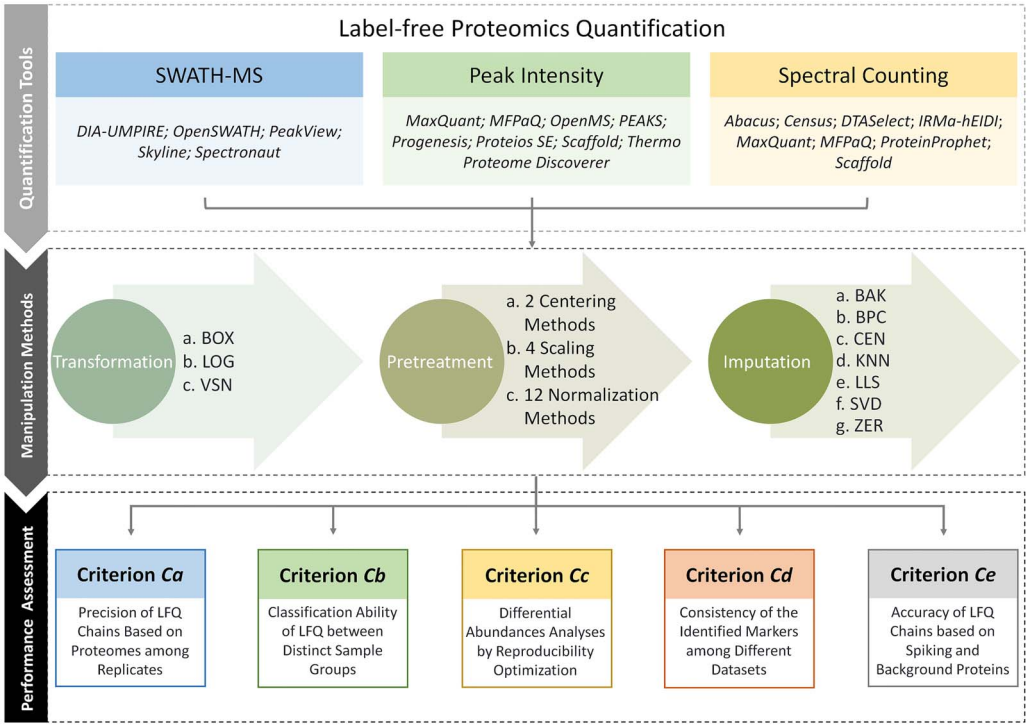
Jianbo Fu†, Qingxia Yang†, Yongchao Luo†, Song Zhang†, Jing Tang, Ying Zhang, Hongning Zhang, Hanxiang Xu and Feng Zhu

Corresponding author. Feng Zhu, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310050, China. E-mail: zhufeng@zju.edu.cn
†Jianbo Fu, Qingxia Yang, Yongchao Luo and Song Zhang contributed equally to this work as co-first authors.

Abstract

The label-free quantification (LFQ) has emerged as an exceptional technique in proteomics owing to its broad proteome coverage, great dynamic ranges and enhanced analytical reproducibility. Due to the extreme difficulty lying in an in-depth quantification, the LFQ chains incorporating a variety of transformation, pretreatment and imputation methods are required and constructed. However, it remains challenging to determine the well-performing chain, owing to its strong dependence on the studied data and the diverse possibility of integrated chains. In this study, an *R* package *EVALFQ* was therefore constructed to enable a performance evaluation on >3000 LFQ chains. This package is unique in (a) automatically evaluating the performance using multiple criteria, (b) exploring the quantification accuracy based on spiking proteins and (c) discovering the well-performing chains by comprehensive assessment. All in all, because of its superiority in assessing from multiple perspectives and scanning among over 3000 chains, this package is expected to attract broad interests from the fields of proteomic quantification. The package is available at https://github.com/idrblab/EVALFQ.

Graphical Abstract



Keywords: label-free quantification, proteomics, comprehensive evaluation, well-performing chains, *R* package

Jianbo Fu is a postdoctoral researcher of the College of Pharmaceutical Sciences in Zhejiang University, China. His current focus includes bioinformatics, computational biology and omics data analysis.
Qingxia Yang is an associate professor of the School of Geographic and Biologic Information in Nanjing University of Posts and Telecommunications, China. Her current focus includes bioinformatics, computational biology and omics data analysis.
Yongchao Luo, Song Zhang, Ying Zhang, Hongning Zhang and Hanxiang Xu are PhD/MD candidates of the College of Pharmaceutical Sciences in Zhejiang University, China. They are interested in bioinformatics.
Jing Tang is an associate professor of the Department of Bioinformatics in Chongqing Medical University, China. Her current focus includes bioinformatics and computational biology.
Feng Zhu is a tenured professor of the College of Pharmaceutical Sciences in Zhejiang University, China. His research laboratory (https://idrblab.org/) has been working in the fields of bioinformatics, OMIC-based drug discovery, system biology and medicinal chemistry.
Received: August 13, 2022. **Revised:** September 24, 2022. **Accepted:** October 8, 2022

Introduction

The proteomic techniques based on mass spectrometry (MS) have been widely used for revealing pathogenic mechanisms [1, 2], monitoring disease development [3], discovering diagnostic markers [4] and so on. There are mainly two approaches in proteomics including label-free quantification (LFQ) and label-based quantification using labeled proteins/peptides (as shown in Supplementary Figure S1 available online at http://bib.oxfordjournals.org/). Among these techniques, LFQ shows unique advantages in the discovery of proteins without the difficult and expensive use of stable isotope labeling [5]. LFQ has emerged exceptional in current proteomics owing to its broad proteome coverage [6], great dynamic ranges [7] and enhanced analytical reproducibility [8]. The LFQ has thus been applied to facilitate the discovery of promising therapeutic targets [9–11], investigate the variations in host microbiota [12, 13] and clarify the metabolism of targeted drugs [14, 15].

However, in-depth and high-quality quantification remains extremely challenging despite recent advances in LFQ technologies [16–19]. Particularly, the challenges are attributed to the asymmetrical distribution of protein abundances [20], the systematic bias inherent in technical processing [21] and the missing values inevitable in an OMICs-based technique [22]. To deal with these challenges, ≥3 transformation, ≥18 pretreatment and ≥7 imputation methods (shown in Supplementary Table S1 available online at http://bib.oxfordjournals.org/) are first constructed to transform data distribution, reduce variability among replicates and impute missing values, respectively; then, thousands of LFQ chains are generated (shown in Supplementary Method S1 available online at http://bib.oxfordjournals.org/) by sequentially integrating all these methods [23]. Because the quantification performance of each LFQ chain is found to be heavily dependent on the nature of the studied data [19], it is essential, but difficult, to identify the well-performing LFQ chains that can facilitate in-depth proteomic quantification [24].

So far, several powerful tools have been developed to analyze or evaluate label-free proteome quantifications [19, 25]. Among these, *Perseus* [25] provides diverse methods but does not conduct any evaluation; *Normalyzer* [26] provides the assessment of normalizations; *NAguideR* [22] evaluates various imputations. Since the LFQ chain involves not only transformation but also pretreatment and imputation, any evaluation that focuses only on a single type of method might not be able to reflect the overall performances of LFQ chains [16]. *ANPELA* is thus the first tool enabling performance assessments for the entire LFQ chain [23]. However, due to its web-based nature, it is impossible for *ANPELA* to conduct the comprehensive scanning of all possible LFQ chains (>3000 as shown in Supplementary Method S1 available online at http://bib.oxfordjournals.org/). Therefore, it is crucial to construct a new tool that realizes a comprehensive evaluation of the quantification performances by scanning all possible (>3000 as shown in Supplementary Method S1 available online at http://bib.oxfordjournals.org/) chains. However, no such tool is available yet.

In this work, an *R* package named *EVALFQ* was thus constructed, which enables the performance evaluation for 3128 LFQ chains. Particularly, this *R* package is able to (1) automatically evaluate the performances using multiple criteria, (2) explore the LFQ accuracy based on spiking proteins and (3) discover the well-performing LFQ chains by comprehensive assessment. Because of its superiority in assessing from multiple perspectives and scanning among thousand chains, *EVALFQ* is expected to attract broad interests from the fields of proteomic quantification.

Materials and methods

Collection of the benchmark datasets analyzed in this study

As shown in Table 1, the first study provided the sequential windowed acquisition of all theoretical fragment ion mass spectra (*SWATH-MS*) proteome data based on the 113 malignant tumorous tissues and 111 adjacent tissues from prostate cancer patients. And the proteomics dataset was quantified by *OpenSWATH*. In the second study, the *peak intensity* proteomic data were detected based on 101 malignant tissues and 98 paired non-tumor tissues from hepatocellular cancer patients and quantified by *MaxQuant*. In the third study, one *spectral counting* proteomic dataset was provided with 17 muscle homogenates of obese

Table 1. Detailed information of the studied benchmarks together with their corresponding acquisition techniques and adopted quantification tools

Study ID	Dataset description	Assumption held	Quantification tool (*mode of acquisition*)	References
PXD004691	113 malignant tumorous tissues and 111 adjacent tissues from prostate cancer patients	A, B, C	OpenSWATH (*SWATH-MS*)	[27]
PXD006512	101 malignant tissues and 98 paired non-tumor tissues from hepatocellular cancer patients	A, B, C	MaxQuant (*Peak Intensity*)	[28]
Katsanos *et al.*	17 muscle homogenates of obesity individuals and 16 muscle homogenates of lean individuals	A, B, C	Scaffold (*Spectral Counting*)	[29]
PXD014956	7 non-mouse proteins of 5 ratios spiked into 5 SGs of C2C12 mouse cell lysate	A, B, C	OpenSWATH (*SWATH-MS*)	[30]
PXD002025	7 wild-type male mice and 5 transgenic male mice overexpressing cytochrome P450 aromatase	A, B, C	Progenesis (*Peak Intensity*)	[31]
PXD002099	3 samples spiking with 2 fmol 48 UPS1 proteins and 3 samples spiking with 50 fmol UPS1	A, B, C	Progenesis (*Peak Intensity*)	[32]
IPX0001804000	103 malignant and 103 paired noncancerous adjacent tissues from lung cancer patients	A, B, C	MaxQuant (*Peak Intensity*)	[33]
PXD000672	18 tumorous tissue biopsies and 18 non-tumorous biopsies from renal cell carcinoma patients	A, B, C	OpenSWATH (*SWATH-MS*)	[34]
PXD002882	21 tissues from Crohn's disease patients and 10 tissues from healthy individuals	A, B, C	MaxQuant (*Peak Intensity*)	[35]

Three assumptions are held for the benchmarks: (A) all proteins are equally important; (B) the level of protein abundance is constant among all samples; (C) the intensities of most proteins are not changed under the studied condition. Study IDs were collected from PRIDE [36], iProX [37] and so on.

individuals and 16 muscle homogenates of lean individuals, which was quantified by *Scaffold*. The *SWATH-MS* proteomic data were provided in the fourth study based on seven non-mouse proteins of five ratios spiked into five sample groups (SGs) of C2C12 mouse cell lysate and quantified by *OpenSWATH*. In the fifth study, the *peak intensity* of the proteomic dataset consisted of seven wild-type male mice and five transgenic male mice overexpressing cytochrome P450 aromatase, which was preprocessed by *Progenesis*. The sixth study provided the *peak intensity* proteome data based on three samples spiking with 2 fmol 48 UPS1 proteins and three samples spiking with 50 fmol UPS1, which was quantified by *Progenesis*. The seventh study gave one *peak intensity* proteomic dataset with 103 malignant and 103 paired noncancerous adjacent tissues from lung cancer patients, which was quantified by *MaxQuant*. The eighth study provided the *SWATH-MS* proteome dataset containing 18 tumorous tissue biopsies and 18 non-tumorous biopsies from renal cell carcinoma patients, which was preprocessed by *OpenSWATH*. Finally, the last study described *peak intensity* proteome data of 21 tissues from Crohn's disease patients and 10 tissues from healthy individuals, which was preprocessed by *MaxQuant*. Moreover, all processed datasets (Table 1) can be downloaded at https://idrblab.org/evalfq/download/All.Processed.Datasets.xlsx.

To test the functions of *EVALFQ*, a variety of benchmarks were collected from renowned proteomic databases [36,37] and a recent publication [29]. In total, nine benchmarks were assessed in this study. As shown in Table 1, the quantification tools and the acquisition modes of these benchmarks were very diverse, which included: *SWATH-MS*, *Peak Intensity* and *Spectral Counting* (quantified by four different quantification software tools). Since the application of each method should obey their individual assumption (as shown in Supplementary Table S1 available online at http://bib.oxfordjournals.org/), three assumptions held for the analyzed benchmarks were also provided in Table 1. These assumptions included: (A) all proteins are equally important; (B) the levels of protein abundances are constant among samples and (C) the intensities of most proteins are unchanged under the studied condition. All data were from PRIDE [36] and iProX [37].

Construction of the LFQ chain by sequential method integration

As shown in Supplementary Table S1 available online at http://bib.oxfordjournals.org/, 3 transformation, 18 pretreatment (centering, scaling and normalization) and 7 imputation methods are available for transforming data distribution, reducing variabilities among replicates and imputing missing values. There were three types of assumptions for these 18 pretreatment methods: (1) all proteins should be equally important, which is the prerequisite for applying the centering and scaling approaches; (2) the level of protein abundance should be constant among all samples, which is the priori hypothesis for some normalization methods including MEA, MED, MAD and TIC and (3) the intensities of the vast majority of the proteins should be unchanged under the studied condition, which is demanded by some other normalization methods including CYC, EIG, LIN, LOW, PQN, QUA, RLR and TMM. LFQ chain is thus composed of five sequentially integrated steps: transformation, centering, scaling, normalization and imputation (Supplementary Table S1 available online at http://bib.oxfordjournals.org/). In other words, a random, comprehensive and sequential integration of all 28 methods can result in 3128 LFQ chains of five steps. The detailed information on the construction of the LFQ chain, by sequentially integrating the methods of five steps, is explicitly explained in Supplementary Method S1 available online at http://bib.oxfordjournals.org/, and the abbreviation of each method was in Supplementary Table S1 available online at http://bib.oxfordjournals.org/.

Independent criteria enabling assessment from multiple perspectives

Five independent criteria available for assessing the performances of LFQ chains were provided in *EVALFQ*. (1) Criterion *Ca*: precision of LFQ based on proteomes among replicates. Criterion *Ca* was popular for evaluating the capacity of reducing intragroup variations among different samples. Its value is larger than 0 and a lower value denotes a more thorough removal of variations. (2) Criterion *Cb*: classification ability of LFQ between distinct SGs. Criterion *Cb* was applied for evaluating the ability to classify multiple classes. Its value is between 0 and 1, and a large value (close to 1) denotes high classification performance. (3) Criterion *Cc*: differential abundances analyses by reproducibility optimization. Criterion *Cc* was effective for evaluating the reproducibility of differential abundances. Its value is between 0 and 1, and a large value (close to 1) denotes high reproducibility. (4) Criterion *Cd*: consistency of the identified markers among different datasets. Criterion *Cd* was used to assess the robustness of markers among different datasets. Its value is between 0 and 1. If it is close to 1, it refers to the highest robustness of the identified markers. (5) Criterion *Ce*: LFQ accuracy based on spiking and background proteins. Criterion *Ce* can reflect the degree of correspondence between the processed data and references. When there were minimized variations, the medians of differences between the processed data and references were equal to 0. All five criteria were well established by the previous publications. Detailed information on all those assessing criteria is provided in Supplementary Method S2 available online at http://bib.oxfordjournals.org/.

Based on these independent criteria, *EVALFQ* enabled the performance assessment of LFQ chains from multiple perspectives [38]. Users can apply one or more criteria discussed above for assessing LFQ chains by parameter settings. Particularly, the performance of 3128 potential LFQ chains can first be ranked separately using each criterion. Different ranking numbers were assigned to each method by the corresponding criteria. Then, an overall ranking of a studied LFQ chain was defined by the sum of the multiple ranking numbers under multiple criteria. The smaller the sum is, the higher an LFQ chain ranks.

Installation and configuration steps of *EVALFQ* package

To ensure the correct installation of *EVALFQ* package, the following tools must be installed in the order listed below. First, install the R language from the R website. The process of installing R depends on the type of operating system. Second, install the RStudio from the RStudio website. Third, install various *R* packages required as dependencies, including *affy*, *Biobase*, *BiocGenerics*, *impute*, *limma*, *metabolomics*, *pcaMethods*, *ProteoMM*, *ROTS* and *vsn*. Some packages can be downloaded from GitHub (https://github.com/), some packages can be downloaded from Bioconductor (http://bioconductor.org/) and the remaining packages are available on CRAN (https://cran.r-project.org). The detailed information on the required packages is shown in Supplementary Method S3 available online at http://bib.oxfordjournals.org/. Fourth, install the *devtools* package and load this package. Fifth, install the *EVALFQ* package. During the installation of *EVALFQ*, an error message stating 'ERROR:

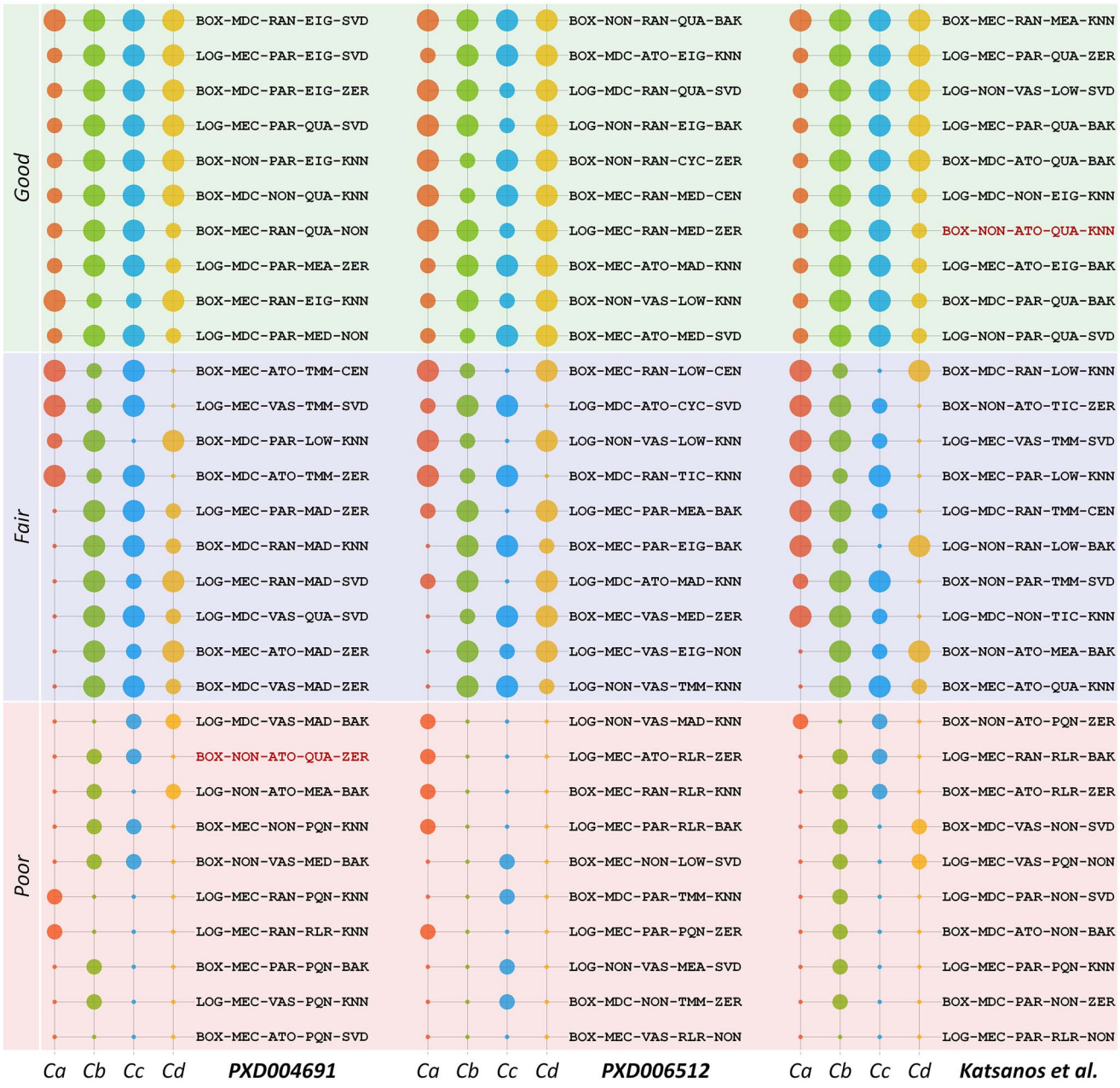


Figure 1. The schematic diagram illustrating the way to identify the well-performing LFQ chains based on multiple criteria. The performances under each criterion (*Ca–Cd*) were shown using the diameters of each circle (long, medium and short diameters indicated 'superior', 'good' and 'poor' performances under each criterion). The overall performances of each LFQ chain were classified into three groups: 'Good ' (performing consistently 'Superior/Good' under all criteria, colored in light green background), 'Fair' (performances among criteria fluctuating dramatically, colored in light blue background) and 'Poor' (performing non-'Superior' under all criteria, colored in pink background). The abbreviations of each method in the LFQ chains are described in Supplementary Table S1 available online at http://bib.oxfordjournals.org/, and the datasets analyzed here included PXD004691, PXD006512 and Katsanos *et al.* (shown in Table 1). The *BOX-NON-ATO-QUA-KNN* (highlighted in red bold font) showed 'Good/Superior' performances under all criteria for Katsanos *et al.* but performed 'Poor' for PXD004691 (in red bold font).

dependency 'Package Name' is not available for package '*EVALFQ*' indicates that the dependency package named 'Package Name' is not installed successfully. Users can reinstall that specific package using the instructions in Supplementary Method S3 available online at http://bib.oxfordjournals.org/.

Implementation and features in the EVALFQ package

EVALFQ is an *R* package based on the General Public License (GPLv3) and is available for free through the GitHub repository. It evaluates the performance of LFQ chains based on multiple criteria, and identifies the well-performing one(s) by a comprehensive scanning of >3000 potential LFQ chains. The manual of the *EVALFQ R* package developed in this study is shown in Supplementary Method S3 available online at http://bib.oxfordjournals.org/. A typical assessment in *EVALFQ* includes three sequential steps:

(*S1*) *File Selection*. The input data could be a variety of files generated by 18 popular quantification software tools, which are based on three modes of acquisition (*SWATH-MS*, *Peak intensity* and *Spectral counting*). *SWATH-MS* is a measurement that allows for the detection and quantification of nearly all detectable

Table 2. Performances of representative LFQ chains on three datasets (collectively assessed by four criteria)

ID	LFQ Chains	Ca	Cb	Cc	Cd
PXD004691	BOX-MDC-RAN-EIG-SVD	9.56E-02	5.13E-01	4.04E-03	2.18E+01
	LOG-MEC-PAR-EIG-SVD	2.20E-01	6.38E-01	3.48E-03	2.57E+01
	BOX-MEC-ATO-TMM-CEN	3.52E-17	5.09E-01	7.90E-04	2.49E+00
	LOG-MEC-VAS-TMM-SVD	1.10E-13	5.09E-01	1.23E-03	3.04E+00
	LOG-MEC-VAS-PQN-KNN	8.56E+01	5.09E-01	8.05E-03	1.60E+01
	BOX-MEC-ATO-PQN-SVD	8.27E-01	5.00E-01	7.14E-03	1.63E+01
PXD006512	BOX-NON-RAN-QUA-BAK	9.19E-02	8.65E-01	9.92E-05	6.26E+01
	BOX-MDC-ATO-EIG-KNN	4.74E-01	1.00E+00	8.93E-05	5.35E+01
	BOX-MEC-RAN-LOW-CEN	5.26E-03	7.43E-01	6.36E-04	7.82E+01
	LOG-MDC-ATO-CYC-SVD	4.93E-01	8.56E-01	9.42E-05	2.97E+01
	BOX-MDC-NON-TMM-ZER	4.73E+01	5.05E-01	3.38E-04	2.20E+00
	BOX-MEC-VAS-RLR-NON	1.37E+00	5.05E-01	8.93E-04	6.05E-01
Katsanos *et al.*	BOX-MEC-RAN-MEA-KNN	2.05E-01	5.76E-01	1.87E-02	1.95E+01
	LOG-MEC-PAR-QUA-ZER	6.58E-01	5.76E-01	1.19E-02	1.80E+01
	BOX-MDC-RAN-LOW-KNN	5.30E-03	5.45E-01	1.67E-01	3.27E+01
	BOX-NON-ATO-TIC-ZER	1.00E-03	5.76E-01	2.00E-02	1.50E+01
	BOX-MDC-PAR-NON-ZER	8.01E-01	5.16E-01	8.65E-02	1.57E+01
	LOG-MEC-PAR-RLR-NON	3.69E+00	5.15E-01	4.69E-02	1.54E+01

All performances were ranked and colored to 'Superior', 'Good' and 'Poor' using the background of dark blue, light blue and light red, respectively. The abbreviations of LFQ chains are described in Supplementary Table S1 available online at http://bib.oxfordjournals.org/. As shown, the performances of different chains varied substantially. Particularly, the performances of some chains were consistently 'Superior' (depicted by dark blue background under all criteria); the performances of another some were always 'Poor' (highlighted by light red background under all criteria); the remaining representative chains performed 'Good' (light blue) or 'Superior' under some criteria but exhibited 'Poor' performance under the others.

peptide fragments in a sample [39]. A list of software is acquired for preprocessing the data of *SWATH-MS*, including *DIA-UMPIRE*, *OpenSWATH*, *PeakView*, *Skyline* and *Spectronaut*. Peak intensity is an acquisition method that allows the comparison of peptide signal intensities at the analytical level of LC–MS [40]. A list of software is acquired for preprocessing the data of peak intensity, including *MaxQuant*, *MFPaQ*, *OpenMS*, *PEAKS*, *Progenesis*, *Proteios SE*, *Scaffold* and *Thermo Proteome Discoverer*. Spectral counting method can count the number of identified peptides or obtained spectral fragments [40,41]. A list of software is acquired for preprocessing the data of spectral counting, including *Abacus*, *Census*, *DTASelect*, *IRMa-hEIDI*, *MaxQuant*, *MFPaQ*, *ProteinProphet* and *Scaffold*. The file format generated by these tools could be readily found in the User Manual. The function *PrepareInuputFiles*() allows the readers to upload their output files generated by those quantification software tools. Detailed description can be found in the online User Manual.

(S2) *Data Analysis*. One of the key features of *EVALFQ* lines is its capacity of providing diverse methods, scanning thousands of LFQ chains and finally discovering the well-performing one. Therefore, the performance evaluation of each LFQ chain based on multiple criteria was realized by two functions: *lfqevalueall*() for treating the data without spiking proteins, and *lfqspikedall*() capable of quantifying the data with spiking proteins. The output files generated by these functions were explicitly described in Supplementary Table S2 available online at http://bib.oxfordjournals.org/.

(S3) *Results Visualization*. With the implementation of function *lfqvisualize*(), a heatmap chart that ranks all 3128 potential LFQ chains was generated to display the performance under each criterion using different colors (dark blue, light blue and pink were used to indicate the performance of 'Superior', 'Good' and 'Poor', respectively). As a schematic diagram, Figure 1 showed a number of chains of the representative performance (10 for consistently 'Superior/Good', 10 for no 'Superior' under any criterion and 10

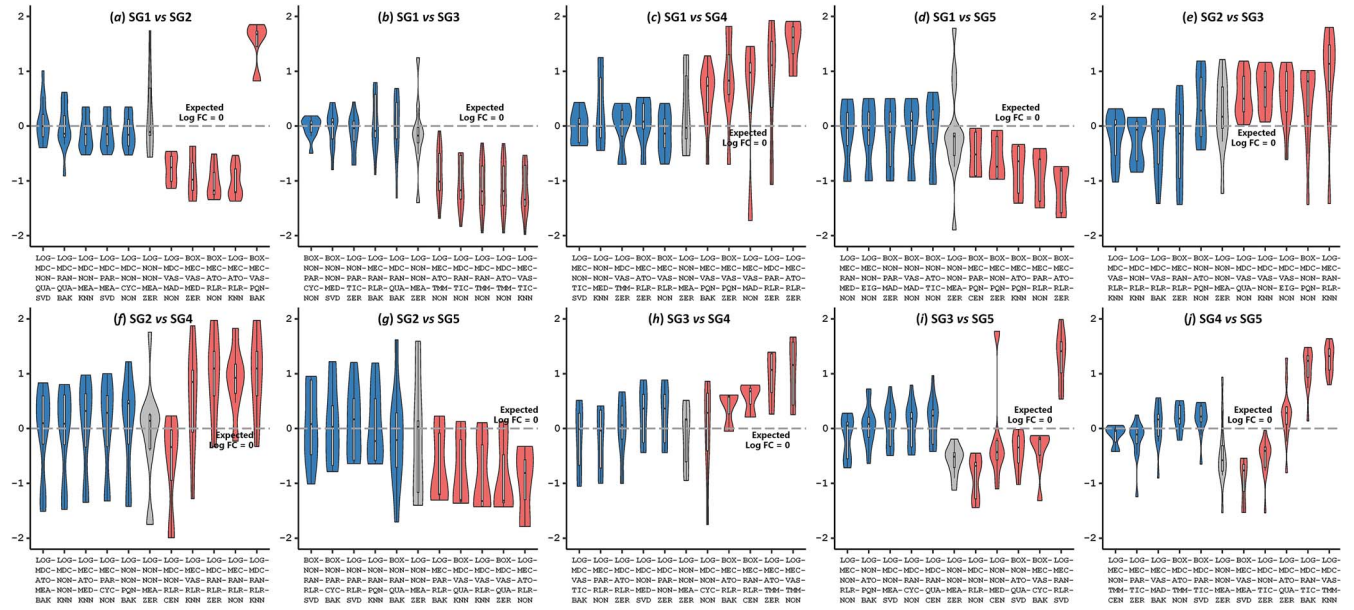


Figure 2. Accuracies of representative LFQ chains that were assessed using five SGs of various spiking protein concentrations. (**A**) SG1 versus SG2, (**B**) SG1 versus SG3, (**C**) SG1 versus SG4, (**D**) SG1 versus SG5, (**E**) SG2 versus SG3, (**F**) SG2 versus SG4, (**G**) SG2 versus SG5, (**H**) SG3 versus SG4, (**I**) SG3 versus SG5 and (**J**) SG4 versus SG5. All samples in each group were collected from the benchmark PXD014956 [30]. The horizontal dashed line represented an expected logFC of zero (completely consistent with spiking proteins). Violin plots in GRAY indicated the LFQ chain that was adopted in the original publication of the benchmark PXD014956 [30]. As illustrated, there are chances to identify an LFQ chain of better performance than the original one (colored in dark blue), and there are still great probabilities to discover the LFQ chain of much worse performance than the original one (colored in red). The abbreviations in each LFQ chain are explicitly described in Table 1.

for very diverse performances among different criteria) for three benchmark datasets. All in all, *EVALFQ* is featured by its capability of identifying well-performing LFQ chains.

Results and discussion

Great variances among the performances of different LFQ chains

To assess the ability of *EVALFQ* in identifying the well-performing LFQ chain, three benchmarks of various modes of acquisitions, PXD004691 [27], PXD006512 [28] and Katsanos *et al.* [29] acquired based on *SWATH-MS*, *Peak Intensity* and *Spectral Counting*, respectively, were collected (shown in Table 1). Table 2 gave the performances of the representative LFQ chains for each dataset (collectively accessed using multiple criteria). As shown, for all these benchmarks, the performances of different chains varied greatly (Case Study S1 explicitly described the variances in the performances of different LFQ chains among criteria). It is likely that some chains may not work well under one or more criteria for the proteomic data quantified by different quantification tools, and it is therefore critical to systematically assess the performance of each LFQ chain based on multiple criteria as realized in *EVALFQ* package.

Well-performing LFQ chains identified based on multiple criteria

The performances of LFQ chains varied greatly under each independent criterion. As illustrated in Supplementary Figures S2A, S3A, S4A, S5A and S6A available online at http://bib.oxfordjournals.org/, the performances of all LFQ chains were ranked solely by Criterion *Ca*, Criterion *Cb*, Criterion *Cc*, Criterion *Cd* and Criterion *Ce* using the benchmark PXD002025 [31], PXD000672 [34], PXD002882 [35], IPX0001804000 [33] and PXD002099 [32], respectively. With the decrease in the performances under one criterion, there were no clear trends in the performances under other criteria. Moreover, as shown in Supplementary Figures S2B, S3B, S4B, S5B and S6B available online at http://bib.oxfordjournals.org/, top-10 LFQ chains identified solely based on Criterion *Ca*, Criterion *Cb*, Criterion *Cc*, Criterion *Cd* or Criterion *Ce* was provided, respectively. For the top-performing chains ranked by one criterion, the performances assessed by other criteria varied greatly. Detailed descriptions for the performances of LFQ chains based on five criteria are provided in Supplementary Method S2 available online at http://bib.oxfordjournals.org/.

To discover the well-performing LFQ chains, *EVALFQ* package integrates a strategy that enables a comprehensive assessment of the performances of >3000 LFQ chains using multiple criteria. As illustrated in Figure 1 and Supplementary Table S3 available online at http://bib.oxfordjournals.org/, this strategy was applied to various benchmarks (PXD004691 [27], PXD006512 [28], Katsanos *et al.* [29]), and the detailed assessment results were provided. Based on the above results, Case Study S2 explicitly showed the necessity of conducting comprehensive performance evaluation among all LFQ chains to identify the ones of good performances. In other words, by comprehensively ranking performances using multiple criteria, the difficulty in finding well-performing LFQ chains can be significantly reduced, which can extensively facilitate the study in current proteomic analysis.

Enhanced accuracy realized by spiking and background proteins

Spiking proteins were well-studied elements that facilitate performance assessment of proteomic quantification [42]. Particularly, the level of spiking protein should be quantified as an expected abundance ratio, while the level of background protein remains unbiased variation [31,43]. By adopting spiking proteins as the golden standard, Case Study S3 provided a vivid description of the

way to enhance quantification accuracy based on both spiking and background proteins. As illustrated in Figure 2, taking the pair of sample-1 versus sample-2 as an example, with reference to the chain (*LOG-NON-NON-MEA-ZER*) adopted in the original publication (the violin plots in gray), five additional chains were found well-performing in guaranteeing the deviation from the expected concentration ratio of spiking proteins (the violin plots in dark blue). All in all, *EVALFQ* showed its ability to verify the accuracy of the LFQ chain by maintaining the true biological variation of spiking proteins, which enabled the identification of the most appropriate LFQ chains using the expected protein abundance ratio as a gold standard. With the advent of the big data era [44–46], *EVALFQ* could be the essential supplement to the available proteomic tools [47–56] and may thus attract broad interests from the related fields.

EVALFQ package is an essential tool for analyzing proteomic data with a lot of advantages. This package covers a wide range of available processing methods, but these methods are limited to those commonly used. To improve the effectiveness of data processing, new methods for proteomics should be implemented constantly, which cannot be automatically completed. Therefore, our team will update this package continuously and manually in the future. Additionally, it takes a lot of time to sequentially evaluate thousands of LFQ chains by utilizing *EVALFQ* package. Especially for large-scale proteomic datasets, the calculating time spent applying this package varies from hundreds to thousands of minutes. In the future, the computational efficiency of *EVALFQ* package will be improved substantially.

Conclusion

This study constructed an *R* package to enable a handy processing and comprehensive evaluation of the LFQ chains. Particularly, it is able to (1) automatically evaluate the performances from multiple perspectives, (2) explore the quantification accuracy using spiking proteins and (3) discover the well-performing quantifications by scanning >3000 LFQ chains. Due to the assessment of over 3000 chains based on multiple criteria from multiple perspectives, this package could discover the well-performing LFQ chains. This package could be an essential and powerful supplement to the available proteomic tools. The *EVALFQ* package is available at https://github.com/idrblab/EVALFQ.

Key Points

- The *EVALFQ* package was developed based on R code for label-free proteome quantification and evaluation.
- The performances of >3000 LFQ chains can be automatically evaluated using multiple criteria from multiple perspectives.
- The quantification accuracy was explored based on spiking proteins in the *EVALFQ* package.
- Using the *EVALFQ* package, the well-performing LFQ chains can be discovered by comprehensive assessment.

Data availability

The implemented code and experimental dataset are available online at https://github.com/idrblab/EVALFQ.

Supplementary Data

Supplementary data are available online at https://academic.oup.com/bib.

Funding

Funded by the National Natural Science Foundation of China (81872798 & U1909208); Natural Science Foundation of Zhejiang Province (LR21H300001); Natural Science Foundation of Jiangsu Province (BK20210597); Leading Talent of the 'Ten Thousand Plan' - National High-Level Talents Special Support Plan of China; Fundamental Research Fund of Central University (2018QNA7023); Key R&D Program of Zhejiang Province (2020C03010); National Key R&D Program of China Synthetic Biology Research (2019YFA0905900); "Double Top-Class" University (181201*194232101); Space Exploration Breeding Grant of Qian Xuesen Lab (TKTSPY-2020-04-03); Scientific Research Grant of Ningbo University (215-432000282); Ningbo Top Talent Proj (215-432094250); Shenzhen Municipal Government grant (NO.2019156, JCYJ20170413113448742, NO.201901); Department of Science & Technology of Guangdong Province (2017B030314083); Westlake Laboratory (Westlake Laboratory of Life Sciences and Biomedicine), Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare, Alibaba Cloud, Information Technology Center of Zhejiang University.

References

1. Leng L, Cao R, Ma J, *et al.* Pathological features of COVID-19-associated lung injury: a preliminary proteomics report based on clinical samples. *Signal Transduct Target Ther* 2020;**5**:240.
2. Huang M, Modeste E, Dammer E, *et al.* Network analysis of the progranulin-deficient mouse brain proteome reveals pathogenic mechanisms shared in human frontotemporal dementia caused by GRN mutations. *Acta Neuropathol Commun* 2020;**8**:163.
3. Egerstedt A, Berntsson J, Smith ML, *et al.* Profiling of the plasma proteome across different stages of human heart failure. *Nat Commun* 2019;**10**:5830.
4. Hoshino A, Kim HS, Bojmar L, *et al.* Extracellular vesicle and particle biomarkers define multiple human cancers. *Cell* 2020;**182**:1044–61.
5. Deracinois B, Flahaut C, Duban-Deweer S, *et al.* Comparative and quantitative global proteomics approaches: an overview. *Proteomes* 2013;**1**:180–218.
6. Shu T, Ning W, Wu D, *et al.* Plasma proteomics identify biomarkers and pathogenesis of COVID-19. *Immunity* 2020;**53**:1108–22.
7. Doblmann J, Dusberger F, Imre R, *et al.* apQuant: accurate label-free quantification by quality filtering. *J Proteome Res* 2019;**18**:535–41.
8. Yu Q, Wu W, Tian X, *et al.* Unraveling proteome changes of Holstein beef *M. semitendinosus* and its relationship to meat discoloration during post-mortem storage analyzed by label-free mass spectrometry. *J Proteomics* 2017;**154**:85–93.
9. Gillette MA, Satpathy S, Cao S, *et al.* Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma. *Cell* 2020;**182**:200–25.
10. Zou Q, Ma Q. The application of machine learning to disease diagnosis and treatment. *Math Biosci* 2019;**320**:108305.
11. Wang L, Xuan Z, Zhou S, *et al.* A novel model for predicting lncRNA-disease associations based on the lncRNA-miRNA-disease interactive network. *Curr Bioinform* 2019;**14**:269–78.
12. Gehrig JL, Venkatesh S, Chang HW, *et al.* Effects of microbiota-directed foods in gnotobiotic animals and undernourished children. *Science* 2019;**365**:eaau4732.

13. Qu K, Guo F, Liu X, *et al.* Application of machine learning in microbiology. *Front Microbiol* 2019;**10**:827.
14. Martin JK, Sheehan JP, Bratton BP, *et al.* A dual-mechanism antibiotic kills Gram-negative bacteria and avoids drug resistance. *Cell* 2020;**181**:1518–32.
15. Ji J, Tang J, Xia K, *et al.* LncRNA in tumorigenesis microenvironment. *Curr Bioinform* 2019;**14**:640–1.
16. Tang J, Fu J, Wang Y, *et al.* Simultaneous improvement in the precision, accuracy, and robustness of label-free proteome quantification by optimizing data manipulation chains. *Mol Cell Proteomics* 2019;**18**:1683–99.
17. Zhang B, Kall L, Zubarev RA. DeMix-Q: quantification-centered data processing workflow. *Mol Cell Proteomics* 2016;**15**:1467–78.
18. Muller F, Fischer L, Chen ZA, *et al.* On the reproducibility of label-free quantitative cross-linking/mass spectrometry. *J Am Soc Mass Spectrom* 2018;**29**:405–12.
19. Navarro P, Kuharev J, Gillet LC, *et al.* A multicenter study benchmarks software tools for label-free proteome quantification. *Nat Biotechnol* 2016;**34**:1130–6.
20. Aguilan JT, Kulej K, Sidoli S. Guide for protein fold change and p-value calculation for non-experts in proteomics. *Mol Omics* 2020;**16**:573–82.
21. Schubert OT, Rost HL, Collins BC, *et al.* Quantitative proteomics: challenges and opportunities in basic and applied research. *Nat Protoc* 2017;**12**:1289–94.
22. Wang S, Li W, Hu L, *et al.* NAguideR: performing and prioritizing missing value imputations for consistent bottom-up proteomic analyses. *Nucleic Acids Res* 2020;**48**:e83.
23. Tang J, Fu J, Wang Y, *et al.* ANPELA: analysis and performance assessment of the label-free quantification workflow for metaproteomic studies. *Brief Bioinform* 2020;**21**:621–36.
24. Chawade A, Sandin M, Teleman J, *et al.* Data processing has major impact on the outcome of quantitative label-free LC-MS analysis. *J Proteome Res* 2015;**14**:676–87.
25. Tyanova S, Temu T, Sinitcyn P, *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* 2016;**13**:731–40.
26. Chawade A, Alexandersson E, Levander F. Normalyzer: a tool for rapid evaluation of normalization methods for omics data sets. *J Proteome Res* 2014;**13**:3114–20.
27. Zhu Y, Weiss T, Zhang Q, *et al.* High-throughput proteomic analysis of FFPE tissue samples facilitates tumor stratification. *Mol Oncol* 2019;**13**:2305–28.
28. Jiang Y, Sun A, Zhao Y, *et al.* Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. *Nature* 2019;**567**:257–61.
29. Kras KA, Langlais PR, Hoffman N, *et al.* Obesity modifies the stoichiometry of mitochondrial proteins in a way that is distinct to the subcellular localization of the mitochondria in skeletal muscle. *Metabolism* 2018;**89**:18–26.
30. Barkovits K, Pacharra S, Pfeiffer K, *et al.* Reproducibility, specificity and accuracy of relative quantification using spectral library-based data-independent acquisition. *Mol Cell Proteomics* 2020;**19**:181–97.
31. Valikangas T, Suomi T, Elo LL. A systematic evaluation of normalization methods in quantitative label-free proteomics. *Brief Bioinform* 2018;**19**:1–11.
32. Pursiheimo A, Vehmas AP, Afzal S, *et al.* Optimization of statistical methods impact on quantitative proteomics data. *J Proteome Res* 2015;**14**:4118–26.
33. Xu JY, Zhang C, Wang X, *et al.* Integrative proteomic characterization of human lung adenocarcinoma. *Cell* 2020;**182**:245–61.
34. Guo T, Kouvonen P, Koh CC, *et al.* Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. *Nat Med* 2015;**21**:407–13.
35. Mottawea W, Chiang CK, Muhlbauer M, *et al.* Altered intestinal microbiota-host mitochondria crosstalk in new onset Crohn's disease. *Nat Commun* 2016;**7**:13419.
36. Perez-Riverol Y, Csordas A, Bai J, *et al.* The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* 2019;**47**:D442–50.
37. Ma J, Chen T, Wu S, *et al.* iProX: an integrated proteome resource. *Nucleic Acids Res* 2019;**47**:D1211–7.
38. Fu J, Zhang Y, Wang Y, *et al.* Optimization of metabolomic data processing using NOREVA. *Nat Protoc* 2022;**17**:129–51.
39. Anjo SI, Santa C, Manadas B. SWATH-MS as a tool for biomarker discovery: from basic research to clinical applications. *Proteomics* 2017;**17**:1600278.
40. Nikolov M, Schmidt C, Urlaub H. Quantitative mass spectrometry-based proteomics: an overview. *Methods Mol Biol* 2012;**893**:85–100.
41. Arike L, Peil L. Spectral counting label-free proteomics. *Methods Mol Biol* 2014;**1156**:213–22.
42. Ramus C, Hovasse A, Marcellin M, *et al.* Benchmarking quantitative label-free LC-MS data processing workflows using a complex spiked proteomic standard dataset. *J Proteomics* 2016;**132**:51–62.
43. Risso D, Ngai J, Speed TP, *et al.* Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol* 2014;**32**:896–902.
44. Wang Y, Zhang S, Li F, *et al.* Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Res* 2020;**48**:D1031–41.
45. Yin J, Sun W, Li F, *et al.* VARIDT 1.0: variability of drug transporter database. *Nucleic Acids Res* 2020;**48**:D1042–50.
46. Yin J, Li F, Zhou Y, *et al.* INTEDE: interactome of drug-metabolizing enzymes. *Nucleic Acids Res* 2021;**49**:D1233–43.
47. Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* 2016;**11**:2301–19.
48. Bruderer R, Bernhardt OM, Gandhi T, *et al.* Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol Cell Proteomics* 2015;**14**:1400–10.
49. Tsou CC, Avtonomov D, Larsen B, *et al.* DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat Methods* 2015;**12**:258–64.
50. Tang J, Mou M, Wang Y, *et al.* MetaFS: performance assessment of biomarker discovery in metaproteomics. *Brief Bioinform* 2020;**1**:bbaa105.
51. Li F, Zhou Y, Zhang X, *et al.* SSizer: determining the sample sufficiency for comparative biological study. *J Mol Biol* 2020;**432**:3411–21.
52. Yang Q, Li B, Tang J, *et al.* Consistent gene signature of schizophrenia identified by a novel feature selection strategy from comprehensive sets of transcriptomic data. *Brief Bioinform* 2020;**21**:1058–68.
53. Zhang J, Xin L, Shan B, *et al.* PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol Cell Proteomics* 2012;**11**:M111.010587.
54. Rost HL, Rosenberger G, Navarro P, *et al.* OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat Biotechnol* 2014;**32**:219–23.

55. Li B, Tang J, Yang Q, *et al*. NOREVA: normalization and evaluation of MS-based metabolomics data. *Nucleic Acids Res* 2017;**45**: W162–70.

56. Yang Q, Wang Y, Zhang Y, *et al*. NOREVA: enhanced normalization and evaluation of time-course and multi-class metabolomic data. *Nucleic Acids Res* 2020;**48**:W436–48.