Simultaneous Improvement in the Precision, Accuracy, and Robustness of Label-free Proteome Quantification by Optimizing Data Manipulation Chains

Authors

Jing Tang, Jianbo Fu, Yunxia Wang, Yongchao Luo, Qingxia Yang, Bo Li, Gao Tu, Jiajun Hong, Xuejiao Cui, Yuzong Chen, Lixia Yao, Weiwei Xue, and Feng Zhu

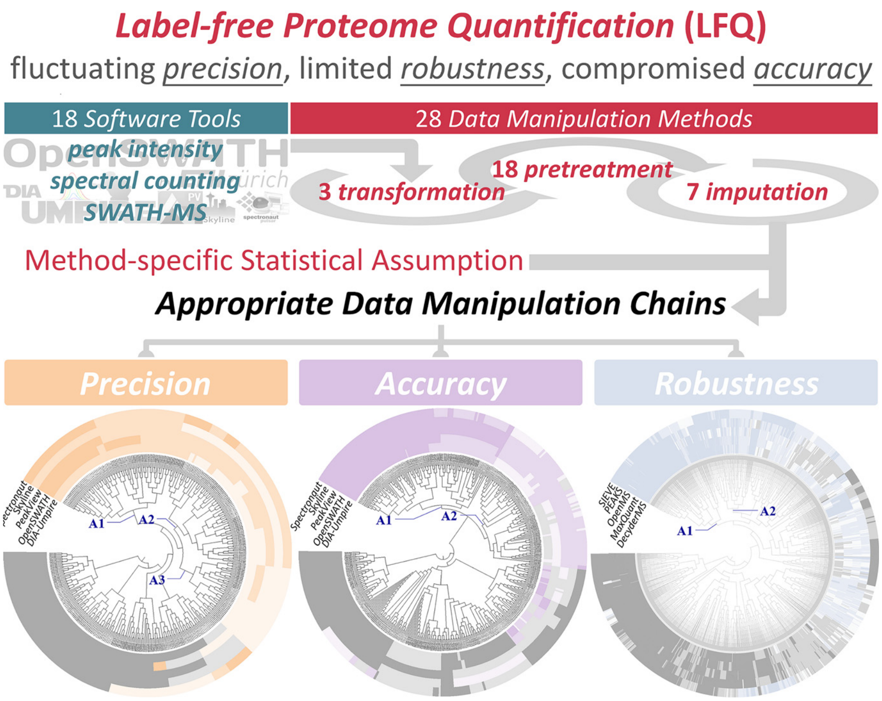
Correspondence

zhufeng@zju.edu.cn;
prof.zhufeng@gmail.com

In Brief

High-quality label-free proteome quantification (LFQ) is valuable for clinical and pharmaceutical studies yet remains extremely challenging despite technical advances. Particularly, fluctuating precision, limited robustness, and compromised accuracy are known issues. Here, we described and validated a new strategy enabling the discovery of the LFQs of simultaneously enhanced precision, robustness, and accuracy from thousands of LFQ manipulation chains. In the proof-of-concept study, this strategy showed superior ability in identifying well-performing LFQs. An online tool incorporating this novel strategy was also developed.

Graphical Abstract



Highlights

- High-quality LFQ is valuable technique yet remains extremely challenging.
- Fluctuating precision, limited robustness, and compromised accuracy are known issues.
- We proposed a strategy collectively improving LFQ precision, robustness, and accuracy.
- An online tool incorporating this novel strategy was also developed.

Tang et al., 2019, Molecular & Cellular Proteomics *18*, 1683–1699
August 2019 © 2019 Tang et al. Published under exclusive license by The American Society for Biochemistry and Molecular Biology, Inc.
https://doi.org/10.1074/mcp.RA118.001169

Simultaneous Improvement in the Precision, Accuracy, and Robustness of Label-free Proteome Quantification by Optimizing Data Manipulation Chains*[S]

Jing Tang‡§¶§§, Jianbo Fu‡§§, Yunxia Wang‡§§, Yongchao Luo‡, Qingxia Yang‡§, Bo Li§, Gao Tu‡§, Jiajun Hong‡, Xuejiao Cui§, Yuzong Chen‖, Lixia Yao**, Weiwei Xue§, and Feng Zhu‡§‡‡

The label-free proteome quantification (LFQ) is multistep workflow collectively defined by quantification tools and subsequent data manipulation methods that has been extensively applied in current biomedical, agricultural, and environmental studies. Despite recent advances, in-depth and high-quality quantification remains extremely challenging and requires the optimization of LFQs by comparatively evaluating their performance. However, the evaluation results using different criteria (precision, accuracy, and robustness) vary greatly, and the huge number of potential LFQs becomes one of the bottlenecks in comprehensively optimizing proteome quantification. In this study, a novel strategy, enabling the discovery of the LFQs of simultaneously enhanced performance from thousands of workflows (integrating 18 quantification tools with 3,128 manipulation chains), was therefore proposed. First, the feasibility of achieving simultaneous improvement in the precision, accuracy, and robustness of LFQ was systematically assessed by collectively optimizing its multistep manipulation chains. Second, based on a variety of benchmark datasets acquired by various quantification measurements of different modes of acquisition, this novel strategy successfully identified a number of manipulation chains that simultaneously improved the performance across multiple criteria. Finally, to further enhance proteome quantification and discover the LFQs of optimal performance, an online tool (https://idrblab.org/anpela/) enabling collective performance assessment (from multiple perspectives) of the entire LFQ workflow was developed. This study confirmed the feasibility of achieving simultaneous improvement in precision, accuracy, and robustness. The novel strategy proposed and validated in this study together with the online tool might provide useful guidance for the research field requiring the mass-spectrometry-based LFQ technique. ***Molecular & Cellular Proteomics 18: 1683–1699, 2019. DOI: 10.1074/mcp.RA118.001169.***

Mass-spectrometry (MS)-based proteomics have emerged as one of the most powerful techniques to detect the correlation of complex molecular network rewiring with biological phenotypes (1), identify the human protein interactome (2), discover novel therapeutic targets (3), and characterize new protein therapeutics (4). Compared with qualitative proteomics, the quantitative MS-based proteomics is distinguished by its ability to give detailed level of protein intensities and can thus provide a more complete picture of cellular process and enable network-centered study (5). Among the approaches currently available for quantifying proteins, the label-free proteome quantification (LFQ)[1] shows unique advantages in (*a*) the simultaneous discovery of proteins without the labor-intensive and costly procedures of stable isotope labeling (6), (*b*) the capacity to process the large sample cohort (6, 7), and (*c*) its applicability to samples from any source (5, 8). Due to its distinguishing features above, the LFQ has become increasingly important in and has been extensively applied to a broad range of research fields, such as system biology (9), drug discovery (3), and agricultural (10), medical (11), and environmental (12) sciences.

However, in-depth and high-quality quantification remains extremely challenging despite recent advances in LFQ technique (13). The challenges include snowballing missing values (14), fluctuating precision (15), limited robustness (16), and compromised accuracy (6), among others. On one hand, a variety of powerful quantification tools integrating novel align-

From the ‡College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China; §School of Pharmaceutical Sciences, Chongqing University, Chongqing 401331, China; ¶ Department of Bioinformatics, Chongqing Medical University, Chongqing 400016, China; ‖ Department of Pharmacy, National University of Singapore, Singapore 117543, Singapore; ** Department of Health Sciences Research, Mayo Clinic, Rochester MN 55905

Received October 29, 2018, and in revised form, April 28, 2019
Published, MCP Papers in Press, May 16, 2019, DOI 10.1074/mcp.RA118.001169

© 2019 Tang et al. Published under exclusive license by The American Society for Biochemistry and Molecular Biology, Inc.

ment and feature generation methods have been constructed to maximally fill in missing blanks and elevate the reproducibility of LFQ (17–19). On the other hand, many computational algorithms have been proposed to minimize the instrumental and experimental fluctuations (5), enhance the stability of the identified candidate markers (20), and reduce the extremely large dynamic range of the protein abundance (21). Specifically, ≥18 quantification software tools and ≥3 transformation, ≥18 pretreatment, and ≥7 missing-value imputation methods have been proposed and sequentially integrated into the quantification workflow to overcome the above challenges to the extent possible (supplemental Table S1). Recent investigation has revealed that the quality of LFQ relies heavily on the selection of tools or methods and, in turn, greatly affects the retrieval of reliable biological interpretation from the proteomic data (6). Therefore, it is essential to evaluate the correctness and further optimize the qualities of LFQs by comparatively benchmarking their performance (22).

To date, several studies have been conducted to evaluate the performance variations in quantification tools by assessing the number of proteins quantified (23, 24), the number of missing values produced (24), and the precision (25) or accuracy (26). Some well-performing tools are thus identified based on the criterion adopted by each study, but their identification results vary greatly (PEAKS and Progenesis were found to perform well in quantifying proteins (23) and reproducing missing values (24), respectively; OpenMS and MaxQuant were discovered superior in quantification precision (23) and accuracy (25, 26), respectively). These findings indicate significant variations among the evaluation results by different criteria. Moreover, the performance of seven normalization methods has been assessed using the quantification precision (27). However, the random, comprehensive, and sequential integration of all quantification tools, transformation, pretreatment, and imputation methods can result in thousands of potential LFQs (supplemental Table S1). The multiplicity and complexity therefore become one of the bottlenecks in the comprehensive assessment of all potential LFQs.

Precision (28) and robustness (29) are two well-established criteria for assessing the quality of LFQs that should be collectively considered to not only reduce proteome variations among replicates but also elevate consistency among the different lists of identified markers (24, 27). In the meantime, it was also necessary to consider both the quantification precision (28) and the deviations of spiked proteins from their expected abundance ratio (quantification accuracy (6)) for current proteomic analyses (27, 28). Due to the variation in the assessment results by different criteria and the vast number of potential LFQs (discussed in the previous paragraph), it is essential to assess the feasibility of identifying the LFQs that simultaneously enhance the precision, accuracy, and robustness and to design novel strategy for "thoroughly" evaluating (6, 24, 30) and collectively optimizing the performance of LFQs. Ideally, an additional tool with the unique functions of performance assessments (discussed in the preceding) should be constructed to facilitate current proteomic analyses. However, no such study has been conducted yet.

Herein, the feasibility of achieving simultaneous improvements in the precision, accuracy, and robustness of an LFQ was systematically assessed by collectively optimizing its multistep analyzing chain, and a new strategy was proposed by scanning quantification tools and all possible combinations of data manipulation chains. Based on the representative benchmark data acquired using different quantification measurements, this strategy successfully identified a number of data manipulation chains with simultaneous enhancement of performance in terms of multiple criteria. To facilitate proteome quantifications and discover the LFQs of the optimal performance, an online tool enabling collective performance assessment was developed.

MATERIALS AND METHODS

The Modes of Acquisition and Their Corresponding Quantification Tools Studied in This Work—The LFQ specified in this study was a multistep workflow that included the data preprocessing by various quantification tools and subsequent data manipulation chains. There were ≥18 quantification tools popular in current proteomics research, the majority of which were especially designed for preprocessing the data acquired by specific mode of acquisition (data-dependent (DDA) and data-independent (DIA) acquisitions). For the DDA mode, there were two quantification measurements: peak intensity and spectral counting (31), and the measurement sequential window acquisition of all theoretical mass spectra (SWATH-MS) belonged to the DIA mode (32). As described in supplemental Table S1, only two quantification tools (MaxQuant and MFPaQ) were capable of preprocessing the data of multiple quantification measurements. The subsequent data manipulation chain included a sequential process of three steps: transformation, pretreatment, and missing-value imputation (supplemental Table S1). Overall, the LFQ analyzed here is a multistep workflow collectively defined by sequential integration of quantification tools, transformation, pretreatment, and missing-value imputation methods.

Constructing the Manipulation Chains by Sequentially Integrating the Manipulation Methods—To the best of our knowledge, there were ≥3 transformation, ≥18 pretreatment (2 centering, 4 scaling, and 12 normalization methods), and ≥7 missing-value imputation methods sequentially integrated to construct a multistep (three-step in general and five-step in detail) manipulation chain. In particular, the 28 methods in the manipulation chain included three transformation: Box-Cox, LOG, and VSN; 18 pretreatment: (two centering: mean and median; four scaling: auto, Pareto, vast, and range; and 12 normalization: mean, median, MAD, TIC, cyclic Loess, linear baseline scaling, RLR, LOWESS, EigenMS, PQN, quantile, and TMM); and seven imputation: background, BPCA, censored, KNN, LLS, SVD, and zero. As a transformation method integrating with subsequent normalization technique, the VSN was unique in determining data-dependent

[1] The abbreviations used are: LFQ, label-free proteome quantification; MS, mass spectrometry; DDA, data-dependent acquisition; DIA, data-independent acquisition; PMAD, pooled intragroup median absolute deviation; logFC, logarithmic fold change; ROTS, reproducibility-optimized test statistic; PRIDE, The PRIDE PRoteomics IDEntifications database; SWATH-MS, sequential window acquisition of all theoretical mass spectra.

transformation parameters by having a built-in transformation. For the convenience of the discussions in this study, each manipulation method was abbreviated by a three-letter code as shown in Table I. Within a particular step, if none of the method was used, a three-letter code, NON, was applied to indicate the nonapplication of any method in this step. The representative application of the manipulation methods in current proteomics was systematically reviewed and provided in supplemental Table S1.

Moreover, the manipulation methods were reported as based on their own statistical assumption about the data, which might make them inappropriate for manipulating some proteomic data (33, 34). Therefore, the corresponding statistical/biological assumption and purpose of data manipulation of methods in each step were systematically reviewed and are provided in Table I. Taking pretreatment methods as the example, there were generally three types of assumptions: (A-α) all proteins were assumed to be equally important, which was required by the appropriate application of both centering and scaling methods (35); (A-β) the level of protein abundance was assumed to be constant among all samples, which was a priori assumption for some normalization methods, including MEA, MED, MAD, and TIC (27, 36–38); and (A-γ) the intensities of the vast majority of the proteins were assumed to be unchanged under the studied conditions, which was demanded by some other normalization methods such as CYC, LIN, LOW, PQN, QUA, RLR, and TMM (27, 39–42). Among 12 normalization methods, the EIG was the only one with no priori assumption about the relative strength of signals due to each source of variation (43). It is important to emphasize that due to the distinct assumptions, some methods may be fundamentally inappropriate for certain datasets and cannot be assessed in this study (33, 34). Therefore, before any performance assessment in the Discussion section of this study, the nature of the studied dataset was first analyzed and whether the method's assumption held for these data was then discussed. Moreover, in the online tool developed for proteomic quantification and performance assessment, a default setting for confirming whether the method's assumption held for the analyzed dataset is provided in both the online and the local versions of the constructed tool.

Based on the preanalyses on the nature of the studied dataset and its obedience toward the method's assumption, 28 manipulation methods were screened, and only the ones with their assumption held for studied datasets were kept for the subsequent assessments based on performance. Theoretically, a random, comprehensive, and sequential integration of 27 methods (excluding VSN) could result in 3,120 manipulation chains of five steps ($2 \times 3{\times}5 \times 13 \times 8 = 3{,}120$, taking noncentering, nonscaling, nonnormalization, and nonimputation into account, which have been widely adopted in previous publications (44–46)). Transformation was reported to be essential prior to the downstream analysis in any proteomic study (47); nontransformation was thus not allowed in both the analyses of this study and the online tool. Moreover, because the VSN was unique in having a built-in transformation and subsequent normalization technique, it should not combine with other transformation or pretreatment methods and could only be combined with imputation. These combinations therefore resulted in eight additional manipulation chains. As a result, there were 3,128 potential manipulation chains, and detailed descriptions of all manipulation methods can be found in Supplementary Methods.

Multiple Criteria for Assessing the Performance of a Given Data Manipulation Chain—Several well-established and widely adopted criteria in current proteomics were collected in this study for ensuring a systematic performance assessment of a given data manipulation chain. Three criteria (precision, robustness, and accuracy) were used for quantitative assessment, while two others (classification capacity and differential abundance analysis in proteomics) were employed as qualitative evaluation. A description of these criteria, together with their corresponding metrics, can be found in Supplementary Methods.

(a) Precision Measuring the Reduction in Proteome Variation among Replicates—The quantification precision of the LFQ was profoundly affected by different modes of acquisition, various types of software for preprocessing raw proteomic data, and diverse methods for data manipulation, which could be evaluated by the pooled intragroup median absolute deviation (PMAD) of the protein intensities among replicates (6, 48). PMAD reflected LFQ's ability to reduce the variation among replicates and thus enhance technical reproducibility (28). Lower PMAD denoted more thorough removal of experimentally induced noise and indicated better precision of the corresponding LFQ and manipulation chain (15).

(b) Robustness among Different Sets of Protein Biomarkers Identified from Different Datasets—The robustness of the biomarker discovery from proteomic data was usually assessed by the popular metric: consistency score, which calculated the level of reproducibility among the different lists of biomarkers identified from the different partitions of the studied dataset (16, 29). A higher consistency score value indicated the more robust results in the biomarker discovery (29). Herein, the studied datasets were randomly sampled 50 times to generate multiple subdatasets. Then, each protein was ranked based on its statistical significance measured by *q*-value and fold change. Third, the top-ranked proteins in each subdataset were selected as markers. Finally, a consistency score was calculated based on its reported and well-established equation (16).

(c) Accuracy Assessing the Deviation of Spiked Proteins from Their Expected Abundance Ratio—To adjust/validate the quantification accuracy of LFQs, additional experimental data (*e.g.* spiked proteins) were created and applied as the golden references (6, 48), and the expected log fold changes (logFCs) for both the spiked and background proteins were essential for assessing quantification accuracy (the expected logFC for the background proteins should equal to zero) (24). Herein, the logFCs of protein intensities for both spiked and background proteins between distinct sample groups were first calculated. Then, the mean squared error was applied to assess the level of correspondence between the quantification and the expected logFC. The performance could be reflected by how well the quantification logFCs corresponded to the expected values using the references (24). The deviations in both quantification and expected logFCs would be zero with the minimized deviation (24).

(d) Qualitative Criteria for Assessing the Performance of Data Manipulation Chain—The qualitative criteria frequently adopted by previous report and applied in this study for assessing LFQ's performance included the LFQ's (1) classification capacity between distinct groups (49) and (2) differential abundance analysis in proteomics based on reproducibility optimization (50). On one hand, an appropriate manipulation chain was expected to retain or even enlarge the variation in proteomic data between distinct sample groups (49, 51), and a heatmap hierarchically clustering samples based on their protein abundances was thus used as an effective measure for assessing LFQ's classification capacity (49). On the other hand, to avoid overfitting/confounding, the distribution of *p* values of protein intensities between distinct sample groups should be examined using differential abundance analysis in proteomics (50) (ideally, one expected an uniform distribution for the bulk of nondifferentially expressed proteins, with a peak in the interval of [0.00, 0.05] corresponding to proteins with differential intensity (50)), and a volcano plot coloring proteins by differential intensity was thus drawn to depict the total number of the proteins of differential abundance (24). In the OMIC (any research field of biological study ending in -omics such as genomics, transcriptomics, proteomics or metabolomics) study exploring the mechanism underlying complex processes, a limited num-

TABLE I

Various manipulation methods including 3 transformation, 18 pretreatment (2 centering, 4 scaling, and 12 normalization methods), and 7 imputation methods together with their purpose of data manipulation and/or corresponding statistical/biological assumptions. All manipulation methods were abbreviated using three-letter code. As a transformation method integrating with subsequent normalization technique, the VSN determined data-dependent transformation parameters by having a built-in transformation (47, 79)

Classes	Abb.	Manipulation Method	Purpose/Assumption of the Manipulation Method
Transformation	**BOX**	Box-Cox Transformation	Making asymmetric data fulfill the normality assumption in a regression model by converting the protein abundances into a more symmetric distribution (80).
	LOG	Log Transformation	Converting the distribution of ratios of abundance values of proteins into a more symmetric (almost normal distribution) and minimizing the effect of proteins with extreme abundance (39).
	VSN	Variance Stabilization Normalization	Having a built-in transformation (47) and making individual observations more directly comparable (81); assuming that most of the proteins across different samples are not differentially expressed (81).
Pretreatment			
Centering	**MEC**	Mean-Centering	Converting all the intensities to fluctuations around zero instead of around the mean of the protein intensities; assuming all proteins are equally important (35).
	MDC	Median-Centering	Making all the intensities to fluctuations around zero instead of around the median of the protein intensities; assuming all proteins are equally important (35).
Scaling	**ATO**	Auto Scaling	Adjusting each protein abundance for systematic variance using the standard deviation of each protein of all samples as scaling factor (82); assuming all proteins are equally important (35).
	PAR	Pareto Scaling	Scaling each protein abundance for systematic variance using the square root of the standard deviation of each protein of all samples as scaling factor (83); assuming all proteins are equally important (35).
	VAS	Vast Scaling	Adjusting each protein abundance for systematic variance using the coefficient of variation of each protein of all samples as scaling factor (84); assuming all proteins are equally important (35).
	RAN	Range Scaling	Scaling each protein abundance for systematic variance using the abundance range of each protein of all samples as scaling factor (85); assuming all proteins are equally important (35).
Normalization	**MEA**	Mean Normalization	Ensuring the protein abundance values from all studied samples directly comparable with each other (86); assuming the mean level of the protein abundance is constant for all samples (27).
	MED	Median Normalization	Making the protein intensities from all individual samples directly comparable with each other (27, 86); assuming the median level of the protein abundance is constant for all samples (36).
	MAD	Median Absolute Deviation	Ensuring the comparability of protein intensities among all samples (86); assuming the median level of the protein abundance and the spread of abundances are the same in all samples (37).
	TIC	Total Ion Current	Making the protein intensities from all samples directly comparable with each other (86); assuming the total area under the protein abundance curve is constant among samples (38).
	CYC	Cyclic Loess	Assuming that the intensities of the vast majority of the proteins are not changed in control and case groups (33, 87) and the systematic bias is nonlinearly dependent on the protein abundances (27).
	LIN	Linear Baseline Scaling	Assuming that the abundances of the majority of the proteins in samples are unchanged under the studied condition (33, 88) and the systematic bias is linearly dependent on the protein intensities (39).
	RLR	Robust Linear Regression	Assuming that the intensities of the majority of the proteins are not changed in control and case groups (87, 88) and the systematic bias is linearly dependent on the magnitude of protein abundances (27).
	LOW	Locally Weighted Scatterplot Smoothing	Assuming that the abundances of the majority of the proteins are unchanged under the studies circumstances (40, 87) and the systematic bias is nonlinearly dependent on the protein intensities (40).
	EIG	EigenMS	Overcoming the problems caused by the heterogeneity in the protein intensities of studied samples (43, 89); Does not require any assumption about the relative strength of signals due to each source of variation (43).
	PQN	Probabilistic Quotient Normalization	Ensuring the comparability of protein intensities among all samples (86); assuming that the majority of the protein intensities does not vary for the studied classes (41).

TABLE I—*continued*

Classes	Abb.	Manipulation Method	Purpose/Assumption of the Manipulation Method
	QUA	Quantile Normalization	Making the protein intensities from all samples directly comparable with each other (86); assuming that the majority of protein intensity signals are unchanged among samples (40).
	TMM	Trimmed Mean of M Values	Ensuring the protein abundance values from all studied samples directly comparable with each other (86); assuming the majority of proteins are not differentially expressed between control and case groups (42).
Imputation	**BAK**	Background Imputation	Assuming that the protein values are missing because of having small concentrations in the sample and thus cannot be detected during the MS run (27).
	BPC	Bayesian Principal Component Imputation	Imputing based on the variational Bayesian framework that does not force orthogonality between the principal components (90).
	CEN	Censored Imputation	Imputing the lowest intensity values in the dataset by assuming that the missing of protein values is because of being below detection capacity (27).
	KNN	K-nearest Neighbor Imputation	Finding *k* most similar proteins (*k*-nearest neighbors) and using a weighted average over these *k* proteins to estimate the missing protein values (27, 91).
	LLS	Local Least Squares Imputation	Representing a studied protein that has missing values as a linear combination of a number of proteins similar to this particular protein (92).
	SVD	Singular Value Decomposition	Applying this imputation method to the data to obtain sets of mutually orthogonal expression patterns of all proteins in the data (91).
	ZER	Zero Imputation	Imputing the missing intensities of the studied proteins by directly replacing these missing values with a number of zeros (27).

ber of the proteins of differential abundance might lead to false discovery (52). Thus, to assess LFQ's classification capacity, the number of protein intensities in each sample was first reduced by feature selection. First, the differential significance of each protein between distinct sample groups measured by adjusted *p* value was calculated using the reproducibility-optimized test statistic (ROTS) package (53); then, the significant features (adjusted *p* value <0.05) were selected for subsequent heatmap analyses. Then, the proteins (rows) and samples (columns) were clustered based on their similarity in the profile of protein abundances. A corresponding process was described by a previous report (49). Moreover, to conduct differential abundance analysis in proteomics, differential significance of protein intensities between distinct sample groups measured by *p* value was first calculated using ROTS in R package (54). Then, the distribution of *p* values was visualized by histogram; skewed distribution might indicate overfitting/confounding in the studied manipulation chain (55).

Optimizing the Performance of Manipulation Chain by Assessing from Multiple Perspectives—Each criterion made the performance assessment possible from its own perspective, and the combination of multiple criteria could thus ensure the comprehensive evaluation of a given manipulation chain. On one hand, quantification precision and robustness were collectively considered to evaluate the performance of chain on the proteomic datasets acquired by either DDA or DIA. On the other hand, precision and accuracy were evaluated for DDA-based or DIA-based proteomic datasets. Moreover, an online tool was developed to provide the performance assessment from quantitative and qualitative perspectives (all criteria described in previous sections could be simultaneously evaluated). To optimize the performance of LFQ, potential manipulation chains with suitable assumptions appropriate to the studied datasets were systematically scanned, and the chains top-ranked by single or multiple criteria were identified as demonstrating the optimal performance.

Identification of the Well-Performing Manipulation Chains Using Hierarchical Clustering—PMAD values, consistency scores, or deviations from the expected abundance ratio of manipulation chains across different benchmark datasets (preprocessed by different quantification tools) were first calculated. Hierarchical clustering of the chains with calculable results of all benchmarks was conducted to identify the chains consistently performing well across all benchmarks. Particularly, the PMAD values, consistency scores, or deviations from the expected abundance ratio of a given chain among *N* sets of benchmark data were used to construct *N*-dimensional vectors. Then, hierarchical clustering was adopted to investigate the relationship among these vectors, and therefore among the corresponding chains. Manhattan distance was adopted to measure the distance between any two vectors (56). To view the hierarchical tree among chains, the tree generator *iTOL* was used to generate and display the tree structure (57).

Diverse and Representative Benchmark Datasets Collected for the Analyses in This Study—In total, 58 benchmark datasets from 11 studies were collected and applied to evaluate the performance of manipulation chains. As shown in Table II, the first study provided the peak intensity proteome data based on the cerebrospinal fluids from 10 Alzheimer's disease patients and 10 nondemented controls (23). There were five proteomic benchmark datasets (preprocessed by five quantification tools: DecyderMS, MaxQuant, OpenMS, PEAKS, and Sieve) in this study. The second study described a spectral-counting-based controlled, spiked proteomic data for which the ground truth of variant protein was known (58). Particularly, 48 UPS1 proteins were spiked into the yeast lysate with five different concentrations (0.5, 5, 12.5, 25, and 50 fmol/μg), and the samples were preprocessed using four quantification tools: IRMa-hEIDI, MaxQuant, MFPaQ, and Scaffold. For each tool, a random combination of five concentrations resulted in 10 datasets with two distinct concentrations. As a result, 40 datasets in total were collected for four quantification tools. The proteomics datasets of the third through fifth studies (59–61) were acquired by the DDA (peak intensity) technique and quantified by MaxQuant. Meanwhile, the datasets from the sixth through eighth studies (54, 62, 63) were acquired by DDA (peak intensity) technique and quantified by Progenesis. These six studies therefore resulted in six benchmark datasets. The ninth study gave one peak intensity proteomic dataset with six purified proteins spiked with two distinct concentrations (64), which was quantified by OpenMS. The 10th study

TABLE II

Detailed information of the studied benchmarks (ordered by their appearances in "Results and Discussion"). Three types of assumptions held for these benchmarks: (A-α) all proteins are equally important; (A-β) the level of protein abundance is constant among all samples; (A-γ) the intensities of the majority of the proteins are not changed under the studied conditions. The assumption A-α held for all benchmarks, while both assumption A-β and A-γ could not hold for the datasets of the fourth study. The corresponding quantification measurements and applied quantification software tool(s) were also provided

No.	Benchmark Studies	PRIDE or Other IDs	Description of Benchmark Study	Assumption Held	No. of Proteins	Quantification Measurement (Applied Quantification Tools)
1	*PLoS One* 11: e0150672, 2016	Shevchenko	10 Alzheimer's disease patients 10 nondemented controls	**A-α**, **A-β**, **A-γ**	4,967	Peak Intensity (DecyderMS, MaxQuant, OpenMS, PEAKS, Sieve)
2	*Data Brief* 6:286–294, 2015	PXD001819	48 UPS1 proteins spiked into the yeast lysate with five concentrations	**A-α**, **A-β**, **A-γ**	752	Spectral Counting (IRMa-hEIDI, MaxQuant, MFPaQ, Scaffold)
3	*Nat. Commun*. 7:13419, 2016	PXD002882	21 Crohn's disease patients 10 healthy individuals	**A-α**, **A-β**, **A-γ**	4,169	Peak Intensity (MaxQuant)
4	*Cell Host Microbe* 23:27–40, 2018	PXD006129	14 western-style diet mice 14 chow-fed mice	**A-α**, **A-β**, **A-γ**	3,243	Peak Intensity (MaxQuant)
5	*Microbiome* 5:144, 2017	PXD006224	60 metabolic-phase fecal samples 24 equilibrium-phase fecal samples	**A-α**, **A-β**, **A-γ**	9,761	Peak Intensity (MaxQuant)
6	*Proteomics* 16:2937–2944, 2016	PXD002885	8 hESC fresh cells in protocol A 8 hESC fresh cells in protocol C	**A-α**, **A-β**, **A-γ**	198	Peak Intensity (Progenesis)
7	*J. Proteome Res*. 11:4118–4126, 2015	PXD002099	48 UPS1 proteins spiked with distinct concentrations (2 *vs* 50 fmol/μl)	**A-α**, **A-β**, **A-γ**	1,442	Peak Intensity (Progenesis)
8	*J. Proteome Res*. 9:761–776, 2010	CPTAC-ST6	48 UPS1 proteins spiked with distinct concentrations (0.25 *vs* 20 fmol/μl)	**A-α**, **A-β**, **A-γ**	1,570	Peak Intensity (Progenesis)
9	*J. Proteome Res*. 16:2964–2974, 2017	PXD006336	6 purified proteins spiked with distinct concentrations (sample group 1 *vs* 2)	**A-α**, **A-β**, **A-γ**	22,719	Peak Intensity (OpenMS)
10	*Cell Rep*. 18:3219–3226, 2017	PXD003972	20 wild type mouse samples 20 knock-in mice expressing GRB2	**A-α**, **A-β**, **A-γ**	901	SWATH-MS (OpenSWATH)
11	*Nat. Biotechnol*. 34:1130–1136, 2016	PXD002952	3 mixtures of 30% yeast *vs* 5% *E. coli* 3 mixtures of 15% yeast *vs* 20% *E. coli*	**A-α**	5,731	SWATH-MS (DIA-Umpire, Skyline, OpenSWATH, PeakView, Spectronaut)

provided the SWATH-MS-based DIA dataset containing 20 samples of wild-type mouse and 20 knock-in mice expressing GRB2 (29), which was preprocessed by OpenSWATH. Finally, the last study described SWATH-MS-based DIA data of two distinct mixture groups (three mixtures of 30% yeast versus 5% *Escherichia coli* and another three mixtures of 15% yeast versus 20% *E. coli*) (6). There were five benchmark datasets (preprocessed by five quantification tools: DIA-Umpire, Skyline, OpenSWATH, PeakView, and Spectronaut) in this study.

RESULTS AND DISCUSSION

Dependence of Quantification Precision on the Selection of Multistep Manipulation Chain—The LFQ precision measured by the pooled intragroup median absolute deviation (PMAD) across replicate groups was widely adopted as an effective measurement of quantification performance (28). To assess the level of dependence of precision on the multistep manipulation chains, two benchmark datasets acquired by either peak intensity or spectral counting were first collected from study 1 (23) and study 2 (58) of Table II. Then, the precision of these datasets was improved by the collective optimization of the multistep manipulation chains. As shown on the left side of Table III, the precision of representative chains for peak intensity benchmark (Table II, study 1) varied greatly (PMADs were from 0.0558 to 3.9600). Based on the precision levels defined by PMAD values (superior: <0.14 (6), good: 0.14~0.3 (65), fair: 0.3~0.7 (66), and poor: >0.7 (66)), the chain adopted by the original study (23) to manipulate the collected dataset (LOG-[NON-NON-MED]-NON, the first line in Table III highlighted in gray) performed consistently "fair" across five quantification tools (PMADs from 0.4708 to 0.6430). Clearly, some chains performed better than the one used in original study (with the LOG-[MEC-RAN-PQN]-BPC performing "superior" for all quantification tools). Similarly, the precision of the representative chains for spectral counting dataset (Table II, study 2 of distinct concentrations: 12.5 versus 25 fmol/μg of the spiked UPS1 protein) is provided in supplemental Table S2. As reported in the original study (58) of this dataset, only log transformation was used (the first line in supplemental Table S2, highlighted in gray color), and it performed consistently "good" across four quantification tools. Still, the precision of some chains was found to surpass that of the original one (with LOG-[MDC-RAN-EIG]-KNN performing "superior" across all quantification software tools).

Based on the preanalyses on the nature of the studied datasets and their obedience toward the assumption of manipulation methods, all methods were appropriate for manipulating the datasets of Table II, studies 1 and 2. On one hand, comprehensive evaluation of the precision of 3,128 potential

TABLE III

The level of dependence of precision and robustness on the selection of manipulation chains assessed based on the benchmark dataset of the first study in Table I (23). These studied LFQs were collectively defined by five quantification tools and eight representative manipulation chains. As reported in original study (23), the log transformation together with median normalization were applied before any subsequent analysis, which defined the manipulation chain of that study as LOG-[NON-NON-MED]-NON (highlighted in gray background). The precision levels were defined by the PMAD values (superior: <0.14, good: 0.14~0.3, fair: 0.3~0.7, and poor: >0.7), and the robustness levels were assigned by the ranking of consistency scores (high: top 20%, medium: 20~50%, and low: bottom 50%). Each manipulation method within an LFQ was abbreviated by a three-letter code that was systematically defined in Table I

Representative Chains for Data Manipulation	Precision: Reproducibility among technical replicates (PMAD)					Robustness: Rank of manipulation chains by consistency scores				
	DecyderMS	MaxQuant	PEAKS	OpenMS	Sieve	DecyderMS	MaxQuant	PEAKS	OpenMS	Sieve
LOG-[NON-NON-MED]-NON	5.56E−01 (Fair)	5.41E−01 (Fair)	6.43E−01 (Fair)	4.80E−01 (Fair)	4.71E−01 (Fair)	457 (High)	358 (High)	369 (High)	313 (High)	687 (Medium)
LOG-[MDC-RAN-EIG]-KNN	1.31E−01 (Superior)	1.40E−01 (Superior)	1.33E−01 (Superior)	1.33E−01 (Superior)	1.20E−01 (Superior)	176 (High)	120 (High)	104 (High)	163 (High)	122 (High)
LOG-[MEC-NON-EIG]-BAK	6.45E−01 (Fair)	7.41E−01 (Poor)	7.85E−01 (Poor)	8.04E−01 (Poor)	4.06E−01 (Fair)	193 (High)	387 (High)	1,421 (Low)	936 (Medium)	75 (High)
LOG-[MEC-RAN-PQN]-CEN	1.64E−01 (Good)	1.94E−01 (Good)	2.64E−01 (Good)	4.64E−01 (Fair)	5.76E−02 (Superior)	1,409 (Low)	1,744 (Low)	1,879 (Low)	1,489 (Low)	2,152 (Low)
LOG-[MEC-RAN-PQN]-BPC	8.19E−02 (Superior)	8.38E−02 (Superior)	8.78E−02 (Superior)	8.41E−02 (Superior)	5.58E−02 (Superior)	2,239 (Low)	2,290 (Low)	2,319 (Low)	2,239 (Low)	2,298 (Low)
BOX-[MDC-PAR-EIG]-ZER	2.61E+00 (Poor)	3.02E+00 (Poor)	2.00E+00 (Poor)	2.74E+00 (Poor)	3.96E+00 (Poor)	174 (High)	138 (High)	278 (High)	228 (High)	2 (High)
BOX-[MEC-RAN-NON]-KNN	2.23E−01 (Good)	2.29E−01 (Good)	2.23E−01 (Good)	2.25E−01 (Good)	2.19E−01 (Good)	474 (High)	473 (High)	379 (High)	451 (High)	332 (High)
BOX-[MEC-ATO-RLR]-BPC	1.64E+00 (Poor)	1.24E+00 (Poor)	1.49E+00 (Poor)	1.46E+00 (Poor)	1.10E+00 (Poor)	2,210 (Low)	2,203 (Low)	2,237 (Low)	2,032 (Low)	2,341 (Low)

manipulation chains for study 1 (peak intensity) found that >1,000 chains (for each quantification tool) demonstrated an enhanced precision compared with the one (LOG-[NON-NON-MED]-NON) adopted by the original study (23). As shown in Fig. 1*A*, 1,022 chains with a consistently better precision than LOG-[NON-NON-MED]-NON for all quantification tools were identified. Ten example chains with substantially enhanced precision are provided in supplemental Table S3 (the significant differences in precision between the original chain and example one can be found in Fig. 1*B*). On the other hand, for study 2 of distinct concentrations (12.5 versus 25 fmol/μg) of the spiked UPS1 proteins (spectral counting), the evaluation of precision of >3,000 potential chains identified 1,095 (Fig. 1*C*) as performing consistently better than the one adopted in the original report (58). Supplemental Table S4 provides the example chains of greatly enhanced precision, and the significant differences in precision between the original chain and the example ones can also be observed in Fig. 1*D* for this spectral-counting-based benchmark. All in all, these findings indicate that the precision was highly dependent on the multistep manipulation chain, and a comprehensive assessment can facilitate the discovery of the well-performing chains.

The identification of the well-performing chains was conducted in previous sections based on the datasets quantified by various tools for a single experiment. However, each tool might generate the dataset of a tool-specific property, which could result in different chains appropriate for each tool. In other words, it would be essential to assess the dependences of the well-performing chains on different tools. Thus, a variety of datasets from multiple experiments but quantified by the same tool were collected. Particularly, eight tools for DDA-based proteomics (MFPaQ, MaxQuant, OpenMS, PEAKs, Progenesis QI, Proteios S.E., Scaffold, and Thermo Proteome Discoverer) were first searched in the PRoteomics IDEntifications (PRIDE) database (67). Then, several criteria were applied to guarantee the availability and processability of the collected raw proteomic data, which included (1) label-free quantification, (2) complete set of raw data files, and (3) clear descriptions on sample groups. The application of these criteria on PRIDE datasets further yielded three quantification tools with multiple (≥2) datasets, which included MaxQuant, OpenMS, and Progenesis. Third, eight benchmarks of these three tools (three benchmarks for both MaxQuant and Progenesis; two benchmarks for OpenMS) were collected for the subsequent analyses (Table II, study 1 and studies 3–9).

Based on the eight sets of data, the dependence of well-performing manipulation chains on the selection of quantification tool was assessed using the precision of >3,000 chains. As shown in Fig. 2, the clustering analyses among chains across multiple datasets were done for MaxQuant (Fig. 2*A*), Progenesis (Fig. 2*B*), and OpenMS (Fig. 2*C*). Second, three neighboring partitions of varied performance were identified, which included partition α containing the chains of consistently superior precision (PMADs <0.14) across multi-

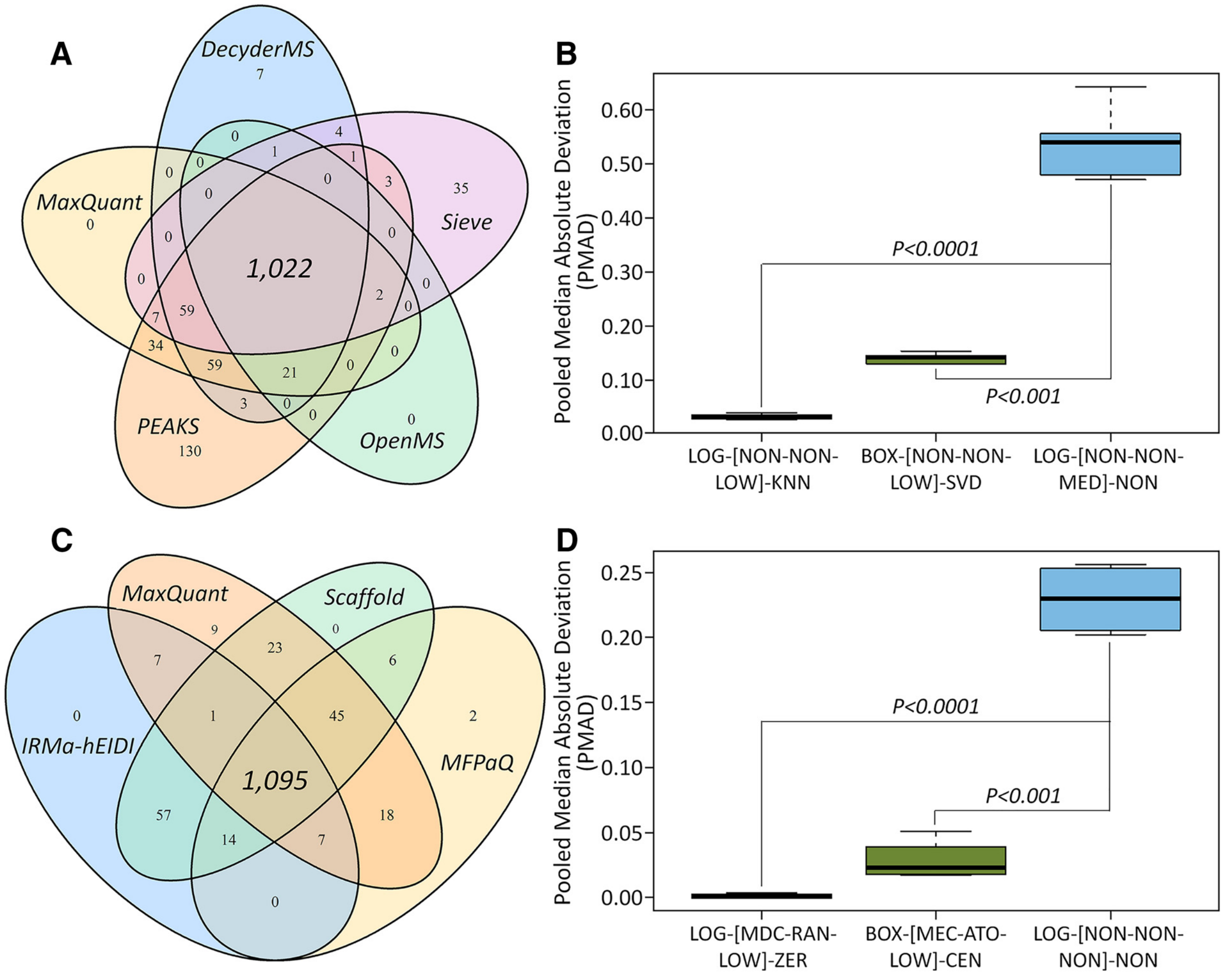


FIG. 1. **Comprehensive assessment of precision of 3,128 potential manipulation chains and comparison to the chains successfully adopted by previous studies (23, 58).** On one hand, a log transformation together with the median normalization were applied before any analysis in the Table II, study 1 (23), which defined manipulation chain of that study as LOG-[NON-NON-MED]-NON. (*A*) The Venn diagram illustrating the number of chains (combined with five quantification tools) performing better in precision than LOG-[NON-NON-MED]-NON. (*B*) Significant difference in precision between LOG-[NON-NON-MED]-NON and two example chains. On the other hand, only log transformation was used to manipulate the spectral counting data from Table II, study 2 (58), the manipulation chain of which could therefore be defined as LOG-[NON-NON-NON]-NON. (*C*) The Venn diagram showing the number of chains (combined with four quantification tools) performing better than the chain adopted by the original study (LOG-[NON-NON-NON]-NON). (*D*) The significant differences were observed between LOG-[NON-NON-NON]-NON and two example chains (LOG-[MDC-RAN-LOW]-ZER and BOX-[MEC-ATO-*LOW]-CEN*). Each manipulation method within a chain was abbreviated by a three-letter code that was defined in Table I.

ple datasets, partition β including the chain of good precision (PMADs <0.3) across multiple datasets, and partition γ consisting of the chains of fair precision (PMADs <0.7) for multiple datasets. Finally, the Venn diagrams showing the manipulation chains shared by different quantification tools or demonstrating tool-specific characteristics were provided for the chains within partition α (Fig. 2*D*), partition α and β (Fig. 2*E*), and partition α and β and γ (Fig. 2*F*). As demonstrated, the majority of these well-performing chains were shared by all quantification tools, while there were still dozens of chains performing well only for a single tool. Moreover, the number of chains performing well for both OpenMS and Progenesis was much larger than that for both MaxQuant and Progenesis or for both MaxQuant and Progenesis.

Inconsistency among the Quantification Performance Assessed from Multiple Perspectives—Besides precision (28), several other criteria could be used to assess the performance of LFQ (24), which include: the quantification accuracy (6) and robustness (29), classification capacity (49), and differential abundance analysis (50). Given the huge amount of potential manipulation chains and complicated nature of the studied datasets, the level of consistency among the performance assessed by different criteria was essential for proteomic quantification. Thus, the benchmark dataset (Table II, study 10 (29)) acquired using SWATH-MS was collected, and the quantification performance of three example chains (as illustrated in Fig. 3) were assessed by multiple criteria. Particularly, due to the lack of spiked proteins in the collected dataset, the perform-

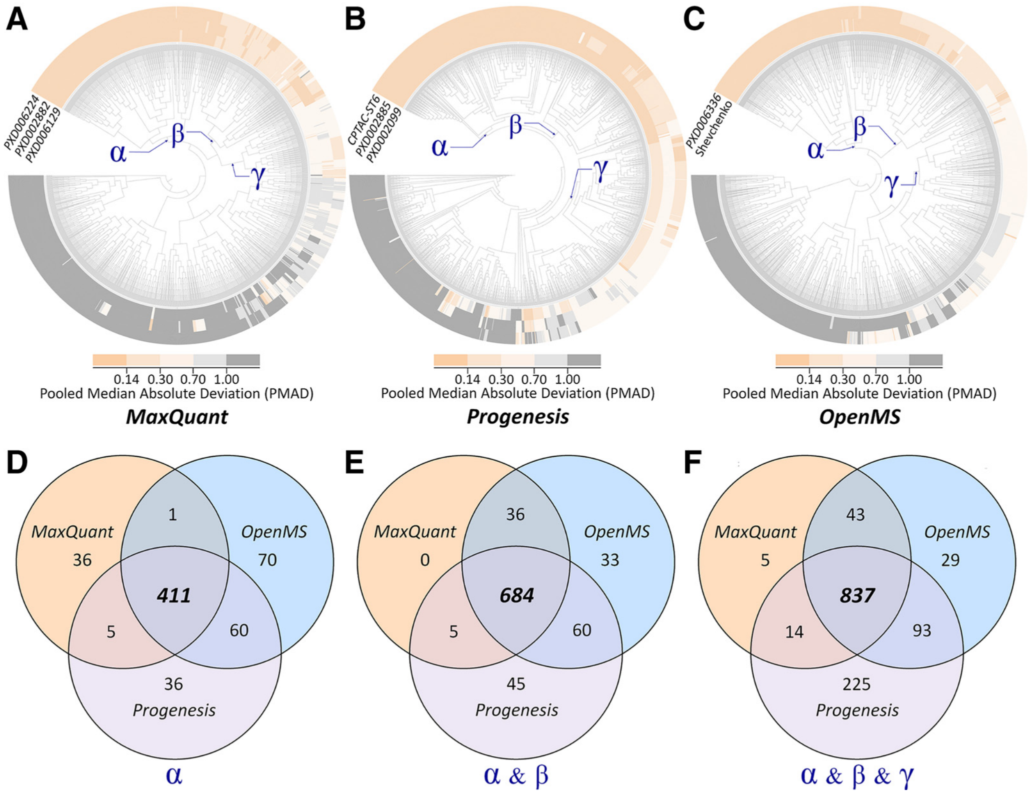


FIG. 2. **The dependence of well-performing manipulation chains on the selection of quantification tool comprehensively assessed by the precision of 3,128 potential chains.** First, the clustering analyses among manipulation chains across multiple datasets were conducted for three quantification tools: (*A*) MaxQuant, (*B*) Progenesis, and (*C*) OpenMS. Then, three neighboring partitions of varied performance were identified, which included partition αcontaining the chains of consistently superior precision (PMADs <0.14) across multiple datasets, partition βincluding the chain of good precision (PMADs <0.3) across multiple datasets, partition γ consisting of the chains of fair precision (PMADs <0.7) for multiple datasets. Finally, the Venn diagrams showing the chains shared by multiple tools or demonstrating tool-specific characteristics were provided for the chains within (*D*) partition α, (*E*) partition α and β, and (*F*) partition α and β and γ.

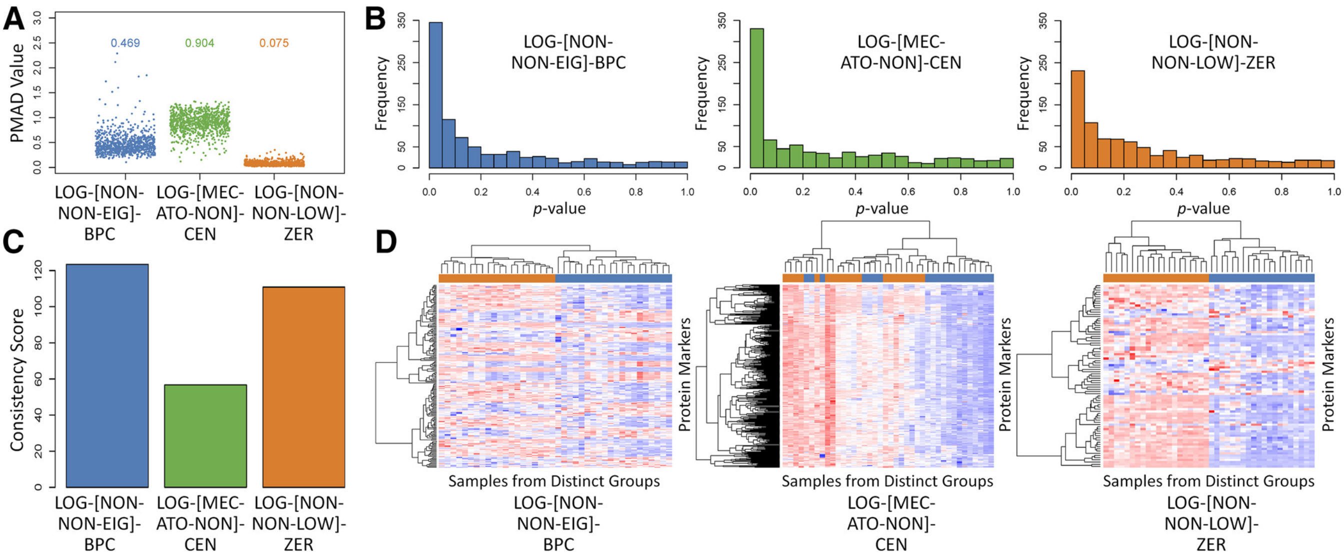


FIG. 3. **The performance of three representative manipulation chains (LOG-[NON-NON-EIG]-BPC, LOG-[MEC-ATO-NON]-CEN and LOG-[NON-NON-LOW]-ZER) assessed by multiple criteria:** (*A*) precision, (*B*) differential abundance analysis, (*C*) robustness, and (*D*) classification capacity. Each manipulation method within a chain was abbreviated by a three-letter code that was defined in Table I.

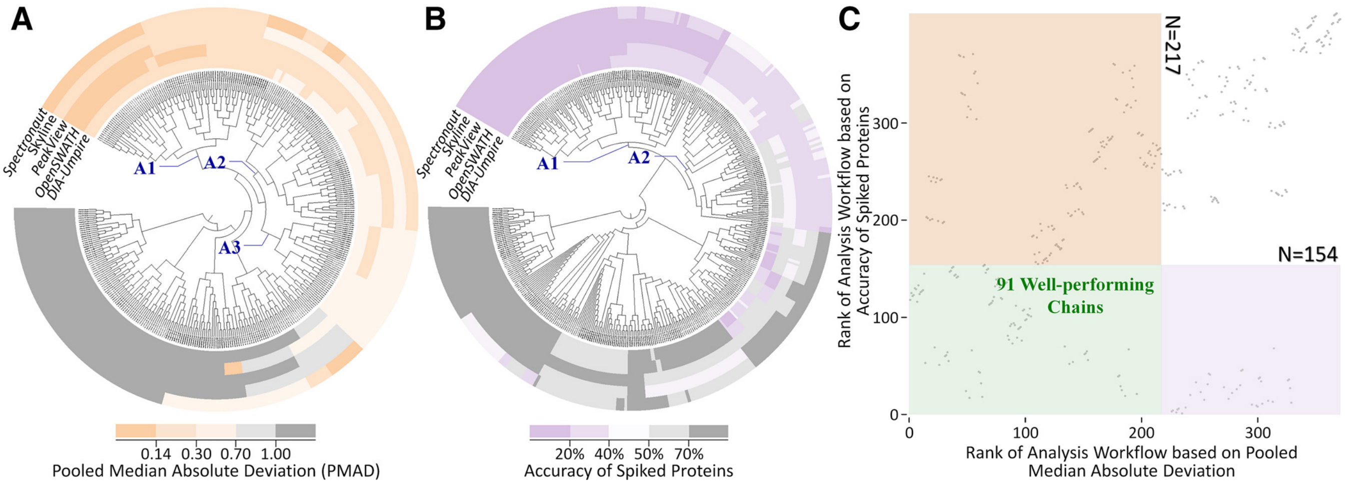


FIG. 4. **The strategy proposed in this study to discover the manipulation chains of simultaneously improved precision and accuracy based on the benchmarks from Table II, study 11.** First, clustering analyses among chains across five quantification tools were conducted for (*A*) precision and (*B*) accuracy. Second, a two-dimensional scatter plot (*C*) was drawn to show the ranks of each manipulation chain (represented by a gray dot) collectively determined by precision (horizontal axis) and accuracy (vertical axis). The pink, violet, and green areas in *C* denoted the chains of good precision (A1+A2+A3 in *A*), good accuracy (A1+A2 in *B*), and good performance for both, respectively. As a result, 91 chains (within the green region of *C*) were found to perform well under both criteria (precision and accuracy).

ance of the example chains were assessed only by four criteria: precision (Fig. 3*A*), differential abundance analysis (Fig. 3*B*), robustness (Fig. 3*C*), and classification capacity (Fig. 3*D*). As illustrated, the performance of the different chains evaluated by the same criterion varied greatly, and there was substantial inconsistency among the performance of a given chain assessed using different criteria. For example, LOG-[MEC-ATO-NON]-CEN performed the worst in precision (Fig. 3*A*) but was top-ranked for differential abundance analysis (Fig. 3*B*).

The inconsistency among the quantification performance assessed by multiple criteria can also be found in Table III. For all quantification tools, the precision of LOG-[MEC-RAN-PQN]-BPC was “superior,” but its robustness was always “low. Similarly, as shown in supplemental Table S2, the precision of BOX-[NON-VAS-EIG]-BAK was consistently “poor”, but its accuracy was always “high”. However, the chains performing well under both criteria could also be found (*e.g.* the LOG-[MDC-RAN-EIG]-KNN performed well under both precision and robustness in Table III). Moreover, as shown in supplemental Fig. S1, some chains (*e.g.* BOX-[NON-PAR-TIC]-KNN) performed consistently well across all criteria, while others (*e.g.* BOX-[MEC-VAS-PQN]-BAK) always lacked quantification capacities under multiple criteria. Because these criteria complement one another, a collective consideration of them may lead to simultaneous improvement in precision, accuracy, or robustness. In other words, a collective optimization of LFQ based on the assessment of >3,000 chains may facilitate the discovery of well-performing LFQs.

Simultaneously Enhancing Precision and Accuracy of Data Manipulation by Chain Optimization—

*(1) For the Proteomic Data Acquired by Data-Independent Acquisition (DIA)—*As one of the most advanced DIA techniques, the SWATH-MS-based proteomics was recently applied to provide more comprehensive detection and accurate quantitation of proteins compared with the traditional technique (68). In order to identify the LFQs of the simultaneously improved quantification precision and accuracy, five DIA-based benchmarks of Table II, study 11 (6), preprocessed by five quantification tools (DIA-Umpire, OpenSWATH, PeakView, Skyline, and Spectronaut), were collected to reveal the difference in precision and accuracy among various chains. By analyzing the nature of the five studied datasets and their obedience toward the assumption of the manipulation methods shown in Table I, all normalization methods (excluding EIG) and VSN transformation were inappropriate to manipulate studied data because (1) the level of protein abundance could not be assumed as constant among all studied samples and (2) the intensities of the vast majority of proteins could not be assumed as unchanged. In other words, because both assumptions (A-β and A-γ) could not hold for the studied datasets, the methods based on these assumptions were fundamentally inappropriate for the datasets and were excluded from the >3,000 potential chains. As a result, clustering analysis was used to identify the relationships among the performance of the remaining 480 potential manipulation chains. As illustrated in Fig. 4*A*, quantification precision of the majority of these 480 chains was consistent across five quantification tools, and the chains within partition A (A1, A2, and A3) performed consistently well across five quantification tools. Similar to precision, the accuracy of most chains were also consistent across these tools (Fig. 4*B*), and the chains within partition A (A1 and A2) performed consistently well across five tools. Based on the above assessments by two criteria, Fig. 4*C* provided the ranks of each manipulation chain (indicated by gray dot) collectively defined by precision (hor-

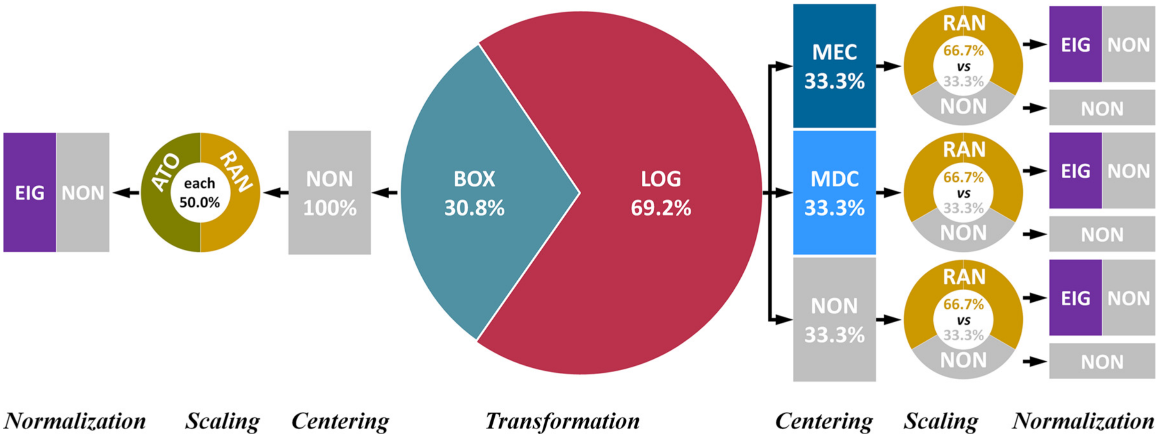


FIG. 5. **The distribution of manipulation methods in 91 well-performing chains identified based on five DIA-based datasets from Table II, study 11.** The manipulation method was abbreviated by three-letter code that was defined in Table I. For the seven imputation methods together with the nonimputation (NON), they demonstrated the exactly equal chances within the 91 well-performing chains, which showed that the selection of different imputation methods (even NON) had nothing to do with the performance under this circumstance. Therefore, the imputation methods were not displayed in this distribution. All normalization methods together with VSN transformation were found inappropriate for manipulating the five DIA-based benchmarks, and the assumptions of these methods did not hold for the studied dataset.

izontal axis) and accuracy (vertical axis). The pink, violet, and green regions indicate the chains of good precision (partition A in Fig. 4*A*), good accuracy (partition A in Fig. 4*B*), and good performance for both precision and accuracy, respectively. This finding revealed the feasibility of identifying the chains of simultaneously improved precision and accuracy, which was known as the key issue for SWATH-MS-based quantitative proteomics (69). Moreover, there were 91 manipulation chains (within the green region of Fig. 4*C*) identified as well-performing under both criteria (precision and accuracy). The analysis on the distribution of the manipulation methods in these 91 chains could facilitate the discovery of the chains suitable for SWATH-MS-based proteomics.

Therefore, the distributions of manipulation methods in the identified chains were systematically analyzed. As shown in Fig. 5, 63 out of 91 (69.2%) chains were based on LOG transformation, and the remaining adopted BOX transformation. On one hand, two centering methods (MEC and MDC) and the noncentering (NON) showed equal chance to combine with LOG, which denoted that the selection of different centering methods (even noncentering) could hardly influence LFQ's performance when combining with LOG. On the other hand, only NON was in combination with BOX transformation, which indicated that BOX did not prefer to combine with any centering approach under this circumstance. Among four scaling methods together with the nonscaling (NON), only two (RAN and NON) were suitable for combining with LOG. If RAN was applied, the EIG normalization together with the nonnormalization (NON) became the ones of good performance. Moreover, the combinations of BOX-ATO and BOX-RAN were found equally suitable for quantifying this SWATH-MS dataset, and the EIG normalization and NON were still found to perform well. When the vast majority of the proteins were differentially expressed between distinct sample groups, EIG was reported to be not only effective in reducing the intragroup variations (good precision) but also suitable for normalizing the data of spiked proteins (good accuracy) (27), which was consistent with the finding of this study. For seven imputation methods together with the nonimputation (NON), they showed equal chance within those 91 chains, which indicated that the selection of different imputation (even NON) had nothing to do with the quantification performance in this situation.

(2) For the Proteomic Data Acquired by DDA—For DDA-based proteomics, it was also necessary to consider both quantification precision and accuracy (27, 28). Therefore, four DDA-based benchmark datasets from Table II, study 2 of distinct concentrations (12.5 versus 25 fmol/μg) of spiked UPS1 protein (58), preprocessed by four quantification tools (IRMa-hEIDI, MaxQuant, MFPaQ, and Scaffold), were collected to uncover the difference in both precision and accuracy among manipulation chains. Based on the preanalyses on the nature of the four studied datasets and their obedience toward the assumption of manipulation methods, all methods were appropriate to manipulate these benchmarks in the first place. Then, the clustering analysis was used to reveal the relationship among the performance of 3,128 chains. Similar to Fig. 4, these chains of consistently good performance in precision and accuracy were partitioned into a clustered area (A1, A2, and A3) in supplemental Fig. S2*A* and another clustered area (A1 and A2) in supplemental Fig. S2*B*, respectively. The regions in supplemental Fig. S2*C* colored in pink, violet, and green indicate the chains of good precision (areas of A1, A2, and A3 in supplemental Fig. S2*A*), good accuracy (areas of A1 and A2 in supplemental Fig. S2*B*), and good performance in precision and accuracy, respectively. The red dot in supplemental Fig. S2*C* denotes the chain used in the original study (58), which was among those identified 728 chains of

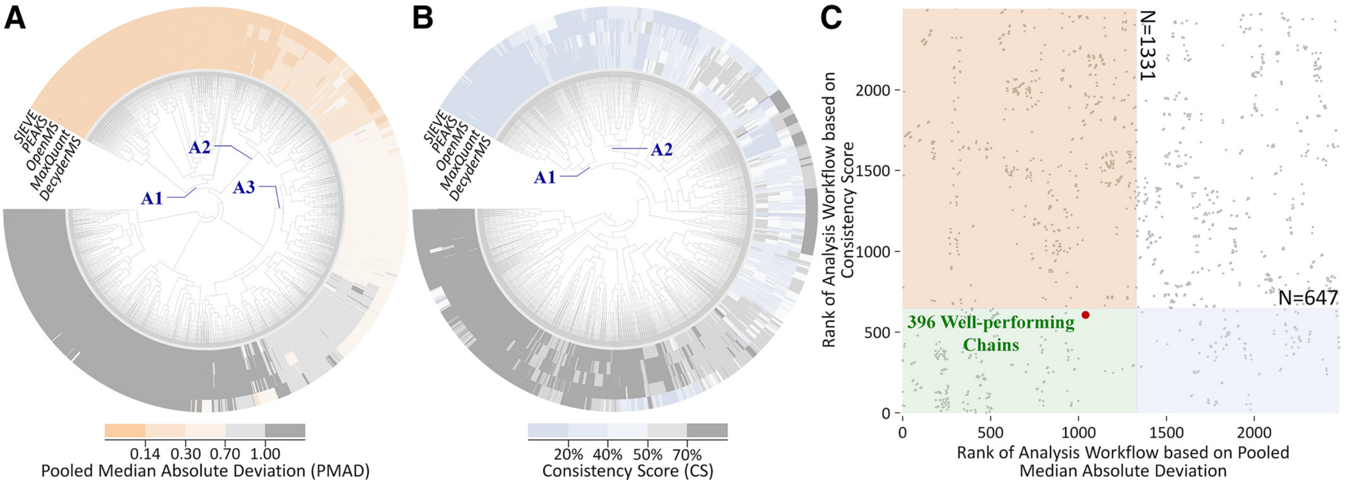


FIG. 6. **The strategy proposed in this study to discover the manipulation chains of simultaneously improved precision and robustness based on benchmarks collected from Table II, study 1.** First, clustering analysis among manipulation chains across five quantification tools was conducted for (*A*) precision and (*B*) robustness. Second, a two-dimensional scatter plot (*C*) was drawn to provide the ranks of each chain (represented by gray dot) collectively determined by precision (horizontal axis) and robustness (vertical axis). The pink, blue, and green areas in *C* indicated the chains of good precision (A1+A2+A3 in *A*), good robustness (A1+A2 in *B*), and good performance for both precision and robustness, respectively. As a result, 396 chains (within the green area of *C*) were found to perform well under both criteria (precision and robustness).

consistently good performance. However, hundreds of chains were identified as performing better than the original one in both precision and accuracy (shown in the lower left corner of supplemental Fig. S2*C*). Besides the four benchmarks analyzed previously, there were nine additional benchmarks of the distinct concentrations of the spiked UPS1 proteins (58). Herein, these nine benchmarks were analyzed using the same strategy discussed previously, and nine sets of manipulation chains performing well in both precision and accuracy were identified (supplemental Figs. S3–S11). Together with the first benchmark dataset of 12.5 versus 25 fmol/μg, the intersection of 10 sets of well-performing chains led to 133 chains well-performing under both criteria (precision and accuracy).

Moreover, a systematic analysis on the distributions of manipulation methods in the identified 133 chains was also conducted. As shown in supplemental Fig. S12, 84 out of the 133 (63.2%) identified chains were based on LOG transformation, and the remaining used BOX transformation. When combining with LOG, TMM was the only normalization method performing well under both precision and accuracy. Two centering methods (MEC and MDC) and the noncentering (NON) showed equal chances to combine with LOG-scaling-normalization, which indicated that the selection of different centering methods (even noncentering) could hardly influence the quantification performance in this situation, but the centering would be accompanied by different scaling methods (PAR, RAN, and NON were suitable for all centering). When combining with BOX, TMM, QUA, and TIC normalization stood out from all normalization methods. Only noncentering was integrated with BOX transformation, which demonstrated that BOX did not prefer to combine with any centering method under this circumstance. To the best of our knowledge, the TIC, TMM, and QUA have been frequently applied for normalizing the spectral-counting-based proteomic datasets (42, 70–72). Moreover, a comprehensive literature search in PubMed by combining the name of the remaining normalization methods with "spectral counting" and "proteomics" resulted in few publications, which may indicate the incapability of these methods in simultaneously improving precision and accuracy. For seven imputation methods together with non-imputation (NON), they showed exactly equal chances within those 133 well-performing chains, which denoted that the selection of different imputation methods (even NON) still had nothing to do with the performance in this situation.

Collectively Improving Precision and Robustness of Data Manipulation by Chain Optimization—Precision (28) and robustness (29) (with distinct underlying theory) are two well-established criteria for assessing manipulation chain's performance. Particularly, both criteria should be collectively considered not only to reduce proteome variations among replicates but also to enhance consistencies among different lists of identified biomarkers (24, 27). In this study, five proteomic benchmark datasets from Table II, study 1 (23), preprocessed by various quantification tools (DecyderMS, MaxQuant, OpenMS, PEAKS, and Sieve), were first collected to reveal the difference in both precision and robustness among the chains. Based on the preanalysis on the nature of these five studied datasets and their obedience toward the assumption of manipulation methods, all methods were appropriate for manipulating the datasets in the first place. Then, clustering analysis was conducted to discover the relation among the performance of different chains. As shown in Fig. 6*A*, the precision of the majority of >3,000 chains were consistent among quantification tools, and the chains within partition A

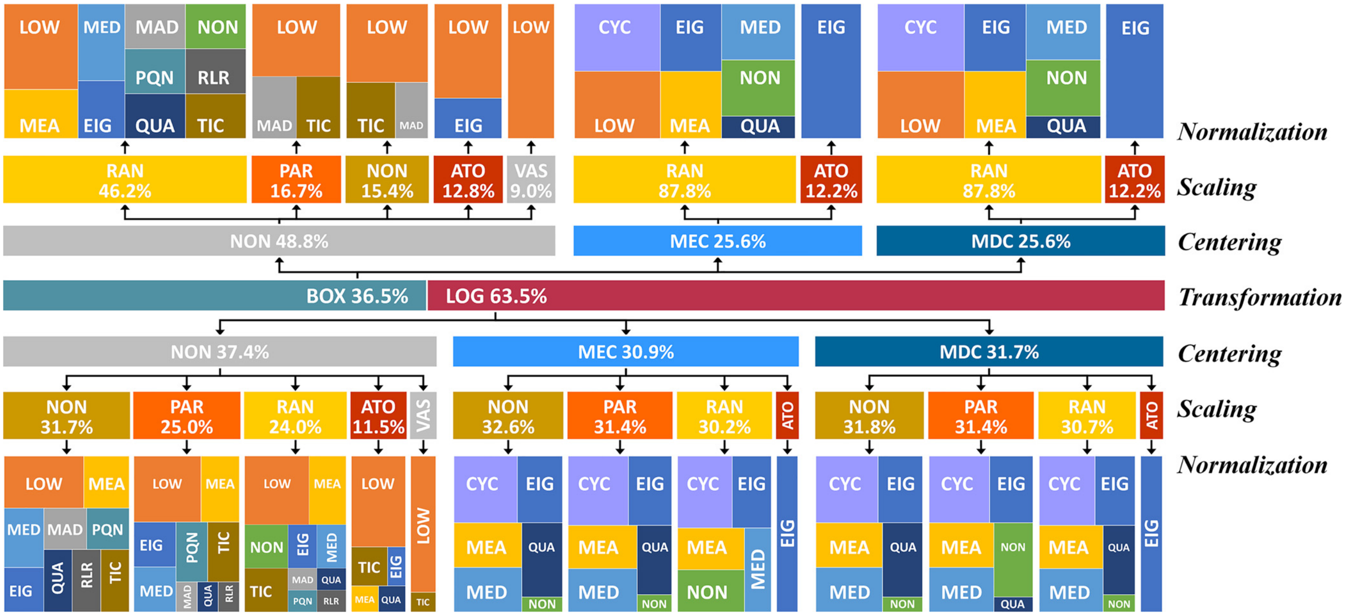


FIG. 7. **The distribution of manipulation methods in the 396 well-performing chains identified based on the datasets from Table II, study 1.** The manipulation methods were abbreviated by three-letter codes (defined in Table I). For the seven imputation methods together with the nonimputation (NON), they showed similar chances within those 396 chains, which indicated that the selection of different imputation methods (even the NON) may have slight influence on the performance under this circumstance. Therefore, the imputation methods were not displayed in this distribution.

(A1, A2, and A3) performed consistently well in five tools. Similar to precision, the robustness of most manipulation chains was consistent for five tools (Fig. 6*B*), and the chains in partition A (A1 and A2) performed consistently well across five different tools. Based on the preceding evaluation from two perspectives, Fig. 6*C* illustrates the ranks of each chain (represented by a gray dot) collectively determined by precision (horizontal axis) and robustness (vertical axis). Pink, blue, and green regions in Fig. 6*C* indicate those chains of good precision (partition A in Fig. 6*A*), good robustness (partition A in Fig. 6*B*), and good performance for both precision and robustness, respectively. Moreover, the red dot in Fig. 6*C* denotes the chain adopted in the original study (23), which was among the identified 396 chains of consistently good performance. However, hundreds of chains were also discovered to perform better than the original one in both precision and accuracy (shown in the lower left corner of Fig. 6*C*). This finding reveals the feasibility of identifying the manipulation chains of collective enhancements in precision and robustness, which is essential for current proteomics (24, 27).

Furthermore, the analysis on the distribution of manipulation methods in those identified 396 chains was conducted. As illustrated in Fig. 7, 251 out of these 396 (63.5%) identified chains were based on LOG transformation, and the remaining used BOX transformation. If the centering methods (MEC/MDC) were applied, both RAN and ATO scaling showed great chance of good performance, and PAR scaling together with the nonscaling (NON) demonstrated high chance of good performance when combining with LOG. For normalization, several methods showed great chance of good performance, including CYC, EIG, MED, MEA, QUA, and sometimes LOW. If noncentering was applied, all scaling methods had the chance of good performance, and the normalization methods showing a good performance included LOW, MEA, EIG, TIC, and MAD. Similar to the previous discussion, seven imputation methods together with the nonimputation (NON) showed a similar chance within 396 well-performing chains, which indicated that the selection of different imputation methods (even NON) might have little impact on the performance in this situation. As a transformation method integrating the normalization technique, VSN performed well under this circumstance regardless of the selection of imputation methods (even NON).

Development of an Online Tool for Proteomic Quantification and Performance Assessment—Because the underlying theories of multiple criteria were distinct from each other, a collective consideration of the quantification precision, accuracy, and robustness as well as qualitative criteria could achieve the most comprehensive assessment and were reported to be essential for modern proteomic research (24, 27). However, because of the dataset-dependent nature of performance assessment and the lack of suitable datasets for assessing the feasibility of the proposed strategy, a powerful tool for proteomics quantification and performance assessment was urgently needed. So far, several LFQ-relevant tools have been available and popular in proteome quantification (supplemental Table S5). Gmine (73) and Perseus (74) integrated various manipulation methods in their quantification process, but no performance assessment function was provided. LFQbench (6) and msCompare (75) were recognized for evaluating the performance of three to five tools. The Normalyzer (28), SPANS (76), and GiaPronto (77) were distinguished as able to assess one to eight pretreatment methods. Because a typical LFQ combined both quantification tool and manipulation method, any assessment focusing solely on the tool or the method could not fully reflect the overall performance of LFQ. Moreover, none of these tools could systematically assess the performance (as highly recommended by the previous studies (24, 27)) based on multiple criteria of distinct underlying theories. Because the performance of given LFQs was susceptible to the studied dataset (78), it was essential to find the most appropriate tool together with a manipulation chain for particular datasets. However, it was challenging to perform such discovery as there were large numbers of potential workflows and the multifaceted nature of the evaluation criteria.

Therefore, a web-based tool (official site: https://idrblab.org/anpela/; mirror sites: http://idrblab.cn/anpela/, http://idrb.zju.edu.cn/anpela/) was developed as the first tool enabling the performance assessment of the entire LFQ manipulation chain (collectively assessed by five well-established criteria) in this study. This tool not only automatically detects the diverse formats of data generated by all quantification tools but also provides the most complete set of manipulation methods among available web-based and standalone tools (supplemental Table S5). Due to its unique abilities in discovering well-performing chains, performing assessment from multiple perspectives, and validating quantification accuracy based on spiked proteins, it has great application potentials for any proteomic studies requiring LFQ technique. This online tool can be readily accessed by users with no login requirement and is freely accessible using a variety of popular web browsers such as Google Chrome, Mozilla Firefox, Safari, and Internet Explorer (10 or later), and the procedures for using this online tool are fully illustrated and documented in its online tutorial.

Due to the tremendous computational workload required for assessing >3,000 chains, it is extremely time- and resource-consuming to make this comprehensive assessment service online. Therefore, an alternative way of enabling the evaluation on a user's local computer was provided as standalone software, which can be downloaded from both the official and mirror sites and provides the same sets of assessment metrics and plots as that of the online version. The exemplar input and output files can be downloaded together with the source code of this standalone tool. The installation and configuration of R environment together with a number of packages required before running the tool are provided in the User Manual (downloadable from its website), which can help the users to get familiar with this tool as soon as possible.

CONCLUSION

Based on the in-depth analyses in this study, some common trends in the identified well-performing chains were discovered, which might give recommendation for researchers in relevant fields. As shown in Figs. 2*A*-2*C*, although there were manipulation chains dependent heavily on the studied datasets (great variations in the colors of PMAD), >50% of the analyzed chains showed a consistent precision (the same colors of PMAD) across multiple datasets. These results indicate that there were manipulation chains performing consistently well/badly regardless of the analyzed datasets. Similarly, as provided in Figs. 4*A* and 4*B*, although there were chains dependent on the applied software tools (variations in the branch colors), many studied chains gave consistent performance (the same branch colors) across multiple software tools. These results denoted that, similar to various datasets, there were manipulation chains performing consistently well/badly regardless of the applied software tools. Moreover, the well-performing manipulation chains identified for DIA and DDA datasets were shown in Fig. 5 and supplemental Fig. S12, respectively. As shown, although there were some common features between the identified manipulation methods, there were great variations in these figures. These results denoted that the well-performing methods also depended on the acquisition methods. Based on the preceding analyses, variations could be frequently induced by different datasets, various acquisition methods, and diverse software tools, which made it extremely essential to develop tool for both proteomic quantification and comprehensive performance assessment. Thus, a tool was developed to enable the performance assessments of the entire data manipulation chain (collectively assessed by five well-established criteria) in this study. This online tool not only automatically detected the diverse formats of data generated by quantification tools but also provided the most complete set of manipulation methods among all available web-based and standalone tools.

DATA AVAILABILITY

The mass spectrometry proteomics data were previously published (see Table II) and have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifiers PXD001819, PXD002882, PXD006129, PXD006224, PXD002885, PXD002099, PXD006336, PXD003972 and PXD002952.

* Funded by the National Key Research and Development Program of China (2018YFC0910500), National Natural Science Foundation of China (81872798), Fundamental Research Funds for Central Universities (2018QNA7023, 10611CDJXZ238826, 2018CDQYSG0007, and CDJZR14468801), and Innovation Project on Industrial Generic Key Technologies of Chongqing (cstc2015zdcy-ztzx120003). The authors declare no competing interests.

S This article contains supplemental material Tables S1-S5 and Figs. S1-S12.

‡‡ To whom correspondence should be addressed: College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058,

China. E-mail: zhufeng@zju.edu.cn or prof.zhufeng@gmail.com.

§§ These authors contributed equally to this work as co-first authors.

Author contributions: J.T., J.F., Y.W., and Y.L. performed research; J.T., Q.Y., and B.L. contributed new reagents/analytic tools; J.T., J.F., Y.W., G.T., J.H., X.C., Y.C., L.Y., and W.X. analyzed data; F.Z. designed research; and F.Z. wrote the paper.

REFERENCES

1. Lobingier, B. T., Hüttenhain, R., Eichel, K., Miller, K. B., Ting, A. Y., von Zastrow, M., and Krogan, N. J. (2017) An approach to spatiotemporally resolve protein interaction networks in living cells. *Cell* **169,** 350–360.e12
2. Thul, P. J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Blal, H. A., Alm, T., Asplund, A., Björk, L., Breckels, L. M., Bäckström, A., Danielsson, F., Fagerberg, L., Fall, J., Gatto, L., Gnann, C., Hober, S., Hjelmare, M., Johansson, F., Lee, S., Lindskog, C., Mulder, J., Mulvey, C. M., Nilsson, P., Oksvold, P., Rockberg, J., Schutten, R., Schwenk, J. M., Sivertsson, A., Sjostedt, E., Skogs, M., Stadler, C., Sullivan, D. P., Tegel, H., Winsnes, C., Zhang, C., Zwahlen, M., Mardinoglu, A., Pontén, F., von Feilitzen, K., Lilley, K. S., Uhlén, M., and Lundberg, E. (2017) A subcellular map of the human proteome. *Science* **356,** eaal3321
3. van Rooden, E. J., Florea, B. I., Deng, H., Baggelaar, M. P., van Esbroeck, A. C. M., Zhou, J., Overkleeft, H. S., and van der Stelt, M. (2018) Mapping in vivo target interaction profiles of covalent inhibitors using chemical proteomics with label-free quantification. *Nat. Protoc.* **13,** 752–767
4. Li, K. S., Shi, L. Q., and Gross, M. L. (2018) Mass spectrometry-based fast photochemical oxidation of proteins (FPOP) for higher order structure characterization. *ACC. Chem. Res.* **51,** 736–744
5. Distler, U., Kuharev, J., Navarro, P., and Tenzer, S. (2016) Label-free quantification in ion mobility-enhanced data-independent acquisition proteomics. *Nat. Protoc.* **11,** 795–812
6. Navarro, P., Kuharev, J., Gillet, L. C., Bernhardt, O. M., MacLean, B., Röst, H. L., Tate, S. A., Tsou, C. C., Reiter, L., Distler, U., Rosenberger, G., Perez-Riverol, Y., Nesvizhskii, A. I., Aebersold, R., and Tenzer, S. (2016) A multicenter study benchmarks software tools for label-free proteome quantification. *Nat. Biotechnol.* **34,** 1130–1136
7. Cretu, D., Prassas, I., Saraon, P., Batruch, I., Gandhi, R., Diamandis, E. P., and Chandran, V. (2014) Identification of psoriatic arthritis mediators in synovial fluid by quantitative mass spectrometry. *Clin. Proteomics* **11,** 27
8. Li, Z., Adams, R. M., Chourey, K., Hurst, G. B., Hettich, R. L., and Pan, C. L. (2012) Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos. *J. Proteome Res.* **11,** 1582–1590
9. Rieckmann, J. C., Geiger, R., Hornburg, D., Wolf, T., Kveler, K., Jarrossay, D., Sallusto, F., Shen-Orr, S. S., Lanzavecchia, A., Mann, M., and Meissner, F. (2017) Social network architecture of human immune cells unveiled by quantitative proteomics. *Nat. Immunol.* **18,** 583–593
10. Min, C. W., Lee, S. H., Cheon, Y. E., Han, W. Y., Ko, J. M., Kang, H. W., Kim, Y. C., Agrawal, G. K., Rakwal, R., Gupta, R., and Kim, S. T. (2017) In-depth proteomic analysis of Glycine max seeds during controlled deterioration treatment reveals a shift in seed metabolism. *J. Proteomics* **169,** 125–135
11. Frantzi, M., Latosinska, A., Flühe, L., Hupe, M. C., Critselis, E., Kramer, M. W., Merseburger, A. S., Mischak, H., and Vlahou, A. (2015) Developing proteomic biomarkers for bladder cancer: Towards clinical application. *Nat. Rev. Urol.* **12,** 317–330
12. Komatsu, S., Han, C., Nanjo, Y., Altaf-Un-Nahar, M., Wang, K., He, D., and Yang, P. (2013) Label-free quantitative proteomic analysis of abscisic acid effect in early-stage soybean under flooding. *J. Proteome Res.* **12,** 4769–4784
13. Hogrebe, A., von Stechow, L., Bekker-Jensen, D. B., Weinert, B. T., Kelstrup, C. D., and Olsen, J. V. (2018) Benchmarking common quantification strategies for large-scale phosphoproteomics. *Nat. Commun.* **9,** 1045
14. Zhang, B., Käll, L., and Zubarev, R. A. (2016) DeMix-Q: Quantification-centered data processing workflow. *Mol. Cell. Proteomics* **15,** 1467–1478
15. Müller, F., Fischer, L., Chen, Z. A., Auchynnikava, T., and Rappsilber, J. (2018) On the reproducibility of label-free quantitative cross-linking/mass spectrometry. *J. Am. Soc. Mass. Spectrom.* **29,** 405–412
16. Wang, X., Gardiner, E. J., and Cairns, M. J. (2015) Optimal consistency in microRNA expression analysis using reference-gene-based normalization. *Mol. Biosyst.* **11,** 1235–1240
17. Shen, X., Shen, S., Li, J., Hu, Q., Nie, L., Tu, C., Wang, X., Poulsen, D. J., Orsburn, B. C., Wang, J., and Qu, J. (2018) IonStar enables high-precision, low-missing-data proteomics quantification in large biological cohorts. *Proc. Natl. Acad. Sci. U.S.A.* **115,** E4767–E4776
18. Tyanova, S., Temu, T., and Cox, J. (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11,** 2301–2319
19. Tsou, C. C., Avtonomov, D., Larsen, B., Tucholska, M., Choi, H., Gingras, A. C., and Nesvizhskii, A. I. (2015) DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat. Methods* **12,** 258–264
20. Barschke, P., Oeckl, P., Steinacker, P., Ludolph, A., and Otto, M. (2017) Proteomic studies in the discovery of cerebrospinal fluid biomarkers for amyotrophic lateral sclerosis. *Expert. Rev. Proteomics* **14,** 769–777
21. Huang, Q., Yang, L., Luo, J., Guo, L., Wang, Z., Yang, X., Jin, W., Fang, Y., Ye, J., Shan, B., and Zhang, Y. (2015) SWATH enables precise label-free quantification on proteome scale. *Proteomics* **15,** 1215–1223
22. Gatto, L., Hansen, K. D., Hoopmann, M. R., Hermjakob, H., Kohlbacher, O., and Beyer, A. (2016) Testing and validation of computational methods for mass spectrometry. *J. Proteome Res.* **15,** 809–814
23. Khoonsari, P. E., Häggmark, A., Lönnberg, M., Mikus, M., Kilander, L., Lannfelt, L., Bergquist, J., Ingelsson, M., Nilsson, P., Kultima, K., and Shevchenko, G. (2016) Analysis of the cerebrospinal fluid proteome in Alzheimer's disease. *PloS One* **11,** e0150672
24. Välikangas, T., Suomi, T., and Elo, L. L. (2018) A comprehensive evaluation of popular proteomics software workflows for label-free proteome quantification and imputation. *Brief Bioinform.* **19,** 1344–1355
25. Al Shweiki, M. R., Mönchgesang, S., Majovsky, P., Thieme, D., Trutschel, D., and Hoehenwarter, W. (2017) Assessment of label-free quantification in discovery proteomics and impact of technological factors and natural variability of protein abundance. *J. Proteome Res.* **16,** 1410–1424
26. Ramus, C., Hovasse, A., Marcellin, M., Hesse, A. M., Mouton-Barbosa, E., Bouyssié, D., Vaca, S., Carapito, C., Chaoui, K., Bruley, C., Garin, J., Cianferani, S., Ferro, M., Van Dorssaeler, A., Burlet-Schiltz, O., Schaeffer, C., Couté, Y., and Gonzalez de Peredo, A. (2016) Benchmarking quantitative label-free LC-MS data processing workflows using a complex spiked proteomic standard dataset. *J. Proteomics* **132,** 51–62
27. Välikangas, T., Suomi, T., and Elo, L. L. (2018) A systematic evaluation of normalization methods in quantitative label-free proteomics. *Brief Bioinform.* **19,** 1–11
28. Chawade, A., Alexandersson, E., and Levander, F. (2014) Normalyzer: A tool for rapid evaluation of normalization methods for omics data sets. *J. Proteome Res.* **13,** 3114–3120
29. Caron, E., Roncagalli, R., Hase, T., Wolski, W. E., Choi, M., Menoita, M. G., Durand, S., García-Blesa, A., Fierro-Monti, I., Sajic, T., Heusel, M., Weiss, T., Malissen, M., Schlapbach, R., Collins, B. C., Ghosh, S., Kitano, H., Aebersold, R., Malissen, B., and Gstaiger, M. (2017) Precise temporal profiling of signaling complexes in primary cells using SWATH mass spectrometry. *Cell Rep.* **18,** 3219–3226
30. Li, B., Tang, J., Yang, Q., Li, S., Cui, X., Li, Y., Chen, Y., Xue, W., Li, X., and Zhu, F. (2017) NOREVA: Normalization and evaluation of MS-based metabolomics data. *Nucleic Acids Res.* **45,** W162–W170
31. Gao, B. B., Stuart, L., and Feener, E. P. (2008) Label-free quantitative analysis of one-dimensional PAGE LC/MS/MS proteome: Application on angiotensin II-stimulated smooth muscle cells secretome. *Mol. Cell. Proteomics* **7,** 2399–2409
32. Gupta, S., Ahadi, S., Zhou, W., and Röst, H. (2019) DIAlignR provides precise retention time alignment across distant runs in DIA and targeted proteomics. *Mol. Cell. Proteomics* **18,** 806–817
33. Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., and Mann, M. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13,** 2513–2526
34. Parca, L., Beck, M., Bork, P., and Ori, A. (2018) Quantifying compartment-associated variations of protein abundance in proteomics data. *Mol. Syst. Biol.* **14,** e8131
35. van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K., and van der Werf, M. J. (2006) Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics* **7,** 142

36. De Livera, A. M., Dias, D. A., De Souza, D., Rupasinghe, T., Pyke, J., Tull, D., Roessner, U., McConville, M., and Speed, T. P. (2012) Normalizing and integrating metabolomics data. *Anal. Chem.* **84,** 10768–10776
37. Fundel, K., Haag, J., Gebhard, P. M., Zimmer, R., and Aigner, T. (2008) Normalization strategies for mRNA expression data in cartilage research. *Osteoarthritis Cartilage* **16,** 947–955
38. Smolinska, A., Hauschild, A. C., Fijten, R. R., Dallinga, J. W., Baumbach, J., and van Schooten, F. J. (2014) Current breathomics—A review on data pre-processing techniques and machine learning in metabolomics breath analysis. *J. Breath Res.* **8,** 027105
39. Callister, S. J., Barry, R. C., Adkins, J. N., Johnson, E. T., Qian, W. J., Webb-Robertson, B. J., Smith, R. D., and Lipton, M. S. (2006) Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J. Proteome Res.* **5,** 277–286
40. Adriaens, M. E., Jaillard, M., Eijssen, L. M., Mayer, C. D., and Evelo, C. T. (2012) An evaluation of two-channel ChIP-on-chip and DNA methylation microarray normalization strategies. *BMC Genomics* **13,** 42
41. Tobin, J., Walach, J., de Beer, D., Williams, P. J., Filzmoser, P., and Walczak, B. (2017) Untargeted analysis of chromatographic data for green and fermented rooibos: Problem with size effect removal. *J. Chromatogr. A* **1525,** 109–115
42. Branson, O. E., and Freitas, M. A. (2016) A multi-model statistical approach for proteomic spectral count quantitation. *J. Proteomics* **144,** 23–32
43. Leek, J. T., and Storey, J. D. (2007) Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3,** 1724–1735
44. Guo, T., Kouvonen, P., Koh, C. C., Gillet, L. C., Wolski, W. E., Röst, H. L., Rosenberger, G., Collins, B. C., Blum, L. C., Gillessen, S., Joerger, M., Jochum, W., and Aebersold, R. (2015) Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. *Nat. Med.* **21,** 407–413
45. Liu, Y., Buil, A., Collins, B. C., Gillet, L. C., Blum, L. C., Cheng, L. Y., Vitek, O., Mouritsen, J., Lachance, G., Spector, T. D., Dermitzakis, E. T., and Aebersold, R. (2015) Quantitative variability of 342 plasma proteins in a human twin population. *Mol. Syst. Biol.* **11,** 786
46. Wu, J. X., Song, X., Pascovici, D., Zaw, T., Care, N., Krisp, C., and Molloy, M. P. (2016) SWATH mass spectrometry performance using extended peptide MS/MS assay libraries. *Mol. Cell. Proteomics* **15,** 2501–2514
47. Rausch, T. K., Schillert, A., Ziegler, A., Lüking, A., Zucht, H. D., and Schulz-Knappe, P. (2016) Comparison of pre-processing methods for multiplex bead-based immunoassays. *BMC Genomics* **17,** 601
48. Kuharev, J., Navarro, P., Distler, U., Jahn, O., and Tenzer, S. (2015) In-depth evaluation of software tools for data-independent acquisition based label-free quantification. *Proteomics* **15,** 3140–3151
49. Griffin, N. M., Yu, J., Long, F., Oh, P., Shore, S., Li, Y., Koziol, J. A., and Schnitzer, J. E. (2010) Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nat. Biotechnol.* **28,** 83–89
50. Risso, D., Ngai, J., Speed, T. P., and Dudoit, S. (2014) Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* **32,** 896–902
51. Williams, K. E., Lemieux, G. A., Hassis, M. E., Olshen, A. B., Fisher, S. J., and Werb, Z. (2016) Quantitative proteomic analyses of mammary organoids reveals distinct signatures after exposure to environmental chemicals. *Proc. Natl. Acad. Sci. U.S.A.* **113,** E1343–E1351
52. Blaise, B. J. (2013) Data-driven sample size determination for metabolic phenotyping studies. *Anal. Chem.* **85,** 8943–8950
53. Elo, L. L., Filén, S., Lahesmaa, R., and Aittokallio, T. (2008) Reproducibility-optimized test statistic for ranking genes in microarray studies. *IEEE/ACM Trans Comput. Biol. Bioinform.* **5,** 423–431
54. Pursiheimo, A., Vehmas, A. P., Afzal, S., Suomi, T., Chand, T., Strauss, L., Poutanen, M., Rokka, A., Corthals, G. L., and Elo, L. L. (2015) Optimization of statistical methods impact on quantitative proteomics data. *J. Proteome Res.* **14,** 4118–4126
55. Karpievitch, Y. V., Dabney, A. R., and Smith, R. D. (2012) Normalization and missing value imputation for label-free LC-MS analysis. *BMC Bioinformatics* ***13,*** S5
56. Barer, M. R., and Harwood, C. R. (1999) Bacterial viability and culturability. *Adv. Microb. Physiol.* **41,** 93–137
57. Letunic, I., and Bork, P. (2016) Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44,** W242–W245
58. Ramus, C., Hovasse, A., Marcellin, M., Hesse, A. M., Mouton-Barbosa, E., Bouyssié, D., Vaca, S., Carapito, C., Chaoui, K., Bruley, C., Garin, J., Cianferani, S., Ferro, M., Dorssaeler, A. V., Burlet-Schiltz, O., Schaeffer, C., Couté, Y., and Gonzalez de Peredo, A. (2016) Spiked proteomic standard dataset for testing label-free quantitative software and statistical methods. *Data Brief* **6,** 286–294
59. Mottawea, W., Chiang, C. K., Mühlbauer, M., Starr, A. E., Butcher, J., Abujamel, T., Deeke, S. A., Brandel, A., Zhou, H., Shokralla, S., Hajibabaei, M., Singleton, R., Benchimol, E. I., Jobin, C., Mack, D. R., Figeys, D., and Stintzi, A. (2016) Altered intestinal microbiota-host mitochondria crosstalk in new onset Crohn's disease. *Nat. Commun.* **7,** 13419
60. Schroeder, B. O., Birchenough, G. M. H., Ståhlman, M., Arike, L., Johansson, M. E. V., Hansson, G. C., and Bäckhed, F. (2018) Bifidobacteria or fiber protects against diet-induced microbiota-mediated colonic mucus deterioration. *Cell Host Microbe* **23,** 27–40.e7
61. Tilocca, B., Burbach, K., Heyer, C. M. E., Hoelzle, L. E., Mosenthin, R., Stefanski, V., Camarinha-Silva, A., and Seifert, J. (2017) Dietary changes in nutritional studies shape the structural and functional composition of the pigs' fecal microbiome-from days to weeks. *Microbiome* **5,** 144
62. Govaert, E., Van Steendam, K., Scheerlinck, E., Vossaert, L., Meert, P., Stella, M., Willems, S., De Clerck, L., Dhaenens, M., and Deforce, D. (2016) Extracting histones for the specific purpose of label-free MS. *Proteomics* **16,** 2937–2944
63. Tabb, D. L., Vega-Montoto, L., Rudnick, P. A., Variyath, A. M., Ham, A. J., Bunk, D. M., Kilpatrick, L. E., Billheimer, D. D., Blackman, R. K., Cardasis, H. L., Carr, S. A., Clauser, K. R., Jaffe, J. D., Kowalski, K. A., Neubert, T. A., Regnier, F. E., Schilling, B., Tegeler, T. J., Wang, M., Wang, P., Whiteaker, J. R., Zimmerman, L. J., Fisher, S. J., Gibson, B. W., Kinsinger, C. R., Mesri, M., Rodriguez, H., Stein, S. E., Tempst, P., Paulovich, A. G., Liebler, D. C., and Spiegelman, C. (2010) Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *J. Proteome Res.* **9,** 761–776
64. Weisser, H., and Choudhary, J. S. (2017) Targeted feature detection for data-dependent shotgun proteomics. *J. Proteome Res.* **16,** 2964–2974
65. Chong, P. K., Gan, C. S., Pham, T. K., and Wright, P. C. (2006) Isobaric tags for relative and absolute quantitation (iTRAQ) reproducibility: Implication of multiple injections. *J. Proteome Res.* **5,** 1232–1240
66. Simula, M. P., Cannizzaro, R., Marin, M. D., Pavan, A., Toffoli, G., Canzonieri, V., and De Re, V. (2009) Two-dimensional gel proteome reference map of human small intestine. *Proteome Sci.* **7,** 10
67. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Pérez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yilmaz, S., Tiwary, S., Cox, J., Audain, E., Walzer, M., Jarnuczak, A. F., Ternent, T., Brazma, A., and Vizcaíno, J. A. (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47,** D442–D450
68. Röst, H. L., Rosenberger, G., Navarro, P., Gillet, L., Miladinović, S. M., Schubert, O. T., Wolski, W., Collins, B. C., Malmström, J., Malmstrom, L., and Aebersold, R. (2014) OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat. Biotechnol.* **32,** 219–223
69. Gillet, L. C., Navarro, P., Tate, S., Rost, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11,** O111.016717
70. Madeira, J. P., Alpha-Bazin, B., Armengaud, J., Omer, H., and Duport, C. (2016) Proteome data to explore the impact of pBClin15 on *Bacillus cereus* ATCC 14579. *Data Brief* **8,** 1243–1246
71. Milac, T. I., Randolph, T. W., and Wang, P. (2012) Analyzing LC-MS/MS data by spectral count and ion abundance: Two case studies. *Stat Interface* **5,** 75–87
72. Yee, K. M., Feener, E. P., Madigan, M., Jackson, N. J., Gao, B. B., Ross-Cisneros, F. N., Provis, J., Aiello, L. P., Sadun, A. A., and Sebag, J. (2015) Proteomic analysis of embryonic and young human vitreous. *Invest. Ophthalmol. Vis. Sci.* **56,** 7036–7042
73. Proietti, C., Zakrzewski, M., Watkins, T. S., Berger, B., Hasan, S., Ratnatunga, C. N., Brion, M. J., Crompton, P. D., Miles, J. J., Doolan, D. L., and Krause, L. (2016) Mining, visualizing and comparing multidimensional biomolecular data using the Genomics Data Miner (GMine) web-server. *Sci. Rep.* **6,** 38178

74. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., Mann, M., and Cox, J. (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13,** 731–740
75. Hoekman, B., Breitling, R., Suits, F., Bischoff, R., and Horvatovich, P. (2012) msCompare: a framework for quantitative analysis of label-free LC-MS data for comparative candidate biomarker studies. *Mol. Cell. Proteomics* **11,** M111.015974
76. Webb-Robertson, B. J., Matzke, M. M., Jacobs, J. M., Pounds, J. G., and Waters, K. M. (2011) A statistical selection strategy for normalization procedures in LC-MS proteomics experiments through dataset-dependent ranking of normalization scaling factors. *Proteomics* **11,** 4736–4741
77. Weiner, A. K., Sidoli, S., Diskin, S. J., and Garcia, B. (2017) GiaPronto: A one-click graph visualization software for proteomics datasets. *Mol. Cell. Proteomics* **17,** TIR117.000438
78. Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J. R., Vázquez-Baeza, Y., Birmingham, A., Hyde, E. R., and Knight, R. (2017) Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5,** 27
79. Karp, N. A., Huber, W., Sadowski, P. G., Charles, P. D., Hester, S. V., and Lilley, K. S. (2010) Addressing accuracy and precision issues in iTRAQ quantitation. *Mol. Cell. Proteomics* **9,** 1885–1897
80. Lo, K., and Gottardo, R. (2012) Flexible mixture modeling via the multivariate t distribution with the Box-Cox transformation: An alternative to the skew-t distribution. *Stat. Comput.* **22,** 33–52
81. Lin, S. M., Du, P., Huber, W., and Kibbe, W. A. (2008) Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Res.* **36,** e11
82. Wang, S. Y., Kuo, C. H., and Tseng, Y. J. (2013) Batch normalizer: A fast total abundance regression calibration method to simultaneously adjust batch and injection order effects in liquid chromatography/time-of-flight mass spectrometry-based metabolomics data and comparison with current calibration methods. *Anal. Chem.* **85,** 1037–1046
83. Wang, X., Zhang, A., Han, Y., Wang, P., Sun, H., Song, G., Dong, T., Yuan, Y., Yuan, X., Zhang, M., Xie, N., Zhang, H., Dong, H., and Dong, W. (2012) Urine metabolomics analysis for biomarker discovery and detection of jaundice syndrome in patients with liver disease. *Mol. Cell. Proteomics* **11,** 370–380
84. Di Guida, R., Engel, J., Allwood, J. W., Weber, R. J., Jones, M. R., Sommer, U., Viant, M. R., and Dunn, W. B. (2016) Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling. *Metabolomics* **12,** 93
85. Smilde, A. K., van der Werf, M. J., Bijlsma, S., van der Werff-van der Vat, B. J., and Jellema, R. H. (2005) Fusion of mass spectrometry-based metabolomics data. *Anal. Chem.* **77,** 6729–6736
86. Craig, A., Cloarec, O., Holmes, E., Nicholson, J. K., and Lindon, J. C. (2006) Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *Anal. Chem.* **78,** 2262–2267
87. Ballman, K. V., Grill, D. E., Oberg, A. L., and Therneau, T. M. (2004) Faster cyclic loess: normalizing RNA arrays via linear models. *Bioinformatics* **20,** 2778–2786
88. Wang, B., Wang, X. F., and Xi, Y. (2011) Normalizing bead-based microRNA expression data: A measurement error model-based approach. *Bioinformatics* **27,** 1506–1512
89. Karpievitch, Y. V., Taverner, T., Adkins, J. N., Callister, S. J., Anderson, G. A., Smith, R. D., and Dabney, A. R. (2009) Normalization of peak intensities in bottom-up MS-based proteomics using singular value decomposition. *Bioinformatics* **25,** 2573–2580
90. Stacklies, W., Redestig, H., Scholz, M., Walther, D., and Selbig, J. (2007) pcaMethods—A bioconductor package providing PCA methods for incomplete data. *Bioinformatics* **23,** 1164–1167
91. Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., and Altman, R. B. (2001) Missing value estimation methods for DNA microarrays. *Bioinformatics* **17,** 520–525
92. Kim, H., Golub, G. H., and Park, H. (2005) Missing value estimation for DNA microarray gene expression data: Local least squares imputation. *Bioinformatics* **21,** 187–198