Published in final edited form as:

J Proteome Res. 2010 November 5; 9(11): 5698–5704. doi:10.1021/pr100508p.

Improved LC-MS/MS Spectral Counting Statistics by Recovering Low Scoring Spectra Matched to Confidently Identified Peptide Sequences

Jian-Ying Zhou[‡], Athena A. Schepmoes[‡], Xu Zhang[‡], Ronald J. Moore[‡], Matthew E. Monroe[‡], Jung Hwa Lee^{‡,§}, David G. Camp II[‡], Richard D. Smith[‡], and Wei-Jun Qian^{‡,*}

[‡]Biological Science Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99352

[§]Department of Chemistry and Center for Electro- and Photo-Responsive Molecules, Korea University, Seoul 136-701, Korea

Abstract

Spectral counting has become a popular method for LC-MS/MS based proteome quantification; however, this methodology is often not reliable when proteins are identified by a small number of spectra. Here we present a simple strategy to improve spectral counting based quantification for low abundance proteins by recovering low quality or low scoring spectra for confidently identified peptides. In this approach, stringent data filtering criteria were initially applied to achieve confident peptide identifications with low false discovery rate (e.g., < 1% at peptide level) after LC-MS/MS analysis and database search by SEQUEST. Then, all low scoring MS/MS spectra that match to this set of confidently identified peptides were recovered, leading to more than 20% increase of total identified spectra. The validity of these recovered spectra was assessed by the parent ion mass measurement error distribution, retention time distribution, and by comparing the individual low score and high score spectra that correspond to the same peptides. The results support that the recovered low scoring spectra have similar confidence levels in peptide identifications as the spectra passing the initial stringent filter. The application of this strategy of recovering low scoring spectra significantly improved the spectral count quantification statistics for low abundance proteins, as illustrated in the identification of mouse brain region specific proteins.

Keywords

Spectral count; LC-MS/MS; false negative; quantification

Introduction

Spectral counting by analyzing the total number of tandem mass spectra assigned to each protein is a common strategy for determining relative protein abundances in liquid chromatography-tandem mass spectrometry (LC-MS/MS)-based proteomics^{1–4}. Previous studies have demonstrated that the spectral counts of proteins correlate linearly with protein

*To whom correspondence should be addressed: Dr. Wei-Jun Qian, Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, P. O. Box 999, MSIN: K8-98, Richland, WA 99352, Weijun.Qian@pnl.gov.

SUPPORTING INFORMATION AVAILABLE

A complete list of protein spectral count information that shows specific brain region distributions is available in a Microsoft Excel worksheet that may be downloaded free of charge at <http://pubs.acs.org>.

abundances in complex samples 10^2 – 10^5 over a linear dynamic range of >2 orders of magnitude⁵. Although this quantitative strategy is reliable for measuring large changes in relatively abundant proteins identified with many peptides or spectra, it is limited by the inability to confidently quantify differences in low abundance proteins because of the small number of spectra typically acquired for such proteins^{5–6}.

A number of different strategies have been utilized to improve the reliability of spectral count quantification^{7–13}. For example, multiple technical replicates¹² or extensive multidimensional peptide fractionations prior to LC-MS/MS¹⁴ are often applied to increase the number of spectral count measurements. In another example, Zybaylov et al.¹³ used a normalized spectral abundance factor calculated from the total spectral count number and the length of the protein to quantify the differential membrane proteome expression in *Saccharomyces cerevisiae* under different conditions. Choi, et al.⁹ developed a novel statistical framework (QSpec) for significance analysis of spectral counting with extensions to a variety of experimental design factors and adjustments for protein properties. Elsewhere, Griffin, et al.¹⁰ developed a normalized label-free quantitative method termed the normalized spectral index that combined peptide count, spectral count, and fragment-ion MS/MS intensity for more reliable quantification.

Despite these developments, spectral counting-based quantification is inherently limited by low spectral counts for low abundance proteins that are often of greatest interest in biological and clinical studies. When determining relative abundances of these low abundance proteins, it is especially important to include all MS/MS spectra that identify proteins. However, many low scoring spectra that may correctly identify peptides are often filtered out (i.e., false negative identifications) in LC-MS/MS experiments due to the application of stringent filtering criteria, in particular after database searching using a target-decoy search strategy¹⁵ to control the false discovery rate (FDR) of peptide identifications. If these false negative spectra are recovered, thereby increasing the total number of spectra, then spectral counts for low abundance proteins should also increase.

In this work, we report a strategy that improves spectral counting statistics by recovering low quality spectra that identify those confidently identified peptides, but with scores lower than the filtering cutoffs. With this strategy, database searching results initially are filtered using stringent criteria to derive a list of confidently identified peptides. Next, all low scoring spectra with sequences matching to the confidently identified peptides are recovered and combined with initial filtered spectra, thereby increasing the number of spectra (n) that can be utilized for quantification. Results from LC-MS/MS analyses of mouse brain tissue samples demonstrate that the recovered low scoring spectra have similar mass accuracy and retention time distributions as those that passed the initial stringent filtering criteria, supporting that the confidence levels of these two sets of spectra are comparable in terms of peptide identifications. By affording a 20% increase in the number of spectra, application of this strategy significantly reduces the standard deviation (proportional to $n^{-1/2}$) for replicate analyses and improves the spectral count quantification.

Experimental Methods

Protein Extraction and Digestion

Mouse brain tissues served as model samples for this study. Cortex, cerebellum, striatum, and remainder of the brain (ROB) tissues dissected from four adult C57BL/6J male mice (9 weeks, 21–27 g) were obtained from Jackson Laboratories (Bar Harbor, ME). All tissues were homogenized in 50 mM NH_4HCO_3 (pH 7.8), and protein concentration was determined by BCA assay (Pierce, Rockford, IL). Aliquots of homogenates from the same brain region from four mice were pooled for protein digestion. Approximately 300 μg

pooled sample was subsequently treated with 50% (v/v) trifluoroethanol (Sigma-Aldrich) for 2 h at 60 °C, 5 mM tributylphosphine (Sigma) for 0.5 h at 60 °C, 40 mM iodoacetamide for 1 h at 37 °C, and then diluted five-fold with 50 mM NH_4HCO_3 . Samples were digested using sequencing-grade trypsin (Promega, Madison, WI) for 3 h at 37 °C at a 1:50 trypsin to protein ratio (w:w), with 1 mM CaCl_2 added during the digestion. Once digested, samples were cleaned using a SPE C-18 column (Supelco, Bellefonte, PA) and then dried using a speed-vac concentrator. All samples were stored at -80°C until time for further analysis.

Strong Cation Exchange (SCX) Fractionation

SCX fractionation of digested peptides was performed using an 1100 series HPLC system (Agilent Technologies, Wilmington, DE) at a flow rate of 200 $\mu\text{L}/\text{min}$. 150 μg tryptic peptides from each sample were resuspended in buffer A (25% acetonitrile, 10 mM ammonium formate, pH 3.0) and loaded onto a 2.1×200 mm (5 μm , 300 Å) Polysulfoethyl A LC column (PolyLC, Columbia, MD) preceded by a 2.1×10 mm guard column. After loading peptides onto the column, the mobile phase consisted of 100% A for 10 min, a 40-min linear gradient from 0 to 50% B (25% acetonitrile, 500 mM ammonium formate, pH 6.8), a 10-min linear gradient from 50 to 100% B, and then 100% B for 20 min. Using an automated fraction collector, 25 fractions were collected for each sample. Each fraction was lyophilized prior to LC-MS analysis.

Reversed-phase Capillary LC-MS/MS Analysis

Peptides were analyzed using a custom-built automated four-column high-pressure capillary LC system coupled on-line to either a linear ion trap (LTQ; Thermo Scientific, San Jose, CA) or an LTQ-Orbitrap mass spectrometer (Thermo Scientific) via a nanoelectrospray ionization interface manufactured in-house. The reversed-phase capillary column was prepared by slurry-packing 3- μm Jupiter C18 bonded particles (Phenomenex, Torrance, CA) into a 65-cm-long, 75- μm -inner diameter fused silica capillary (Polymicro Technologies, Phoenix, AZ). After loading 2.5 μg of peptides onto the column, the mobile phase was held at 100% A (0.1% formic acid) for 20 min, followed by a linear gradient from 0 to 70% buffer B (0.1% formic acid in 90% acetonitrile) over 85 min. Each full MS scan (m/z 400–2000) was followed by collision-induced MS/MS spectra (normalized collision energy setting of 35%) for the 10 most abundant ions. The dynamic exclusion duration was set to 1 min, the heated capillary was maintained at 200 °C, and the ESI voltage was held at 2.2 kV.

Data Analysis

LC-MS/MS raw data were converted into .dta files using Extract_MS_n (version 3.0) in Bioworks Cluster 3.2 (Thermo Fisher Scientific, Cambridge, MA), and the SEQUEST algorithm (version 27, revision 12) was used to independently search all MS/MS spectra against the mouse International Protein Index (IPI) database that had 51,489 total protein entries (version 3.35, released October 24, 2007). The FDR was estimated using a decoy-database searching methodology¹⁵. Search parameters and filtering criteria (Table 1) were applied to limit the FDR at the unique peptide level to <1%¹⁶. Identified proteins were grouped to a non-redundant protein list using ProteinProphetTM software, after which one protein IPI number was randomly selected to represent each corresponding protein group that consisted of a number of database entries. Only those proteins or protein groups with two or more unique peptide identifications were considered as confident identifications.

Results and Discussion

Strategy for Recovering Low Scoring Spectra

In this study, a strategy for recovering low scoring spectra was applied to increase the number of spectral counts used for relative quantification (Figure 1). We hypothesized that low scoring spectra that matched sequences corresponding to confidently identified peptides should have a similar probability of identifying correct peptides. To test this hypothesis, we first applied a set of stringent data filtering criteria (Table 1) to achieve confident (FDR <1%) peptide identifications, after which we recovered all low scoring spectra that matched to the same set of confidently identified peptides as opposed to accept other peptides mapped to proteins that were identified based on the set of confident peptides. The recovered spectra were combined with the initial set of spectra prior to spectral counting-based quantification.

To evaluate the reliability of this strategy for improving spectral counting quantification of low abundance species, we used tissues from cerebellum, cortex, striatum, and ROB regions of the mouse brain as testing samples. Proteins from each brain region were digested by trypsin, and the resulting peptides were separated into 25 fractions by off-line SCX chromatography prior to LC-MS/MS analysis on an LTQ instrument. A portion of these fractions (striatum and ROB) were also analyzed using an LTQ-Orbitrap instrument to evaluate parent ion mass measurement error distributions for both low scoring and high scoring spectra.

In total, 43,093 unique peptides were identified with a FDR of 1.1% from the 100 LTQ LC-MS/MS datasets acquired for the four brain regions, from which 4668 proteins were identified by ≥ 2 unique peptides. 159,010 spectra passed the initial stringent filtering, and an additional 35,186 spectra (22% of the total spectra passed the filter) that matched the same set of identified sequences were recovered. As shown in Figure 2, spectra were recovered similarly for both low and high abundances proteins.

Similarly, the 50 Orbitrap LC-MS/MS datasets yielded 22,009 unique peptides identified with a FDR of 0.6%, using the same criteria as applied to the LTQ datasets, from which 2881 proteins were covered by ≥ 2 unique peptides. 68,270 Orbitrap spectra passed the initial filtering criteria, and an additional 14,291 low scoring spectra (~21% of the total spectra passed the stringent filter) were recovered following this simple strategy.

Evaluation of the Quality of Recovered Spectra

To evaluate the quality of recovered spectra matched to confidently identified sequences, we initially examined the parent ion mass measurement error (a major parameter used for peak matching in label-free quantification¹⁷) of all Orbitrap spectra. Figure 3 shows that ~85% of the recovered spectra had mass errors <6 ppm, which is comparable to ~93% of the stringently filtered spectra, but significantly different from 5.3% of spectra that matched to sequences in the reversed database. The slightly higher percentage of recovered spectra observed outside the 6 ppm mass error window may be associated with the difficulty of obtaining accurate mass measurements for low signal peaks, i.e., the MS signals may have poor signal-to-noise ratios for these spectra.

To further evaluate the quality of recovered spectra, we compared LC retention times (in terms of scan number) of recovered spectra with retention times of original filtered spectra. Figure 4 shows good correlation between the average scan number of recovered and filtered spectra from individual LC-MS/MS datasets. The linear coefficient of determination (R^2) is > 0.98 in both cases, which suggests there is no significant difference in retention times for the vast majority of peptides. We estimate that ~10% of the recovered spectra are possible

outliers based on the fact that the retention times for these spectra were off the average retention time of the filtered spectra by more than 10% of the entire separation time (~10 min). However, mass error examination of these possible outliers revealed that ~80% still exhibited < 6 ppm mass error, suggesting that these spectra still have similar confidence in terms of identifications compared to those filtered good quality spectra. The reason that many peptides are identified across a relatively wide retention time window is due to chromatography peak tailing for abundant species.

We also examined filtered and recovered spectra obtained for specific peptides. The comparison of the LTQ and LTQ-Orbitrap spectra with theoretical fragmentation allowed us to evaluate the quality of individual recovered spectra. For most of these peptides, the recovered spectra presented an MS/MS pattern that matched to the filtered spectra, but generally with lower signal intensity, as well as lower XCorr and ΔC_n values. This trend is exemplified in Figure 5A that compares the worst recovered spectrum (spectrum b) with the best filtered spectrum (spectrum a) for peptide AGFAGDDAPR within the LTQ dataset. All of the major fragment peaks in the filtered spectra (Figure 5B) are observed in the recovered spectra (Figure 5C) and with a similar distribution pattern; however, the signal intensity of recovered spectrum (6.7E2) is nearly three orders of magnitude lower than that of filtered spectrum (8.2E5).

The observations that the recovered spectra have similar mass error and retention time distributions, as well as well-matched major MS/MS fragment patterns as the filtered spectra suggest that the recovered low scoring spectra have FDRs comparable to the filtered spectra.

Improvement in Spectral Count Quantification

To illustrate the effect of the recovered spectra on spectral counting-based relative quantification, we utilized data obtained from technical replicates during sample analysis. The $\log_2$ ratio of spectral counts between replicates was calculated as:

$$\text{Log}_2 \text{ ratio} = \text{Log}_2[(\text{spectral count in replicate A} + 0.5)/(\text{spectral count in replicate B} + 0.5)]$$

We add 0.5 to all spectral counts to provide valid $\log_2$ ratio for cases with zero spectral counts in one of the replicates. If a protein had an absolute $\log_2$ ratio >1.6, then we considered that this protein was observed with false changes in abundance because of data variations.

In triplicate LC-MS/MS analyses on the LTQ instrument, an average of 1135 proteins were identified for each replicate following application of stringent filtering criteria. Low scoring spectra were recovered for 465 (41.0%) of these proteins, of which 376 were identified initially by at least two high scoring spectra. The 376 proteins were further separated into two groups based on total spectral counts after data recovery, i.e. ≤ 10 counts (231 proteins) and > 10 counts (145 proteins). In Figure 6, the percentages of observed false changes in abundance based on the use of high scoring spectra only without the recovered spectra are ~10%, ~15%, and ~0.2% for the group of 376 proteins, proteins in the ≤ 10 counts group, and proteins in the > 10 counts group, respectively. Note that when the recovered spectra are combined with the high-scoring spectra, the percentages of observed false changes in abundance are reduced to ~2%, ~4%, and ~0.2%, respectively, for the three groups. An even more significant decrease in the percentage of false changes in abundance was observed when proteins identified by single high-quality spectra were included (data not shown). These results suggest that the use of recovered spectra improves the reliability of spectral count quantification for low abundance proteins, but has little effect on high abundance proteins. This observation is anticipated since the variations or standard deviations for spectral counting based quantification in replicated analyses should be proportional to $n^{-1/2}$;

thus, the increased number of total spectra has a significant impact on the reliability and confidence of quantification for low abundance or low spectral count proteins.

To further illustrate the effect of recovered spectra on spectral count quantification, we compared protein abundance differences in different regions of the mouse brain. Of the 4225 proteins with at least two unique peptides identified in cerebellum, cortex, and striatum by LTQ, 2684 proteins exhibited increased spectral counts after recovering low scoring spectra that matched to confidently identified peptides. Among these proteins, 496, 168, and 776 proteins showed region specific distribution (spectral counts in one region contributes to >50% of total) in cerebellum, cortex, and striatum, respectively. The detailed spectral count information for all these proteins in different brain regions is provided in Supplemental Table 1. Figure 7 shows three example proteins with improved confidence in quantification. While the addition of recovered spectra does not significantly affect the spectral count distribution among the different brain regions for most proteins, the increased numbers of spectral counts make the relative quantification more reliable than the original data when proteins were only identified by a few spectral counts per region.

We further compared the spectral counting data with mRNA abundance data available from the Allen Brain Atlas (<http://www.brain-map.org/>) for 30 proteins showing significant region specific distribution. Among these proteins, 27 showed similar distribution pattern between the two datasets. No mRNA data were available for two of these proteins and one other protein did not show good correlation between the two datasets. Figure 8 presents the comparison between the spectral counting and mRNA data for 9 region specific proteins. Purkinje cell protein-2 (PCP2) is known to be specifically associated with Purkinje cell in the cerebellum¹⁸, Phosphodiesterase 10a (PDE10a) is highly expressed in the striatum, consistent with its known localization to striatal medium-sized spiny projection neurons and the response of these neurons to cortical stimulation synaptic structures^{19–20}. All these data support that the spectral counts after this recovering strategy provide reliable quantification.

Conclusion

In LC-MS/MS-based proteome profiling, false negative peptide identifications generated by spectral filtering after database searching is a commonly encountered issue. In this work, we presented a simple strategy for recovering low scoring spectra, but matched with confidently identified peptides, to improve relative quantification using spectral counts. Following application of stringent data filtering criteria to identify peptides with high confidence, discarded spectra that matched to the confidently identified peptides were recovered for further relative quantitative analysis. Similarities in mass error distribution, retention time distribution, and MS/MS fragmentation patterns between the recovered spectra and spectra that passed stringent filtering criteria support that the confidence levels of these two sets of spectra for identifying peptides are comparable. The addition of recovered spectra to the initial set of filtered spectra improves spectral counting statistics and the confidence and accuracy of spectral count quantification for low abundance proteins.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Abbreviations

SCX	strong cation exchange
1D	one-dimensional

2D	two-dimensional
LTQ	linear ion trap quadrupole
IPI	International Protein Index
Xcorr	cross-correlation score
ΔC_n	delta correlation
FDR	false discovery rate

Acknowledgments

Portions of this research were supported by the NIH National Institute of Diabetes and Digestive and Kidney Diseases (R01 DK074795) and NIH National Center for Research Resources (RR18522). Experimental work was performed in the Environmental Molecular Sciences Laboratory, a U. S. Department of Energy (DOE) Office of Biological and Environmental Research national scientific user facility on the Pacific Northwest National Laboratory (PNNL) campus in Richland, Washington. PNNL is multi-program national laboratory operated by Battelle for the DOE under Contract No. DE-AC05-76RLO 1830.

References

1. Liu H, Sadygov RG, Yates JR 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem.* 2004; 76(14):4193–4201. [PubMed: 15253663]
2. Qian WJ, Jacobs JM, Camp DG II, Monroe ME, Moore RJ, Gritsenko MA, Calvano SE, Lowry SF, Xiao W, Moldawer LL, Davis RW, Tompkins RG, Smith RD. Comparative proteome analyses of human plasma following in vivo lipopolysaccharide administration using multidimensional separations coupled with tandem mass spectrometry. *Proteomics.* 2005; 5(2):572–584. [PubMed: 15627965]
3. Paoletti AC, Parmely TJ, Tomomori-Sato C, Sato S, Zhu D, Conaway RC, Conaway JW, Florens L, Washburn MP. Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proc Natl Acad Sci U S A.* 2006; 103(50):18928–18933. [PubMed: 17138671]
4. Zybaylov B, Coleman MK, Florens L, Washburn MP. 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(19):6218–6224. [PubMed: 16194081]
5. Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, Resing KA, Ahn NG. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics.* 2005; 4(10):1487–1502. [PubMed: 15979981]
6. Usaitė R, Wohlschlegel J, Venable JD, Park SK, Nielsen J, Olsson L, Yates JI JR. 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(1):266–275. [PubMed: 18173223]
7. Sun A, Zhang J, Wang C, Yang D, Wei H, Zhu Y, Jiang Y, He F. Modified Spectral Count Index (mSCI) for Estimation of Protein Abundance by Protein Relative Identification Possibility (RIPpro): A New Proteomic Technological Parameter. *J Proteome Res.* 2009
8. Fu X, Gharib SA, Green PS, Aitken ML, Frazer DA, Park DR, Vaisar T, Heinecke JW. Spectral index for assessment of differential protein expression in shotgun proteomics. *J Proteome Res.* 2008; 7(3):845–854. [PubMed: 18198819]
9. Choi H, Fermin D, Nesvizhskii AI. Significance analysis of spectral count data in label-free shotgun proteomics. *Mol Cell Proteomics.* 2008; 7(12):2373–2385. [PubMed: 18644780]
10. Griffin NM, Yu J, Long F, Oh P, Shore S, Li Y, Koziol JA, Schnitzer JE. Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nat Biotechnol.* 2010; 28(1):83–89. [PubMed: 20010810]

11. Lu P, Vogel C, Wang R, Yao X, Marcotte EM. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol.* 2007; 25(1): 117–124. [PubMed: 17187058]
12. Pavelka N, Fournier ML, Swanson SK, Pelizzola M, Ricciardi-Castagnoli P, Florens L, Washburn MP. Statistical similarities between transcriptomics and quantitative shotgun proteomics data. *Mol Cell Proteomics.* 2008; 7(4):631–644. [PubMed: 18029349]
13. Zybailov B, Mosley AL, Sardi ME, Coleman MK, Florens L, Washburn MP. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J Proteome Res.* 2006; 5(9):2339–2347. [PubMed: 16944946]
14. Wolters DA, Washburn MP, Yates JR 3rd. An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem.* 2001; 73(23):5683–5690. [PubMed: 11774908]
15. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods.* 2007; 4(3):207–214. [PubMed: 17327847]
16. Qian WJ, Kaleta DT, Petritis BO, Jiang H, Liu T, Zhang X, Mottaz HM, Varnum SM, Camp DG 2nd, Huang L, Fang X, Zhang WW, Smith RD. Enhanced detection of low abundance human plasma proteins using a tandem IgY12-SuperMix immunoaffinity separation strategy. *Mol Cell Proteomics.* 2008; 7(10):1963–1973. [PubMed: 18632595]
17. Zimmer JS, Monroe ME, Qian WJ, Smith RD. Advances in proteomics data analysis and display using an accurate mass and time tag approach. *Mass Spectrom Rev.* 2006; 25(3):450–482. [PubMed: 16429408]
18. Guan J, Luo Y, Denker BM. Purkinje cell protein-2 (Pcp2) stimulates differentiation in PC12 cells by Gbetagamma-mediated activation of Ras and p38 MAPK. *Biochem J.* 2005; 392(Pt 2):389–397. [PubMed: 15948714]
19. Fujishige K, Kotera J, Omori K. Striatum- and testis-specific phosphodiesterase PDE10A isolation and characterization of a rat PDE10A. *Eur J Biochem.* 1999; 266(3):1118–1127. [PubMed: 10583409]
20. Threlfell S, Sammut S, Menniti FS, Schmidt CJ, West AR. Inhibition of Phosphodiesterase 10A Increases the Responsiveness of Striatal Projection Neurons to Cortical Stimulation. *J Pharmacol Exp Ther.* 2009; 328(3):785–795. [PubMed: 19056933]
21. Danielson PE, Watson JB, Gerendasy DD, Erlander MG, Lovenberg TW, de Lecea L, Sutcliffe JG, Frankel WN. Chromosomal mapping of mouse genes expressed selectively within the central nervous system. *Genomics.* 1994; 19(3):454–461. [PubMed: 7910582]

Figure 1. Strategy of recovering low-scoring spectra

After database searching by SEQUEST and stringent data filtering to achieve confident peptide identifications (green line) with a FDR < 1%, many low scoring spectra that correctly identify peptides (in green) are discarded. Recovering these low scoring spectra that match to confidently identified peptides and adding them to the set of spectra that pass initial filtering increases the number of spectral counts for relative quantification.

Figure 2. Spectra recovery rate for proteins with different abundances based on their original spectral counts

All proteins were separated into 6 bins based on their original counts as 1–5, 6–10, 11–20, 21–40, 41–80 and over 80 spectra. The percent of spectra recovered (%) of each bin was calculated by $(\text{total recovered spectral counts}) / (\text{total filtered spectral counts}) \times 100\%$.

Figure 3. Recovered spectra have similar mass error distribution with high scored spectra
 The mass errors of 68270 filtered spectra, 14291 recovered spectra and 204311 spectra referring to the reversed protein sequences from 50 Orbitrap runs are classified into different bins (bin size = 0.8). The percentages of spectra are calculated by (scan counts in the bin) / (total scan counts in the category) $\times$ 100. **Filtered**: spectra passed the stringent filter; **Recovered**: spectra recovered by the identified peptides; **Reverse ID**: spectra referring to the decoy database.

Figure 4. Recovered spectra have similar LC retention time with high scored spectra

The average scan numbers of recovered spectra of all individual peptides are plotted against filtered spectra. Peptides with RSD of filtered scan number >20% are not included. Both datasets are collected from single LC-MS/MS runs. (A), data from Orbitrap. (B), data from LTQ.

Figure 5. Recovered spectra have similar fragment pattern with high scored spectra

All spectra of peptide AGFAGDDAPR from 100 LTQ runs are plotted by XCorr and ΔCn scores. Spectra of the best score (spectrum a) and worst score (spectrum b) marked by arrows are further compared with the theoretical spectra in (B, C).

Figure 6. Reduced percentages of false changes observed in replicate experiments

The $\log_2$ ratios between replicates are calculated as $\log_2[(\text{replicate A} + 0.5) / (\text{replicate B} + 0.5)]$. Proteins with $\log_2$ ratio > 1.6 are considered as false positives. The percentage of false changes in abundance is calculated for each individual protein category. **Filtered**, spectral counting using the spectra that passed the filter. **Filtered + Recovered**, spectral counting using both filtered and recovered spectra.

Figure 7. Improved quantification confidence for selected low abundance region specific brain proteins

(A) Glyoxylate reductase/hydroxypyruvate reductase (GRHPR); (B) Histone H1.4 (HIST1H1E); (C) Acyl carrier protein (NDUFAB1).

Figure 8. Consistence between spectral count data and mRNA data in the brain region specific distribution

Pcp2, purkinje cell protein 2. **Grid2**, glutamate receptor delta-2 subunit. **Psd2**, isoform 3 of pH and SEC7 domain-containing protein 2. **Cdh11**, cadherin-11. **D10Bwg1379e**, brefeldin a-inhibited guanine nucleotide-exchange protein3. **Gda**, guanline deaminase. **Rgs9**, isoform 2 of regulator of G-protein signaling 9. **Gnal**, guanine nucleotide binding protein, alpha stimulating, olfactory type, isoform 2. **Strn**, striatin, calmodulin binding protein. (A), spectral count data. (B), mRNA level information extracted from Allen Brain Atlas (<http://www.brain-map.org>).

Table 1

SEQUEST filtering criteria used for peptide identifications

Charge state	ΔCn	XCorr	Tryptic ends
1+	≥ 0.05	≥ 1.7	Fully
1+	≥ 0.1	≥ 1.5	Fully
1+	≥ 0.05	≥ 3.0	Partially
1+	≥ 0.16	≥ 2.8	Partially
2+	≥ 0.05	≥ 2.8	Fully
2+	≥ 0.1	≥ 2.7	Fully
2+	≥ 0.16	≥ 2.3	Fully
2+	≥ 0.05	≥ 3.8	Partially
2+	≥ 0.1	≥ 3.7	Partially
2+	≥ 0.16	≥ 3.5	Partially
3+	≥ 0.05	≥ 3.5	Fully
3+	≥ 0.1	≥ 3.3	Fully
3+	≥ 0.16	≥ 3	Fully
3+	≥ 0.05	≥ 4.6	Partially
3+	≥ 0.1	≥ 4.5	Partially
3+	≥ 0.16	≥ 4.3	Partially