MS-EmpiRe Utilizes Peptide-level Noise Distributions for Ultra-sensitive Detection of Differentially Expressed Proteins

Authors
Constantin Ammar, Markus Gruber, Gergely Csaba, and Ralf Zimmer

In Brief
Peptide measurements from quantitative LC-MS/MS experiments need to be computationally processed to determine whether their corresponding proteins are significantly regulated between conditions. We show that this processing can have drastic influences on the number of detected regulated proteins. We present a new computational tool, MS-EmpiRe, which provides strongly increased sensitivity while retaining strict control of the FDR. MS-EmpiRe provides an easy end-to-end workflow for any type of quantitative proteomics data and is available as an R package with examples and a manual.

Highlights
- New software to determine differentially expressed proteins in quantitative MS experiments.
- Applicable to any modern quantitative MS setup and study type, including clinical setups.
- Ultra-sensitive at strict FDR control, up to 1000 additional proteins in a single comparison.
- Easy to use, fast processing and readily available package, including user friendly manual.
Mass spectrometry based proteomics is the method of choice for quantifying genome-wide differential changes of protein expression in a wide range of biological and biomedical applications. Protein expression changes need to be reliably derived from many measured peptide intensities and their corresponding peptide fold changes. These peptide fold changes vary considerably for a given protein. Numerous instrumental setups aim to reduce this variability, whereas current computational methods only implicitly account for this problem. We introduce a new method, MS-EmpiRe, which explicitly accounts for the noise underlying peptide fold changes. We derive data set-specific, intensity-dependent empirical error fold change distributions, which are used for individual weighing of peptide fold changes to detect differentially expressed proteins (DEPs).

In a recently published proteome-wide benchmarking data set, MS-EmpiRe doubles the number of correctly identified DEPs at an estimated FDR cutoff compared with state-of-the-art tools. We additionally confirm the superior performance of MS-EmpiRe on simulated data. MS-EmpiRe requires only peptide intensities mapped to proteins and, thus, can be applied to any common quantitative proteomics setup. We apply our method to diverse MS data sets and observe consistent increases in sensitivity with more than 1000 additional significant proteins in deep data sets, including a clinical study over multiple patients. At the same time, we observe that even the proteins classified as most insignificant by other methods but significant by MS-EmpiRe show very clear regulation on the peptide intensity level. MS-EmpiRe provides rapid processing (< 2 min for 6 LC-MS/MS runs (3 h gradients)) and is publicly available under github.com/zimmerlab/MS-EmpiRe with a manual including examples. Molecular & Cellular Proteomics 18: 1880–1892, 2019. DOI: 10.1074/mcp.RA119.001509.
Differential Quantification with MS-EmpiRe

| Table I | The two state-of-the-art differential quantification methods MaxLFQ (with t-test) and MSqRob compared against MS-EmpiRe |
|---------|---------------------------------------------------------------------------------------------------------------|
| **Normalization** | Median/Mode of peptide fold change distributions (only replicate samples) | Minimization of peptide fold changes (all samples) | Estimation of sample specific contribution via linear model |
| **Protein fold change estimation** | Maximum likelihood estimation from peptide fold change distribution | Fold change of MaxLFQ protein intensities | Estimation of condition specific contributions via linear model |
| **Statistical test** | Modified Stouffer’s method | t-test | Moderated t-test |
| **Test statistic** | Sum of Z-transformed, background normalized peptide fold changes | Difference of MaxLFQ protein intensities | Estimation of condition specific contributions via linear model |
| **Variance estimation** | Empirical and Replicate based peptide distributions | Protein-level sample variance | Empirical Bayes Estimation |
| **Outlier detection** | Signal scoring based on outlier distribution, downscaling of outlier peptides | Median of peptide fold changes | M-Estimation with Huber weights |

A common computational approach for differential expression analysis is to derive protein level fold changes from the peptide fold changes and to apply statistical tests such as the t-test to assign a significance value to it. This protein level approach is for example implemented in the Perseus pipeline (10). Peptide level models have been proposed (11, 12) and have recently been shown to offer superior performance compared with protein level approaches (13, 14, 15). A recent implementation is given in the MSqRob package (14). The majority of peptide level models are based on linear regression, which can be problematic for data with strong distortion, outliers, or small peptide numbers. We propose an orthogonal approach, consisting of the direct assessment of peptide level noise, which we term **Mass Spectrometry analysis using Empirical and Replicate based statistics (MS-EmpiRe)**. We introduce empirically generated, intensity dependent error fold change distributions and utilize this for between-sample normalization and to derive differential expression probabilities for each peptide. We then show that these probabilities can be combined to the protein level via a modification to Stouffer’s method (16). The data for MS-EmpiRe can be measured with a variety of quantitative proteomics setups, as we need only peptide intensities grouped to proteins as input. We test the performance of the method on a recently published proteome wide benchmarking data set of O’Connell, Gygi and coworkers (17), containing LFQ as well as TMT-MS3 data. With MS-EmpiRe we observe up to 121% more sensitivity compared with the approach reported in O’Connell et al. We additionally compare our approach with the peptide level tool MSqRob and see similar performance increases. On simulated differential expression changes, we see similar performance results as on the benchmarking set and we additionally demonstrate MS-EmpiRe’s superior classification abilities on state-of-the-art proteomics data sets. MS-EmpiRe is available as an R package on GitHub (github.com/zimmerlab/MS-EmpiRe).

**Materials and Methods**

**MS-EmpiRe Compared with State-of-the-Art Approaches**—We compare our method with two current methods, MaxLFQ with Perseus (8, 10) and MSqRob (13, 14), which have different strategies of solving the challenges associated with MS based protein quantification (Table I). For the following descriptive comparison, we focus on the principle issues of differential quantification: Normalization between different experimental samples, the estimation of the protein fold change, the statistical test applied, derivation of the corresponding test statistic (which represents the protein level change), estimation of the variance parameter(s) of the statistical test and outlier correction. In MaxLFQ, normalization is carried out by minimizing the sum of peptide level fold changes with run specific normalization factors. The sum is taken over all runs, also between conditions, with the underlying assumption that most of the proteins do not change. Several statistical tests can be applied to MaxLFQ data, where the...
of RNA sequencing data, which are subject to sample specific effects. In the context of proteomics, the total amount of proteins that is processed per run has a significant effect on the signal measured per peptide. Therefore, raw signals originating from two different samples are rarely comparable. To correct for sample-specific effects, all signals of a sample are typically multiplied by a scaling factor. The variance estimation is hence carried out via these empirical, data set-specific and intensity dependent background distributions. This allows a refined weighing of the influence of peptide noise on a given fold change. Outlier detection is carried out by applying a system of linear equations to identify scaling factors such that the change of peptide signals between any two samples (and fractions) is minimized.

All previously mentioned normalization procedures rely on the assumption that most of the signals do not change between any two samples, and because of that they try to minimize the change of all peptide signals. We use a similar approach, but only for samples from the same experimental condition (i.e., replicate samples, which should indeed measure the same peptide values) and use a different factor to normalize between conditions:

Normalization within a condition in MS-EmpiRe is done by single linkage clustering as described in Fig. 1. Each cluster contains either one or multiple samples. We start with as many clusters as we have replicates and successively merge the two most similar clusters until we end up with one cluster that contains all samples. Similarity between any two clusters is defined as follows: For two clusters, we compute a fold change distribution. We build every possible sample pair between the two clusters and compute the fold change for every peptide that was detected in both samples of a pair. The variance of this distribution is used to determine the similarity between clusters whereas the median is used as an estimate for a systematic signal shift. To merge two clusters c₁ and c₂, we scale all signals of samples in c₁ from c₂ by the median. This step yields a new cluster that contains all samples from c₁ and c₂ and in which all samples are shifted onto each other. Single linkage clustering is applied to each condition separately. Samples from two different conditions are then shifted in a similar way, the difference is the selection of the shift parameter. Because we can no longer assume that peptides do not change, except for experimental noise, we propose to use the most probable fold change from the distribution instead of the median. This choice is similar to the idea of centralization proposed by Zien et al. (22). Instead of enforcing a minimal change between all peptides, this shift only targets most
peptides and is still in accordance with the assumption that most proteins do not change. The shift parameter can also be defined by the user using prior knowledge for certain subsets of proteins for example.

After the normalization, signals for the same peptide between any two samples are comparable.

Empirical Error Distributions—Our goal is to detect DEPs between different conditions. However, only peptide level measurements are available from current standard MS experiments and protein level changes have to be inferred. We argue that each peptide level change should be assessed in context of the noise associated with the measurement. MS-EmpiRe is therefore centered around replicate based empirical error distributions (Fig. 2A and 2B). The empirical error distribution is fully based on the data and derived as follows: We compute the log2 fold change of every peptide signal between any two replicate pairs in each condition. As the log2 fold change between replicate samples should be zero, each deviation from zero can be seen as an error. This results in one large collection of errors, approximately $C \times \frac{N(N-1)}{2} \times P$ for C conditions with N replicates each and P detected peptides. Because we observed that the variance of peptide measurements depends on signal strength (Fig. 3E) we decided to split the complete distribution into intensity dependent sub-distributions. Each of the resulting subdistributions contains just a subset of all peptide fold changes. For the construction we sort the peptides ascending to their mean signal strength. We slide a window over the sorted list of peptides to determine the relevant subset for each distribution. The window size and how far it is shifted in each step are

---

**Fig. 2. Schematic of the MS-EmpiRe workflow.** A) All identified peptides from a proteomics run are sorted by their mean intensities. B) The peptides are the split into subgroups based on their intensity. For each subgroup, the error fold changes of the individual peptides are calculated. An error fold change simply denotes the log2 fold change of a peptide between two replicate conditions. All error fold changes within a subgroup form an empirical error distribution. Distributions corresponding to lower intensity peptides show a larger variance than for high intensity peptides. C) When a protein is tested for differential expression, each peptide gets assigned an empirical error distribution. Peptides of similar intensities can get the same distribution assigned. D) For each peptide fold change, the probability that this fold change happened by chance (e.g. the p value) is assessed from the empirical distribution. This means that the same fold change will get a much lower p value when the distribution is wide as compared with when it is narrow. To make this value manageable, the p value is then transformed to a Z-value, by transferring the mass of the empirical probability distribution to a standard normal distribution. E) The Z-values for each peptide are corrected for outliers. F) The corrected Z-values can directly be summed to the protein level, and the corresponding protein-level p value can be obtained as well as the FDR after multiple testing correction.
Differential Quantification with MS-EmpiRe

parameters that can be controlled by the user. Adjusting it can increase the resolution of the subdistributions at the cost of computational time. The default window size is 10% of the total number of peptide measurements with a maximum of 1200. The size of the shift is set to 1/8 of the actual window size. Note that each peptide may appear in multiple subdistributions if the shift is smaller than the window size. To assign each peptide to only one subdistribution, we save the mean signal of the first and last element of each distribution. We then calculate the distance of the mean signal to the start and end of each distribution for every peptide. Each peptide is then assigned to a distribution such that the minimum of those two distances is maximized, that is,

$$\max(\min(\{s_{\text{start}} - s, s - s_{\text{end}}\}))$$  \hspace{1cm} (Eq. 1)

After this step we have a collection of empirical error distributions that describe the observed measurement errors in relation to signals (Fig. 2C).

Any observed peptide fold change can now be put into context of the background noise. This allows us to determine the probability of the peptide fold change under the corresponding empirical error distribution. We denote this probability as the empirical p-value (see also supplemental Figs. S1 and S2).

Merging Scores Over Replicates—We can now determine the empirical p-value for every peptide between any two samples. What we rather want, however, is the same information for whole proteins between two conditions including replicate data. This means we must express the empirical p-value in terms of a score that we can combine over replicates as well as peptides. Further, the score should be able to distinguish between negative and positive fold changes. This way we can identify groups of peptides that consistently show the same direction of change between multiple replicate pairs. One score fulfilling these criteria is the Z-value, that is, a score that follows a standard normal distribution. We can transform an observed fold change into the corresponding Z-value as follows:

$$Z_p = \phi^{-1}(\rho_{\text{emp}})$$  \hspace{1cm} (Eq. 2)

where $\phi^{-1}$ is the inverse of the cumulative distribution function of the standard normal distribution and $\rho_{\text{emp}}$ is the empirical p value. This is analogous to Stouffer’s method (16) for combined probability tests.

This means we can transform any empirical error distribution to a standard normal distribution (Fig. 2D). In the following sections we will show how those Z-values can be transformed to joint probabilities over replicate data as well as multiple replicates.

To distinguish between background noise and signals, usually not only 2 samples, but N versus M replicate measurements from two different conditions are compared. Those yield up to $N \times M$ scores per peptide which are merged to make a protein-level statement between the two conditions. Under the null hypothesis of no change, each of the $N \times M$ Z-values follows a standard normal distribution. Under this assumption, we can simply compute the sum of the $N \times M$ standard normally distributed Z-values which follows a normal distribution as well. Looking at the sum is reasonable for the following reasons: It should become extreme only if we have multiple measurements that consistently deviate in the same direction. Too few too weak deviations are canceled out by the nondeviating measurements. The same is true for strong deviations in different directions. The mean of the resulting normal distribution is zero as it is the sum over the individual means. In general, the variance of a random variable that is the sum over multiple random variables is the sum over the full covariance matrix of the variables, that is,

$$\var\left(\sum_{j=0}^{N \times M} X_j\right) = \sum_{i=0}^{N \times M} \sum_{j=0}^{N \times M} \cov(X_i, X_j)$$  \hspace{1cm} (Eq. 3)

The means and variances are known for each of the variables because they follow a standard normal distribution. We are also able to compute the covariances for the dependent variables. This is necessary because some of the possible sample comparisons are not independent, any two sample pairs that share either the first or second sample. It can be shown that the covariance of two Z-value random variables that share one of the samples is 0.5 (supplementary Material, Section 1 and supplemental Fig. S3).

For each peptide, we can now assess the unexpectedness under the previously derived background distribution over all sample pairs.

Correcting for Outlier Measurements—One problem about the sum described in the previous section, is that it is susceptible to single outlier measurements. A single extreme Z-value can be enough to make the resulting sum significant. This is because of the null hypothesis that each of the sample pair comparisons must not be differential. We therefore introduce a correction to estimate the probability, that a single outlier shifts the distribution toward higher values (Fig. 2E). For this correction, we estimate the Z-value of the peptide when it is not regulated ($Z_{\text{normed}}$) and subtract it from the original Z-value ($Z_{\text{orig}}$). $Z_{\text{normed}}$ is estimated as follows: We compute all possible fold changes of the peptide between two conditions (replicate 1 versus replicate 1, replicate 1 versus replicate 2, etc.). This results in a (very small) fold change distribution. Analogous to section 3.2, we use the median of this distribution as a scaling factor and shift all signals of the second condition by the median. This minimizes the difference of signals between the two conditions and simulates a nonregulated peptide. We again compute the summed Z-value for those shifted peptides, that is, $Z_{\text{normed}}$. If the peptide measurements were differentially regulated before the shift, $Z_{\text{normed}}$ would be less extreme than $Z_{\text{orig}}$. If the shift does not change the signal, $Z_{\text{orig}}$ and $Z_{\text{normed}}$ are the same. We can hence introduce a new value $Z_{\text{cor}} = Z_{\text{orig}} - Z_{\text{normed}}$, which denotes the difference between a regulated and a nonregulated shift. We now want to use the distribution of $Z_{\text{cor}}$ to estimate, how unlikely an observed $Z_{\text{cor}}$ value is. The higher abs($Z_{\text{cor}}$) is, the more extreme the original measurement was.

However there exists no closed form for the distribution of $Z_{\text{cor}}$. We therefore sample such a distribution by simulating a set of nondifferential measurements. For each simulated measurement, we compute $Z_{\text{cor}}$. Like Eq. 2, we look up the cumulative probability of a measured $Z_{\text{cor}}$ in the simulated distribution of $Z_{\text{cor}}$ and transform its empirical p value to a Z-value. Whether this correction is needed may depend on the data. The null distribution of the corrected score is a standard normal distribution.

Correcting for Outlier Peptides—The previous section shows how to correct for outlier measurements of single intensities. Peptides that show such a more extreme change over many samples than the remaining peptides of the same protein must be accounted for separately. To detect outlier peptides, we compute fold changes for every peptide and sample pair of the same protein. If the median of a single peptide is more extreme than the 25%/75% quantile of the whole protein distribution and protein and peptide are shifted in the same direction, it is marked as an outlier.

To compute $Z_{\text{normed}}$, for an outlier peptide, we shift the signals by the 25%/75% quantile of the protein distribution instead of the median of the peptide distribution. This modified shift results in a less extreme $Z_{\text{cor}}$ for the peptide.

Combining the Peptide Scores—What we have so far is a Z-value that expresses how likely it is for a peptide fold change over all possible replicate comparisons to occur by chance. Each of those Z-values follows a standard normal distribution.

Like the first step of merging the peptide scores over all replicate pairs, we can join these scores for all peptides from the same protein (Fig. 2F). In contrast to measurements for the same peptide from different sample pairs, peptide measurements can be regarded as
independent measurements for the same protein. This means that under the null hypothesis that every peptide score is a standard normal distributed variable, the sum of such peptide scores is distributed

$$P \sum_{i=0}^P Z_i \sim N(0,P)$$  \hspace{1cm} (Eq. 4)

with P being the number of different peptides mapped to a certain protein. Using this sum of peptide scores we can now express the probability of a protein under the null hypothesis of no change while taking into account all replicate measurements. To correct for multiple testing, we finally apply the Bonferroni-Hochberg false discovery adjustment.

Reprocessing of the Proteome Wide Benchmarking Data Set—We downloaded the raw data of the study of O’Connell et al. (17) from the PRIDE repository PXD007683 and processed the TMT as well as the LFQ data set with MaxQuant (23) version 1.6.0.16 with standard settings and the respective quantification set (11 plex TMT-MS3 and LFQ). The LFQ data consisted of 11 single-shot runs and for the TMT data 10 runs corresponding to 10 fractions were available. The mapping of the raw files is available in supplemental Table S1. Each set contained three conditions: 10% yeast spike-in, 5% yeast spike-in, 3.3% yeast spike-in. For 10% yeast spike-in, three replicates were measured, for 5% and 3.3%, four replicates were measured. The data sets were searched against a combined yeast (7904 entries) and human (20,317 entries) database downloaded from Uniprot (April, 2018, reviewed). Specific digestion was set for trypsin with two missed cleavages allowed. Carbamidomethylation of Cysteine was set as a fixed modification and Oxidation of Methionine and N termi-nus Acetylation were set as variable modifications. 20 ppm mass tolerance was set for precursor ions and 0.5 m/z was set for fragment ions. Results were filtered to 1% FDR at the peptide spectrum match (PSM) and protein level. For the LFQ data, the “match between runs” option was set (default configuration). We compared the number of identified proteins with the results reported from O’Connell et al. and observed only slight differences, possibly because of different databases and MaxQuant versions (−2% for TMT and −1% for LFQ for our protein). We found around 400 proteins after filtering in the MaxLFQ with Perseus setup, which agrees well with O’Connell et al. As we used MaxQuant instead of SEQUEST (24) for the TMT analysis, we had a noticeably lower identification rate (around 100 proteins less compared with the SEQUEST results presented in the main text of O’Connell et al.). This is similar to the number of MaxQuant results for TMT, reported in the supplement of O’Connell et al.

Processing of the Different Proteomics Studies—We tested and compared MS-EmpiRe on three different proteomics studies of Sharma et al. (25), Ramond et al. (26) and Ping et al. (27). We downloaded the MaxQuant processed data of Sharma et al. and Ramond et al. directly from the corresponding PRIDE repositories. For the Ping et al. data, we downloaded the raw files from the PRIDE repository and processed the TMT data analogous to the method described above.

Filtering of the Benchmarking Data Set—Among the different tools, we noticed large differences in the number of proteins that are submitted to statistical testing. MaxLFQ with Perseus showed the most conservative filtering, whereas MSQRob was most permissive. The decision whether to accept proteins with only a single quantified peptide value had the most impact on the filtering. With only one peptide per protein, a misidentified peptide can immediately lead to a false classification. As in MS-EmpiRe a peptide needs to be consistently quantified over multiple replicates to gain significance, the probability for such an event decreases and we hence decided to use a less conservative filtering of only one peptide per protein. We also compared the one-peptide with the two-peptide approach and observed no significant effects on the FDR (see supplemental Fig. S4).

RESULTS AND DISCUSSION

Fold Change Based Normalization Reveals Structure of the Benchmarking Data Set—We used the benchmarking data set from the study of O’Connell et al. (17), where yeast is spiked into human cell lysate at different concentrations (10%, 5%, and 3.3%) (Fig. 3A). Hence, when comparing the abundance of yeast proteins e.g. in the 10% sample with the 5% sample, one expects a fold change of 2 for each yeast protein. The two other combinations 5% versus 3.3 and 10% versus 3.3% give a fold change of 1.5 and 3, respectively. When applying a differential quantification algorithm, the changing proteins are known (e.g. the yeast proteins) and, thus, measures like specificity and sensitivity can be assessed. The samples were measured twice, once using a TMT-MS3 approach and once using label free quantification (LFQ). To visualize the normalization procedure employed by MS-EmpiRe, we used the LFQ data set. Between every sample pair, we calculated the log2 fold change for each individual peptide. This resulted in a distribution of fold changes for each sample pair, which was either a between replicate, or a between sample distribution. For the between replicate distribution we would only expect deviations from zero because of measurement errors or biological variation. Therefore, we call this distribution the empirical error distribution. For the between sample distribution,
we would only expect systematic deviations from zero for the regulated proteins. In Fig. 3B, the between sample distributions are displayed before normalization. For clarity, human proteins (which should not change at all) and yeast proteins (which have systematic changes applied to them) are displayed separately. We can already see some trends in the distributions, which underlines the fact that the fold change-based view is an intuitive measure for quantitative data sets. When applying subsequent between-replicate and between-sample normalization, as described in the methods section, we obtain the visibly clustered distributions displayed in Fig. 3C. The human peptides (around 90% of peptides) are not shifted and distributed around 0. The yeast proteins are aligned around the shift that was experimentally applied to them (i.e., the log₂ transformation of 1.5, 2, or 3). If too much or too little yeast had been applied to one of the samples, this would reflect in a stronger deviation in a subset of the replicate distributions. This is not the case in our data set and we see—with slight deviations—an alignment around the desired value (dashed lines). In a real life example, we would not expect systematic changes around one fold change in one direction, but larger spread deviations in both directions. The example, however, visualizes that a fold change-based approach on quantitative data sets is an effective procedure to normalize data sets without altering the structure of the underlying data. In general, the distributions reveal a ubiquitous problem in MS based proteomics data. The data is so noisy, that a lot of the measured yeast peptides do not even show regulation. This is most striking for the 1.5 set, where around one quarter of the peptides show no regulation, or even regulation into the wrong direction. Hence there is no way to classify these peptides correctly by themselves. Because usually, multiple replicates and peptides exist for a protein, the quantification of a protein can be seen as multiple drawings from such a distribution. This underlines that peptide fold changes should always be analyzed in the context of the data set specific noise.

Fig. 3. Experimental setup and fold change based metrics. A, Benchmarking setup for quantitative assessment of fold changes taken from O’Connell et al. (17). Different amounts of yeast lysate are spiked into human cell lysate. The three groups contain 10%, 5% and 3.3% yeast lysate, respectively. B, Peptide level fold change distribution between all conditions, before normalization. C, The distribution after fold change based normalization. D, Intensity dependent peptide fold changes between two replicates of the LFQ data (error fold changes), displayed as smoothed density scatter plot. E, Error fold changes for 10 intensity regions displayed as box plots. Each box contains the same number of data points. The quantiles correspond to the fractions 0.05, 0.15, 0.50, 0.85, 0.95.
Assessment of Empirical Error Distributions Underlines the Importance of Context Dependent Fold Changes—As already described in the methods section, one of the key features of the MS-EmpiRe algorithm is the quantitative assessment (and subsequent utilization) of fold change consistency. For this we make use of the fact that the log fold change between replicate measurements should be 0, which allows us to derive empirical error distributions containing the fold changes of peptides between replicate measurements. These distributions have already been discussed in the previous section and the human peptides displayed in Fig. 3C are a good example. However, when looking at the empirical error distributions over a whole data set, as in the previous section, we neglect the well-known fact that low-intensity peptides are subject to significantly more variation than high intensity peptides. An intuitive way to visualize this is by plotting the error fold changes against the mean intensities of the peptides, as displayed in Fig. 3D. In this density scatter plot, we see that most peptides are actually of low intensity and in the low intensity region a large spread of the fold changes is visible. Furthermore, we see global dependence of the error fold changes on the intensity and high intensity peptides are much less prone to deviate strongly from zero. Nevertheless, outliers exist for all intensities, which underlines the need for a more quantitative assessment. This is depicted in Fig. 3E, where empirical error distributions are given for distinct intensity bins. The original fold change distribution is split into 10 boxes and each box contains equal amounts of data points. We see that the lowest intensity box is particularly noisy, with around 40% of the data above a fold change of 1.5 (log₂ fold change of around 0.6). Hence, if a low intensity peptide has a 50% intensity increase from one condition to the other, the probability that it is an irrelevant change is at around 40%. When we consider the highest intensity box, this value drops from 40% to around 5%. Identical fold changes hence have a very different meaning, depending on the context. The MS-EmpiRe approach transforms an observed fold change in the intensity region a large spread of the fold changes is visible. Furthermore, we see global dependence of the error fold changes on the intensity and high intensity peptides are much less prone to deviate strongly from zero. Nevertheless, outliers exist for all intensities, which underlines the need for a more quantitative assessment. This is depicted in Fig. 3E, where empirical error distributions are given for distinct intensity bins. The original fold change distribution is split into 10 boxes and each box contains equal amounts of data points. We see that the lowest intensity box is particularly noisy, with around 40% of the data above a fold change of 1.5 (log₂ fold change of around 0.6). Hence, if a low intensity peptide has a 50% intensity increase from one condition to the other, the probability that it is an irrelevant change is at around 40%. When we consider the highest intensity box, this value drops from 40% to around 5%. Identical fold changes hence have a different meaning, depending on the context. The MS-EmpiRe approach transforms an observed fold change in the context of its corresponding empirical error distribution and, therefore, quantitatively accounts for this phenomenon. Especially, as every data set carries its own noise, and e.g. TMT-MS3 data shows significantly reduced noise (see supplemental Fig. S5) the empirical assessment for each data set is crucial.

MS-EmpiRe Shows Up To 121% Sensitivity Increase in an Experimental Benchmarking Set—The key question addressed in the setup of O’Connell et al. (17) is, how many proteins can be detected as differentially expressed in a proteome wide benchmarking setup. Analogous to their paper, we assessed how many of the experimentally shifted yeast proteins we were able to detect via MaxLFQ coupled to the Perseus pipeline. We then compared this approach with our MS-EmpiRe approach and the more recent tool MSqRob. Perseus was executed analogous to the settings given in O’Connell et al. (reverse and contaminant filtering, at least two replicate measurements per protein and two sided homoscedastic t-test with Benjamini-Hochberg correction) and MSqRob was executed with default settings. An FDR of 5% was set for all approaches. In Fig. 4A and 4B, we show the results of the benchmark for the more challenging LFQ setup. The number of proteins available for testing differs markedly, depending on how conservative the corresponding tool is in filtering peptides for quantification (see also methods section). MS-EmpiRe clearly outperforms MaxLFQ+t-test in terms of sensitivity, with up to 120% more DEP detections in the fold change 1.5 setup. When comparing MS-EmpiRe with MSqRob, it seems that MSqRob is slightly more sensitive. However, for MSqRob the observed FDR (i.e., the number of human proteins detected) is between 9% to 15% instead of the required 5%. MaxLFQ+t-test and MS-EmpiRe also violate the FDR in the fold change 2 setup, but only by 1 and 2 percentage points, respectively. To make the sensitivity analysis more comparable, we show an “FDR corrected” bar for MSqRob, where we set the FDR cutoff of MSqRob to a more stringent value, such that the actual FDR is also at around 5%. In this setup MS-EmpiRe outperforms MSqRob in terms of sensitivity in all cases and detects more than twice the number of proteins of MSqRob in the most challenging fold change 1.5 setup. As MS-EmpiRe and MaxLFQ+t-test both violated the FDR for the fold change 2 data set, we looked at the corresponding data in more detail. We saw that many of the misclassified human proteins in the data set were particularly tough cases, where many peptides over many replicate conditions show consistent up- or down-regulation (see supplemental Table S2). Of course, cases like this are included in the FDR estimation. However, the condition seems to have more cases of consistent protein up-regulation than expected. We considered “tuning” the FDR estimator to be more conservative, but took into account that the FDR violation was only mild and that neither the other benchmarking conditions, nor the simulations showed further FDR violations. As slight regulation or systematic distortions might always occur under experimental settings, we decided to leave the model unchanged. The setup shown in Fig. 4A contains the input sets after the individual filtering applied by each method. This corresponds to a real-life application of the methods, but also reduces the comparability of the classification capabilities. We hence compared each method on the same set of peptides, which consisted of the intersection of all input peptides. This led to a significant decrease in the number of proteins detected. Interestingly, the drastic reduction of input peptides strongly increased the number of detected proteins for MSqRob for the 1.5 set. This implies, that MSqRob is prone to give an over-optimistic scoring to proteins with sparse peptide evidence and hence more stringent filtering might be appropriate. In the fold change 1.5 set, MSqRob also does not violate the FDR constraint. In the two other sets, the peptide filtering does not seem to suffice to control the FDR and for MS-EmpiRe the fold change 2 set still violates the FDR.
slightly. Nevertheless, MS-EmpiRe is the most sensitive method over all sets. When comparing the methods on the less challenging TMT data set in Fig. 4, we see overall high sensitivity, which is also discussed in O'Connell et al. Also here, we see a substantial increase in sensitivity of around 20% compared with the \( t \)-test, when considering the 1.5-fold change set.

**MS-EmpiRe Identifies Up To 1200 Additional Significant Proteins in Quantitative MS Data Sets**—To test the performance of MS-EmpiRe in different experimental settings, we applied our method to public MS data sets from three different studies. The first study by Sharma et al. (25) was a deep (\( \sim 12,000 \) proteins) LFQ proteomics study of neuronal cell development. For the second data set, we chose the LFQ study of Ramond et al. (26) that was also tested in the MSqRob study (13). The data is from a single knockout experiment in Francis cellula plants (\( \sim 1000 \) proteins). The last data set by Ping et al. (27) was a deep (\( \sim 10,000 \) proteins) TMT-MS3 study of Alzheimer's disease (AD), Parkinson's disease (PD) and co-morbid (ADPD) patients. We applied MS-EmpiRe, MaxLFQ + \( t \)-test and MSqRob to the LFQ data sets and MS-EmpiRe as well as the \( t \)-test to the TMT data set. MSqRob was excluded as it currently does not support TMT data. The overall performance is better because of higher depth from sample fractionation, lower noise and fewer missing values. Quantification on the protein level hence already works well, still MS-Empire shows a significant sensitivity increase of around 19% for fold change 1.5.

**Fig. 4. Assessment of the differential detection performance on the benchmarking setup of O'Connell et al.** A, Number of proteins detected in the LFQ data by the MaxLFQ + \( t \)-test setup, MSqRob and MS-EmpiRe. The light shades show the numbers of yeast proteins accessible for testing in the setups, which differ for every method. As MSqRob shows high FDR rates (bottom plot), an FDR-corrected bar is introduced for MSqRob. MaxLFQ + \( t \)-test shows low sensitivity at good error rate control like MSqRob. MaxLFQ + \( t \)-test is very conservative for the fold change 1.5 setup, with no false positives (no bar visible). MS-EmpiRe increases the detection substantially in all cases with good error rate control. This is most pronounced in the most challenging fold change 1.5 setup. B, Number of proteins detected when using an intersected input set. Because of this conservative approach the numbers and differences are lower in general, nevertheless MS-EmpiRe is the best performing method. C, Comparison of MaxLFQ + \( t \)-test and MS-EmpiRe on a TMT data set. MSqRob was excluded as it currently does not support TMT data. The overall performance is better because of higher depth from sample fractionation, lower noise and fewer missing values. Quantification on the protein level hence already works well, still MS-Empire shows a significant sensitivity increase of around 19% for fold change 1.5.
especially in the deep data sets, we found a highly increased number of DCPs. In the Sharma et al. data set, up to 1,200 additional DCPs are identified by MS-EmpiRe. In the clinical data of Ping et al., up to 1,000 additional DCPs can be identified when using MS-EmpiRe.

Set-based Comparison Reveals Strong Differences Between Significant Proteins for Each Method—In the previous section, we have shown that the absolute numbers of DCPs are quite different for the individual methods. To check how consistent the results of the methods are, we analyzed the overlaps of the sets of DCPs. To assess these overlaps, we restricted on a set of proteins, where regulation should be very clear: we considered all proteins, where the log2 fold change estimate was larger than 1 and the FDR was below 0.05 in at least one method.

Fig. 5. Application of MS-EmpiRe, MaxLFQ+t-test and MSqRob to three different quantitative LC-MS/MS data sets. A, Number of DCPs detected in the three different data sets by the three different methods. Each bar represents the number of proteins found in the set determined by the dots below. MS-EmpiRe is the most sensitive approach. B, Overlaps of protein hits on a subset of very clear hits for DIV5 versus DIV15. MS-EmpiRe shows large overlaps with MaxLFQ+t-test and MSqRob, whereas the overlap between MaxLFQ+t-test and MSqRob is small. C, Investigation of the proteins called by only one method. Peptide fold change plots for the proteins with the largest FDR difference to the other two methods. A consistent shift of the boxes above or below log2 fold change 0 (black dashed line) indicates regulation. Left (MS-EmpiRe): protein D37ZPP3–2 (FDR_emp < 0.01, FDR_msqr < 0.45, FDR_mlfq < 0.83). Almost all peptides imply clear up-regulation. Middle (MSqRob): protein Q8C878 (FDR_emp < 0.95, FDR_msqr < 0.01, FDR_mlfq < 0.83). We see varying up- and down-regulation. Right (MaxLFQ+t-test): protein Q9JJL8 (FDR_emp < 0.74, FDR_msqr < 0.61, FDR_mlfq < 0.01). We see varying up- and down-regulation. D) MaxQuant protein intensities versus fold change plot for protein P04424 in the clinical data set (FDR_emp < 0.01, FDR_ttest = 0.99). MS-EmpiRe can clearly resolve the small but systematic fold changes. Many more validation plots for all methods tested can be found under https://www.bio.ifi.lmu.de/files/gruber/empire/.
Differential Quantification with MS-EmpiRe

comparing days in vitro (DIV) 5 with 15. The set of proteins detected by all methods is the largest, but still consists of only 30% of all DCPs. MS-EmpiRe has large overlaps with MSqRob and with MaxLFQ + t-test, whereas the exclusive overlap between MaxLFQ + t-test and MSqRob is very small. This provides further confidence in many hits of MS-EmpiRe and indicates that MS-EmpiRe can close the gap between MaxLFQ + t-test and MSqRob. The remaining sets are hits detected by only one method. Here, MS-EmpiRe has the largest set. The one-method sets must be treated with special care (high chance for false positives) and they will be investigated in more detail the next section. The intersection sets for the other studies and conditions are displayed in supplemental Fig. S6. For most other conditions, the detection rate of MaxLFQ + t-test/TMT + t-test is very low and hence the overlap between all three methods is also very small. There is again a large overlap between MSqRob and MS-EmpiRe and large one-method sets. For the Ramond et al. data sets, seven of the eight proteins passing the threshold are in the combined set of all three methods. This indicates, that MS-EmpiRe is not over-sensitive in data sets with little regulation happening (few DEPs).

A Detail View On the Quantitative Data Validates the Proteins Called by MS-EmpiRe—In contrast to benchmarking data sets, for “real-life” quantitative MS-data sets, a ground truth is not available. We cannot easily decide, whether a DCP is a DEP (i.e., actually regulated). What is possible, however, is to visualize all the quantitative data available for a protein to allow manual inspection in detail. As the quantitative MS data is on the peptide level, we visualize the peptide intensities using peptide fold change plots. For this, we assess the fold changes between all replicates of a peptide in condition 1 and in condition 2 and represent them as a box plot. The fold change plot for a given protein has as many boxes as peptides measured and each box contains as many fold changes as there are replicate pairs between the conditions. In Fig. 5C we show for each method the protein with largest FDR difference to the two other methods. This means the selected protein has a significant (small) FDR in one method whereas both other methods assign it a very insignificant (high) FDR. For MS-EmpiRe, we see highly consistent fold changes in one direction, indeed indicating a DEP. In contrast to the proteins detected by MaxLFQ + t-test and MSqRob, a distinct change of the protein is visible, adding further confidence into the MS-EmpiRe results. For a comprehensive check, we provide visualizations of all proteins detected in the intersection data sets on the website https://www.bio.lif.ifi.lmu.de/files/gruber/empire/, allowing in detail inspection of the DCPs. For the clinical data set of Ping et al., we see that the DCPs show small fold changes in general but very consistent regulation on the peptide level (Fig. 5D). This indicates that the precise quantification of TMT-MS3 together with the MS-EmpiRe approach is a powerful combination for the detection of biomarkers or clinically relevant proteins.

On the web pages we also provide further plots on the data sets: comparative volcano plots, comparative FC-scatter plots and plots of the peptide intensities as well as the MaxLFQ intensities for each protein.

In Silico Benchmarking Shows High Sensitivity and Conservative FDR Estimation—The importance of experimental benchmarking setups for quantitative proteomics cannot be overstated. Without reference standards, it is impossible to estimate the performance of experimental and computational methods. Unfortunately, performing an experimental benchmarking is cumbersome as it requires very precise mixing and sample handling. Additionally, only constant fold changes can be applied to a given setup, which does not reflect an actual regulatory scenario. To complement the experimental benchmarking, we generated an in silico benchmarking set, as described in the methods section. In short, we used the human background proteins measured in O’Connell et al. as replicate measurements and we divided six replicate measurements into two groups. We then applied in silico intensity changes on the protein and peptide level to one of the groups and compared the two groups in a differential quantification context. As we know which proteins are “artificially regulated,” we can assess measures like sensitivity and specificity analogous to the experimental benchmarking setup discussed in the previous sections. We simulated two setups: one like the one in O’Connell et al., where we always applied the same fold change (with some noise) to a sub-fraction of the proteome, including a 10% fraction. Additionally, we simulated a more realistic scenario, where the in silico expression changes were not always the same, but were drawn from a distribution. We designed the distribution to be bimodal such that up- and down-regulation was possible. The results for sensitivity and precision (i.e., specificity) for LFQ data are depicted in Fig. 6. The boxes result from changing different fractions of the proteome (individual simulations where 5%, 10%, . . . , 40% of the proteome are changed). Surprisingly, we noticed that the fraction of proteome changing significantly influences the sensitivity of the applied statistical test, especially for the MaxLFQ + t-test setup. For example, when 30% of the proteome is changing with a fold change of 1.5, this is better detected by a statistical test as when only 5% of the proteome is changing with a fold change of 1.5. The reason for this is apparently a loss of significance after multiple testing correction. As multiple testing correction can be seen as a shifting of the p values into the direction of a uniform distribution, stronger deviations from the uniform distribution (e.g. many regulated proteins) are less strongly affected. In supplemental Fig. S7 we see that the protein level scoring underlying the t-test does not allow a very distinct discrimination between regulated and nonregulated proteins as compared with MS-EmpiRe, which explains the losses in sensitivity with MaxLFQ + t-test. The clearer distinction between regulated and nonregulated proteins by the peptide level tools is also reflected in the fact, that the peptide level tools MS-
**Differential Quantification with MS-EmpiRe**

![Graph showing LFQ Simulation in fixed and dynamic conditions](image)

**Fig. 6. In silico benchmarking of MS-EmpiRe, MaxLFQ + t-test and MSqRob.** The x-tics represent the median fold change by which the data is shifted. The boxes contain the sensitivity/specificity results, when different fractions of the proteome are changed. Each box contains eight values corresponding to 5% up to 40% of the proteome changing in 5% steps. Sensitivity and specificity for different fold changes upon constant (left) as well as dynamic (right) proteome changes are shown. A clear dependence on the fraction of the proteome changing is visible. As in the benchmarking set, MaxLFQ + t-test shows low sensitivity with good error rate control. MSqRob shows high sensitivity but often violates the error estimation and MS-EmpiRe changing is visible. As in the benchmarking set, MaxLFQ + t-test shows low sensitivity with good error rate control. MSqRob shows high sensitivity but often violates the error estimation and MS-EmpiRe shows high sensitivity with good error rate control.

EmpiRe and MSqRob show less dependence on the proteome fraction in terms of sensitivity. In general, the results of the in silico simulations in Fig. 6 show a similar picture as compared with the experimental benchmarking setup. MS-EmpiRe and MaxLFQ + t-test show conservative error estimation, which however comes at the cost of drastically reduced sensitivity for MaxLFQ + t-test. MS-EmpiRe is the most sensitive tool and MSqRob detects only slightly less proteins. However, MSqRob shows problems in terms of error rate control, especially for setups with strong fold changes. This might be because of an over-optmistic error estimation of MSqRob because of the many clear classification cases. Comparing the fixed setup with the setup where we generate dynamic noise, we see that the overall identification rate decreases, whereas the general trends for all three methods are very similar. Error rate estimation does not decrease and hence all methods show the desired response toward high noise in the data.

**CONCLUSION**

Current Mass Spectrometry proteomics publications often report the number of quantified proteins for a given proteomics setup. This number however, can be misleading, as the number of quantified proteins does not necessarily reflect the number of proteins that can actually be detected in a differential quantification experiment. This is especially the case for more noise-prone proteomics setups, such as LFQ. Given the popularity of such setups, increasing the depth of a proteomics experiment at the level of differential detection becomes an ever more important issue. Considering recent studies (13, 14), it is likely that the future development of differential quantification will increasingly use peptide level information. Peptide level tools like MSqRob show significantly improved sensitivity, though we have shown in this study that FDR control is still difficult in a proteome-wide setup. Conversely, the popular MaxLFQ + t-test pipeline also implements the most conservative approach in the experimental benchmarking setup. With MS-EmpiRe, we introduce a new peptide level tool that shows high sensitivity at accurate error rate estimation. While FDR estimation for MS-EmpiRe is almost as accurate as for MaxLFQ + t-test in the experimental setup, it even outperforms MaxLFQ + t-test in the in silico simulation. We have shown that MS-EmpiRe gives up to 2-fold increase in sensitivity for small fold changes (1.5-fold change), which can be highly relevant for biological applications. Even though a fold change of only 1.5 is challenging for a proteomics setup, such a change may already reflect a drastic alteration in a biological system. The key to the sensitivity of MS-EmpiRe is the direct modeling of errors on the peptide fold change level. This gives an immediate statistical weight to individual peptide fold changes, which are then transferred to the protein level via basic statistics and without additional optimizations or parameters. This simple approach relies on the only assumption of consistency between replicates. Therefore, our method heavily relies on good replicate measurements. Even though MS-EmpiRe can work with only two replicate measurements per condition, ideally three or more replicate samples should be available, to obtain accurate error estimates. For MS proteomics data, where robust workflows exist and the creation of replicate measurements is a standard, we believe that this requirement matches well with current experimental practices. From our perspective, the consistency of replicates is a minimalistic and reasonable assumption that can be made for proper processing of proteomics data. A possible deviation from replicate consistency might occur, when uncontrolled factors in a biological experiment change between replicates. In our setup, this might lead to an underestimation of DEPs. However, replicate-inconsistent setups are highly critical and should be handled with care. Based on our analysis, we conclude that MS-EmpiRe is currently the most sensitive tool for differential protein detection. MS-EmpiRe requires as inputs only peptide intensities and protein identifications and, thus, applicable to virtually any modern proteomics measurement. MS-EmpiRe is an easy-to-use option for proteomics researchers and helps to improve the quality and biological insight gained from MS proteomics studies.

**Acknowledgments—**C.A. was supported by a PhD scholarship of the Deutsche Forschungsgemeinschaft (DFG, Graduate School QBM). This work was partially funded by SFB 1123/Z2 of the DFG.
DATA AVAILABILITY

All proteomics data sets used in this study are publicly available through the ProteomeXchange Consortium (http://www.proteomexchange.org/) via the corresponding PRIDE partner repositories (28): O’Connell et al. (17): PXD007683. Sharma et al. (25): PXD001250. Ramond et al. (26): PXD001584. Ping et al. (27): PXD007160. All relevant quantitative protein and peptide intensity tables are available via: https://www.bio.ifi.lmu.de/software/msempire/index.html.

[5] This article contains supplemental material.

To whom correspondence should be addressed. E-mail: zimmer@ifi.lmu.de.

These authors contributed equally to this work.

Author contributions: C.A., M.G., G.C. performed research; C.A., M.G., G.C., and R.Z. analyzed data; C.A., M.G., G.C., and R.Z. wrote the paper.

REFERENCES

1. Bantscheff, M., Lemeer, S., Savitski, M. M., and Kuster, B. (2012) Quantitative mass spectrometry in proteomics: Critical review update from 2007 to the present. Anal. Bioanal. Chem. 404, 939–965
2. Olsen, J. V. (2005) Parts per Million Mass Accuracy on an Orbitrap Mass Spectrometer via Lock Mass Injection into a C-trap. Mol. Cell. Proteomics 4, 2010–2021
3. Ding, L., Rad, R., Gygi, S. P., and Haas, W. (2011) MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. Nat. Methods 8, 937–940
4. 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, doi: 10.1074/mcp.O111.016717. Epub 2012 Jan 18
5. Bruderer, R., Bernhardt, O. M., Gandhi, T., Xuan, Y., Sondermann, J., Schmidt, M., Gomez-Varela, D., and Reiter, L. (2017) Optimization of experimental parameters in data-independent mass spectrometry significantly increases depth and reproducibility of results. Mol. Cell. Proteomics 41, 2296–2309
6. Meier, F., Geyer, P. E., Virreira Winter, S., Cox, J., and Mann, M. (2018) BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes. Nat. Methods 15, 440–448
7. Abaye, D. A., Pullen, F. S., and Nielsen, B. V. (2011) Peptide polarity and the position of arginine as sources of selectivity during positive electrospray ionization mass spectrometry. Rapid Commun. Mass Spectrom. 25, 3597–3608
8. 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
9. Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., and Hamon, C. (2003) Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal. Chem. 75, 1895–1904
10. Tyano, 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 (proteomics) data. Nat. Methods 13, 731–740
11. Clough, T., Key, M., Ott, I., Ragg, S., Schadow, G., and Vittek, O. (2009) Protein quantification in label-free LC-MS experiments. J. Proteome Res. 8, 5275–5284
12. Karpiuk, Y., Stanley, J., Tavener, T., Huang, J., Adkins, J. N., Ansorg, C., Heffron, F., Metz, T. O., Qian, W. J., Yoon, H., Smith, R. D., and Dabney, A. R. (2009) A statistical framework for protein quantitation in bottom-up MS-based proteomics. Bioinformatics 25, 2028–2034
13. Goeminne, L. J., Argentinii, A., Martens, L., and Clement, L. (2015) Summarization vs peptide-based models in label-free quantitative proteomics: Performance, pitfalls, and data analysis guidelines. J. Proteome Res. 14, 2457–2465
14. Goeminne, L. J., Gevaert, K., and Clement, L. (2016) Peptide-level robust ridge regression improves estimation, sensitivity, and specificity in data-dependent quantitative label-free shotgun proteomics. Mol. Cell. Proteomics 15, 657–668
15. Koh, H. W., Swa, H. L., Fermir, D., L’er, S. G., Gunaratne, J., and Choi, H. (2015) EBprot: Statistical analysis of labeling-based quantitative proteomics data. Proteomics, 15, 2580–2591
16. Rosenthal, R. (1978) Combining results of independent studies. Psychol. Bull. 85, 185–193
17. O’Connell, J. O., Paulin, J. A., O’Brien, J. J., and Gygi, S. P. (2018) Proteome-wide evaluation of two common protein quantification methods. J. Proteome Res. 17, 1934–1942
18. Smyth, G. K. (2005) Limma: linear models for microarray data. In Bioinformatics and computational biology solutions using R and Bioconductor (R. Gentleman, V. Carey, S. Dudoit, R. Irizarry and W. Huber, eds.), pp. 397–420, Springer, New York.
19. Anders, S., and Huber, W. (2010) Differential expression analysis for sequence count data. Genome Biol. 11, R106
20. Robinson, M. D., McCarthy, D. J., and Smyth G. K. (2009) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140
21. Robinson, M. D., and Oshlack A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biology, 11, R25
22. Zien,A., Aigner, T., Zimmer, R., and Lengauer, T. (2001) Centralization: A new method for the normalization of gene expression data. Bioinformatics, 17, S323–S331
23. Cox J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 26, 1367–1372
24. Eng, J. K., McCormack, A. L., and Yates, J. R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. 5, 976–989
25. Sharma, K., Schmitt, S., Bergner, C. G., Tyanova, S., Kannaiyan, N., Manrique-Hoyos, N., Kongi, K., Cantuti, L., Hanisch, U. K., Philips, M. A., Roessner, M. J., Mann, M., and Simons, M. (2015) Cell type–and brain region–resolved mouse brain proteome. Nature Neurosci. 18, 1819
26. Ramond, E., Gesbert, G., Guerrera, I. C., Chhuon, C., Dupuis, M., Rigard, M., Henry, T., Barel, M., and Charbit A. (2015) Importance of host cell arginine uptake in francisella phagosomal escape and ribosomal protein amounts. Mol. Cell. Proteomics 14, 870–881
27. Ping, L., Duong, D. M., Yin, L., Gearing, M., Lah, J. J., Levey, A. I., and Seyfried, N. T. (2016) Global quantitative analysis of the human brain proteome in Alzheimer’s and Parkinson’s disease. Scientific data 5, 180036
28. Vizcaíno, J. A., Côté, R. G., Csordas, A., Dianes, J. A., Fabregat, A., Foster, J. M., Griss, J., Alpi, E., Birim, M., Contell, J., O’Keeley, G., Schoenegger, A., Ovelleiro, D., Pérez-Riverol, Y., Reisinger, F., Rios, D., Wang, R., Hermajob, H. (2013) The Proteomics Identifications (PRIDE) database and associated tools: status in 2013, Nucleic Acids Res. 41, D1063–D1069