MCP Papers in Press. Published on April 10, 2010 as Manuscript M900628-MCP200

Addressing accuracy and precision issues in iTRAQ quantitation

Natasha A. Karp[1], Wolfgang Huber[2], Pawel G. Sadowski[3], Philip D. Charles[3], Svenja V. Hester[3], and Kathryn S. Lilley[3]

1: Wellcome Trust Sanger Institute, Cambridge, England.

2: European Bioinformatics Institute, EMBL Outstation: Hinxton, England

3: Department of Biochemistry, Cambridge University, Cambridge, England.

Correspondence: Dr Kathryn Lilley,

Cambridge System Biology Centre,

Dept of Biochemistry,

University of Cambridge,

Tennis Court Road,

Cambridge.

CB2 1QR

Email: k.s.lilley@bioc.cam.ac.uk

Tel: +44-1223-760253

Fax: +44-1223-333345

Keywords: iTRAQ, VSN, normalisation, heterogeneity of variance, accuracy, precision

Running title: Addressing accuracy and precision in iTRAQ quantitation

Abbreviations:

Q-Q: quantile-quantile

ECA: *Erwinia carotovora*

CV: coefficient of variation

RI plot: Ratio- intensity plot

iTRAQ: isobaric tags for relative or absolute quantitation

VSN: variance stabilising transformation

HCD: higher collision decomposition

FWHM: full width at half maximum

Summary

iTRAQ is a mass spectrometry technology that allows quantitative comparison of protein abundance by measuring peak intensities of reporter ions released from iTRAQ tagged peptides by fragmentation during MS/MS. However, current data analysis techniques for iTRAQ struggle to report reliable relative protein abundance estimates and suffer with problems of precision and accuracy. The precision of the data is affected by variance heterogeneity: low signal data have higher relative variability, however low abundant peptides dominate datasets. Accuracy is compromised as ratios are compressed towards one, leading to under-estimation of the ratio. This manuscript investigates both issues and proposes a methodology which combines the peptide measurements to give a robust protein estimate even when the data for the protein are sparse or at low intensity. Our data indicate that ratio compression arises from contamination during precursor ion selection, which occurs at a consistent proportion within an experiment and thus results in a linear relationship between expected and observed ratios. We propose that a correction factor can be calculated from spiked proteins at known ratios. Secondly, we demonstrate that variance heterogeneity is present in iTRAQ data sets irrespective of the analytical packages, LC-MS/MS instrumentation and iTRAQ labelling kit (4-plex or 8-plex) employed. We propose using an additive-multiplicative error model for peak intensities in MS/MS quantitation and demonstrate that a variance stabilising transformation (VSN) is able to address the error structure and stabilise the variance across the entire intensity range. The resulting uniform variance structure simplifies the downstream analysis. Heterogeneity of variance consistent with an additive-multiplicative model has been reported in other MS-based quantitation including fields outside of proteomics; consequently the VSN methodology has the

potential to increase the capabilities of MS in quantitation across diverse areas of biology and chemistry.

1. Introduction

Different techniques are being used and developed in the field of proteomics to allow quantitative comparison of samples between one state and another. These can be divided into gel [1-4] or mass spectrometry based [5-8] techniques. Comparative studies have found that each technique has strengths and weaknesses and plays a complementary role in proteomics [9, 10]. There is significant interest in stable isotope labelling strategies of proteins or peptides, as with every measurement there is the potential to employ an internal reference allowing relative quantitation comparison, which significantly increases sensitivity of detection of change in abundance. Isobaric labelling techniques such as tandem mass tags [11, 12] or isobaric tags for relative or absolute quantitation (iTRAQ) [13, 14] allow multiplexing of six and eight separately labelled samples within one experiment respectively. In contrast to most other quantitative proteomic methods, where precursor ion intensities are measured, here, the measurement and ensuing quantitation of iTRAQ reporter ions occurs after fragmentation of the precursor ion. Differentially labelled peptides are selected in MS as a single mass precursor ion, as the size difference of the tags is equalised by a balance group. The reporter ions are only liberated in MS/MS, after the reporter ion and balance groups fragment from the labelled peptides during collision induced dissociation (CID). iTRAQ has been applied to a wide range of biological applications from bacteria under nitrate stress [15] to mouse models of cerebellar dysfunction [16].

For the majority of MS based quantitation methods (including MS/MS-based methods like iTRAQ), the measurements are made at the peptide level and then combined to compute a summarised value for the protein from which they arose. An advantage is that the protein can be

identified and quantified from data of multiple peptides, often with multiple values per distinct peptide, thereby enhancing confidence in both identity and the abundance. However, question arises of how to summarise the peptide readings to obtain an estimate of the protein ratio. This will involve some sort of averaging, and we need to consider the distribution of the data, in particular these three aspects: (i) are the data centred around a single mode (which would be related to the true protein quantitation) or are there phenomena that make them multi-modal? (ii) are the data approximately symmetric (non-skewed) around the mode? (iii) are there outliers? In the case of multi-modality, it is recommended that an effort be made to separate the various phenomena into their separate variables and to dissect the multimodality. Li *et al.* developed ASAPratio for ICAT data that included a complex data combination strategy. Peptide abundance ratios are calculated by combining data from multiple fractions across MS runs, and then averaging across peptides to give an abundance ratio for each parent protein [17]. GPS Explorer, a software package developed for iTRAQ, assumes normality in the peptide ratio for a protein once an outlier filter is applied [18]. The iTRAQ package ProQuant, assumes that peptide ratio data for a protein follow a log-normal distribution [19]. Averaging can be via mean [20], weighted average [21, 22] or weighted correlation [23]. Some of these methods try to take into account the varying precision of the peptide measurements. There are many different ideas of how to process peptide data but as yet no systematic study has been completed to guide analysis and ensure the methods being utilised are appropriate.

The quality of a quantitation method can be considered in terms of precision, which refers to how well repeated measurements agree with each other, and accuracy, which refers to how much they on average deviate from the true value. Both of these types of variability are inherent to the

measurement process. Precision is affected by random errors, non-reproducible and unpredictable fluctuations around the true value. (In)accuracy, by contrast, is caused by systematic biases that go consistently in the same direction. In iTRAQ, systematic biases can arise due to inconsistencies in iTRAQ labelling efficiency and protein digestion [22]. Typically, ratiometric normalisation has been used to address this tag bias, where all peptide ratios are multiplied by a global normalisation factor determined to centre the ratio distribution on one [19, 22]. Even after such normalisation, concerns have been raised that iTRAQ has imperfect accuracy with ratios shrunken towards one and this under-estimation have been reported across multiple MS platforms [23-27]. It has been suggested that this under-estimation arises from co-eluting peptides with similar m/z values, which are co-selected during ion selection and co-fragmented during collision induced dissociation [23, 27]. As the majority of these will be at one to one ratio across the reporter ion tags (as required for normalisation in iTRAQ experiments), they will contribute a background value equally to each of the iTRAQ reporter ion signals and diminish the computed ratios.

With regard to random errors, iTRAQ data are seen to exhibit heterogeneity of variance, that is, the signal's variance depends on its mean. In particular, the coefficient of variation (CV), is higher in data from low intensity peaks than in data from high intensity peaks [16, 22, 23]. This has also been observed in other MS based quantitation techniques when quantifying from the MS signal [28-30]. Different approaches have been proposed to model the variance heterogeneity. Pavelka *et al.* used a power law global error model (PLGEM) in conjunction with quantitation data derived from spectral counts [31]. Other authors have proposed that the higher CV at low signal arises from the majority of MS instrumentation measuring ion counts as whole numbers

[32]. Anderle *et al.* described a two-component error model in which Poisson statistics of ion counts being measured as whole numbers dominate at the low intensity end of the dynamic range and multiplicative effects at the high intensity end, and demonstrated its fit to a label free LC-MS quantitation data [28]. Previously, Rocke and Lorenzata proposed a two component additive-multiplicative error model in environmental toxin monitoring study utilizing gas chromatography MS in the 1990's [29].

How can the variance heterogeneity be addressed in the data analysis? Some of the current approaches include: outlier removal [18, 25], weighted means [21, 22], inclusion filters [16, 22], logarithmic transformation [19] and a weighted correlation analysis [23]. Outlier removal methods, for example using Dixon's test, assume a normal distribution for which there is little empirical basis. The inclusion filter method, where low intensity data are excluded, reduces the protein coverage considerably if the heterogeneity is to be significantly reduced. The weighted mean method results in higher intensity readings contributing more to the weighted mean than readings from low intensity readings. Filtering, outlier removal and weighted methods are of limited use for peptides for which only a few low intensity readings were made; however, such cases typically dominate the datasets. Even with a logarithmic transformation, heterogeneity has been reported for iTRAQ data [16, 19, 22]. Current methods struggle to address the issue and to maintain sensitivity.

Here, we investigate the data analysis issues that relate to precision and accuracy in quantitation and propose a robust methodology that is designed to make use of all data without ad hoc filtering rules. The additive-multiplicative model, mentioned above, motivates the so-called

generalised logarithm transformation, a transformation that addresses heterogeneity of variance by approximately stabilizing the variance of the transformed signal across its whole dynamic range [34]. Huber *et al.* provided an open source software package, VSN, which determines the data-dependent transformation parameters. Here we report that the application of this transformation is beneficial for the analysis of iTRAQ data. We investigate the error structure of iTRAQ quantitation data, using different peak identification and quantitation packages, LC-MS/MS data collection systems, and both the 4-plex and 8-plex iTRAQ systems. The usefulness of the VSN transformation to address heterogeneity of variance is demonstrated. Furthermore, we consider the correlations between multiple, peptide-level readings for the same protein and propose a method to summarise them to a protein abundance estimate. We consider same-same comparisons to assess the magnitude of experimental variability, and then use a set of complex biological samples whose biology has been well characterised to assess the power of the method to detect true differential abundance. We assess the accuracy of the system with a four protein mix at known ratios spanning a fold change expression range of 1 to 4. From this we propose a methodology to address iTRAQ's accuracy issues.

2. Experimental procedures

Table 1 summarises the datasets used in this analysis. Detailed experimental procedural information is available in supplementary information. To evaluate experimental variability in the iTRAQ system we prepared same-same datasets, for which an aliquot of the same sample was labelled by each of the available isobaric tags and then combined prior to peptide separation and quantitation. Same-same datasets were collected for different sample types, quantitation systems, MS/MS systems and for both the 4- and 8-plex labelling system. To investigate iTRAQ

accuracy, an experiment was prepared with a background of proteins at unchanging level but with the addition of four spiked proteins of known ratios (Table 2). Two of the proteins were present at one to one to allow data normalisation to adjust for tag differences. To examine the approach on a complex biological system with biological differences, iTRAQ data were collected from yeast grown under various nutritional limiting conditions.

Suggested location for Table 1 and 2

3. Results

3.1 Raw data analysis

3.1.1 The data sampling characteristics of iTRAQ

Examining the peptide level data highlights an unbalanced peptide sampling; some peptides are sampled many times whilst the majority of peptides are only sampled once or twice (Supplementary Figure 1). At the protein level, this leads to some proteins having only one reading whilst others have hundreds. The majority of these peptide readings are low volume and hence to maximize the sensitivity of the study it is desirable to keep these peptides for the data analysis (Supplementary Figure 2). The volume distribution arises as sampling of peptides is not random, but rather occurs as a result of a data dependent selection process in the MS for the high intensity peaks beyond any exclusion list/dynamic exclusion process applied. This limits iTRAQ to relative level comparison only (i.e. comparing ratios).

3.1.2 Fragmentation behaviour

To assess biases and variability in fragmentation, we examined the ratio between reporter ion maximum intensity and the 145 Da peak maximum intensity in the phosphorylase B dataset, whose high sampling depth allowed analysis at the peptide level. The 145 Da peak arises from incomplete fragmentation and is composed of the balance group attached to the 114-117 Da reporter group. The mass of 145 Da is common to all of the four 4-plex tags. We considered the top 31 sampled peptides (which comprise 50% of the dataset). The data were filtered, by removing peptide readings if they contained missing reporter ion values occurred, or if two or more of the reporter ion peak maximum intensities were below 15 counts. First, we found fragmentation efficiency to be peptide dependent; this is shown by the different ratios for different peptides between the reporter ion and the 145 Da peak intensities (Supplementary Data). Second, fragmentation efficiency was consistent across tags within an experiment run (data not shown); this is unsurprising, since with the iTRAQ system, fragmentation of the four reporter ion occurs simultaneously. These results provide further support for preferring relative level comparisons over raw measurements.

3.1.3 Heterogeneity of variance: the variance-mean dependence

The previous sections used the standard i-Tracker filtering method. To understand the variance behaviour fully, all quantified peptides were included in the analyses from this point on. Ratio-intensity plots (RI plot) were used to assess the distribution of ratios as a function of average signal strength (Figure 1).While panels A and B show that the centre of the distribution of log-ratios has no significant intensity dependent systematic bias, in agreement with the findings of Hu *et al. [16]*, the width of the distribution is significantly larger at low intensities than at high intensities. This heterogeneity of variance has previously been seen in iTRAQ data collected

with a 4700 Proteomics Analyzer (Applied Biosystems, USA) and analysed with GPS explorer (Applied Biosystems, USA) (ABI) [16], and independently with data collected with a QSTAR when analysed with ProQuant v1.1 [21].

Suggest location for figure 1

The logarithm transformation, has previously been suggested for iTRAQ data with the objective of addressing the heterogeneity of variance [19, 21]. However, Figure 1A and B show that the logarithm transformation does not sufficiently stabilize the variance. To further investigate the error structure, the relationship between the mean and the variance on the log scale for the four tags for each peptide reading after normalisation was assessed for each dataset; a representative case is shown in Figure 2. The plot is consistent with the additive-multiplicative (two-component) error model: a multiplicative component, with a leading exponent of 1 on the log-log plot, dominates at high intensities. At low intensities, the variance tends to a constant, signal-independent value due to an additive component.

Suggested location for figure 2

3.2. Variance stabilising transformation

Many measurements in physics and chemistry follow a multiplicative error model. Consider a quantity whose true value is x, and measurements of which result in observed values $x(1+\varepsilon)$, where ε represents small, positive or negative random numbers. Then, the standard deviation of the measurements is x times the standard deviation of ε. When transformed to the logarithmic

scale, the measured values are (to good approximation) log $x + \varepsilon$, and the standard deviation is simply the standard deviation of ε, independent of the true value x. This transformation is therefore referred to as a variance stabilising transformation, and the concept can be generalised to the additive-multiplicative error model. The variance stabilising transformation in that case is called the generalised logarithm [35-37], resembles the usual logarithm transformation at the upper end of the intensity scale (where multiplicative effects dominate), a linear transformation at the lower end (where additive effects dominate) and interpolates smoothly in between (Supplementary Figure 3). Furthermore, if we define the usual log-ratio between two peak intensities I_1 and I_2 as

$$q = \log \frac{I_1}{I_2} \quad \text{[Eq1]}$$

then a generalised log-ratio can be defined as

$$\rho = \log \frac{I_1 + \sqrt{I_1^2 + c^2}}{I_2 + \sqrt{I_2^2 + c^2}} \quad \text{[Eq2]}$$

Here, c is a data-dependent constant; more specifically, it depends on the mean and standard deviation of the additive error component. For values of I_1 and I_2 both much larger than c, the generalised log-ratio ρ simplifies to the usual log-ratio. ρ is compressed towards 0 compared to q, i.e., its absolute value $|\rho|$ is always smaller than $|q|$. The size of this *shrinkage* depends on the size of I_1 and I_2, becoming more pronounced as I_1 or I_2 get smaller. The VSN software [34] can be used to fit the parameter c of the generalised logarithm. In addition, it allows a simultaneous affine-linear (shift and scale) normalisation to adjust the data for systematic, label or sample processing associated biases. Like other global normalisation methods, the VSN algorithm uses

the assumption that the majority (50% or more) of intensity values are truly not changing in expression.

The result of the generalised log (variance stabilisation normalisation, VSN) transformation on the *Erwinia* same-same dataset is shown in Figure 1 C and D: the variation of the generalised log-ratio is independent of the signal strength. The coefficient of variance (CV) is frequently used as a measure of variability. Figure 3 shows, with the blue dots, the heterogeneity of the CV as a function of average signal seen with log2 transformed data. In the proteomics community, techniques are frequently compared via single CV summary values [21]. However, as Figure 3 shows, due to the heterogeneity and intensity-dependence of the CV, summaries such as "median CV" are generally too simplistic, and may be misleading. Figure 3 also contrasts this behavior with that of the data transformed by the VSN transformation: their CV is approximately constant, and it coincides with the CV of the logarithm transformed data at high intensity levels. For medium or low intensities, the CV of the VSN transformed data is reduced compared to the logarithm transformed data.

In the *Erwinia* dataset of Figure 3, the high intensity convergence is not reached, which suggests that the MS data were not collected over the full dynamic range. However, it was seen with the phosphorylase B dataset (Supplementary Figure 4). The phosphorylase B sample is a simpler sample with individual peptides at much higher signal strengths. The CV following VSN transformation was smaller with the phosphorylase B dataset compared to the *Erwinia* datasets, but was similar between the *Erwinia* experiments. This is unsurprising as the *Erwinia* datasets had an additional separation stage (SCX) and were derived from a more complex sample.

Suggested location for figure 3

To demonstrate that the phenomena reported here are independent of analysis software and MS/MS collection system, the phosphorylase B QSTAR 4-plex data were re-processed with Mascot V2.2.0 (Matrix Science, London) and the LTQ-OrbitrapXL 4-plex data were processed with Proteome Discoverer (Thermo Fisher, USA) and Mascot V2.2.0 (Matrix Science, London) using the default settings for quantitation. For both packages, the data behaviour was essentially identical to that reported above: no significant systematic intensity dependent bias of the mean, presence of heterogeneity of variance, consistency with the additive-multiplicative error model, and variance stabilisation by the generalised logarithm transformation as fit by the VSN software (Supplementary Figures 4, 5 & 6).

To increase throughput, an 8-plex iTRAQ version has been released (Applied Biosystems, USA). The 8-plex version relies on the same amine-labelling chemistry of peptides as with the 4-plex reagents. The 8-plex version has a modified tag compared with the 4-plex version and a larger balance group. Additional reporter ions at 113, 118, 119, and 121 m/z are liberated during CID of the 8-plex tags enabling increased multiplexing of samples in experiments quantifying protein expression. To ensure that the behaviour being addressed is universal to iTRAQ irrespective of which sets of tags are employed, same-same phosphorylase B 8-plex labelled sample was injected on the LTQ-OrbitrapXL. The data were processed with Mascot V2.2.0 (Matrix Science, London) for quantitation using the default settings for quantitation. Again, data behaviour was essentially identical to that reported above (Supplementary Figure 7).

3.3 From peptides to protein: A complex structure

While the measurements are made at the peptide level, interest often lies at the protein level, and a method is needed to summarize the peptide-level readings into a single, robust relative abundance estimate for each protein. A variety of approaches have been suggested for this task, which differ in how they address the different potential biases and the potentially different amount of confidence (precision) in each peptide-level reading. Here, we first discuss these issues, then present our approach.

3.3.1.1 Fraction effect

We define a 'fraction effect' within a peptide as a significant dependence between the measured ratio and the fraction in which the reading was taken. The top ten sampled peptides from *Erwinia* dataset B and C were examined for a fraction effect by grouping the VSN transformed data by fraction and using a one-way ANOVA to assess for a significant difference in the mean between groups. Only fractions which were sampled more than three times were included in the analysis. 45% of the peptides had a statistically significant difference between fraction groups and the percentage of the variance explained by the fraction effect varied between 37.0 and 86.5% (average 57.1%). From this analysis we conclude for these peptides that the fraction effect was significant. With the phosphorylase B dataset, no statistically significant differences were seen between repeat injections, indicating that the fraction effect arose not from the repeat injections but rather from the separation (SCX) stage. Note that a fraction effect was also seen with log transformed or raw data when a Kruskal–Wallis test was used (data not shown). These results indicate that the error within fraction group for a peptide is smaller than the error between

fraction groups and is arising not from the repeat injections but from additional variance from the repeated SCX separation.

3.3.1.2 Peptide effect

We define a 'peptide effect' within a protein as a significant dependency between the measured ratio and the precursor ion (i.e. peptide). Due to the fraction effect, insufficient numbers of readings were obtained per peptide to consider a peptide effect for the *Erwinia* datasets. The phosphorylase B datasets, however, were designed to get multiple readings for each peptide, and an ANOVA was used to test for a significant difference in the mean ratio between peptides. Only peptides which were sampled more than three times were included in the analysis. The percentage of the variance explained by the peptide effect varied between 13.1 and 78.5 % (average 54.0%) depending on the tag combination examined. This peptide effect was observed for both MS instrumentations and for all software packages analysed. These results indicate that the measurement error within a peptide group for a protein is smaller than the error between peptide groups.

3.3.1.3 Intensity effect

As described in the previous section, the variance, and hence the readings' confidence intervals, are different in different parts of the dynamic range. It is uncommon to have a large number of replicate readings for each peptide, hence estimating that variance directly is impractical. We proposed applying the VSN variance stabilising transformation, which puts the data on a scale on which intensity effects on the variance are removed, but are traded for a intensity-dependent conservative bias, that is, shrinkage towards ratios of one when the intensities are small.

3.3.1.4 Data distribution

To investigate the data distribution and ensure the appropriate application of statistical tools, we plotted frequency histograms and normal quantile-quantile (Q-Q) plots for the readings of the top ten sampled peptides from the phosphorylase B dataset after VSN normalisation (Figure 4). The data distributions were localised and uni-modal, resembling a combination of a normal distribution with outliers. This was found for data obtained with or without the i-Tracker standard low volume filter (discard if less than 3 of the reporter ions above a threshold of 15 counts)[39].

Suggested location for figure 4

3.3.2 Estimating the protein ratio

First, we compute a robust central tendency measure for each protein, such as the trimmed average of the VSN-transformed peak intensities of all the peptides belonging to the protein. Differences between these quantities for different conditions then measure the differential abundance of the protein between the conditions. In doing so, we ignore the fraction and peptide effects described in Sections 3.3.1.1-2, and accept the conservative variance-bias trade-off of the generalised log-ratio described in Section 3.3.1.3. While it is conceivable that a mathematical model could be constructed that explicitly models and adjusts for these effects, such an approach would likely be complicated by unbalanced data structure (Section 3.1.1), often with few readings at each level, and by fragility to outliers and model misspecification. Here, we argue that while ignoring these effects might potentially incur suboptimal estimates, the disadvantage

is by far offset in practice, at least with data from current experiments, by the simplicity and robustness of the above approach.

Figure 5 shows the CV, at the protein level, of protein abundance estimates, where peptide data were combined with a 20% trimmed mean. The CV was calculated at the protein level using the ratio obtained from the six different possible tag combinations. For comparison, the CVs of protein abundance estimates are also shown when the ordinary logarithm transformation was used instead of VSN's generalised logarithm. With VSN, the CV showed no signal strength dependence and was generally lower than with the logarithm.

Suggested location for figure 5

3.3.3 Selecting a significance threshold

In the simplest situation, iTRAQ is used in a pair-wise comparison [10, 18, 21, 40, 41]. A protein is deemed to be differentially abundant if measured ratios exceed a certain threshold. The threshold is chosen such that it encompasses the majority of technical variation in a same-same comparison. This analysis approach assumes that the samples being compared are representative of the population and takes no account of biological variation. The thresholds that encompass 90 and 95% of the experimental variation were found to be reproducible across different tag combinations (Figure 6, Table 3). For the *Erwinia* datasets, a ±1.1 fold-change threshold encompassed 95% of the experimental variation after a trimmed mean estimation of protein ratio using VSN transformed data was employed. Thus, the experimental variation is so low that

proteins with low changes in expression will be detectable in a pair-wise comparison, although the researcher will need to assess such a change is biologically significant.

Suggested location for figure 6 and table 3

3.4 Validation: application to real data

Both a log transformation with ratiometric normalisation and the VSN transformation were applied to data from a biological study comparing yeast grown under various nutritional limiting conditions [42]. The variability of the data from yeast samples was found to be higher than in the *Erwinia* study (Table 4). For both the VSN and log transformed data, when biological differences were present, they were reflected by the protein ratios (Figure 7). For example, for the carbon versus nitrogen limited samples, the top ten proteins, as judged by the largest fold change, were searched for function information in the *Saccharomyces cerevisiae* genome database. 8 of the 10 proteins had database information indicating change in expression triggered by carbon or nitrogen source change limitation. The findings were also in agreement with the transcriptome, endometabolome and exometabolome metabolic control analysis of yeast grown under nutritional limiting conditions by Castrillo *et al.* [42].

Suggested location for figure 7

The VSN transformed data identified considerably more proteins as having significant change in expression compared to the log transformed data (Table 5). The greater sensitivity with the VSN method arose from the reduced variability of the peptide readings used to estimate the protein

ratio (Supplementary Figure 9), and was also reflected by the lower sensitivity threshold with VSN (Table 4).

Suggested location for table 4 & 5.

3.5 Accuracy

The above analysis has focused on the precision of the iTRAQ technology. Concerns have been raised that iTRAQ might have problems with accuracy, by systematically underestimating ratios [23-27]. To assess this question, we prepared a sample with proteins at known ratios. Our findings confirmed that there is systematic ratio under-estimation in iTRAQ quantitation. However, we observed a linear relationship between the observed and the expected ratio at the protein level over the four-fold range difference examined (Figure 8). Consistent with under-estimation, the gradient of the linear relationship was less than one and the under-estimation became more obvious for larger ratio changes. This effect was seen both on data collected with a Q-STAR and a QTof Premier, suggesting that the effect is ubiquitous and not dependant on the MS technology used. It has been suggested that this under-estimation arises from contaminating peptides with similar m/z ratios during ion selection prior to collision induced dissociation [23, 27]. A quantitative model reveals that when the relative amount of contamination is the same within an experiment, a linear relationship between observed and true ratios is expected, and would be independent of signal strength. We observe this in our data: RI plots at the peptide and protein level show no systematic deviations in the ratio observed with signal strength (data not shown). If the relative amount of contamination increases, the under-estimation becomes more pronounced. In fact, this was seen when the isolation width was increased in a study of iTRAQ

labelled BSA digest [23]. To investigate the effect of contamination within the ion-selection process, the selection window settings used with the QTof Premier were changed as described in the Methods section. No statistically significant difference was seen between the three settings (data not shown), in agreement with that, the quantitative model predicted that even a two-fold increase in the contamination (10 to 20%) would only result in a minor impact on the linear relationship seen (Supplementary Figure 10). We conclude that we were not able achieve such strong changes in contamination levels using the ranges of ion selection parameters we employed, suggesting that factors other than the ion selection window give rise to this effect.

The under-estimation could arise from the MS, or the protein or the sample complexity or a mixture of all three. Our results, whilst limited to two proteins changing in ratio, indicate that the under-estimation is independent of the protein. The peptides for a protein were found to be scattered randomly around the estimated ratio suggesting that a peptide specific component is not significant in the degree of under-estimation. No difference was seen between the ratios when three different amounts of sample were injected, which suggests that peptide ion abundance is not a crucial component to the degree of under-estimation. While it is conceivable that larger changes in sample complexity might trigger differences, in the system used in this study, the sample complexity was reasonably high in all cases, with utilisation of minimal prefractionation of the peptides by a single short chromatography run prior to MS analysis. Further studies to pinpoint the true source of under estimation are beyond the remit of this work.

Kuzyk *et al.* (2009) reported that an intensity dependent bias was seen at high ratio changes (≥5:1) with a Q-STAR and was possible with a LC-MALDI TOF/TOF at 10:1 ratio [26]. This

bias led to greater under-estimation. For the QToF Premier known ratio mix data, no significant intensity dependent bias was seen (Supplementary Figure 11); however an intensity dependent deviation in the ratio reported was observed in the Q-STAR data, including peptides at a ratio of 1:1 (Supplementary Figure 12). The bias was not seen in the 1:1 ratio with the *Erwinia* sample which has a more typical sample complexity and dynamic range (section 3.1.3). This issue with the Q-STAR needs further study that is beyond the scope of this manuscript, but highlights a need to be cautious with high signal intensity data which arise when relatively large amounts of a few proteins are labelled using standard protocols.

Discussion

Both accuracy and precision of measurements in quantitative analyses rely on reproducible and exact values being returned from the experiment. The iTRAQ ratio data exhibit heterogeneity of variance, where the variance is higher for low intensity signals. This is a significant problem, as low signals dominate the datasets, and in a typical iTRAQ experiment many proteins have only a few peptide readings. Furthermore, the commonly used requirement of a minimum of two peptides for confident identification of a protein results in the desire to keep as many readings as possible in an analysis. Consequently, methods which discard peptide readings below a threshold significantly limit the depth of proteins sampled in a study. Other methods, such as weighted mean or weighted regression, also aim to address the issue of heterogeneity of variance; however these methods do not work well for proteins with few peptide readings, which dominate iTRAQ studies

A two component error model consisting of an additive and multiplicative component is proposed to account for the variance structure. The presence of both components was verified with both the 4-plex and 8-plex iTRAQ tag systems, independent of analytical software and LC-MS/MS instrumentation used. The additive-multiplicative error model suggests that an appropriate data transformation will be useful, the so-called generalised logarithm (or: glog) transformation, which stabilises the variance across the entire intensity range. After such transformation, the decoupling of the variance from the signal significantly simplifies the downstream analysis as each peptide reading for a protein can be treated equally. Furthermore, it allows using low-intensity readings (rather than, say, discarding them). In data from a biological system, low intensity readings may be among the most interesting ones, when a peptide is seen at low abundance in some of the biological samples and higher in others. The price that we pay for using variance stabilisation is that ratios of small peak areas are compressed towards 1 (or, glog-ratios towards 0). This is a conservative effect and is called the "variance-bias" trade-off where a (hopefully large) improvement in precision is traded for the (hopefully small) cost of a bias. For the datasets of interest, we feel that this trade-off is justified, giving the benefit of being able to include all peptides and having robust estimates for all proteins even if few peptides are present.

The additive-multiplicative error structure has also been reported with quantitation by other MS based methodologies, and the additive component may arise from the integration of count based signal inherent with the majority of MS instrumentation [29, 31, 32] and/or the presence of a small basal unspecific background signal. As a consequence, heterogeneity of variance is, to varying degree, likely to be an inherent feature of all peptide quantitation methodologies and estimation that uses the glog transformation may play a useful role for these techniques.

The VSN software for fitting the error model and transformation parameters is available freely and with open source as a package for the statistics and programming environment R, downloadable from the Bioconductor website (http://bioconductor.org). To apply this software, all that is required are the raw reporter ion areas at the peptide level.

iTRAQ, like other MS based quantitation techniques, faces the problem of how to combine readings from multiple peptides to estimate an abundance ratio for the parent protein. Nesvizhskii and Aebersold have suggested that inconsistent relative abundance ratios from distinct peptides may point to the presence of novel biological significant forms (e.g. novel splice variants, product of protein degradation, post translational modification etc.) [44]. It is thus worth considering the distribution of readings for a protein at a peptide level. This has been incorporated into a freely available visualisation package for the R environment that compares expression changes for the peptides from the same protein [13]. In our same-same data, all differences arose purely from technical effects. Some sub-structure was identified in peptide readings where readings from a specific peptide or fraction clustered. Ideally a hierarchical process which takes a central tendency measure at each level would be used to estimate the overall protein ratio. In our opinion there are too few readings in a typical study at each level for this approach to be robust - outliers would be too influential on the result. We therefore propose the use of a trimmed mean as a robust measure of central tendency for the VSN-transformed peptide readings for a protein as these readings were found to be uni-modal in distribution with some outliers. In the case of proteins with only a few peptides a standard mean would be

calculated as there is no alternative in this situation. This can be combined with visual inspection to assess whether the assignment of peptides to a parent protein is appropriate.

The simplest iTRAQ experiment is a pair-wise comparison between sample types looking for changes above a threshold determined from experimental variation assessed by looking at same-same comparison. For both raw and log transformed data, the threshold is difficult to determine as it should have an intensity dependent element. This is complicated even more by the fact that on the protein level, the estimated protein ratios are obtained from peptides at various intensities; consequently the majority of current methodologies fail to consider this problem. With the VSN transformation this intensity dependence is removed and 90% and 95% thresholds were found to lead to reproducible results across tag combinations. The thresholds varied with sample type but were low and indicated the sensitivity of the technology to expression changes. In practice, of course, the experimenter will use larger thresholds that also take into account biological variation. The thresholds reported here is not intended as a universal benchmark, and the reality is that for each new system (be it MS, chromatography or sample) a new same-same sample study should be run. If the compared samples are such that *most* protein abundances are the same across samples, then the distribution of observed glog-ratios can also be used to set the significance threshold. A threshold methodology was applied to a biological study and the iTRAQ findings were in keeping with those published for this system.

Compared to the previous iTRAQ data processing methodologies, we showed that the VSN processed data are more precise and sensitive to detecting changes. The advantages of the VSN methodology will be greatest in situations where hypothesis tests are used to detect changes in

expression. Such tests are particularly useful in studies which include biological replicates to ensure the differences highlighted arise from a treatment difference rather than from a sampling effect. Underlying the more powerful hypothesis tests are assumptions such as normality and homogeneity of variance, which tend to be more appropriate with the VSN transformed data.

The study on a sample with known ratios in two independent MS systems confirmed that the iTRAQ technology does have an accuracy problem: ratios tend to be under-estimated. The experiments here, spanning a one to four fold ratio, suggest that this effect is independent of signal strength and leads to a linear relationship between the observed and the expected ratio which goes through the origin. Data modelling supports the suggestion of Bantscheff *et al* [23] that this under-estimation arises from contamination in the precursor ion selection process and indicates that a linear relationship would be obtained when the proportion of contamination is consistent within an experiment. With this linear relationship a single correction factor can be calculated to adjust for this under-estimation from readings of known proteins which span a range of expected ratios. We, therefore, recommend that for a typical sample, a similar experiment to that described here is carried out and a gradient value estimated from the linear relationship is employed as a correction factor for their system. Alternatively, if sample complexity is thought to influence this relationship, we envisage a kit could be developed which consists of a mixture of proteins at known ratios which are added to samples prior to iTRAQ labelling, and which would allow the calculation of the correction factor.

To support further development in data analysis, raw data for an example same-same study (ErwiniaC), the yeast study, and the spiked study are downloadable from the PRIDE database

[45], and the PRIDE identification numbers are provided in supplementary information. Excel spread sheets including the quantitation data for both raw and normalized are also available in supplementary information.

In summary, this manuscript proposes methodologies to address the precision and accuracy limitations of iTRAQ. The accuracy issue, arising from contamination during precursor ion selection specific to MS/MS quantitation, can be addressed by calculation of a correction factor from spiked samples, whilst the precision issue can be addressed by the VSN transformation. This then allows a robust estimation of the ratio at the protein level as all peptides have near equivalent precision. Together these methodologies will allow iTRAQ to provide robust quantitative data even when a protein is quantified from only 2 peptides. The potential application of the VSN method in MS studies is not restricted to iTRAQ quantitation or even to proteomics, as many MS based applications have reported precision problems related to heterogeneity of variance.

Acknowledgements

This work was supported by a BBSRC grant (BB/C50694/1), which funded Dr. N.A. Karp as a BBSRC Research Associate. We would like to thank Ian Foulds and George Salmond for provision of *Erwinia carotovora* samples, and June Petty, Juan Castrillo and Stephen Oliver for the provision of the *Saccharomyces cerevisiae* data set. We would also like to thank Gary Woffendin and Michaela Scigelova from Thermo Finnigan for running samples on the LTQ-Orbitrap XL and for help with data analysis.

4. References

1. Fievet, J., et al., *Assessing factors for reliable quantitative proteomics based on two-dimensional gel electrophoresis.* Proteomics, 2004. **4**(7): p. 1939-49.
2. Smejkal, G.B., M.H. Robinson, and A. Lazarev, *Comparison of fluorescent stains: relative photostability and differential staining of proteins in two-dimensional gels.* Electrophoresis, 2004. **25**(15): p. 2511-9.
3. Yan, J.X., et al., *Fluorescence two-dimensional difference gel electrophoresis and mass spectrometry based proteomic analysis of Escherichia coli.* Proteomics, 2002. **2**(12): p. 1682-98.
4. Hu, Y., et al., *Proteome analysis of Saccharomyces cerevisiae under metal stress by two-dimensional differential gel electrophoresis.* Electrophoresis, 2003. **24**(9): p. 1458-1470.
5. Gygi, S.P., et al., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.* Nat Biotechnol, 1999. **17**(10): p. 994-9.
6. Zhou, H., et al., *Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry.* Nat Biotechnol, 2002. **20**(5): p. 512-5.
7. Yao, X., et al., *Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus.* Anal Chem, 2001. **73**(13): p. 2836-42.
8. Everley, P., et al., *Quantitative cancer proteomics: stable isotope labeling with amino acids in cell culture (SILAC) as a tool for prostate cancer research.* Mol Cell Proteomics, 2004. **3**(7): p. 729-735.

9. Wu, W.W., et al., *Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF*. J Proteome Res, 2006. **5**(3): p. 651-8.

10. Wolff, S., et al., *Gel-free and gel-based proteomics in Bacillus subtilis: a comparative study*. Mol Cell Proteomics, 2006. **5**(7): p. 1183-92.

11. Dayon, L., et al., *Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags.* Anal Chem, 2008. **80**(8): p. 2921-31.

12. Thompson, A., et al., *Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS.* Anal Chem, 2003. **75**(8): p. 1895-904.

13. Choe, L., et al., *8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease.* Proteomics, 2007. **7**(20): p. 3651-60.

14. Ow, S.Y., et al., *Quantitative shotgun proteomics of enriched heterocysts from Nostoc sp. PCC 7120 using 8-plex isobaric peptide tags.* J Proteome Res, 2008. **7**(4): p. 1615-28.

15. Redding, A.M., et al., *Study of nitrate stress in Desulfovibrio vulgaris Hildenborough using iTRAQ proteomics.* Brief Funct Genomic Proteomic, 2006. **5**(2): p. 133-43.

16. Hu, J., et al., *Optimized proteomic analysis of a mouse model of cerebellar dysfunction using amine-specific isobaric tags.* Proteomics, 2006. **6**(15): p. 4321-34.

17. Li, X.J., et al., *Automated statistical analysis of protein abundance ratios from data generated by stable-isotope dilution and tandem mass spectrometry.* Anal Chem, 2003. **75**(23): p. 6648-57.

18. Choe, L.H., et al., *A comparison of the consistency of proteome quantitation using two-dimensional electrophoresis and shotgun isobaric tagging in Escherichia coli cells.* Electrophoresis, 2005. **26**(12): p. 2437-49.

19. Boehm, A.M., et al., *Precise protein quantification based on peptide quantification using iTRAQ.* BMC Bioinformatics, 2007. **8**: p. 214.

20. Unwin, R.D., et al., *Quantitative proteomic analysis using isobaric protein tags enables rapid comparison of changes in transcript and protein levels in transformed cells.* Mol Cell Proteomics, 2005. **4**(7): p. 924-35.

21. Gan, C.S., et al., *Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ).* J Proteome Res, 2007. **6**(2): p. 821-7.

22. Lin, W.T., et al., *Multi-Q: a fully automated tool for multiplexed protein quantitation.* J Proteome Res, 2006. **5**(9): p. 2328-38.

23. Bantscheff, M., et al., *Robust and Sensitive iTRAQ Quantification On an LTQ Orbitrap Mass Spectrometer.* Mol Cell Proteomics, 2008. **7**(9): p. 1702-1713.

24. DeSouza, L.V., et al., *Absolute quantification of potential cancer markers in clinical tissue homogenates using multiple reaction monitoring on a hybrid triple quadrupole/linear ion trap tandem mass spectrometer.* Anal Chem, 2009. **81**(9): p. 3462-70.

25. Keshamouni, V.G., et al., *Differential protein expression profiling by iTRAQ-2DLC-MS/MS of lung cancer cells undergoing epithelial-mesenchymal transition reveals a migratory/invasive phenotype.* J Proteome Res, 2006. **5**(5): p. 1143-54.

26. Kuzyk, M.A., et al., *A comparison of MS/MS-based, stable-isotope-labeled, quantitation performance on ESI-quadrupole TOF and MALDI-TOF/TOF mass spectrometers.* Proteomics, 2009. **9**(12): p. 3328-3340.

27. Ow, S., et al., *iTRAQ Underestimation in Simple and Complex Mixtures:"The Good, the Bad and the Ugly".* Journal of Proteomic Research, 2009. **Early release**.

28. Anderle, M., et al., *Quantifying reproducibility for differential proteomics: noise analysis for protein liquid chromatography-mass spectrometry of human serum.* Bioinformatics, 2004. **20**(18): p. 3575-82. Epub 2004 Jul 29.

29. Rocke, D.M. and A. Lorenzato, *A two-component model for measurement error in analytical chemistry.* Technometrics, 1995. **37**(2): p. 176-184.

30. Du, P., et al., *A noise model for mass spectrometry based proteomics.* Bioinformatics, 2008. **24**(8): p. 1070-7.

31. Pavelka, N., et al., *Statistical similarities between transcriptomics and quantitative shotgun proteomics data.* Mol Cell Proteomics, 2008. **7**(4): p. 631-44.

32. Enke, C., *The science of chemical analysis and the technique of mass spectometry.* International journal of mass spectrometry, 2001. **212**(1-3): p. 1-11.

33. Rocke, D.M. and B. Durbin, *A model for measurement error for gene expression arrays.* J Comput Biol, 2001. **8**(6): p. 557-69.

34. Huber, W., et al., *Variance stabilization applied to microarray data calibration and to the quantification of differential expression.* Bioinformatics, 2002. **18**: p. 96S-104S.

35. Hahne, F., et al., *Bioconductor Case Studies (Use R)*. 2008: Springer

36. Huber, W., et al., *Parameter estimation for the calibration and variance stabilization of microarray data.* Stat Appl Genet Mol Biol, 2003. **2**: p. Article3.

37. Huber, W., et al., *Variance stabilization applied to microarray data calibration and to the quantification of differential expression.* Bioinformatics, 2002. **18 Suppl 1**: p. S96-104.

38. Aitchison, J. and J. Brown, *Lognormal distribution: with special references to its uses in economics*. 1st ed. ed. 1957, Cambridge: Cambridge University Press. 194.

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

40. Chen, X., et al., *Amino acid-coded tagging approaches in quantitative proteomics.* Expert Rev Proteomics, 2007. **4**(1): p. 25-37.

41. DeSouza, L., et al., *Search for cancer markers from endometrial tissues using differentially labeled tags iTRAQ and cICAT with multidimensional liquid chromatography and tandem mass spectrometry.* J Proteome Res, 2005. **4**(2): p. 377-86.

42. Castrillo, J.I., et al., *Growth control of the eukaryote cell: a systems biology study in yeast.* J Biol, 2007. **6**(2): p. 4.

43. Hartig, A., et al., *Differentially regulated malate synthase genes participate in carbon and nitrogen metabolism of S. cerevisiae.* Nucleic Acids Res, 1992. **20**(21): p. 5677-86.

44. Nesvizhskii, A.I. and R. Aebersold, *Interpretation of shotgun proteomic data: the protein inference problem.* Mol Cell Proteomics, 2005. **4**(10): p. 1419-40.

45. Jones, P., et al., *PRIDE: new developments and new datasets.* Nucleic Acids Res, 2008. **36**(Database issue): p. D878-83.

Figure legends

Figure 1: Ratio-intensity plots (RI plot) for the log-ratio of 115 to 114 reporter ions against the average reporter ion signal in the *Erwinia* same-same dataset B. These plots are used to assess the distribution of ratios as a function of signal strength. In panels A and B, the *y*-axis shows the logarithm (base 2) of the ratios; in panel A, the *x*-axis is proportional to the mean and in panel B to the rank of the mean. Choosing the rank of the mean for the abscissa distributes the data points evenly along the *x*-axis and helps with the visual assessment of distribution width; when the *x*-axis is simply the mean, the uneven distribution of the data along the *x*-axis range can confound the visual assessment. In panels C and D, VSN-ratios (or generalised log-ratios, see Section 3.2.1) are shown on the y-axis; in panel C the *x*-axis shows the mean and in panel D, the rank of the mean. For the display of panels A and B (logarithm transformation), data were filtered to remove zero and negative values, whereas all data are shown in panels C and D (VSN transformation).

Figure 2: The relationship between the logarithm (base 2) of the mean signal and the logarithm (base 2) of its variance. The solid line shows a smooth regression line calculated by local polynomial regression. Shown is *Erwinia* dataset B. No prior intensity-based data filtering was performed.

Figure 3: CV versus signal intensity comparison for the log2 (blue) and VSN (black) transformed data for *Erwinia* dataset B. No prior intensity-based data filtering was performed.

Figure 4: Example Normal Q-Q plot as a typical distribution profile for a peptide from the phosphorylase B dataset with standard i-Tracker low volume filtering.

Figure 5: Comparison of the CV versus mean behaviour for the log2 versus VSN transformed data at the protein level when a trimmed mean (20%) approach was used. The solid line show a moving average calculated with a local polynomial regression. One data point from the log

dataset has been removed for visual clarity as it had a high CV of 1.5. This analysis was completed on the *Erwinia* dataset B where no intensity-based data filtering had been performed beyond the peptides being unique for a protein and the peptide being confident in its assignment to a protein.

Figure 6: Example of reproducibility of the protein VSN ratio in a same-same experiment across the various tag combination. The protein ratio has been estimated by calculating a 20% trimmed mean using all the unique peptide readings for a protein. This analysis was completed on the *Erwinia* dataset B where no filtering was done beyond the peptides being unique for a protein and confidence of the peptide to protein assignment. The tag combinations are represented as follows: 115-114 by a dot, 116-114 by a triangle, 117-114 by a cross, 116-115 by an X, 117-115 by a diamond and 117-116 by an upside-down triangle.

Figure 7: A box-percentile plot comparing the protein ratio distribution across various sample comparisons where the protein ratio has been calculated as a 20% trimmed mean from the peptides contributing to that protein. A: For VSN transformed data and B: For log2 transformed data with ratiometric normalisation. X versus Y indicates that sample X values have been divided by sample Y values.

Figure 8: The observed versus expected protein ratio for VSN normalised data from the known ratio samples processed with the Q-STAR. The dotted line indicates the equivalence relationship between observed and expected. No intensity-based data filtering was performed.

Tables

Study Type	Sample Type (dataset name)	iTRAQ system	LC- MS/MS system	Quantitation system
Same-Same	*Erwinia (ErwiniaB)*	4-plex	Q-STAR	iTRACKER Mascot
	Erwinia (ErwiniaC)	4-plex	Q-STAR	iTRACKER Mascot
	Erwinia Phosphorylase B	4-plex	OrbitrapXL	Mascot
		8-plex	QSTAR	Mascot
			OrbitrapXL	Mascot
Known ratio	Proteins at known ratios	4-plex	Q-STAR Premier	iTRACKER
Biological unknown	Yeast grown under nutritional limiting	4-plex	Q-STAR	iTRACKER

Table 1: A summary of the various datasets used within this manuscript.

Protein	114	115	116	117
BSA	1	2	3	4
CytC	1	1	1	1
Enolase	1	1	1	1
PhosphorylaseB	4	3	2	1

Table 2: A breakdown of the proteins included in the known ratio preparation and the level of these protein for each TRAQ reporter tag.

Dataset	Percentile	Mean log ratio (VSN1-VSN2)	95% confidence intervals for the mean log ratio		Mean fold change	95% confidence intervals for mean fold change	
			Lower	Upper		Lower	Upper
Erwinia B	2.5	-0.15	-0.16	-0.13	0.90	0.89	0.91
	5	-0.12	-0.13	-0.10	0.92	0.92	0.93
	95	0.11	0.11	0.12	1.08	1.08	1.09
	97.5	0.14	0.14	0.15	1.11	1.10	1.11
Erwinia C	2.5	-0.12	-0.13	-0.11	0.92	0.91	0.93
	5	-0.09	-0.10	-0.09	0.94	0.93	0.94
	95	0.09	0.09	0.09	1.06	1.06	1.06
	97.5	0.11	0.11	0.12	1.08	1.08	1.09

Table 3: For both same-same *Erwinia* datasets, the table details the average experimental thresholds as a VSN ratio and a fold change that would encompass either 95 or 90% of the naturally occurring technical variation. The average threshold was calculated for each dataset as with four tags there are seven possible pair-wise comparisons and a 95% confidence interval reported to give a measure of consistency. This analysis was completed where no filtering had occurred beyond the peptides being unique for a protein and the peptide being confident in its assignment to a protein.

	Log2 transformed		VSN transformed	
Percentile position	Protein ratio	Fold change	Protein ratio	Fold change
2.5	-0.37	0.77	-0.25	0.84
5	-0.29	0.82	-0.2	0.87
95	0.27	1.2	0.2	1.15
97.5	0.33	1.26	0.24	1.18

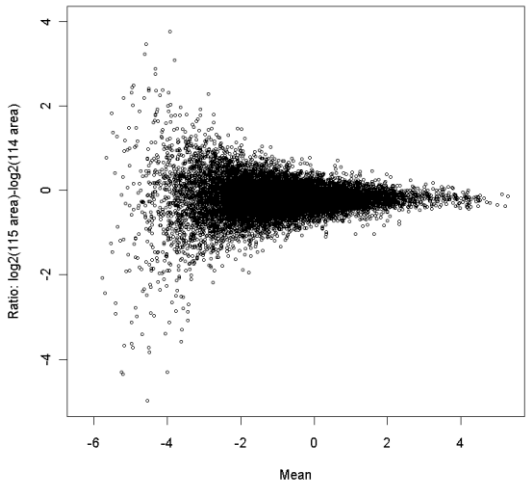
Table 4: Thresholds calculated from the same-same sample comparison of the pooled sample in the yeast study for various percentile positions for both log and VSN transformed data.

	Number of proteins		Protein identified as having significant changes in expression (%)			
			Log2		VSN	
Samples compared	VSN	LOG	Up-regulated	Down-regulated	Up-regulated	Down-regulated
Nitrogen limited versus carbon limited	1042	1038	7.10	17.18	14.26	22.74
Carbon limited versus sulphate limited	1042	1040	21.02	11.13	25.29	14.23

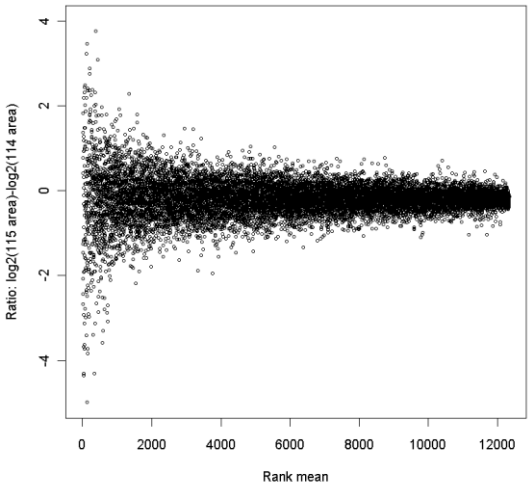
Nitrogen limited versus sulphate limited	1042	1036	7.10	4.51	5.89	5.41
Sulphate limited versus phosphate limited	923	923	2.71	1.63	0.29	0.18

Table 5: Percentage of proteins identified as having significant changes in expression between the various samples compared when 97.5 and 2.5% threshold were used.

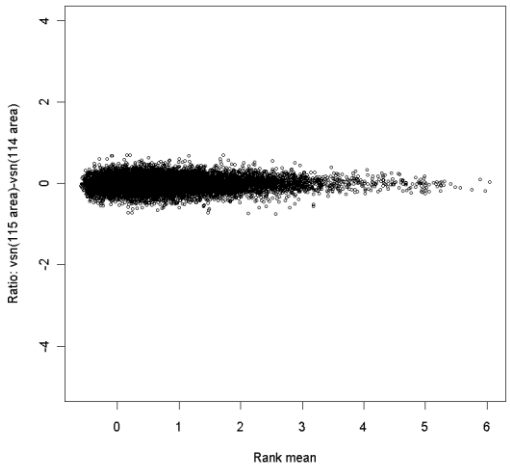
Figures



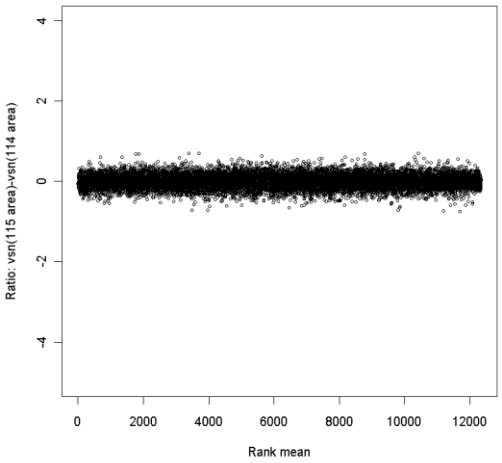
A



B



C



D

Figure 1: Ratio-intensity plots (RI plot) for the log-ratio of 115 to 114 reporter ions against the average reporter ion signal in the *Erwinia* same-same dataset B. These plots are used to assess the distribution of ratios as a function of signal strength. In panels A and B, the *y*-axis shows the logarithm (base 2) of the ratios; in panel A, the *x*-axis is proportional to the mean and in panel B to the rank of the mean. Choosing the rank of the mean for the abscissa distributes the data points evenly along the *x*-axis and helps with the visual assessment of distribution width; when the *x*-axis is simply the mean, the uneven distribution of the data along the *x*-axis range can confound the visual assessment. In panels C and D, VSN-ratios (or generalised log-ratios, see Section 3.2.1) are shown on the y-axis; in panel C the *x*-axis shows the mean and in panel D, the rank of the mean. For the display of panels A and B (logarithm transformation), data were filtered to remove zero and negative values, whereas all data are shown in panels C and D (VSN transformation).

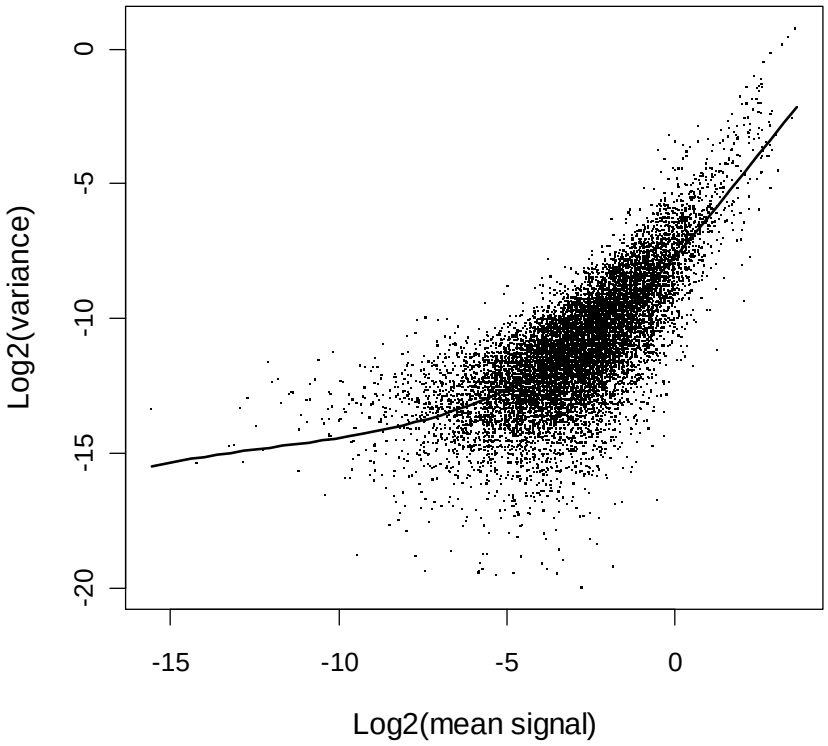


Figure 2: The relationship between the logarithm (base 2) of the mean signal and the logarithm (base 2) of its variance. The solid line shows a smooth regression line calculated by local polynomial regression. Shown is *Erwinia* dataset B. No prior intensity-based data filtering was performed.

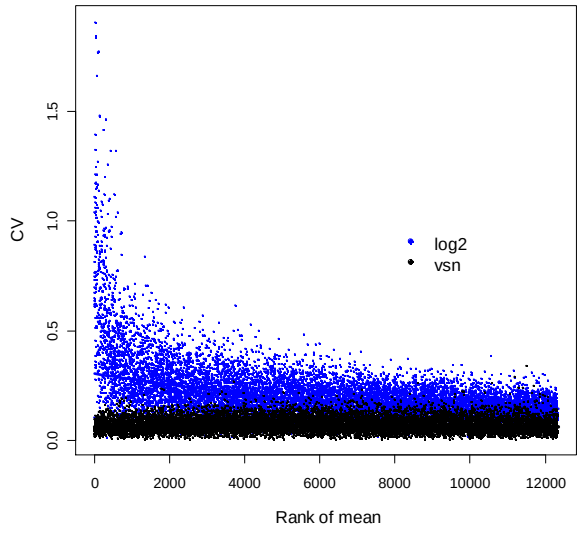


Figure 3: CV versus signal intensity comparison for the log2 (blue) and VSN (black) transformed data for *Erwinia* dataset B. No prior intensity-based data filtering was performed.

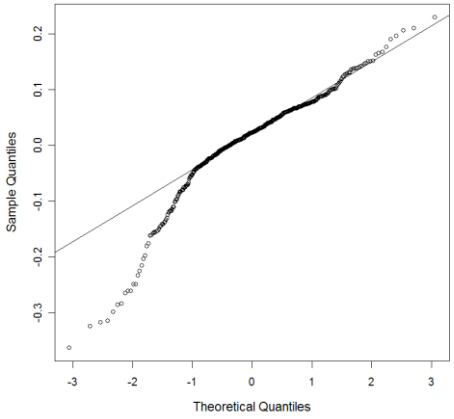


Figure 4: Example Normal Q-Q plot as a typical distribution profile for a peptide from the phosphorylase B dataset with standard i-Tracker low volume filtering.

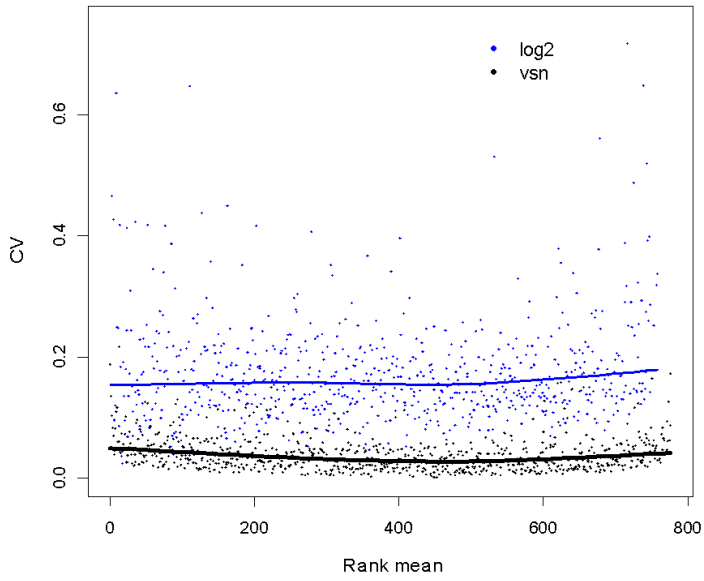


Figure 5: Comparison of the CV versus mean behaviour for the log2 versus VSN transformed data at the protein level when a trimmed mean (20%) approach has been used. The solid line show a moving average calculated with a local polynomial regression. One data point from the log dataset has been removed for visual clarity as it had a high CV of 1.5. This analysis was completed on the *Erwinia* dataset B where no intensity-based data filtering had been performed beyond the peptides being unique for a protein and the peptide being confident in its assignment to a protein.

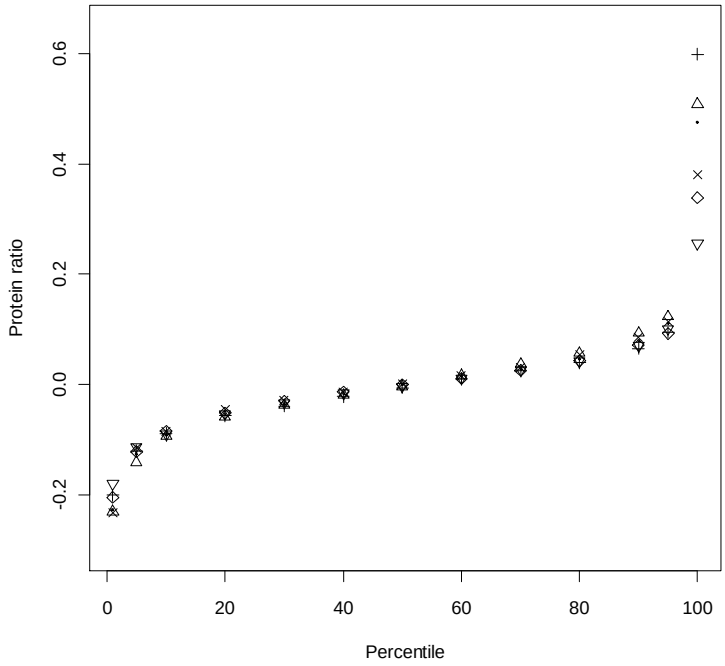


Figure 6: Example of reproducibility of the protein VSN ratio in a same-same experiment across the various tag combination. The protein ratio has been estimated by calculating a 20% trimmed mean using all the unique peptide readings for a protein. This analysis was completed on the *Erwinia* dataset B where no filtering was done beyond the peptides being unique for a protein and confidence of the peptide to protein assignment. The tag combinations are represented as follows: 115-114 by a dot, 116-114 by a triangle, 117-114 by a cross, 116-115 by an X, 117-115 by a diamond and 117-116 by an upside-down triangle.

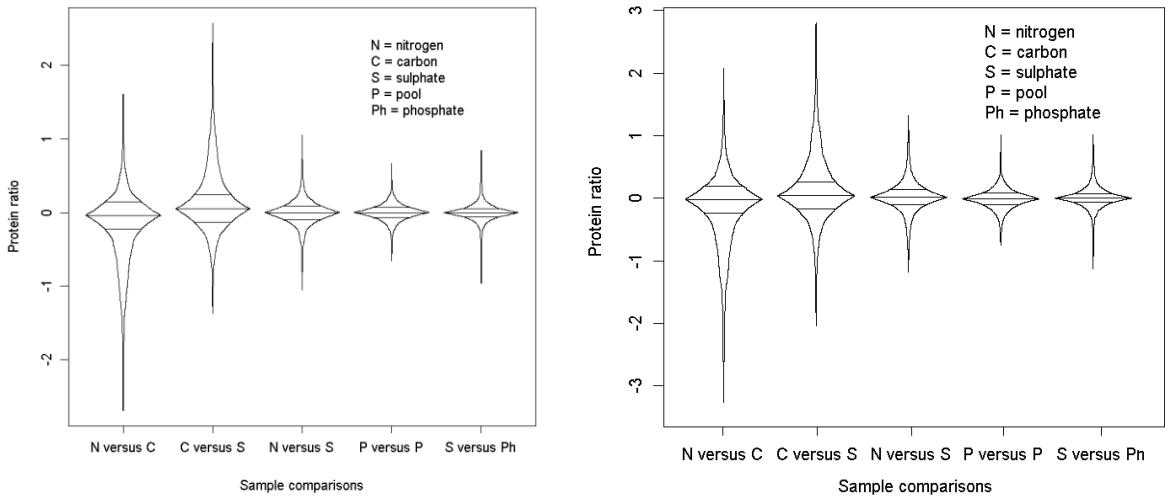


Figure 7: A box-percentile plot comparing the protein ratio distribution across various sample comparisons where the protein ratio has been calculated as a 20% trimmed mean from the peptides contributing to that protein. A: For VSN transformed data and B: For log2 transformed data with ratiometric normalisation. X versus Y indicates that sample X values have been divided by sample Y values.

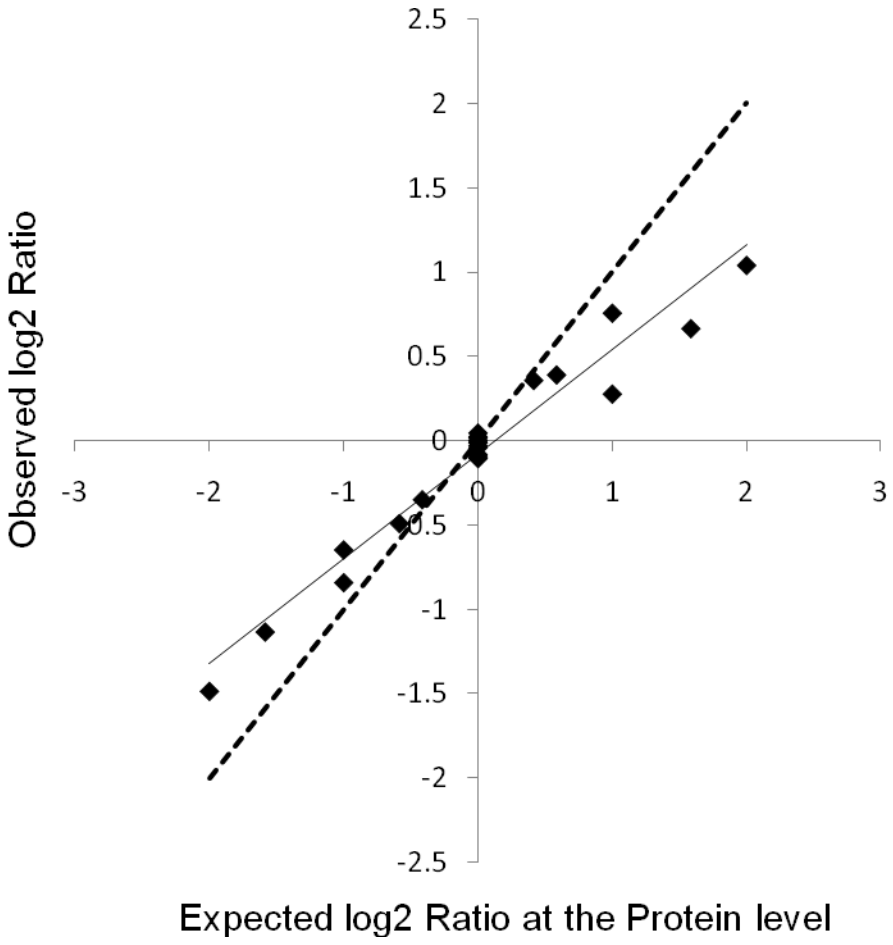


Figure 8: The observed versus expected protein ratio for VSN normalised data from the known ratio samples processed with the Q-STAR. The dotted line indicates the equivalence relationship between observed and expected. No intensity-based data filtering was performed.