


ANPELA: analysis and performance assessment of the label-free quantification workflow for metaproteomic studies

Jing Tang, Jianbo Fu, Yunxia Wang, Bo Li, Yinghong Li, Qingxia Yang, Xuejiao Cui, Jiajun Hong, Xiaofeng Li, Yuzong Chen, Weiwei Xue and Feng Zhu 

Corresponding author: Feng Zhu, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, P. R. China. Tel.: +86-571-88208444; Fax: +86-571-88208444; E-mail: zhufeng@zju.edu.cn; prof.zhufeng@gmail.com

Abstract

Label-free quantification (LFQ) with a specific and sequentially integrated workflow of acquisition technique, quantification tool and processing method has emerged as the popular technique employed in metaproteomic research to provide a comprehensive landscape of the adaptive response of microbes to external stimuli and their interactions with other organisms or host cells. The performance of a specific LFQ workflow is highly dependent on the studied data. Hence, it is essential to discover the most appropriate one for a specific data set. However, it is challenging to perform such discovery due to the large number of possible workflows and the multifaceted nature of the evaluation criteria. Herein, a web server ANPELA (<https://idrblab.org/anpela/>) was developed and validated as the first tool enabling performance assessment of whole LFQ workflow (collective assessment by five well-established criteria with distinct underlying theories), and it enabled the identification of the optimal LFQ workflow(s) by a comprehensive performance ranking. ANPELA not only automatically detects the diverse formats of data generated by all quantification tools but also provides the most complete set of processing methods among the available web servers and stand-alone tools. Systematic validation using metaproteomic benchmarks revealed ANPELA's capabilities in (1) discovering well-performing workflow(s), (2) enabling assessment from multiple perspectives and (3) validating LFQ accuracy using spiked proteins. ANPELA has a unique ability to evaluate the performance of whole LFQ workflow and enables the discovery of the optimal LFQs by the comprehensive performance ranking of all 560 workflows. Therefore, it has great potential for applications in metaproteomic and other studies requiring LFQ techniques, as many features are shared among proteomic studies.

Key words: label-free quantification; LFQ workflow; metaproteomics; microbiota; performance assessment

Jing Tang, Jianbo Fu, Yunxia Wang, Bo Li, Yinghong Li, Qingxia Yang, Xuejiao Cui, Jiajun Hong, Xiaofeng Li are students of the College of Pharmaceutical Sciences in Zhejiang University, China and jointly cultivated by the School of Pharmaceutical Sciences in Chongqing University, China. They are interested in the area of bioinformatics and proteomics-based target identification.

Feng Zhu is a professor of the College of Pharmaceutical Sciences in Zhejiang University, China. He got his PhD degree from the National University of Singapore, Singapore. His research group (<https://idrblab.org/>) has been working in the fields of bioinformatics, OMIC-based drug discovery, system biology and medicinal chemistry. Welcome to visit his personal website (<https://idrblab.org/Peoples.php>) at any time.

Submitted: 30 September 2018; Received (in revised form): 19 November 2018

© The Author(s) 2019. 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

Introduction

Microbial community (MC) is characterized by widely diverse species and highly dynamic compositions [1, 2] that make it closely associated with the pathogenesis of various diseases (such as cancer [3], inflammatory disease [4] and metabolic disorder [5]), the productivity of agricultural plants [6, 7], the biogeochemical cycle of complex environmental matrices [8, 9] and so on. Investigations of the protein abundance in specific MC have enabled an unprecedented view of the adaptive response of microbes to external stimuli or their interaction with other organisms or host cells [10, 11]. Due to its superior ability to provide quantitative and dynamic information on microbiome's functional activity by directly profiling protein expression [12], metaproteomics stands out from other OMIC (any research field of biological study ending in -omics such as genomics, proteomics or metabolomics) techniques and becomes popular in current MC studies [13–18]. Thus far, several methods have been developed to quantify metaproteomic data [19]. Due to the simplicity of its experimental design and its ability to process large sample cohorts [20], label-free quantification (LFQ) is the most widely employed approach able to monitor the intestinal responses to gut microbiota [21–23], decipher the diversity and dynamics of bacterial proteome [24], reveal the abnormal abundance of proteins [25] and so on.

However, the low precision [26], poor reproducibility [27, 28], inaccuracy [29] and false discovery rate [30, 31] of the LFQ of microbial proteins have emerged as the key 'technical challenges' in recent metaproteomic study [27–29, 32–35]. These challenges may be attributed to (1) heterogeneity among microbial samples [36], (2) vast dynamic range of protein abundance [29], (3) highly random nature of experimental sampling [32], (4) systematic variations among sample preparations or experimental runs [12, 26, 32] and (5) enlarged search space in metaproteomic study [2, 30, 31]. The substantial variations in microbial signaling or phenotypes are frequently reported to originate from subtle changes in microbial samples [37–39], but these changes are rarely detected if an LFQ does not perform well [27]. Thus, a variety of acquisition techniques [40, 41], quantification tools [41, 42] and subsequent processing methods (transformation, normalization, missing value filtering and imputation) [43, 44] have been developed and extensively used to increase precision, enhance reproducibility and improve accuracy of LFQ in current proteomic studies (Supplementary Table S1).

Till now, 3 acquisition techniques (2 modes of acquisition), ≥ 18 quantification tools (Table 1) and dozens of subsequent processing methods (≥ 4 transformation, ≥ 16 normalization and ≥ 8 filtering/imputation) have been sequentially integrated to form numerous LFQ workflows. As reported, different LFQ workflows can produce considerably different or, in some cases, contradictory results [45], and the suitability of a specific workflow depends heavily on the particular data set analyzed [46]. It is now of great interest to compare the performances among various quantification tools [20, 33] and different label-free techniques [47, 48], and the selection of the optimal workflow for a specific data set is urgently needed [49], which is however a very challenging task in metaproteomics. First, the feasibility of discovering the optimal LFQ workflow is severely hampered by the huge number of possible combinations. Taking the subsequent processing methods as example, there are over 500 possible combinations of transformation, normalization and filtering/imputation approaches. Second, the criterion used to assess LFQ's performances is critical but also a great challenge for selecting the optimal workflow [40, 50]. LFQ's performance is

primarily evaluated by 'precision' (intragroup variations) [50], but a variety of new criteria could also be applied [20, 51, 52]. As these criteria complement one another, it is highly recommended to collectively employ these criteria as a 'thorough' evaluation of each workflow [46, 47, 52].

Several powerful LFQ-related tools (Supplementary Table S2) are currently available [53–58]. Perseus [53] and Gmine [54] integrate various processing methods in their corresponding analysis chain, but no assessment on the performances of LFQ is conducted. 'LFQbench' [20] and 'msCompare' [55] are recognized as evaluating the performances of 3–5 quantification tools [20, 55], and 'Normalyzer' [56], 'SPANS' [57] and 'GiaPronto' [58] are distinguished for being able to assess 1–8 normalization methods [56–58]. Since a typical LFQ workflow integrates both quantification tools and processing methods, any assessment focusing solely on the tool or the method could not fully reflect the overall performance of the whole workflow. Moreover, none of the available tools could systematically evaluate the performance, as highly recommended by previous studies [46, 47, 52], based on all those proposed criteria (each with distinct underlying theory) [20, 47, 59–63], and it is impossible for these tools to discover the workflows of optimal performance by comprehensively ranking all (more than 500) possible quantification combinations. Thus, it is essential to develop new tools to assess the performance of the whole LFQ workflow from multiple perspectives and identify the optimal one based on comprehensive performance ranking.

Herein, an open access server ANPELA was constructed and validated to enable the performance assessment of whole LFQ workflow (18 quantification tools accompanied by the combination of 28 processing methods). Five well-established criteria [20, 59–63] were collectively used for the first time to facilitate comprehensive assessment from multiple perspectives. ANPELA not only automatically detects the diverse formats of data generated by all quantification tools but also comprehensively provides the most complete set of processing methods among the available web servers or stand-alone applications. Moreover, it enables the identification of the optimal LFQ workflow(s) based on a comprehensive performance ranking. To validate its ability to (1) discover well-performing workflow(s), (2) enable assessment from multiple perspectives and (3) validate LFQ accuracy using spiked proteins, 16 benchmark data sets were collected and 4 independent metaproteomic cases were studied. In conclusion, ANPELA is unique for its ability to evaluate the performance of the whole LFQ workflow and enable the discovery of the optimal one(s) by comprehensive performance ranking, which can thus greatly facilitate proteome quantification for current metaproteomic research.

Materials and methods

Quantification tools and subsequent processing methods used in ANPELA

Eighteen quantification tools for preprocessing mass spectrometry (MS)-based raw data acquired by three acquisition techniques (Peak intensity, Spectral counting and SWATH-MS) were provided in ANPELA (Supplementary Method S1), which included Abacus [64], Census [65], DIA-Umpire [66], DTASelect [67], IRMaHEIDI [68], MFPaQ [68], MaxQuant [69], OpenMS [70], OpenSWATH [20], PEAKS [71], ProteinProphet [72], Proteios SE [73], Progenesis [44], SWATH 2.0 [20], Skyline [74], Spectronaut [75], Scaffold [33] and Thermo Proteome Discoverer [33]. Moreover, 28 subsequent processing methods were used for analyzing metaproteomic

Table 1. Overview on 18 quantification tools together with their corresponding mode of acquisition, acquisition technique and quantification metric. APEX, absolute protein expression index; Average-Top3, average intensity of the three highest intense peptides; emPAI, exponentially modified protein abundance index; iBAQ, intensity-based absolute quantification; IRP, internal reference peptide; MaxLFQ, delayed normalization and maximal peptide ratio extraction; NoC, intensity of nonconflicting peptides; NSAF, normalized spectral abundance factors; PAI, protein abundance index; Raw, raw spectral counts; Summed-Top3, summed intensity of the three highest intense peptides; TIC, total ion current; TPA, total protein approach; TRIC, transfer of identification confidence

| Mode of acquisition | Acquisition technique | Representative quantification tools | Quantification metric provided |
|------------------------------------|------------------------------------|---------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data-dependent acquisition (DDA) | Spectral counts (MS2) | Abacus | NSAF [47, 64, 130] |
| | | Census | Raw [65] |
| | | DTASelect | Raw [67] |
| | | IRMa-hEIDI | Raw [129] |
| | | MaxQuant | emPAI; APEX [47, 69, 130] |
| | | MFPaQ | PAI [131] |
| | | ProteinProphet | Raw [132] |
| | | Thermo Proteome Discoverer | NSAF [33, 47, 130] |
| | | Scaffold | emPAI; NSAF; Total spectra; Weighted spectra [133] |
| | | Data-independent acquisition (DIA) | Peak intensity (MS1) |
| MFPaQ | TIC; Average TIC [134] | | |
| OpenMS | TIC; Top three TIC [135] | | |
| PEAKS | Average-Top3; Summed-Top3 [136] | | |
| Progenesis | Average-Top3; NoC [33, 137] | | |
| Proteios SE | Summed-Top3 [48,73] | | |
| Scaffold | Average-Top3; iBAQ; TIC [138] | | |
| Skyline | Average-Top3 [139] | | |
| Thermo Proteome Discoverer | Minora algorithm [140, 141] | | |
| Census | Weighted overall intensities [142] | | |
| Data-independent acquisition (DIA) | SWATH-MS | DIA-Umpire OpenSWATH PeakView Spectronaut Skyline | Data-centric iBAQ abundance measures and ‘Top N’ metrics [65] Peptide-centric TRIC algorithm [143, 144] Peptide-centric extracted ion chromatograms alignment [145] Peptide-centric ‘target-decoy’ search strategy [75] Peptide-centric IRP method [146] |

data (Supplementary Method S2), which included 4 transformation (Box-cox [76], Cube root [77], Log transform [78] and Power [79]), 16 normalization (Auto Scaling [80], Cyclic Loess [81], EigenMS [50], Linear Baseline Scaling [82], Lowess [83], Median [84], Mean [85], Median Deviation [86], Pareto Scaling [87], PQN [88], Quantile [82], Robust Linear Regression [50], TIC [89], TMM [90], VSN [91] and Z-score [92]) and 8 filtering or imputation (Back [44], Basic [44], Bpca [44], Censor [44], Knn [93], Lls [44], Svd [93] and Zero [94]). To make the discussion in this study clearly understood, a three-letter abbreviation was assigned to each method (Table 2), and the brief descriptions of these methods were also provided in Table 2.

Criteria and corresponding assessment metrics for evaluating the performance of LFQ workflow

Besides the criterion ‘precision’ (intragroup variations) primarily applied for assessing the performance of LFQ [50], a variety of criteria have been developed and applied in the LFQ-related studies [20, 51, 52]. Due to the independent nature of these criteria [47], they are known to complement one another and highly recommended to be collectively employed for ‘thoroughly’ evaluating specific LFQ workflow [46, 47, 52]. Particularly, two independent criteria, ‘precision’ and ‘accuracy’, were collectively considered to enable the in-depth evaluation of LFQ [20, 44, 50], and three independent criteria were further integrated for the first time to comparatively evaluate various LFQ methods [47]. Moreover, an appropriate LFQ is expected to retain or even increase the difference in proteomic data between two distinct sample classes [59,

95]. In the meantime, the reproducibility of identified markers is also recognized as another independent criterion for assessing the performance of LFQ [52]. Thus, five independent criteria were used here to assess the performance of an LFQ workflow in ANPELA (each measured by a variety of assessment metrics; Supplementary Method S3).

(a) Precision of LFQ based on the proteomes among replicates

Different acquisition techniques, various tools for preprocessing raw proteomics data and diverse approaches for data processing (transformation, normalization, filtering and missing value imputation) could profoundly affect the precision of LFQ [20], which was assessed by pooled coefficient of variation (PCV) of the reported protein intensities among replicates [20, 96]. PCV was thus adopted as an assessment metric reflecting LFQ’s ability to reduce variations among replicates [97] and to enhance technical reproducibility [56]. A lower value of PCV indicates a more complete reduction of unwanted signals (denoting improved LFQ precision) [98].

(b) Classification ability of LFQ between distinct sample groups

An appropriate LFQ is expected to retain or even increase the difference in metaproteomic data between two distinct sample groups [59, 95]. Two-way clustering of proteins identified from these two groups is thus used as an assessment metric to assess LFQ classification ability [95, 99]. First, the total number of protein intensities in each sample is reduced by feature selection. Then proteins (rows) and samples (columns) are clustered by

Table 2. Abbreviated names of the subsequent processing methods used in this study together with their corresponding full names and brief descriptions

| | Processing method (abbreviation) | Brief description of each method | |
|-----------------------------------------------|-------------------------------------------------------|---------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------|
| Transformation | Box-Cox Transformation (BOX) | Transforming data based on linearity, normality and homoscedasticity assumption [76] | |
| | Cube Root Transformation (CUB) | Improving normality distribution of simple count data [77] | |
| | Log Transformation (LOG) | Almost routinely carried out for reaching a more symmetric distribution [78] | |
| | None (NON) | No transformation method applied | |
| | Power Transformation (POW) | Stabilizing variance and making data more normal distribution-like [79] | |
| Normalization | Auto Scaling (ATO) | Scaling protein intensities based on the standard deviation of OMIC data [80] | |
| | Cyclic Loess (CYC) | Estimating a regression surface using multivariate smoothing procedure [81] | |
| | EigenMS (EIG) | Preserving true differences based on treatment effect and singular value decomposition [50] | |
| | Linear Baseline Scaling (LIN) | Mapping each spectrum to the baseline based on a constant linear relationship [82] | |
| | Locally Weighted Scatterplot Smoothing (LOW) | Normalizing a proteomic data set by compensating for non-linear bias [83] | |
| | Mean Normalization (MEA) | Normalizing the data by mean value of all signals to eliminate background effect [84] | |
| | Median Absolute Deviation (MAD) | A measure of the spread of the data and used to estimate the sample standard deviation [85] | |
| | Median Normalization (MED) | Scaling the samples so that they have the same median [86] | |
| | None (NON) | No normalization method applied | |
| | Pareto Scaling (PAR) | Reducing the weight of large fold changes in protein intensities by standard deviation [87] | |
| | Probabilistic Quotient Normalization (PQN) | Transforming the spectra based on an overall estimation on the most probable dilution [88] | |
| | Quantile Normalization (QUA) | Achieving the same distribution of protein intensities across all samples [82] | |
| | Robust Linear Regression (RLR) | Used for transference to rescale one reference interval to another scale [50] | |
| | Total Ion Current (TIC) | Summing all the separate ion currents carried by the ions of different m/z [89] | |
| | Trimmed Mean of M Values (TMM) | Estimating scale factors between samples for differential expression analysis [90] | |
| | Variance Stabilization Normalization (VSN) | A non-linear method for keeping the variance constant over the entire data range [50] | |
| | Z-score normalization (ZSC) | Normalizing data based on the mean and standard deviation [92] | |
| | Imputation | Background Imputation (BAK) | Simulating the situation where protein values are missing [44] |
| | | Bayesian Principal Component Imputation (BPC) | Providing capacity to auto-select the parameters used in the estimation [44] |
| | | Censored Imputation (CEN) | Imputing non-missing completely at random missing values by lowest intensity value [44] |
| K-nearest Neighbor Imputation (KNN) | | Imputing values based on K proteins similar to the proteins with missing values [93] | |
| Local Least Squares Imputation (LLS) | | Missing value imputation by a linear combination of similar genes identified by KNN [44] | |
| None (NON) | | No imputation/filtering method applied | |
| Singular Value Decomposition (SVD) | | Estimating the missing values based on a linear consideration [93] | |
| Zero Imputation (ZER) | | Replacing the missing values with zeros deemed to the simplest imputation method [94] | |

their similarities in a protein intensity profile. Details can be found in [Supplementary Method S3](#).

(c) Differential expression analysis based on reproducibility optimization

To avoid overfitting/confounding in LFQ, distribution of *P*-values of protein intensities between two distinct sample groups is

examined [100, 101]. Unbiased variation is expected for the majority of proteins (without clear differentiation of expression), with an obvious peak in the [0.00, 0.05] interval corresponding to proteins with differential intensities [101]. Moreover, a volcano plot of proteins with differential intensity provides a glimpse of the total number of differentially expressed proteins [102]. In a metaproteomic study exploring mechanisms underlying complex biological process, the limited number of differentially

expressed proteins may result in false discovery [103]. Thus, the statistical differences (P -value) of protein intensities between distinct groups are first estimated by reproducibility-optimized test statistic (ROTS) [61, 104]. Distribution of P -values is then provided, and a skewed distribution may indicate overfitting and/or confounding [105].

(d) Reproducibility of the identified protein markers among different data sets

Consistency score (CS) is a popular assessment metric representing the reproducibility of identified markers [52, 106] and is used to assess the level of consistency among the protein markers discovered from the various portions of a given data set [107]. A higher CS denotes a more stable discovery of markers [52]. First, a random sampling is conducted on the studied data set to generate multiple sub-data sets. Then each protein is ranked based on its significance (q -value) and absolute fold change. Finally, those top-ranked proteins in each sub-data set are selected as markers before the calculation of the CS (Supplementary Method S3).

(e) Accuracy of LFQ based on spiked and background proteins

Extra experimental data (e.g. spiked proteins) are frequently generated and used as a reference to validate or adjust LFQ's performances [20, 96], and it is therefore straightforward to know the expected log fold changes (logFCs) for both spiked and background proteins (the expected logFC for background proteins is zero) [102]. First, the differentially expressed proteins are identified by ROTS. Then the true positive rate (TPR) for the successful discovery of spiked protein is calculated. A higher TPR corresponds to a greater accuracy. Moreover, the logFCs of protein intensity (for both spiked and background proteins) between distinct sample groups are measured, and the correlation between measured and expected logFCs is then assessed by mean squared error. A higher level of correlation indicates a more accurate LFQ workflow [102].

Each criterion above evaluated the LFQ performance based on its own underlying theory, and the combination of multiple criteria resulted in a systematic assessment. The assessment results are directly provided as tables or illustrated as plots online, all of which can be downloaded from the ANPELA web page. More information can be found in Supplementary Method S3.

Details of web server implementation

Official (<https://idrblab.org/angepela/>) and mirror (<http://idr.zju.edu.cn/angepela/>) sites of ANPELA are deployed on a server running 'Cent OS Linux v7.0 operating system', 'Apache HTTP web server v2.2.15' and 'Apache Tomcat servlet container' (<http://httpd.apache.org>). The web interface was constructed by R Package v3.4.1 (Shiny v0.13.1) running on the Shiny-server v1.4.1.759 (<http://www.rstudio.com/shiny>). A number of R packages were utilized in the background processes including affy, AUC, DiffCorr, DT, e1071, fastlo, ggfortify, ggsci, ggplot2, gplots, impute, limma, LPE, metabolomics, MetNorm, NOISeq, pcaMethods, png, RcmdrMisc, ROTS, rmarkdown, ropis, shiny, shinyBS, shinydashboard, shinyRGL, statTarget and vsn. ANPELA website can be readily accessed by all users with no login requirement and by a variety of popular web browsers including Google Chrome, Mozilla Firefox, Safari and Internet Explorer (10 or later). For using the stand-alone version of ANPELA downloaded directly from the official and mirror sites, the installation of R envi-

ronment is required before its running. The 'User Manual' of the stand-alone version of tool was provided in Supplementary Method S4.

The input formats required by ANPELA

The input files required by ANPELA are the standard output raw files of the 18 quantification tools, together with a label file indicating the category of each sample. The standard output raw files of quantification tools are described in the 'Tutorial' panel of ANPELA. Alternatively, a standard format designed by ANPELA is also accepted for assessing the in-house tool or predefined analytical workflow, which should be in 'csv' format with the dimension of $m \times n$ (m and n indicate the numbers of metaproteomic features and microbial samples, respectively). Additionally, an additional file providing the concentrations of known proteins (such as spiked proteins) is required to assess the performances of LFQ based on the last criterion. Specifically, the extra file should contain the 'class of samples' and the 'Sample ID'. The 'Sample ID' should be unique and defined by the preference of the ANPELA user, and the 'class of samples' refers to the group of 'Sample ID'. The input sample file can be found in the 'Tutorial' panel of ANPELA.

The formats and visualization of ANPELA output files

The output files of ANPELA include the following: (1) a variety of statistical measures (e.g. coefficient of variation, median absolute deviation, CS and PCV value), (2) the histograms, boxplots and matrixplots before and after transformation, normalization, filtering and missing value imputation, (3) Venn diagrams assessing LFQ's reproducibility, (4) a map of PCV distribution of protein intensities among replicates, (5) the P -value distribution, heatmap and volcano plot based on the identified markers, (6) boxplots of the deviations between the measured and expected logFCs of spiked and background proteins and (7) bar chart and receiver operating characteristic curve measuring quantification accuracy of both spiked and background proteins. The resulting LFQ files and assessment documents in a compressed ZIP format can be downloaded from ANPELA.

Results and discussion

The standard workflow of ANPELA

The main components of the ANPELA framework include four steps: (α) **uploading of the raw metaproteomic data**. Various file formats generated by all 18 quantification tools can be accepted, and the users are asked to upload specific files containing the data generated by those tools, together with a label file indicating the class of each sample. If the users want to process their own data before the ANPELA analysis, they can upload the data in a unified format defined by ANPELA. (β) **data transformation and normalization**. Transformation aims at reducing the impacts of very large value of intensities and making them more comparable or normally distributed [108, 109], and normalization can remove variability during the separate sample preparations and Mass Spectrometry (MS) runs [46, 50, 78]. (γ) **data filtering and imputation**. In total, eight popular approaches of metaproteomic studies were integrated. Missing values are frequently encountered in metaproteomic data sets and can greatly hamper subsequent OMIC analysis (with no calculation result for some extreme cases) [83]. (δ) **multifaceted performance assessment**. Each LFQ workflow is collectively evaluated from

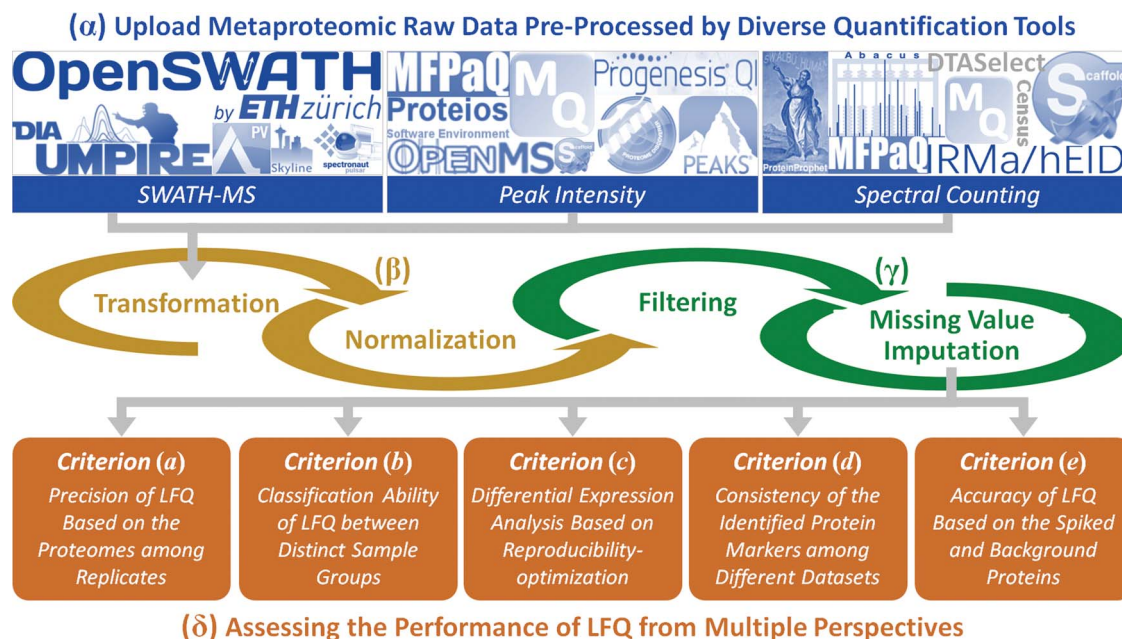


Figure 1. The standard workflow of ANPELA: (α) uploading the raw metaproteomic data preprocessed by 18 quantification tools; (β) data transformation and normalization; (γ) data filtering and imputation; and (δ) the multifaceted performance evaluation from five different perspectives.

five perspectives. A series of assessment metrics (represented by digital numbers or statistical plots) are provided. **Figure 1** illustrates the standard workflow of ANPELA, and a detailed demo on its usage can be found in the ‘Tutorial’ panel.

As known, a web server is accessible online and can work regardless of users’ operating system and platform, providing advantage over stand-alone applications in terms of accessibility and software update. However, the web server has the disadvantage of being slower than the stand-alone applications due to the time cost of web connection and the shared nature of computational resources [110]. Therefore, the function of ANPELA was realized by not only an online web server but also a local version downloadable directly from ANPELA sites. On one hand, the online web server provided the unique function of evaluating the performance of the whole LFQ workflow from multiple perspectives; on the other hand, the downloaded local tool enabled the discovery of the optimal LFQ(s) by comprehensively ranking all (more than 500) possible LFQ workflows. The official (<https://idrblab.org/angepela/>) and mirror (<http://idrblab.zju.edu.cn/angepela/>) sites of ANPELA could be readily assessed. Additionally, the source code of the local tool could be downloaded from the home page of these sites and run under R environment without extra installation. The ‘User Manual’ of the local version of ANPELA was also provided in **Supplementary Method S4**.

ANPELA facilitates the discovery of well-performing LFQ workflows for metaproteomic study

To comprehensively evaluate the level of dependence of LFQ workflows on different metaproteomic data sets, PRIDE database [111] was screened by searching the keywords including ‘Microbiota’, ‘Microbiome’ and ‘Metaproteomic’, which resulted in 106 metaproteomics-related records. Then the corresponding studies of the records were systematically reviewed. By considering several additional criteria [LFQ, the availability of raw intensity data files and the protein database or library to search against, the well-defined parameters (such as isolation scheme

and range of retention time) and the clear description on distinct sample groups], eight representative metaproteomic data sets were finally selected for further analysis: PXD006224 [42] (60 metabolic phase & 24 equilibrium phase fecal samples), PXD002882 [112] (21 Crohn’s disease patients & 10 healthy individuals), PXD006129 [113] (14 western style diet & 14 chow-fed mice), PXD006070 [114] (9 corn & 9 grass silage-based samples), PXD003028 [115] (8 people before & 8 people after their breakfast), PXD000987 [116] (8 transverse colon & 8 descending colon samples), PXD005929 [117] (3 surface-exposed & 3 whole cell extracts) and PXD006810 [118] (3 NleB1-infected and 3 wild-type cells). Detailed information of these benchmark data sets could be found in **Supplementary Table S3**.

Herein, the PCV of protein intensities among replicates for each data set was calculated and applied to assess the ‘precision’ of LFQ. A lower PCV value denoted a more complete reduction of the unwanted signal, which indicated improved LFQ ‘precision’ [20, 98]. The performance of each workflow could be categorized into four groups based on the following PCV values: superior (<0.14) [20, 119], good (0.14~0.3) [119, 120], fair (0.3~0.7) [121] and poor (>0.7) [121]. For 20 representative workflows (as shown in **Table 3**), their performances on different data sets varied significantly. Taking the first workflow (BOX-MAD-SVD) shown in **Table 3** as an example, its performance was superior, good, fair and poor in 1, 2, 4 and 1 data sets, respectively. The variation in LFQ performances among eight metaproteomic data sets are further illustrated in **Supplementary Figures S1** and **S2**. As shown, there was a clear data set-dependent characteristic for the analyzed representative workflows. Thus, for a particular metaproteomic study, it was necessary to assess the performance of each LFQ to discover the well-performing ones, and ANPELA was developed to provide this unique function by assessing each workflow as a whole analysis chain.

Besides metaproteomics, it was of great interest to evaluate the LFQ workflow on further proteomics data sets. As newly emerging technique [122], the ‘sequential windowed acquisition of all theoretical fragment ion mass spectra’ (SWATH-MS)

Table 3. Evaluation of different LFQ workflows based on the value of PCV of eight metaproteomic benchmark data sets. The performance of a workflow was categorized into four groups based on the following PCV values: superior (<0.14, double underline), good (0.14~0.3, single underline), fair (0.3~0.7, dotted underline) and poor (>0.7, no underline). The workflow abbreviations are provided in Table 2, and those benchmark data sets are in descending order by their total number of samples (the number of cases versus that of the controls, as listed in the brackets under each data set ID)

| Workflow | PXD006224 (60:24) | PXD002882 (21:10) | PXD006129 (14:14) | PXD006070 (9:9) | PXD003028 (8:8) | PXD000987 (4:4) | PXD005929 (3:3) | PXD006810 (3:3) |
|-------------|----------------------|----------------------|----------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| BOX-MAD-SVD | <u>0.05</u> | <u>0.15</u> | <u>0.47</u> | <u>0.38</u> | <u>0.44</u> | <u>0.64</u> | <u>0.27</u> | 1.31 |
| BOX-EIG-KNN | <u>0.26</u> | <u>0.48</u> | <u>0.32</u> | <u>0.19</u> | <u>0.26</u> | <u>0.22</u> | <u>0.10</u> | <u>0.28</u> |
| BOX-QUA-BAK | <u>0.43</u> | <u>0.76</u> | <u>0.65</u> | <u>0.54</u> | <u>0.54</u> | <u>0.46</u> | <u>0.22</u> | <u>0.64</u> |
| BOX-VSN-CEN | <u>0.09</u> | <u>0.02</u> | <u>0.14</u> | <u>0.09</u> | <u>0.10</u> | <u>0.09</u> | <u>0.04</u> | <u>0.43</u> |
| CUB-EIG-KNN | <u>0.29</u> | <u>0.52</u> | <u>0.36</u> | <u>0.21</u> | <u>0.29</u> | <u>0.25</u> | <u>0.11</u> | <u>0.31</u> |
| CUB-MAD-SVD | <u>0.05</u> | <u>0.15</u> | <u>0.47</u> | <u>0.38</u> | <u>0.44</u> | <u>0.64</u> | <u>0.27</u> | 1.31 |
| CUB-RLR-CEN | 0.75 | 0.93 | 0.85 | <u>0.59</u> | <u>0.61</u> | <u>0.53</u> | <u>0.24</u> | <u>0.55</u> |
| CUB-VSN-BAK | <u>0.11</u> | <u>0.18</u> | <u>0.15</u> | 0.76 | <u>0.12</u> | <u>0.11</u> | <u>0.05</u> | <u>0.60</u> |
| LOG-EIG-CEN | <u>0.10</u> | <u>0.24</u> | <u>0.13</u> | <u>0.12</u> | <u>0.12</u> | <u>0.09</u> | <u>0.03</u> | <u>0.12</u> |
| LOG-PQN-ZER | 0.90 | 0.95 | 1.27 | 1.07 | 0.87 | <u>0.68</u> | <u>0.29</u> | 1.34 |
| LOG-TIC-BAK | <u>0.21</u> | <u>0.31</u> | <u>0.23</u> | <u>0.45</u> | <u>0.21</u> | <u>0.17</u> | <u>0.09</u> | 0.72 |
| LOG-VSN-SVD | <u>0.06</u> | <u>0.15</u> | <u>0.48</u> | <u>0.39</u> | <u>0.45</u> | <u>0.65</u> | <u>0.27</u> | 1.31 |
| NON-EIG-KNN | 1.08 | 1.55 | 1.07 | <u>0.69</u> | 0.96 | 0.82 | <u>0.32</u> | 0.82 |
| NON-MAD-SVD | <u>0.05</u> | <u>0.15</u> | <u>0.47</u> | <u>0.38</u> | <u>0.44</u> | <u>0.64</u> | <u>0.27</u> | 1.30 |
| NON-VSN-BAK | <u>0.11</u> | <u>0.23</u> | <u>0.17</u> | <u>0.12</u> | <u>0.12</u> | <u>0.11</u> | <u>0.04</u> | <u>0.14</u> |
| NON-ZSC-CEN | <u>0.64</u> | 1.60 | 0.94 | 1.68 | 1.39 | 1.56 | <u>0.44</u> | 1.06 |
| POW-CYC-BAK | 1.25 | 2.95 | 2.46 | 1.20 | 4.55 | 10.83 | 1.00 | 6.08 |
| POW-LOW-BAK | 2.26 | <u>0.41</u> | 4.71 | 4.26 | 4.08 | 4.41 | 15.87 | 6.64 |
| POW-TMM-ZER | 1.68 | 1.07 | 1.11 | 1.47 | 4.07 | 2.65 | 11.16 | 1.07 |
| POW-VSN-CEN | <u>0.19</u> | <u>0.26</u> | <u>0.28</u> | <u>0.25</u> | <u>0.32</u> | <u>0.43</u> | <u>0.14</u> | <u>0.33</u> |

was reported to provide much more comprehensive detection and accurate quantitation of proteins compared to traditional acquisition techniques [122]. Therefore, seven SWATH-MS proteomic data sets were collected from the PRIDE database [111] following the similar process of metaproteomic data sets collection. These benchmark data sets included PXD001064 [123] (72 monozygotic & 44 dizygotic twins blood samples), PXD003972 [124] (20 wild-type & 20 GRB2 knock-in mice), PXD004880 [125] (18 'Down syndrome' patients & 18 healthy individuals plasma samples), PXD000672 [126] (18 tumorous & 18 non-tumorous kidney tissue biopsies), PXD006106 [127] (10 formaldehyde-treated & 10 -untreated HeLa), PXD003278 [128] (6 siRNA-treated & 6 PRPF8-depleted Cal51 cell samples) and PXD002952 [20] (3 yeast 30% versus *Escherichia coli* 5% & 3 yeast 15% versus *E. coli* 20% mixtures). Detailed information of these 7 data sets could be found in Supplementary Table S3, and the performances of the same set of 20 representative LFQ workflows in Table 3 on these 7 SWATH-MS proteomic data sets were provided in Supplementary Table S4. Similar to the findings of metaproteomic data sets, the performances of a given LFQ on seven different sets of SWATH-MS data varied greatly. Taking the first workflow (BOX-MAD-SVD) as an example, it performed 'superior', 'good', 'fair' and 'poor' in 5, 0, 1 and 1 data sets, respectively (Supplementary Table S4), which showed similar data set-dependent characteristic as for those metaproteomic data sets.

As shown in Table 3 and Supplementary Table S4, some workflows were found to provide mainly 'superior', 'good', 'fair' or 'poor' PCV performance. These findings inspired us to further investigate whether there was any general guideline on the selection of consistently well- or poorly performed workflows across various data sets. To achieve this, those eight metaproteomic and seven SWATH-MS proteomic data sets were further analyzed to assess the 'robustness' of LFQ performance. First, the 'robustness' of LFQ performance between any 2 of 15 studied data sets (defined by the amounts of overlapped workflows

with the same level of performances between two data sets) was calculated. Since the number of 2 combinations from those 15 studied data sets (C_2^{15}) equaled to 105, 4 error bars (in dark gray) were drawn in Supplementary Figure S3A to indicate the amounts of the overlapped workflows of consistently 'superior', 'good', 'fair' or 'poor' performance between any 2 of those 15 studied data sets. Then the 'robustness' of LFQ performance among 8 metaproteomic and among all 15 studied data sets was calculated, and the amounts of the overlapped workflows performing consistently 'superior', 'good', 'fair' or 'poor' were shown in Supplementary Figure S3A (blue bars for the metaproteomic data sets and orange bars for all 15 studied data sets). Obviously, with the accumulation of the analyzed data sets (from 2 to 8 to 15 in Supplementary Figure S3A), the amounts of overlapped LFQs decreased substantially (from ~161 to 53 to 33 and from ~242 to 129 to 97 for 'superior' and 'poor', respectively), and there was no overlapped workflow consistently performing 'good' and 'fair' across 8 and 15 data sets. These indicated a dramatic reduction of the 'robustness' of LFQ performance with the accumulation of the analyzed data sets.

As shown in Supplementary Figure S3A, there were 33 and 97 workflows performing consistently 'superior' and 'poor' across all 15 studied data sets. To investigate the probabilities of each processing method appearing in the consistently 'superior' performing workflows, Supplementary Figure S3B was drawn. The occurrence probabilities of some processing methods were found to be higher than the others. Moreover, Supplementary Figure S3C was provided to show the frequency of each method in consistently 'poor' performing workflows, and some methods were found to appear more frequently than others. However, considering the dramatically reduced amounts of overlapped workflows (from ~161 to 33 for 'superior', from ~242 to 97 for 'poor') with the accumulation of the analyzed data sets (Supplementary Figure S3A), it should be very difficult to summarize a general guideline for selecting the

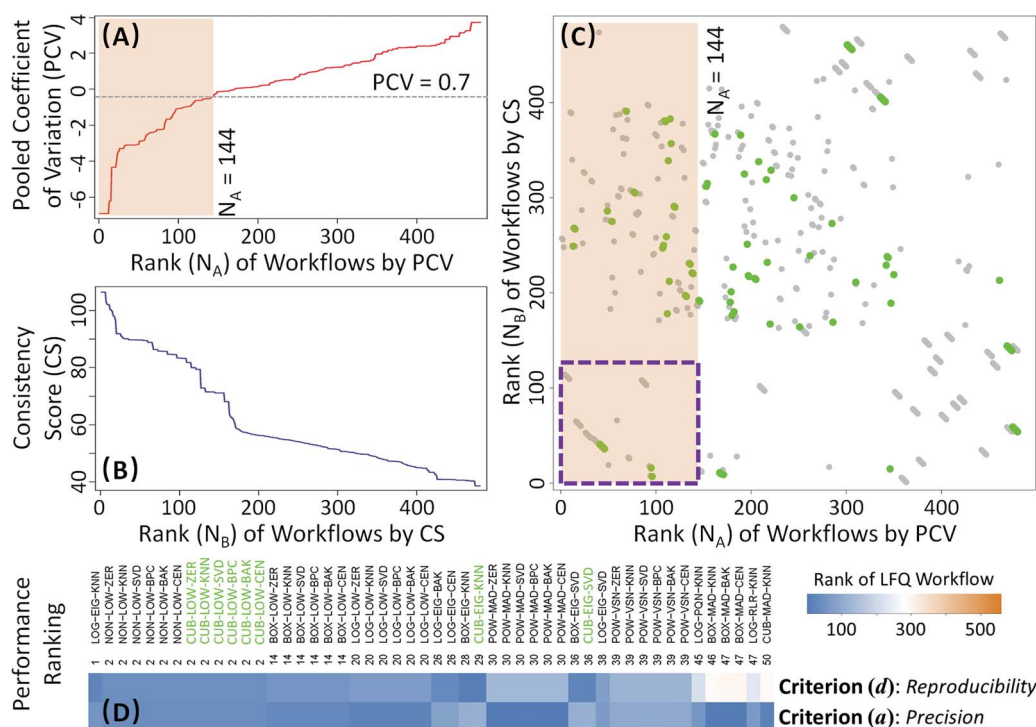


Figure 2. Performance assessment of LFQ workflows with the benchmark microbiome data set (PXD006224 [42]) from multiple perspectives. (A) precisions of different LFQ workflows were assessed by PCVs. Cut-off (≤ 0.7) for ‘desired’ PCVs [121] resulted in 144 workflows of good performance (in light orange background). (B) reproducibility of various workflows was evaluated by the CSs. (C) performance of workflows was collectively assessed by both PCVs and ‘CSs’. The dots indicate the rank of each LFQ workflow, as evaluated by PCVs and ‘CSs’, with the possible workflow used in Tilocca’s pioneering study [42] shown as green dots. The dots accumulated within the purple dash line square were considered as the well-performing workflows under both criteria (a) precision and (d) reproducibility. (D) the top 50 overall ranked workflows by both criteria together with their ranks by each independent criterion.

consistently well- or poorly performing workflows, especially when more data sets are included for further analysis. Thus, for a specific data set, an independent performance assessment on various LFQ workflows is required, and tools like ANPELA are recommended.

Feasibility of discovering the LFQ performing consistently well under two distinct criteria

Five well-established criteria [20, 59–62] (each with distinct underlying theory) are now available to evaluate the performance of LFQ workflows. Herein, the benchmark data set PXD006224 [42] was collected to analyze and compare the assessment results based on different criteria. As demonstrated in Figure 2A and B, two distinct criteria, (a) precision and (d) reproducibility, were utilized to assess and rank the performance of 560 LFQ workflows. In particular, Figure 2A ranked the LFQ workflows based on their ‘precision’ (measured by PCV value). The cut-off (≤ 0.7) for a ‘desired’ PCV [121] resulted in 144 workflows with good performance (light orange). Figure 2B ordered the LFQ workflows by their ‘reproducibility’ (measured by ‘CS’). Both figures demonstrated significant variations among the performances of different LFQ workflows (from $e^{-6.9}$ to $e^{3.7}$ for PCVs; from 38.6 to 106.5 for ‘CSs’). Due to these variations, the performance of each LFQ workflow should be assessed before any metaproteomic study, and ANPELA could be a handy tool for providing such important information.

Moreover, the performance of LFQ collectively assessed by both criteria, (a) precision and (d) reproducibility, was shown in Figure 2C. As illustrated, among those 144 highest-ranked work-

flows by PCV value, only 44 (30.6%, accumulated within the purple dash line square of Figure 2C) were found to perform well under the criterion (d) reproducibility. Considering the large number of possible workflows (560 in total, shown as dots in Figure 2C), only a small fraction ($\sim 7.9\%$) could be considered as the well-performing ones by both criteria. Furthermore, the workflow used in Tilocca’s pioneering study [42] was identified and mapped in Figure 2C (green dots). Since Tilocca’s study mentioned only the transformation method applied, all workflows (96 in total) containing that method were highlighted as green dot. Eight (8.3%) out of these 96 were considered as well performing by both criteria (within purple dash line square). Supplementary Table S5 provided a full list of LFQ workflows ranked independently by the ‘precision’ or ‘reproducibility’. As shown, the ranking results of two criteria were inconsistent with each other. Due to the independent nature of these two criteria [47], the collective consideration of both criteria was proposed here to give an overall ranking to each workflow. Particularly, the overall ranking was defined by the sum of the ranks for two criteria (the smaller the sum is, the higher the LFQ workflow ranks). Taking the LFQ workflow (LOG-EIG-KNN) in Supplementary Table S5 as an example, its independent ranks by both criteria were not so high (ranked 47th and 19th by criterion (a) and (d), respectively), but it was the one of the best overall ranking due to its well-balanced performances between two criteria. Figure 2D gave the top 50 overall ranked workflows together with their ranks by independent criteria (the workflows highlighted in green indicated those LFQs used in Tilocca’s study [42]).

Similar to the above analysis, the same strategy was applied to another SWATH-MS-based benchmark data set PXD001064

[123]. The performance ranks by two distinct criteria, (a) *precision* and (d) *reproducibility*, were illustrated in [Supplementary Figure S4A](#) and B, and the performances of LFQs collectively evaluated by both criteria were shown in [Supplementary Figure S4C](#). Similar to metaproteomic data set PXD006224 analyzed above, among the 280 highest-ranked workflows by one criterion (PCV values), only 14 (5.0%, accumulated in the purple dash line square of [Supplementary Figure S4C](#)) were found to be highly ranked by another ('CSs'). That is to say, among all 560 workflows (the dots in [Supplementary Figure S4C](#)), only very small fraction (~2.5%) could be considered as the well-performing ones by both criteria. The original workflows (five in total) used in Liu's study [123] for analyzing PXD001064 data set were highlighted as green dots in [Supplementary Figure S4C](#). As shown, three out of these five were suggested by criterion (a) to be the 'desired' workflows, but none of the five workflows was found to be well performing under both criteria (no green dot in purple dash line square). [Supplementary Figure S4D](#) gave the top 50 overall ranked workflows together with their ranks by each independent criterion. In summary, based on the analysis on two benchmark data sets (one metaproteomic and another SWATH-MS proteomic data sets), the selection of the most suitable criterion for a specific biological problem should be determined by the innate nature of the given data set. As these criteria complement one another [20, 52, 59–62], it is essential to collectively employ them for assessing the performance of LFQ workflow, which further made ANPELA distinguished from other available tools.

ANPELA enables the performance assessment of LFQ workflows from multiple perspectives

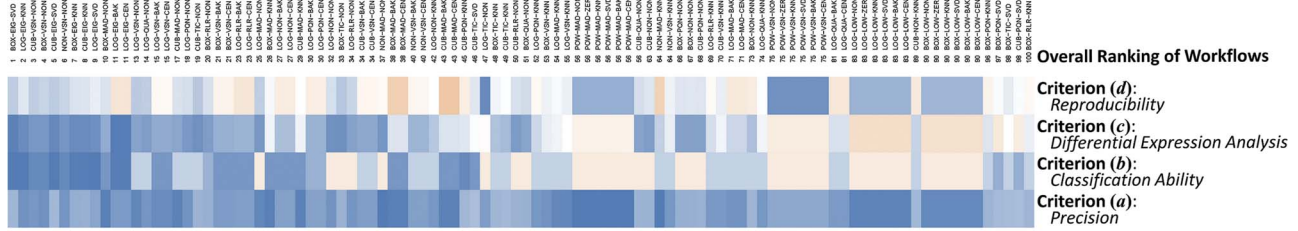
The assessment and subsequent performance ranking from multiple perspectives in ANPELA was realized by a variety of metrics. These metrics included PCV value for criterion (a) *precision* [20], classification accuracy for criterion (b) *classification ability* [95], level of uniformity of P-value distribution for criterion (c) *differential expression analysis* [101], CS value for criterion (d) *reproducibility* [107] and degree of correlation between the measured and expected logFCs of protein intensity for criterion (e) *accuracy* [50]. Based on these metrics, the performances of 560 LFQ workflows could be ranked separately, and five ranking numbers were assigned to each workflow by the five corresponding criteria. Due to the independent nature of the five criteria [47], the collective consideration of multiple criteria was proposed in this study and realized in ANPELA for providing the overall ranking to all 560 workflows. Particularly, the overall ranking of a given workflow was defined by the sum of multiple ranking numbers under multiple criteria (the smaller the sum is, the higher the LFQ workflow ranks). Taking the metaproteomic benchmark data set PXD006129 (including 14 western style diet & 14 chow-fed mice) [113] as an example, the overall ranking of the performances of 560 workflows on this particular data set was calculated and the top 100 ranked LFQs were provided in [Figure 3A](#). Since there is no spiked protein in this data set, only four criteria (a), (b), (c) and (d) were collectively assessed. As shown, the performance ranking under each criterion was represented by colored rectangles (the top-ranked LFQ was colored by exact blue. With the increase of ranking number, the color changed gradually toward orange with the last-ranked LFQ shown in exact orange). Based on the comprehensive ranking shown in [Figure 3A](#), the selection range of LFQ workflows could be significantly narrowed down, and it is thus possible for the readers to discover the optimal workflows for their own data set.

Because of the tremendous computational workload required for assessing 560 workflows, it is too time- and resource-consuming to make the comprehensive assessment service online. Therefore, an alternative way of enabling the evaluation on user's local computer was provided as a stand-alone ANPELA. This version could be downloaded from both official (<https://idrblab.org/anpela/>) and mirror (<http://idrb.zju.edu.cn/anpela/>) sites and provided the same sets of assessment metrics and plots as that of the online ANPELA. Moreover, this local version tool could further provide an overall performance ranking for all 560 workflows in the format of not only a figure (shown in [Figure 3](#)) but also a 'CSV' table (including detail results of assessing metrics, ranks by independent criteria and overall ranking). The exemplar input/output files could be simultaneously downloaded together with the source code of the stand-alone tool from ANPELA websites. The installation of R environment was required before running ANPELA and was provided in 'User Manual' (downloadable from websites) and [Supplementary Method S4](#). Both 'User Manual' and exemplar files could help users to get familiar with this tool as soon as possible.

ANPELA validates the accuracy of LFQ workflows based on spiked and background proteins

Spiked protein is frequently adopted to evaluate the performance of an LFQ workflow [52, 129]. A workflow with desirable accuracy should not only maintain unbiased variation in background proteins between sample groups (the expected logFC between the case and control equals to zero) but also make the variation in spiked proteins biased by viewing the expected abundance ratios as golden standard [50, 101]. Herein, the ability of ANPELA to validate the accuracy of LFQ workflows based on spiked proteins was analyzed using benchmark data PXD002099 with known 'ground truth' of variant proteins (48 UPS1 proteins spiked into yeast proteome digest with five different concentrations 2, 4, 10, 25 and 50 fmol/μL) [61]. A random combination of any 2 of these different concentrations could therefore result in 10 pairs of sample group with distinct concentration. Taking the pair of 10 versus 25 fmol/μL as an example, the accuracy of the 20 representative LFQ workflows were determined by considering the spiked ([Figure 4A](#)) and background ([Figure 4B](#)) proteins. All workflows were found to be able to maintain the unbiased variations of background proteins between sample groups, but only some workflows could make the variation in spiked proteins biased according to the expected abundance ratio ($\log_{FC_{expected}} = 1.32$ in this case). With reference to the logFC of unnormalized data (the gray boxplot in [Figure 4A](#)), several workflows demonstrated good performance in guaranteeing biased variation in spiked proteins. The variance stabilization normalization (VSN) described in Valikangas's study [50] as 'performed well in differential expression analysis' was identified well performing (the 6th ranked as shown in [Figure 4](#)) in this study. The well-performing workflows (in [Figure 4](#)) also included LOG-MED-NON, LOG-RLR-BAK, NON-EIG-ZER, NON-RLR-NON and LOG-CYC-BAK. For the remaining nine pairs of sample group, the accuracy of the same set of 20 representative workflows were analyzed and illustrated in [Supplementary Figures S5–S13](#). As shown, the accuracy of different representative workflows for a specific concentration varied greatly, and the accuracy ranks of a particular workflow for different pairs also changed substantially. [Table 4](#) gave an overview of the LFQ workflows performing better than unnormalization (indicated by circle).

A. Metaproteomic Dataset PXD006129 (14 western style diet & 14 chow-fed mice)



B. Metaproteomic Dataset PXD002099 (with spiked proteins)

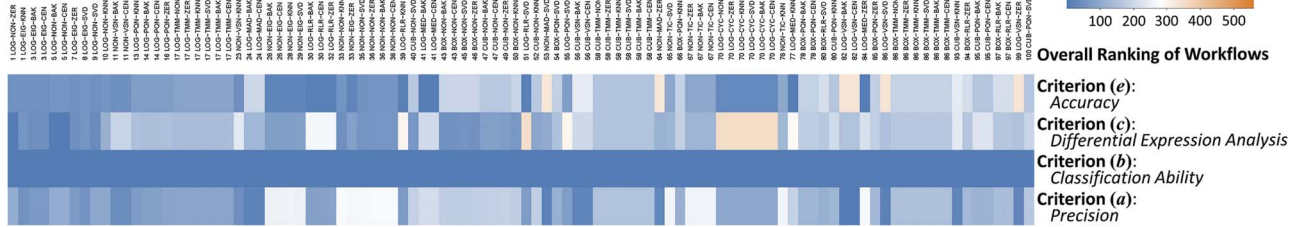


Figure 3. Overall ranking of the performances of LFQ workflows on the metaproteomic data sets (A) PXD006129 with 14 western style diet and 14 chow-fed mice and (B) PXD002099 containing 48 UPS1 proteins spiked with five different concentrations. Since there is no spiked protein in data set PXD006129, only four criteria (a), (b), (c) and (d) were collectively assessed. Since the sample sizes of both case and control in PXD002099 were very small (3 versus 3), it was inappropriate to assess the quantification performance base on criterion (d) reproducibility, and only four criteria (a), (b), (c) and (e) were considered. The performance ranking under each criterion was represented by colored rectangles (the top-ranked LFQ was colored by exact blue). With the increase of ranking number, the color changed gradually toward orange with the last-ranked LFQ shown in exact orange.

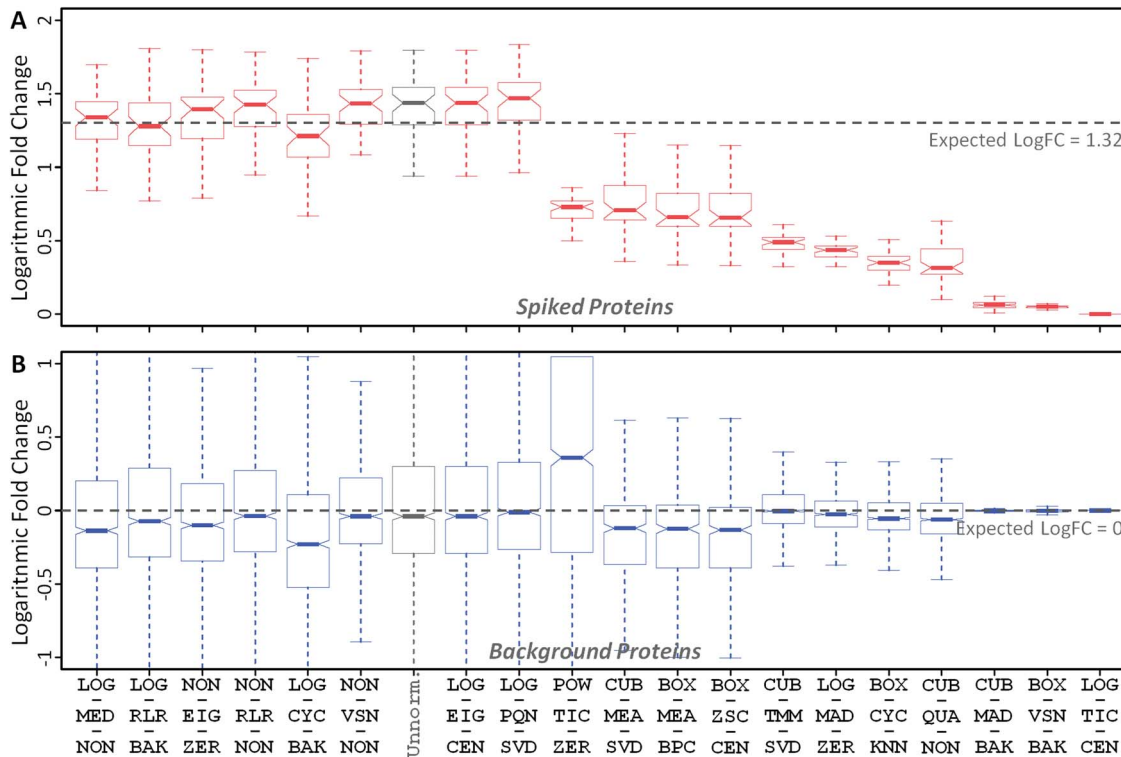


Figure 4. The accuracy of 20 representative LFQ workflows based on (A) spiked and (B) background proteins for two distinct data sets of different protein concentration (10 versus 25 fmol/μL) in PXD002099. The workflows were ordered by the level of deviations between the median and expected logFCs of spiked proteins. Expected logFCs for the background and spiked proteins were 0 and 1.32, respectively. The workflow abbreviation was provided in Table 2. ‘Unnorm.’ denotes unnormalized data.

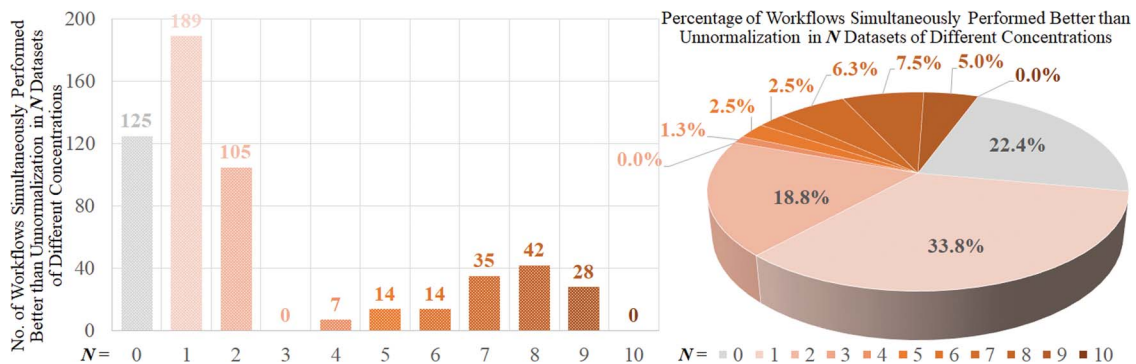
No LFQ workflow was found performing consistently better across all pairs of sample group, and the number of these well-performing workflows for different data set pairs varied from 1 to 15.

Supplementary Table S6 provided the degree of correlation (between the measured and expected logFCs of protein intensity)

of LFQ workflow for the data sets of different concentrations (2 versus 4 fmol/μL, 2 versus 10 fmol/μL, 2 versus 50 fmol/μL, 4 versus 10 fmol/μL, 4 versus 25 fmol/μL, 4 versus 50 fmol/μL, 10 versus 25 fmol/μL, 10 versus 50 fmol/μL and 25 versus 50 fmol/μL). The number in each cell of Supplementary Table S6 indicated the median deviation of spiked proteins ±

Table 4. Overview of the performance of 20 representative LFQ workflows assessed by criterion (e) accuracy. The LFQ workflows performing better and worse than unnormalization were represented by circle and cross, respectively

| Workflow | 2 versus | 2 versus | 2 versus | 2 versus | 4 versus | 4 versus | 4 versus | 10 versus | 10 versus | 25 versus |
|-------------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|
| | 4 | 10 | 25 | 50 | 10 | 25 | 50 | 25 | 50 | 50 |
| NON-VSN-NON | × | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| LOG-CYC-BAK | × | ○ | ○ | ○ | ○ | ○ | ○ | ○ | × | ○ |
| LOG-MED-NON | × | ○ | ○ | ○ | ○ | ○ | ○ | ○ | × | ○ |
| LOG-PQN-SVD | × | ○ | ○ | ○ | ○ | ○ | ○ | × | ○ | ○ |
| LOG-RLR-BAK | × | ○ | ○ | ○ | ○ | ○ | ○ | ○ | × | × |
| NON-RLR-NON | × | ○ | ○ | ○ | ○ | ○ | ○ | ○ | × | × |
| NON-EIG-ZER | × | × | ○ | ○ | ○ | × | × | ○ | × | ○ |
| BOX-MEA-BPC | × | ○ | × | × | ○ | × | × | × | × | × |
| BOX-VSN-BAK | × | ○ | × | × | ○ | × | × | × | × | × |
| BOX-ZSC-CEN | × | ○ | × | × | ○ | × | × | × | × | × |
| CUB-MEA-SVD | × | ○ | × | × | ○ | × | × | × | × | × |
| LOG-EIG-CEN | ○ | × | × | × | × | × | × | × | × | ○ |
| BOX-CYC-KNN | × | × | × | × | ○ | × | × | × | × | × |
| CUB-QUA-NON | × | × | × | × | ○ | × | × | × | × | × |
| CUB-TMM-SVD | × | × | × | × | ○ | × | × | × | × | × |
| LOG-MAD-ZER | × | × | × | × | ○ | × | × | × | × | × |
| CUB-MAD-BAK | × | × | × | × | × | × | × | × | × | × |
| LOG-TIC-CEN | × | × | × | × | × | × | × | × | × | × |
| POW-TIC-ZER | × | × | × | × | × | × | × | × | × | × |

**Figure 5.** Numbers and percentages of the LFQs performing better than unnormalization among all workflows assessed by criterion (e) accuracy.

standard deviation. The number highlighted in green and bold was the workflow of better accuracy than unnormalization. As shown, no LFQ workflow was discovered to perform consistently better than unnormalization across all 10 pairs of sample group, and the majority of the workflows performed well only in very limited pairs (~75% of the 560 were well performing in no more than 2 pairs; Figure 5). The inconsistency of the performances among different pairs indicated the data set-dependent nature of LFQ workflow, and a case-by-case performance assessment for a given data set is thus required. In sum, ANPELA demonstrated good capacity for validating the accuracy of an LFQ workflow by preserving the true biological variation of spiked proteins, which could be used to discover the optimal workflow for a metaproteomic study with the gold standard of the expected protein abundance ratio.

Besides ‘accuracy’, other four criteria could be collectively considered in ANPELA. A metaproteomic benchmark data set PXD002099 (48 UPS1 proteins spiked into the yeast proteome with five different concentrations) [61] was therefore collected as example. The overall ranking of the performances of 560 workflows on this data set was calculated and the top 100 ranked LFQs were provided in Figure 3B. Since the sample sizes of the

cases and controls were small (3 versus 3), it was inappropriate to assess the quantification performance using criterion (d) reproducibility, and only criteria (a), (b), (c) and (e) were considered. Based on the ranking in Figure 3B, the selection range of LFQs was greatly narrowed down by collectively considering those independent criteria.

Conclusions

ANPELA allows the users to directly analyze the standard output files of popular quantification tools, and its multifaceted strategies for performance assessments could collectively improve the reproducibility, precision and accuracy of LFQ in metaproteomic studies. ANPELA has a unique ability to evaluate the performance of the whole workflow and enables the identification of the optimal LFQ(s) by the comprehensive performance ranking of all 560 workflows. Thus, it has great potential for applications in metaproteomic and other studies requiring LFQ techniques, as many features are shared among proteomic studies. To enhance the accessibility of ANPELA, a mirror site (<http://idr.b.zju.edu.cn/angepela/>) was constructed as the backup of the official ANPELA site (<https://idrblab.org/angepela/>).

Author Contributions

F.Z. conceived the idea and supervised the work. J.T., J.F. and Y.W. performed the research. J.T., B.L., Y.L. and J.F. constructed the web server, developed the software and wrote the scripts. J.T., J.F., Y.W., Y.L., Q.Y., X.C., J.H., X.L., Y.C. and W.X. prepared and analyzed the data. F.Z. wrote the manuscript.

Key Points

- An online tool ANPELA is presented to enable performance assessment of the whole LFQ workflow (collective assessment by five well-established criteria with distinct underlying theories).
- ANPELA enables the identification of the optimal LFQ(s) based on the comprehensive performance ranking of all 560 workflows.
- ANPELA not only automatically detects the diverse formats of data generated by all quantification tools but also provides the most complete set of processing methods among the available web servers and stand-alone tools.
- Systematic validation using metaproteomic benchmarks revealed ANPELA's capabilities. It is publicly accessible at <https://idrblab.org/angepela/>.

Supplementary Data

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

Funding

National Natural Science Foundation of China (81872798); National Key Research and Development Program of China (2018YFC0910500); Innovation Project on Industrial Generic Key Technologies of Chongqing (cstc2015zdcy-ztxx120003); Fundamental Research Funds for the Central Universities (2018QNA7023, 10611CD)XZ238826, 2018CDQYSG0007, CDJZR14468801, CDJKXB14011).

References

1. Widder S, Allen RJ, Pfeiffer T, et al. Challenges in microbial ecology: building predictive understanding of community function and dynamics. *ISME J* 2016;10:2557–68.
2. Cheng K, Ning Z, Zhang X, et al. MetaLab: an automated pipeline for metaproteomic data analysis. *Microbiome* 2017;5:157.
3. Lynch SV, Pedersen O. The human intestinal microbiome in health and disease. *N Engl J Med* 2016;375:2369–79.
4. Mallick H, Ma S, Franzosa EA, et al. Experimental design and quantitative analysis of microbial community multiomics. *Genome Biol* 2017;18:228.
5. Martinez MC, Andriantsitohaina R. Extracellular vesicles in metabolic syndrome. *Circ Res* 2017;120:1674–86.
6. Castrillo G, Teixeira PJ, Paredes SH, et al. Root microbiota drive direct integration of phosphate stress and immunity. *Nature* 2017;543:513–8.
7. Philippot L, Raaijmakers JM, Lemanceau P, et al. Going back to the roots: the microbial ecology of the rhizosphere. *Nat Rev Microbiol* 2013;11:789–99.
8. Liu J, Williams PC, Geisler-Lee J, et al. Impact of wastewater effluent containing aged nanoparticles and other components on biological activities of the soil microbiome, Arabidopsis plants, and earthworms. *Environ Res* 2018;164:197–203.
9. Maier S, Tamm A, Wu D, et al. Photoautotrophic organisms control microbial abundance, diversity, and physiology in different types of biological soil crusts. *ISME J* 2018;12:1032–46.
10. Broberg M, Doonan J, Mundt F, et al. Integrated multi-omic analysis of host-microbiota interactions in acute oak decline. *Microbiome* 2018;6:21.
11. Wang DZ, Kong LF, Li YY, et al. Environmental microbial community proteomics: status, challenges and perspectives. *Int J Mol Sci* 2016;17:E1275.
12. Zhang X, Ning Z, Mayne J, et al. In vitro metabolic labeling of intestinal microbiota for quantitative metaproteomics. *Anal Chem* 2016;88:6120–5.
13. Brunkwall L, Orho-Melander M. The gut microbiome as a target for prevention and treatment of hyperglycaemia in type 2 diabetes: from current human evidence to future possibilities. *Diabetologia* 2017;60:943–51.
14. Eymann C, Lassek C, Wegner U, et al. Symbiotic interplay of fungi, algae, and bacteria within the lung lichen *Lobaria pulmonaria* L. Hoffm. as assessed by state-of-the-art metaproteomics. *J Proteome Res* 2017;16:2160–73.
15. Tanca A, Abbondio M, Palomba A, et al. Potential and active functions in the gut microbiota of a healthy human cohort. *Microbiome* 2017;5:79.
16. Kleiner M, Dong X, Hinzke T, et al. Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities. *Proc Natl Acad Sci U S A* 2018;115:E5576–84.
17. Zhang X, Deeke SA, Ning Z, et al. Metaproteomics reveals associations between microbiome and intestinal extracellular vesicle proteins in pediatric inflammatory bowel disease. *Nat Commun* 2018;9:2873.
18. Korpela K, Salonen A, Vepsäläinen O, et al. Probiotic supplementation restores normal microbiota composition and function in antibiotic-treated and in caesarean-born infants. *Microbiome* 2018;6:182.
19. von Bergen M, Jehmlich N, Taubert M, et al. Insights from quantitative metaproteomics and protein-stable isotope probing into microbial ecology. *ISME J* 2013;7:1877–85.
20. 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.
21. Lichtman JS, Sonnenburg JL, Elias JE. Monitoring host responses to the gut microbiota. *ISME J* 2015;9:1908–15.
22. Ferrer M, Raczowska BA, Martinez-Martinez M, et al. Phenotyping of gut microbiota: focus on capillary electrophoresis. *Electrophoresis* 2017;38:2275–86.
23. Tang J, Zhang Y, Fu J, et al. Computational advances in the label-free quantification of cancer proteomics data. *Curr Pharm Des* 2018. doi: 10.2174/1381612824666181102125638.
24. El-Rami F, Nelson K, Xu P. Proteomic approach for extracting cytoplasmic proteins from *Streptococcus sanguinis* using mass spectrometry. *J Mol Biol Res* 2017;7:50–7.

25. Kleiner M, Young JC, Shah M, et al. Metaproteomics reveals abundant transposase expression in mutualistic endosymbionts. *MBio* 2013;4:e00223–e00213.
26. Li Z, Adams RM, Chourey K, et al. Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos. *J Proteome Res* 2012;11:1582–90.
27. Cretu D, Prassas I, Saraon P, et al. Identification of psoriatic arthritis mediators in synovial fluid by quantitative mass spectrometry. *Clin Proteomics* 2014;11:27.
28. Barschke P, Oeckl P, Steinacker P, et al. Proteomic studies in the discovery of cerebrospinal fluid biomarkers for amyotrophic lateral sclerosis. *Expert Rev Proteomics* 2017;14:769–77.
29. Huang Q, Yang L, Luo J, et al. SWATH enables precise label-free quantification on proteome scale. *Proteomics* 2015;15:1215–23.
30. Ivanov MV, Lobas AA, Karpov DS, et al. Comparison of false discovery rate control strategies for variant peptide identifications in shotgun proteogenomics. *J Proteome Res* 2017;16:1936–43.
31. Cappadona S, Baker PR, Cutillas PR, et al. Current challenges in software solutions for mass spectrometry-based quantitative proteomics. *Amino Acids* 2012;43:1087–108.
32. Wang H, Shi T, Qian WJ, et al. The clinical impact of recent advances in LC-MS for cancer biomarker discovery and verification. *Expert Rev Proteomics* 2016;13:99–114.
33. Al Shweiki MR, Monchgesang S, Majovsky P, et al. Assessment of label-free quantification in discovery proteomics and impact of technological factors and natural variability of protein abundance. *J Proteome Res* 2017;16:1410–24.
34. Shen X, Shen S, Li J, et al. IonStar enables high-precision, low-missing-data proteomics quantification in large biological cohorts. *Proc Natl Acad Sci U S A* 2018;115:E4767–76.
35. Lyutvinskiy Y, Yang H, Rutishauser D, et al. In silico instrumental response correction improves precision of label-free proteomics and accuracy of proteomics-based predictive models. *Mol Cell Proteomics* 2013;12:2324–31.
36. Vandenkoornhuysen P, Quaiser A, Duhamel M, et al. The importance of the microbiome of the plant holobiont. *New Phytol* 2015;206:1196–206.
37. Legewie S, Bluthgen N, Schafer R, et al. Ultrasensitization: switch-like regulation of cellular signaling by transcriptional induction. *PLoS Comput Biol* 2005;1:e54.
38. Gatto L, Hansen KD, Hoopmann MR, et al. Testing and validation of computational methods for mass spectrometry. *J Proteome Res* 2016;15:809–14.
39. Li H, He Y, Ding G, et al. dbDEPC: a database of differentially expressed proteins in human cancers. *Nucleic Acids Res* 2010;38:D658–64.
40. Chen B, Zhang D, Wang X, et al. Proteomics progresses in microbial physiology and clinical antimicrobial therapy. *Eur J Clin Microbiol Infect Dis* 2017;36:403–13.
41. Lassek C, Burghartz M, Chaves-Moreno D, et al. A metaproteomics approach to elucidate host and pathogen protein expression during catheter-associated urinary tract infections (CAUTIs). *Mol Cell Proteomics* 2015;14:989–1008.
42. Tilocca B, Burbach K, Heyer CME, et al. Dietary changes in nutritional studies shape the structural and functional composition of the pigs' fecal microbiome-from days to weeks. *Microbiome* 2017;5:144.
43. Heyer R, Kohrs F, Reichl U, et al. Metaproteomics of complex microbial communities in biogas plants. *J Microbiol Biotechnol* 2015;8:749–63.
44. Valikangas T, Suomi T, Elo LL. A comprehensive evaluation of popular proteomics software workflows for label-free proteome quantification and imputation. *Brief Bioinform* 2017. doi:<https://doi.org/10.1093/bib/bbx054>.
45. Khoonsari PE, Haggmark A, Lonnberg M, et al. Analysis of the cerebrospinal fluid proteome in Alzheimer's disease. *PLoS One* 2016;11:e0150672.
46. Weiss S, Xu ZZ, Peddada S, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* 2017;5:27.
47. Bubis JA, Levitsky LI, Ivanov MV, et al. Comparative evaluation of label-free quantification methods for shotgun proteomics. *Rapid Commun Mass Spectrom* 2017;31:606–12.
48. Ning K, Fermin D, Nesvizhskii AI. Comparative analysis of different label-free mass spectrometry based protein abundance estimates and their correlation with RNA-Seq gene expression data. *J Proteome Res* 2012;11:2261–71.
49. 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.
50. Valikangas T, Suomi T, Elo LL. A systematic evaluation of normalization methods in quantitative label-free proteomics. *Brief Bioinform* 2018;19:1–11.
51. Hawley AK, Brewer HM, Norbeck AD, et al. Metaproteomics reveals differential modes of metabolic coupling among ubiquitous oxygen minimum zone microbes. *Proc Natl Acad Sci U S A* 2014;111:11395–400.
52. Kashyap PC, Chia N, Nelson H, et al. Microbiome at the frontier of personalized medicine. *Mayo Clin Proc* 2017;92:1855–64.
53. Tyanova S, Temu T, Sinitcyn P, et al. The Perseus computational platform for comprehensive analysis of (pro)teomics data. *Nat Methods* 2016;13:731–40.
54. Proietti C, Zakrzewski M, Watkins TS, et al. Mining, visualizing and comparing multidimensional biomolecular data using the Genomics Data Miner (GMine) web-server. *Sci Rep* 2016;6:38178.
55. Hoekman B, Breitling R, Suits F, et al. msCompare: a framework for quantitative analysis of label-free LC-MS data for comparative candidate biomarker studies. *Mol Cell Proteomics* 2012;11(6):M111.015974.
56. Chawade A, Alexandersson E, Levander F. Normalizer: a tool for rapid evaluation of normalization methods for omics data sets. *J Proteome Res* 2014;13:3114–20.
57. Webb-Robertson BJ, Matzke MM, Jacobs JM, et al. A statistical selection strategy for normalization procedures in LC-MS proteomics experiments through dataset-dependent ranking of normalization scaling factors. *Proteomics* 2011;11:4736–41.
58. Weiner AK, Sidoli S, Diskin SJ, et al. GiaPronto: a one-click graph visualization software for proteomics datasets. *Mol Cell Proteomics* 2018;17:1426–31.
59. Williams KE, Lemieux GA, Hassis ME, et al. Quantitative proteomic analyses of mammary organoids reveals distinct signatures after exposure to environmental chemicals. *Proc Natl Acad Sci U S A* 2016;113:E1343–51.
60. Collins BC, Hunter CL, Liu Y, et al. Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. *Nat Commun* 2017;8:291.

61. Pursiheimo A, Vehmas AP, Afzal S, et al. Optimization of statistical methods impact on quantitative proteomics data. *J Proteome Res* 2015;14:4118–26.
62. Dowle AA, Wilson J, Thomas JR. Comparing the diagnostic classification accuracy of iTRAQ, peak-area, spectral-counting, and emPAI methods for relative quantification in expression proteomics. *J Proteome Res* 2016;15:3550–62.
63. Zhu F, Li XX, Yang SY, et al. Clinical success of drug targets prospectively predicted by in silico study. *Trends Pharmacol Sci* 2018;39:229–31.
64. Fermin D, Basur V, Yocum AK, et al. Abacus: a computational tool for extracting and pre-processing spectral count data for label-free quantitative proteomic analysis. *Proteomics* 2011;11:1340–5.
65. Park SK, Venable JD, Xu T, et al. A quantitative analysis software tool for mass spectrometry-based proteomics. *Nat Methods* 2008;5:319–22.
66. Tsou CC, Avtonomov D, Larsen B, et al. DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat Methods* 2015;12:258–64.
67. Tabb DL, McDonald WH, Yates JR, III. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res* 2002;1:21–6.
68. Ramus C, Hovasse A, Marcellin M, et al. Spiked proteomic standard dataset for testing label-free quantitative software and statistical methods. *Data Brief* 2016;6:286–94.
69. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 2008;26:1367–72.
70. Sturm M, Bertsch A, Gropf C, et al. OpenMS—an open-source software framework for mass spectrometry. *BMC Bioinformatics* 2008;9:163.
71. Ma B, Zhang K, Hendrie C, et al. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom* 2003;17:2337–42.
72. Nesvizhskii AI, Keller A, Kolker E, et al. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* 2003;75:4646–58.
73. Hakkinen J, Vincic G, Mansson O, et al. The proteios software environment: an extensible multiuser platform for management and analysis of proteomics data. *J Proteome Res* 2009;8:3037–43.
74. MacLean B, Tomazela DM, Shulman N, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 2010;26:966–8.
75. 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 micro-tissues. *Mol Cell Proteomics* 2015;14:1400–10.
76. Sakia RM. The Box–Cox transformation technique—a review. *J R Stat Soc Ser D Stat* 1992;41:169–78.
77. Ho EN, Kwok WH, Leung DK. Control of the misuse of testosterone in castrated horses based on an international threshold in plasma. *Drug Test Anal* 2015;7:414–9.
78. De Livera AM, Dias DA, De Souza D, et al. Normalizing and integrating metabolomics data. *Anal Chem* 2012;84:10768–76.
79. Li L, Wu J, Ghosh JK, et al. Estimating spatiotemporal variability of ambient air pollutant concentrations with a hierarchical model. *Atmos Environ (1994)* 2013;71:54–63.
80. Gromski PS, Xu Y, Hollywood KA, et al. The influence of scaling metabolomics data on model classification accuracy. *Metabolomics* 2015;11:684–95.
81. Kultima K, Nilsson A, Scholz B, et al. Development and evaluation of normalization methods for label-free relative quantification of endogenous peptides. *Mol Cell Proteomics* 2009;8:2285–95.
82. Kohl SM, Klein MS, Hochrein J, et al. State-of-the art data normalization methods improve NMR-based metabolomic analysis. *Metabolomics* 2012;8:146–60.
83. Karpievitch YV, Dabney AR, Smith RD. Normalization and missing value imputation for label-free LC-MS analysis. *BMC Bioinformatics* 2012;13(Suppl 16):S5.
84. Callister SJ, Barry RC, Adkins JN, et al. Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J Proteome Res* 2006;5:277–86.
85. Andjelkovic V, Thompson R. Changes in gene expression in maize kernel in response to water and salt stress. *Plant Cell Rep* 2006;25:71–9.
86. Matzke MM, Waters KM, Metz TO, et al. Improved quality control processing of peptide-centric LC-MS proteomics data. *Bioinformatics* 2011;27:2866–72.
87. van den Berg RA, Hoefsloot HC, Westerhuis JA, et al. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* 2006;7:142.
88. Dieterle F, Ross A, Schlotterbeck G, et al. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabolomics. *Anal Chem* 2006;78:4281–90.
89. Gaspari M, Chiesa L, Nicastrì A, et al. Proteome speciation by mass spectrometry: characterization of composite protein mixtures in milk replacers. *Anal Chem* 2016;88:11568–74.
90. Lin Y, Golovnina K, Chen ZX, et al. Comparison of normalization and differential expression analyses using RNA-seq data from 726 individual *Drosophila melanogaster*. *BMC Genomics* 2016;17:28.
91. Li B, Tang J, Yang Q, et al. Performance evaluation and online realization of data-driven normalization methods used in LC/MS based untargeted metabolomics analysis. *Sci Rep* 2016;6:38881.
92. Geiger T, Wehner A, Schaab C, et al. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol Cell Proteomics* 2012;11:M111.014050.
93. Troyanskaya O, Cantor M, Sherlock G, et al. Missing value estimation methods for DNA microarrays. *Bioinformatics* 2001;17:520–5.
94. Gan X, Liew AW, Yan H. Microarray missing data imputation based on a set theoretic framework and biological knowledge. *Nucleic Acids Res* 2006;34:1608–19.
95. Griffin NM, Yu J, Long F, et al. Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nat Biotechnol* 2010;28:83–9.
96. Kuharev J, Navarro P, Distler U, et al. In-depth evaluation of software tools for data-independent acquisition based label-free quantification. *Proteomics* 2015;15:3140–51.

97. Chignell JF, Park S, Lacerda CMR, et al. Label-free proteomics of a defined, binary co-culture reveals diversity of competitive responses between members of a model soil microbial system. *Microb Ecol* 2018;**75**:701–19.
98. 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.
99. Dubinkina VB, Tyakht AV, Odintsova VY, et al. Links of gut microbiota composition with alcohol dependence syndrome and alcoholic liver disease. *Microbiome* 2017;**5**:141.
100. Arroyo V, Gines P. Prostaglandins and the treatment of hepatorenal syndrome in cirrhosis. *J Hepatol* 1990;**11**: 142–4.
101. 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.
102. Komatsu S, Han C, Nanjo Y, et al. Label-free quantitative proteomic analysis of abscisic acid effect in early-stage soybean under flooding. *J Proteome Res* 2013;**12**: 4769–84.
103. Blaise BJ. Data-driven sample size determination for metabolic phenotyping studies. *Anal Chem* 2013;**85**:8943–50.
104. Suomi T, Seyednasrollah F, Jaakkola MK, et al. ROTS: an R package for reproducibility-optimized statistical testing. *PLoS Comput Biol* 2017;**13**:e1005562.
105. Min CW, Lee SH, Cheon YE, et al. In-depth proteomic analysis of Glycine max seeds during controlled deterioration treatment reveals a shift in seed metabolism. *J Proteomics* 2017;**169**:125–35.
106. Li B, Tang J, Yang Q, et al. NOREVA: normalization and evaluation of MS-based metabolomics data. *Nucleic Acids Res* 2017;**45**:W162–70.
107. Wang X, Gardiner EJ, Cairns MJ. Optimal consistency in microRNA expression analysis using reference-gene-based normalization. *Mol Biosyst* 2015;**11**:1235–40.
108. Zevin AS, Xie IY, Birse K, et al. Microbiome composition and function drives wound-healing impairment in the female genital tract. *PLoS Pathog* 2016;**12**:e1005889.
109. Xia J, Wishart DS. Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. *Nat Protoc* 2011;**6**:743–60.
110. Lee DY, Saha R, Yusufi FN, et al. Web-based applications for building, managing and analysing kinetic models of biological systems. *Brief Bioinform* 2009;**10**: 65–74.
111. Vizcaino JA, Csordas A, del-Toro N, et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res* 2016;**44**:D447–56.
112. 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.
113. Schroeder BO, Birchenough GMH, Stahlman M, et al. Bifidobacteria or fiber protects against diet-induced microbiota-mediated colonic mucus deterioration. *Cell Host Microbe* 2018;**23**:27–40.
114. Deusch S, Camarinha-Silva A, Conrad J, et al. A structural and functional elucidation of the rumen microbiome influenced by various diets and microenvironments. *Front Microbiol* 2017;**8**:1605.
115. Grassl N, Kulak NA, Pichler G, et al. Ultra-deep and quantitative saliva proteome reveals dynamics of the oral microbiome. *Genome Med* 2016;**8**:44.
116. van der Post S, Hansson GC. Membrane protein profiling of human colon reveals distinct regional differences. *Mol Cell Proteomics* 2014;**13**:2277–87.
117. Poppleton DI, Duchateau M, Hourdel V, et al. Outer membrane proteome of *Veillonella parvula*: a diderm firmicute of the human microbiome. *Front Microbiol* 2017;**8**:1215.
118. Scott NE, Giogha C, Pollock GL, et al. The bacterial arginine glycosyltransferase effector NleB preferentially modifies Fas-associated death domain protein (FADD). *J Biol Chem* 2017;**292**:17337–50.
119. Chong PK, Gan CS, Pham TK, et al. Isobaric tags for relative and absolute quantitation (iTRAQ) reproducibility: implication of multiple injections. *J Proteome Res* 2006;**5**: 1232–40.
120. La Delfa NJ, Freeman CC, Petrucci C, et al. Equations to predict female manual arm strength based on hand location relative to the shoulder. *Ergonomics* 2014;**57**:254–61.
121. Simula MP, Cannizzaro R, Marin MD, et al. Two-dimensional gel proteome reference map of human small intestine. *Proteome Sci* 2009;**7**:10.
122. Anjo SI, Santa C, Manadas B. SWATH-MS as a tool for biomarker discovery: from basic research to clinical applications. *Proteomics* 2017;**17**:1600278.
123. Liu Y, Buil A, Collins BC, et al. Quantitative variability of 342 plasma proteins in a human twin population. *Mol Syst Biol* 2015;**11**:786.
124. Caron E, Roncagalli R, Hase T, et al. Precise temporal profiling of signaling complexes in primary cells using SWATH mass spectrometry. *Cell Rep* 2017;**18**:3219–26.
125. Sullivan KD, Evans D, Pandey A, et al. Trisomy 21 causes changes in the circulating proteome indicative of chronic autoinflammation. *Sci Rep* 2017;**7**:14818.
126. 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.
127. Tan SLW, Chadha S, Liu Y, et al. A class of environmental and endogenous toxins induces BRCA2 haploinsufficiency and genome instability. *Cell* 2017;**169**: 1105–18.
128. Liu Y, Gonzalez-Porta M, Santos S, et al. Impact of alternative splicing on the human proteome. *Cell Rep* 2017;**20**: 1229–41.
129. 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.
130. McIlwain S, Mathews M, Bereman MS, et al. Estimating relative abundances of proteins from shotgun proteomics data. *BMC Bioinformatics* 2012;**13**:308.
131. Mouton-Barbosa E, Roux-Dalvai F, Bouyssie D, et al. In-depth exploration of cerebrospinal fluid by combining peptide ligand library treatment and label-free protein quantification. *Mol Cell Proteomics* 2010;**9**:1006–21.
132. von Haller PD, Yi E, Donohoe S, et al. The application of new software tools to quantitative protein profiling via isotope-coded affinity tag (ICAT) and tandem mass spectrometry: II. Evaluation of tandem mass spectrometry methodologies for large-scale protein analysis, and the application of statistical tools for data analysis and interpretation. *Mol Cell Proteomics* 2003;**2**:428–42.
133. Kosanam H, Thai K, Zhang Y, et al. Diabetes induces lysine acetylation of intermediary metabolism enzymes in the kidney. *Diabetes* 2014;**63**:2432–9.

134. Roux-Dalvai F, Gonzalez de Peredo A, Simo C, et al. Extensive analysis of the cytoplasmic proteome of human erythrocytes using the peptide ligand library technology and advanced mass spectrometry. *Mol Cell Proteomics* 2008;**7**:2254–69.
135. Weisser H, Nahnsen S, Grossmann J, et al. An automated pipeline for high-throughput label-free quantitative proteomics. *J Proteome Res* 2013;**12**:1628–44.
136. 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.
137. Qi D, Brownridge P, Xia D, et al. A software toolkit and interface for performing stable isotope labeling and top3 quantification using Progenesis LC-MS. *OMICS* 2012;**16**:489–95.
138. Searle BC. Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics* 2010;**10**:1265–9.
139. Schilling B, Rardin MJ, MacLean BX, et al. Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation. *Mol Cell Proteomics* 2012;**11**:202–14.
140. Kaur I, Kaur J, Sooraj K, et al. Comparative evaluation of the aqueous humor proteome of primary angle closure and primary open angle glaucomas and age-related cataract eyes. *Int Ophthalmol* 2018. doi: [10.1007/s10792-017-0791-0](https://doi.org/10.1007/s10792-017-0791-0).
141. Sialana FJ, Wang AL, Fazari B, et al. Quantitative proteomics of synaptosomal fractions in a rat overexpressing human DISC1 gene indicates profound synaptic dysregulation in the dorsal striatum. *Front Mol Neurosci* 2018;**11**:26.
142. Park SK, Aslanian A, McClatchy DB, et al. Census 2: isobaric labeling data analysis. *Bioinformatics* 2014;**30**:2208–9.
143. Rost HL, Liu Y, D'Agostino G, et al. TRIC: an automated alignment strategy for reproducible protein quantification in targeted proteomics. *Nat Methods* 2016;**13**:777–83.
144. Fu J, Tang J, Wang Y, et al. Discovery of the consistently well-performed analysis vha in for SWATH-MS based pharmacoproteomic quantification. *Front Pharmacol* 2018;**9**:681.
145. Lambert JP, Ivosev G, Couzens AL, et al. Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition. *Nat Methods* 2013;**10**:1239–45.
146. Sherrod SD, Myers MV, Li M, et al. Label-free quantitation of protein modifications by pseudo selected reaction monitoring with internal reference peptides. *J Proteome Res* 2012;**11**:3467–79.