Bioinformatics, 32(7), 2016, 1040–1047
doi: 10.1093/bioinformatics/btv675
Advance Access Publication Date: 20 November 2015
Original Paper

OXFORD

Gene expression

iPQF: a new peptide-to-protein summarization method using peptide spectra characteristics to improve protein quantification

Martina Fischer and Bernhard Y. Renard*

Research Group Bioinformatics (NG 4), Robert Koch Institute, 13353 Berlin, Germany

*To whom correspondence should be addressed.

Associate Editor: Burkhard Rost

Received on July 1, 2015; revised on October 8, 2015; accepted on November 10, 2015

Abstract

Motivation: Isobaric labelling techniques such as iTRAQ and TMT are popular methods for relative protein abundance estimation in proteomic studies. However, measurements are assessed at the peptide spectrum level and exhibit substantial heterogeneity per protein. Hence, clever summarization strategies are required to infer protein ratios. So far, current methods rely exclusively on quantitative values, while additional information on peptides is available, yet it is not considered in these methods.

Methods: We present *iPQF* (**i**sobaric **P**rotein **Q**uantification based on **F**eatures) as a novel peptide-to-protein summarization method, which integrates peptide spectra characteristics as well as quantitative values for protein ratio estimation. We investigate diverse features characterizing spectra reliability and reveal significant correlations to ratio accuracy in spectra. As a result, we developed a feature-based weighting of peptide spectra.

Results: A performance evaluation of *iPQF* in comparison to nine different protein ratio inference methods is conducted on five published MS2 and MS3 datasets with predefined ground truth. We demonstrate the benefit of using peptide feature information to improve protein ratio estimation. Compared to purely quantitative approaches, our proposed strategy achieves increased accuracy by addressing peptide spectra reliability.

Availability and implementation: The *iPQF* algorithm is available within the established R/Bioconductor package *MSnbase* (version $\geq$ 1.17.8).

Contact: renardB@rki.de

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Mass spectrometry based proteomics has evolved as the method of choice for identification and quantification of proteins (Domon and Aebersold, 2006), and major advances were achieved in the development of new quantification techniques. Isobaric labelling techniques such as iTRAQ and TMT have gained much popularity, allowing for simultaneous absolute and relative protein quantification in different samples within a single run (Gygi *et al.*, 1999; Ong *et al.*, 2002; Ross *et al.*, 2004; Thompson *et al.*, 2003). This enables the investigation of changes in protein abundance across various conditions, which is crucial for the study of regulation processes, diagnostics research and biomarker studies. Thereby, accuracy in protein ratio estimates plays an essential role. However, accuracy problems in iTRAQ and TMT data have been demonstrated by different studies (Gan *et al.*, 2007; Hultin-Rosenberg *et al.*, 2013; Karp *et al.*, 2010; Kirchner *et al.*, 2010; Mahoney *et al.*, 2011) and reliable protein ratio estimation remains a challenging task.

Several steps are involved in the quantification process. First, peptides are identified and quantified by iTRAQ or TMT reporter ions in the MS/MS spectra. Factors contributing directly to the variability of peptide quantitative estimates include: efficiency of protein

digestion and labeling, co-eluting peptides, reporter ion peak detection, intensity assessment, label interference and a limited dynamic range of the instrument (Burkhart *et al.*, 2011; Vaudel *et al.*, 2010). A frequently reported bias is the underestimation of ratios and its compression towards one, which is supposed to arise from co-eluting peptides (Bantscheff *et al.*, 2008; Ow *et al.*, 2009; Sandberg *et al.*, 2014). Several approaches address these issues by proposing specific sample preparations (Burkhart *et al.*, 2011), intensity calculation methods and correction strategies (Boehm *et al.*, 2007; Shadforth *et al.*, 2005; Vaudel *et al.*, 2010). Further MS3 data acquisition is considered as a new promising strategy to reduce and potentially eliminate the peptide interference effect (Ting *et al.*, 2011).

The next major step in this process is the inference from peptides to proteins. Measurements of label intensities are assessed at the spectra level and subsequently a summarization strategy is needed to estimate the corresponding protein ratios. Generally, all peptide spectra assigned to a protein are assumed to share the same expression profile. Indeed substantial variance heterogeneity is observed due to random and systematic biases (Bauer *et al.*, 2012; Karp *et al.*, 2010). The question arises how a peptide-to-protein summarization method can appropriately address this existing variance heterogeneity. Different studies demonstrated that the coefficient of variance is dependent on the absolute signal intensity, suffering from higher variation in low-intensity than in high-intensity data (Hundertmark *et al.*, 2009; Hultin-Rosenberg *et al.*, 2013; Karp *et al.*, 2010; Mahoney *et al.*, 2011). Therefore different summarization methods were developed to account for these intensity-dependent effects by filtering for low intensity peptides (Hu *et al.*, 2006), weighting peptides according to their absolute intensities (Hultin-Rosenberg *et al.*, 2013, Lin *et al.*, 2006) or by applying a variance stabilization method (Huber *et al.*, 2002; Karp *et al.*, 2010). Other approaches examine the error structure and the underlying ratio distributions and develop noise models accordingly (Breitwieser *et al.*, 2011; Karp *et al.*, 2010; Zhou *et al.*, 2014). Further, standard statistical concepts, such as averaging by mean or median, are still one of the most commonly used methods to find protein ratio estimates from a range of peptide quantities. Multiple tools and comprehensive iTRAQ quantification pipelines either offer or are exclusively based on simple median or weighted mean calculations for protein ratio inference (Boehm *et al.*, 2007; Lin *et al.*, 2006; Onsongo *et al.*, 2010). Additionally, strategies for filtering outlying peptide ratios are frequently proposed, including methods like Grubb's and Dixon's test (Choe *et al.*, 2005; Li *et al.*, 2003). A different category of approaches requires the integration of replicate samples or spike-in proteins to enable an assessment of the internal experimental variation (Hultin-Rosenberg *et al.*, 2013).

All these summarization methods have in common that they only focus on quantitative peptide information in order to infer protein quantities. So far, the main feature, which is extensively studied and related to the reliability of peptide quantities, is the absolute intensity signal. However, there are several additional characteristics of peptides available, which are known to have an impact on the overall reliability of a specific peptide and its measurements.

In this work, we identify and investigate the impact of diverse peptide spectra features such as charge state, sequence length, identification score, mass and a distance metric within uniquely and redundantly measured spectra. We examine how these features correlate with the variance heterogeneity and to which extent they are related to quantification accuracy in spectra. Our aim is to find a combination of feature criteria that allows inferring ratio reliability by using the complementary strength of the features. As a result, we developed *iPQF* which integrates the information of peptide spectra characteristics with given quantitative values. We show the added value of peptide spectra feature information to improve protein ratio estimation.

The proposed algorithm can be combined with any purely quantitative approach. In addition, a fundamental intention was not to disregard any information, but rather to keep peptide spectrum matches and down weight unreliable spectra according to the features instead of losing information by filtering. Further, no internal replicates or specific sample setup in the design of iTRAQ and TMT experiments is required which may restrict applicability.

Finally, we evaluate the performance of our approach on five different published iTRAQ and TMT datasets providing a ground truth of known peptide and protein quantities. Thereby, we consider three MS2 datasets with minimal amount of biases, one MS2 dataset showing a high peptide interference effect as well as one MS3 dataset. A comparison study with nine commonly used peptide-to-protein summarization methods is conducted. To our knowledge, this is also the most comprehensive comparison study of summarization methods.

2 Methods

2.1 Peptide spectra features

Considering the relative quantification values of peptide spectra being assigned to the same protein, a substantial heterogeneity is observed (shown in Supplementary Fig. S1). The objective of this work is to investigate whether the observed peptide variation can be explained by underlying peptide spectra characteristics. Thereby, we aim to relate diverse features of spectra to the quality of their quantitative information. As a result, the reliability of given peptide spectra can be inferred and protein quantification can be improved by accounting for it.

In order to study the impact of features on the quantification accuracy, we assess the deviance of ratios from the spectra to a given ground truth by calculating the Euclidian distance across all iTRAQ/TMT labels, subsequently referred to as *quantification error*. Next, a correlation study is conducted by calculating Spearman's correlation coefficient between feature values and the peptide spectra quantification error.

We examine the impact of the following peptide features: identification score, sequence length, charge state, mass, absolute ion intensity, modification state and a distance metric within uniquely and redundantly measured spectra as explained below. The shared status of a peptide is not considered and corresponding spectra are discarded, as the negative impact of an incorrectly assigned peptide may be larger than the potential gain of an additional peptide for protein ratio estimation.

Here, we define a group of *redundant* spectra as several MS/MS events for one peptide, while *unique* spectra are referred to peptides quantified exactly by one MS/MS event. For redundant peptide spectra of a protein, which are subject to the same conditions in the MS experiment, an even higher ratio similarity across labels is expected than among different sequence fragments of a protein. Hence, a peptide spectrum exhibiting ratios diverging from all other ratios in the redundant spectra group is suspected to be less reliable. For each protein we form different groups according to its different redundant spectra and one group pooling all uniquely measured spectra. The idea is that not only the number of spectra per protein matters, but also the degree of ratio similarity within these groups. For each peptide spectrum we compute the mean Euclidian distance of its ratios to the ratios of all other spectra belonging to the same group.

The identification score indicates the correctness of the peptide spectrum match. A low score implies less reliable peptide identification and consequently an uncertainty in the peptide to protein assignment, potentially resulting in an incorrect peptide ratio for the protein ratio calculation.

The impact of absolute ion intensity was already intensively studied and is well known as a key indicator for the reliability of ratio estimates. It has been shown that the accuracy of peptide ratio estimates depends strongly on the involved absolute intensities (Carrillo *et al.*, 2010; Hundertmark *et al.*, 2009; Karp *et al.*, 2010; Mahoney *et al.*, 2011). Low intensities are expected to be subject to noise and ratios exhibit large variations, while ratio estimates converge to the true value as intensity increases. Here, we calculate the mean absolute intensity across all labels for each spectrum.

Peptide modifications in iTRAQ experiments occur mainly due to enzymatic or sample preparation related reactions. A slightly increased false-positive protein identification rate was reported by allowing more modifications to be present (Tenga and Lazar, 2011). Further, varying peptide expression behavior in a protein and shifted ratios were observed due to modifications. In our investigation, we distinguish between modified and unmodified peptide spectra without further distinguishing specific types of modification.

The features charge state, mass and sequence length are interrelated and have direct or indirect impact on peptide identification. Higher charge states give rise to a variety of possible fragments carrying diverse amounts of charges. The peptide search space needs to be expanded accordingly and the risk of false-positive identifications is increased as a consequence. Further, long sequence peptides tend to show a bias to higher identification scores compared to short sequences dependent on the identification tool. The importance of these features and their crucial role has also been shown in other work (Keller *et al.*, 2002; Käll *et al.*, 2007; Fusaro *et al.*, 2009).

2.2 iPQF algorithm

The proposed algorithm iPQF (**i**sobaric **P**rotein **Q**uantification based on **F**eatures) is a peptide-to-protein summarization method. For each peptide spectrum, it requires peptide identification, reporter ion intensities and assignment to the respective protein. Next, a summarization strategy is required to combine given peptide spectra quantities to estimate protein quantities.

iPQF presents a novel approach by using information of spectra features to evaluate peptide spectra ratios. Spectra receive weights and contribute to the protein quantification according to their reliability.

The algorithm is conducted protein-wise, which means individual protein quantifications are not influenced by other protein quantifications in the dataset. However, the number of identified peptide spectra per protein is important; we recommend a minimum of three spectra for protein quantification.

The algorithm consists of six steps, which are calculated for each protein (Fig. 1) (see also example process in the Supplementary Material): (i) **Feature assessment**: Feature values are computed for each of the seven different features for all peptide spectra belonging to the specific protein. (ii) **Spectra ranking per feature**: Peptide values obtained for each feature are ranked from most to least reliable feature value based on knowledge of associated low and high quantification errors which was acquired in our correlation study. Hence, if a peptide spectrum receives a high rank for a specific feature, this means its reported quantification is considered more reliable by this feature compared to a spectrum showing a lower rank.

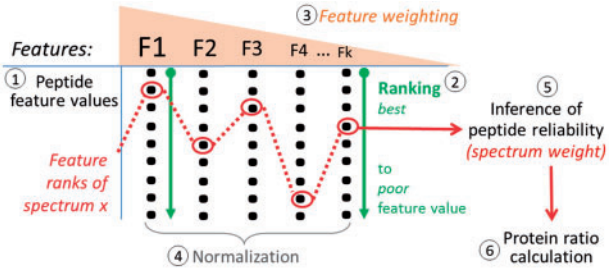

Fig. 1. Outline of the iPQF summarization method for protein ratio inference using a feature-based weighting of peptide spectra. Six steps are conducted for each protein to estimate the reliability of its underlying spectra ratios

(iii) **Feature weighting**: For each peptide spectrum we obtain several ranks, one for each presented feature, and each rank individually states the quantification reliability of the spectrum. Yet the explanation power of the features is different, and the impact of the diverse features is weighted according to strong and weak correlations observed with quantification errors (see results in Section 4.1). We propose a default weighting order of features based on consensus observations in the different datasets and prove its robustness (for a more detailed explanation and the robustness analysis refer to Supplementary Material) (iv) **Normalization of ranks**: We normalize the ranks of each feature by the overall number of spectra to ensure the ranks to be within the range of zero and one. (v) **Inference of overall peptide spectra reliability**: The feature ranks obtained for each spectrum are combined to receive an overall reliability measure called peptide spectrum weight. We do so by calculating a classic average rank per spectrum and normalize it by the weighted sum of all features. As a result, peptide spectra receiving weights close to one represent reliable ratios to enable the inference of true protein ratios, while peptide spectra weights decreasing to zero refer to a reduced confidence in its given quantification values. (vi) **Protein ratio calculation**: A weighted mean approach using squared peptide spectrum weights is conducted to estimate the underlying protein ratio.

Further, iPQF protein estimates can be additionally coupled to pure quantitative strategies using a mean approach, referred to as *combined iPQF* approach here. Generally, we recommend applying the algorithm based on relative spectra intensities in order to estimate protein ratios instead of using absolute intensities. The variance in absolute intensities can be large, while relative intensities are more robust.

2.3 Implementation

The introduced iPQF algorithm is implemented in R (version $\geq$ 3.1.3), and was integrated into the existing R/Bioconductor package *MSnbase* (version $\geq$ 1.17.8) (Gatto and Lilley, 2012), which offers a variety of processing functions for iTRAQ data (see MSnbase vignette). Further, the algorithm is designed for optional combination with any summarization method, which focuses exclusively on quantitative values, to combine strengths of both approaches.

3 Experimental setup

3.1 Dataset description

We evaluate peptide quantification data from five different published MS2 and MS3 datasets based on iTRAQ and TMT experiments, which have predefined protein fold-changes. Thereby we

consider three MS2 datasets with smaller fold changes and minimal interference effect as well as one MS2 dataset affected by high peptide interference events. Overall, the datasets hold diverse data characteristics concerning the dataset size, the number of identified spectra per protein, the expected ratios and the range of peptide feature values, thus covering different possible protein-peptide scenarios.

1. Dataset (MS2) from Hultin-Rosenberg *et al.*: Peptides from a lung cancer cell line A549 were labeled with iTRAQ 8-plex tags according to a 2:2:1:1:2:2:1:1 fold change. Here, the dataset showing most identifications in the publication was chosen, which is based on a 400 μg loaded peptide amount, prefractionated by IPG-IEF and analyzed on a LTQ Orbitrap Velos (Thermo Scientific). Peptide spectra identification and protein inference was performed using Proteome Discoverer 1.1 with Mascot 2.2 (Matrix Science), and identified peptides below a 1% protein FDR level were quantified. Further, peptide intensities were isotope impurity corrected.
2. Dataset (MS2) from Breitwieser *et al.*: A 4-plex iTRAQ experiment was designed with human plasma proteins holding constant ratios of 1:1:1:1 and two spiked-in proteins, a rat ceruplasmin being mixed in 1:2:5:10 ratio concentrations and a mouse ceruplasmin with 10:5:2:1 ratios. MS analysis was conducted on a hybrid LTQ Orbitrap XL (Thermo Scientific) coupled to a HPLC nanoflow system (Agilent 1200). Peptide spectra were searched and quantified using Mascot 2.3 and Phenyx 2.6.1 and only concordant peptide identifications were kept. Protein inference was set to hold an FDR level of 1%.
3. Dataset (MS2) from Zhou *et al.*: Replicate samples from mouse cell lysates were created with equal concentrations, labeled with iTRAQ 8-plex reagents (expected ratios 1:1:1:1:1:1:1:1) and measured by a TripleTOF 5600 (Absciex). The ProteinPilot software was used for peptide spectra identification and quantification, holding the protein FDR below 1%.

   4.-5. Datasets with MS3 and MS2 spectra from Ting *et al.*: A 6-plex TMT experiment was designed with a two-proteome mixture model containing human cell lines and yeast Lys-C digests to study the peptide interference effect. Yeast peptides were mixed according to 10:4:2.5:10:4:2.5 ratios and human peptides with equal amounts (1:1:1) were added to the first three labels. The MS2 dataset presents compressed yeast ratios in the first three labels due to human peptide interference, while in the MS3 dataset the interference effect is almost eliminated. Samples were measured on a LTQ Orbitrap Velos. The focus here is on the yeast peptide and protein identification and quantification which was performed by Sequest with a protein FDR of 1.5%.

All datasets were filtered for shared peptides, contaminants, and for spectra showing missing or zero intensities in one of the iTRAQ/TMT labels. An additional filtering was applied in case of MS3 dataset (4) due to extreme outliers in the dataset (see also filtering by Ting *et al.*), using a less restrictive approach than in the original publication and discarding only spectra deviating more than tenfold from the expected ratios which are biologically not reasonable (Supplementary Fig. S2d). Peptide spectra intensities were normalized according to the median intensity present in each label for dataset (2) and (3). No normalization was applied in the case of dataset (1) and (4)–(5), as this would contrast with the fold-change setting defined for all peptides. Further, protein identifications based on the support of only one or two peptide spectra are not considered for quantification and evaluation here. As a result of the preprocessing, 624 proteins based on 5885 peptide spectra are considered in dataset (1), 145 proteins with 13 758 spectra in dataset (2), 2811 proteins with 217 822 spectra in dataset (3) and 781 proteins with 8934 spectra in MS3 dataset (4) (processing and analysis of the corresponding dataset (5) with MS2 spectra representing the impact of peptide interference can be found in the Supplementary Material).

The aim of this work is the computation of accurate protein ratios based on relative peptide intensities; here, we focus mainly on the relative intensity level of proteins and peptide spectra and not on the absolute intensities. Hence, ratios are calculated for all datasets. For datasets (2) and (3), a ratio of a spectrum is defined by dividing its absolute intensity of one iTRAQ label by its summed intensities of all labels. This is a robust approach, as it satisfies greater label independence in the ratio calculation and peptide ratios are not exclusively based on one specific label. In case of dataset (1), we followed the ratio computation in the corresponding publication, in which intensities were divided by the mean intensity of iTRAQ label 113 and 114. For datasets (4)–(5) we relied on the provided ratios.

3.2 Method comparison

All introduced datasets come with predefined ratios for spectra and proteins, thus allowing the performance evaluation of diverse peptide-to-protein summarization methods. In order to compare the different summarization methods and to assess their accuracy in estimating protein ratios, we consider the *protein estimation error*. The error is defined as the squared differences of the protein ratio estimates to the ground truth with subsequent summation across labels.

We investigate and compare up to nine commonly used peptide-to-protein summarization methods with our proposed iPQF approach. Protein ratios are estimated based on given peptide spectra ratios for each label individually in all presented methods:

- *Median*: The estimated protein ratio corresponds to the median of peptide ratios being assigned to the protein.
- *Mean*: The mean is used instead of the median.
- *Mean (Top5, Top3)*: A group of five or three spectra showing the highest absolute intensities are selected respectively and the mean is applied (Searle, 2010; Silva *et al.*, 2006).
- *Tukey's Median Polish*: An additive model is iteratively fitted to the ratios until the sum of absolute residuals falls below a significantly small threshold. The sum of the resulting overall median and label effect, given by the model, is used to estimate the protein ratio. (Gatto *et al.*, 2012; Tukey *et al.*, 1977)
- *Sum of intensities*: The absolute peptide spectra intensities of one protein are summed for each label. Protein ratios are calculated on the basis of the intensity sums. (Carrillo *et al.*, 2010)
- *Total Least Squares*: The objective to find the protein ratio is to fit a straight line between peptide spectra ratios of two different labels. Different from linear regression, here orthogonal distances are minimized between ratios and an optimal line. (Carrillo *et al.*, 2010; Huffel and Vandewalle, 1991)
- *isobar*: A noise model is built which estimates the underlying noise variance dependent on absolute spectra intensities. The inverse of the noise variance serves as a weighting factor for individual peptide spectra. Protein ratios are subsequently computed by a weighted average approach. (Breitwieser *et al.*, 2011)

Additionally, a comparison to protein quantification results of Mascot and ProteinPilot, which were provided in the supplements of the corresponding publications, is included for datasets (1) and (3), respectively.

4 Results

First, we demonstrate the correlation between peptide spectra features and quantification accuracy. Second, we evaluate the performance of the iPQF algorithm in comparison to nine summarization methods. Results are provided for the different MS2 datasets as well as for the additional MS3 dataset.

4.1 Peptide feature correlation study

The distributions of peptide ratios measured by the different labels are shown in Supplementary Figure S2a–e. Ratio values are spread around the ground truth values of the corresponding dataset. Considering the quantification error per spectrum, defined by the Euclidian distance of the measured ratios to the expected ratios, a right skewed distribution is observed in all datasets. The two spike-in proteins of dataset (2) (Breitwieser *et.al*) each exhibit a group of strongly diverging peptide spectra ratios from the ground truth, which causes a second peak in the quantification error distribution (Supplementary Fig. S2b).

The correlation of quantification errors to peptide spectra features is analyzed to study the feature impact on ratio accuracy. The corresponding Spearman's correlation coefficients are reported in Table 1. All correlation coefficients are assessed to be statistically significant by using Spearman's rho statistic to estimate a rank-based measure of association. Overall, correlations observed are strikingly consistent across the three MS2 datasets, despite different sample complexity, experimental setups, different instrumentation and different analysis software used. Further, even with an additional isolation and fragmentation step resulting in an MS3 scan, the same correlation trend with slightly decreased correlation coefficients is observed.

The most meaningful feature reflecting ratio accuracy is the proposed similarity metric within redundantly and uniquely measured spectra groups, holding positive correlations between 0.52 and 0.72. Hence, a small mean Euclidian distance of a specific peptide spectrum to spectra belonging to the same redundant or unique group, respectively, implies a small quantification error. However, the error increases with the peptide spectrum diverging from its group (Fig. 2A, Supplementary Fig. S3–4). Further, ratio accuracy is decreasing with increasing charge of a peptide spectrum, especially apparent in the most common range between a charge state of two and four (Fig. 2B, Supplementary Fig. S5). The increase of noise and ratio variation in low absolute ion intensity data has been shown before and is also confirmed in this study (Supplementary Fig. S6). A consistently increasing ratio error is observed with increasing sequence length from mainly 5 to 30 amino acids, illustrated by a positive correlation between 0.17 and 0.39 (Fig. 2C, Supplementary Fig. S7). The high interrelation between length and mass of a peptide is also clearly reflected by similar correlation coefficients to the quantification error, further supporting both features as indicators of ratio reliability (Supplementary Fig. S8). Correlation of the identification score varies between the datasets due to the different scoring systems, datasets (1) and (2) are based on Mascot, while dataset (3) relies on ProteinPilot and datasets (4)–(5) on Sequest. Generally higher scores correspond to smaller ratio errors; however, it is interesting to observe that error variation increases at the same time (see further details in Supplementary Fig. S9). For the group of modified spectra the ratio error appears to be increased in all datasets compared to non-modified spectra (Supplementary Fig. S10).

Table 1. Correlation study of peptide spectra features to relative quantification error

Peptide features	Spearman's correlation coefficient			
	Dataset 1 (MS2) (Hultin-Rosenberg *et al.*)	Dataset 2 (MS2) (Breitwieser *et al.*)	Dataset 3 (MS2) (Zhou *et al.*)	Dataset 4 (MS3) (Ting *et al.*)
Redundancy metric	0.65	0.71	0.72	0.52
Uniquely measured metric	0.61	0.67	0.68	0.55
Charge state	0.54	0.38	0.18	0.14
Ion intensity	−0.49	−0.59	−0.64	−0.23
Sequence length	0.39	0.29	0.25	0.17
Mass	0.38	0.30	0.20	0.16
Identification score	−0.13	−0.34	0.09	0.08
Modification	0.14	0.11	0.22	0.13

The visualization of peptide feature-error-correlations displays a homogenous trend for all datasets, notably for MS2 as well as MS3 data (refer to Supplementary Fig S11). The impact of peptide interference events on feature-error-correlations is shown by means of the dataset (5) with MS2 spectra (see Supplementary Table 1 and Fig. S12). Additionally, the two spike-in rat and mouse proteins of dataset (2) are shown separately in Supplementary Fig. S13. In particular, short peptide sequences are assigned to the rat protein and the observed outlier peptide group consists exclusively of redundantly measured spectra showing low absolute intensities. Further, a study of inter-correlations between features reveals a strong and expected relation structure among features such as length, mass, charge state and score (Supplementary Fig. S14). However, despite significant correlations of individual features, the combination of features is crucial to eliminate pitfalls of single features and make use of opposed strength. The proposed iPQF approach combines the information from all different features to obtain overall ratio reliability for each spectrum.

4.2 Evaluation of protein summarization methods

For evaluation of peptide-to-protein summarization methods, we rely on diverse datasets, in particular concerning the number of peptide spectra per protein (Supplementary Fig. S15). Dataset (1) and (4) consist of a large number of proteins being supported by predominantly three to ten or respectively twenty peptide spectra, while dataset (2) comprises only 145 proteins based on a range of three to over hundred spectra. Dataset (3) is an overall large dataset holding a median of 26 spectra per protein and diverse cases of several hundred spectra.

We present a performance evaluation of iPQF and nine additional protein ratio inference methods, which are primarily based on quantitative peptide information only. The accuracy of each method is described by the *protein estimation error*, which is assessed for each protein of a dataset, and a statistical summary is displayed in form of boxplots. Method comparisons are shown for three MS2 datasets in Figure 3. We present two forms of iPQF results, the *pure* form of iPQF using spectra feature information only and a *combined* form in which iPQF is coupled to one of the quantitative approaches, here shown for iPQF combined with the MedianPolish approach. The combined form illustrates the added value of feature information to quantitative approaches.

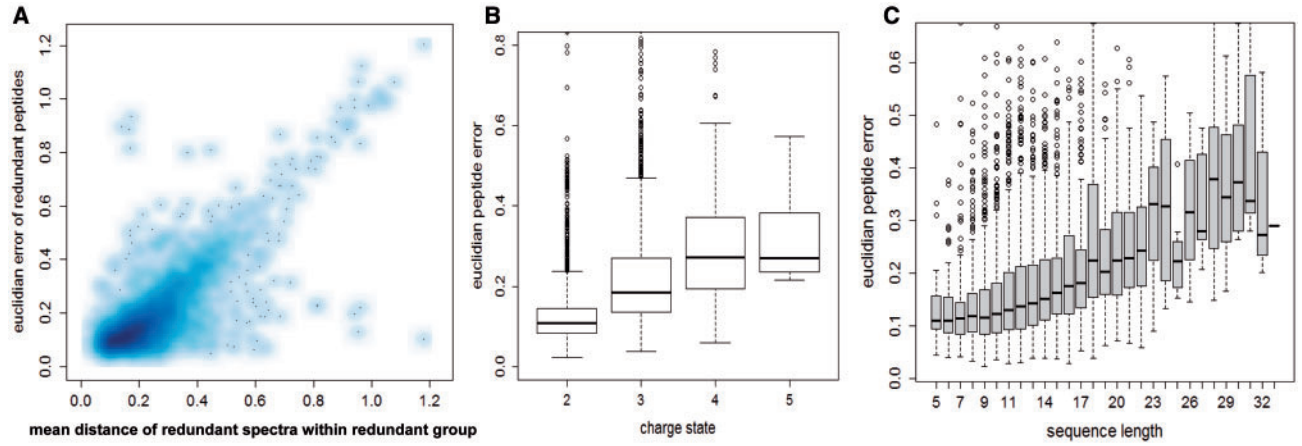

Fig. 2. Correlation of spectra features to quantification error, shown for three selected features of dataset (1) (Hultin-Rosenberg *et al.*). The impact of the features (A) redundancy metric, (B) charge state and (C) sequence length on spectra ratio accuracy is displayed. A significant trend is observed in all cases

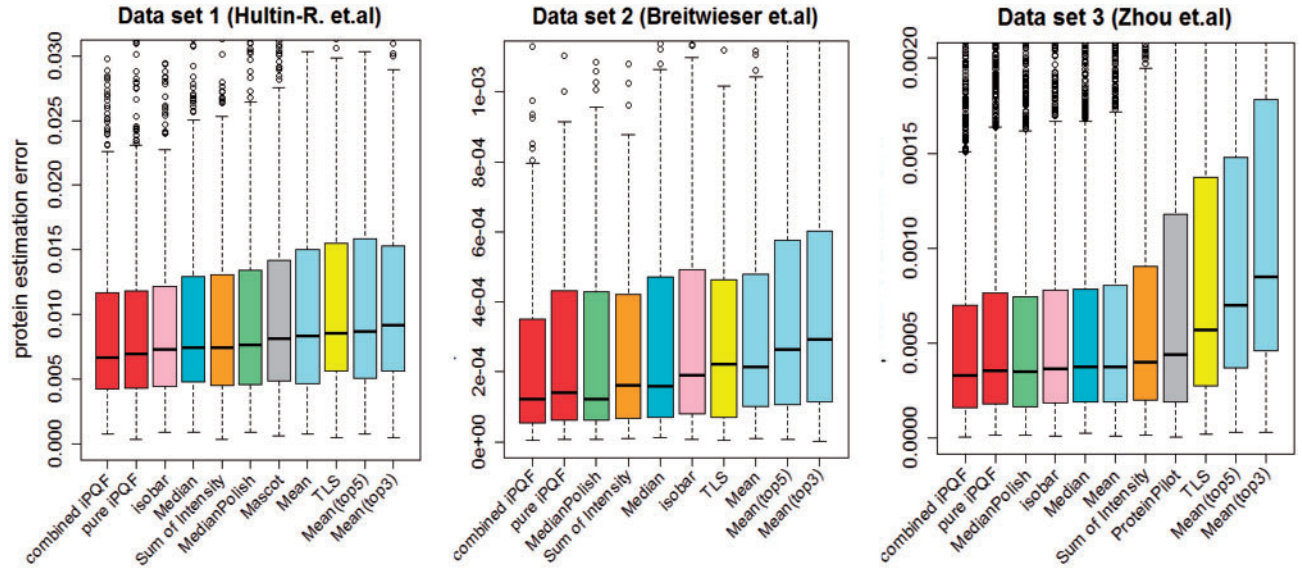

Fig. 3. Performance evaluation of iPQF approaches and nine summarization strategies shown for three different MS2 datasets (dataset (1): 624 proteins with 5885 spectra; dataset (2): 145 proteins with 13758 spectra; dataset (3): 2811 protein with 217822 spectra). Boxplots display the protein estimation error of each method applied (note that methods are ordered according to error size). Improved and robust protein ratio accuracy is observed for the iPQF approaches in all three datasets

Overall, the pure iPQF approach shows better performance in dataset (1) and comparable performance to the other summarization methods in dataset (2) and (3), proving the importance of feature information. The combined iPQF approach exhibits the best protein ratio accuracy of all methods in each of the three different datasets. Further, iPQF approaches prove robustness, while other methods vary in performance dependent on the dataset applied.

In particular, feature information is of high value in dataset (1), which is dominated by small peptide spectra numbers per protein corresponding to sparsely available quantitative information. Thus, iPQF approaches perform best using all additional knowledge to weight spectra, while the diverse mean-based approaches struggle most due to high variation within protein profiles based on few spectra. The more robust and sophisticated approaches show an intermediate performance. As spectra numbers vary more in dataset (2), the pure iPQF approach becomes comparable to the other methods; however, the combined iPQF improves over all methods by throughout lower quantiles including a significantly reduced upperquartile of the estimation error. The large spectra numbers in dataset (3) result in similar quantiles of estimation errors of most methods, even the mean approach performs equivalent to the more robust median and all other sophisticated approaches. In contrast, mean (top5/top3) methods restrict themselves to few peptide spectra with high absolute intensity and have a significant performance loss. Also in this dataset, the combined iPQF achieves improved ratio accuracy, shown by consistently lower quantiles. Generally, the commercial and commonly used tools Mascot and ProteinPilot do not show competitive performance, here.

Evaluation of iPQF in MS3 data is presented in Figure 4 and also confirms superior performance of the combined iPQF approach, while pure iPQF shows comparable results to other approaches. Generally high protein estimation errors are observed due to many

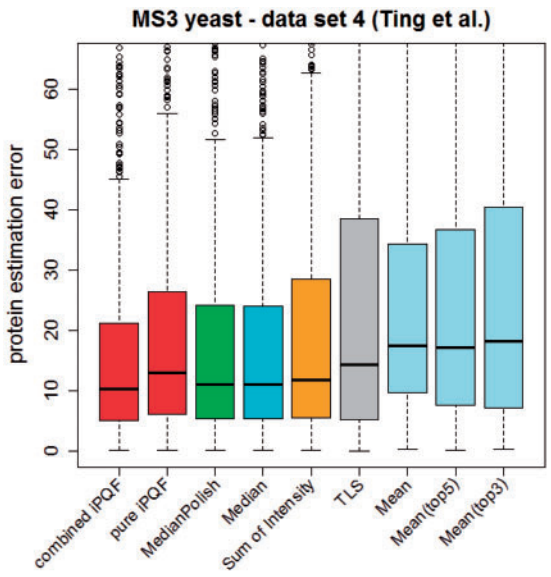

Fig. 4. Performance of iPQF approaches and seven other summarization strategies in a MS3 data setting (Note that isobar could not be run on dataset 4). A reduced protein estimation error is attained for the combined iPQF strategy, confirming the benefit of features also in MS3 data (781 proteins with 8934 spectra)

outlying ratios in the dataset which significantly impact the performance of the mean based approaches. A performance comparison of the methods on dataset (5) being affected by peptide interference also supports the integration of feature information (Supplementary Fig. S16).

Further, we evaluate accuracy details of the methods by considering specific deviation ranges from the ground truth ratios and assess the amount of protein ratios which could be estimated within this deviation range. A superior sensitivity can be observed for the iPQF approaches (Supplementary Fig. S17). Additionally, the AUC measure (area under the curve) is provided for all methods, showing the highest AUC for the combined iPQF (Supplementary Table 2).

5 Discussion

Inference of protein ratios based on heterogeneous peptide spectra measurements remains a crucial issue, which receives little consideration in most quantification pipelines. In this work, we present a new summarization strategy *iPQF*, which integrates spectra characteristics with quantitative values for protein ratio estimation. We investigate different peptide spectra features and reveal significant correlations between features and quantification accuracy. As a result, we are able to show the added value of feature information to achieve improved protein ratio accuracy.

Peptide spectra features contain valuable information in addition to pure quantitative information. Since no individual feature shows near-perfect correlation to quantification error, the combination of features can be crucial to compensate for failures of individual features and to make use of their diverse strengths. Overall, it is unlikely that a peptide spectrum is mischaracterized by a large set of features at the same time.

In particular, proteins with a high diversity of underlying feature values profit from the approach, while feature uniformity naturally reduces the impact by giving similar weights to spectra. This is primarily relevant for proteins holding a small to medium number of peptide spectrum matches exhibiting ratio variation. Here particularly, benefit of the iPQF approach is shown. In cases of large numbers of peptide spectra the feature impact is decreased and approaches using the mean already perform considerably well.

Another prerequisite for successful protein ratio estimation is that peptide ratio measurements are spread around the true protein ratio value. The best protein quantification method still remains dependent on given peptide quantities, and cannot work if peptide values coherently and systematically diverge from the ground truth. Feature and quantification error correlation will also not necessarily be sufficiently strong in these divergent cases as the error is strongly biased.

A major issue in iTRAQ and TMT datasets is the peptide interference effect which causes the underestimation of ratios and its compression towards one. MS3 data acquisition has proven to significantly reduce the interference effect. Evaluation of iPQF approaches also confirms the applicability in MS3 data settings and still shows a robust performance under interference impact compared to other methods.

The flexible design of the algorithm enables further extensions. One option is to join results of a purely quantitative method with estimations obtained by iPQF to benefit from both strategies. Here, we provide a combination of iPQF with MedianPolish and show significant improvements over both individual methods in our results. The advantage of a joined approach is that in case of few peptide spectra per protein additional feature knowledge can compensate for the sparse information in the quantitative setup, while more sophisticated summarization strategies can be applied with rich quantitative information available. Further, a different option is to exclusively employ the spectra feature-based reliability measure provided by iPQF and integrate it in existing summarization approaches. Beyond this, new and relevant features of interest can be easily added to the implemented feature framework.

Generally, the idea of a feature-based spectra weighting is transferable beyond iTRAQ data. While our studies only focus on feature-error-correlations in iTRAQ and TMT data, similar findings are expected for SILAC as well as label-free data. Algorithmic steps of iPQF are technically applicable to quantitative proteomic methods requiring peptide summarization, but careful evaluation in the context of the specific experiment is necessary.

Moreover our proposed approach is independent of using replicate samples or spike-in proteins, independent of the instrument, and the selected multiplex. Further, in contrast to modelling approaches mostly requiring larger numbers of peptide spectra, iPQF is also applicable in small settings. Also no assumption concerning underlying ratio distributions or specific data criteria is required. Hence, we also chose replicate independent summarization methods and use corresponding settings in tools, such as isobar (Breitwieser *et al.*, 2011), for evaluation comparison.

In addition, a fundamental intention was to keep peptide spectra by applying a feature based weighting instead of losing information by filtering. Filtering of low-intensity spectra or outlier ratios is commonly performed; however this significantly reduces the protein coverage as few peptide readings per protein typically dominate the datasets (Karp *et al.*, 2010). Further, defining a cutoff for outlier filtering is a critical issue as important information is potentially discarded.

Overall, we provide a broad performance comparison of nine different protein ratio inference methods on five published datasets with predefined ground truth. To the best of our knowledge, an overall benchmark study of current methods assessing diverse biases and impact on protein ratio accuracy in iTRAQ/TMT data is missing. Here, we also provide a basis for future comparison of summarization methods.

6 Conclusion

The goal of the protein quantification process is the inference of protein quantities based on peptide quantities. However, peptide ratios assigned to a protein exhibit substantial heterogeneity and require clever summarization strategies. We present *iPQF*, which integrates peptide spectra characteristics as well as quantitative values for protein ratio estimation. The novelty of the algorithm is to weight spectra according to their feature reliability. Comprehensive evaluation of iPQF in comparison to other summarization methods yields a superior and robust performance. As a result, the benefit of feature information to achieve improved protein ratio accuracy is shown.

Acknowledgements

We are very grateful to L. Gatto (University of Cambridge) for insightful comments and for making iPQF available in MSnbase. We thank K. Trappe and M. Lindner for critical and helpful comments on the manuscript.

Funding

The authors acknowledge financial support by Deutsche Forschungsgemeinschaft (DFG), grant number RE3474/2-1 to BYR.

Conflict of Interest: none declared.

References

Bantscheff,M. *et al.* (2008) Robust and sensitive iTRAQ quantification on an LTQ Orbitrap mass spectrometer. *Mol. Cell. Proteomics*, **7**, 1702–1713.

Bauer,C. *et al.* (2012) PPINGUIN: Peptide profiling guided identification of proteins improves quantitation of iTRAQ ratios. *BMC Bioinform.*, **13**, 34.

Boehm,A.M. *et al.* (2007) Precise protein quantification based on peptide quantification using iTRAQ. *BMC Bioinform.*, **8**, 214.

Breitwieser,F.P. *et al.* (2011) General statistical modeling of data from protein relative expression isobaric tags. *J. Proteome Res.*, **10**, 2758–2766.

Burkhart,J.M. *et al.* (2011) iTRAQ protein quantification: a quality-controlled workflow. *Proteomics*, **11**, 1125–1134.

Carrillo,B. *et al.* (2010) Methods for combining peptide intensities to estimate relative protein abundance. *Bioinformatics*, **26**, 98–103.

Choe, L.H. *et al.* (2005) A comparison of the consistency of proteome quantitation using two-dimensional electrophoresis and shotgun isobaric tagging in *Escherichia* coli cells. *Electrophoresis*. **26**, 2437–2449.

Domon,B. and Aebersold,R. (2006) Mass spectrometry and protein analysis. *Science*, **312**, 212–217.

Fusaro,V.A. *et al.* (2009) Prediction of high-responding peptides for targeted protein assays by mass spectrometry. *Nat. Biotechnol.*, **27**, 190–198.

Gan,C.S. *et al.* (2007) Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ). *J. Proteome Res.*, **6**, 821–827.

Gatto, L. and Lilley K.S. (2012) MSnbase – an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. *Bioinformatics*, **28**, 288–289.

Gygi,S.P. *et al.* (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.*, **17**, 994–999.

Hu,J. *et al.* (2006) Optimized proteomic analysis of a mouse model of cerebellar dysfunction using amine-specific isobaric tags. *Proteomics*, **6**, 4321–4334.

Huber,W. *et al.* (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, **18**, 96–104.

Hultin-Rosenberg,L. *et al.* (2013) Defining, comparing and improving iTRAQ quantification in mass spectrometry proteomics data. *Mol. Cell. Proteomics*, **12**, 2021–2023.

Hundertmark,C. *et al.* (2009) MS-specific noise model reveals the potential of iTRAQ in quantitative proteomics. *Bioinformatics*, **25**, 1004–1011.

Karp,N.A. *et al.* (2010) Addressing accuracy and precision issues in iTRAQ quantitation. *Mol. Cell. Proteomics*, **9**, 1885–1897.

Käll,L. *et al.* (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods*, **4**, 923–925.

Keller,A. *et al.* (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.*, **74**, 5383–92.

Kirchner,M. *et al.* (2010) Computational protein profile similarity screening for quantitative mass spectrometry experiments. *Bioinformatics*, **26**, 77–83.

Li,X.Y. *et al.* (2003) Automated statistical analysis of protein abundance ratios from data generated by stable-isotope dilution and tandem mass spectrometry. *Anal. Chem.*, **75**, 6648–6657.

Lin,W.T. *et al.* (2006) Multi-Q: a fully automated tool for multiplexed protein quantitation. *J. Proteome Res.*, **5**, 2328–2338.

Mahoney,D.W. *et al.* (2011) Relative quantification: characterization of bias, variability and fold changes in mass spectrometry data from iTRAQ-labeled peptides. *J. Proteome Res.*, **10**, 4325–4333.

Ong,S.E. *et al.* (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics*, **1**, 376–386.

Onsongo,G. *et al.* (2010) LTQ-iQuant: a freely available software pipeline for automated and accurate protein quantification of isobaric tagged peptide data from LTQ instruments. *Proteomics*, **10**, 3533–3538.

Ow,S.Y. *et al.* (2009) iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly". *J. Proteome Res.*, **8**, 5347–5355.

Ross,P.L. *et al.* (2004) Multiplexed protein quantitation in Saccharomyces cerevisiaeusing amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics*, **3**, 1154–1169.

Sandberg,A. *et al.* (2014) Quantitative accuracy in mass spectrometry based proteomics of complex samples: the impact of labeling and precursor interference. *Proteomics*, **96**, 133–144.

Searle,B.C. (2010) Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics*, **10**, 1265–1269.

Shadforth, I.P. *et al.* (2005) i-Tracker: for quantitative proteomics using iTRAQ (TM). *BMC Genomics*, **6**, 145.

Silva,J.C. *et al.* (2006) Absolute quantification of proteins by LCMS^{E}. A virtue of parallel ms acquisition. *Mol. Cell. Proteomics*, **5**, 144–156.

Tenga,M. and Lazar J.M. (2011) Impact of peptide modifications on iTRAQ quantitation accuracy. *Anal. Chem.*, **83**, 701–701.

Ting,L. *et al.* (2011) MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat. Methods*, **8**, 937–940.

Thompson,A. *et al.* (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.*, **75**, 1895–1904.

Van Huffel,S. and Vandewalle,J. (1991) *The Total Least Squares Problem: Computational Aspects and Analysis*. Society for Industrial and Applied Mathematics, Philadelphia.

Vaudel,M. *et al.* (2010) Peptide and protein quantification: a map of the minefield. *Proteomics*, **10**, 650–670.

Zhou,C. *et al.* (2014) A hierarchical statistical modeling approach to analyze proteomic isobaric tag for relative and absolute quantitation data. *Bioinformatics*, **30**, 549–558.