analytical chemistry

Article

pubs.acs.org/ac

pQuant Improves Quantitation by Keeping out Interfering Signals and Evaluating the Accuracy of Calculated Ratios

Chao Liu,[†,‡] Chun-Qing Song,[§] Zuo-Fei Yuan,[†] Yan Fu,[†] Hao Chi,[†,‡] Le-Heng Wang,[†] Sheng-Bo Fan,[†,‡] Kun Zhang,[†,‡] Wen-Feng Zeng,[†,‡] Si-Min He,[†,‡] Meng-Qiu Dong,*[,§] and Rui-Xiang Sun*[,†,‡]

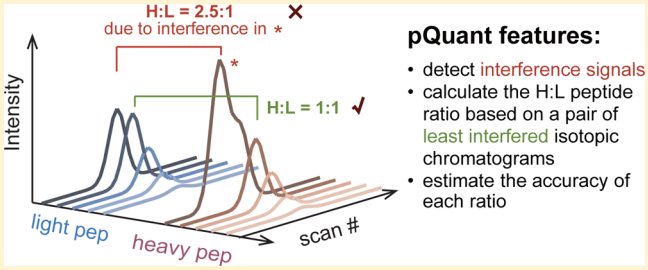
[†]Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing 100190, China

[‡]University of Chinese Academy of Sciences, Beijing 100049, China

[§]National Institute of Biological Sciences, Beijing, Beijing 102206, China

Supporting Information

ABSTRACT: In relative protein abundance determination from peptide intensities recorded in full mass scans, a major complication that affects quantitation accuracy is signal interference from coeluting ions of similar m/z values. Here, we present pQuant, a quantitation software tool that solves this problem. pQuant detects interference signals, identifies for each peptide a pair of least interfered isotopic chromatograms: one for the light and one for the heavy isotope-labeled peptide. On the basis of these isotopic pairs, pQuant calculates the relative heavy/light peptide ratios along with their 99.75% confidence intervals (CIs). From the peptides ratios and their CIs, pQuant estimates the protein ratios and associated CIs by kernel density estimation. We tested pQuant, Census and MaxQuant on data sets obtained from mixtures (at varying mixing ratios from 10:1 to 1:10) of light- and heavy-SILAC labeled HeLa cells or ^{14}N- and ^{15}N-labeled *Escherichia coli* cells. pQuant quantitated more peptides with better accuracy than Census and MaxQuant in all 14 data sets. On the SILAC data sets, the nonquantified "NaN" (not a number) ratios generated by Census, MaxQuant, and pQuant accounted for 2.5−10.7%, 1.8−2.7%, and 0.01−0.5% of all ratios, respectively. On the ^{14}N/^{15}N data sets, which cannot be quantified by MaxQuant, Census and pQuant produced 0.9−10.0% and 0.3−2.9% NaN ratios, respectively. Excluding these NaN results, the standard deviations of the numerical ratios calculated by Census or MaxQuant are 30−100% larger than those by pQuant. These results show that pQuant outperforms Census and MaxQuant in SILAC and ^{15}N-based quantitation.



Much progress has been made in mass spectrometry (MS)-based quantitative proteomics in recent years, as evidenced by numerous applications, such as biomarker discovery,[1] study of chromatin assembly and disassembly,[2] identification of insulin signaling targets,[3] and protein post-translational modification (PTM).[4]

Among the most commonly used quantitative strategies are full MS scan-based quantitation methods, such as SILAC (stable isotope labeling with amino acids in cell),[5] ^{15}N-labeling,[6] and ^{18}O-labeling.[7] In these strategies, proteins are metabolically labeled with stable isotopes, digested into peptides, and then analyzed using liquid chromatography (LC)-MS/MS. Quantitation software tools are designed to extract the intensities of pairs of light (L, unlabeled) and heavy (H, labeled) peptides from full MS scans. The relative abundance ratio of a protein between two conditions is then calculated based on the ratios of its constituent peptides.[8]

For high-complexity samples such as whole cell lysates, it is not uncommon that a peptide coelutes with another peptide or a nonpeptide contaminant of a similar m/z value.[9] The interference caused by coeluting ions of similar m/z values can seriously compromise the accuracy of quantitation.[10,11] We examined two leading quantitation software tools Census[12] and MaxQuant,[13] and found that a lot of the peptide quantitation results are "NaN" (not a number) or outliers (far from the sample mixing ratio) (Tables 1−3), and many of them are due to interference signals. For example, in our SILAC data of 1:1 (H/L) mixed HeLa cells, Census outputs 167 NaN and 420 outlier ratios that are larger than 2 or smaller than 0.5. Among the 587 ratios, only 19 (3%) correspond to low-abundance peptides; the other 568 (97%) have obvious signals of both the light peptide and the heavy peptide in the chromatograms, and most of them are quantified erroneously because of interference from coeluting isobaric ions.

For those that are not NaN ratios, it is desirable to know which ones are accurate and which ones not. However, in spite of the pressing need, few of the existing quantitation tools provide accuracy evaluation for calculated abundance ratios of peptides and proteins. Census attempts to assess the accuracy of a peptide

Received: December 31, 2013
Accepted: May 5, 2014
Published: May 5, 2014

   dx.doi.org/10.1021/ac404246w | *Anal. Chem.* 2014, 86, 5286−5294

Table 1. Comparison of the pQuant and Census Quantitation Results on the SILAC Data[a]

			peptide level[b]					protein group level[c]			
				NaN ratios		numerical ratios				numerical ratios	
sample-mixing ratio (H/L)	expected $\log_2$(ratio)	quant. software	no. all ratios	no.	%	median	S.D.	no. all ratios	no. NaN ratios	median	S.D.
1:1	0.00	pQuant	6750	4	0.06	0.01	0.32	1219	0	0.01	0.32
		Census		167	2.47	0.08	0.66		23	0.11	0.62
1:2	−1.00	pQuant	8852	8	0.09	−0.86	0.34	1705	2	−0.83	0.30
		Census		248	2.80	−0.81	0.84		37	−0.71	0.79
2:1	1.00	pQuant	9198	14	0.15	1.06	0.38	1820	0	1.07	0.41
		Census		312	3.39	1.18	0.67		37	1.21	0.62
1:5	−2.32	pQuant	8804	20	0.23	−2.18	0.50	1993	4	−2.12	0.49
		Census		554	6.29	−2.06	1.30		81	−1.74	1.14
5:1	2.32	pQuant	5722	23	0.40	2.50	0.75	1667	2	2.51	0.75
		Census		441	7.71	2.64	0.99		82	2.74	0.89
1:10	−3.32	pQuant	6879	29	0.42	−3.18	0.76	1833	6	−3.06	0.71
		Census		735	10.70	−2.84	1.75		116	−2.25	1.48
10:1	3.32	pQuant	4654	25	0.53	3.14	0.78	1524	4	3.11	0.81
		Census		475	10.20	3.47	1.06		93	3.47	0.97

[a]Calculated ratios are $\log_2$-transformed. [b]In quantification of each peptide, only the PSM with the highest identification score was used as the starting point to reconstruct chromatograms. [c]To guarantee a fair comparison, we calculate the protein ratios at the group level by taking the median of the peptide ratios.

ratio by its determinant score (R^2),[12] and in some studies, R^2 is even used to recalibrate ratios.[14,15] However, R^2 is essentially a value that describes how well the chromatogram of a light peptide correlates with that of the corresponding heavy peptide; it is not a measure of accuracy per se. A ratio far from the correct value sometimes has a high R^2 value (Figure 1).

Here, we describe a new quantitation software tool called pQuant, in which we strive to minimize the interference of coeluting ions of similar m/z values. [pQuant can be downloaded from http://pfind.ict.ac.cn/software/pQuant/index.html.] pQuant reconstructs a chromatogram for each isotopic peak, generating one set of isotopic chromatograms for the light peptide and another set for the heavy peptide. pQuant then calculates the peptide ratio and the associated confidence interval (CI) based on the least interfered isotopic chromatogram of the light peptide and the least interfered chromatogram of the heavy peptide. Finally, a protein ratio is calculated from peptide ratios by kernel density estimation. The accuracy of a protein ratio is estimated based on the number of quantified peptides of the protein and the confidence intervals of these peptide ratios. We show that pQuant greatly improves full MS scan-based, proteome-wide quantitation. pQuant results greatly reduces the number of NaN ratios and outlier ratios (Tables 1−3). In our 1:1 SILAC data, pQuant outputs only 4 NaN ratios and 80 outliers (ratios >2 or <0.5), that is, of the 587 inaccurate Census ratios described above, pQuant successfully corrected 503 (86%).

EXPERIMENTAL SECTION

Sample Preparation. HeLa cells were grown in DMEM deficient medium (Invitrogen) containing 10% dialyzed FBS (Invitrogen), supplemented with [^{13}C6, ^{15}N2] labeled L-lysine and [^{13}C6, ^{15}N4] labeled L-arginine (Cambridge Isotope Laboratories, Inc.) or normal lysine and arginine. The heavy- and light- isotopic labeled whole-cell lysates were mixed at seven different ratios 10:1, 5:1, 2:1, 1:1, 1:2, 1:5, and 1:10, by total protein amounts. The *Escherichia coli* strain MG1655 was cultured at 37 °C overnight in M9 medium containing either ^{15}N-labeled (≥99% atomic enrichment of ^{15}N) or unlabeled ammonium. The ^{14}N- (i.e., unlabeled) and ^{15}N-labeled cultures were mixed at seven different ratios 10:1, 3:1, 1.5:1, 1:1, 1:1.5, 1:3, and 1:10 by total cell numbers (OD_{600} mL).

Mass Spectrometry. All experiments were performed on an LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific, San Jose, CA). A MudPIT[16] column was placed in-line with an Agilent 1200 quaternary HPLC (Agilent, Palo Alto, CA). Eluted peptides were sprayed into the mass spectrometer and analyzed using a modified 5-step MudPIT method. After the first reverse phase gradient (MudPIT step 1), the remaining four steps each started with a salt pulse containing 20, 50, 100, or another 100% of buffer C before the reverse phase gradient. Full MS scans were acquired in the Orbitrap (400−2000 m/z, R = 60000), and each full MS scan was followed by data-dependent MS/MS scans in the linear ion trap on the eight most intense ions at 35% normalized collision energy. Ions already selected once for MS/MS were dynamically excluded for 30 s.

Database Searching. For the SILAC data, MS/MS spectra were searched against the concatenated forward and reversed IPI human database (version 3.68) using ProLuCID,[17] Andromeda (a search engine embedded in MaxQuant, version 1.4)[18] and pFind,[19−21] respectively. Carbamidomethylation of cysteines were included as a fixed modification. The estimated false discovery rate (FDR) was no more than 1% for identified spectra. The filtered SILAC identification results were quantified using Census (version 1.57), MaxQuant, and pQuant at their respective default settings for SILAC. The ^{14}N/^{15}N data were analyzed similarly, except that an *E. coli* database (Escherichia_coli_K_12_substr_MG1655 from NCBI) and appropriate settings for ^{14}N/^{15}N labeling were used. Supporting Information Table 1 lists the key parameters of the software tools that we used.

Overview of pQuant. As shown in Supporting Information Figure 1, the workflow of pQuant consists of three steps: extraction of peptide signals, quantitation of peptide ratios, and quantitation of protein ratios.

Step 1: Extraction of Peptide Signals. This module carries out three computational tasks: (1) data preparation, (2)

dx.doi.org/10.1021/ac404246w | Anal. Chem. 2014, 86, 5286−5294

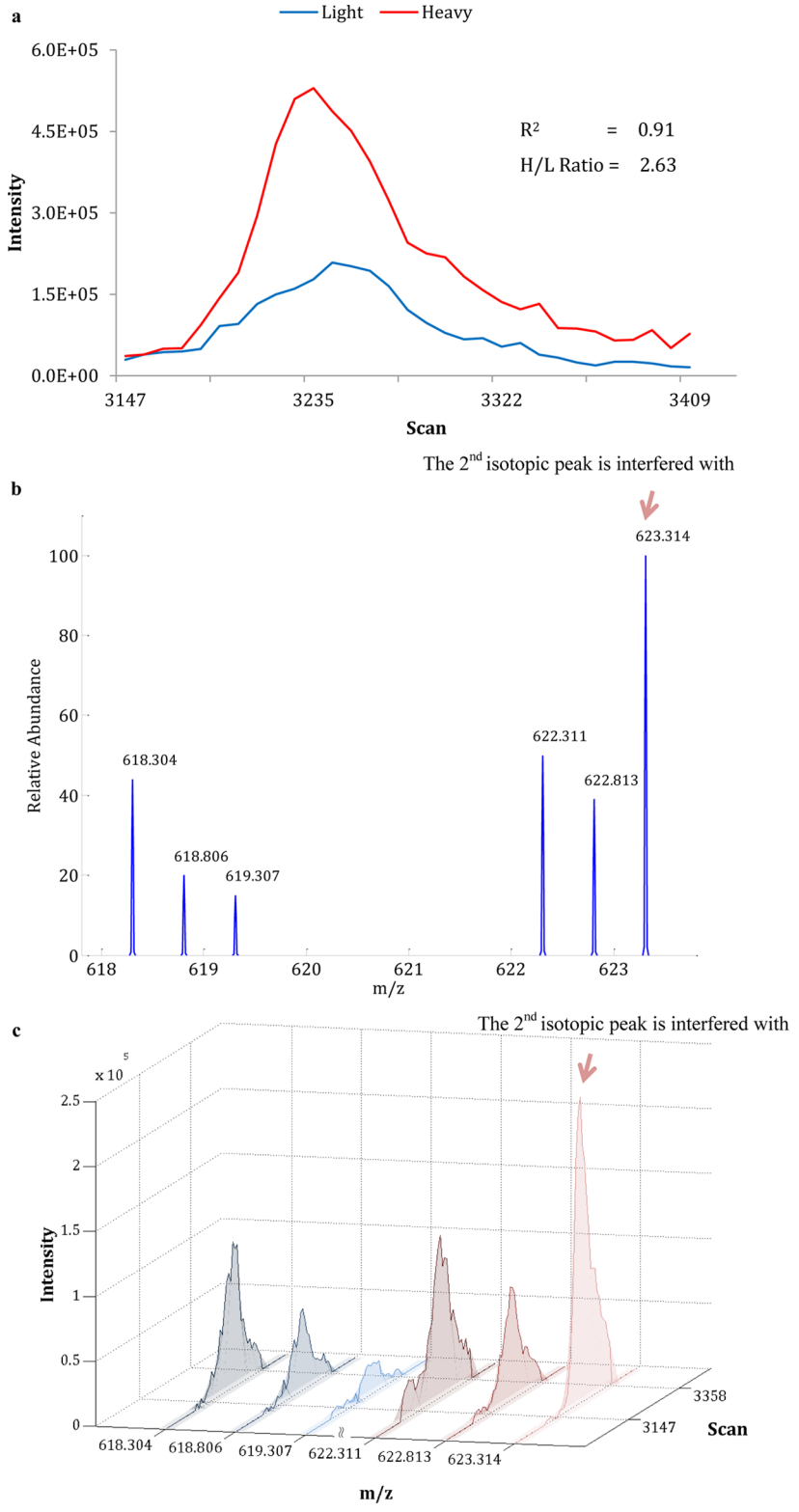


Figure 1. Quantitation results for peptide GHTQQDPEVPK from the H/L = 1:1 mixed SILAC data set. (a) The light (blue) and heavy (red) peptide chromatograms constructed by Census. The determinant score R^2 is 0.91, suggesting good correlation between the chromatograms and that the regression H/L ratio of 2.63 should be trustworthy. However, the expected ratio is 1:1. (b) A full-scan spectrum (scan 3218) showing the precursor peptides. There is obvious interference in the 2nd isotopic peak of the heavy peptide (indicated by an arrow), and the interference signal is probably included in the chromatogram reconstruction by Census. (c) The isotopic chromatograms of the monoisotopic, 1st, and 2nd isotopic peaks of the light (blue) or heavy (red) peptide are shown along the *m*/*z*, elution time (represented by scan numbers) and intensity dimensions. pQuant automatically determines the start and end points of all the isotopic chromatograms. The start point is Scan 3147 and the end point is Scan 3358. Quantitation is based on two isotopic chromatograms (one for the light peptide and one for the heavy peptide) that are least affected. Here pQuant uses the monoisotopic chromatograms of the light and the heavy peptide, and the calculated ratio is 1.09 with a normalized CI of [90%, 110%], very close to 1:1.

dx.doi.org/10.1021/ac404246w | *Anal. Chem.* 2014, 86, 5286−5294

identifying experimental isotopic peaks in full MS scans, and (3) constructing isotopic chromatograms.

(1) Data Preparation. First, MS files are indexed. Then high-confidence peptide-spectrum matches (PSMs) produced by identification software tools are read into the program.

(2) Identifying Experimental Isotopic Peaks. For each input peptide sequence, pQuant calculates the theoretical distribution of isotopic peaks using a stepwise convolution algorithm[22] and identifies experimental isotopic peaks in a range of MS scans where the peptide may be expected. Because the trigger MS scan (the MS scan immediately before the MS/MS spectrum that identifies a peptide) is often off the center of the chromatogram and most of our peptides have a chromatographic peak width of less than 1 min, a 2 min window (defined by users) centered at the trigger MS scan is applied to include all the MS scans containing the peptide.

For each MS scan in this window, pQuant applies a ppm-level m/z tolerance window (±10 ppm for our data sets) around the theoretical m/z values of the isotopic peaks of the peptide to select the experimental peaks. After the removal of low-intensity peaks below a user-defined threshold (1E4 in this study), if more than one peak is found in any of the theoretical mass windows, pQuant will decide which one is better by allowing only one peak per window and comparing all possible combinations of the candidate isotopic peaks to the theoretical pattern. The one with an intensity distribution pattern most similar to the theoretical pattern is kept and its constituent peaks are regarded as experimental isotopic peaks of the peptide in the current MS scan. In this way, pQuant distinguishes true peptide signals from noise and other peptides.

To guard against possible incomplete labeling of stable isotopes, pQuant estimates for each peptide the atomic enrichment ratio as described before.[23] pQuant simulates the theoretical isotopic distributions at varying atomic enrichment ratios (100%, 99%, 98%, etc.), calculates the similarity between the experimental isotopic distribution and a theoretical isotopic distribution, and determines the enrichment value that generates the best match.

(3) Constructing Isotopic Chromatograms. Some quantitation tools including Census construct a pair of chromatograms for each peptide: the intensities of all isotopic peaks of the light peptide are summed to generate one chromatogram and another one for the corresponding heavy peptide. These are referred to as "peptide chromatograms" in this paper. In contrast, MaxQuant constructs chromatograms for individual isotopic peaks of the light and the heavy peptides. These "isotopic chromatograms" are used in pQuant.

Next, pQuant determines the start and end points of isotopic chromatograms. For each isotopic peak, its experimental intensities along the retention time axis in contiguous MS scans are assembled into a profile. This profile extends both left and right from the trigger MS scan until the intensity drops below 10% of the apex of the extending profile. Because an MS2 scan is usually triggered before the intensity of the precursor ion reaches the apex, the left and right extensions are typically asymmetric, and there is often an upward tail in the extending profile. Thus, it is necessary to split the profile at the local minima nearest to the growing ends, and what remains in the middle is a temporary isotopic chromatogram for the next step. This algorithm is resistant to burrs and tailing of chromatographic peaks, and the computing complexity is low. However, this process alone is not sufficient to determine the start and end points of an isotopic chromatogram. We have observed that the chromatograms of the fourth and fifth isotopic peaks are generally not as wide as the chromatogram of the monoisotopic peak unless they are interfered by other ions, in which case the extended parts of the fourth or fifth isotopic chromatogram are discarded. To accommodate slight retention time shifts between a pair of light and heavy peptides, pQuant shifts these chromatograms across each other until the correlation coefficient reaches a maximum.

Step 2: Quantitation of Peptide Ratios. We use the vector $\boldsymbol{t}^l = (t^l_1, ..., t^l_N)$ to represent the theoretical isotopic abundances of the light peptide, and the vector $\boldsymbol{t}^h = (t^h_1, ..., t^h_M)$ to represent the theoretical isotopic abundances of the heavy peptide, in which N is the number of the isotopic peaks of the light peptide, and M is the number of the isotopic peaks of the heavy peptide. Isotopic peaks with relative abundances less than 5% of the base peak are exlcuded, so usually, $3 \le N \le 5$ and $3 \le M \le 5$. The intensity of the ith isotopic peak in the jth scan of the *light* peptide is indicated by $p^l_{i,j}$. For the monoisotopic peak, $i = 1$; for the first isotopic peak, $i = 2$; the experimental isotopic intensities of the light peptide in the j-th scan is $(p^l_{1,j}, ..., p^l_{N,j})$. The monoisotopic chromatogram of the light peptide is represented by the vector $\boldsymbol{p}^l_1 = (p^l_{1,1}, ..., p^l_{1,K})$. K is the number of scans in the chromatogram. The same symbols are also used for the heavy peptide except that the superscript is "h". These terms are visualized in Supporting Information Figure 2.

Next, all the isotopic chromatograms are normalized. The normalized chromatograms of the light peptide are $\boldsymbol{L}_1 = \boldsymbol{p}^l_1(\sum t^l_i/t^l_1)$, $\boldsymbol{L}_2 = \boldsymbol{p}^l_2(\sum t^l_i/t^l_2)$, ..., $\boldsymbol{L}_N = \boldsymbol{p}^l_N(\sum t^l_i/t^l_N)$. The normalized chromatograms of the heavy peptide are $\boldsymbol{H}_1 = \boldsymbol{p}^h_1(\sum t^h_i/t^h_1)$, $\boldsymbol{H}_2 = \boldsymbol{p}^h_2(\sum t^h_i/t^h_2)$, ..., $\boldsymbol{H}_M = \boldsymbol{p}^h_M(\sum t^h_i/t^h_M)$. Once normalized, any two chromatograms, one from the light peptide and the other from the heavy peptide, can be used to calculate the H/L ratio of the peptide.

Given two chromatograms, for example, one is the monoisotopic chromatogram of the light peptide, and the other is the monoisotopic chromatogram of the heavy peptide, let $\mathbf{X} = \boldsymbol{L}_1 = (X_1, ..., X_j, ..., X_K)$ and $\mathbf{Y} = \boldsymbol{H}_1 = (Y_1, ..., Y_j, ..., Y_K)$. pQuant calculates the ratio using the regression model $\mathbf{Y} = a\mathbf{X} + e$ where a is the H/L ratio and e indicates a Gaussian noise. The optimal value of a is solved using the least-squares method as $\hat{a} = \sum X_j Y_j / \sum X_j X_j$. The confidence interval for a is calculated as follows. On the basis of the properties of the least-squares estimator, the estimated standard error of $\hat{a}$ is $\hat{\sigma} = (K^{-1} \cdot \sum (Y_j - \hat{a}X_j)^2 / \sum X_j^2)^{1/2}$. Because $(\hat{a} - a)/\hat{\sigma} \rightarrow N(0,1)$, the so-called "asymptotic normality" of the least-squares estimator, the $(1 - \alpha)$ confidence interval for $\hat{a}$ is approximately $[\hat{a} - z_{\alpha/2}\hat{\sigma}, \hat{a} + z_{\alpha/2}\hat{\sigma}]$. In pQuant, we set $\alpha = 0.0025$, so, $z_{\alpha/2} = 3$, and the 99.75% confidence interval for an H/L ratio $\hat{a}$ is $[\hat{a} - 3\hat{\sigma}, \hat{a} + 3\hat{\sigma}]$.

To guarantee that the CIs of different ratios are comparable, we transform the original CIs $[\hat{a} - 3\hat{\sigma}, \hat{a} + 3\hat{\sigma}]$ to normalized CIs $[(\hat{a} - 3\hat{\sigma})/\hat{a}, (\hat{a} + 3\hat{\sigma})/\hat{a}]$ and express the latter in percentages, for example, [80%, 120%]. CI refers to "normalized 99.75% confidence interval" hereinafter unless indicated otherwise.

As such, pQuant calculates $N \times M$ ratios and their corresponding CIs. To determine which ratio is the most accurate, we follow a basic idea that CI negatively correlates with the amount of interference experienced by a peak (Figure 2). For example, as illustrated in Figure 1, from the monoisotopic chromatogram of the light peptide and that of the heavy peptide, the calculated CI is [90%, 110%], and using the second isotopic chromatogram of the heavy peptide the calculated CI is [70%, 130%]. The latter CI value is larger because the second isotopic chromatogram of the heavy peptide has interfering signals. For each peptide, pQuant outputs the ratio with the narrowest CI as

dx.doi.org/10.1021/ac404246w | Anal. Chem. 2014, 86, 5286−5294

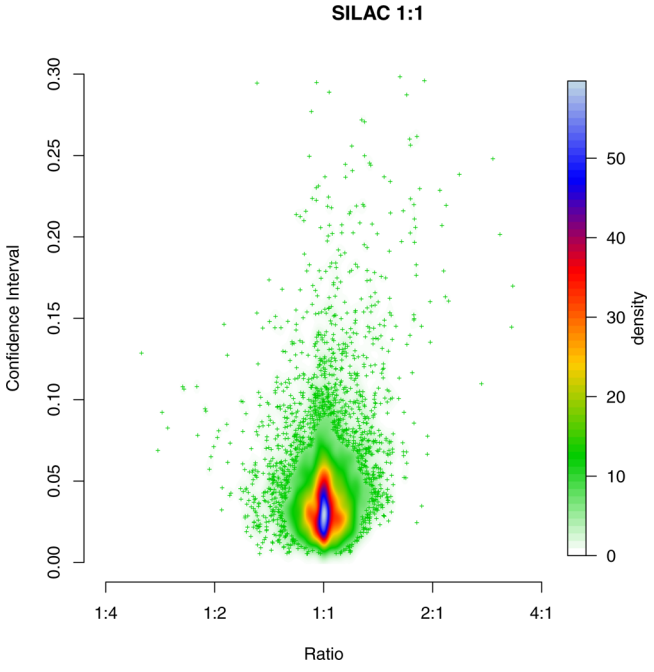


Figure 2. Scatter plot of peptide ratios and CIs from the 1:1 SILAC data. For each peptide, its log2-transformed ratio is shown on the *X*-axis, and the *Y*-axis value represents the corresponding confidence interval using the $\hat{\sigma}$ value. The 99.75% confidence interval for an H/L ratio $\hat{a}$ is $[\hat{a} - 3\hat{\sigma}, \hat{a} + 3\hat{\sigma}]$. Ratios with small CIs cluster tightly around 1:1, and those with larger CIs do not, indicating that ratios with small CIs are mostly accurate.

the peptide ratio. The associated CI value is also included in the output as an accuracy indicator.

pQuant and MaxQuant use the same regression model $\mathbf{Y} = a\mathbf{X} + e$. Census uses another regression model $\mathbf{Y} = a\mathbf{X} + b + e$, for which the optimal values of a and b can be estimated using the least-squares method:

$$\begin{cases} \hat{a} = \sum\left(X_j - \dfrac{\sum X_j}{K}\right)\left(Y_j - \dfrac{\sum Y_j}{K}\right) \Big/ \sum\left(X_j - \dfrac{\sum X_j}{K}\right)^2 \\ \quad \approx \sum X_j Y_j \Big/ \sum X_j X_j \\ \hat{b} = \dfrac{\sum Y_j}{K} - \hat{a}\cdot\dfrac{\sum X_j}{K} \approx 0 \end{cases}$$

Mathematically, the model used by Census is regression with constant term, and the model used by pQuant is regression through the origin (RTO). Textbooks rarely discuss RTO other than to caution against dropping the constant term from a regression, but RTO is appropriate or even necessary in some circumstances.[24,25] Theoretically, RTO is a special case of regression with constant term.[26−30] In reality, before and after a peptide is eluted, the signal of the peptide should be null, therefore RTO is more appropriate than the regression with constant term. Besides, the latter is vulnerable to interfering signals, sometimes resulting in $a < 0$ (see examples in Results and the mathematical explanation by Hocking et al.[31]).

Step 3: Quantitation of Protein Ratios. For shotgun proteomics, protein ratios are calculated from peptide ratios. Using a large data set, we tested many of the methods that had been used,[8] including simple average with or without outliers, linear regression with or without outliers, taking the median of peptide ratios, and computing a weighted average by considering peptide intensity. We find that if a protein has many quantified peptides, the calculated protein ratios using different methods are all similar and fairly reliable (Figure 3a). However, if a protein has only two quantified peptides and one of peptide ratio is inaccurate, they all fail (Figure 3b).

In pQuant, we use kernel density estimation,[32] a widely used nonparametric method, to derive protein ratios and estimate the accuracy of protein ratios from the peptide ratios and their CIs. It is relatively simple and intuitive; but for its effective use, parameters such as the kernel function and bandwidth need to be carefully selected. We find that the commonly used Gaussian kernel function is appropriate in this application:

$$f_{\text{pep}}(x) = \frac{1}{T}\cdot\frac{1}{\sqrt{2\pi}\,\sigma_{\text{pep}}} e^{-(x-\mu_{\text{pep}})^2/2\sigma_{\text{pep}}^2} \quad (1)$$

where μ_{pep} is the peptide ratio $\hat{a}$, σ_{pep} is the estimated standard deviation of $\hat{a}$, and T is the number of peptide hits of a protein. The protein ratio distribution can be calculated by $F(x) = \sum f_{\text{pep}}(x)$ in which $F(x)$ is an irregular curve but can be fitted to a Gaussian distribution:

$$f_{\text{pro}}(x) = \frac{1}{\sqrt{2\pi}\,\sigma_{\text{pro}}} e^{-(x-\mu_{\text{pro}})^2/2\sigma_{\text{pro}}^2} \quad (2)$$

where μ_{pro} is the estimated protein ratio, and σ_{pro} is the estimated standard deviation of μ_{pro}. The normalized 99.75% confidence interval of μ_{pro} is $[(\mu_{\text{pro}} - 3\sigma_{\text{pro}})/\mu_{\text{pro}}, (\mu_{\text{pro}} + 3\cdot\sigma_{\text{pro}})/\mu_{\text{pro}}]$.

RESULTS AND DISCUSSION

Performance evaluation of pQuant. We compared pQuant with Census and MaxQuant. To guarantee a fair comparison, we adopted the following procedure based on previously suggested rules:[33] (1) Light and heavy samples are of the same biological state to minimize the difference between samples. Thus, most of the measured ratios should be close to the sample mixing ratios. Our SILAC data and $^{14}N/^{15}N$ data were prepared by following this rule. (2) All ratios must be log transformed (see Supporting Information) and without post-quantitation normalization. The overlap of the quantitation results by two quantitation tools is used for comparison to avoid the differences introduced by identification. Peptides belonging to contaminant proteins (according to http://www.maxquant.org/contaminants.zip) are removed. (3) NaN ratios are analyzed. These refer to the NaN or null results by MaxQuant and the N/A results by Census. pQuant output "Infinite" or "−Infinite" when the light or the heavy peptide has extremely low or no signal in MS scans[34,35] (Supporting Information Figure 3). These ratios must be analyzed separately, because they are not positive rational numbers and cannot be used to calculate standard deviation or mean ratios. (4) After the removal of NaN ratios, we compare the deviation of the median ratio from the sample-mixing ratio and also the standard deviations of ratios between software tools. Both the median ratio and the standard deviation of ratios reflect accuracy. For the same data set, if two quantitation tools obtain the same expected median ratio, the one with a smaller standard deviation outperforms the other.

Comparison of pQuant and Census. The performance of pQuant was compared to Census using the SILAC data (Table 1) and the $^{14}N/^{15}N$ data (Table 2). The NaN ratios are only 0.06−

dx.doi.org/10.1021/ac404246w | Anal. Chem. 2014, 86, 5286−5294

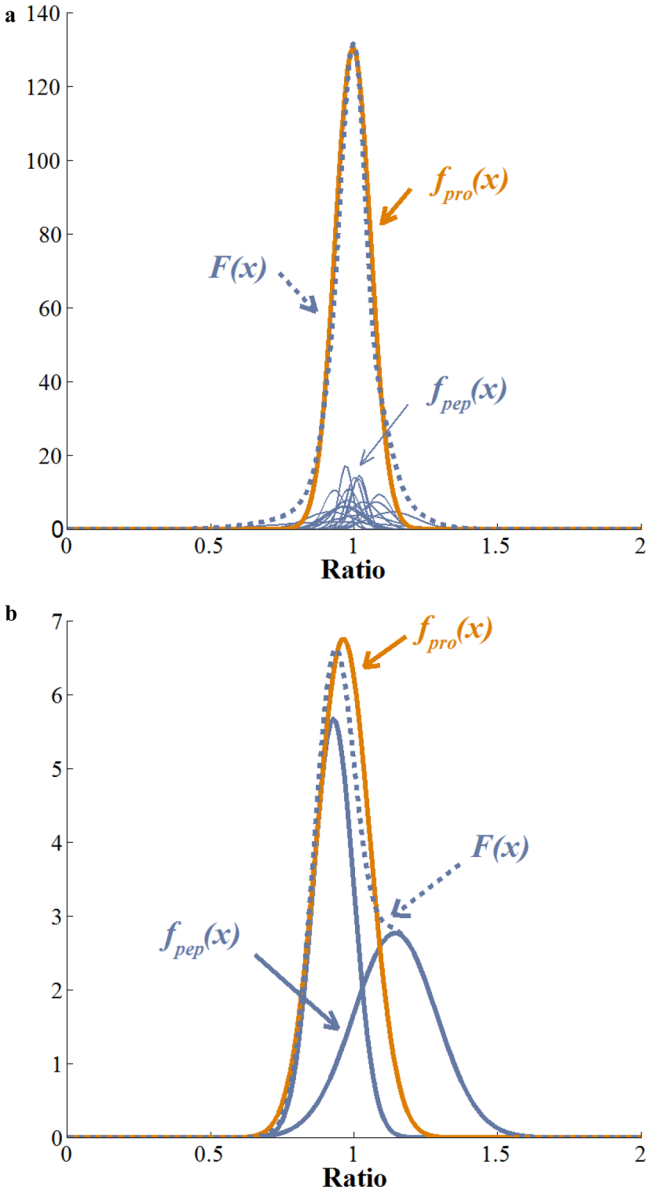


Figure 3. Quantitation and evaluation of protein ratios. (a) Protein IPI00909239.1 has 20 quantified peptides and their quantitation ratios are all relatively accurate, concentrating at 1:1. In this case, the calculated protein ratios are quite close using three different methods (0.99 by the kernel density estimation, 0.98 by taking the median, and 0.99 by averaging). (b) Protein IPI:IPI00909763.1 has only 2 quantified peptides with different ratios, 0.93 and 1.14. The normalized CIs of these two ratios are [77%, 123%] and [62%, 138%], respectively. Without peptide accuracy evaluation, it is difficult to determine which peptide ratio is more reliable in estimating the protein ratio. Using kernel density estimation, the calculated protein ratio is 0.96, close to the peptide ratio with a smaller CI, and the normalized CI of the protein ratio is [74%, 126%]. In both panels, the solid blue, dashed blue, and solid orange traces correspond to $f_{pep}(x)$, F, and $f_{pro}(x)$ described in the Experimental Section.

2.93% of all ratios in the pQuant results, much less compared to those in the Census results (0.85−10.7%).

Census uses the regression model $\mathbf{Y} = a\mathbf{X} + b + e$ to calculate ratios and output NaN if $a \leq 0$. pQuant uses the regression model $\mathbf{Y} = a\mathbf{X} + e$ to calculate ratios and output NaN if $a \leq 0$. When either the light or the heavy peptide has extremely low or no signal in MS scans, that is, the **X** or **Y** values are primarily zeros, an output of NaN is reasonable (Supporting Information Figure 3). However, for about 90% of the Census NaN ratios, there are obvious signals for both the light peptide and the heavy peptide in the MS scans. In these cases, interference signals are to blame (Supporting Information Figure 4). We also examined all the NaN ratios reported by pQuant and found that they all corresponded to low abundance peptides, that is, either the light or the heavy peptide or both have extremely low or zero signal in MS scans.

After the removal of the NaN ratios from the quantitation results, the medians of the ratios given by pQuant and Census are almost the same. However, the standard deviations of the Census ratios are on average twice as big as their pQuant counterparts. This is attributed to the outlier ratios in the Census results (an example in Figure 1). For the 1:1 SILAC data, there are a total of 175 Census ratios larger than 4 or smaller than 0.25. In comparison, only eight such outlier ratios are produced by pQuant.

The distributions of the peptide ratios computed by pQuant and Census from 14 data sets are displayed in Supporting Information Figures 5 and 6. The Census results for the 1:10 (H:L) data sets have the most NaN ratios and the largest standard deviations (Tables 1 and 2). Our investigation found that again the Census regression model ($\mathbf{Y} = a\mathbf{X} + b + e$) is to blame. It happens that **Y** represents the light peptide and **X** represents the heavy peptide. Generally, b is negligible, but as illustrated in Supporting Information Figure 7, b can be a big value, which causes a to be much smaller than it should be. This is the most damaging when a is expected to be as large as 10. Cornbleet et al. addressed this problem mathematically[36] and a simple solution is exchanging the values of **Y** and **X** in the regression model and calculating the inverse ratio.

Census results are affected by interference signals because the intensities of all isotopic peaks of a light peptide are summed, and so are those of the heavy peptide. This strategy has the benefit of being able to tolerate incomplete heavy isotope labeling, because incomplete labeling hardly affects the summed intensity of the entire isotopic envelope, although it greatly alters the intensity of individual isotopic peaks.[15,37] Without interference from coeluting ions of similar m/z values, summing the intensities of all isotopic peaks is perfectly fine, but if there is interference, the interference signal would be summed into the peptide signal and distort the quantitation result. In pQuant, the enrichment value of the heavy isotope label is calculated for each peptide. This value and the experimental intensity of the least interfered isotopic peak are used to simulate the intensities of other isotopic peaks. Thus, interference is minimized and accuracy is improved.

Comparison of pQuant and MaxQuant. Given that MaxQuant is incompatible with $^{14}N/^{15}N$ data, we only compared pQuant to MaxQuant on the SILAC data (Table 3).

NaN accounted for 1.83−2.74% of the peptide quantitation results by MaxQuant and only 0.01−0.48% of those by pQuant. It is difficult to trace why MaxQuant outputs a NaN ratio, but these NaN ratios are marked ISO or MSMS, meaning that either the light or the heavy peptide has extremely low or no signal in the MS scans. Nonetheless, for about 80% of MaxQuant NaN ratios there are obvious signals of both the light peptide and the heavy peptide in the MS scans. So, these NaN ratios are likely caused by interference. Supporting Information Figure 8 illustrates one such example.

We examined all the NaN ratios in the pQuant results and found that they all corresponded to low abundance peptides. Furthermore, we did not find a single pQuant NaN ratio that has

dx.doi.org/10.1021/ac404246w | Anal. Chem. 2014, 86, 5286−5294

Table 2. Comparison of the pQuant and Census Quantitation Results on the ^{15}N-Labeling Data[a]

			peptide level[b]					protein group level[c]			
				NaN ratios		numerical ratios				numerical ratios	
sample-mixing ratio (H/L)	expected $\log_2$(ratio)	quant. software	no. all ratios	no.	%	median	S.D.	no. all ratios	no. NaN ratios	median	S.D.
1:1	0.000	pQuant	3274	12	0.37	−0.12	0.46	401	0	−0.12	0.50
		Census		28	0.85	−0.03	0.72		0	−0.01	0.62
2:3	−0.585	pQuant	3183	13	0.41	−0.69	0.47	426	0	−0.69	0.52
		Census		38	1.19	−0.58	0.86		0	−0.51	0.65
3:2	0.585	pQuant	3318	10	0.30	0.38	0.54	386	0	0.40	0.54
		Census		59	1.78	0.52	0.75		0	0.56	0.58
1:3	−1.585	pQuant	2711	16	0.59	−1.69	0.77	374	0	−1.69	0.56
		Census		71	2.62	−1.56	1.14		1	−1.43	0.76
3:1	1.585	pQuant	2490	19	0.76	1.48	0.52	382	0	1.48	0.58
		Census		61	2.45	1.63	0.73		1	1.62	0.60
1:10	−3.32	pQuant	2452	61	2.49	−3.47	0.78	473	0	−3.64	0.69
		Census		246	10.0	−3.18	1.77		4	−2.64	1.14
10:1	3.32	pQuant	2666	78	2.93	3.16	0.69	447	5	3.16	0.63
		Census		180	6.75	3.47	1.03		5	3.38	0.77

[a]Ratios in this table are all $\log_2$-transformed. [b]In quantification of each peptide, only the PSM with the highest identification score was used as the starting point to reconstruct chromatograms. [c]To guarantee a fair comparison, we calculate the protein ratios at the group level by taking the median of the peptide ratios.

Table 3. Comparison of the pQuant and MaxQuant Quantitation Results on the SILAC Data[a]

			evidence level[b]					protein group level[c]			
				NaN ratios		numerical ratios				numerical ratios	
sample-mixing ratio (H/L)	expected $\log_2$(ratio)	quant. software	no. all ratios	no.	%	median	S.D.	no. all ratios	no. NaN ratios	median	S.D.
1:1	0.00	pQuant	6408	1	0.01	0.01	0.27	1402	1	0.01	0.21
		MaxQuant		153	2.39	0.01	0.34		6	0.03	0.26
1:2	−1.00	pQuant	7010	6	0.08	−0.86	0.31	1549	1	−0.86	0.28
		MaxQuant		137	1.95	−0.89	0.39		1	−0.86	0.36
2:1	1.00	pQuant	7495	12	0.16	1.05	0.35	1655	1	1.06	0.34
		MaxQuant		137	1.83	1.06	0.46		2	1.09	0.39
1:5	−2.32	pQuant	7723	20	0.26	−2.18	0.45	1697	4	−2.18	0.41
		MaxQuant		191	2.47	−2.18	0.68		4	−2.18	0.67
5:1	2.32	pQuant	6753	22	0.33	2.44	0.66	1674	2	2.48	0.64
		MaxQuant		123	1.82	2.40	0.96		8	2.47	0.78
1:10	−3.32	pQuant	6617	32	0.48	−3.18	0.70	1596	2	−3.18	0.60
		MaxQuant		181	2.74	−3.06	1.16		6	−3.06	1.17
10:1	3.32	pQuant	6128	24	0.39	3.08	0.82	1597	3	3.11	0.74
		MaxQuant		125	2.04	2.97	1.20		4	3.04	0.99

[a]Ratios in this table are all $\log_2$-transformed. [b]MaxQuant uses "evidence" to indicate a SILAC pairs in MS. A peptide sequence may have more than one evidence, and each evidence corresponds to an individual ratio. [c]To guarantee a fair comparison, we calculate the protein ratios in group level by taking the median of the peptide ratios.

a Census or MaxQuant ratio close to the sample-mixing ratio. In a deliberate test we let pQuant quantify falsely matched peptides (peptides from the reversed protein sequences) on the 1:1 ^{14}N/^{15}N data set, and 97.7% of them are quantified as NaN by pQuant. In comparison, for peptides identified with high confidence, only 0.01−2.93% of them have NaN ratios by pQuant (Tables 1−3).

After the removal of the NaN ratios, the values of most ratios calculated by pQuant and MaxQuant are almost the same except for a few seriously interfered peptides. Supporting Information Figure 9 illustrates an outlier ratio calculated by MaxQuant. Because of these outlier ratios, the standard deviations of the MaxQuant ratios are on average 30% larger than those of the pQuant ratios. The medians of the peptide or protein ratios calculated by the two software tools are all close to the sample-mixing ratios. For example, among the pQuant ratios from the 1:1 SILAC data, 98.9% are in the interval [0.5, 2], and the middle 50% (between the upper and lower quartiles) are in the interval [0.91, 1.10]. This result shows that most of the H/L ratios are close to the expected ratio of 1:1. As the sample-mixing ratio moves away from 1:1, the fraction of ratios that are well off the expected value increases, no matter which software tool is used (Supporting Information Figures 6 and 10). This phenomenon was reported previously.[38] Supporting Information Figure 10 illustrates the distributions of the ratios from all samples.

As shown above, MaxQuant is meticulously refined. pQuant and MaxQuant have two techniques in common. First, the isotopic chromatograms in pQuant are similar to the "3D peaks" in MaxQuant. Second, both MaxQuant and pQuant use the regression model $\mathbf{Y} = a\mathbf{X} + e$ to calculate peptide ratios. These

dx.doi.org/10.1021/ac404246w | Anal. Chem. 2014, 86, 5286−5294

two techniques greatly improve quantitation accuracy. The most important technical improvement of pQuant over MaxQuant is that pQuant selects two isotopic chromatograms with the least amount of interference, one for the light peptide and one for the heavy peptide, to calculate the H/L ratios. MaxQuant does not make such distinction and may include interference signals into ratio calculation. Furthermore, MaxQuant does not evaluate the accuracy of peptide ratios or protein ratios. As illustrated in Supporting Information Figure 8, it is difficult to discern which signals are used for ratio calculation in MaxQuant.

Protein Ratios Suggesting Large Abundance Change. Most quantitative experiments aim to identify proteins whose abundance levels differ significantly between two conditions. For this purpose, statistical methods such as outlier testing,[39] Student's t test,[14] Bayes analysis,[40] analysis of variance (ANOVA),[41] Bonferroni family-wise Type I error rate-controlling method,[42,43] Benjamini–Hochberg FDR-controlling method,[44] and q-value methods[45,46] have been adopted. Invariably, they all start from protein ratios computed by quantitation software tools. The presence of inaccurate protein abundance ratios and a lack of accuracy evaluation to keep them out could distort the results of these statistical methods. Suppose that protein ratios in a proteome-wide quantification experiment follow a log-normal distribution,[14] quantitation errors may lead to a false positive result (a protein wrongly thought of to have a large abundance change) by shifting a ratio away from the center to either tail, or lead to a false negative result by affecting a ratio in the opposite direction.

Accurate protein ratios require accurate peptide ratios, therefore inaccurate peptide ratios are usually detected using an outlier test and discarded. However, this is not always effective, even for proteins with multiple peptide measurements.[8] Besides, a large number of proteins (e.g., 30% of the proteins in our SILAC data set) have only one or two peptides quantified, so inaccurate ratios cannot be detected using an outlier testing method. Furthermore, the accuracy of protein ratios is hardly ever evaluated.

As illustrated in Tables 1–3, NaN protein ratios are rare in pQuant results, and the standard deviations of protein ratios calculated by pQuant are smaller than those by Census and MaxQuant. In the Census and MaxQuant results, there are more NaN or outlier ratios for proteins, suggesting significant abundance changes. However, after manual examination of the original data, we found that most of them are questionable. Census relies on the determinant score (R^2) as a surrogate for accuracy evaluation of peptide ratios. If only the peptide ratios with high R^2 (>0.9, for example) are used to calculate protein ratios, the proportion of incorrect protein ratios will decrease but the total number of quantified proteins will also decrease and a subset of proteins that truly have large abundance changes will be lost.

Lastly, it is worth emphasizing that good identification results are essential for quantitation. A protein ratio will be more accurate if it has more correctly identified peptides.[47] Meanwhile, falsely identified proteins (or peptides) will have NaN or outlier ratios, and these proteins may be falsely reported as having significant abundance changes. Therefore, excellence in data analysis for quantitative proteomics requires perfection in both the identification and the quantitation software tools.

CONCLUSIONS

Minimizing the interference of coeluting ions of similar m/z values is one of the major advantages of pQuant. This interference problem is a very serious one and the most difficult to solve in quantitative proteomics. Coeluting interference signals frequently cause a large deviation (10-fold or more) in relative protein abundance ratios. Other problems such as arginine-to-proline conversion in SILAC samples,[48] incomplete labeling,[49] retention time shifts between labeled and unlabeled peptides, and unequal total protein input of light and heavy samples, normally introduce a deviation of 1.5-fold or less, and they can be avoided to a large extent by doing experiments carefully. In our data sets, the arginine-to-proline conversion was hardly detectable, the efficiency of ^{15}N-labeling was about 99%, pairs of light and heavy peptides coeluted nearly perfectly, and the vast majority of the experimental ratios were close to the expected values. In contrast, the interference of coeluting ions of similar m/z was pervasive in all our data, and it is also common for label-free quantitation[11] as well as MS/MS based quantitation (e.g., iTRAQ).[10,50] Before the LC-MS technology is sophisticated enough to preclude this problem, computationally detecting and keeping out interference are critical in all quantitation experiments.

The other advantage of pQuant is accuracy evaluation. Although accuracy evaluation should be an integral part of quantitation, it was largely ignored in previous software-development efforts for quantitative proteomics. pQuant evaluates both peptide and protein ratios. Furthermore, pQuant provides normalized CI rather than a score, because normalized CI intuitively and uniformly reports the accuracy for each peptide or protein ratio.[51]

Through these improvements, more peptides and proteins are quantitated with higher accuracy using pQuant, as we show in this paper with a total of 14 data sets. pQuant supports different full MS scan-based quantitation strategies such as SILAC and ^{15}N-labeling. Moreover, it is compatible with various MS data formats and identification software tools.

ASSOCIATED CONTENT

Supporting Information

Additional materials as described in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: dongmengqiu@nibs.ac.cn.
*E-mail: rxsun@ict.ac.cn.

Author Contributions

C.L. and C.-Q.S. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Long Wu, Kun He, and Hai-Feng Chen from the Institute of Computing Technology, Chinese Academy of Sciences, for valuable discussions, Dr. Tie-Mei Li from National Institute of Biological Sciences (NIBS), Beijing, for help with the manuscript, and Drs. Li-Ming Sun and Mo Xu, also from NIBS, Beijing, for providing the HeLa cell samples. R.-X.S. was supported by the National Key Basic Research and Development Program of China (973) under Grant No. 2012CB910602. M.-Q.D. was supported by National Scientific Instrumentation Grant Program 2011YQ09000506. S.-M.H. was supported by the National Key Basic Research and Development Program of China (973) under Grant No. 2010CB912701. This work was

also supported by the CAS Knowledge Innovation Program under Grant No. KGCX1-YW-13 and ICT-20126033; and the National Natural Science Foundation of China under Grant No. 30900262.

REFERENCES

(1) Zhao, Y.; Lee, W. N.; Xiao, G. G. *Expert Rev. Proteomics* **2009**, *6*, 115−118.
(2) Xu, M.; Long, C.; Chen, X.; Huang, C.; Chen, S.; Zhu, B. *Science* **2010**, *328*, 94−98.
(3) Dong, M. Q.; Venable, J. D.; Au, N.; Xu, T.; Park, S. K.; Cociorva, D.; Johnson, J. R.; Dillin, A.; Yates, J. R., 3rd. *Science* **2007**, *317*, 660−663.
(4) Xu, P.; Duong, D. M.; Seyfried, N. T.; Cheng, D.; Xie, Y.; Robert, J.; Rush, J.; Hochstrasser, M.; Finley, D.; Peng, J. *Cell* **2009**, *137*, 133−145.
(5) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. *Mol. Cell Proteomics* **2002**, *1*, 376−386.
(6) Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 6591−6596.
(7) Stewart, I. I.; Thomson, T.; Figeys, D. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 2456−2465.
(8) Carrillo, B.; Yanofsky, C.; Laboissiere, S.; Nadon, R.; Kearney, R. E. *Bioinformatics* **2010**, *26*, 98−103.
(9) Yuan, Z. F.; Liu, C.; Wang, H. P.; Sun, R. X.; Fu, Y.; Zhang, J. F.; Wang, L. H.; Chi, H.; Li, Y.; Xiu, L. Y.; Wang, W. P.; He, S. M. *Proteomics* **2011**, *12*, 226−235.
(10) Ting, L.; Rad, R.; Gygi, S. P.; Haas, W. *Nat. Methods* **2011**, *8*, 937−940.
(11) Yang, C.; Yu, W. *J. Proteome Res.* **2010**, *9*, 2705−2712.
(12) Park, S. K.; Venable, J. D.; Xu, T.; Yates, J. R., 3rd. *Nat. Methods* **2008**, *5*, 319−322.
(13) Cox, J.; Mann, M. *Nat. Biotechnol.* **2008**, *26*, 1367−1372.
(14) Ting, L.; Cowley, M. J.; Hoon, S. L.; Guilhaus, M.; Raftery, M. J.; Cavicchioli, R. *Mol. Cell Proteomics* **2009**, *8*, 2227−2242.
(15) MacCoss, M. J.; Wu, C. C.; Liu, H.; Sadygov, R.; Yates, J. R., 3rd. *Anal. Chem.* **2003**, *75*, 6912−6921.
(16) Washburn, M. P.; Ulaszek, R.; Deciu, C.; Schieltz, D. M.; Yates, J. R., 3rd. *Anal. Chem.* **2002**, *74*, 1650−1657.
(17) Xu, Tao; V, J. D.; Kyu Park, S.; Cociorva, D.; Lu, B.; Liao, L.; Wohlschlegel, J.; Hewel, J.; Yates, J. R., III. *Mol. Cell Proteomics* **2006**, *5*, S174.
(18) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. *J. Proteome Res.* **2011**, *10*, 1794−1805.
(19) Fu, Y.; Yang, Q.; Sun, R.; Li, D.; Zeng, R.; Ling, C. X.; Gao, W. *Bioinformatics* **2004**, *20*, 1948−1954.
(20) Li, D.; Fu, Y.; Sun, R.; Ling, C. X.; Wei, Y.; Zhou, H.; Zeng, R.; Yang, Q.; He, S.; Gao, W. *Bioinformatics* **2005**, *21*, 3049−3050.
(21) Wang, L. H.; Li, D. Q.; Fu, Y.; Wang, H. P.; Zhang, J. F.; Yuan, Z. F.; Sun, R. X.; Zeng, R.; He, S. M.; Gao, W. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 2985−2991.
(22) Rockwood, A. L.; Haimi, P. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 415−419.
(23) MacCoss, M. J.; Wu, C. C.; Matthews, D. E.; Yates, J. R., 3rd. *Anal. Chem.* **2005**, *77*, 7646−7653.
(24) Casella, G. *Am. Stat.* **1983**, *37*, 147−152.
(25) Chambers, R. L.; Dunstan, R. *Biometrika* **1986**, *73*, 597−604.
(26) Motulsky, H.; Christopoulos, A. *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting*; Oxford University Press: Oxford, U.K., 2004.
(27) Legendre, P.; Desdevises, Y. *J. Theor. Biol.* **2009**, *259*, 727−743.
(28) Kozak, A.; Kozak, R. A. *For. Chron.* **1995**, *71*, 326−330.
(29) Strong, F. C. *LC−GC* **1992**, *10*, 828−828.
(30) Bonate, P. L. *LC−GC* **1992**, *10*, 378−379.
(31) Hocking, R. R. *Methods and Applications of Linear Models: Regression and the Analysis of Variance*; Wiley: New York, 1996.
(32) Parzen, E. *Ann. Math. Stat.* **1962**, *33*, 1065−1076.
(33) Colaert, N.; Vandekerckhove, J.; Martens, L.; Gevaert, K. *Methods Mol. Biol.* **2011**, *753*, 373−398.
(34) Filiou, M. D.; Varadarajulu, J.; Teplytska, L.; Reckow, S.; Maccarrone, G.; Turck, C. W. *Proteomics* **2012**, *12*, 3121−3128.
(35) Webhofer, C.; Zhang, Y.; Brusis, J.; Reckow, S.; Landgraf, R.; Maccarrone, G.; Turck, C. W.; Filiou, M. D. *J. Proteomics* **2013**, *88*, 27−33.
(36) Cornbleet, P. J.; Gochman, N. *Clin. Chem.* **1979**, *25*, 432−438.
(37) Venable, J. D.; Wohlschlegel, J.; McClatchy, D. B.; Park, S. K.; Yates, J. R., 3rd. *Anal. Chem.* **2007**, *79*, 3056−3064.
(38) Huang, X.; Tolmachev, A. V.; Shen, Y.; Liu, M.; Huang, L.; Zhang, Z.; Anderson, G. A.; Smith, R. D.; Chan, W. C.; Hinrichs, S. H.; Fu, K.; Ding, S. J. *J. Proteome Res.* **2011**, *10*, 1228−1237.
(39) Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B. *Anal Bioanal Chem.* **2007**, *389*, 1017−1031.
(40) Margolin, A. A.; Ong, S. E.; Schenone, M.; Gould, R.; Schreiber, S. L.; Carr, S. A.; Golub, T. R. *PLoS One* **2009**, *4*, e7454.
(41) Karpievitch, Y.; Stanley, J.; Taverner, T.; Huang, J.; Adkins, J. N.; Ansong, C.; Heffron, F.; Metz, T. O.; Qian, W. J.; Yoon, H.; Smith, R. D.; Dabney, A. R. *Bioinformatics* **2009**, *25*, 2028−2034.
(42) Holm, S. *Scand. J. Stat.* **1979**, *6*, 65−70.
(43) Perneger, T. V. *BMJ [Br. Med. J.]* **1998**, *316*, 1236−1238.
(44) Benjamini, Y.; Hochberg, Y. *J. R. Stat Soc. Ser., B* **1995**, *57*, 289−300.
(45) Storey, J. D. *Ann. Stat.* **2003**, *31*, 2013−2035.
(46) Listgarten, J.; Emili, A. *Mol. Cell Proteomics* **2005**, *4*, 419−434.
(47) Bantscheff, M.; Lemeer, S.; Savitski, M. M.; Kuster, B. *Anal Bioanal Chem.* **2012**, *404*, 939−965.
(48) Park, S. K.; Liao, L.; Kim, J. Y.; Yates, J. R., 3rd. *Nat. Methods* **2009**, *6*, 184−185.
(49) Cappadona, S.; Baker, P. R.; Cutillas, P. R.; Heck, A. J.; van Breukelen, B. *Amino Acids* **2012**, *43*, 1087−1108.
(50) Shirran, S. L.; Botting, C. H. *J. Proteomics* **2010**, *73*, 1391−1403.
(51) Pan, C.; Kora, G.; McDonald, W. H.; Tabb, D. L.; VerBerkmoes, N. C.; Hurst, G. B.; Pelletier, D. A.; Samatova, N. F.; Hettich, R. L. *Anal. Chem.* **2006**, *78*, 7121−713.