Relative, Label-free Protein Quantitation: Spectral Counting Error Statistics from Nine Replicate MudPIT Samples

Bret Cooper,^a Jian Feng,^b and Wesley M. Garrett^c

^a Soybean Genomics and Improvement Laboratory, USDA-ARS, Beltsville, Maryland, USA

^b Snoqualmie, Washington, USA

^c Animal Biosciences and Biotechnology Laboratory, USDA-ARS, Beltsville, Maryland, USA

Nine replicate samples of peptides from soybean leaves, each spiked with a different concentration of bovine apotransferrin peptides, were analyzed on a mass spectrometer using multidimensional protein identification technology (MudPIT). Proteins were detected from the peptide tandem mass spectra, and the numbers of spectra were statistically evaluated for variation between samples. The results corroborate prior knowledge that combining spectra from replicate samples increases the number of identifiable proteins and that a summed spectral count for a protein increases linearly with increasing molar amounts of protein. Furthermore, statistical analysis of spectral counts for proteins in two- and three-way comparisons between replicates and combined replicates revealed little significant variation arising from run-to-run differences or data-dependent instrument ion sampling that might falsely suggest differential protein accumulation. In these experiments, spectral counting was enabled by PANORAMICS, probability-based software that predicts proteins detected by sets of observed peptides. Three alternative approaches to counting spectra were also evaluated by comparison. As the counting thresholds were changed from weaker to more stringent, the accuracy of ratio determination also changed. These results suggest that thresholds for counting can be empirically set to improve relative quantitation. All together, the data confirm the accuracy and reliability of label-free spectral counting in the relative, quantitative analysis of proteins between samples. (J Am Soc Mass Spectrom 2010, 21, 1534–1546) © 2010 American Society for Mass Spectrometry

Multidimensional protein identification technology (MudPIT), a qualitative mass spectrometry (MS) method for the detection of many proteins in a sample, is becoming ever more quantitative [1, 2]. Instrumentation improvements for mass precision, sensitivity, selectivity, and speed, along with advances in software and computational approaches, are changing the ways that amounts of proteins are measured relative to one another in a single sample or between samples [1]. These quantitative applications are enabling systems biology studies and have led to broader understandings of protein dynamics in cells [3–6].

Relative protein quantitation in MudPIT takes two forms: labeled and label-free. Labels include isotopes or isobaric tags that shift the mass of a labeled protein or peptide. If two differentially labeled samples are analyzed concurrently, an expected shift in the mass spectrum can be observed for a parent or daughter ion from one of the samples and the relative amounts of proteins or peptides determined from the differential ion peak

heights (ion signal strength), ion peak areas, or numbers of tandem mass spectra produced [3, 7–13]. In the label-free approaches, parent ion peaks, parent ion mass counts, or numbers of tandem mass spectra can also be used to determine amounts of protein or peptide in a sample [4, 14–20]. However, label-free samples cannot be analyzed concurrently in a mass spectrometer because there is no predictable mass shift to enable a direct comparison. Therefore, while the label-free methods may be more convenient in that they save financial or time costs associated with labeling, they are thought to be less accurate than labeling methods [1]. In reality, label and label-free methods are independently supported by empirical data that confirm their quantitative accuracy within a relative context.

In this light, we have independently pursued a label-free method of counting tandem mass spectra for deducing protein accumulation changes between samples. Our decision to pursue a label-free method over labeling was based on the difficulty of incorporating isotopic labels into the metabolic cycles of live agricultural pathogens and plants that we study and our lack of resources for developing and using labeling chemistries. Instead, we invented a software platform, PANORAMICS, that allows us to analyze large amounts

Address reprint requests to Dr. B. Cooper, Soybean Genomics and Improvement Laboratory, 10300 Baltimore Ave., Bldg. 006, Rm. 213, Beltsville, MD 20705, USA. E-mail: bret.cooper@ars.usda.gov

of MudPIT data, and we were able to configure the program to perform tandem mass spectral counting [21]. We now regularly use MudPIT to identify hundreds and thousands of proteins from pathogenic fungi and their host plants and have used PANORAMICS to perform spectral counting and show alterations in protein accumulation that may be associated with developmental cues and disease resistance [4, 22].

To justify our approach in the past, we relied upon the reports of others who demonstrated that tandem mass spectral counts for standard protein samples can rise linearly with increasing molar amounts of protein over at least 2 orders of magnitude ($\log_{10}$) [15, 17, 20]. Now, to verify our own particular spectral counting approach, we have analyzed different molar amounts of bovine apotransferrin in the backgrounds of nine complex but replicate mixtures of soybean peptides. We expected that the experiments would allow us to test three things: (1) whether our method could distinguish a single statistically significant protein accumulation change in the midst of a large background of proteins; (2) whether there were any statistically significant accumulation changes among background soybean proteins—theoretically, there should be none and any observed changes would be considered erroneous and point to flaws in the method; (3) if altering rules for selecting spectra for counting would produce quantitative results that are more or less accurate than expected values. The following results from these tests confirm the reliability of spectral counting for the relative quantitative assessment of protein accumulation between multiple samples using MudPIT.

Methods

Peptide Samples

Glycine max cv. Williams 82 was grown during the summer months in an insect-free, controlled-environment greenhouse. This variety is important because it is the parent to many soybean production cultivars grown in the U.S. and because its genome is sequenced [23]. The first, fully-opened trifoliate leaves from 3 to 4 wk-old plants ~12 inches tall were collected from 50 plants. Leaves were frozen in liquid nitrogen and ground in acetone/10% trichloroacetic acid/0.07% β -mercaptoethanol. The slurry was centrifuged at $35,000 \times g$ and the pellet washed three times in acetone/0.07% β -mercaptoethanol. Proteins from 20 mg of dried pellet were resolubilized in 8 M urea/100 mM Tris-HCl pH 8.5/1% β -dodecyl maltoside, filtered through a 5 μ m membrane, precipitated in 25% trichloroacetic acid, centrifuged at $21,000 \times g$ and washed three times in acetone. The final protein pellet was resuspended in 8 M urea/100 mM Tris-HCl pH 8.5. Protein concentrations were estimated by bicinchoninic assay (Pierce, Rockford, IL, USA). Nine samples each containing 1 mg soluble protein were reduced, carboxyamidomethylated, and digested with trypsin [24]. The

digested samples were desalted using solid-phase extraction with SPEC-PLUS PT C18 columns (Varian, Lake Forrest, CA, USA). The samples were combined and then equally divided into 1 mg protein aliquots among nine separate vials. Different molar amounts of tryptic digests of bovine apotransferrin (Michrom Bioresources, Auburn, CA, USA) were added to the vials.

Mass Spectrometry

Peptides from vials were separated on home-made 75 μ m i.d. fused-silica columns with a 5 μ m tip and packed first with reverse phase C18 resin (Aqua, 5 μ m; Phenomenex, Torrance, CA, USA) followed by strong cation exchange resin (Luna, 5 μ m; Phenomenex) [24]. A 12-step elution procedure consisting of step-wise increasing concentrations of salt solution followed by increasing gradients of organic mobile phase was used [24]. Solvent flow was 200 nL/min and was controlled with an Accela HPLC pump (Thermo Fisher Scientific, Waltham, MA, USA) and with a T-split junction where 2100 V electricity was applied [24]. The eluent was electrosprayed directly into the orifice of an LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific) controlled by Xcalibur 2.0.7 software (Thermo Fisher Scientific). A parent-ion scan was performed in the Orbitrap over the range of 400–1600 m/z at 30,000 resolution, with 500,000 automatic gain control (AGC), 500 ms ion injection time, and 1 μ scan. Lock-mass was enabled [25]. Data-dependent MS² was performed in the linear ion trap with 10,000 AGC and 100 ms ion injection times with 1 μ scan and on the 10 most intense MS ions with a minimum signal of 1000. An isolation width of 2 m/z and normalized collision energy of 35% were used. Dynamic exclusion was used with repeat count of 1, 30 s repeat duration, a list of 500, list duration of 2 min, and exclusion mass width of ± 0.7 Da.

Mascot Searching

MS² spectrum data files were extracted from the raw data with Bioworks 3.3.1 (Thermo Fisher Scientific) using the parameters 600–4500 mass range, 1 group scan, 1 minimum group count, and 5 minimum ion counts. Sets of combined MS² spectra were searched with Mascot 2.2.04 [26]. Search parameters were for tryptic digests, one possible missed cleavage, fixed amino acid modification [+57, C], variable amino acid modifications [+16, M], monoisotopic mass values, ± 10 ppm parent ion mass tolerance, ± 0.8 Da fragment ion mass tolerance and #13C = 1 enabled. The searched database consisted of soybean genome [23] protein reference sequences (Glyma1.pep.fa from www.phytozome.net; 75,781 records), the apotransferrin record (gi 2501351), and common contaminants (75,951 records total).

Protein Identification

Mascot output files were processed by PANORAMICS, a probability-based program that determines the likelihood that peptides are correctly assigned to proteins [21]. PANORAMICS considers for each spectrum up to 10 ranking peptide matches made by Mascot with Ions/Identity score differences not less than -10 and calculates the probabilities that these matches are correct. Generally, peptide matches with higher Ions scores receive higher probabilities while peptide matches with lower Ions scores receive lower probabilities. However, the database size and the length and charge state of each peptide sequence are also factored. A probability for a protein or a group of proteins with the same set of matching peptides is then calculated from the corresponding peptide probabilities. Generally, high protein probabilities are a product of several high-scoring peptide matches, but lower-scoring and lower-ranking peptides can also contribute a smaller amount of probability to protein identification. Full probability values for distinct peptides (unique to a protein or group) are used in protein probability calculations whereas portions of the probabilities for shared peptides (those common to different groups) are distributed across the sharing groups such that greater portions are distributed to the most probable proteins. As a result of the probability calculations and the groupings, parsimony is achieved. Protein groups were ultimately treated as single proteins and we arbitrarily chose one record from the list of all records in the group as the one to represent the group. Proteins with probabilities $> 95\%$ were considered for spectral counting.

Spectral Counting

PANORAMICS counted a spectrum assigned to a peptide if an ion score produced a positive probability, which required the Mascot Ions/Identity score difference to be not less than -10 and if the score was top-ranking (note that the rule for counting only top-ranking matches is different from the model for calculating probabilities where all ranking matches are considered). A distinct peptide count was based on the total number of spectra satisfying the preceding criteria while a shared peptide count was divided by the number of protein groups in which it was shared. An experimental version of PANORAMICS was also produced to count spectra assigned to peptides regardless of Mascot Ion score rank (ALLCOUNT), to count spectra to all peptides with a top-ranking Mascot Ions score regardless the Ions/Identity score difference (TOPCOUNT) and to count spectra assigned to peptides if the Ions score was top-ranking and if the Ions/Identity score difference was not less than zero (IDENTCOUNT). Spectral counts assigned to each peptide in an identified protein or protein group were summed.

Statistical Analysis

Quantitative analysis was limited to the same sets of proteins identified between replicates or combinations of replicates. A G-test was used to assess the statistical differences between summed spectral counts for each protein [17], with the hypothesis being that the summed spectral count of any protein A was equal between replicates or combinations. Normalization values were the total sum of spectral counts for all considered proteins in a replicate or combination of replicates. The corresponding P value was calculated from χ^2 distribution with 1° of freedom for a two-way comparison or 2° of freedom for a three-way comparison. The Benjamini and Hochberg correction was applied such that only a protein with P value $< q(j/t)$ was considered as having differential summed spectral counts between samples. In this equation, j is the rank of the protein in the list of t proteins sorted in ascending order by P value and q is the false discovery rate. In essence, the correction uses the P value to establish a false discovery rate q . Here, q was set for false discovery rates of 5% and 1%. Alternatively, a t -test was used to compare the normalized summed spectral counts for a protein across three replicates to the same in another set of three replicates. Normalization values were the total sum of spectral counts for all considered proteins in a replicate. The Benjamini and Hochberg correction was also applied to the t -test P values.

Theory

Rationale for Quantitative Method

Our relative quantitative proteomic method begins with digesting proteins in a sample and generating peptide tandem mass spectra. The spectra are then matched to candidate peptide sequences by Mascot and the peptides are “reassembled” into proteins using the probability model of PANORAMICS. Proteins exceeding a defined probability cut-off are considered significant; we typically choose 95%. Then, spectral counts for the peptides comprising each protein are summed to generate a protein spectral count. Summed spectral counts for the same protein identified between samples can then be compared with each other, and a G-test is used to determine significant differences. Several decisions pertaining to this method are detailed:

1. We sum the spectral count for all peptides comprising a protein as opposed to counting one or a few representative “proteotypic” peptides for a protein [27, 28]. Our choice logically fits the PANORAMICS probability model that determines which proteins are then quantitatively examined: Protein probabilities are a product of all significant peptide matches that contribute to protein identification. Therefore, the spectra for those peptides are counted. A comparison of counting spectra of all significant pep-

tides versus proteotypic peptides is not part of this study.

2. PANORAMICS does not count all spectra submitted to and analyzed by Mascot. A spectrum is counted if its corresponding Mascot Ions score produced a positive peptide probability, if the Mascot Ions score was top-ranking among possible candidate peptides and if the Mascot Ions score/Identity score difference was not less than -10 . For example, PANORAMICS does not count a spectrum if the Ions score for a peptide match is 28 and the Identity score is 40. A study on the effects of alternative counting methods is presented below.
3. Because of random sampling effects associated with high-throughput mass spectrometry [15], it is usually difficult to determine if the absence of any protein from a dataset is a result of it not being in the sample or just a by-product of the chance of not detecting it. The latter is most likely for these experiments, as we show later. Additionally, in these experiments, a protein could be identified with a probability of 95.1% in one sample, but 94.9% in another. As a result of a cut-off for accepting proteins identified with probabilities of 95% or greater, the protein would remain in the first dataset but not the other. In the past, we have accepted lower protein identification probabilities as long as there was a higher probability for identification in another sample [4]. Here, however, we have limited our analysis of spectral count variability to proteins common between datasets and exceeding the 95% cut-off.
4. Spectra can be matched to peptide sequences that are distinct to a protein or shared among a set of proteins (usually members of a family of proteins) [29]. Logically, if a peptide is shared among two different protein groups A and B and there are two spectral counts, then the counts could be distributed all to A, all to B, or split between A and B if A and B coexist in the sample. Unfortunately, because the peptide organization and distribution information is lost upon digestion, there is no way to know how the counts should be truly assigned. Several labs have experimented with distributing counts of shared peptides across proteins based on estimations [30–33], yet these schemes rely on best guesses. However, these studies have shown that relative protein quantitation is more accurate when shared peptides are included in the counting scheme in some fashion. With these findings in mind, we have decided to include spectral counts for shared peptides but divide the count by the number of protein groups in which the shared peptides appear, to prevent shared peptides from being counted more than once. This unbiased distribution creates several circumstances that should be noted. Assume protein group A with probability 97% and protein group B with probability 94% share a peptide. In this study, A is considered for analysis

but B is not because of the 95% cut-off. However, the shared peptide count for A is still divided by 2 even though B is no longer considered in the dataset. While it may be possible to reassign the remaining count of the shared peptide in B to A, we do not do so for the sake of simplicity. As a result, possibly a slight “under-counting” error is created for peptide A. We then assume this under-counting error is relegated to the protein group(s) that shares the peptide. Certainly, this potential error is minimized in protein groups with a preponderance of distinct peptides. For protein groups preponderant with shared peptides, these proteins are highly homologous and may have nearly indistinguishable summed spectral counts as a result, despite the over-count. If there are differences, these should theoretically be predicated on the counts of any distinct peptides. For this study, we remain consistent with the counting of shared peptides, but recognize that this topic is worthy of future consideration.

5. We evaluate statistical differences between summed spectral counts for common proteins between samples using a G-test. The G-test has a number of features that make it appealing for analyzing MudPIT data: (a) multiple MudPIT analyses are often required to overcome sampling limitations inherent to shotgun mass spectrometry [15, 34]. By pooling the spectra from the replicate analyses and performing a unified Mascot search, the number of peptides associated with each protein increases (thereby increasing spectral count as well) and the total number of proteins identified increases too. The drawback is that means and variances of spectral counts cannot be estimated for proteins among the replicates once files are combined. This in turn eliminates the use of statistical analyses that rely on these values to make comparisons. However, the G-test does not need to estimate the means and variances, but instead directly tests a null hypothesis that an observed count results from sampling from a distribution of counts in a population. Clearly, combining datasets favors the G-test. (b) The G-test considers population size, and spectral counts are normalized to the totals for the datasets as part of the analysis. Therefore, it is not necessary to normalize to any single reference standard. (c) Some statistical tests are limited to pair-wise comparisons, but the G-test can be generalized to suit examining multiple conditions [17]. This favors proteomics where several samples subjected to different treatments might be compared with a control. The following results will show the reliability of the G-test in determining significant differences and its advantage over a *t*-test. Readers interested in comparisons between other statistical tests used to evaluate differences in spectral counts are referred to Zhang et al. 2006 [17].
6. Finally, in this study the soybean replicate samples analyzed were theoretically equivalent except for

having different amounts of a single bovine protein, apotransferrin. Therefore the samples were theoretically free from uncontrolled variability. Furthermore, the data for the samples was treated equivalently except where noted below. However, there were other factors that were not entirely controllable that had the potential to affect spectral counts. To our best ability, we used good standard pipetting practices and calibrated instruments to reduce error, packed all columns with the same preparation of separation materials, and packed the columns with tolerances no less than 0.5 cm and no greater than 1 cm to the lengths describe above. The mass spectrometer was tuned and all experiments run subsequently with respect to the same tune file parameters. Other factors that had the potential to impact count, but remained static for these experiments, were number of ions selected for tandem mass spectrometry and parameters for dynamic exclusion. While our parameters for dynamic exclusion were based on those that have been optimized for relative quantitation by spectral counting [35], possibly future research could be performed to refine these parameters or those for ion selection.

Results and Discussion

Proteins from Nine Replicate Samples

Each of the nine replicate soybean peptide samples was spiked with a different molar amount of tryptic peptides from a bovine apotransferrin tryptic digest, ranging from 0.005 to 2.5 pmol and analyzed by MudPIT. We found that the summed spectral counts for apotransferrin increased linearly with the amounts of apotransferrin between 0.5 and 2.5 pmol (Figure 1; Table 1, column “ Σ spectral count apotransferrin”). A best fit line with $R^2 = 0.9585$ revealed consistent reproducibility of spectral counts between duplicated analyses on 0.5, 1.0, and 2.5 pmolar amounts (Figure 1). Apotransferrin was not detected above 95% probability in samples with 0.005 or 0.05 pmol, which establishes our lower limit of detection between 0.05 and 0.5 pmol. This is similar to Old et al. [20], who examined an apotransferrin dynamic range between 0.05 and 5.0 pmol and found the lower limit of detection to be between 0.05 and 0.3 pmol. In addition, Old et al. [20] observed a loss of spectral count linearity above 30. This was not the case for us and is likely due to the faster duty-cycle of our mass spectrometer and our adjusted dynamic exclusion parameters.

A more interesting picture emerged when evaluating the soybean proteins present in the samples. Despite the samples theoretically being comprised of the same soybean peptide mixtures, the numbers of proteins found above a 95% confidence level ranged from 1474 to 2127 proteins between replicates (Table 1, column “proteins > 95%”). This run-to-run variability is most

Figure 1. Relationship between number of spectra associated with peptides for apotransferrin and the molar amount of apotransferrin added to a sample. Soybean peptide samples were spiked with 0.005, 0.05, 0.5, 1.0, and 2.5 pmol apotransferrin and were analyzed by MudPIT twice (except 0.005, which was only analyzed once). Apotransferrin at 0.005 and 0.05 pmolar amounts were not resolved by this method. Spectral count values for the pairs of samples with 0.5, 1.0, and 2.5 pmol apotransferrin in Table 1 correspond with this figure.

likely a result of ion sampling variability inherent to MudPIT [15, 34].

While the numbers of proteins detected between samples varied by 30%, the total numbers of spectra collected in each replicate run varied only by 10% (ranging from 252,035 to 276,007 spectra; Table 1, column “total spectra collected”). In addition, there was no perfect correlation between number of proteins found and total number of spectra collected. In fact for replicate no. 5, there was an instrument/software communication glitch in five of the separation steps of that run that prevented the recording of all spectra observed. We included this imperfect run in this analysis as a way of deliberately introducing variability. Interestingly, while replicate no. 5 had the smallest number of total spectra collected, it did not yield the smallest number of identified proteins. Rather, replicate no. 7 yielded the smallest number of proteins, yet it had nearly 10,000 more spectra than replicate no. 5. Therefore, other factors besides total numbers of spectra collected, e.g., total numbers of good peptide spectra generated, have more direct bearing on number of total proteins detected.

When spectra from the replicates were combined (pooled), the number of identified proteins increased (Table 1, column “proteins > 95%”). This demonstrates the benefit to protein discovery when combining spectra from replicate runs. As proteins from each of nine replicates were sequentially combined, the numbers of new proteins increased by ~200 per replicate to a final total of 3502 (Figure 2). The saturation point for protein discovery was not reached even after nine replicate analyses, meaning that additional analyses would be necessary for deeper coverage of leaf proteins.

Table 1. Protein and spectral count statistics for replicate and combined replicate files

Soybean replicate	Apo transferrin pmol	Total spectra collected	Proteins >95% ^a	Σ Spectral count for proteins >95% ^a	Σ Spectral count apotrans ferrin	Σ Spectral count ALL COUNT (proteins >95%)	Σ Spectral count TOP COUNT (proteins >95%)	Σ Spectral count IDENT COUNT (proteins >95%)
1	0.5	276,007	1,829	21,774	17	30,027	25,083	16,489
2	0.05	272,263	1,796	22,018	0 (nf)	30,041	25,113	14,422
3	0.005	270,060	1,474	18,827	0 (nf)	—	—	—
4	0.5	270,991	1,539	20,671	14	28,830	24,205	15,318
5	0.05	252,035	1,542	20,540	0 (nf)	—	—	—
6	2.5	255,497	1,908	24,456	60	33,816	28,190	18,551
7	1	262,736	1,516	20,873	36	29,018	24,375	15,230
8	1	261,060	1,740	23,200	37	—	—	—
9	2.5	266,927	2,127	26,079	69	—	—	—
2+3	0.055	542,323	2,019	41,759	0 (nf)	—	—	—
1+2	0.55	548,270	2,271	44,654	18	—	—	—
4+5	0.55	523,026	2,071	42,478	16	—	—	—
2+4	0.55	543,254	2,057	43,864	15	—	—	—
1+5	0.55	528,042	2,161	43,360	19	—	—	—
3+7	1.005	532,796	1,878	40,714	36	—	—	—
2+8	1.05	533,323	2,297	46,343	38	—	—	—
4+7	1.5	533,727	1,929	42,669	50	—	—	—
3+6	2.505	525,557	2,225	44,345	60	—	—	—
3+9	2.505	536,987	2,418	46,075	69	—	—	—
5+6	2.55	507,532	2,198	46,113	62	—	—	—
1+6	3	531,504	2,366	47,292	77	—	—	—
1+2+3	0.555	818,330	2,430	64,685	18	—	—	—
1+2+5	0.6	800,305	2,508	66,557	20	—	—	—
1+3+4	1.005	817,058	2,380	63,668	31	—	—	—
1+3+7	1.505	808,803	2,365	63,890	53	—	—	—
4+5+8	1.55	784,086	2,470	67,179	53	—	—	—
7+8+5	2.05	775,831	2,361	67,195	75	—	—	—
4+5+6	3.05	778,523	2,533	68,284	76	—	—	—
3+6+7	3.505	788,293	2,423	66,656	96	—	—	—
2+8+9	3.55	800,250	2,808	73,760	107	—	—	—
6+9+2	5.05	794,687	2,828	74,884	130	—	—	—

nf = not found.

^aProtein probability as determined by PANORAMICS.

Statistical Analysis of Replicate Variation

Since it was clear that there was variability between replicates in terms of proteins discovered, we were concerned that this variability might also affect the accuracy of spectral counting. Because label-free spectral counting relies on samples being analyzed consecutively rather than concurrently, we worried that it might be difficult to distinguish the error inherent to the run-to-run variability from bona fide changes among true experimental samples.

To test whether ion sampling error could adversely influence spectral counting, we performed two- and three-way comparisons of replicates using a G-test. There was not full congruency of overlapping proteins between replicates. For example, replicates no. 1 and no. 6 had 1829 and 1908 proteins, respectively, but there were only 1374 proteins in common between them (Table 2, column “common proteins”). These differences appear to be a product of run-to-run variability (potential sources will be investigated in a later section). As for the overlapping proteins, eight exhibited differ-

entially significant spectral counts when the false discovery rate was set at 5% [Table 2, column “proteins with different spectral counts (G-test, sig. > 95%)”]. Since zero was theoretically expected, these proteins with differential counts were falsely discovered by this method. However, the overall false discovery rate, empirically measured by dividing the number of falsely discovered proteins by the total number of proteins for which no variation was expected, was just 0.6% and within the expected false discovery rate set by the *q*-value [Table 2, “variance per common protein (sig. > 95%)”]. For several other combinations examined, between one and 39 proteins exhibited differential spectral counts [Table 2, column “proteins with different spectral counts (G-test, sig. > 95%)”], and the empirical false discovery rates ranged from 0.1% to 3.8% [Table 2, column “variance per common protein (sig. > 95%)”]. For the three-way comparisons, the sets of overlapping proteins were generally smaller, as expected. However, the number of proteins with significantly different spectral counts rose somewhat, with the empirical false

Figure 2. Number of proteins identified at >95% confidence level with each added replicate file. As spectral datasets from each replicate run are combined and then analyzed by Mascot and PANORAMICS, the number of proteins identified increases, thus attributing to the benefit of performing replicate analyses and combining spectral datasets. This graph suggests that more than nine replicate analyses would be required for comprehensive soybean protein identification from these replicate samples.

discovery rate reaching 7.5% in one example [Table 2, column “variance per common protein (sig. > 95%)”]. Nonetheless, the average empirical false discovery rates are in line with the preset 5% q -values; given the limited number of trials, we believe the empirical false discovery rates are very close to those predicted and constrained by the statistical tests.

Earlier, we showed that the numbers of proteins detected increased as spectra from replicates were pooled (Table 1, column “proteins > 95%”). Hence, we thought that pooling spectra from two or three replicates and comparing pooled data to other merged replicates would result in increased numbers of proteins that were common between them. This was generally true, as combinations of two replicates had between 1500 and 1700 proteins in common, and combinations of three replicates had ~1900 proteins in common for two-way comparisons (Table 2, column “common proteins”). The overlap was smaller for three-way comparisons, but the number of overlapping proteins still increased as more spectra were pooled. We therefore assumed that pooling spectra from replicates might give the effect of increased statistical sampling which could theoretically reduce the amount of variation caused by ion sampling. Surprisingly, spectral count variability did not bottom out. Instead, empirical false discovery rates remained at the levels predicted by the statistical tests [Table 2, column “variance per common protein (sig. > 95%)”]. In other words, combining spectral datasets may have reduced spectral count variability for some proteins, but the addition of other proteins to the dataset kept the overall false discovery rate within an expected amount of error.

We sought to determine if there was any counting bias associated with proteins falsely discovered as having significantly different spectral counts. We examined the distribution of counts for all 1716 proteins common to 6+9+2, 1+3+7 and 4+5+8

combined replicate files (Figure 3). Only seven proteins had summed spectral counts greater than 500; 81% of proteins had summed spectral counts less than 51, with 10–25 counts being most prevalent [note that the spectral count linearity of apotransferrin could be determined with 69 or fewer counts (also see below)]. Next, we evaluated the distribution of spectral counts for the 84 proteins with significantly different spectral counts in the 6+9+2 file. Most of the proteins with variable counts had counts 51 or greater (Figure 3). This means that while the false discovery rate applies to the dataset as a whole, error is mostly centralized to proteins having 51 counts or more. Since most of the proteins evaluated have counts less than 51, it can be expected that these proteins exhibit less statistical deviation.

Statistical Analysis of True Variation

The ability to significantly evaluate the different amounts of apotransferrin between samples was also examined. Spectral count ratios as small as 1 to 3.5 were found to be significantly different ($q < 0.01$), but ratios of 1 to 1.7 were not (Table 2, columns “apotransferrin summed spectral count” and “apotransferrin count ratio”). Our smallest ratios are better than the smallest 1 to 5 ratios most accurately measured in prior studies [15, 17, 20]. When replicates were combined and examined in three-way comparisons, the discriminating power was more refined. Ratios of 1 to 3.5 to 2 were found to be significantly different, as were ratios 1 to 0.9 to 3.3, ratios 1 to 2.4 to 4.2, and ratios 1.4 to 1 to 2 [Table 2, columns “apotransferrin summed spectral count” and “apotransferrin count ratio”]. Hence, more subtle spectral count differences were resolvable in three-way comparisons, but the improved discrimination was also associated with slightly increased empirical false discovery rates.

It should be noted that while these spectral counts ratios were found to be significantly different, the observed ratios did not accurately represent the true molar ratios of apotransferrin spiked into the samples. For example, a 1 to 5 M difference was observed as a 1 to 3.5 spectral count ratio, while a molar difference of 1 to 5 to 2 was measured as a 1 to 4.3 to 2.6 spectral count ratio (Table 2, columns “apotransferrin count ratio” and “apotransferrin molar ratio”). In other words, the observed spectral count ratios were slightly suppressed compared with the actual molar ratios of apotransferrin. Likely, dynamic exclusion results in spectral count suppression, but an alternative cause based on the spectral counting scheme will be presented later. Notwithstanding, the G-test was sufficient for distinguishing as low as 3.3-fold spectral counts changes in two-way comparisons and 2-fold changes in three-way comparisons.

Table 2. Protein and spectral count statistics for comparisons between replicate and combined replicate files

Replicates	Common proteins	Total spectra >95% common proteins	Proteins with different spectral counts (G-test, sig. >95%)	Variance per common protein (sig. >95%)
Two-way comparisons				
1 to 2	1329	18,869/19,289	1	0.001
1 to 6	1374	19,323/21,370	8	0.006
2 to 7	1040	17,359/17,748	3	0.003
3 to 4	1167	16,976/18,371	2	0.002
4 to 5	1030	17,283/17,100	39	0.038
6 to 7	1193	20,003/18,800	34	0.028
4 to 1+2	1334	19,346/36,679	2	0.001
2 to 3+6	1544	20,898/39,440	10	0.006
7 to 1+6	1316	19,165/37,363	26	0.020
1+2 to 5+6	1662	39,506/41,528	81	0.049
1+2 to 3+6	1712	40,084/40,129	11	0.006
4+5 to 3+6	1637	38,635/39,828	15	0.009
5+6 to 3+7	1469	39,221/36,594	118	0.080
2+4 to 3+6	1600	39,836/39,303	31	0.019
2+3 to 4+7	1507	37,217/38,355	14	0.009
1+2+3 to 4+5+6	1905	59,398/61,807	41	0.022
1,2,3 to 4,5,6	774	14,806/15,050/13,814/14,967/ 14,879/16,008	—	—
1+2+5 to 3+6+7	1900	59,495/60,279	16	0.008
1,2,5 to 3,6,7	793	14,280/13,369/14,793/14,453/ 14,423/15,468	—	—
1+3+7 to 4+5+6	1877	58,510/61,394	27	0.014
1,3,7 to 4,5,6	729	14,937/15,240/15,130/13,990/ 16,332/15,428	—	—
Three-way comparisons				
1 to 2 to 3	1163	18,280/18,700/17,159	7	0.006
1 to 2 to 5	1077	17,413/17,702/17,673	39	0.036
1 to 6 to 7	1070	17,197/19,005/17,795	26	0.024
2 to 4 to 6	1069	17,921/17,824/19,006	62	0.058
4 to 6 to 7	988	17,147/18,300/17,539	74	0.075
1+2 to 5+6 to 3+7	1343	36,087/37,930/35,244	116	0.086
2+4 to 1+5 to 3+6	1378	37,163/36,145/36,805	61	0.044
1+2 to 3+6 to 5+6	1556	38,440/38,785/40,340	53	0.034
1+2 to 4+5 to 3+6	1468	37,979/37,149/38,143	25	0.017
3+9 to 2+8 to 4+7	1433	37,874/38,884/37,831	39	0.027
1+3+4+ to 7+8+5 to 6+9+2	1683	55,399/58,843/62,385	122	0.072
4+5+6 to 1+3+7 to 2+8+9	1733	59,440/56,550/62,481	59	0.034
6+9+2 to 1+3+7 to 4+5+8	1716	62,866/56,377/58,729	84	0.049

*Apotransferrin; nf = not found; NA = not applicable.

G-Test Rigor

As mentioned earlier, it has been demonstrated that the G-test is a useful statistical method for analyzing differential spectral counts between multiple samples [17]. To further confirm the rigors of the G-test, we tested its ability to balance disparate datasets of spectral counts that might arise as result of sample handling errors or sample loading mistakes. Theoretically, some balance can be restored statistically by normalization. To test this, we compared a single replicate to two merged replicates (Table 2, replicates 4 to 1+2, 2 to 3+6, and 7 to 1+6), such that the merged file contained nearly double the number of spectra. In this instance, if the G-test normalizes correctly, there should be no significant differences found. Three out of three tests confirmed that this was indeed the case.

The empirical false discovery rates of 0.1%, 0.6%, and 2.0% were within the predicted range and were similar to those observed in the one-to-one comparisons [Table 2, column “variance per common protein (sig. > 95%)”].

The G-test was also compared with the *t*-test in terms of statistical reliability for evaluating spectral counts. The *t*-test allows the comparison of two conditions, but requires that each condition have at least three replicates to estimate means and variances. Therefore, we were not able to evaluate replicates on a one-to-one or two-to-two basis like we did with the G-test. This in itself points to the advantage of the G-test since only single replicates were needed to statistically resolve apotransferrin spectral count differences in two- and three-way comparisons (Table 2).

Table 2. Continued

Proteins with different spectral counts (G-test, sig. >99%)	Variance per common protein (sig. >99%)	Proteins with different spectral counts (t-test, sig. >95%)	Apotransferrin summed spectral count	Apotransferrin count ratio	Apotransferrin molar ratio
1	0.001	—	17 versus 0 (nf)	NA	1:0.1
6	0.004	—	17 versus 60 (sig. >99%)	1:3.5	1:5
2	0.002	—	0 (nf) versus 36	NA	0.1:1
2	0.002	—	0 (nf) versus 14	NA	0.01:1
18	0.017	—	14 versus 0 (nf)	NA	1:0.1
8	0.007	—	60 versus 36 (not sig. >95%)	1.7:1	5:2
1	0.001	—	14 versus 18 (not sig. >95%)	1:1.3	1:1.1
7	0.005	—	0 (nf) to 60	NA	0.1:5
9	0.007	—	36 versus 77 (not sig. >95%)	1:1.9	2:6
42	0.025	—	18 versus 62 (sig. >99%)	1:3.4	1:1.5.1
6	0.004	—	18 versus 60 (sig. >99%)	1:3.3	1:1.5
9	0.005	—	16 versus 60 (sig. >99%)	1:3.8	1:1.5
63	0.043	—	62 versus 36 (not sig. >95%)	1.7:1	5:1.2
19	0.012	—	15 versus 60 (sig. >99%)	1:4	1:1.5
5	0.003	—	0 (nf) versus 50	NA	0.1:3
24	0.013	—	18 versus 76 (sig. >99%)	1:4.2	1:1.6.1
—	—	0	not in all 6	NA	NA
13	0.007	—	20 versus 96 (sig. >99%)	1:4.8	1:2.7
—	—	0	not in all 6	NA	NA
14	0.007	—	53 versus 76 (not sig. >95%)	1:1.4	1:2
—	—	1*	15,16,17 versus 60,61,62 (sig. >99%)	1:3.8	1:5 (mock)
4	0.003	—	17 versus 0 (nf) versus 0 (nf)	NA	1:0.1:0.01
24	0.022	—	17 versus 0 (nf) versus 0 (nf)	NA	1:0.1:0.1
10	0.009	—	17 versus 60 versus 36 (sig. >99%)	1:3.5:2.1	1:5:2
34	0.032	—	0 (nf) versus 14 versus 60	NA	0.1:1:5
35	0.035	—	14 versus 60 versus 36 (sig. >99%)	1:4.3:2.6	1:5:2
68	0.051	—	18 versus 62 versus 36 (sig. >99%)	1:3.4:2	1:1.5:1.2
39	0.028	—	15 versus 19 versus 60 (sig. >99%)	1:1.3:4	1:1.1:1.5
33	0.021	—	18 versus 60 versus 62 (sig. >99%)	1:3.3:3.4	1:1.5:5.1
16	0.011	—	18 versus 16 versus 60 (sig. >99%)	1:0.9:3.3	1:1.1:1.5
20	0.014	—	69 versus 38 versus 50 (not sig. >95%)	1.8:1:1.3	5:2.1:3
57	0.034	—	31 versus 75 versus 130 (sig. >99%)	1:2.4:4.2	2:4.1:10.1
30	0.017	—	76 versus 53 versus 107 (sig. >95%)	1.4:1:2	6.1:3.7:1
47	0.027	—	130 versus 53 versus 53 (sig. >99%)	2.5:1:1	10.1:3:3.1

Furthermore, when constrained to comparing a set of three replicates to another set of three replicates by the *t*-test, there were many proteins not found across all six replicates. These proteins could not be evaluated by the *t*-test because a mean and variance could not be calculated when a protein was missing in any one replicate. Given this constraint of analyzing proteins found across all six replicates, the number of proteins was half of what could be analyzed by the G-test [Table 2, column “common proteins” for replicates 1,2,3 to 4,5,6; 1,2,5 to 3,6,7 and 1,3,7 to 4,5,6]. Nevertheless, no variation in spectral counts was observed for soybean proteins as measured by the *t*-test [Table 2, column “proteins with different spectral counts (*t*-test, sig. > 95%)”].

To evaluate true count differences, we turned to apotransferrin. But, because apotransferrin was not

among proteins with probabilities > 95% in all replicates, we inserted mock spectral counts in the files missing them. Under this contrived circumstance, the *t*-test determined the apotransferrin spectral count ratios of 1 to 3.8 to be statistically different (Table 2, replicate 1,3,7 to 4,5,6, column “apotransferrin summed spectral count”). This discrimination power agrees with the G-test in similar circumstances.

We conclude that the G-test is statistically rigorous for determining differential spectral counts in two- and three-way comparisons. In addition, the G-test allowed us to combine spectra from different samples, enabling the evaluation of twice as many proteins as the *t*-test while maintaining reasonable control of the false discovery rate. Finding significant differences among a few samples has important consequences in terms of biolog-

Figure 3. Distribution of summed spectral counts for all proteins common between combined replicates 6+9+2, 1+3+7, and 4+5+8 and those that had significantly different summed spectral counts. Proteins with summed spectral counts were binned into the spectral count groups shown. The percentage of all proteins (>95% probability) common to the combined replicate datasets that fall into each bin are shown in black. Percentages of a subset of proteins with statistically different spectral counts between combined replicates are shown in gray. The figure shows that the distribution of proteins with statistically different spectral counts is not the same as that for the entire dataset. This suggests that most spectral count variation is observed among the most highly abundant proteins.

ical discovery, especially when working with limited, precious samples.

On Proteins Unique to a Replicate

While there appears to be little statistically significant summed spectral count variability between the same proteins found between replicates, there remains a portion of proteins left unstudied that are found in some replicates but not others. For example, replicates no. 3 and no. 4, respectively, had 1474 and 1539 detected with a probability of 95% or greater, but only 1167 proteins were common between them. This means that 307 proteins were unique to replicate no. 3 and 372 were unique to replicate no. 4. Because biological study would be improved if the amount of overlap could be greater, we sought for a reason to explain why some proteins are only found in one sample but not another. One possible reason would be the probability cut-offs imposed. But when we examined the set of proteins from replicate no. 3 identified at a probability level between 95% and 80%, only 97 of the unique proteins from replicate no. 4 were found. Thus, lower scoring can only account for about a third of the differences.

Next, we evaluated the summed spectral counts for the unique proteins to see if there was a difference in the magnitude of counts for unique proteins versus common proteins. More than 88% of the proteins from replicate no. 4 had summed spectral counts less than 10. Contrasted to Figure 3, where only 25% of common proteins had summed spectral counts less than 10, it is plausible that infrequent observation of peptides explains why some proteins are found in one replicate but not another. To take it further, we examined the

Table 3. Protein and spectral count statistics of altered spectral counting schemes for comparisons between replicate and combined replicate files

Replicates compared	Common proteins	Spectra ALLCOUNT (proteins > 95%)	Spectra TOPCOUNT (proteins > 95%)	Spectra IDENTCOUNT (proteins > 95%)	PANORAMICS proteins with different spectral counts (G-test, sig. > 95%)	ALLCOUNT proteins with different spectral counts (G-test, sig. > 95%)	TOPCOUNT proteins with different spectral counts (G-test, sig. > 95%)
1 to 2	1329	25,512/26,011	21,695/22,019	14,533/12,756	1	12	11
1 to 6	1374	26,189/29,039	22,273/24,704	15,411/13,885	7	39	11
2 to 7	1040	23,392/24,252	19,805/20,670	11,609/13,196	3	14	9
6 to 7	1193	27,044/25,888	22,983/21,976	15,411/13,885	34	73	46
4 to 6 to 7	988	23,614/24,671/ 23,955	20,013/20,956/ 20,454	12,836/14,135/ 13,047	74	125	85
2 to 4 to 6	1069	23,788/24,385/ 25,291	20,096/20,640/ 21,455	11,785/13,199/ 14,461	62	129	77

nf = not found.

numbers of shared and distinct peptides used to identify the unique proteins in replicate no. 4. There were 312 proteins identified by more distinct peptides than shared peptides; 263 had no shared peptides at all, and 215 were solely identified by single distinct peptides. Taken together, these numbers suggest that these unique proteins were predominantly identified by single peptides at low abundance. Since the replicates that we evaluated were theoretically equivalent, it would stand to reason that run-to-run variability inherent to MudPIT, such as subtle peptide separation differences of each hand-made column and instrument ion selectivity based on second-to-second fluctuations of ions in traps, may sufficiently explain why certain peptides at low abundance can be found in one replicate but not another. Clearly, as previously discussed, and delineated in [Tables 1](#) and [2](#), and illustrated in [Figure 2](#), some of this variation can be tempered by pooling spectra from replicate runs which effectively increases the number proteins that overlap between two sets of data.

On Methods of Counting

PANORAMICS was originally designed to provide a solution for determining which sets of peptide spectra most likely identify a given protein [21]. Its ability to report the counts of peptide spectra was added only after we realized the program may also enable this kind of quantitative analysis. Hence, the spectral counting scheme was predicated on the probability model for identifying proteins. As a result, this created a unique basis for counting. Unlike other approaches where only peptides exceeding a high match score or probability threshold are counted for each protein [36], PANORAMICS counts high and low scoring spectra because even low-scoring spectra have some small probability of correctly identifying a protein. In addition,

when Mascot matches more than one peptide sequence to a spectrum (up to 10 possible ranked sequences), PANORAMICS calculates probabilities for each match. Because of the probability model, these different peptide sequences for the same spectrum are assigned to separate protein groups with different probabilities of identification attributed to them. While this is logical in terms of probability, we implemented a rule to count only the top-ranking matches to prevent any spectrum matched to multiple peptides in the probability model from being counted more than once (note that this is different from the rule of dividing the counts of shared peptides to prevent their overcounting).

Clearly, just as we were able to establish a set of rules for our counting method, it is possible to design other counting schemes too. These alternative schemes may result in different counting results. For example, the Yates lab counts spectra if the top-ranked peptide match scores do not exceed a false-positive rate of 5% as determined by reverse database searching [37]. Interestingly, setting such a threshold may lead to an undercount. For example, spectra that fall just below the threshold are not counted even if the match is reasonable for identification. This could happen with PANORAMICS as well. Imagine if spectrum 1 matches peptide A with an Ions/Identity score difference of 5 and the next consecutive spectrum 2 also matches peptide A but with an Ions/Identity score difference of -12. In this case, spectrum 1 is counted by PANORAMICS but spectrum 2 is not. Possibly, spectrum 1 and spectrum 2 represent stronger and weaker fragmentation patterns for the same peptides and that the real count is 2. But because of the threshold, the count for this peptide is under-reported. This type of undercount can occur regardless if one uses PANORAMICS or the Yates' lab method.

Table 3. Continued

IDENTCOUNT proteins with different spectral counts (G-test, sig. > 95%)	Apotransferrin spectra and ratio PANORAMICS	Apotransferrin spectra and ratio ALLCOUNT	Apotransferrin spectra and ratio TOPCOUNT	Apotransferrin spectra and ratio IDENTCOUNT	Apotransferrin molar ratio
22	17 versus 0 (nf)	22 versus 0 (nf)	22 versus 0 (nf)	9 versus 0 (nf)	1:0.1
6	17 versus 60 (sig. > 99%) [1:3.5]	22 versus 74 (sig. > 99%) [1:3.4]	22 versus 69 (sig. > 99%) [1:3.1]	9 versus 53 (sig. > 99%) [1:5.9]	1:5
33	0 (nf) versus 36 60 versus 36	0 (nf) versus 45 74 versus 45	0 (nf) versus 42 69 versus 42	0 (nf) versus 19 53 versus 19	0.1 to 1
21	(not sig.) [1.7:1]	(not sig.) [1.6:1]	(not sig.) [1.6:1]	(sig. > 95%) [2.8:1]	2.5:1
46	14 versus 60 versus 36 (sig. > 99%) [1:4.3:2.6]	17 versus 74 versus 45 (sig. > 99%) [1:4.4:2.6]	16 versus 69 versus 42 (sig. > 99%) [1:4.3:2.6]	5 to 53 to 19 (sig. > 99%) [1:10.6:3.8]	1:5:2
80	0 (nf) versus 14 versus 60	0 (nf) versus 17 versus 74	0 (nf) versus 16 vs. 69	0 (nf) versus 5 versus 53	0.1:1:5

Since it is apparent that different rules or thresholds can affect spectral counting, we implemented three other counting schemes into PANORAMICS and tested whether alterations in counting affected variability of protein counts between replicates. Note that the changes made were to the counting rules, not to the ways probabilities were determined, meaning that each program detected the same proteins with the same probabilities but that the proteins had different summed spectral counts. The first scheme, ALLCOUNT, counted all spectra assigned to peptides regardless of Mascot scoring rank. This allowed the situation where second, third etc. ranking matches could be counted. In other words, this allowed multiple counts per spectrum (up to 10 per spectrum, a parameter of Mascot) and established the maximum count possible given the data. The second scheme, TOPCOUNT, counted spectra to all peptides with a top-ranking Mascot Ions score regardless the Ions/Identity score difference, meaning that it would count a top-ranked match with a very low Mascot score. The third scheme, IDENTCOUNT, was designed to mimic as closely as possible within the PANORAMICS framework a Yates lab counting scheme [37]. Counts were assigned to peptides if the Ions score was top-ranking and if the Ions/Identity score difference was not less than zero. In Mascot, if the Ions/Identity score difference is zero, the false-positive rate for the match is 5%.

Compared with PANORAMICS, ALLCOUNT counted ~9000 more spectra, TOPCOUNT ~4000 more spectra, and IDENTCOUNT ~5000 less spectra on select single replicate files (Table 1). Thus, it can be seen that the counting scheme for PANORAMICS results in a total count that is less than with TOPCOUNT (counting all of the top ranked peptide spectra in the Mascot .dat file) but more than with IDENTCOUNT (counting the top-ranked spectra with Mascot probabilities of 95% or greater than being correct). In other words, PANORAMICS counts many but not all lower scoring peptides that may also contribute some reasonable probability to protein identification.

The effect of the different counts on spectral count variability was examined between single replicate files using the G-test, as before. Compared with PANORAMICS counting, there were more proteins with significantly different spectral counts using the ALLCOUNT and TOPCOUNT schemes (Table 3). For IDENTCOUNT, there were instances where there were fewer proteins with count variability compared with PANORAMICS, but there were also instances where there were more. Generally, as more rigorous thresholds and counting rules were applied, ranging from counting regardless of match rank or match probability (ALLCOUNT, least stringent), to counting only top matches regardless of probability (TOPCOUNT), to counting only top matches with some probability of being correct (PANORAMICS), to counting only top matches with >95% probability of being correct (IDENTCOUNT, most stringent), the false discovery rates dropped.

However, as the counting thresholds were changed from weaker to more stringent, the accuracy of ratio determination for the true positive apotransferrin also changed (Table 3). Spectral counts for a 1:5 M ratio of apotransferrin were determined to be significantly different using all counting schemes, but the counts provided by ALLCOUNT and TOPCOUNT predicted ratios (1 to 3.4 and 1 to 3.1, respectively) slightly smaller than PANORAMICS (1 to 3.5) whereas the counts from IDENTCOUNT (1: 5.9) overestimated the true ratio. In three-way comparisons, spectral counts for 1:5:2 M ratios of apotransferrin were also significantly different using all counting schemes. But PANORAMICS, ALLCOUNT and TOPCOUNT counts more closely reflected the true molar ratio while IDENTCOUNT overestimated the counts with ratios of 1 to 10.6 to 3.8. The only advantage to the stringent thresholds of IDENTCOUNT was the ability to statistically resolve 2.8-fold changes, but the sacrifice was over-estimation for the larger ratios. Therefore, based on these tests, the original configuration for spectral counting in PANORAMICS appears to be a sufficient compromise for obtaining reasonably accurate count ratios between samples without generating statistically unacceptable amounts of false differences when low-scoring spectra are included in the count.

Conclusions

To be beneficial to biology, relative protein quantitation by label-free tandem mass spectral counting must not only be able to resolve true protein accumulation changes between two or more samples, but it must have a low false discovery rate. Here, we have confirmed both by evaluating replicate samples of soybean peptides spiked with different molar amounts of bovine apotransferrin. The numbers of tandem mass spectra that were collected for apotransferrin increased linearly with amounts of apotransferrin and the empirical false discovery rate was nearly the same as the false discovery rate predicted by the Benjamini and Hochberg correction, a conservative method for calculating q -values. Since more accurate statistical methods are available for calculating q -values [38], it is possible to further limit false discovery rates and improve upon MudPIT for identifying subtle changes in the accumulation of proteins between samples.

We have also examined four methods for counting spectra and discovered that changing the count scheme can influence the false-discovery rate for background soybean proteins. However, we observed that including some poorer-scoring spectra improves the accuracy of measuring molar differences of proteins between samples. This is likely because including true but poorer-scoring spectra shifts the observed count toward the theoretical absolute count. Consequently, the results also suggest that thresholds for counting can be empirically set to improve relative quantitation using the label-free method of spectral counting.

Acknowledgments

The authors are grateful to Kimberly Campbell at USDA-ARS, Beltsville, MD for the cultivation of soybean plants.

References

- Mueller, L. N.; Brusniak, M. Y.; Mani, D. R.; Aebersold, R. An Assessment of Software Solutions for the Analysis of Mass Spectrometry Based Quantitative Proteomics data. *J. Proteome Res.* **2008**, *7*, 51–61.
- Mann, M. Comparative Analysis to Guide Quality Improvements in Proteomics. *Nat. Methods* **2009**, *6*, 717–719.
- Petyuk, V. A.; Qian, W. J.; Chin, M. H.; Wang, H.; Livesay, E. A.; Monroe, M. E.; Adkins, J. N.; Jaitly, N.; Anderson, D. J.; Camp, D. G. II; Smith, D. J.; Smith, R. D. Spatial Mapping of Protein Abundances in the Mouse Brain by Voxelation Integrated with High-Throughput Liquid Chromatography-Mass Spectrometry. *Genome Res.* **2007**, *17*, 328–336.
- Lee, J.; Feng, J.; Campbell, K. B.; Scheffler, B. E.; Garrett, W. M.; Thibivilliers, S.; Stacey, G.; Naiman, D. Q.; Tucker, M. L.; Pastor-Corrales, M. A.; Cooper, B. Quantitative Proteomic Analysis of Bean Plants Infected by a Virulent and Avirulent Obligate Rust Fungus. *Mol. Cell. Proteom.* **2009**, *8*, 19–31.
- Malmstrom, J.; Beck, M.; Schmidt, A.; Lange, V.; Deutsch, E. W.; Aebersold, R. Proteome-Wide Cellular Protein Concentrations of the Human Pathogen *Leptospira interrogans*. *Nature* **2009**, *460*, 762–765.
- Zybailov, B.; Friso, G.; Kim, J.; Rudella, A.; Rodriguez, V. R.; Asakura, Y.; Sun, Q.; van Wijk, K. J. Large Scale Comparative Proteomics of a Chloroplast Clp Protease Mutant Reveals Folding Stress, Altered Protein Homeostasis, and Feedback Regulation of Metabolism. *Mol. Cell. Proteom.* **2009**, *8*, 1789–1810.
- Zybailov, B.; Coleman, M. K.; Florens, L.; Washburn, M. P. Correlation of Relative Abundance Ratios Derived from Peptide Ion Chromatograms and Spectrum Counting for Quantitative Proteomic Analysis Using Stable Isotope Labeling. *Anal. Chem.* **2005**, *77*, 6218–6224.
- Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative Analysis of Complex Protein Mixtures Using Isotope-Coded Affinity Tags. *Nat. Biotechnol.* **1999**, *17*, 994–999.
- 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.* **2002**, *1*, 376–386.
- Choe, L.; D'Ascenzo, M.; Relkin, N. R.; Pappin, D.; Ross, P.; Williamson, B.; Guertin, S.; Pribil, P.; Lee, K. H. Eight-plex Quantitation of Changes in Cerebrospinal Fluid Protein Expression in Subjects Undergoing Intravenous Immunoglobulin Treatment for Alzheimer's Disease. *Proteomics* **2007**, *7*, 3651–3660.
- Paoletti, A. C.; Parmely, T. J.; Tomomori-Sato, C.; Sato, S.; Zhu, D.; Conaway, R. C.; Conaway, J. W.; Florens, L.; Washburn, M. P. Quantitative Proteomic Analysis of Distinct Mammalian Mediator Complexes Using Normalized Spectral Abundance Factors. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18928–18933.
- Pan, C.; Kora, G.; Tabb, D. L.; Pelletier, D. A.; McDonald, W. H.; Hurst, G. B.; Hettich, R. L.; Samatova, N. F. Robust Estimation of Peptide Abundance Ratios and Rigorous Scoring of Their Variability and Bias in Quantitative Shotgun Proteomics. *Anal. Chem.* **2006**, *78*, 7110–7120.
- Dong, M. Q.; Venable, J. D.; Au, N.; Xu, T.; Park, S. K.; Cociorva, D.; Johnson, J. R.; Dillin, A.; Yates, J. R. III. Quantitative Mass Spectrometry Identifies Insulin Signaling Targets in *C. elegans*. *Science* **2007**, *317*, 660–663.
- Chelius, D.; Zhang, T.; Wang, G.; Shen, R. F. Global Protein Identification and Quantification Technology Using Two-Dimensional Liquid Chromatography Nanospray Mass Spectrometry. *Anal. Chem.* **2003**, *75*, 6658–6665.
- Liu, H.; Sadygov, R. G.; Yates, J. R. III. A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. *Anal. Chem.* **2004**, *76*, 4193–4201.
- Wang, G.; Wu, W. W.; Zeng, W.; Chou, C. L.; Shen, R. F. Label-Free Protein Quantification Using LC-Coupled Ion Trap or FT Mass Spectrometry: Reproducibility, Linearity, and Application with Complex Proteomes. *J. Proteome Res.* **2006**, *5*, 1214–1223.
- 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.* **2006**, *5*, 2909–2918.
- Lu, P.; Vogel, C.; Wang, R.; Yao, X.; Marcotte, E. M. Absolute Protein Expression Profiling Estimates the Relative Contributions of Transcriptional and Translational Regulation. *Nat. Biotechnol.* **2007**, *25*, 117–124.
- Xia, Q.; Wang, T.; Park, Y.; Lamont, R. J.; Hackett, M. Differential Quantitative Proteomics of *Porphyromonas gingivalis* by Linear Ion Trap Mass Spectrometry: Non-label Methods Comparison, *q*-Values, and LOWESS Curve Fitting. *Int. J. Mass Spectrom.* **2007**, *259*, 105–116.
- Old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G. Comparison of Label-Free Methods for Quantifying Human Proteins by Shotgun Proteomics. *Mol. Cell. Proteom.* **2005**, *4*, 1487–1502.
- Feng, J.; Naiman, D. Q.; Cooper, B. Probability Model for Assessing Proteins Assembled from Peptide Sequences Inferred from Tandem Mass Spectrometry data. *Anal. Chem.* **2007**, *79*, 3901–3911.
- Cooper, B.; Neelam, A.; Campbell, K. B.; Lee, J.; Liu, G.; Garrett, W. M.; Scheffler, B.; Tucker, M. L. Protein Accumulation in the Germinating *Uromyces appendiculatus* Uredospore. *Mol. Plant Microbe Interact.* **2007**, *20*, 857–866.
- Schmutz, J.; Cannon, S. B.; Schlueter, J.; Ma, J.; Mitros, T.; Nelson, W.; Hyten, D. L.; Song, Q.; Thelen, J. J.; Cheng, J.; Xu, D.; Hellsten, U.; May, G. D.; Yu, Y.; Sakurai, T.; Umezawa, T.; Bhattacharyya, M. K.; Sandhu, D.; Valliyodan, B.; Lindquist, E.; Peto, M.; Grant, D.; Shu, S.; Goodstein, D.; Barry, K.; Futrell-Griggs, M.; Abernathy, B.; Du, J.; Tian, Z.; Zhu, L.; Gill, N.; Joshi, T.; Libault, M.; Sethuraman, A.; Zhang, X. C.; Shinozaki, K.; Nguyen, H. T.; Wing, R. A.; Cregan, P.; Specht, J.; Grimwood, J.; Rokhsar, D.; Stacey, G.; Shoemaker, R. C.; Jackson, S. A. Genome Sequence of the Palaeopolyploid Soybean. *Nature* **2010**, *463*, 178–183.
- Florens, L.; Washburn, M. P. Proteomic Analysis by Multidimensional Protein Identification Technology. *Methods Mol. Biol.* **2006**, *328*, 159–175.
- Olsen, J. V.; de Godoy, L. M.; Li, G.; 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.* **2005**, *4*, 2010–2021.
- Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *Electrophoresis* **1999**, *20*, 3551–3567.
- Silva, J. C.; Gorenstein, M. V.; Li, G. Z.; Vissers, J. P.; Geromanos, S. J. Absolute Quantification of Proteins by LCMSE: A Virtue of Parallel MS Acquisition. *Mol. Cell. Proteom.* **2006**, *5*, 144–156.
- Mallick, P.; Schirle, M.; Chen, S. S.; Flory, M. R.; Lee, H.; Martin, D.; Ranish, J.; Raught, B.; Schmitt, R.; Werner, T.; Kuster, B.; Aebersold, R. Computational Prediction of Proteotypic Peptides for Quantitative Proteomics. *Nat. Biotechnol.* **2007**, *25*, 125–131.
- Nesvizhskii, A. I.; Aebersold, R. Interpretation of Shotgun Proteomic Data: The Protein Inference Problem. *Mol. Cell. Proteom.* **2005**, *4*, 1419–1440.
- Dost, B.; Bandeira, N.; Li, X.; Shen, Z.; Briggs, S.; Bafna, V. Shared Peptides in Mass Spectrometry Based Protein Quantification. in *RECOMB 2009, LNCS 5541*, Batzoglou, S. Ed.; Springer-Verlag: Berlin Heidelberg, 2009; p. 356–371.
- Sardiu, M. E.; Gilmore, J. M.; Carrozza, M. J.; Li, B.; Workman, J. L.; Florens, L.; Washburn, M. P. Determining Protein Complex Connectivity Using a Probabilistic Deletion Network Derived from Quantitative Proteomics. *PLoS One* **2009**, *4*, e7310.
- Mosley, A. L.; Florens, L.; Wen, Z.; Washburn, M. P. A Label-Free Quantitative Proteomic Analysis of the *Saccharomyces cerevisiae* nucleus. *J. Proteom.* **2009**, *72*, 110–120.
- Zhang, Y.; Wen, Z.; Washburn, M. P.; Florens, L. Refinements to Label-Free Proteome Quantitation: How to Deal with Peptides Shared by Multiple Proteins. *Anal. Chem.* **2010**, *82*, 2272–2281.
- Tabb, D. L.; Vega-Montoto, L.; Rudnick, P. A.; Variyath, A. M.; Ham, A. J.; Bunk, D. M.; Kilpatrick, L. E.; Billheimer, D. D.; Blackman, R. K.; Cardasis, H. L.; Carr, S. A.; Clauser, K. R.; Jaffe, J. D.; Kowalski, K. A.; Neubert, T. A.; Regnier, F. E.; Schilling, B.; Tegeler, T. J.; Wang, M.; Wang, P.; Whiteaker, J. R.; Zimmerman, L. J.; Fisher, S. J.; Gibson, B. W.; Kinsinger, C. R.; Mesri, M.; Rodriguez, H.; Stein, S. E.; Tempst, P.; Paulovich, A. G.; Liebler, D. C.; Spiegelman, C. Repeatability and Reproducibility in Proteomic Identifications by Liquid Chromatography-Tandem Mass Spectrometry. *J. Proteome Res.* **2010**, *9*, 761–776.
- Zhang, Y.; Wen, Z.; Washburn, M. P.; Florens, L. Effect of Dynamic Exclusion Duration on Spectral Count Based Quantitative Proteomics. *Anal. Chem.* **2009**, *81*, 6317–6326.
- Zybailov, B.; Mosley, A. L.; Sardiu, M. E.; Coleman, M. K.; Florens, L.; Washburn, M. P. Statistical Analysis of Membrane Proteome Expression Changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* **2006**, *5*, 2339–2347.
- Usaita, R.; Wohlschlegel, J.; Venable, J. D.; Park, S. K.; Nielsen, J.; Olsson, L.; Yates III, J. R. Characterization of Global Yeast Quantitative Proteome Data Generated from the Wild-Type and Glucose Repression *Saccharomyces cerevisiae* Strains: The Comparison of Two Quantitative Methods. *J. Proteome Res.* **2008**, *7*, 266–275.
- Storey, J. D.; Tibshirani, R. Statistical Significance for Genomewide Studies. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9440–9445.