



RESEARCH ARTICLE

Evaluation of Normalization Methods on GeLC-MS/MS Label-Free Spectral Counting Data to Correct for Variation during Proteomic Workflows

Emine Gokce,¹ Christopher M. Shuford,¹ William L. Franck,² Ralph A. Dean,² David C. Muddiman¹

¹W. M. Keck FT-ICR Mass Spectrometry Laboratory, Department of Chemistry, North Carolina State University, Raleigh, North Carolina 27695, USA

²Center for Integrated Fungal Research, North Carolina State University, Raleigh, North Carolina, USA

Abstract

Normalization of spectral counts (SpCs) in label-free shotgun proteomic approaches is important to achieve reliable relative quantification. Three different SpC normalization methods, total spectral count (TSpC) normalization, normalized spectral abundance factor (NSAF) normalization, and normalization to selected proteins (NSP) were evaluated based on their ability to correct for day-to-day variation between gel-based sample preparation and chromatographic performance. Three spectral counting data sets obtained from the same biological conidia sample of the rice blast fungus *Magnaporthe oryzae* were analyzed by 1D gel and liquid chromatography-tandem mass spectrometry (GeLC-MS/MS). Equine myoglobin and chicken ovalbumin were spiked into the protein extracts prior to 1D-SDS-PAGE as internal protein standards for NSP. The correlation between SpCs of the same proteins across the different data sets was investigated. We report that TSpC normalization and NSAF normalization yielded almost ideal slopes of unity for normalized SpC versus average normalized SpC plots, while NSP did not afford effective corrections of the unnormalized data. Furthermore, when utilizing TSpC normalization prior to relative protein quantification, *t*-testing and fold-change revealed the cutoff limits for determining real biological change to be a function of the absolute number of SpCs. For instance, we observed the variance decreased as the number of SpCs increased, which resulted in a higher propensity for detecting statistically significant, yet artificial, change for highly abundant proteins. Thus, we suggest applying higher confidence level and lower fold-change cutoffs for proteins with higher SpCs, rather than using a single criterion for the entire data set. By choosing appropriate cutoff values to maintain a constant false positive rate across different protein levels (i.e., SpC levels), it is expected this will reduce the overall false negative rate, particularly for proteins with higher SpCs.

Key words: Label-free, Proteomics, Mass spectrometry, Total spectral count normalization, Normalized spectral abundance factor, Quantification, Spectral counting, Significance test, Fold-change, GeLC-MS/MS, nanoLC-MS, Orbitrap XL

Emine Gokce and Christopher M. Shuford contributed equally to this research.

Electronic supplementary material The online version of this article (doi:10.1007/s13361-011-0237-2) contains supplementary material, which is available to authorized users.

Correspondence to: David C. Muddiman; e-mail: david_muddiman@ncsu.edu

Received: 13 June 2011
Revised: 17 August 2011
Accepted: 17 August 2011
Published online: 24 September 2011

Introduction

Label-free quantification for proteomic analyses has gained popularity throughout the last decade. Advantages of label-free approaches compared to label incorporated methods (e.g., SILAC [1, 2]) include simplicity of sample preparation and applicability to any organism. Additionally, reduced sample complexity allows for an increase in the number of peptides sequenced, which results in a greater dynamic range and more comprehensive proteome coverage [3, 4]. Spectral counting [5] and ion abundance [6–10] have been used for label-free quantification and are known to correlate with protein abundance [11]. Furthermore, an algorithm combining spectral counting and ion abundance measurements was developed by Feener and co-workers [12], enabling quantification for an increased number of proteins.

The number of spectral counts (SpCs) for a protein is simply the number of MS/MS spectra that result in identification of its proteolytic peptides. In bottom-up proteomic strategies, data-dependent MS/MS acquisition software selects peptide ions based upon their abundance and charge state, which typically favors identification of more abundant peptides/proteins. Applying dynamic exclusion for previously selected peptides limits the number of SpCs for those abundant peptides and enables the selection of peptides of lower abundance, resulting in higher protein sequence coverage, more confident protein identifications, and increased depth of proteome coverage [13].

Normalization of SpCs is performed to reduce the variance observed between samples and replicates. The variability in SpCs can be caused by numerous factors including sample preparation, gel-to-gel variance (if a gel-based approach is part of the proteomic workflow), and changes in chromatography. Carvalho et al. [14] pointed out the importance of normalizing spectral counting data in order to quantify proteins; however, the most effective normalization method for label-free spectral counting has yet to be elucidated.

For total spectral count (TSpC) normalization [15], the technical replicate with the highest number of TSpC is chosen and the remaining technical replicates for that sample are normalized to it. Subsequently, the values across different samples are normalized to the sample with the highest technical replicate TSpC. The normalization is done for each protein individually and comparisons of average normalized SpCs of the same proteins are made between samples. Comparison of absolute or normalized SpCs between proteins to determine their relative abundance is generally precluded by that fact that longer proteins yield a higher number of spectral counts on average than shorter proteins. To account for this, Washburn and co-workers [16–18] developed a method, termed normalized spectral abundance factor (NSAF), in which the SpC for a given protein are divided by its length (L) to give a spectral abundance factor (SAF). To account for variations between runs, the SAF for a given protein (e.g., SpC/L) is subsequently normalized to the sum of all SAFs for proteins identified within that run to

create a normalized SAF (i.e., NSAF) that can be used to compare the relative abundance of proteins both between and within samples. In the latter method, the authors are using the assumption that the sum of all SAF should be conserved between replicates to correct for differences in sampling rates. In TSpC normalization, the assumption is the sum of all SpC (i.e., the TSpC) should be conserved between runs/samples. In comparison, normalization to selected proteins (NSP) does not assume the sampling rate should be conserved between replicates/samples; rather, it relies on the premise that the total SpC for a standard protein should be conserved between replicates/samples if it is present at the same concentration. In practice, the standard protein(s) can be an endogenous house-keeping protein or an exogenous protein; the latter has the advantage of knowing precisely the amount of protein added to each sample. Any change observed in the SpC for the standard protein is assumed to reflect the variation between replicates and samples for the entire identified proteome. Consequently, the relative changes in SpC for the standard protein between replicates/samples are used as correction factors to normalize the SpC of all proteins.

Significant biological change in protein abundance is determined either by fold-change [19, 20], by significance testing [21–23], or by using a combination of both [14]. Even though SpC datasets do not necessarily meet the criteria for a normal distribution, Student's t -tests are often applied in such experiments to determine statistical significance. Zhang et al. [21] performed a control experiment with yeast samples and six spike-in proteins at three different concentrations to calculate the false positive rates (FPR) for different significance tests. When only one replicate was utilized and thus the assumption of a normal distribution was obviously false, the G -test provided the lowest FPRs. However, when three replicates were utilized, t -testing performed similarly.

We are utilizing label-free quantification to gain proteomic insights related to pathogenicity of the fungus *Magnaporthe oryzae* (*M. oryzae*). *M. oryzae* causes rice blast disease destroying millions of hectares of rice each year, resulting in losses valued at billions of dollars [24]. Since half of the world's human population relies on rice as a nutrition source [25], understanding fungal development as it relates to disease progression is important for development of control strategies. Dean and coworkers sequenced the whole genome of *M. oryzae* in 2005 [26] providing a reliable database for proteomic approaches. Using 1D gel and liquid chromatography-mass spectrometry (GeLC-MS), we would like to study the development of the fungus *M. oryzae*, wild type and mutants, at different time points in its life cycle. Since the production of biological samples over a wide range of conditions (e.g., time, treatments, mutants) is difficult and downstream proteomic workflow is very time consuming, establishing a confident normalization method is critical. Thus, we have used *M. oryzae* conidia spectral counting data to compare the normalization methods TSpC, NSAF and NSP in their ability to account for variance

between samples due to differences in sample preparation and chromatographic performance.

Experimental

Sample Preparation

M. oryzae conidia were harvested from 8 d old minimal medium plates. Three biological replicates, each containing 2 million conidia, were pooled to account for biological variance. Conidia were lysed by bead beating in a 1X PBS (Fisher Scientific, Pittsburgh, PA, USA), 2 M urea (Sigma Aldrich, St. Louis, MO, USA), and 0.1 % SDS buffer (Bio-Rad, Hercules, CA, USA) to create the biological sample from which all experiments were derived. Protein concentration was determined via a BCA assay (Thermo Fisher Scientific, Rockford, IL, USA). Samples 1 and 2, derived from the same biological sample, were prepared and processed on different days. Chicken myoglobin (Sigma Aldrich) and equine ovalbumin (Sigma Aldrich) were chosen as spike-in proteins and 25 ng of each was added to 50 µg of total protein for each sample. The samples were loaded onto 10%–20% gradient 1D-SDS PAGE gels (Bio-Rad). It is necessary to start out with more material in regards to the number of fractions, the recovery, and the number of injections of a complex protein/peptide mixture. Fifty µg is the amount of material loaded for the size gels utilized and it allows for adequate amounts of peptide material to be recovered (post-digestion). After Coomassie staining (Bio-Rad), 10 fractions of gel bands were excised and in-gel digestion [27] was performed on each fraction. Briefly, each gel-fraction was destained with 100 µL of 50:50 ammonium bicarbonate (Sigma Aldrich)/acetonitrile (ACN) (Burdick and Jackson, Muskegon, USA). Reduction was performed with 100 µL of 10 mM dithiothreitol (Sigma Aldrich) at 56 °C for 30 min, alkylation with 100 µL of 90 mM iodoacetamide (Sigma Aldrich) in the dark for 30 min and digestion overnight at 37 °C with trypsin (protein: protease ratio of 5:1). Acetonitrile was added and discarded between each step to dehydrate the gel pieces. To extract the peptides, 200 µL of 5% formic acid (Sigma Aldrich) in ACN was added to each fraction and incubated for 15 min at 37 °C. The supernatants were transferred into new tubes. ACN (100 µL) was added to the gel pieces and supernatants were combined; this was repeated once more for each fraction. Sample 2' was produced by pooling one-third of the volume of adjacent in-gel digested fractions of sample 2 together to give a sample with only 5 gel fractions, yet having undergone the same sample processing as the 10 gel fraction sample. All samples were dried down and stored at –20 °C until nanoLC-MS/MS analysis.

NanoLC-MS/MS

A 75 µm i.d. IntegraFrit capillary (New Objective, Woburn, MA, USA) trap was packed to 5 cm with Magic C18AQ packing material (Michrom BioResources, Auburn, CA, USA). A 75 µm i.d. PicoFrit capillary column (New

Objective, Woburn, MA, USA) was packed 15 cm with the same packing material. Separation was carried out using a nanoLC-1D+ system from Eksigent (Dublin, CA, USA) with a continuous, vented column configuration as previously reported by our group [28]. A 2 µL (200 ng) sample was aspirated into a 10 µL loop and loaded onto the trap. Only 200 ng were analyzed per injection so as not to overload the nanoLC column. The flow rate was set to 350 nL/min for separation on the analytical column. Mobile phase A was composed of 98% H₂O (Burdick and Jackson), 2% ACN and 0.2% formic acid (Sigma) and mobile phase B was composed of 98% ACN, 2% H₂O, and 0.2% formic acid. A 1 h linear gradient from 5% to 50% B was performed. All measurements were performed at room temperature and three technical replicates of each sample were run to allow for statistical comparisons between samples, which are necessary for label-free quantification.

A hybrid LTQ-Orbitrap XL MS (Thermo Fisher Scientific, Bremen, Germany) was used to perform MS analysis. For data-dependent acquisition, the parameters recently published by our group to be optimal for achieving maximum proteome coverage were used verbatim [29]. External calibration was performed following manufacturer's instructions and using the manufacturer's calibration mix and lock mass internal calibration using polydimethylcyclosiloxane (*m/z* 445.120025) was enabled [30].

Data Analysis

Data analysis was performed by searching each .RAW file, independently, against a concatenated target-reverse *M. oryzae* database (MG8_GeneCall10.fasta) from the Broad Institute using MASCOT Distiller version 2.3.01 (Matrix Science Inc., Boston, MA, USA). MASCOT parameters were ±5 ppm peptide ion tolerance, ±0.6 Da MS/MS fragment ion tolerance, and two allowed missed cleavages. Carbamidomethylation of cysteine was set as a fixed modification and oxidation of methionine and deamidation of glutamine and asparagine were variable modifications. Peptide lists (.dat files) were created for each .RAW file by MASCOT. ProteoIQ version 2.1.01_SILAC_beta08 (Bio-Inquire, Athens, GA, USA) was used to create five different label-free spectral counting projects: (a) Sample 1, (b) Sample 2, (c) Sample 2', (d) combination of samples 1 and 2, and (e) combination of samples 2 and 2'. A 1% protein FDR was applied to each project, independently (i.e., the FDR was calculated based on the cumulative results of the sample files included in that particular project) [31]. Log₂ SpC ratios were calculated and a pairwise *t*-test was performed on the proteins identified in sample 1 and sample 2.

Results and Discussion

The experimental workflow is shown in Figure 1. Sample 1 and sample 2 are derived from the same biological sample

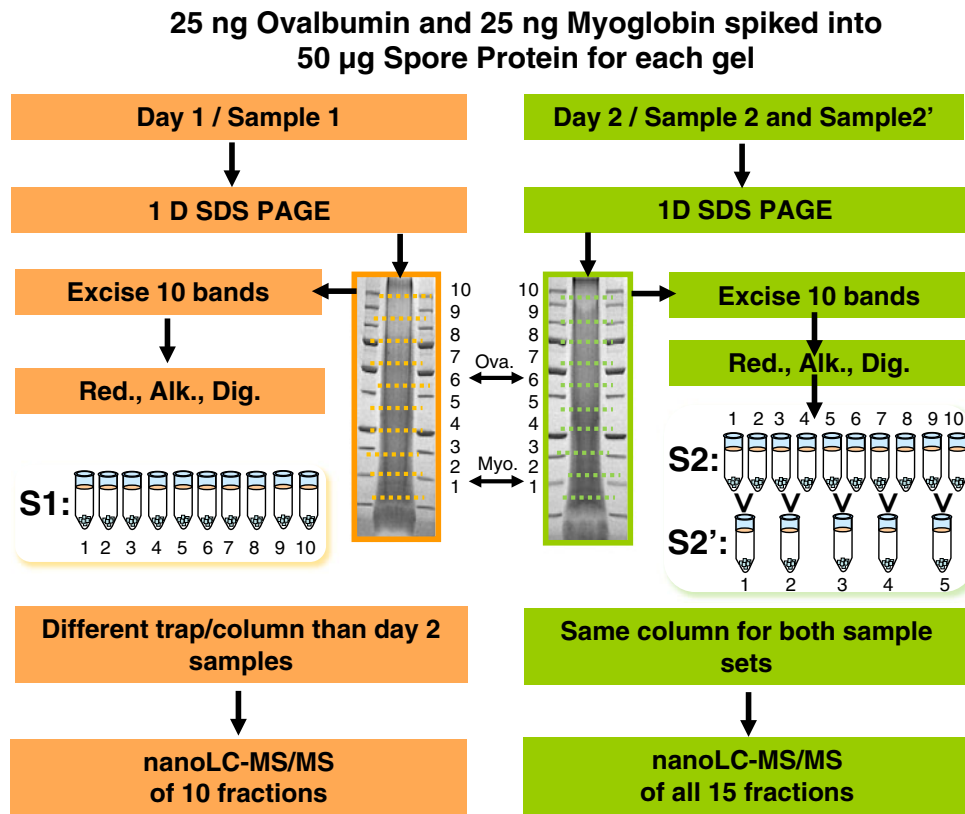


Figure 1. Experimental workflow. The samples were prepared and processed on two different days. Myoglobin (25 ng) and ovalbumin (25 ng) were added to 50 µg of *M. oryzae* conidial protein. One-dimensional SDS-PAGE separation and in-gel digestion were performed. Ten fractions from d 1 and 10 fractions from d 2 were analyzed in triplicate by nanoLC-MS using different traps and columns for each sample set. Additionally, adjacent fractions from d 2 sample were pooled and also analyzed in triplicate by nanoLC-MS

but were prepared and processed after protein extraction on different days. *M. oryzae* conidial protein spiked with equine myoglobin and chicken ovalbumin was loaded onto 1D-SDS-PAGE gels and in-gel digestion was performed. The 10 fractions of each sample were analyzed by nanoLC-MS/MS using different traps and analytical columns for each sample. Additionally, one-third of the final volume of adjacent in-gel digested fractions of sample 2 was combined to produce sample 2' (five fractions) and were also analyzed by nanoLC-MS.

Sample 1 yielded 76,638 TSpC (the sum of three technical replicates) and 1185 proteins identified (see Figure 2). The TSpC number for sample 2 was 95,025 and 1477 proteins were identified. The number of shared proteins between samples 1 and 2 was 1121. Sample 1 and sample 2 contained 64 and 356 unique proteins, respectively. The differences in protein identifications (24%) and number of TSpCs (19%) between samples 1 and 2 were caused by sample processing on different days, reagent quality, gel-to-gel variance, and use of a different trap and column. A slightly higher percentage of difference in the number of proteins identified was observed in a study by Cooper et al. [22]. They reported differences up to 30% in the number of proteins identified between nine replicate

soybean peptide samples spiked with different amounts (0.005 to 2.5 pmol) of tryptic peptides from bovine apotransferrin tryptic digest, separated by MudPIT and analyzed on a LTQ-Orbitrap XL mass spectrometer.

In order to determine if normalization methods can recover from variables such as large differences in sample complexity, we mimicked such a situation by doubling the sample complexity of sample 2 by pooling adjacent gel fractions to create sample 2'. The number of TSpC for sample 2' was 49,067, about half of the TSpC of sample 2, which was expected due to the decrease in the number of fractions (10 fractions for sample 2 and only five fractions for sample 2'). Moreover only 1087 proteins were identified from sample 2' and were a subset of the total population of protein identifications from sample 2.

SpC scatter plots for proteins from combined analysis of sample 1 and 2 are shown in Figure 3. The first plot shows the unnormalized SpCs for each protein (sum of the three technical replicates) versus the average SpCs of the two samples. A regression line slope of 1 is anticipated in the absence of biological variation as is the case here. The unnormalized scatter plots show slopes of 0.901 and 1.099 for samples 1 and 2 respectively, indicating that sample handling on different days, gel to gel variance, difference in

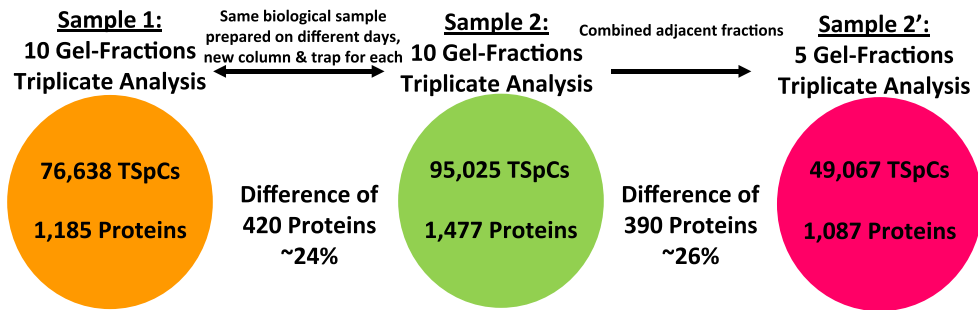


Figure 2. Venn diagrams showing the TSpCs and the total protein numbers identified at 1% protein FDR in each sample. Sample 2 yielded the highest number of identified proteins with 1477. While a difference of 19% in the number of TSpCs between sample 1 and 2 was observed, a reduction in TSpCs by 50% was noticed when adjacent fractions were combined and only half the number of fractions were analyzed

reagent quality, and slightly different chromatography have some effect on SpC reproducibility. Normalizing the data with TSpC normalization corrected the slopes to 1.0093 for sample 1 and 0.9907 for sample 2, while normalization with NSAF corrected the slopes to 1.001 and 0.9878, respectively. NSP yielded some improvements by normalization to the spike-in proteins myoglobin and/or ovalbumin. NSP to myoglobin and ovalbumin yielded slopes of 1.0262 for sample 1 and 0.9738 for sample 2. Slopes of 0.968 for sample 1 and 1.032 for sample 2 were observed by normalizing to myoglobin and

normalizing to ovalbumin resulted in slopes of 1.0696 for sample 1 and 0.9304 for sample 2. Normalization factors can be found in the Online Resource 1–Supplementary Tables S1 and S2.

The sample complexity was doubled in the case of sample 2' to simulate a drastic change in sample complexity and to evaluate the ability of normalization methods to compensate for that. In the unnormalized scatter plots, shown in Figure 4, the slope of sample 2' (0.6968) was almost half of the slope of sample 2 (1.3032). The

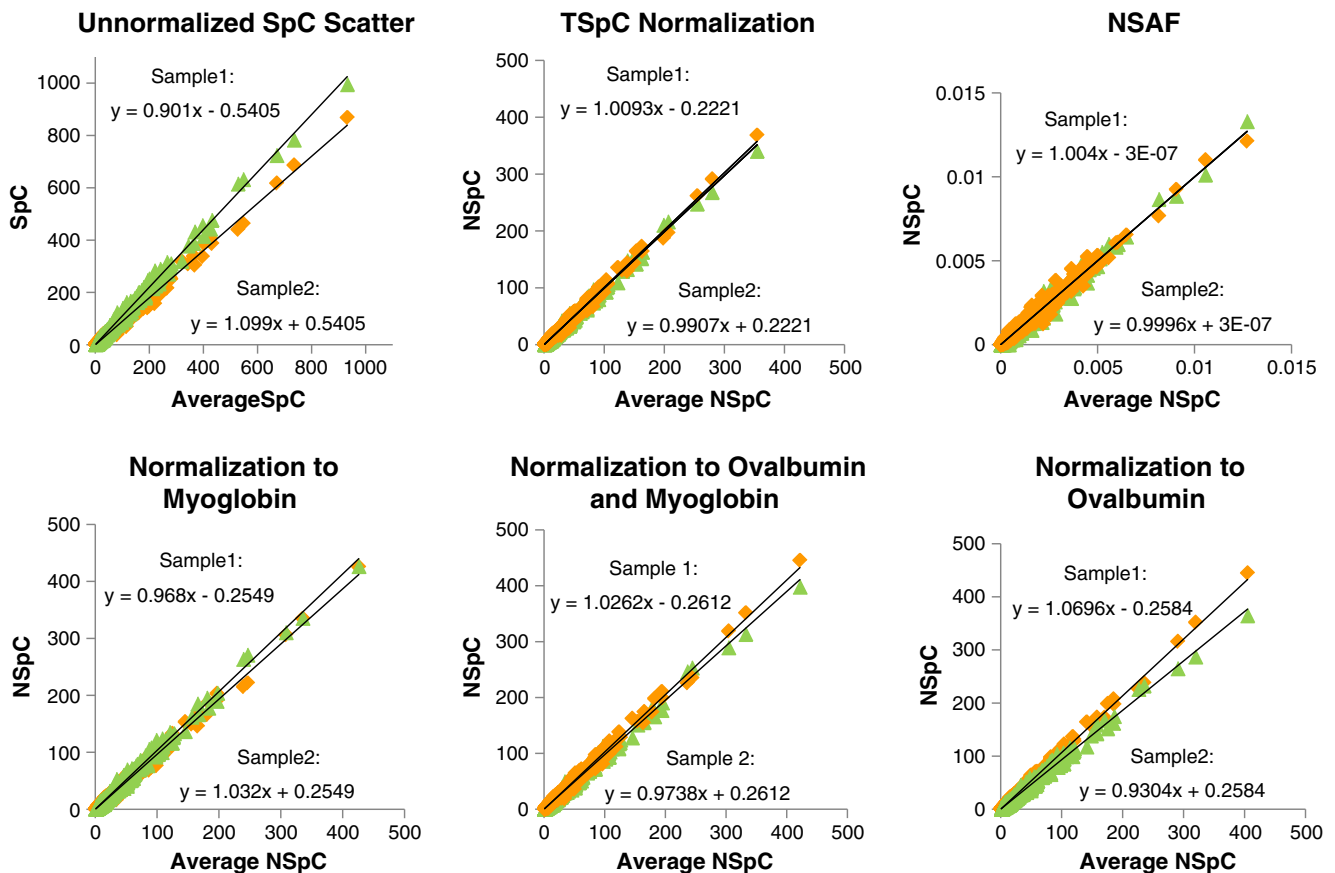


Figure 3. SpC scatter plots from combined analysis of sample 1 and 2. Unnormalized SpCs and normalized SpC (NSpC) data for each protein are plotted versus the average SpCs for the protein derived from both samples. NSAF normalization with slopes of 1.001 and 0.999 for samples 1 and 2 corrects best in comparison to NSP and TSpC normalization

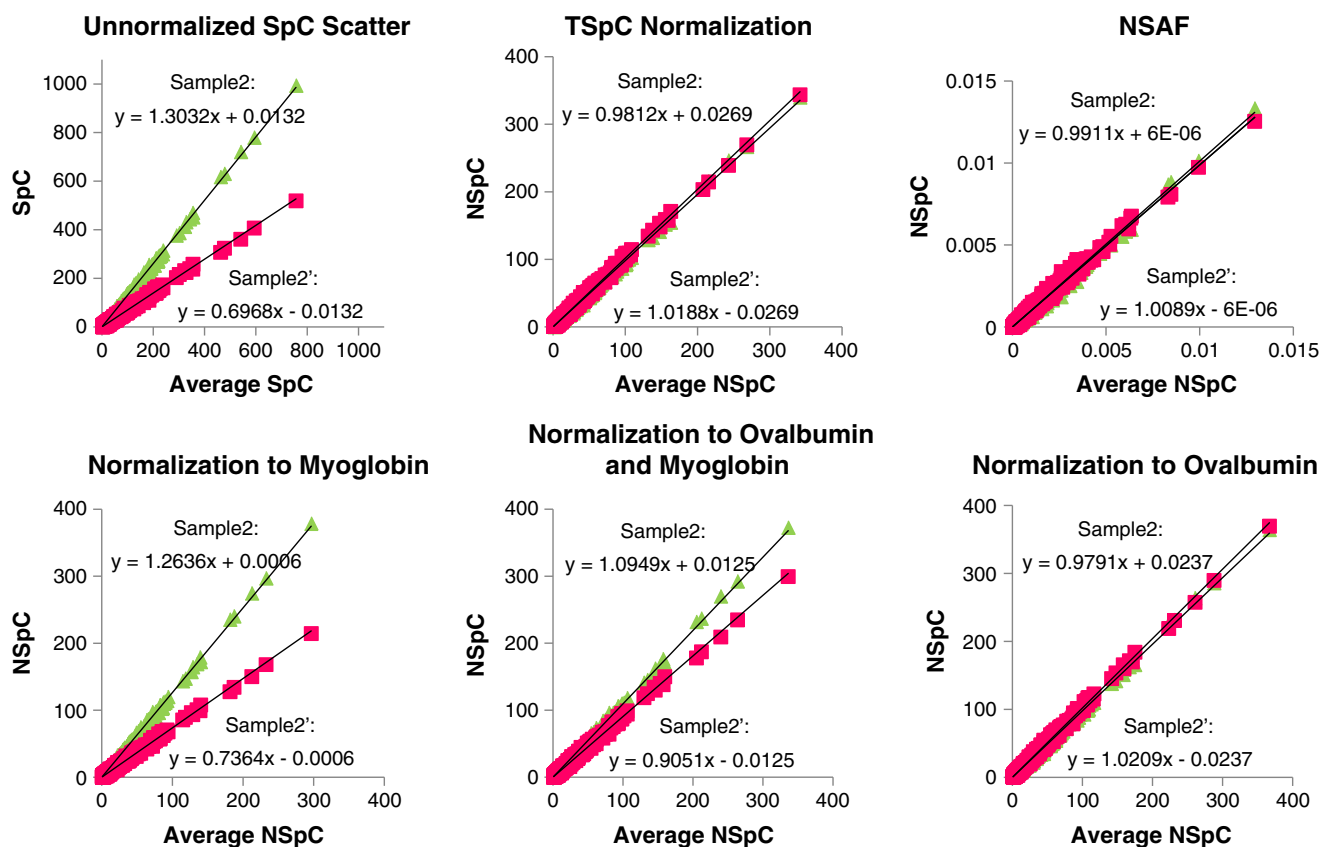


Figure 4. Unnormalized SpC data and NSpCs from combined analysis of sample 2 and sample 2' are plotted versus the average SpCs for each protein. The gross error, mimicked by doubling the complexity of the samples, was adjusted best by NSAF normalization. The unnormalized slope of sample 2' was corrected to 1.0122. Normalizing to myoglobin showed almost no effect, based on the large variation in SpCs smaller proteins experience and therefore lose their reliability

normalized plots show that normalization can correct even for such drastic differences in sample complexity. The slope for sample 2' gets corrected to 1.0188 with TSpC normalization and to 1.0122 with NSAF. In spite of good sequence coverage (see Online Resource 1–Supplementary Figure S1), and the use of two spike-in proteins with different attributes, NSP did not perform as well as TSpC and NSAF normalization; however, better performance was observed for ovalbumin compared to myoglobin. Slopes of 0.9791 for sample 2 and 1.0209 for sample 2' were observed for normalization to ovalbumin, whereas normalization to myoglobin yielded slopes of 1.2636 for sample 2 and 0.7364 for sample 2'. Myoglobin was identified with 21 SpCs (8.98 NSpCs) in sample 1, 21 SpCs (7.18 NSpCs) sample 2, and 20 SpCs (13.24 NSpCs) sample 2' across all three replicates. In comparison, ovalbumin was identified with 63 SpCs (26.66 NSpCs) in sample 1, 88 SpCs (30.1 NSpCs) in sample 2, and 45 SpCs (29.81 NSpCs) in sample 2' correlating with the pattern of TSpCs in each sample (76,638, 95,025, 49,067) (see also Online Resource 1–Supplementary Figure S2 and Supplementary Table S3). This observation suggests that as a larger protein ovalbumin had a greater number of SpC and, thus, was more sensitive to the variations between samples.

As an additional metric, we evaluated which normalization method gives rise to the lowest variance across the technical replicates. TSpC normalization and NSAF resulted in lower median coefficient of variations (CV) for the samples, while the median CVs using NSP were significantly higher when normalized to both spike-in proteins (see Online Resource 1–Supplementary Figure S3). These data indicate that TSpC and NSAF are the superior normalization techniques compared with NSP. This observation is most likely due to the fact that former methods utilize the entire identified proteome for normalization, which allows for better correction of variability within the similar biological samples. NSP may be better suited to instances in which global protein expression differences exist between two biological samples [20].

After identifying the optimal normalization method(s) we wanted to determine the best means for detecting true biological change between two samples. In previous concepts, one specific threshold has been applied to an entire data set in order to define what proteins are changing significantly. However, results from this study and others have shown higher SpC proteins may require separate criteria for detecting significant change due to their lower

variance; conversely, that the high variance of low SpC proteins should preclude them from quantification. Moreover, various methods such as significance testing or applying fold-change thresholds have been utilized with various criteria that are often selected arbitrarily and with disregard to their true predictive value.

To demonstrate this, we produced the volcano plots, shown in Figure 5, with the data obtained here between samples 1 and 2. Volcano plots have long been utilized in the genomic microarray analysis to quickly identify species that have both large and highly significant changes and were more recently applied to proteomic spectral counting data sets by Yates and co-workers [14]. In these plots, the

expression change for a given protein is plotted on the x-axis while the corresponding statistical significance is plotted on the y-axis. In Figure 5a, the plot shows when applying a standard P value cutoff of 0.05, 276 of 1511 proteins identified (~18%) between samples 1 and 2 would be falsely discovered to have changed in abundance. Similarly, 290 such proteins (~19%) would be falsely discovered to have changed if the traditional 2-fold threshold for expression change had been utilized. Interestingly, we also found that these two methods lead to false discovery of different proteins, in particular, different level proteins. As indicated by the Venn diagram in Figure 5a, only 100 of the same proteins were falsely discovered between the 0.05 P value

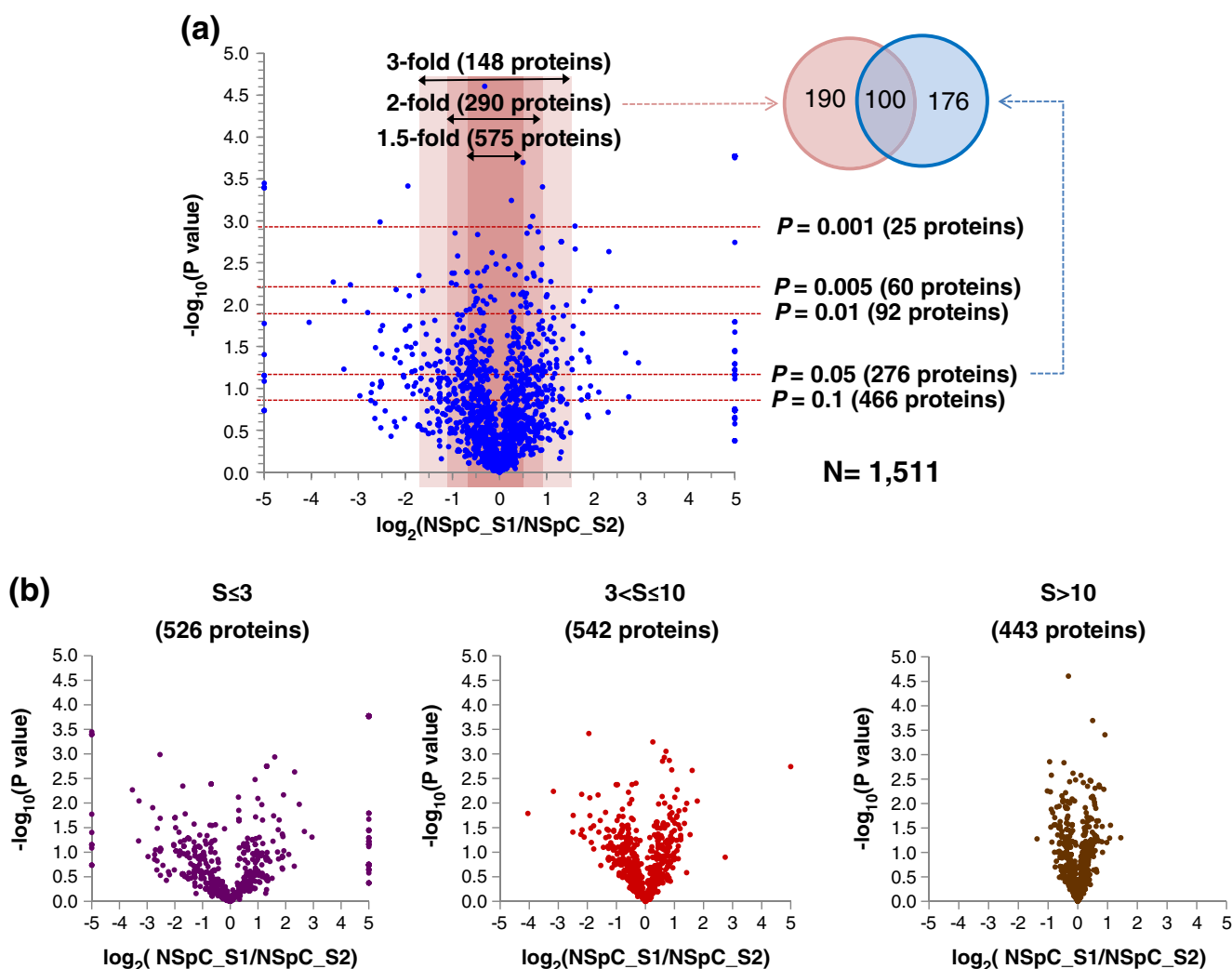


Figure 5. Volcano plots for comparing the normalized SpC between sample 1 and sample 2. The \log_2 expression ratio is plotted versus the $-\log_{10}$ of the P value obtained from significance testing (pairwise t -test). **(a)** Plot comparing all proteins identified between samples 1 and 2. Proteins outside the given fold-change limits or above the P value cutoffs are considered to have significantly changed. The absolute number of proteins meeting each criterion are given in parenthesis. The Venn diagram shows the overlap in the number of proteins deemed to have changed when applying either a 2-fold change cutoff or P value cutoff or 0.05. **(b)** Plots comparing proteins at different SpC levels. The proteins in each plot are defined by S , the SpC obtained per replicate injection

cutoff and 2-fold expression change threshold. Upon closer inspection, nearly all of proteins falsely discovered by the 2-fold cutoff were low abundance (i.e., low SpC) proteins, while those discovered by the 0.05 P value cutoff were slightly biased towards higher abundance proteins. This phenomenon can be seen in Figure 5b, which depicts volcano plots for different SpC levels: $S \leq 3$, $3 < S \leq 10$, and $S > 10$, where S indicates the SpC per replicate (six in this case, three per sample). These plots show the fold-change distribution narrows as S increases, indicating more abundant proteins have lower variance data than less abundant proteins.

In Tables 1 and 2, narrower SpC per replicate (S) bins were utilized to calculate the false positive rates (FPR) achieved at different SpC levels for different stringency criteria. Given we could be certain no biological change had occurred between samples 1 and 2, we were able to define the FPR for a given criterion as the number of proteins discovered (falsely) in that SpC bin by the total number of proteins (N) falling within that same bin. When comparing different fold-change cutoffs (Table 1), there again appears to be a strong propensity for low SpC proteins to be falsely discovered. Across all fold-change cutoffs, the proteins with lower SpC have the higher FPR. Although a higher fold-change cutoff could be applied to reduce the overall FPR for all proteins, these data suggest little confidence could be placed in results for low abundance even when higher stringency cutoffs are applied. For example, if a 2.5-fold cutoff were utilized, the FPR higher abundance proteins ($S > 5$) would be acceptable, but the FPR for lower proteins ($S \leq 5$) would be no better than 10% and would be particularly poor, 39%, for very low abundance proteins ($S < 1.67$).

In Table 2, which shows the FPR for different stringency P values, the opposite trend is observed. Excluding proteins with a low number of SpC ($S < 1.67$), the FPR increases slightly as S proceeds from low values to higher values, indicating that more abundant proteins are more apt to yield a false positive at a given P value. If a single P value were selected, such as 0.01, this data shows both high abundance proteins ($S > 20$) and very low abundance proteins ($S < 1.67$) would yield the majority of false positives. The importance of these observations is 2-fold: first, the large variance of low SpC proteins increases their probability to have erroneously large fold-changes and, secondly, high SpC proteins have a greater propensity to yield low P values simply as a result of their lower variance.

Given the different protein levels are uniquely affected by the two testing methods, we sought to apply dual constraints in order to better control the FPR for all protein levels. In Table 3 are shown different combinations of P value and fold-change cutoffs applied to the entire dataset as well as to different SpC levels. In general, researchers seek to maintain a low FPR; however, using constraints that yield too low of an FPR will result in a high false negative rate (FNR). Thus, we think it reasonable that a FPR of 10% be sought, so as not to exclude too many true positives that would occur during future experiments. The combination that best accomplishes this for the entire dataset, regardless of SpC, is a P value and fold-change cutoff of 0.1 and 2.5, respectively. Notice, this combination utilizes stringent fold-change cutoffs and lax P value cutoffs. This is because the majority of proteins in the dataset have a relatively low SpC (median $S=5.3$) and are more strongly affected by fold-change cutoffs. Consequently, this combination results in undesirably low FPR higher SpC proteins ($S > 5$) and undesirably high FPR for lower SpC proteins ($S \leq 2$). Even when excluding very low abundant proteins from the data set (i.e., when considering only proteins having $S \geq 1.67$), similar outcomes are reach. It is apparent, when looking at the other combinations, there is no single combination that will satisfy the 10% target FPR across all SpC levels. As a result, we suggest applying different constraints to the different SpC levels.

Since large fold-change cutoffs likely result in a large FNR for higher abundance proteins, we suggest applying more stringent P value cutoffs and less stringent fold-change cutoffs for this set of proteins. For instance, we would not set the fold-change cutoff any higher than 1.5-fold for proteins having $S > 10$ as the chances of observing a larger fold-change, particularly a fold increase, for high abundance proteins is reduced due to the low linear dynamic range of spectral counting. Conversely, we would suggest utilizing more stringent fold-change cutoffs and less stringent P value cutoffs for low SpC proteins (i.e., $S \leq 10$). For very low SpC proteins ($S < 1.67$) we would consider excluding these from consideration or applying both stringent P value and fold-change cutoffs due to their disposition to yield false positives. Here we defined very low SpC proteins as those with less than 1.67 SpC per replicate injection, or an average of five total spectral counts between two samples. Earlier studies by Old et al. [11] and Collier et al. [3] also proposed a cutoff of 5 or more

Table 1. False Positive Rates for SpC Bin Widths at Different Fold-Change Cutoffs

Fold-change Cutoff	SpC/Rep (S)						
	$S < 1.67$	$1.67 \leq S \leq 2$	$2 < S \leq 5$	$5 < S \leq 10$	$10 < S \leq 15$	$15 < S \leq 20$	$20 < S$
1.25	80.9	73.7	68.1	53.5	51.8	34.7	24.8
1.5	77.3	59.6	45.0	27.1	21.3	8.3	6.1
2	59.4	28.9	19.6	8.0	2.1	0.0	0.9
2.5	44.2	14.9	10.1	4.0	0.7	0.0	0.4
3	39.4	13.2	6.7	2.3	0.0	0.0	0.0
N	251	114	404	299	141	72	230

Table 2. False Positive Rates for SpC Bin Widths at Different P value Cutoffs

P value Cutoff	SpC/Rep (S)						
	S<1.67	1.67≤S≤2	2<S≤5	5<S≤10	10<S≤15	15<S≤20	20<S
0.1	30.5	25.4	27.7	31.1	36.9	30.6	36.5
0.05	19.8	14.0	14.9	18.1	21.3	16.7	24.3
0.01	10.7	2.6	3.7	4.3	5.7	6.9	9.6
0.005	10.7	0.9	1.7	2.7	3.5	5.6	3.9
0.001	7.8	0.0	0.2	0.7	1.4	0.0	0.4
N	251	114	404	299	141	72	230

total spectral counts across 2 samples, both having triplicate injections, to ensure accurate quantification. Gammulla et al. [32] utilized even more stringent criteria, allowing for quantification of proteins having six or more spectral counts in each sample when triplicate injections were performed. We chose to define our constraints and bin widths using SpC per replicate (S), such that they would be independent of the number injections and samples. Consequently, comparisons will be able to be drawn in future experiments regardless of the number of samples or replicates.

It should be emphasized that these results were obtained in the context of a Gel-based proteomics experiments, which may have led to higher false positive rates than other sample preparation techniques (e.g., MudPIT) due to inherent differences in their reproducibility. Additionally, the use of technical replicates rather than biological replicates here may have contributed to the low variation observed for higher SpC proteins. As such, we would caution readers when applying these same criteria to their data. Instead, we would recommend performing similar control experiments in order

to define the variability specific to their lab, protocol, and sample type.

Conclusions

TSpC normalization, NSAF, and NSP for label-free spectral counting data were investigated on the in-gel tryptic digest of *M. oryzae*. Normalization to TSpC and NSAF normalization revealed very good correlations and low variance for all data sets. With normalization, correcting for variance caused by sample preparation, gel to gel variance, chromatographic performances, and even drastic changes in sample complexity was possible. We evaluated further that accurate quantification is dependent on the number of SpCs. When applying different constraints for significance tests and/or fold-change cutoffs, we observed biases in the FPR across different SpC levels. In particular, we observed higher SpC proteins to have lower variance data and, as a result, required less stringent fold-change cutoffs to achieve accurate quantification. Conversely, lower SpC proteins showed less

Table 3. False Positive Rates for Combined P value and Fold-Change Cutoffs Applied on Different SpC Bin Width

P value Cutoff	Fold-change Cutoff	SpC/Rep (S)								
		ALL	S≥1.67	S<1.67	1.67≤S≤2	2<S≤5	5<S≤10	10<S≤15	15<S≤20	20<S
0.1	1.25	25.7	25.7	25.5	23.7	26.7	28.4	29.1	25.0	19.6
0.1	1.5	18.8	17.5	25.5	22.8	22.3	20.4	17.0	8.3	5.7
0.1	2	10.5	7.7	24.7	16.7	12.6	7.4	2.1	0.0	0.9
0.1	2.5	7.4	4.4	22.3	11.4	7.2	4.0	0.7	0.0	0.4
0.1	3	6.4	3.4	21.1	10.5	5.9	2.3	0.0	0.0	0.0
0.05	1.25	15.0	15.0	15.1	12.3	13.9	17.4	17.0	13.9	14.3
0.05	1.5	11.3	10.5	15.1	11.4	13.1	12.4	9.9	6.9	4.3
0.05	2	6.6	5.1	14.3	9.6	8.7	5.4	0.7	0.0	0.4
0.05	2.5	5.0	3.3	13.5	7.9	5.2	3.7	0.0	0.0	0.0
0.05	3	4.4	2.6	13.5	7.9	4.2	2.3	0.0	0.0	0.0
0.01	1.25	5.2	4.1	10.4	1.8	3.7	4.0	5.0	4.2	5.7
0.01	1.5	3.8	2.5	10.4	1.8	3.5	2.7	2.1	1.4	1.7
0.01	2	2.7	1.3	9.6	1.8	2.7	1.0	0.0	0.0	0.4
0.01	2.5	2.3	1.0	8.8	1.8	2.0	0.7	0.0	0.0	0.0
0.01	3	2.3	1.0	8.8	1.8	2.0	0.7	0.0	0.0	0.0
0.005	1.25	3.4	2.1	10.4	0.9	1.7	2.3	2.8	4.2	1.7
0.005	1.5	2.9	1.4	10.4	0.9	1.5	2.0	1.4	1.4	0.9
0.005	2	2.0	0.5	9.6	0.9	1.0	0.3	0.0	0.0	0.0
0.005	2.5	1.7	0.3	8.8	0.9	0.5	0.3	0.0	0.0	0.0
0.005	3	1.7	0.3	8.8	0.9	0.5	0.3	0.0	0.0	0.0
0.001	1.25	1.5	0.3	7.6	0.0	0.2	0.3	1.4	0.0	0.0
0.001	1.5	1.5	0.2	7.6	0.0	0.2	0.3	0.7	0.0	0.0
0.001	2	1.3	0.1	7.6	0.0	0.2	0.0	0.0	0.0	0.0
0.001	2.5	1.3	0.1	7.6	0.0	0.2	0.0	0.0	0.0	0.0
0.001	3	1.3	0.1	7.6	0.0	0.2	0.0	0.0	0.0	0.0
N		1511	1260	251	114	404	299	141	72	230

reproducibility and required higher fold-change cutoffs in combination with significance testing to ensure accurate quantification. Consequently, we suggest applying different constraints to different SpC levels in order to circumvent these biases and maintain a constant FPR for all proteins.

Acknowledgments

The authors thank Genna Andrews for thoughtful inputs to the paper. They also thank the National Science Foundation Cellular Processes Research Program (grant number MCB-0918611) and Plant Genome Research Program (grant number DBI-092391), the W. M. Keck Foundation, and North Carolina State University for their financial support.

References

- Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., Mann, M.: Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteom.* 376–386 (2002)
- Ong, S.E., Kratchmarova, I., Mann, M.: Properties of C-13-Substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J. Proteome Res.* 173–181 (2003)
- Collier, T.S., Sarkar, P., Franck, W.L., Rao, B.M., Dean, R.A., Muddiman, D.C.: Direct comparison of stable isotope labeling by amino acids in cell culture and spectral counting for quantitative proteomics. *Anal. Chem.* 82(20), 8696–8702 (2010)
- Collier, T.S., Randall, S.M., Sarkar, P., Rao, B.M., Dean, R.A., Muddiman, D.C.: Comparison of stable-isotope labeling with amino acids in cell culture and spectral counting for relative quantification of protein expression. *Rapid Commun. Mass Spectrom.* 25(17), 2524–2532 (2011)
- Liu, H.B., Sadygov, R.G., Yates, J.R.: A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* 76(14), 4193–4201 (2004)
- Wang, W.X., Zhou, H.H., Lin, H., Roy, S., Shaler, T.A., Hill, L.R., Norton, S., Kumar, P., Anderle, M., Becker, C.H.: Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal. Chem.* 75(18), 4818–4826 (2003)
- Wiener, M.C., Sachs, J.R., Deyanova, E.G., Yates, N.A.: Differential mass spectrometry: A label-free LC-MS method for finding significant differences in complex peptide and protein mixtures. *Anal. Chem.* 76(20), 6085–6096 (2004)
- Bondarenko, P.V., Chelius, D., Shaler, T.A.: Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. *Anal. Chem.* 74(18), 4741–4749 (2002)
- Chelius, D., Bondarenko, P.V.: Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J. Proteome Res.* 1(4), 317–323 (2002)
- Johnson, K.L., Mason, C.J., Muddiman, D.C., Eckel, J.E.: Analysis of the low molecular weight fraction of serum by LC-Dual ESI-FT-ICR mass spectrometry: Precision of retention time, mass, and ion abundance. *Anal. Chem.* 76(17), 5097–5103 (2004)
- Old, W.M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K.G., Mendoza, A., Sevinsky Jr., Resing, K.A., Ahn, N.G.: Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell. Proteom.* 4(10), 1487–1502 (2005)
- Gao, B.B., Stuart, L., Feener, E.P.: Label-free quantitative analysis of one-dimensional PAGE LC/MS/MS proteome. *Mol. Cell. Proteom.* 7(12), 2399–2409 (2008)
- Lundgren, D.H., Hwang, S.I., Wu, L.F., Han, D.K.: Role of spectral counting in quantitative proteomics. *Expert Rev. Proteom.* 7(1), 39–53 (2010)
- Carvalho, P.C., Fischer, J.S., Chen, E.I., Yates, J.R. III., Barbosa, V.C.: PatternLab for proteomics: A tool for differential shotgun proteomics. *BMC Bioinformatics*, 9 (2008)
- Dong, M.Q., Venable, J.D., Au, N., Xu, T., Park, S.K., Cociorva, D., Johnson, J.R., Dillin, A., Yates, J.R.: Quantitative Mass Spectrometry Identifies Insulin Signaling Targets in *C. Elegans*. *Science* 317(5838), 660–663 (2007)
- Zybailov, B., Mosley, A.L., Sardi, M.E., Coleman, M.K., Florens, L., Washburn, M.P.: Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* 5(9), 2339–2347 (2006)
- Florens, L., Carozza, M.J., Swanson, S.K., Fournier, M., Coleman, M.K., Workman, J.L., Washburn, M.P.: Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods* 40(4), 303–311 (2006)
- Sardi, M.E., Cai, Y., Jin, J.J., Swanson, S.K., Conaway, R.C., Conaway, J.W., Florens, L., Washburn, M.P.: Probabilistic assembly of human protein interaction networks from label-free quantitative proteomics. *Proc. Natl. Acad. Sci. U.S.A.* 105(5), 1454–1459 (2008)
- Nagy, I., Sun, N., Pan, C.P., Nickell, S., Mann, M., Baumeister, W.: Quantitative proteome and transcriptome analysis of the archaeon *Thermoplasma acidophilum* cultured under aerobic and anaerobic conditions. *J. Proteome Res.* 9(9), 4839–4850 (2010)
- Kuo, J., Han, X., Hsiao, C., Yates, J.R.I.I.I., Waterman, C.M.: Analysis of the myosin-II-responsive focal adhesion proteome reveals a role for beta-pix in negative regulation of focal adhesion maturation. *Nat. Cell. Biol.* 13(4), 383–U109 (2011)
- Zhang, B., VerBerkmoes, N.C., Langston, M.A., Uberbacher, E., Hettich, R.L., Samatova, N.F.: Detecting differential and correlated protein expression in label-free shotgun proteomics. *J. Proteome Res.* 5(11), 2909–2918 (2006)
- Cooper, B., Feng, J., Garrett, W.M.: Relative, Label-Free Protein Quantitation: Spectral Counting Error Statistics from Nine Replicate MudPIT Samples. *J. Am. Soc. Mass Spectrom.* 21(9), 1534–1546 (2010)
- Carvalho, P.C., Fischer, J.S.G., Perales, J., Yates, J.R., Barbosa, V.C., Bareinboim, E.: Analyzing marginal cases in differential shotgun proteomics. *Bioinformatics* 27(2), 275–276 (2011)
- Ou, S.H.: Pathogen variability and host-resistance in rice blast disease. *Annu. Rev. Phytopathol.* 18, 167–187 (1980)
- Khush, G.S.: What it will take to feed 5.0 billion rice consumers in 2030. *Plant Mol. Biol.* 59(1), 1–6 (2005)
- Dean, R.A., Talbot, N.J., Ebbole, D.J., Farman, M.L., Mitchell, T.K., Orbach, M.J., Thon, M., Kulkarni, R., Xu, J.R., Pan, H.Q., Read, N.D., Lee, Y.H., Carbone, I., Brown, D., Oh, Y.Y., Donofrio, N., Jeong, J.S., Soanes, D.M., Djonovic, S., Kolomiets, E., Rehmeier, C., Li, W.X., Harding, M., Kim, S., Lebrun, M.H., Bohnert, H., Coughlan, S., Butler, J., Calvo, S., Ma, L.J., Nicol, R., Purcell, S., Nusbaum, C., Galagan, J.E., Birren, B.W.: The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* 434(7036), 980–986 (2005)
- Shevchenko, A., Wilm, M., Vorm, O., Mann, M.: Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. *Anal. Chem.* 68(5), 850–858 (1996)
- Andrews, G.L., Shuford, C.M., Burnett, J.C., Hawkrigde, A.M., Muddiman, D.C.: Coupling of a vented column with splitless NanoRPLC-ESI-MS for the improved separation and detection of brain natriuretic peptide-32 and its proteolytic peptides. *J. Chromatogr. B* 877(10), 948–954 (2009)
- Andrews, G.L., Dean, R.A., Hawkrigde, A.M., Muddiman, D.C.: Improving proteome coverage on a LTQ-orbitrap using design of experiments. *J. Am. Soc. Mass Spectrom.* 22(4), 773–783 (2011)
- Olsen, J.V., de Godoy, L.M.F., Li, G.Q., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., Mann, M.: Parts per Million Mass Accuracy on an Orbitrap Mass Spectrometer via Lock Mass Injection into a C-Trap. *Mol. Cell. Proteom.* 4(12), 2010–2021 (2005)
- Weatherly, D.B., Atwood, J.A., Minning, T.A., Cavola, C., Tarleton, R.L., Orlando, R.: A heuristic method for assigning a false-discovery rate for protein identifications from mascot database search results. *Mol. Cell. Proteom.* 4(6), 762–772 (2005)
- Gammulla, C.G., Pascovici, D., Atwell, B.J., Haynes, P.A.: Differential metabolic response of cultured rice (*Oryza sativa*) cells exposed to high- and low-temperature stress. *Proteomics* 10(16), 3001–3019 (2010)