Published in final edited form as:

J Proteome Res. 2017 December 01; 16(12): 4523–4530. doi:10.1021/acs.jproteome.7b00362.

Quality Assessments of Long-Term Quantitative Proteomic Analysis of Breast Cancer Xenograft Tissues

Jian-Ying Zhou[†], Lijun Chen[†], Bai Zhang[†], Yuan Tian[†], Tao Liu[‡], Stefani N. Thomas[†], Li Chen[†], Michael Schnaubelt[†], Emily Boja[§], Tara Hiltke[§], Christopher R. Kinsinger[§], Henry Rodriguez[§], Sherri R. Davies^{||}, Shunqiang Li^{||}, Jacqueline E. Snider^{||}, Petra Erdmann-Gilmore^{||}, David L. Tabb[⊥], R. Reid Townsend^{||}, Matthew J. Ellis^{||}, Karin D. Rodland[‡], Richard D. Smith[‡], Steven A. Carr[#], Zhen Zhang[†], Daniel W. Chan[†], and Hui Zhang^{*,†}

[†]Department of Pathology, Johns Hopkins University, Baltimore, Maryland 21231, United States

[‡]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99352, United States

[§]Office of Cancer Clinical Proteomics Research, National Cancer Institute, Bethesda, Maryland 20892, United States

^{||}Department of Internal Medicine, Washington University School of Medicine, St. Louis, Missouri 63110, United States

[⊥]Department of Biomedical Informatics, Vanderbilt University Medical School, Nashville, Tennessee 37232, United States

[#]The Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, United States

Abstract

Clinical proteomics requires large-scale analysis of human specimens to achieve statistical significance. We evaluated the long-term reproducibility of an iTRAQ (isobaric tags for relative

*Corresponding Author: huizhang@jhu.edu.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteo-me.7b00362. Figure S1: Sample loading test. Figure S2: Effect of collision energy on iTRAQ quantification. Figure S3: PSMs filtering by reporter ion intensity. (PDF)

ORCID

Tao Liu: 0000-0001-9529-6550

Stefani N. Thomas: 0000-0003-1679-5453

David L. Tabb: 0000-0001-7223-578X

Richard D. Smith: 0000-0002-2381-2349

Hui Zhang: 0000-0001-8726-7098

Author Contributions

J-Y.Z., L.C., Y.T., S.N.T.: Sample preparation and data generation. J-Y.Z., B.Z., L.C., D.L.T., M.S.: Data analysis. T.L., E.B., T.H., C.R.K., H.R., S.R.D., R.R.T., M.J.E., K.D.R., R.D.S., S.A.C., H.Z., Z.Z., D.W.C.: CPTAC group discussion. S.R.D., S.L., J.E.S., P.E.-G.: Preparation of breast cancer xenograft tissues as reference materials to assess the system stability during the proteomic analyses of human tumors. H.Z., Z.Z., D.W.C.: Co-principal Investigators of Johns Hopkins Proteome Characterization Center.

Notes

The authors declare no competing financial interest.

All of the mass spectrometry data were generated at Johns Hopkins and are deposited at the CPTAC Data Coordinating Center as raw files for public access at <https://cptac-data-portal.georgetown.edu/cptac/s/S021> (Studies: TCGA Ovarian Cancer CompRef Samples; Mass Spectrometry Site: Johns Hopkins University, JHUZ).

and absolute quantification)-based quantitative proteomics strategy using one channel for reference across all samples in different iTRAQ sets. A total of 148 liquid chromatography tandem mass spectrometric (LC–MS/MS) analyses were completed, generating six 2D LC–MS/MS data sets for human-in-mouse breast cancer xenograft tissues representative of basal and luminal subtypes. Such large-scale studies require the implementation of robust metrics to assess the contributions of technical and biological variability in the qualitative and quantitative data. Accordingly, we derived a quantification confidence score based on the quality of each peptide-spectrum match to remove quantification outliers from each analysis. After combining confidence score filtering and statistical analysis, reproducible protein identification and quantitative results were achieved from LC–MS/MS data sets collected over a 7-month period. This study provides the first quality assessment on long-term stability and technical considerations for study design of a large-scale clinical proteomics project.

Graphical abstract

Keywords

clinical proteomics; quantification; iTRAQ; tumor tissues; Cancer Biology and Disease Human Proteome Project

INTRODUCTION

Proteomics entails high-throughput protein quantification to identify proteins associated with biological hypotheses¹ and is increasingly used in clinical research to discover protein

candidates for clinical applications.² By analyzing the proteomes of disease-related tissues or body fluids, quantitative proteomics can identify proteins, protein modifications, or protein complexes for disease detection, prognosis, monitoring, and development of novel therapeutic targets.³ Currently, there are several LC–MS/MS-based methods available for proteomics quantification,^{1,4} including label-free quantification,^{5–9} targeted proteomics using selected reaction monitoring (SRM) or multiple reaction monitoring (MRM),¹⁰ stable isotope label-based approaches such as isotope-coded affinity tag (ICAT),¹¹ isobaric tags for relative and absolute quantification (iTRAQ),¹² tandem mass tags (TMT),¹³ and stable isotope labeling by amino acids in cell culture (SILAC).¹⁴ Label-free comparative proteomic strategies include spectral counting and peak area or intensity comparison.^{15–17} Overall, label-based quantitative approaches such as using iTRAQ or TMT tags have better quantification accuracy and reproducibility.^{18–20}

Label-free and label-based quantification approaches are routinely used for clinical proteomics studies.^{21–23} With iTRAQ labeling, quantitative information is obtained from the intensities of low mass reporter ions in the same MS/MS scan. In the past few years, the Clinical Proteomic Tumor Analysis Consortium (CPTAC)^{24,25} under the direction of the U.S. National Cancer Institute (NCI) has achieved the standardization of proteomic workflows on different MS platforms for the robust proteomic analysis of clinical specimens.^{26–28} Large-scale projects such as CPTAC require comparisons across >100 genomically characterized clinical specimens in a single study, which exceeds the capability of traditional iTRAQ-based quantification wherein a maximum of eight different iTRAQ channels can be compared in a sample set. Various novel approaches to increase the multiplexing of quantitative proteomic studies have been developed including NeuCode SILAC, 5-plex SILAC, and the use of isotopologues.^{29–32} However, these approaches are not amenable for mammalian tissues due to the requirement for metabolic labeling. It should also be noted that increasing the scale of multiplexing often results in decreased quantitative accuracy.^{33,34}

Experimental factors potentially introduce variability to the final protein quantitation results,^{35–37} including protein extraction and digestion efficiency, iTRAQ labeling efficiency, sample fractionation, sample storage, details of instrumentation and methodology, peptide ionization, ion suppression, LC separation resolving power, instrument stability, LC buffer change, column change, data analysis method, and so on. Most of these factors are inevitable, especially for the extended periods of time that are required for the analysis of very large sample sets. Thus it is important to establish a strategy to minimize the impact of these factors on the selection of protein candidates. Considerable efforts have been made to achieve reliable iTRAQ quantification results by data analysis, such as excluding peptide-spectrum matches (PSMs) with reporter ion intensities below 5000,³⁸ variability-stabilizing transformation,³⁹ weighting the peptide data according to intensity,⁴ or acquiring data using a triple-stage mass spectrometry (MS3) strategy.³⁴ The MS3 strategy was reported to be quite efficient in eliminating the ratio distortion introduced by the interference of the contaminating near-isobaric ions.²³ However, this strategy can potentially reduce the overall sensitivity and proteome coverage of the study because only a small percentage of MS1 precursor ions are converted into the MS3 reporter ions. Moreover, the majority of the results mentioned above were achieved from the analysis of standard protein mixtures using

channel-to-channel comparison in the same iTRAQ data set. Not much has been done to evaluate the reliability of iTRAQ quantification for a large number of clinical samples analyzed over a long period of time.

We evaluated the possibility of analyzing the relative protein abundances in large sets or a theoretically unlimited number of clinical samples using one iTRAQ channel as the reference⁴⁰ for normalization across all iTRAQ data sets. The long-term reliability of this strategy was evaluated by analyzing complex breast cancer xenograft tissues^{41,42} using 2D-LC-MS/MS analysis. After data filtering by quantification confidence scores determined from the quality of each individual PSM and statistical analysis, the observed false-positive rate (FPR) of identifying protein changes was able to be controlled. Reproducible proteomic quantification results of breast cancer xenograft tissues were achieved over a 7-month period on 2D-LC-MS/MS platforms.

EXPERIMENTAL PROCEDURES

Protein Extraction and Trypsin Digest

Approximately 100 mg of each of the cryopulverized breast cancer xenograft tissues⁴³ were sonicated in 1.5 mL of lysis buffer (8 M urea, 0.8 M NH_4HCO_3 , pH 8.0). Protein concentrations of the tissue lysates were determined by a BCA assay (Pierce, Rockford, IL). Proteins were reduced with 10 mM TCEP for 1 h at 37 °C and subsequently alkylated with 12 mM iodoacetamide for 1 h at room temperature in dark. Samples were then diluted 1:4 with deionized water and digested with sequencing-grade modified trypsin (Promega, Madison, WI) at a 1:50 enzyme-to-substrate ratio. After 12 h of digestion at 37 °C, another aliquot of the same amount of trypsin was added to the samples and the samples were further incubated at 37 °C overnight. The digested samples were then acidified with 10% trifluoroacetic acid to pH <3. Tryptic peptides were desalted on reversed-phase C18 SPE columns (Waters, Milford, MA) and dried using a Speed-Vac.

iTRAQ Labeling of Peptides

Desalted peptides were labeled with 4-plex iTRAQ reagents according to the manufacturer's instructions (AB Sciex, Foster City, CA). Two aliquots of peptides from the basal tumor (1 mg for each aliquot) and two aliquots of peptides from the luminal tumor (1 mg for each aliquot) were dissolved in 125 μL of 0.5 M triethylammonium bicarbonate, pH 8.5 solution, and mixed with 5 units of iTRAQ reagent that was freshly dissolved in 375 μL of ethanol. The two technical replicates of basal tumors were labeled by iTRAQ 114 and 115 and the two technical replicates of luminal tumors were labeled by iTRAQ 116 and 117. After 2 h of incubation at RT, 10% TFA solution was added to decrease the pH to <3 to stop the reaction. Peptides labeled by different iTRAQ reagents were then mixed and desalted on C18 SPE columns.

Offline 2D Fractionation

Offline 2D fractionation was performed using basic reversed-phase liquid chromatography (bRPLC) as previously reported.⁴⁴ Approximately 100 μg of 4-plex iTRAQ labeled sample was separated on a reversed phase Zorbax extend-C-18 column (4.6 $\times$ 100 mm column

containing 1.8 μm particles) from Agilent Technology (Santa Clara, CA) using an Agilent 1220 Infinity HPLC System. The solvent consisted of 10 mM ammonium formate (pH 10) as mobile phase A and 10 mM ammonium formate and 90% ACN (pH 10) as mobile phase B. The separation gradient was set as follows: 2% B for 10 min, from 2 to 8% B for 5 min, from 8 to 35% B for 85 min, from 35 to 95% B for 5 min, and 95% B for 25 min. Ninety-six fractions were collected across the LC separation and were concatenated into 24 fractions by combining fractions 1, 25, 49, 73; 2, 26, 50, 74; and so on. The samples were dried in a Speed-Vac and stored at -80°C until LC-MS/MS analysis.

LC-MS/MS Analysis

Peptides were separated on a Dionex Ultimate 3000 RSLCnano system (Thermo Scientific) with a 75 $\mu\text{m} \times 15\text{ cm}$ Acclaim PepMap100 separating column (Thermo Scientific) protected by a 2-cm guard column (Thermo Scientific). The mobile phase flow rate was 300 nL/min and consisted of 0.1% formic acid in water (A) and 0.1% formic acid, 95% acetonitrile (B). The gradient profile was set as follows: 2–22% B for 70 min, 22–29% B for 8 min, 29–95% B for 4 min, 95% B for 8 min. MS analysis was performed using an LTQ Orbitrap Velos Pro mass spectrometer (Thermo Scientific). The spray voltage was set at 2.2 kV. Orbitrap spectra ($\text{AGC } 1 \times 10^6$) were collected from 400 to 1800 m/z at a resolution of 30 000, followed by data-dependent HCD MS/MS (at a resolution of 7500, collision energy 35%, activation time 0.1 ms) of the 10 most abundant ions using an isolation width of 2.0 Da. Charge-state screening was enabled to reject the generation of MS/MS spectra for unassigned and singly charged precursor ions. A dynamic exclusion time of 40 s was used to discriminate against previously selected ions.

Data Analysis

All LC-MS/MS files were analyzed by a cloud-based proteomic pipeline developed in Johns Hopkins University to perform database search for spectrum assignments using MS-GF+ in this study against a combined human and mouse RefSeq database (version 20160914).^{45,46} A decoy database was used to assess the false discovery rate (FDR) at PSM, peptide, and protein levels.⁴⁷ The precursor ion mass tolerance was set as ± 10 ppm, and a maximum of two missed tryptic cleavages were allowed. Carbamidomethylation on Cys, iTRAQ-4plex labeling on Lys, and N-terminus were set as fixed modifications, and oxidation on Met was set as variable modification. Identified proteins were grouped into nonredundant protein groups using 2 PSMs per peptide and 2 peptides per protein group. Proteins identified in all LC-MS/MS experiments (commonly identified proteins) were used in the quantitative analysis for subsequent reproducibility study.

The iTRAQ reporter ions of spectra were analyzed by Proteome Discoverer 1.3 (Thermo Scientific). The extracted iTRAQ reporter ion intensity of each PSM was further analyzed by in-house-developed software to achieve protein-level quantification. The peptide hydrophobicity score was summarized from the score of each amino acid. The confidence score of each PSM was calculated by the following equation

$$\text{confidence score} = (1 - \alpha_1 \times [\text{reporter ion intensity}]^{\alpha_2}) \times (1 - \beta_1 \times e^{\beta_2 \times [\text{MH}^+]}) - \theta$$

where α_1 , α_2 , β_1 , and β_2 are the values from the nonlinear regression of each factor (Figure 2) and γ is the correction factor to address the distribution of all confidence scores. The protein-level quantification confidence score was the sum of the confidence scores from all constituent PSMs. All data were normalized by the median $\log_2$ ratio of all PSMs to the reference iTRAQ channel.

Data Availability

All of the mass spectrometry data were generated at Johns Hopkins and are deposited at the CPTAC Data Coordinating Center as raw files for public access at <https://cptac-data-portal.georgetown.edu/cptac/s/S021> (Studies: TCGA Ovarian Cancer CompRef Samples; Mass Spectrometry Site: Johns Hopkins University, JHUZ).

RESULTS

General Strategy

To investigate the proteomic differences among a large number of clinical samples, we used 4-plex iTRAQ reagents for proteomic quantification. Three channels of each iTRAQ 4-plex set were used to label peptides from individual samples in a random order, and the remaining channel (the 114 channel in our case) was used as a reference sample to normalize data across all data sets (Figure 1). We did not use 8-plex iTRAQ due to the lower proteome coverage reported in previous studies.^{37,48}

We evaluated the long-term reproducibility of this strategy using a basal and a luminal subtype of patient-derived xenograft breast cancer as previously described.^{41,42,49} After labeling an equal amount of basal breast cancer xenograft peptide sample with iTRAQ channels 114 and 115 and luminal breast cancer xenograft peptide sample with iTRAQ channels 116 and 117, the analyses of these samples were conducted as interstitial runs between the analysis of the clinical samples according to the following strategy: The pooled xenograft samples were analyzed by 2D-LC-MS/MS (six sets) after major instrument maintenance, LC buffer change, column change, and long-term sample storage at 4 °C. After LC-MS/MS data collection over 7 months (from April 2013 to October 2013), all data sets were used to evaluate a standard data analysis pipeline to determine the long-term reproducibility of the experimental procedure.

Proteome Profiling Reproducibility

Quantitative iTRAQ information is derived from the PSMs of identified proteins, and the quantitative comparison of a large number of samples from multiple iTRAQ data sets relies on the commonly identified proteins across different data sets. Therefore, the general proteome coverage is closely related to the throughput of proteomic quantification. In this study, after database searching by MS-GF+,^{45,46} 2 PSMs per peptide, and 2 peptides per protein group were used to filter the data, yielding 906 323 PSMs with 0.023% FDR, 95 842 unique peptides with 0.057% FDR at peptide level, and 8759 protein groups with 0.285% FDR at protein level in six 2D-LC-MS/MS data sets. It should be noted that the data for this study were acquired using an Orbitrap Velos mass spectrometer, which is not the most advanced mass spectrometer with greatly enhanced capabilities. However, a multisite study

that evaluated the stability of proteomic differentiation by mass spectrometry indicated that the data acquired on our Orbitrap Velos had a high degree of reproducibility.⁵⁰ The number of proteins that are identified and quantified would be higher using more current instrument models.

PSM-Level Data Filtering

To evaluate the impact of PSM-level factors on quantification confidence, all PSMs from the 2D-LC-MS/MS data sets were used for data analysis pipeline training. We selected three standard deviations (SDs) of the $\log_2$ ratios from the two duplicate iTRAQ channels as a threshold to determine the false-positive level for PSM quantification (Figure 2a). A total of 16 factors associated with PSM quality were evaluated, including reporter ion intensity, precursor ion intensity, m/z , MH^+ , hydrophobicity,⁵¹ SpScore, XCorr, charge state, ion inject time, isolation interference, mass accuracy, missed cleavage, retention time, matched ions, total ions, as well as ion match rate. In addition to the previously described effect of reporter ion intensity³⁶ (Figure 2b), the precursor ion MH^+ , m/z , and hydrophobicity score of each PSM were also closely correlated to the PSM-level FPR (Figure 2c–e).

On the basis of these results, quantification confidence scores derived from the nonlinear regression of PSM factors were used to evaluate the quality of each PSM (Figure 2f,g). We found that the confidence scores generated from multiple factors removed fewer PSMs for quantification than single-factor confidence scores (Figure 2f). This implied that multifactor scores are more efficient than single-factor scores for outlier removal. However, no significant difference in outlier removal efficiency was found between different multifactor scores. The confidence score generated from two PSM factors, such as reporter ion intensity and hydrophobicity or reporter ion intensity and MH^+ , provided almost the same outlier removal efficiency as the confidence score generated from all of the PSM-related factors. Considering the simplicity of confidence score calculation, we selected a confidence score composed of reporter ion intensity and MH^+ for future data filtering.

By dividing the PSMs into 10 bins based on the confidence score comprised of reporter ion intensity and MH^+ , the PSM-level FPR of each bin was evaluated. The correlation between confidence score and PSM-level FPR was determined to be 0.99 (Figure 2g). After removing PSMs with confidence scores ≤ 0.25 , ~3.8% of the total PSMs, including 46% of the total false-positive PSMs, were removed from the PSM pool. Consequently, the quantification FPR was $<1\%$ for the accepted PSMs and 21% for the removed PSMs (Figure 2h). These data indicate that the confidence score filtering is very efficient in isolating low-quality PSMs.

PSM to Protein Data Summary

Twenty-two different strategies with various combinations of data summary methods (mean, median, and sum) and $\log_2$ ratio transformations (PSM-level, peptide-level, and protein-level) were tested to roll up the PSM-level information to protein-level quantification information. In Figure 3, the strategies are presented as (PSM to peptide method)_(peptide to protein method), and “Log2” indicates the $\log_2$ transformation of the intensity ratios. The relative protein abundances for the luminal breast cancer xenograft samples determined from

pairs of iTRAQ sets were used to compare the different strategies. The Log2_Mean_All strategy transformed all PSM-level intensity ratios between the luminal channel and reference channel to log2 ratios and then averaged all log2 ratios to protein ratio. This strategy generated 99 false-positive proteins and a 0.923 set-to-set correlation. The Log2_Median_All strategy transformed all PSM-level intensity ratios to log2 ratios and then calculated the protein ratio by analyzing the median value of all log2 ratios. This strategy, as a result, generated 123 false-positive proteins and a 0.915 set-to-set correlation. The Sum_Sum_log2 strategy directly summed all PSM-level reporter ion intensities to protein-level intensities and then transformed the intensity ratios to log2. Subsequently, the Sum_Sum_log2 strategy yielded an average of 111 false-positive proteins and a 0.915 set-to-set correlation. Interestingly, the Log2_Mean_Mean strategy, in which the reporter ion intensity ratios were log₂-transformed, averaged to the peptide ratios, and then averaged to the protein ratio, provided a slightly lower number of false positives (91) and slightly higher data set-to-data set correlation (0.924) than all of the other methods. Therefore, the Log2_Mean_Mean approach was selected for subsequent data analysis strategies used in this study.

Protein-Level Data Filtering

To control the FPR of protein quantification, pairs of 2D-LC-MS/MS data sets were used to establish a data filtering strategy. For example, in the comparison between data sets generated in May 2013 and June 2013, the quantification SD of the set-to-set comparison (luminal vs luminal) was determined to be 0.3. By using 3SD as the cutoff (log2 ratio 0.90 or -0.90), 113 false-positive proteins were identified from the luminal-to-luminal comparison with quantification FPR ($[\text{number of false positives}]/[\text{total quantified proteins}] \times 100\%$) of 1.57%. With candidate FPR ($[\text{number of false positives from replicates}]/[\text{number of candidates from biological comparison}] \times 100\%$) of 8.16, 1385 protein candidates were identified from the luminal-to-basal comparison. After removing the proteins with confidence score <0.45 (Figure 4a), 5613 proteins were quantified with a quantification FPR of 0.86% (<1%) and a candidate FPR 4.46% (<5%). Approximately 74.3% of the total false positive proteins were removed. By setting the confidence score cutoff at 1.80, 3802 proteins were quantified with a quantification FPR of 0.18% (<0.5%) and a candidate FPR 0.92% (<1%). Approximately 93.8% of false-positive proteins were removed from the initial list. Both quantification FPR and candidate FPR could be further controlled at 0 when a confidence score of 5.1 was used as the cutoff to remove proteins with lower quantification quality. However, only 1637 proteins (22.8% of the total proteins) were quantified after applying this filter.

Other than FPR, the application of confidence score filtering also significantly improved the inter-set correlation of the protein quantification (Figure 4b,c). For example, in the comparison between two data sets (Figure 4c), the set-to-set correlation changed from 0.85 to 0.93 after removing proteins with confidence score 1.8. The majority of the outliers showing significant difference between replicate data sets (red dots in Figure 4c) were removed. However, the confidence score filtering strategy did not significantly improve the quantification between the replicate channels in the same iTRAQ set because the original correlation without filtering was >0.99 (data not shown).

Statistical Analysis for Selecting Proteins with Significant Relative Abundance Changes

A statistical analysis for candidate selection is presented in Figure 4d,e. To identify proteins showing differential relative abundance between the luminal and basal breast cancer xenograft tissue, we used six 2D-LC-MS/MS data sets for training. Three 2D-LC-MS/MS data sets were used for collecting quantitative information for the luminal (L) sample, and the other three data sets were used for the basal (B) sample. All data were extracted by using channel 114 as the reference. To evaluate false-positives, we used the same groups of 2D-LC-MS/MS data sets to analyze luminal versus luminal differences. A protein-level confidence score of 0.5 was used as the filter, and two-tailed t-test was used to assess the statistical significance (Figure 4d,e). Three different strategies were used for the 3 vs 3 comparison: no filter, no t-test; filter, no t-test, and filter with t-test. The no filter, no t-test strategy generated 1230 candidate proteins in the basal versus luminal comparison and 102 false-positive candidates in the luminal versus luminal comparison when using a two-fold change cutoff (absolute $\log_2$ ratio ≥ 1) (Figure 4d). The FPR of candidate selection was 8.29% (Figure 4e). In contrast, the filter, no t-test method identified 1051 candidates and 15 false candidates. The observed FPR of candidate selection from this method was 1.43%. In the filter with t-test method, 788 candidates and 2 false candidates were identified with a candidate selection FPR of 0.25%. In the same method, when using an absolute $\log_2$ ratio ≥ 1.6 , 316 candidates and 0 false candidates were identified, indicating that the candidate selection FPR is 0. These data indicate that the FPR of iTRAQ quantification using a reference channel can be well-controlled by the combination of confidence score filtering and statistical analysis.

Long-Term Reproducibility

The long-term reproducibility of proteomics quantification was evaluated by analyzing training data sets collected over a period of 7 months at different time points during the analysis of TCGA cancer tissue samples. The differences in protein relative abundance between the luminal and basal breast cancer xenografts or between biological replicates were compared across all 2D-LC-MS/MS iTRAQ data by normalizing to channel 114. The intersample variation of proteins quantified in all six 2D-LC-MS/MS data sets was significantly decreased by the application of the confidence filter (Figure 5a,b).

DISCUSSION

The primary purpose of this study was to evaluate the effectiveness of conventional proteomic platforms for identifying protein candidates associated with clinical questions from large sets of TCGA clinical samples analyzed by LC-MS/MS over a long period of time. We used an isobaric labeling-based quantitative proteomics strategy for LC-MS/MS to fit this purpose. By using a reference channel in each set of iTRAQ-labeled samples, the quantitative information for each individual sample generated by different iTRAQ sets can be normalized for statistical analysis. Theoretically, there is no limitation of the sample size for this strategy. The observed FPR of protein candidate selection can be controlled after using the protein quantification confidence score filtering strategy and statistical analysis. The long-term reproducibility was evaluated by analyzing breast cancer samples during a period of 7 months. We should note that the method for using a reference channel for

normalization across the analysis of a large set of samples has been used for other studies.^{52–55} Although our reference strategy was evaluated by 4-plex iTRAQ, it is also compatible with 8-plex iTRAQ and TMTs.¹³

The variability of iTRAQ data has been reported to be mainly associated with reporter ion intensity.^{36,56} Lower intensity scans generate larger quantification CVs. Among the experiment-related factors, the simplest method to increase the reporter ion intensity is to increase the peptide loading amount in LC–MS/MS analyses. In our testing experiments, we observed that when the loading amount increased from 0.1 to 3 μg , a significant increase was observed in the number of quantified proteins and the average PSM intensities (Figure S1a), whereas the general FPR of quantification was decreased (Figure S1b). However, the optimal loading depends upon chromatography-related factors (packing, gradient, column dimensions, etc.) of the LC–MS/MS platforms. Overloading can easily damage the LC column and significantly affect the reproducibility. To increase the reporter ion intensity, another possible experimental approach is to increase the collision energy level to achieve higher fragmentation efficiency (Figure S2). The drawback of this method is that it can potentially decrease the number of identified peptides and the proteome coverage as a result of the overfragmentation of some peptides. In a parallel multiple energy experiment, we used 35% collision energy for the first 60 min of the gradient and 45% energy for the last 30 min of the gradient. Increased reporter ion intensity for large peptides and decreased quantification FPR were observed without a significant decrease in proteome coverage (Figure S2a,b). Thus a stepped normalized collisional energy scheme can improve the reproducibility of iTRAQ quantification, which is consistent with the results from other published studies.^{48,57,58}

It is well-known that the strategy used for data analysis is critical for iTRAQ quantification. In this study, we found that other than the reporter ion intensity, the MH^+ , m/z , and hydrophobicity of each peptide were also closely related to quantification confidence, indicating that iTRAQ quantification has relatively higher variability for large or hydrophobic peptides than for those with average molecular weights and hydrophobicity. Thus for data filtering we introduced a quantification confidence score composed of the quality of each protein's PSMs to control the confidence of proteomics quantification. However, the PSM-level filtering can only provide a very limited improvement in the quantification confidence (Figure S3). Thus we used the protein-level confidence score summarized from all PSMs for data filtering, which isolated false-positive candidates from proteins with confident quantification (Figure 3a). This strategy not only considers the quality of PSMs but also considers the number of high-quality PSMs used for protein-level quantification. Most of the false candidates could be removed by using the confidence score filter. This is important because the downstream validation of protein candidates could require a considerable amount of time and resources.⁵⁹ The confidence score cutoff threshold can be set depending on the number of quantifiable biological replicates. For the proteins quantified in a sufficient number of replicates, the cutoff can be set very low to accept more quantified proteins, and statistical methods can be used to minimize the number of false-positives. Conversely, for the proteins quantified in only one biological replicate, a higher confidence score cutoff to achieve an acceptable candidate FPR is more appropriate.

The application of confidence score filtering significantly improved the quality of the quantification results for samples that were difficult to analyze. Sample contamination (Figure 3b,d) and the low quantity of the clinical samples (data not shown) significantly affected the proteome coverage and quantitation accuracy; however, after using the confidence score filtering, the proteomics quantification data from these samples were still consistent with the better quality samples. Our data also indicated that by using the reference strategy, quantification confidence filtering, and statistical analysis, we were able to consistently identify hundreds of proteins exhibiting differential relative abundance under two different conditions over a long period of time with an FPR approaching 0.

In conclusion, this study evaluated the long-term reproducibility of a quantitative proteomics strategy for the identification of biological question-associated protein candidates from large sets of clinical samples, which required months of LC–MS/MS instrument analysis time. After confidence score filtering and statistical analysis, consistent quantification results can be observed from 2DLC–MS/MS data sets generated over a long period of time. We believe that this strategy will significantly improve the quality of proteomic data, especially for large-scale clinical studies. Additionally, prior to conducting such studies, careful attention should be given to the development of quality-control methods, standard operating procedures, and identification-independent quality metrics.^{50,60}

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by the National Institutes of Health under grants and contracts of the National Cancer Institute, Clinical Proteomics Tumor Analysis Consortium (U24CA160036 and U24CA210985).

References

1. Ong SE, Mann M. Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol.* 2005; 1(5):252–62. [PubMed: 16408053]
2. Zhang H, Chan DW. Cancer biomarker discovery in plasma using a tissue-targeted proteomic approach. *Cancer Epidemiol, Biomarkers Prev.* 2007; 16(10):1915–7. [PubMed: 17932335]
3. Gillette MA, Carr SA. Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. *Nat Methods.* 2012; 10(1):28–34.
4. Li Z, et al. Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos. *J Proteome Res.* 2012; 11(3):1582–90. [PubMed: 22188275]
5. Wang W, et al. Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal Chem.* 2003; 75(18):4818–26. [PubMed: 14674459]
6. Zybaïlov B, et al. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J Proteome Res.* 2006; 5(9):2339–47. [PubMed: 16944946]
7. Zhou JY, et al. Improved LC-MS/MS spectral counting statistics by recovering low-scoring spectra matched to confidently identified peptide sequences. *J Proteome Res.* 2010; 9(11):5698–704. [PubMed: 20812748]
8. Zhang H, et al. High throughput quantitative analysis of serum proteins using glycopeptide capture and liquid chromatography mass spectrometry. *Mol Cell Proteomics.* 2005; 4(2):144–55. [PubMed: 15608340]

9. Zhou JY, et al. Trauma-associated human neutrophil alterations revealed by comparative proteomics profiling. *Proteomics: Clin Appl.* 2013; 7(7–8):571–83. [PubMed: 23589343]
10. Kennedy JJ, et al. Demonstrating the feasibility of large-scale development of standardized assays to quantify human proteins. *Nat Methods.* 2013; 11(2):149–55. [PubMed: 24317253]
11. Gygi SP, et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol.* 1999; 17(10):994–9. [PubMed: 10504701]
12. Ross PL, et al. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics.* 2004; 3(12):1154–69. [PubMed: 15385600]
13. Thompson A, et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem.* 2003; 75(8):1895–904. [PubMed: 12713048]
14. Ong SE, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics.* 2002; 1(5):376–86. [PubMed: 12118079]
15. Anderle M, et al. Quantifying reproducibility for differential proteomics: noise analysis for protein liquid chromatography-mass spectrometry of human serum. *Bioinformatics.* 2004; 20(18):3575–82. [PubMed: 15284095]
16. Kim YJ, et al. Reproducibility assessment of relative quantitation strategies for LC-MS based proteomics. *Anal Chem.* 2007; 79(15):5651–8. [PubMed: 17580949]
17. Old WM, et al. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics.* 2005; 4(10):1487–502. [PubMed: 15979981]
18. Trinh HV, et al. iTRAQ-Based and Label-Free Proteomics approaches for studies of human adenovirus infections. *Int J Proteomics.* 2013; 2013:1–16.
19. Megger DA, et al. Comparison of label-free and label-based strategies for proteome analysis of hepatoma cell lines. *Biochim Biophys Acta, Proteins Proteomics.* 2014; 1844(5):967–76.
20. Bantscheff M, et al. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem.* 2012; 404(4):939–65. [PubMed: 22772140]
21. Crutchfield CA, et al. Advances in mass spectrometry-based clinical biomarker discovery. *Clin Proteomics.* 2016; 13:1. [PubMed: 26751220]
22. Han DK, et al. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat Biotechnol.* 2001; 19(10):946–51. [PubMed: 11581660]
23. Sandin M, Chawade A, Levander F. Is label-free LC-MS/MS ready for biomarker discovery? *Proteomics: Clin Appl.* 2015; 9(3–4):289–94. [PubMed: 25656266]
24. Ellis MJ, et al. Connecting genomic alterations to cancer biology with proteomics: the NCI Clinical Proteomic Tumor Analysis Consortium. *Cancer Discovery.* 2013; 3(10):1108–12. [PubMed: 24124232]
25. Skates SJ, et al. Statistical design for biospecimen cohort size in proteomics-based biomarker discovery and verification studies. *J Proteome Res.* 2013; 12(12):5383–94. [PubMed: 24063748]
26. Zhang H, et al. Integrated proteogenomic characterization of human high grade serous ovarian cancer. *Cell.* 2016; 166(3):755–65. [PubMed: 27372738]
27. Mertins P, et al. Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature.* 2016; 534(7605):55–62. [PubMed: 27251275]
28. Zhang B, et al. Proteogenomic characterization of human colon and rectal cancer. *Nature.* 2014; 513(7518):382–7. [PubMed: 25043054]
29. Hebert AS, et al. Neutron-encoded mass signatures for multiplexed proteome quantification. *Nat Methods.* 2013; 10(4):332–4. [PubMed: 23435260]
30. Merrill AE, et al. NeuCode labels for relative protein quantification. *Mol Cell Proteomics.* 2014; 13(9):2503–12. [PubMed: 24938287]
31. Molina H, et al. Temporal profiling of the adipocyte proteome during differentiation using a five-plex SILAC based strategy. *J Proteome Res.* 2009; 8(1):48–58. [PubMed: 18947249]
32. Tzouros M, et al. Development of a 5-plex SILAC method tuned for the quantitation of tyrosine phosphorylation dynamics. *Mol Cell Proteomics.* 2013; 12(11):3339–49. [PubMed: 23882028]

33. Ow SY, et al. iTRAQ underestimation in simple and complex mixtures: “the good, the bad and the ugly”. *J Proteome Res.* 2009; 8(11):5347–55. [PubMed: 19754192]
34. Ting L, et al. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat Methods.* 2011; 8(11):937–40. [PubMed: 21963607]
35. Evans C, et al. An insight into iTRAQ: where do we stand now? *Anal Bioanal Chem.* 2012; 404(4): 1011–27. [PubMed: 22451173]
36. Hultin-Rosenberg L, et al. Defining, comparing, and improving iTRAQ quantification in mass spectrometry proteomics data. *Mol Cell Proteomics.* 2013; 12(7):2021–31. [PubMed: 23471484]
37. Mahoney DW, et al. Relative quantification: characterization of bias, variability and fold changes in mass spectrometry data from iTRAQ-labeled peptides. *J Proteome Res.* 2011; 10(9):4325–33. [PubMed: 21755926]
38. Hu J, et al. Optimized proteomic analysis of a mouse model of cerebellar dysfunction using amine-specific isobaric tags. *Proteomics.* 2006; 6(15):4321–34. [PubMed: 16800037]
39. Karp NA, et al. Addressing accuracy and precision issues in iTRAQ quantitation. *Mol Cell Proteomics.* 2010; 9(9):1885–97. [PubMed: 20382981]
40. Qian WJ, et al. Large-scale multiplexed quantitative discovery proteomics enabled by the use of an (18)O-labeled “universal” reference sample. *J Proteome Res.* 2009; 8(1):290–9. [PubMed: 19053531]
41. Ding L, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature.* 2010; 464(7291):999–1005. [PubMed: 20393555]
42. Li S, et al. Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts. *Cell Rep.* 2013; 4(6):1116–30. [PubMed: 24055055]
43. Mertins P, et al. Ischemia in tumors induces early and sustained phosphorylation changes in stress kinase pathways but does not affect global protein levels. *Mol Cell Proteomics.* 2014; 13(7):1690–704. [PubMed: 24719451]
44. Yang F, et al. High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. *Expert Rev Proteomics.* 2012; 9(2):129–34. [PubMed: 22462785]
45. Kim S, et al. Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. *J Proteome Res.* 2008; 7:3354–3363. [PubMed: 18597511]
46. Kim S, Pevzner PA. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat Commun.* 2014; 5:5277. [PubMed: 25358478]
47. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods.* 2007; 4(3):207–14. [PubMed: 17327847]
48. Pichler P, et al. Peptide labeling with isobaric tags yields higher identification rates using iTRAQ 4-plex compared to TMT 6-plex and iTRAQ 8-plex on LTQ Orbitrap. *Anal Chem.* 2010; 82(15): 6549–58. [PubMed: 20593797]
49. Zhang B, et al. Proteogenomic characterization of human colon and rectal cancer. *Nature.* 2014; 513(7518):382–7. [PubMed: 25043054]
50. Tabb DL, et al. Reproducibility of differential proteomic technologies in CPTAC fractionated xenografts. *J Proteome Res.* 2016; 15(3):691–706. [PubMed: 26653538]
51. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol.* 1982; 157(1):105–132. [PubMed: 7108955]
52. Kramer G, et al. Accuracy and reproducibility in quantification of plasma protein concentrations by mass spectrometry without the use of isotopic standards. *PLoS One.* 2015; 10(10):e0140097. [PubMed: 26474480]
53. Srinivasan S, et al. iTRAQ quantitative proteomics in the analysis of tears in dry eye patients. *Invest Ophthalmol Visual Sci.* 2012; 53(8):5052–9. [PubMed: 22736608]
54. Geiger T, et al. Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. *Nat Protoc.* 2011; 6(2):147–57. [PubMed: 21293456]
55. Lebert D, et al. DIGESTIF: a universal quality standard for the control of bottom-up proteomics experiments. *J Proteome Res.* 2015; 14(2):787–803. [PubMed: 25495225]
56. Bantscheff M, et al. Robust and sensitive iTRAQ quantification on an LTQ Orbitrap mass spectrometer. *Mol Cell Proteomics.* 2008; 7(9):1702–13. [PubMed: 18511480]

57. Diedrich JK, Pinto AF, Yates JR 3rd. Energy dependence of HCD on peptide fragmentation: stepped collisional energy finds the sweet spot. *J Am Soc Mass Spectrom.* 2013; 24(11):1690–9. [PubMed: 23963813]
58. Pichler P, et al. Improved precision of iTRAQ and TMT quantification by an axial extraction field in an Orbitrap HCD cell. *Anal Chem.* 2011; 83(4):1469–74. [PubMed: 21275378]
59. White FM. The potential cost of high-throughput proteomics. *Sci Signaling.* 2011; 4(160):pe8.
60. Bennett KL, et al. The 2012/2013 ABRF proteomic research group study: assessing longitudinal intralaboratory variability in routine peptide liquid chromatography tandem mass spectrometry analyses. *Mol Cell Proteomics.* 2015; 14(12):3299–309. [PubMed: 26435129]

Figure 1.

Strategy for long-term large-scale clinical proteomics research. After protein extraction and trypsin digestion, peptides were labeled with iTRAQ reagents. Three channels of each iTRAQ 4-plex set were used to label peptides from individual samples in a random order, and the remaining channel was used for a reference sample to normalize the data across all of the iTRAQ data sets. For training purposes, an equal amount of peptides from basal breast cancer xenograft tissues was labeled with iTRAQ channel 114 and 115, and peptides from luminal breast cancer xenograft samples were labeled with iTRAQ channels 116 and 117. The samples were then pooled and analyzed by 1D- or 2D-LC-MS/MS at different time points throughout a period of 7 months.

Figure 2.

PSM-level data filtering. (a) Reproducibility of iTRAQ reporter ion intensity from replicate iTRAQ channels in all 2D-LC-MS/MS data sets. 3SD of the log₂ ratio was used as the cutoff to identify false-positive PSMs. (b–e) Correlation between PSM score cutoff and false-positive rate (FPR) of removed PSMs. (b) Reporter ion intensity. (c) MH⁺. (d) m/z. (e) Hydrophobicity. (f) Data filtering by single or multiple PSM scores. (g–h) Isolation of low-quality PSMs by confidence score composed of reporter ion intensity and MH⁺.

Figure 3.

Summary of PSM-level quantification to protein-level quantification. Strategies are labeled as (PSM to peptide strategy) (peptide to protein strategy). “Log₂” indicates the point at which the intensity values were log₂-transformed. Data summarized from at least 10 pairs of 2D-LC-MS/MS data sets. #, method selected for subsequent data analysis.

Figure 4.

Protein-level data filtering. (a) Isolation of false-positive protein changes generated in the two technical replicate 2D-LC-MS/MS data sets by confidence score filtering.

Quantification FPR refers to the FPR of luminal to luminal comparison. Candidate FPR refers to (number of false-positive proteins in the luminal to luminal comparison)/(number of proteins showing difference between luminal and basal) $\times$ 100%. (b) Improvement in data correlation between two data sets by confidence score filtering. (c) Reproducibility of replicate 2D-LC-MS/MS data sets (May 2013 vs June 2013). (d) 3 vs 3 statistical analysis. 3L/3B: three 2D-LC-MS/MS data sets for luminal breast cancer versus three 2D-LC-MS/MS data sets for basal breast cancer. 3L/3L: three 2D-LC-MS/MS data sets for luminal breast cancer versus another three 2D-LC-MS/MS data sets for luminal breast cancer. Protein level confidence score >0.5 was used as the filter. P values were from two-tailed t -test. (e) FPR analysis for panel d.

Figure 5. Long-term reproducibility of proteome quantification. Statistical analysis of quantified proteins comparing luminal and basal samples with (b) or without (a) the confidence score filter. Data summarized from replicate data sets. Only proteins quantified in all 2D-LC-MS/MS data sets are presented.