Bayesian Confidence Intervals for Multiplexed Proteomics Integrate Ion-Statistics with Peptide Quantification Concordance

Leonid Peshkin,[1] Lillia Ryazanova,[2] Martin Wühr[2, #]

1) Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA
2) Department of Molecular Biology & the Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA

wuhr@princeton.edu

Multiplexed proteomics has emerged as a powerful tool to measure protein expression levels across multiple conditions. The relative protein abundances are inferred by comparing the signal generated by isobaric tags, which encode the samples' origins. Intuitively, the trust associated with a protein measurement depends on the similarity of ratios from different peptides and the signal level of these measurements. Up to this point in the field, peptide-level information has not typically been integrated into confidence, and only the most likely results for relative protein abundances are reported. If confidence is reported, it is based on protein-level measurement agreement between replicates. Here we present a mathematically rigorous approach that integrates peptide intensities and peptide-measurement agreement into confidence intervals for protein ratios (BACIQ). The main advantages of BACIQ are: 1) it removes the need to threshold reported peptide signal based on an arbitrary cut-off, thereby reporting more measurements from a given experiment; 2) confidence can be assigned without replicates; 3) for repeated experiments BACIQ provides confidence intervals for the union, not the intersection, of quantified proteins; 4) for repeated experiments, BACIQ confidence intervals are more predictive than confidence intervals based on protein measurement agreement. Therefore, our method drastically increases the value obtained from quantitative proteomics experiments and will help researchers to interpret their data and prioritize resources. To make our approach easily accessible we distribute it via an R/Stan package.

Introduction

Mass spectrometry based proteomics has undergone a remarkable revolution and is now able to identify ~10,000 proteins in a single experiment (Beck et al., 2015; Deshmukh et al., 2015; Wühr et al., 2014). However, due to the difficulty in predicting ionization efficiency of peptides during electrospray, the signal measured in the mass spectrometer is not a direct readout for a peptide concentration in a sample. Proteomics is well suited to comparing the abundance change of the same peptides/proteins among multiple conditions. In so-called label-free proteomics, the peptide signal is compared between multiple different runs and changes of ~2-fold can be detected as significant (Cox et al., 2014). Even smaller relative protein abundance changes can be detected by encoding multiple conditions with heavy isotopes and analyzing the samples simultaneously. In MS1 based approaches like SILAC the different conditions contain different numbers of heavy isotopes and conditions are encoded by differing peptide masses. However, due to the increase in complexity of the MS1 spectrum with more conditions this approach is only feasible for up to three conditions (Hilger and Mann, 2012). A major breakthrough for proteomics was the introduction of isobaric tags (Thompson

et al., 2003). These tags, which are chemically attached to the peptides, act as barcodes for the different conditions (e.g. replicates, perturbations, or time-points). Each tag has exactly the same mass and only upon fragmentation are the distinct reporter ions released. Due to the identical mass, the MS1 spectrum complexity does not increase in its complexity with more conditions and currently up to 11 conditions can be compared in a single experiment (Jiang et al.). Initially, the co-isolation and co-fragmentation of other peptides led to major artifacts. However, more recently these artifacts have been overcome with the introduction of MultiNotch MS3 (TMT-MS3), QuantMode, and the complement reporter ion approach (TMTc+)(McAlister et al., 2014; Sonnett et al., 2017; Ting et al., 2011; Wenger et al., 2011; Wühr et al., 2012). With these methods, data of superb quality can be generated and changes of less than 10% can be quantified as significant (Wühr et al., 2015).
Despite these impressive capabilities of quantitative multiplexed proteomics, a remarkable shortcoming is the lack of confidence assigned to these measurements. Typically, only the most likely protein ratios are reported. Various factors can distort the measurements: peptide-to-spectra matching uncertainty, enzymatic digestion efficiency, post-translational modifications, and interference (Ting et al., 2011). Generally, there is no sense of how much we can trust the data (Ning et al., 2016). Noise models have been presented to handle peptide-to-protein aggregation in label-free setting. However, the very different nature of multiplexed proteomics data compared to label-free data makes these approaches not easily transferable (Choi et al., 2014; Clough et al., 2012; Zhang et al., 2017). Most proteins are measured via multiple peptide quantification events. Intuitively, one should be able to use both the agreement between quantifications of peptides assigned to a protein and the measured intensities, which are proportional to the number of ions, to assign confidence. Figure S1 summarizes the challenge to express confidence for multiplexed proteomics measurements. However, to our knowledge there is currently no way to integrate all this information in order to express the confidence of protein level quantification. Assigning confidence is important because it allows one to assess significance of changes and enables researchers to prioritize valuable time and resources in follow-up experiments. In a previous study the measurement agreement between peptides assigned to a protein were considered, but the underlying ion-statistics were ignored (Koh et al., 2015). With replicate experiments confidence can be calculated with standard approaches like the t-test or ANOVA (McAlister et al., 2014; Oberg and Mahoney, 2012; Student, 1908). For these approaches protein level measurements typically weigh peptides by ion-statistics but ignore the underlying agreement between peptide measurements. For confidence expressed based on replicate protein-level measurements, the confidence of the measurement can obviously only be expressed for the intersection of protein sets measured in all repeated experiments. Moreover this approach may lead to unwarranted high confidence when multiple experiments have wrong but concordant measurements when each experiment ignores the disagreement at the peptide level. Also, peptides which are measured below an arbitrary level are ignored (Lapek et al., 2017; McAlister et al., 2014).
In this paper we propose a novel mathematically rigorous method for computing and representing the uncertainty of quantitative multiplexed proteomics measurements. Our method is based on a hierarchical Bayesian model of a multiple condition Dirichlet-Multinomial distribution (Beta-Binomial for the two-condition scenario). We call our method BACIQ (**B**ayesian **A**pproach to **C**onfidence **I**ntervals for protein **Q**uantitation). Our approach allows us to represent uncertainty for both individual peptides as well as multi-peptide proteins and takes into account every quantified peptide, regardless of the signal level, thereby increasing the sensitivity of proteomics measurements. Our methods do not require multiple repeated experiments, but if such repeats are available, it integrates the results

providing the output and confidence for the union (not intersection) of all separately measured proteins. The source code was deposited via GitHub https://github.com/Peshkin/BACI-Q

Results

Peptide measurements map to coin-flips

To reason about measurement confidence in multiplexed proteomics, let us begin by discussing the measurement process of a peptide that is labeled with isobaric tags encoding two different conditions (case and control) (Fig. 1A). With multiplexed proteomics we can measure the relative abundance of a peptide between multiple conditions. For the sake of simplicity, we will discuss the two condition case for most of the paper, but all our approaches and the provided software can be generalized to the multi-condition case. Once peptides are labeled with isobaric tags they are combined into one test tube, in which we have a “true ratio” of peptide abundances across samples. The aim of the proteomic experiment then is to recover this “true peptide ratio”. During MS analysis, the peptides get ionized and fragmented. Upon fragmentation, respective fragments of the isobaric tag are released, encoding the different conditions. The relative abundance of the ions encoding the different conditions is used to quantify the relative abundance of the peptides (Fig. 1B). However, the limited number of ions by which measurements are performed will introduce measurement errors due to ion-statistics (Fig. 1C). The error naturally tends to be higher for measurements with lower mass spectrometer signal, which is proportional to the ion-count. More precisely, throughout this paper we will refer to “signal over Fourier transform-noise” acquired in the Orbitrap as "MS-signal". This MS-signal can be extracted from raw files acquired on Orbitrap instruments by dividing the raw signal through the Fourier transform noise. Unlike the raw signal, which is an approximation for the ion-flux, the “signal over Fourier transform-noise” is proportional to the number of ions measured in the Orbitrap (Makarov and Denisov, 2009). The key to reflecting confidence in an estimate of a fraction is in capturing precisely the functional form of the convergence of the measurement to the true ratio as the MS-signal gets higher.

Conversion of Mass Spectrometer signal to charges

To express confidence intervals for peptide ratio measurements we have to be able to convert the MS-signal into the number of ions. To this end, we generated a sample in which all peptides are labeled with the identical 1:1 ratio. We observe that the ratio of the signals in two channels of the mass spec instrument converges to an asymptotic value, just like the fractional outcome of a sequence of coin-tosses converges to a true ratio with the number of tosses. Figure 2A presents a scatter plot where each point represents a single peptide from a dataset of a total of 10534 peptides. The functional form of this dependency for a coin toss represented by a Binomial distribution for the coefficient of variation is $\sqrt{(1-p)/np}$, where ***p*** is the ratio and ***n*** the number of coin-tosses. In order to treat the MS-signal as an outcome of a binomial process, we need to find a conversion factor from a continuous MS-signal into equivalents of “coin flips”. We fit a single parameter ***m*** as a multiplier to an MS-signal ***s*** where ***n = ms*** to the binned data (Fig. 2B). When we perform this analysis on an Orbitrap Lumos with 50K mass resolution, we observe a convertion factor of 2.0. When we performed the equivalent experiments on different instruments and resolutions, we observed different values (Table 1). Makarov et al. previously reported that this conversion factor should scale inversely with the square-root of the Orbitrap resolution (Makarov and Denisov, 2009). Our measurements are in rough agreement with this prediction; when we increase resolution from 15K to 120K on the Lumos we expect the conversion factor to reduce by 2.8-fold, we observe a 3.5 decrease. Based on previous

reports by the same paper, we likely underestimate the number of actual ions by a small factor. Nevertheless, the good fit to the data (Fig. 2B) indicates that the conversion into pseudo-counts allows us to model the relationship between mass spectrometer signal and measurement noise due to ion-statistics and other noise occurring during data acquisition. For a limited number of cases, we have repeated these measurements for various instruments of the same model and obtained very similar results suggesting that for a given instrument model and resolution the conversion factors are invariant.

Assigning confidence intervals for individual peptides

With the conversion factor at hand the ratio estimation is reduced to a well-studied case of an estimation for the Binomial probability of success (Fig. 1) (Gelman et al., 2014). Specifically, the confidence interval for this case is obtained from a Beta distribution (Fig. S2). Figure 2C illustrates the likelihood function of the probability of success parameter of the Binomial distribution $\boldsymbol{q}$ for three peptides at the different levels of total MS-signal as color-coded in Figure 2A. A higher signal gives a tighter distribution.

We next verify that the confidence intervals obtained agree with observations. Computing the 95% confidence intervals we expect that the true answer will lie outside of the confidence interval approximately 5% of the time and will be symmetrically split between over- and under-estimation. Indeed Figure 2D shows that we overestimate 1.98% of the time and under-estimate for 2.50% of the time.

Only considering ion-statistics produces inadequate confidence intervals on the protein level

So far, we have shown that we can adequately express the confidence intervals for peptide measurements. If ion-statistics was the only source of noise, we could sum up all counts from peptides mapped to a protein and express confidence intervals at the protein level. This approach works well for the synthetic case, where all peptides in a mixture were labeled together and show the exactly same ratio (Fig. S3). However, we were worried that in real experiments, other factors like differences in digestion efficiency, labeling problems, erroneous peptide-to-protein assignment, post-translational modifications, chemical interference, and so forth might produce significant additional noise. To test whether only considering ion-statistics is valid for multiplexed proteomics measurements, we revisited our recent publication of nucleocytoplasmic partitioning in the frog oocyte (Wühr et al., 2015) (Fig. 3A). Because in that application we don't know the true answers for proteins, we can't prove that our approach works, but if peptide measurements disagree with each other we can learn that the method is inadequate for the protein level. Indeed, when we evaluate all the peptides mapped to a protein, we observe that their probability distribution often nearly completely exclude each other and are unlikely to come from the integrated confidence interval at the protein level (Fig. 3B). This suggests that besides ion-statistics other significant sources of error contribute to the proteomics measurements and we have to take these sources into account to adequately express confidence intervals at the protein level.

Confidence-intervals at the protein level, that integrate ion-statistics and agreement between peptides mapped to the same protein

Analogously to the single peptide case, let us review the entire proteomics experiment that starts with a "true protein ratio" between the two samples we analyze. We represent the multi-peptide protein case as a generative two-level hierarchical Bayesian model (Fig. 4). The protein is digested into

peptides, which are labeled with isobaric tags. During this process noise can be introduced e.g. with differing digestion or labeling efficiencies between case and control. We represent the first step of the Bayesian model by sampling peptide ratios q_i for the constituent peptides of a given protein from the Beta distribution parameterized according to the “true protein ratio” (Fig. 4A). The second step is sampling the ion statistics for each peptide separately, which is equivalent to the single peptide case we discussed above and in Figure 1. Importantly, different peptides are measured with different MS-signal and therefore with different confidence on the underlying “true peptide ratio” (Fig. 4B).
The target of our estimation is the mean of the Beta-distribution representing the protein, specifically the posterior distribution of that value as the representation of the uncertainty. The entire sampling process maps nicely to the generative process of a well-studied Beta-Binomial distribution (Gelman et al., 2014). Nevertheless, the maximum-likelihood estimate of these distributions is not available in a closed form (Minka, 2000). One approach would be to numerically search for an MLE and estimate the curvature of the likelihood function from the Hessian at the MLE using the asymptotic Normality (Lehmann and Casella, 2006). However unfortunately this approach turned out to be numerically not robust (not shown). A robust alternative approach is accomplished using Monte-Carlo Markov Chain (MCMC) methods implemented in statistical inference language Stan (see Supporting Information)(Carpenter et al., 2016). Naively, it consists of exploring the space of possible protein ratios, computing the likelihood of observed peptide data given a guess at the protein ratio and assembling a large set of plausible ratio samples to use its histogram as posterior likelihood representation. Naturally, such processes are computationally expensive.

Validation of protein level confidence interval estimation

To validate our approach, we produced a standard containing peptides with different labeling ratios. We asked how reliable our approach was in distinguishing proteins which show different expression levels (increase by 20%) and proteins that are unchanged. We mixed equal amounts of human proteins across six samples with E. coli proteins with different mixing ratios in triplicates (Fig. 5A). To make the problem hard, we reduced the number of ions for quantification by limiting ion injection times for the MS3 scan to 22 msec. Thus ideally all E.coli proteins should be identified as "differentially expressed" on a background of human proteins none of which are. This synthetic experiment simulates an essential application of having the complete likelihood function, when confidence intervals representation is used to prioritize the follow-up targets of a proteomic experiment. Assigning a P-value to a claim that a given protein is differentially expressed allows to rank the proteins from the most to the least likely differentially expressed. Sliding threshold on this scale allows building what is known in Machine Learning as an ROC curve – a way to compare probabilistic classifiers tradeoff between false positive and false negative rates. A T-test using 2 or 3 repeats classifies a protein as differentially expressed if a probability that two sets of measurements came from the same distribution is under (1-q%); a BACIQ (our new method to calculate confidence intervals) test classifies a protein as differentially expressed if more than q% of the probability distribution falls to the right of the 0.5 threshold.
Crucially, while the T-test requires at least one repeat, our method can be applied to a single experiment, as well as two and three repeats by merely combining measurements. As shown in Figure 5B even using no replicates, our method substantially outperforms the T-test based classification for two samples and performs similarly to a T-test based classification for triplicates, even though the type of errors is different in trading off false positives for true positives. Combining

only two channels, we uniformly outperform the three replica T-test, and improve even more using all three replicas.

Discussion

We have shown how a hierarchical Beta-Binomial model can be used to adequately reflect uncertainty in quantitative multiplexed proteomics measurements. We presented the method and the implementation of a modeling pipeline which can assign confidence for both single peptide and multiple peptide proteomic measurements. We demonstrated how to estimate a calibration multiplier for a given instrument and mass resolution and then use that multiplier to convert a continuous MS-signal value into discrete event counts suitable for the Beta-Binomial modeling. For the sake of simplicity, we focused on a two-case scenario. However, the same framework can be mathematically expanded to multiple conditions. Confidence for single peptides (Fig. 1) are then adequately modeled with a Dirichelet instead of the Beta distribution. Multiple peptide measurements mapping to one protein (Fig. 3, Fig S5) generalize from the Beta-Binomial model to the Dirichelet-Multinomial model. Stan implementation of the Dirichelet-Multinomial is included with the software.

So far, we have only tested the BACIQ approach for data acquired with TMT-MS3. However, we believe the approach should be easily transferrable to other accurate multiplexed proteomics methods like ModQuant, or TMTc+. The systematic error associated with MS2-based measurements will lead to inadequate confidence intervals for the underlying true protein ratios. However, we suspect that for the prioritization of systematic changes in an experiment, BACIQ could still be useful even when measurements are systematically distorted towards a 1:1 ratio due to interference.

We also expect that with some adaptations BACIQ might be adequate for MS1 based labeled quantitative proteomics methods like SILAC or reductive methylation (Kovanich et al., 2012; Ong et al., 2002). Lastly, we would like to point out that the principles of combining discordant measurements with underlying discrete counts to reflect the level of uncertainty using the Dirichlet-Multinomial model discussed in this paper should be generally applicable to similar non-proteomic measurements such as RNA-Seq.

Materials and methods

Sample preparation

The single proteome standard and the two-proteome interference model was prepared mostly as previously described (Peshkin et al., 2015; Wühr et al., 2015; Wühr et al., 2012). HeLa S3 cells were grown in suspension to 1×10^6 cells/mL. Cells were harvested by spinning 160 rcf for 5 min at room temperature. After two washes with PBS, the pellet was flash frozen in liquid nitrogen. The pellet containing about 600 µg of total protein was resuspended in 1 ml of lysis buffer containing 25 mM HEPES pH 7.2, 2% SDS and protease inhibitors (complete mini., EDTA-free; Roche). Cells were lysed by sonication: 6 pulses, 10sec each, at 75% amplitude.

E. coli cell culture was harvested at 0.5 OD and spun down at 4,000 rcf for 20 min at 4 C. The pellet containing about of 650 ug of total protein was resuspended in 1 ml of lysis buffer containing 8M Urea, 2 M Thiourea, 50 mM HEPES pH 7.2, 2% SDS, 5 mM DTT. Cells were lysed by sonication: 10 pulses, 30 sec each, at 75% amplitude.

200 µL of HeLa lysate was reduced with 5 mM DTT for 20 min at 60 C. Further, both samples - 200µL of HeLa lysate and 200µL of *E.coli* lysate were alkylated with 15 mM *N*- Ethylmaleimide (NEM) for 30 min at room temperature. The excess of NEM was quenched with 5 mM DTT for 10 min at room temperature in both samples. Next, 200 µL of lysate were Methanol-Chloroform precipitated as

previously described (Wessel and Flugge, 1984). Protein concentration was determined using the bicinchoninic acid (BCA) protein assay (Thermo Fisher). The samples were resuspended in 6 M Guanidine Chloride in 10 mM EPPS pH 8.5 with a subsequent dilution to 2 M Guanidine Chloride in 10 mM EPPS pH 8.5 for digestion with Lys-C (Wako, Japan) at room temperature with Lys-C 20 ng/µL overnight. Further the samples were diluted to 0.5 mM Guanidine Chloride in 10 mM EPPS pH 8.5 and digested with Lys-C 20 ng/µL, and Trypsin 10 ng/µl at 37 C overnight. The digested samples were dried using a vacuum evaporator at room temperature and taken up in 200 mM EPPS pH 8.0 for a pH shift which is necessary for optimal labeling conditions. 10 µL of total E. coli or human peptides were labeled with 3µL of TMT 20µg/µL. TMT reagents were dissolved in anhydrous Acetonitrile. TMT Samples were labeled for 2 hours at room temperature. Further, labeled samples were quenched with 0.5% Hydroxylamine solution (Sigma, St. Louis, MO) and acidified with 5% phosphoric acid (pH<2) with subsequent hard spin at 16,000 RCF for 10 min at 4 C. The samples were dried using a vacuum evaporator at room temperature. Dry samples were taken up in HPLC grade water and stage tipped for desalting (Rappsilber et al., 2007). The samples were resuspended in 1% formic acid to 1µg/µL and 1µg of each sample was analyzed with the MultiNotch MS3 approach (McAlister et al., 2014). The samples were labeled with the desired mixing ratios: 1.0:1.0:1.0:1.2:1.2:1.2 for E.coli, and 1.0:1.0:1.0:1.0:1.0:1.0 for HeLa (Figure 5A). Approximately equal amounts of the samples were mixed. To correct for pipetting errors the summed signal for each species was normalized to the desired mixing ratios.

LC/MS analysis

Approximately 1 µL per sample were analyzed by LC-MS. LC-MS experiments were performed on Orbitrap Fusion Lumos (Thermo Fischer Scientific). The instrument was equipped with Easy-nLC 1200 high pressure liquid chromatography (HPLC) pump (Thermo Fischer Scientific). For each run, peptides were separated on a 100 µm inner diameter microcapillary column, packed first with approximately 0.5 cm of 5-µm BEH C18 packing material (Waters) followed by 30 cm of 1.7-µm BEH C18 (Waters). Separation was achieved by applying 4.8%- 24% acetonitrile gradient in 0.125% formic acid and 2% DMSO over 120 min at 350 nL/min at 60C. Electrospray ionization was enabled by applying a voltage of 2.6 kV through a microtee at the inlet of the microcapillary column. The Orbitrap Fusion Lumos was using a MultiNotch-MS3 method (McAlister et al., 2014).The survey scan was performed at resolution of 120k (200m/z) from 350 Thomson (Th) to 1350 Th, followed by the selection of the 10 most intense ions for CID MS2 fragmentation using the quadrupole and a 0.5 Th isolation window. Indeterminate and singly charged, and ions carrying more than six charges were not subjected to MS2 analysis. Ions for MS2 were excluded from further selection for fragmentation for 90 s. MS3 spectra were acquired in the Orbitrap with 120k resolution (200 m/z) simultaneous precursor selection of the five most abundant fragment ions from the MS2 spectrum. For the MS3 spectrum we used an MS2 isolation window of 2 Th. The maximal ion-injection time for the MS3 spectrum was set to 22ms.

Implementation of BACIQ

The model and the inference was implemented using **Stan**® (Carpenter et al., 2016), specifically the R flavor. Stan's sampling functionality was used to execute Monte-Carlo Markov Chain in order to obtain a sample from the posterior distribution over parameters of the model. Figure S4 illustrates the general principles of MCMC. Since for any values of the model's parameters we can efficiently evaluate the likelihood, parameter space is explored while samples are biased towards the areas of

high likelihood. The obtained statistical sample is essentially used for a histogram representation of the posterior to obtain confidence intervals. Additionally, we implemented the inference pipeline as a web-server front end distributing processes among the nodes of a Linux cluster. Two modes of use are available: compute the confidence interval for the ratio of two channels to their sum, or for all ratios of one channels to the sum across channels.

Acknowledgements

We would like to thank Grame McAlister, Alexander Makarov, and Yarden Katz for helpful suggestions and discussions. Thanks to Meera Gupta and Elizabeth Van Itallie for comments on the manuscript. Bob Carpenter and the rest of Stan team for technical support. This work was supported by NIH grants R01GM103785, R01HD073104, Princeton University Startup fund.

Table

Instrument	Resolution	Multiplier (95% Conf. Interval)
Orbitrap Elite	15K	4.5 (4.2, 4.8)
Orbitrap Elite	30K	3.3 (3.1, 3.6)
Orbitrap Elite	60K	2.5 (2.3, 2.6)
Fusion/Lumos	15K	3.4 (2.1, 4.6)
Fusion/Lumos	30K	2.6 (2.0, 3.2)
Fusion/Lumos	50K	2.0 (1.5, 2.5)
Fusion/Lumos	60K	1.8 (1.4, 2.3)
Fusion/Lumos	120K	1.3 (1.0, 1.6)

Table 1. Multiplier for the MS-signal to be converted into discrete Binomial event counts. Shown are the multiplier values and respective confidence intervals for different instruments and typical mass resolution values.

Figures

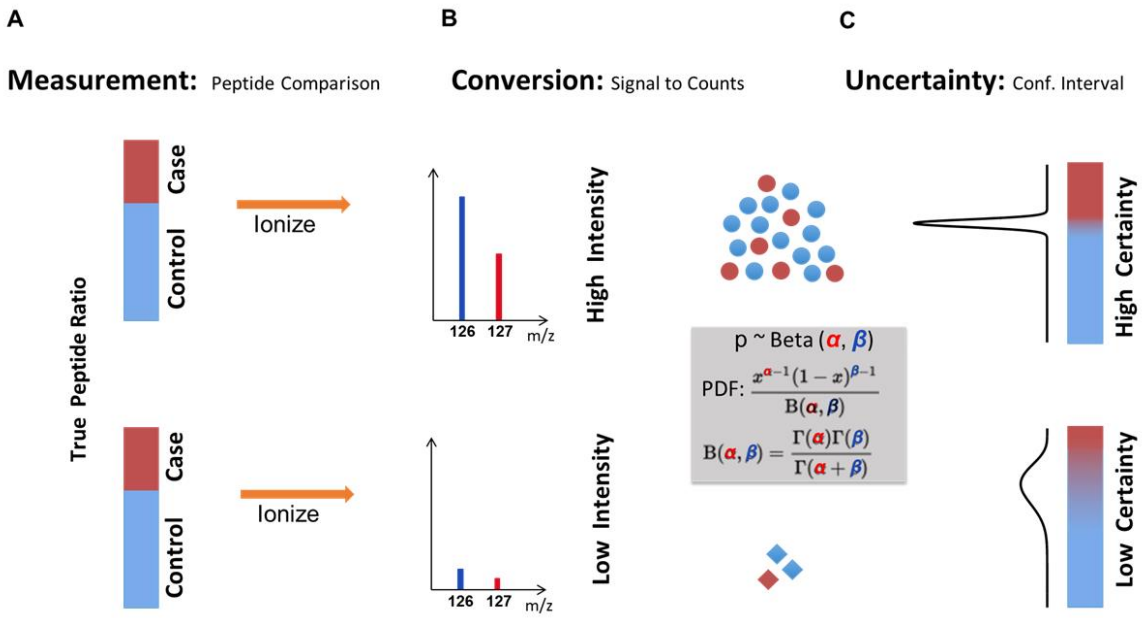

Figure 1. Assigning confidence to the measurement of a single peptide A) Peptides are labeled with isobaric tags encoding the different conditions case and control (red and blue). Shown are examples for two peptides with the identical relative abundance (true ratio) between case and control. B) After ionization and fragmentation the relative intensity of fragments produced by the isobaric tag can be used to quantify relative peptide abundance. These quantification spectra contain the information of the peptide relative abundance but also the total sum of intensities (Signal to FT noise). This signal is proportional to the number of ions measured. Intuitively, the fewer ions we measure the more the measured ion ratio tends to divert from the true ratio between case and control due to ion-statistics C) Once we can convert the signal into number of ions, the problem of estimating the likelihood of the “true ratio” becomes identical to the estimation of a coin's fairness, given a certain number of head and tail measurements. Using a standard Bayesian approach we can express the underlying “true ratio” likelihood as a beta distribution where alpha and beta represent the number of ions measured for the two samples. A higher ion-count (top row) results in a tighter likelihood function and respectively tighter confidence intervals than a low ion-count (bottom row). For more than two cases this approach can be generalized with a Dirichlet-distribution.

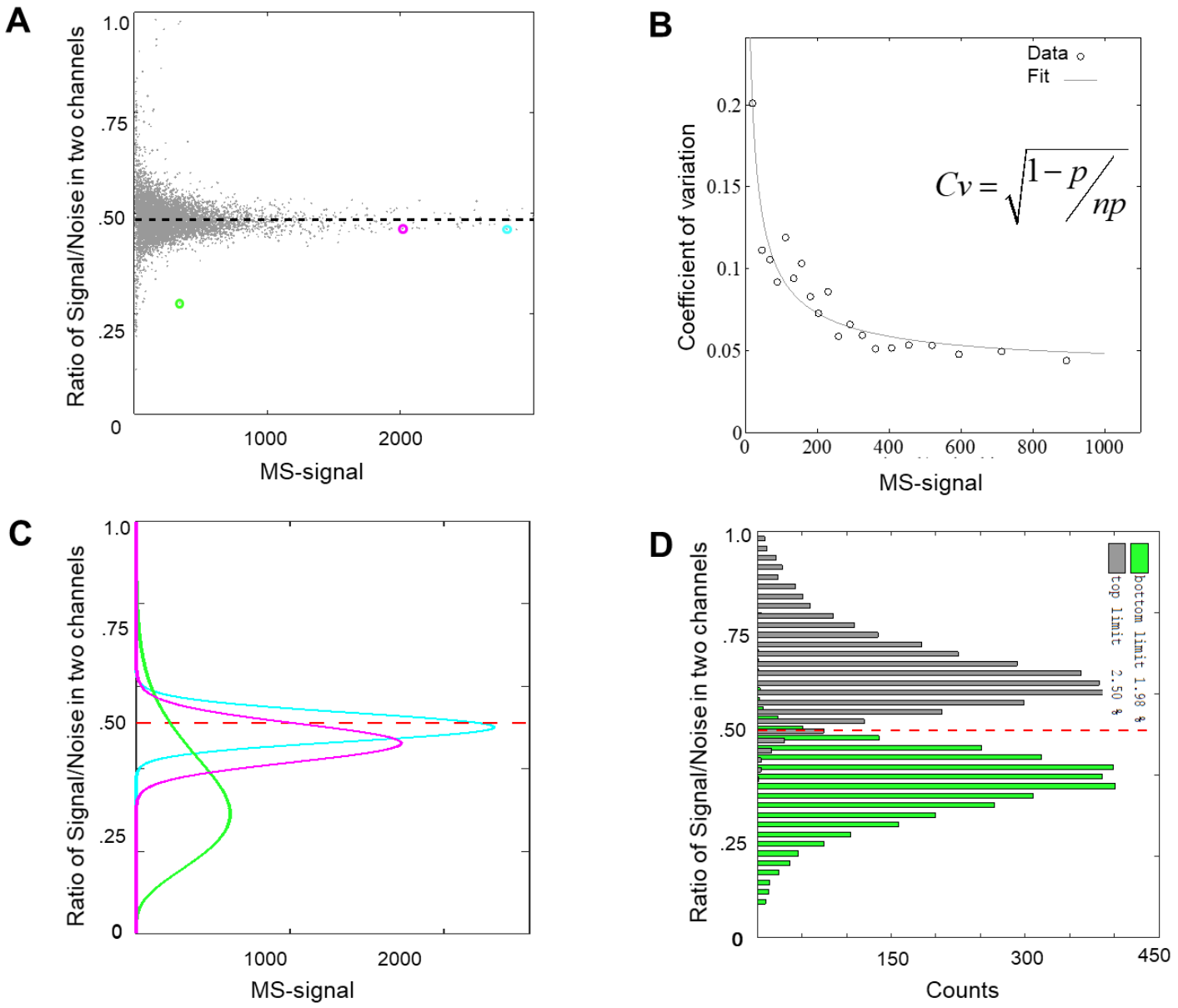

Figure 2: Conversion of MS-signal into counts. A) We generated a sample in which all peptides are labeled with two different TMT-tags and mixed with identical ~1:1 ratio. When we plot the measured ratio versus the summed MS-signal in the two channels, we observe that measurement with higher MS-signal asymptotes to the median answer, which approximates the true underlying ratio (a dashed line). B) Assuming ion-statistics is the dominant noise source, we can fit the binned coefficients of variation (CVs) and obtain the conversion factor of mass spectrometry signal to the number of ions or pseudo-counts. The data shown was obtained on an Orbitrap Lumos with 50K mass resolution. Our best guess for the conversion factor is 2.0 . C) Three examples color-coded to correspond to three peptide data points in sub-figure (A), of the likelihood functions reflecting confidence, expressed as the beta-distribution using the calculated ion-counts. D) Histogram of the upper and lower bound values for the 95% confidence intervals. The observed percentage of peptides for which the true answer is outside of the 95% confidence interval is 1.98% and 2.5% respectively for over- and under-estimation, which are symmetric and in good agreement with the expected total 5% value.

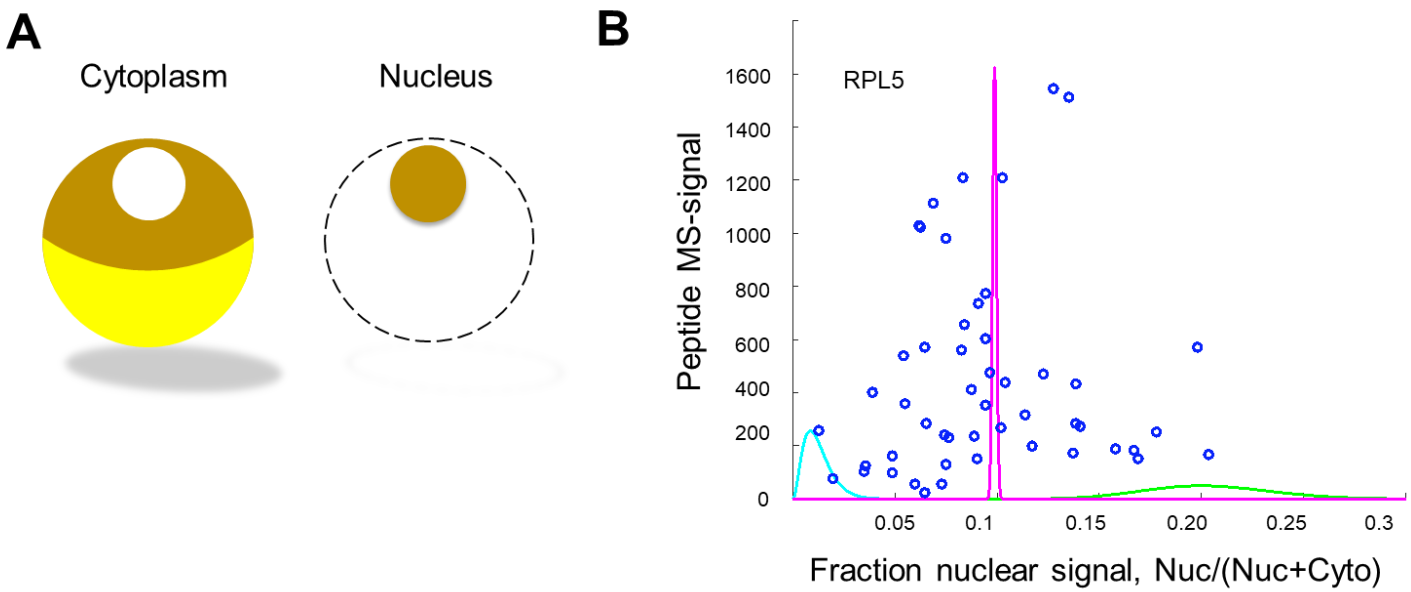

Figure 3. Only considering ion-statistics does not produce accurate confidence intervals at the protein level: A) To evaluate the confidence intervals of peptides from the same protein, we revisited our previously published experiment, where we measured the localization of proteins between nucleus and cytoplasm of the frog oocyte. B) Blue discs show 50 measured peptides (ratio and intensity) assigned to the Ribosomal Protein L5 (RPL5). We show the beta-function likelihoods for two extreme peptides (leftmost in blue and rightmost in green). Note that these peptides likelihood functions are basically mutually exclusive, i.e. most generous confidence intervals would excluding each other. Additionally we show the likelihood based on summing up all the peptides together (magenta) which corresponds to unjustifiably tight confidence. This example illustrates that for the expression of confidence intervals on the protein level, we cannot assume that ion-statistics is the only source of measurement error. Rather, we have to integrate other sources of errors e.g. due to differences in sample handling.

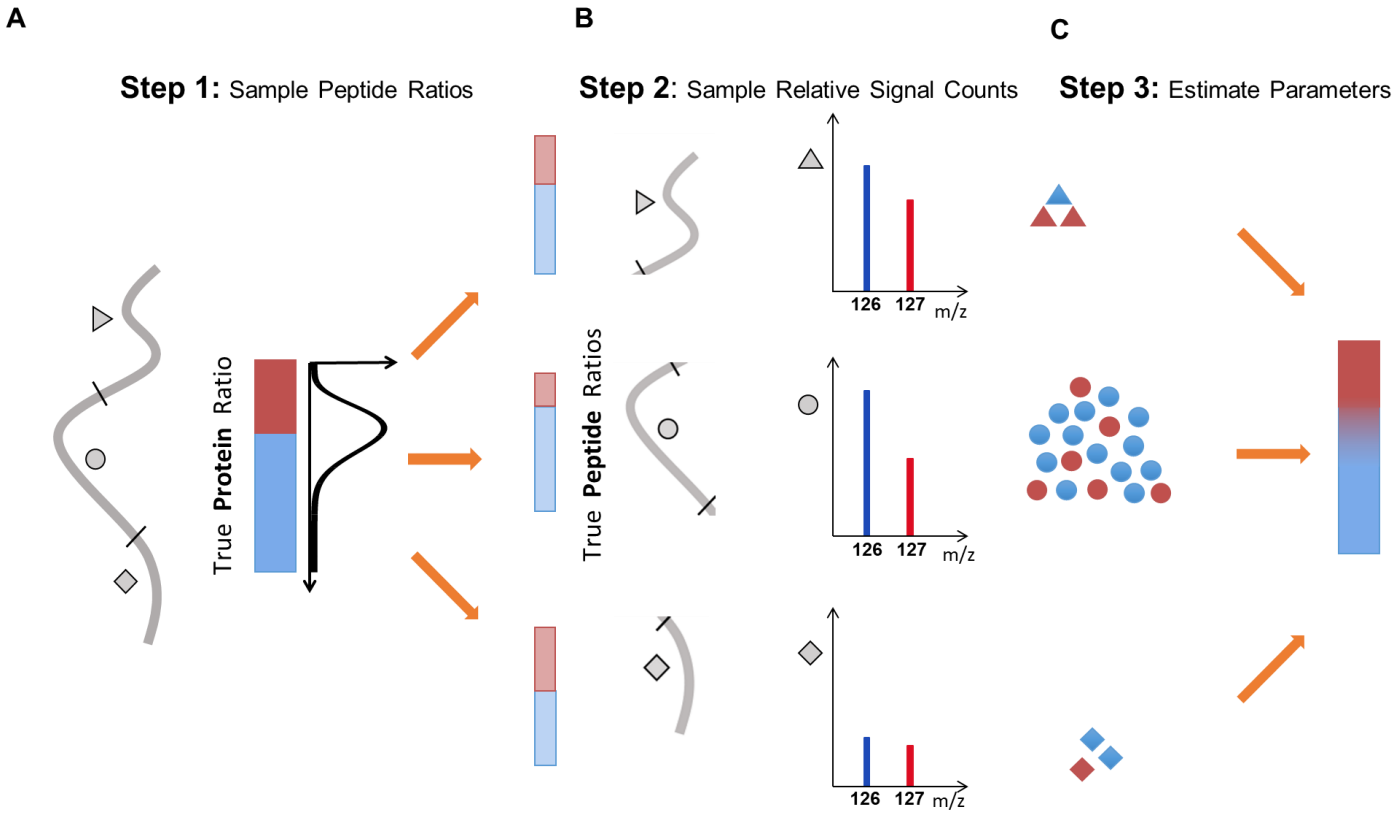

Figure 4. Modeling confidence for proteins with multiple peptide measurements. A) The "true protein ratio" can be distorted due to differences in sample handling (e.g. digestion, isobaric-labeling, post-translational modifications) and give rise to multiple peptides with differing "true peptide ratios". We assume an underlying beta-distribution representing the likelihood function over protein ratio, from which peptide ratios q_i for the constituent peptides of a given protein is sampled as illustrated for i=3; B) Each peptide is measured via the mass spectrometer. This step is equivalent to the model in figure 1. This step is equivalent to sampling from a respective Binomial distribution with probability of success $\boldsymbol{q_i}$. C) The entire data generation process can be adequately described with the beta-binomial process (or dirichlet -multinomial process for more than two cases). We can use this mapping and express confidence as an estimation for the the mean of the Beta distribution representing the protein, specifically the likelihood function of the posterior distribution of that value as the representation of the uncertainty.

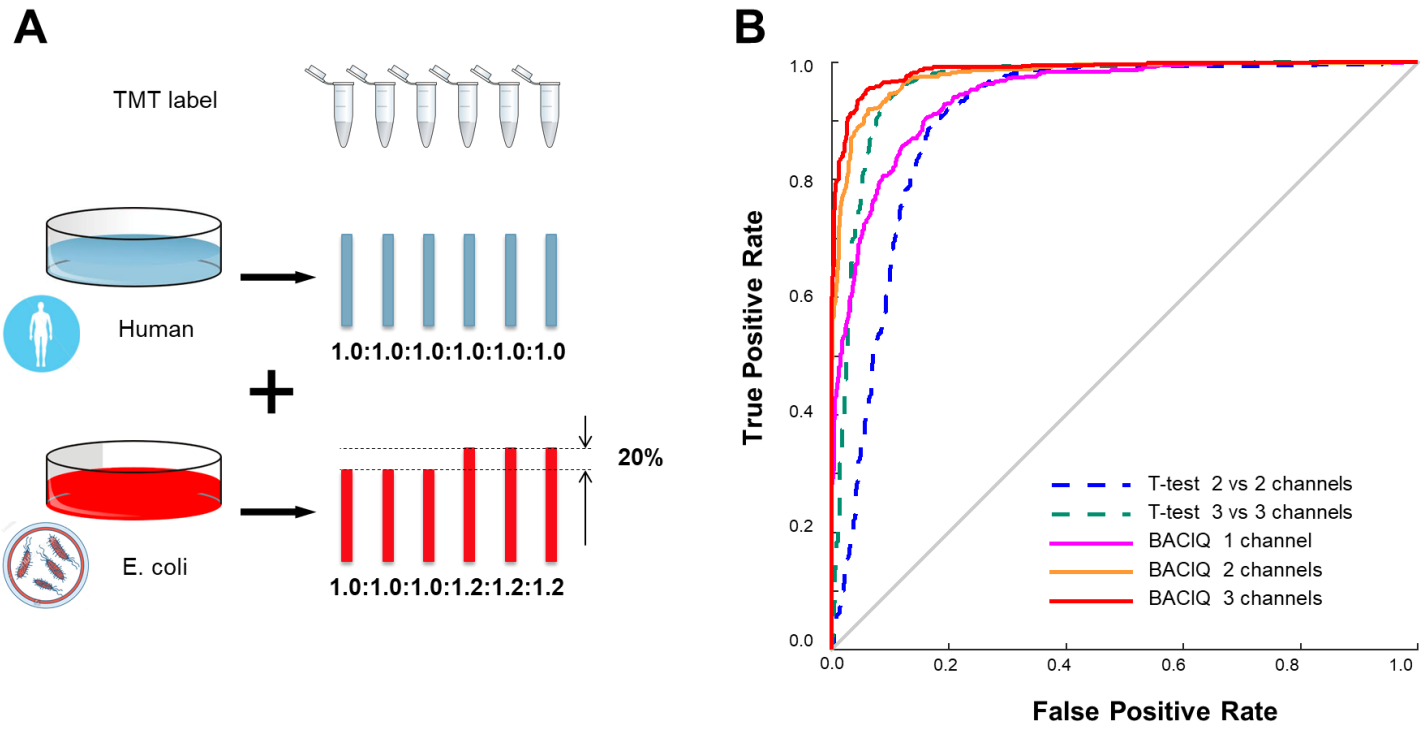

Figure 5. Validating our method with a differential expression experiment. A) Six samples were prepared by mixing material from two species as follows. Six identical *E. coli* samples (i.e. proportions across 6 channels were 1.0 : 1.0 : 1.0 : 1.2 : 1.2 : 1.2) were mixed with a Human sample in two sets of three as shown (i.e. proportions across 6 channels were 1.0 : 1.0 : 1.0 : 1.0 : 1.0 : 1.0). TMT-labeled mixture of two proteomes (*E.coli* and human) was analyzed by LC-MS2. B) Comparison of our method with T-test to detect differentially expressed proteins. BACIQ can detect statistically significant changes even without replicates, while the T-test requires at least one repeat. An ROC plot illustrates a tradeoff between precision and recall, i.e. how many proteins are labeled as differentially expressed. Our method is superior to the T-test when the same number of replicates are used. Even without replicates or method (magenta) outperforms the T-test with two replicates (dashed blue) nearly completely. We achieve close to perfect distinction between differentially and unchanged proteins if we use our method and three replicates (orange and red).

References

Beck, S., Michalski, A., Raether, O., Lubeck, M., Kaspar, S., Goedecke, N., Baessmann, C., Hornburg, D., Meier, F., Paron, I.*, et al.* (2015). The Impact II, a Very High-Resolution Quadrupole Time-of-Flight Instrument (QTOF) for Deep Shotgun Proteomics. Mol Cell Proteomics *14*, 2014-2029.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M.A., Guo, J., Li, P., and Riddell, A. (2016). Stan: A probabilistic programming language. Journal of Statistical Software *20*, 1-37.

Choi, M., Chang, C.Y., Clough, T., Broudy, D., Killeen, T., MacLean, B., and Vitek, O. (2014). MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. Bioinformatics *30*, 2524-2526.

Clough, T., Thaminy, S., Ragg, S., Aebersold, R., and Vitek, O. (2012). Statistical protein quantification and significance analysis in label-free LC-MS experiments with complex designs. BMC Bioinformatics *13 Suppl 16*, S6.

Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Molecular & cellular proteomics : MCP *13*, 2513-2526.

Deshmukh, A.S., Murgia, M., Nagaraj, N., Treebak, J.T., Cox, J., and Mann, M. (2015). Deep proteomics of mouse skeletal muscle enables quantitation of protein isoforms, metabolic pathways, and transcription factors. Mol Cell Proteomics *14*, 841-853.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B. (2014). Bayesian data analysis, Vol 2 (CRC press Boca Raton, FL).

Hilger, M., and Mann, M. (2012). Triple SILAC to determine stimulus specific interactions in the Wnt pathway. J Proteome Res *11*, 982-994.

Jiang, X., Arrey, T., Damoc, E., Scigelova, M., Horn, D., Viner, R., and Huhmer, A.F.R. TMT Workflow on the Q Exactive Series - Instrument Parameter Optimization and Data Analysis in Proteome Discoverer 2.1 Software.

Koh, H.W., Swa, H.L., Fermin, D., Ler, S.G., Gunaratne, J., and Choi, H. (2015). EBprot: Statistical analysis of labeling-based quantitative proteomics data. Proteomics *15*, 2580-2591.

Kovanich, D., Cappadona, S., Raijmakers, R., Mohammed, S., Scholten, A., and Heck, A.J. (2012). Applications of stable isotope dimethyl labeling in quantitative proteomics. Anal Bioanal Chem *404*, 991-1009.

Lapek, J.D., Jr., Greninger, P., Morris, R., Amzallag, A., Pruteanu-Malinici, I., Benes, C.H., and Haas, W. (2017). Detection of dysregulated protein-association networks by high-throughput proteomics predicts cancer vulnerabilities. Nat Biotechnol *35*, 983-989.

Lehmann, E.L., and Casella, G. (2006). Theory of point estimation (Springer Science & Business Media).

Makarov, A., and Denisov, E. (2009). Dynamics of ions of intact proteins in the Orbitrap mass analyzer. J Am Soc Mass Spectrom *20*, 1486-1495.

McAlister, G.C., Nusinow, D.P., Jedrychowski, M.P., Wühr, M., Huttlin, E.L., Erickson, B.K., Rad, R., Haas, W., and Gygi, S.P. (2014). MultiNotch MS3 Enables Accurate, Sensitive, and Multiplexed Detection of Differential Expression across Cancer Cell Line Proteomes. Analytical Chemistry *86*, 7150-7158.

Minka, T. (2000). Estimating a Dirichlet distribution (Technical report, MIT).

Ning, Z., Zhang, X., Mayne, J., and Figeys, D. (2016). Peptide-Centric Approaches Provide an Alternative Perspective To Re-Examine Quantitative Proteomic Data. Anal Chem *88*, 1973-1978.

Oberg, A.L., and Mahoney, D.W. (2012). Statistical methods for quantitative mass spectrometry proteomic experiments with labeling. BMC Bioinformatics *13 Suppl 16*, S7.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (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.

Peshkin, L., Wühr, M., Pearl, E., Haas, W., Freeman, R.M., Jr., Gerhart, J.C., Klein, A.M., Horb, M., Gygi, S.P., and Kirschner, M.W. (2015). On the Relationship of Protein and mRNA Dynamics in Vertebrate Embryonic Development. Developmental Cell *35*, 383-394.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc *2*, 1896-1906.

Sonnett, M., Yeung, E., and Wuhr, M. (2017). Accurate, Sensitive, and Precise Multiplexed Proteomics using the Complement Reporter Ion Cluster. bioRxiv.

Student (1908). The probable error of a mean. Biometrika, 1-25.

Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A.K., and Hamon, C. (2003). Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem *75*, 1895-1904.

Ting, L., Rad, R., Gygi, S.P., and Haas, W. (2011). MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. Nat Methods *8*, 937-940.

Wenger, C.D., Lee, M.V., Hebert, A.S., McAlister, G.C., Phanstiel, D.H., Westphall, M.S., and Coon, J.J. (2011). Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. Nat Methods *8*, 933-935.

Wessel, D., and Flugge, U.I. (1984). A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. Anal Biochem *138*, 141-143.

Wühr, M., Freeman, R.M., Jr., Presler, M., Horb, M.E., Peshkin, L., Gygi, S.P., and Kirschner, M.W. (2014). Deep Proteomics of the Xenopus laevis Egg using an mRNA-Derived Reference Database. Current Biology.

Wühr, M., Güttler, T., Peshkin, L., McAlister, G.C., Sonnett, M., Ishihara, K., Groen, A.C., Presler, M., Erickson, B.K., Mitchison, T.J.*, et al.* (2015). The Nuclear Proteome of a Vertebrate. Current biology : CB *25*, 2663-2671.

Wühr, M., Haas, W., McAlister, G.C., Peshkin, L., Rad, R., Kirschner, M.W., and Gygi, S.P. (2012). Accurate multiplexed proteomics at the MS2 level using the complement reporter ion cluster. Analytical Chemistry *84*, 9214-9221.

Zhang, B., Pirmoradian, M., Zubarev, R., and Kall, L. (2017). Covariation of Peptide Abundances Accurately Reflects Protein Concentration Differences. Mol Cell Proteomics *16*, 936-948.