Software for Quantitative Proteomic Analysis Using Stable Isotope Labeling and Data Independent Acquisition

ARTICLE *in* ANALYTICAL CHEMISTRY · AUGUST 2011

Impact Factor: 5.64 · DOI: 10.1021/ac201555m · Source: PubMed

8 AUTHORS, INCLUDING:

Changhai Tian
University of Nebraska at Omaha
24 PUBLICATIONS **523** CITATIONS
SEE PROFILE

Kai Fu
University of Nebraska at Omaha
98 PUBLICATIONS **3,168** CITATIONS
SEE PROFILE

Jialin Zheng
University of Nebraska Medical Center
123 PUBLICATIONS **3,562** CITATIONS
SEE PROFILE

Scott J Geromanos
Waters Corporation
41 PUBLICATIONS **5,304** CITATIONS
SEE PROFILE

analytical chemistry

ARTICLE

pubs.acs.org/ac

Software for Quantitative Proteomic Analysis Using Stable Isotope Labeling and Data Independent Acquisition

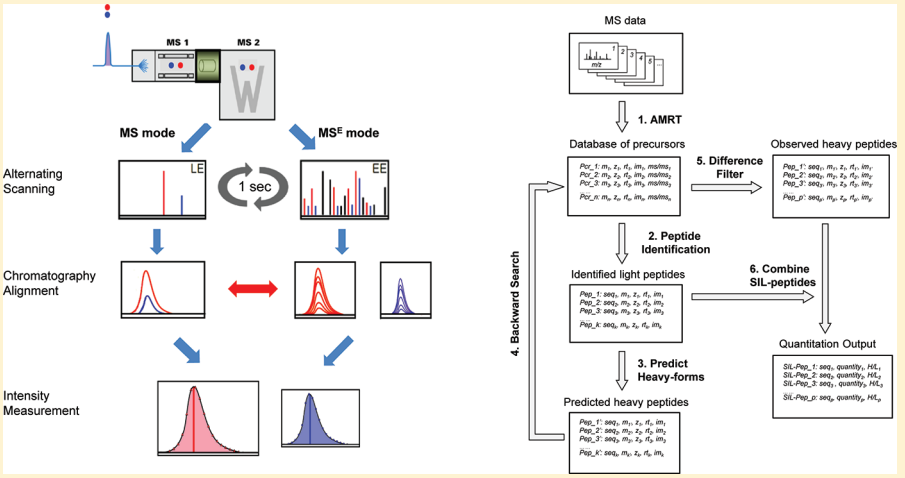
Xin Huang,[†] Miao Liu,[†] Michael J. Nold,[‡] Changhai Tian,[∥] Kai Fu,[†] Jialin Zheng,[∥] Scott J. Geromanos,[‡] and Shi-Jian Ding*[,†,§]

[†]Department of Pathology and Microbiology, [∥]Department of Pharmacology and Experimental Neuroscience, and [§]Mass Spectrometry and Proteomics Core Facility, University of Nebraska Medical Center, Omaha, Nebraska 68198, United States

[‡]Waters Corporation, 34 Maple Street, Milford, Massachusetts 01757, United States

Supporting Information

ABSTRACT:



Many software tools have been developed for analyzing stable isotope labeling (SIL)-based quantitative proteomic data using data dependent acquisition (DDA). However, programs for analyzing SIL-based quantitative proteomics data obtained with data independent acquisition (DIA) have yet to be reported. Here, we demonstrated the development of a new software for analyzing SIL data using the DIA method. Performance of the DIA on SYNAPT G2MS was evaluated using SIL-labeled complex proteome mixtures with known heavy/light ratios (H/L = 1:1, 1:5, and 1:10) and compared with the DDA on linear ion trap (LTQ)-Orbitrap MS. The DIA displays relatively high quantitation accuracy for peptides cross all intensity regions, while the DDA shows an intensity dependent distribution of H/L ratios. For the three proteome mixtures, the number of detected SIL-peptide pairs and dynamic range of protein intensities using DIA drop stepwise, whereas no significant changes in these aspects using DDA were observed. The new software was applied to investigate the proteome difference between mouse embryonic fibroblasts (MEFs) and MEF-derived induced pluripotent stem cells (iPSCs) using $^{16}O/^{18}O$ labeling. Our study expanded the capacities of our UNiquant software pipeline and provided valuable insight into the performance of the two cutting-edge MS platforms for SIL-based quantitative proteomic analysis today.

Quantitative proteomics involves the identification and quantitation of protein components in various biological systems, which hold tremendous potential to reveal large-scale kinetics of proteomes and directly uncover important points in the signaling pathways that control cellular decisions.[1,2] In particular, the integration of liquid chromatography and tandem mass spectrometry (LC-MS/MS), for protein identification and quantitation, has significantly expanded the scale of proteomics studies.[3−5] In order to compare differences in cellular status or between biological systems, the "bottom-up" quantitative proteomics is commonly performed using stable isotope labeling (SIL).[6,7] In these experiments, stable isotopes are introduced into selected proteins or peptides during sample preparation either chemically, enzymatically, or metabolically.[8,9] The major advantage of SIL-based quantitation is that it allows for large-scale and direct comparisons of protein abundances by comparison of the signal intensities of peptides and their stable isotope-labeled analogues.[10,11] In general, SIL methods minimize the variability that can arise during sample processing steps and LC-MS/MS analyses, thereby providing results with less systematic

Received: March 6, 2011
Accepted: August 11, 2011
Published: August 11, 2011

error and higher reproducibility.[12] However, analyzing SIL-based quantitative proteomic data is still technically challenging due to the enormous complexity of the data obtained by a variety of SIL methods, on a diversity of MS platforms using different acquisition methods.[13]

To date, most of the data, generated in quantitative proteomics experiments, were obtained using a data dependent acquisition (DDA) method. In these experiments, an LC-MS/MS data collection cycle starts with a high accuracy, full MS survey of the all precursor ions and is followed by selecting a number of precursor ions for MS/MS fragmentation.[7] The major advantage of DDA is that the fragment ions are derived mostly from a single precursor ion, increasing the specificity of amino acid sequences for that peptide.[14] However, due to the limited and stochastic nature of MS/MS sampling for peptide sequencing on the chromatographic time scale, oftentimes only high abundance precursor ions are fragmented. Consequently, the number of peptides and coverage of identified proteins might be compromised.[15,16] The data independent acquisition (DIA) strategy has been developed to complement the DDA method for proteomic analysis.[17] Instead of a serial selection of precursor ions for data dependent fragmentation, the DIA approach fragments a group of coeluting precursor ions at each given time, enabling a more unbiased detection of all LC-eluted peptides compared to the DDA method.[18−20] The DIA strategy has been recently implemented on two MS platforms: Exactive Orbitrap (Thermo Scientific) and SYNAPT G2 (Waters Co.). The corresponding DIA method was named as all-ion fragmentation[21] (AIF) and LC-MS^E technology,[22−24] respectively.

Many software tools have been produced to analyze SIL-based quantitative proteomics data mainly obtained by the DDA method, such as ASAPRatio,[25] Census,[26] MaxQuant,[27] Vista,[28] and UNiquant.[29] However, programs for analyzing SIL-based quantitative proteomics data obtained with the DIA strategy have yet to be reported. Here, we demonstrated the development of a new addition to our UNiquant software pipeline for analyzing SIL data obtained with LC-MS^E technology. The performance of this software was evaluated using SIL-labeled complex proteome mixtures having either known or unknown ratios. The new software enhanced the capacities of our UNiquant software pipeline for SIL-based quantitative proteomic analysis and enabled us to compare the performance of SYNAPT G2MS using DIA with linear ion trap (LTQ)-Orbitrap using DDA for SIL-based quantitative proteomic analysis, for the first time.

MATERIALS AND METHODS

Sample Preparation. For stable isotope labeling of amino acids in cell culture (SILAC) preparations, human breast cancer cell line MDA-MB-231 was cultured in either SILAC "heavy" (L-[$^{13}C_6$,$^{15}N_4$]-arginine and L-[$^{13}C_6$,$^{15}N_2$]-lysine) or "light" (L-[$^{12}C_6$,$^{14}N_4$]-arginine and L-[$^{12}C_6$,$^{14}N_2$]-lysine) medium for five passages. These cells were lysed in 7 M urea, 2 M thiourea, and 50 mM ammonium bicarbonate. Three proteome mixtures were generated from the heavy (H) and light (L) lysates, with the total protein weight ratios of heavy/light (H/L) = 1:1, 1:5, and 1:10. The proteome mixture samples were reduced by dithiothreitol and alkylated by iodoacetamide. Tryptic digestion was performed, and resulting peptides were desalted and fractionated by solid phase extraction (SPE) using Sep-Pak C_{18} cartridges (Waters Co., Milford, MA). In the desalting step, peptides were washed with 5 mL of 5% acetonitrile with 0.1% trifluoroacetic acid (TFA). Afterward, peptides were eluted with 1 mL of 20%, 30%, 35%, 40%, and 80% acetonitrile/0.1% TFA sequentially. The five SPE-eluted fractions for each proteome mixture were collected and dried in a vacuum centrifuge.

For $^{16}O/^{18}O$ preparations, mouse embryonic fibroblasts (MEFs) were isolated from C57BL/6J mice. MEF-derived induced pluripotent stem cells (iPSCs) were obtained by four-factor-mediated (Oct4, Sox2, Klf4, and c-Myc) retroviral nuclear reprogramming.[30] The MEFs and iPSCs proteins were lysed, reduced, alkylated, and digested using the same protocol as described above. Trypsin-catalyzed $^{16}O/^{18}O$ labeling was carried out to label the tryptic peptides from MEFs (^{16}O) and iPSCs (^{18}O) as previously described.[31] The ^{16}O- and ^{18}O-labeled peptides were mixed in equal amounts. Three technical replicates were prepared (Full methods are available in the Supporting Information.)

LC-MS^E Analysis and Protein Identification. For the $^{16}O/^{18}O$ samples, we employed two-dimensional online reverse phase (RP) chromatography using a nanoACQUITY UPLC system (Waters Co.). The first dimensional separation of the $^{16}O/^{18}O$-labeled cell digests used a hybrid silica, XTerra MS C_{18} 3.5 μm, 150 × 2.1 mm column (Waters Co.). Five fractions were collected from each MEF/iPSC mixture.[32] The second dimensional LC separation used a Symmetry C_{18} 5 μm, 5 mm × 300 μm precolumn and a BEH C_{18} 1.7 μm, 75 μm × 100 mm analytical RP column. For the SILAC-labeled tryptic peptides, we also used a two-dimensional separation. The first dimensional separation used SPE as described above. The second dimensional LC separation was the same as that used in the $^{16}O/^{18}O$ experiments. Mass spectrometry analysis of tryptic peptides was performed using a hybrid quadrupole tandem time-of-flight (Q-TOF) mass spectrometer, SYNAPT G2, equipped with Tri-Wave ion guides that trap and separate ions by ion mobility (Waters Co.). Accurate MS data were collected using alternating, low energy (LE) and elevated energy (EE), survey of acquisitions. ProteinLynx Global Server (PLGS, Waters Co.) was used to analyze the obtained MS data. Briefly, ion detection, clustering, and retention time alignment were processed using an AMRT (accurate mass retention time) method in PLGS.[33] The AMRT data were searched against the Swiss-Prot human protein database using a dual-pass algorithm as previously described.[16] (Full methods are available in the Supporting Information.)

SIL-Based Quantitation of Peptides and Proteins. SIL-based quantitation of proteins/peptides was performed by an in-house developed program written in C# programming language. This program is freely available for academic users upon request. The AMRT data for each LC-MS run were exported to a local database. Included in this output were the weight-averaged monoisotopic mass, charge state, ion drift, charge-state-reduced sum intensity, observed apex retention time, and observed start and stop time of the detected ions. Information for the identified peptides (light precursor species) was exported as well. This contained all the theoretical and experimental properties associated with the identified precursor MS spectrum, such as the unique spectrum id, mass-over-charge (m/z), retention time, peptide sequence, and the identification scores. Theoretical masses of SIL-heavy precursors were determined from the sequence composition of the corresponding identified SIL-light precursors. The predicted SIL-heavy precursor was used to search the AMRT database for an observed ion that matched this SIL-heavy precursor within user-defined thresholds of mass accuracy, elution time, and ion drift. The default settings were as follows: mass accuracy, <5 ppm; difference in retention time,

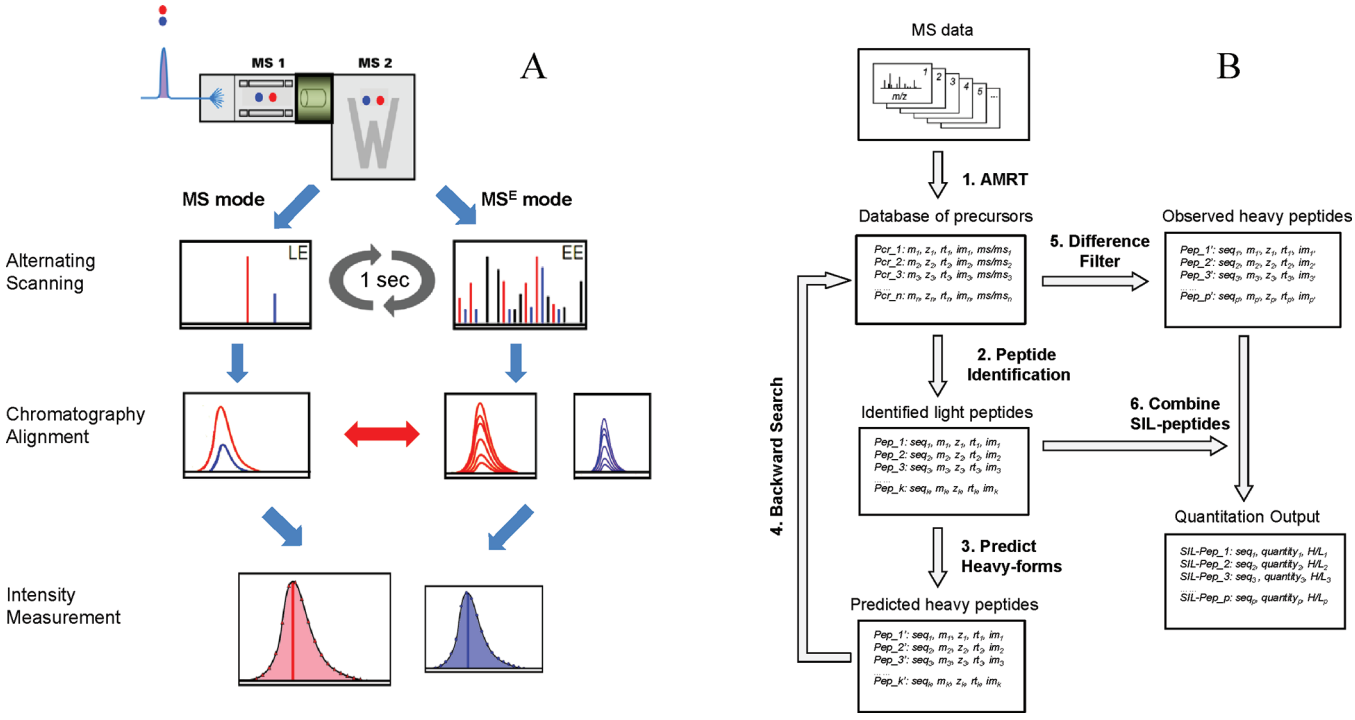


Figure 1. Data independent acquisition strategy for SIL-based quantitative proteomic analysis. (A) Schematic diagram of the LC-MS^E technology. A pair of coeluting SIL-peptides was analyzed through alternating scanning, ion mobility separation, chromatography alignment, and intensities measurement. The low energy (LE) mode of alternating scan enabled the full mass survey of the precursors to be obtained. The elevated energy (EE) mode enabled the MS^E survey of the product ions to be obtained. In the product ion spectrum, the b-ions (black) are identical for both the heavy and light precursors, while a mass shift can be observed in the isotope containing y-ions (red and blue). (B) The data processing flowchat of the SIL-peptide detection and quantitation algorithm. Briefly, MS data were preprocessed using the AMRT method. Attributions (mass, charge, retention time, and ion mobility) of all precursor ions were exported in a local database for peptide identification. The masses of the heavy forms for the identified peptides were determined by the theoretical mass shift of heavy precursors according to the SIL method, and other attributions of the heavy peptides were identical to the light forms. The predicted heavy peptides were searched back in the precursor database by applying the cutoffs dependent on the difference of the theoretical and observed attributions of the peptides. Finally, the observed pairs of SIL-peptides were outputted for quantitation.

<0.05 min; and difference in ion drift, ≤0.5. Intensity information of SIL-heavy precursors was extracted, and the heavy/light pairs were sorted and arranged with the annotated peptide sequence and protein entry. The relative abundance (heavy versus light (H/L) ratio) of each identified peptide was calculated as the sum of the intensities for the heavy precursors divided by the sum of the intensities for the light precursors.[29]

Quantitation of SIL Proteome Mixtures with Known H/L Ratios Using DDA. We adopted a previously published SILAC data set of proteome mixture with known ratios obtained by DDA approach.[29] Briefly, mammalian cells were metabolically labeled with either SILAC heavy (L-[$^{13}C_6$]-arginine and L-[$^{13}C_6$]-lysine) or light (L-[$^{12}C_6$]-arginine and L-[$^{12}C_6$]-lysine) medium. The heavy and light lysates were mixed in three known H/L ratios of 1:1, 1:5, and 1:10. After tryptic digestion, peptides were separated by strong cation chromatography and analyzed using a nanoLC-LTQ-Orbitrap mass spectrometer (Thermo Scientific). The MS analyzer was operated in a DDA mode, with the most intense ions (up to five, depending on precursor intensity) sequentially isolated for CID fragmentation in the linear ion trap. Survey full scan MS spectra (from m/z 375 to 1575) were acquired in the Orbitrap with resolution $R = 100\,000$. SILAC peptides were identified using Mascot (Matrix Science, London, U.K.) to search against a human protein database (IPI human version 3.52) using a decoy technique and quantified using the UNiquant software.[29]

Western Blot Analysis. Preparation of cell lysates for Western blot were carried out as described previously.[34] Briefly, total cell extracts of MEF and iPSC were prepared using M-PER Mammalian Protein Extraction Reagent (Pierce, Rockford, IL). Equal amounts of protein (~20 ug) were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and probed with primary antibodies including: Pcna, β-Actin (Santa Cruz Biotech, Santa Cruz, CA), Oct-4a, Sumo2/3, Ppp2ca (Cell Signaling, Danvers, MA), and Vdac1 (Abcam, Cambridge, MA). Signal intensity of the blotting was measured by ImageJ software.

RESULTS

Software Development for SIL-Based Quantitative Proteomics Using LC-MS^E Technology. A schematic diagram of the LC-MS^E strategy for quantitative proteomic analysis is illustrated in Figure 1A. The coeluting SIL-peptides were injected into the mass spectrometer and scanned by alternating LE mode for precursor ions (MS spectra) and EE modes for product ions (MS^E spectra). All light peptides were identified by database search using PLGS with false discovery/positive rate at 1% at peptide level and 4% at protein level, as previously described.[16] The software for quantitation analysis was developed on the platform of Microsoft .net framework. The programming language was C#. The program was installed on an Intel dual-core PC with 3G RAM. After database searching, the extraction of

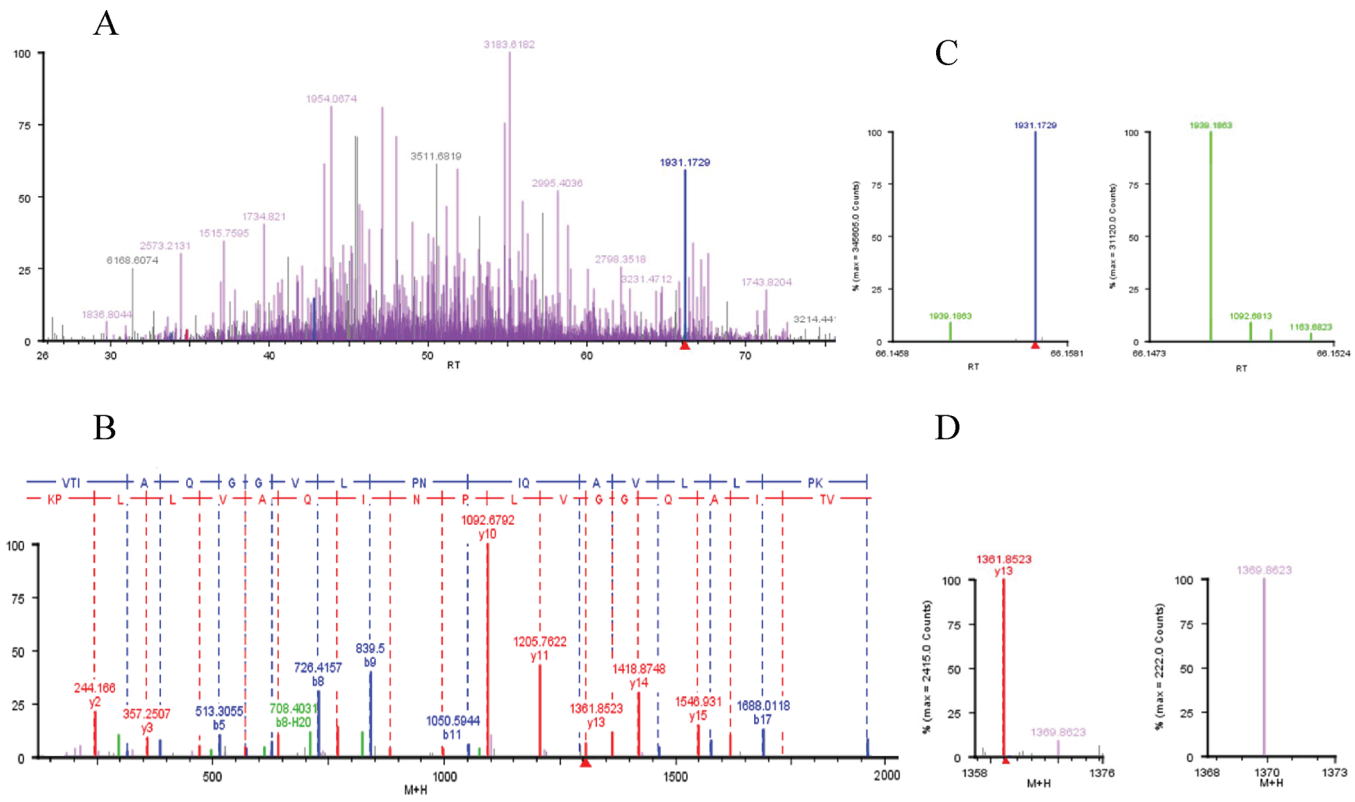


Figure 2. Typical example of LC-MS^E analysis (H/L = 1:10 sample, fraction 1). (A) A chromatogram showing peptide $[M + H]^+$ monoisotopic masses versus retention time (RT) taken in low energy mode. The numbers indicate the masses of selected precursors. (B) A light precursor (blue stick with red triangle, m/z = 1931.1729, RT = 66.16 min) was fragmented in elevated energy mode and identified as "VTIAQGGVLPNIQAVLLPK" from protein entry: Q96QV6, name: histone H2A type 1-A (H2A1A_HUMAN) by database search. The identified b-ions (blue) and y-ions (red) were labeled correspondingly. (C) The coeluting SIL-heavy form (green, right panel) of the identified light precursor (blue, left panel). (D) The y13 product ion from the light form (red, left panel, m/z = 1361.8523) has a corresponding SIL-heavy form (purple, right panel, m/z = 1369.86.23) which comes from the corresponding SIL-heavy precursor. The DIA strategy allows this coeluting heavy SIL-peptide to be observed in both the same precursor ion spectrum as the light form and in the same product ion spectrum.

peptide intensities and subsequent quantification procedures by the program took about 5 min for each LC-MS^E data set. The flowchart of our data analysis procedure is described in Figure 1B. The algorithm for SIL-peptide pair detection is based on the difference of mass accuracy (<5 ppm), elution time (<0.05 min), and ion mobility ($\leq$0.5) between SIL heavy and light precursors. Intensities of the SIL-peptide precursors were summed and used for calculating the relative abundance of every identified peptide and protein. Figure 2 shows a typical LC-MS^E analysis of complex proteome mixtures. The base peak chromatogram of the first fraction of H/L = 1:10 proteome mixture is shown in Figure 2A. In this chromatogram, the peak height represents the total intensity of each unique precursor ion including all charge states and associated isotopes. A precursor with $[M + H]^+$ = 1931.1729 and at retention time (RT) = 66.16 min was subjected to the MS fragmentation at elevated energy mode. The MS^E fragmentation spectrum is illustrated in Figure 2B, from which it was identified as the peptide "VTIAQGGVLPNIQAVLLPK" derived from histone H2A type 1-A (H2A1A_HUMAN). The SIL-heavy form of the light precursor was detected with $[M + H]^+$ = 1939.1863 at RT = 66.15 min. Intensities for the heavy and light precursors were 31 120 and 345 605 counts, respectively (H/L ratio = 0.090, Figure 2C). Furthermore, this abundance relationship of heavy and light species could be obtained from stable isotope-containing y-ions. As shown in Figure 2D, the y13-ions were observed as pairs in the MS^E spectrum. Intensities of heavy and light forms of the y13 product ions were 222 and 2415 counts (H/L ratio = 0.092), respectively.

Identification and Quantification of Proteins from SIL-Labeled Proteome Mixtures with Known Ratios. Since the specificity of identification relies highly on the mass precision of detected precursor/product ions, the accuracy of LC-MS^E determination was examined. The binned frequency of mass error (ppm) for all identified precursor ions and product ions was determined (Supplemental Figure 1, Supporting Information). A mass error of less than $\pm$3 ppm was found for 69.9% of the precursor ions, and a mass error of less than $\pm$10 ppm was found for 92.0% of the precursor ions. The mass error for 60.2% of the product ions was less than $\pm$3 ppm, and for 91.9% of the product ions, the mass error was less than $\pm$10 ppm. Both of these error profiles were normally distributed, and the standard deviations are 5.16 and 5.28 for the precursors and product ions, respectively.

To examine the specificity of the extracted SIL pairs, the binned frequencies of the differences in mass error, retention time, and ion drift for the SIL-precursor ions were plotted. The mass error for more than 92.3% of the SIL-precursor pairs was less than $\pm$5 ppm (Supplemental Figure 2A, Supporting Information). The standard deviation of the mass error distribution was 2.90, which was less variant (σ^2) compared to the total

dx.doi.org/10.1021/ac201555m |*Anal. Chem.* 2011, 83, 6971–6979

Table 1. Summary of the Identified and Quantified Proteins/Peptides from the Three Complex Proteome Mixtures with Known H/L Ratios Using SIL-Based Quantitation

	Protein Identification		Protein Quantitation			
			Proteins		Peptides	
	Proteins	Peptides	Number	Percentage[a] (%)	Number	Percentage[b] (%)
H/L = 1:1	741	7588	593	80.03	3125	41.18
H/L = 1:5	1007	10713	536	53.23	1914	17.87
H/L = 1:10	921	9838	290	31.49	912	9.27
Overlapped[c] in H/L = 1:1 and 1:5	602	4157	397	65.95	1505	36.20
Overlapped in H/L = 1:1 and 1:10	577	3863	239	41.42	784	20.30
Overlapped in H/L = 1:5 and 1:10	721	5238	262	36.34	813	15.52
Overlapped in all three mixtures	553	3385	228	41.23	734	21.68

[a] Percentage is the number of proteins in SIL-based quantitation divided by the number of identified proteins. [b] Percentage is the number of peptides in SIL-based quantitation divided by the number of identified peptides. [c] Overlapped coverage of identified and quantified proteins/peptides is listed accordingly.

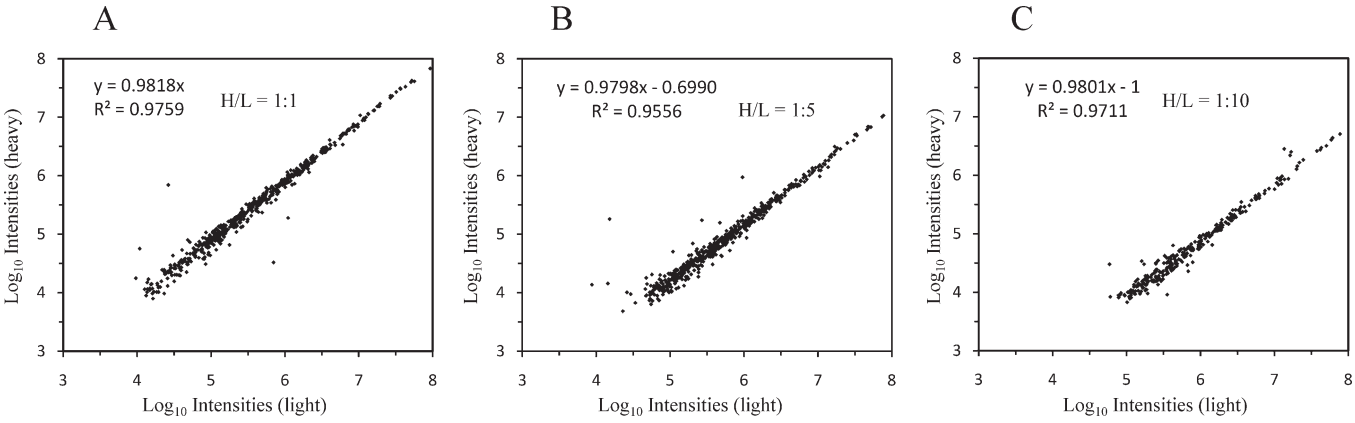


Figure 3. Scatter plots of the $\log_{10}$ intensities of proteins for SIL heavy versus SIL light in the (A) H/L = 1:1, (B) H/L = 1:5, and (C) H/L = 1:10 data. A linear regression line was generated to fit each scatter plot. For the fit, the *y*-axis intercept was set to the true value of H/L log ratio in each proteome mixture. The slope and correlation (R^2) are indicated for each scatter plot.

identified precursors. The frequency of differences in retention time (Supplemental Figure 2B, Supporting Information) and ion drift (Supplemental Figure 2C, Supporting Information) between the SIL-light and -heavy precursor ions were examined. Retention time and ion drift of the coeluting SIL-precursors were highly consistent. The difference in retention time for more than 97.8% of the precursor pairs was less than ±0.05 min, and the difference in ion drift for more than 96.4% of the precursor pairs was less than ±0.5. Interestingly, the difference in ion drift for the SIL-precursors was not normally distributed. The median difference of the ion drift is 0.1. Because separation of precursor ions by ion mobility was based on internal attributes such as their mass, charge, size, and shape, the higher ion drift value for SIL-heavy precursor ions is presumably due to the larger molecular weight compared to the SIL-light precursor ions.

The number of identified peptides and proteins for the three proteome mixtures with known H/L ratios (1:1, 1:5, and 1:10) is summarized in Table 1. A total of 1332 different proteins were identified in all three mixtures, of which 41.5% (553/1332) of proteins were identified universally in every mixture. Averaged ratios of identified peptides per protein were 10.24, 10.63, and 10.68, in the H/L = 1:1, 1:5, and 1:10 mixtures, respectively. This demonstrated a relatively high degree of sequence coverage by the LC-MS^E analysis.

For the SIL-based quantitation, the $\log_{10}$-intensities of heavy versus light proteins in each proteome mixture are scatter plotted in Figure 3. The dynamic range for protein intensities in SIL-based quantitation varied over 4 orders of magnitude for the H/L = 1:1 data and varied over 3 orders of magnitude for the H/L = 1:5 and 1:10 data. Correlations between the heavy versus light proteins were 0.9759, 0.9556, and 0.9711, respectively, in the H/L = 1:1, 1:5, and 1:10 data. Coefficients of regression were 0.9818, 0.9798, and 0.9801, for the three mixtures. A high correlation of the intensities of light and heavy proteins indicates a high consistency of quantitation for the proteins in each proteome mixture. A coefficient of regression close to one indicates that the relative abundance for each peptide was unbiased over different ranges of peptide intensities.

Comparison of SYNAPT G2MS and LTQ-Orbitrap MS for SIL-Based Quantitative Proteomics. Previously, we developed UNiquant software for analyzing the SILAC data obtained with the LTQ-Orbitrap MS with DDA.[29] Here, we compared the accuracy of the peptide quantitation results obtained on the SYNAPT G2MS to the one generated on the LTQ-Orbitrap MS as previously described.[29] As shown in Figure 4, the identified and quantified peptides from all three proteome mixtures with known ratios (H/L = 1:1, 1:5, and 1:10) on both instruments

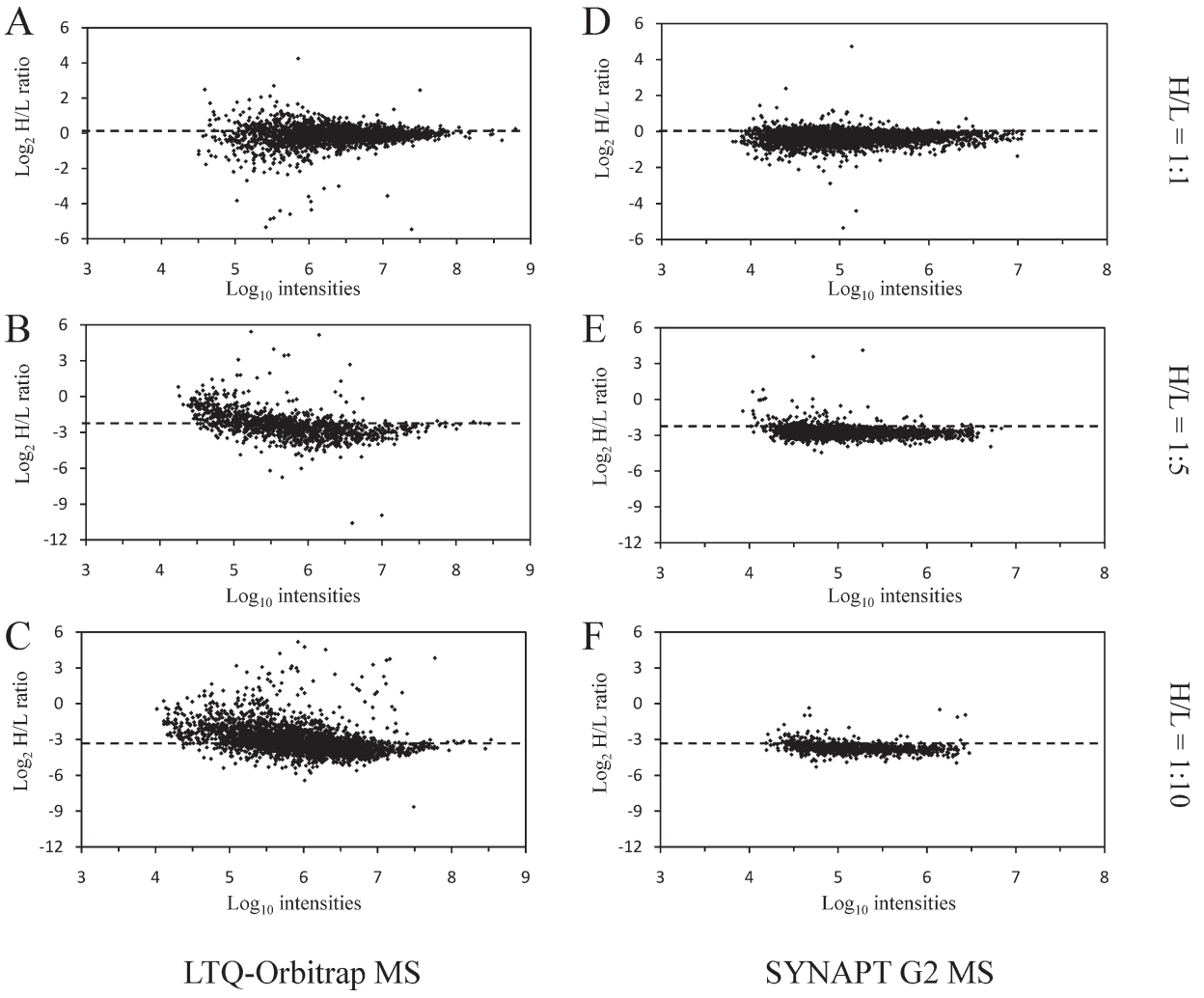


Figure 4. Accuracy of quantitation for SILAC data analysis is compared between the (A−C) LTQ-Orbitrap platform and the (D−F) SYNAPT G2 platform for three proteome mixtures with H/L = 1:1 (A, D), 1:5 (B, E), and 1:10 (C, F). In each scatter plot, the quantified peptides were distributed by their $\log_2$ (H/L) intensity ratios versus $\log_{10}$ (H × L) intensity products. The true log (H/L) ratio is indicated as a dashed line for the H/L = 1:1 ($\log_2$ ratio = 0), 1:5 ($\log_2$ ratio = −2.32), and 1:10 ($\log_2$ ratio = −3.32) mixtures, respectively.

were plotted by their $\log_2$ (H/L) intensity ratios versus the $\log_{10}$ (H × L) intensity products (Figure 4). In the LTQ-Orbitrap analysis, $\log_2$ (H/L) ratios of quantified peptides show a comet-like distribution from the three mixtures (Figure 4A−C). In H/L = 1:5 and 1:10 mixtures, the data in the low intensity region tends to a log ratio of 0 whereas they should have a value of −2.32 for the 1:5 data and a value of −3.32 for the 1:10 data (Figure 4B,C). In the SYNAPT G2 analysis, $\log_2$ (H/L) ratios of quantified peptides show a more uniform distribution for each of the three mixtures (Figure 4D−F). In the H/L = 1:5 and 1:10 mixtures, the $\log_2$ (H/L) ratios are closer to the expected ratios (−2.32 and −3.32). In the H/L 1:1 mixtures, the dynamic ranges ($\log_{10}$ intensities) of both the LTQ-Orbitrap data and the SYNAPT G2 data are about 4 orders of magnitude (Figure 4A,D). In H/L = 1:5 and 1:10 mixtures, the dynamic range of LTQ-Orbitrap data is still 4 orders of magnitude (Figure 4B,C), whereas the range of the data from the SYNAPT G2 drops to 3.5 orders of magnitude in the H/L = 1:5 mixture and to 3.0 in the 1:10 mixture (Figure 4E,F).

Identification and Quantification of Proteins from SIL-Labeled Proteome Mixtures with Unknown Ratios. As a real biological application to evaluate the performance of our software for quantitative proteomics using SIL and LC-MS^E technology, we investigated the proteome difference between MEFs and MEF-derived iPSCs using trypsin-catalyzed $^{16}O/^{18}O$ labeling. We identified 756, 762, and 696 proteins from 5905, 5630, and 5119 unique peptides in three replicates. Among the identified proteins, 77.1% (583/756), 75.6% (576/762), and 75.3% (524/696) were quantifiable in the three replicates. Distribution of the quantified protein $\log_2$ ratios versus protein intensities in the MEF/iPSC data sets is illustrated in Figure 5A. Ten proteins were significantly up-regulated in iPSCs (Z-test, $P < 0.05$, Figure 5A). Up-regulation of selected proteins could be rationalized as follows: Fubp1 (Far upstream element-binding protein 1) activates c-Myc which is one of the four factors used for iPSC formation and it stimulates expression of c-Myc in undifferentiated cells.[35] Crabp1 (cellular retinoic acid binding protein 1) plays an important role in retinoic acid-mediated differentiation and proliferation processes.[36] It might be a direct transcriptional target of c-Myc.[37] Sumo2 and Sumo3 are two proteins with high degree of identity, which are essential for protein SUMO modification. Previous study revealed that sumoylation

dx.doi.org/10.1021/ac201555m |*Anal. Chem.* 2011, 83, 6971–6979

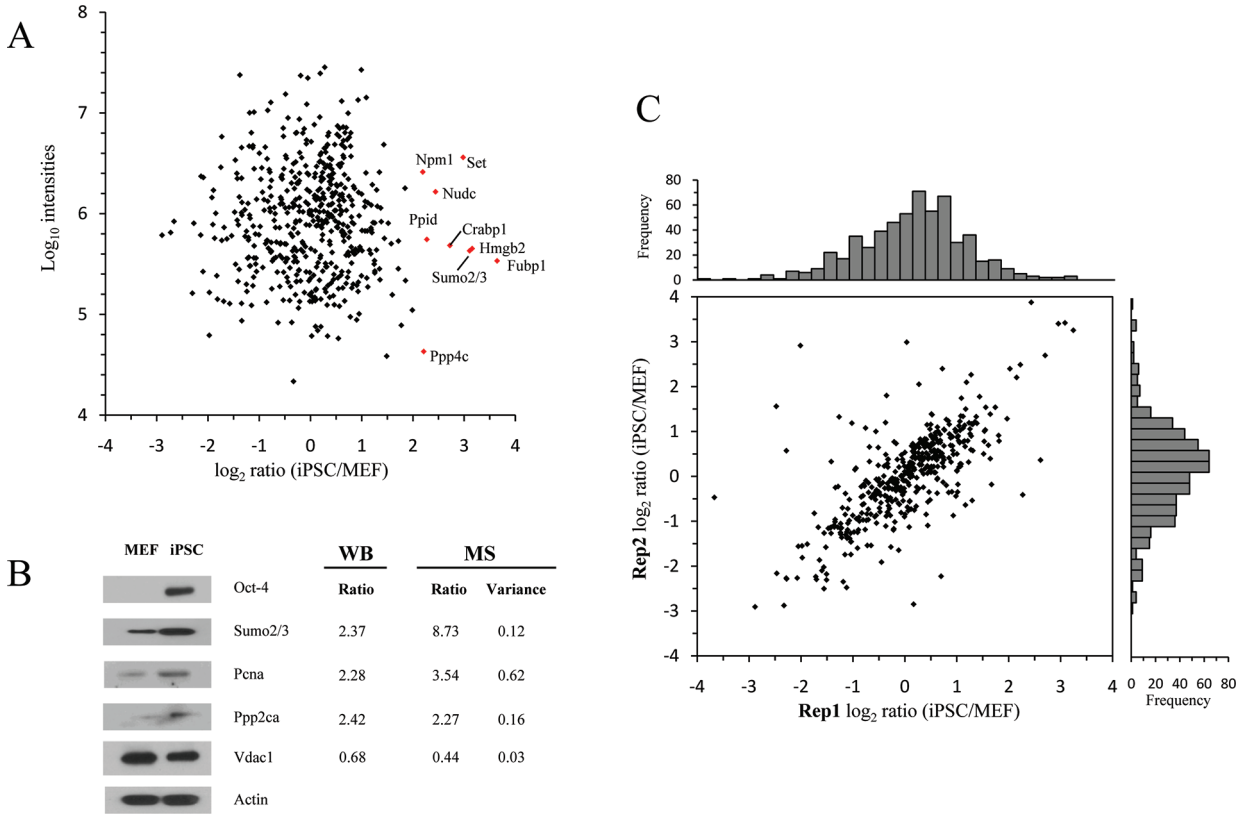


Figure 5. Distribution and correlation of the quantified protein ratios in the MEF/iPSC data sets. (A) Distribution of the quantified protein $\log_2$ ratios (iPSC/MEF) versus protein intensities. Significantly up-regulated proteins ($P < 0.05$, Z-test) are indicated as red points with their names. (B) The relative abundances of proteins quantified by LC-MS^E analysis were validated by Western blotting for the proteins: Sumo2/3, Pcna, Ppp2ca, and Vdac1. Oct-4a is the positive control and Actin is the loading control. Quantified ratios for given proteins by Western blot (WB) are calculated as the difference in OD intensity measured by ImageJ software. Quantified ratios by mass spectrometry (MS) are the averaged relative intensity (iPSC/MEF) in three replicates, and variances are the value of the standard error divided by the mean ratio for each protein. (C) Protein $\log_2$ ratios quantified for two replicates are plotted. Frequencies of the protein $\log_2$ ratios were normally distributed. Ratios quantified in these two replicates were significantly correlated ($P < 0.0001$, Pearson test, $R^2 = 0.5011$).

of Klf4, one of the four factors used for iPSC formation, is essential for its transcriptional activation.[38] Hmgb2 (high mobility group box 2), Nudc (nuclear distribution gene C), and Npm1 (nucleophosmin 1) are important nuclear chaperones. These three proteins may facilitate the formation of DNA circles, nuclear migration, and chromatin remodeling, respectively, during the DNA duplication and reprogramming processes.[39,40] Western blotting was used to confirm the relative change of protein abundance measured by the LC-MS^E method and our software (Figure 5B). Several proteins with different abundance, expression change, and cellular location were selected for validation, including Ppp2ca and Sumo2/3 (cytoplasmic proteins), Vdac1 (mitochondria membrane protein), and Pcna (nucleic protein). Overall, the protein expression changes detected by Western blotting showed good correlation with the LC-MS^E method. The correlation coefficient (R^2) of the $\log_2$ protein ratios (iPSC/MEF) for replicate 1 and replicate 2 LC-MS^E analyses is 0.5011 ($P < 0.0001$, Pearson test, Figure 5C). Similar results were obtained for all other pairings of the three replicates. This demonstrates the reliable proteome quantitation achieved by the whole analytical platform, including our new software.

DISCUSSION

In the current work, we have developed a new program for analyzing SIL data obtained with the LC-MS^E technology for quantitative proteomics studies. The algorithm we developed for isotopic peptide pair detection is based on the differences of accurate mass, elution time, and ion mobility between light and heavy SIL-peptides, which is similar to what we previously described in the UNiquant software with the addition of ion drift.[29] Three complex SILAC proteome mixtures (H/L = 1:1, 1:5, and 1:10) from MDA-MB-231 cell digests and an $^{16}O/^{18}O$-labeled mixture of MEF and MEF-derived iPSC proteomes were used to illustrate the performance of our software for quantitative proteomics analysis using SIL and the LC-MS^E technology.

The LTQ-Fourier transform (FT)/Orbitrap MS with DDA is the major MS platform for SIL-based quantitative proteomic applications today. With this platform, the MS scans are used for peptide quantitation, while MS/MS scans are used for peptide identification. However, a MS scientist has to balance peptide/protein identification, dynamic range, and the precision of protein quantification by carefully selecting the parameters for DDA. It is not uncommon for peptides with low intensities to be detected in the MS mode but to go undetected in MS/MS or vice versa.[17] Also, because the number of precursor ions selected for MS/MS fragmentation is fixed in the DDA mode, the total number of peptides identified for a given protein from complex proteome mixtures is relatively low due to the limited number of MS/MS spectra that can be generated for peptide identification. The DIA approach is an interesting alternative to complement

DDA for SIL-based quantitative proteomic analysis. First, it provides more time for MS/MS fragmentation, making it possible to identify more peptides. Indeed, in this study, we identified an average of more than 10 peptides for each protein using the DIA method. Such high sequence coverage benefits not only protein identification but also the accuracy of protein quantitation. Second, the high resolution MS and MS/MS data makes quantitation possible from both precursor and product ions. We envision that quantitation based on the intensities of both precursor and product ions will further improve the dynamic range and quantitation accuracy for SIL-based quantitative proteomic analysis.

Dynamic range for protein quantitation is one of key features of quantitative proteomics analysis. Using the DIA method, we identified more proteins in the 1:5 and 1:10 mixtures compared to the 1:1 mixture; however, the number of SIL-peptide pairs and the dynamic range of protein intensities decreased in the 1:5 and 1:10 mixtures. The decrease in peptide pairs and dynamic range of protein intensities is mainly due to the loss of low-intensity heavy peptides and to saturation of high-intensity light peptides. That occurred because the peak detection algorithm used a cutoff for acceptable peaks that was based on peak intensity and signal-to-noise ratio. These excluded small peaks, which impacted the number of heavy peptides observed, but ensured that the selected peaks were actually peptides. However, in the DDA analysis of the three complex proteome mixture with known ratios (H/L = 1:1, 1:5, 1:10), the number of peptide pairs did not change significantly (541, 502, and 677 peptide pairs detected for H/L = 1:1, 1:5, 1:10, respectively).[29] Additionally, the dynamic range of protein intensities was consistent (about 4 orders of magnitude) in the DDA analysis of the three complex proteome mixtures with different ratios.

Accurate quantitation of protein abundance is an essential task for MS instruments and its associated data analysis tools. Overall, the SYNAPT G2 with DIA approach showed better quantitation accuracy and reliability than the LTQ-Orbitrap with DDA analysis presumably due to the fundamental difference between these two mass analyzers.[28,41] In a TOF analyzer such as the SYNAPT G2, the signal intensity comes from direct ion counting and many spectra are accumulated up to 10 000 specs per second. Each TOF spectrum usually has a small dynamic range, and a collection of multiple spectra can increase it. If the TOF analyzer is optimized for high sensitivity such as in the case of this study, the SYNAPT G2 instrument gives correct intensity measurements for low intensity ions but saturated readings for high intensity ions. Thus, the very high-intensity ions are discriminated against in the final results because the saturated ions increase internal errors of both measured intensity and mass accuracy. In the Orbitrap analyzer, signal intensities are obtained by Fourier transformation of an ion signal induced on the detection electrodes of the Orbitrap cell.[42] Just one ion signal spectrum is sufficient to obtain a full m/z spectrum with a high dynamic range in ion intensities. In addition to these signal detection differences, the front part ion optics for the two instruments is distinct. The SYNAPT G2 uses a stack ring ion guide while the LTQ-Orbitrap uses a linear quadrupole. These differences produce different ion intensity scales. Together, these differences may explain the difference in the quantification results obtained by the SYNAPT G2 and LTQ-Orbitrap platforms.

For achieving highly accurate quantitation results, we suggest that an instrument dependent, postmeasurement normalization approach is needed for both MS platforms we discussed herein. Specifically, in LTQ-Orbitrap analysis, a nonlinear normalization method such as the locally weighted scatterplot smoothing (Lowess) method is recommended.[43] Supplemental Figure 3 (Supporting Information) shows the distribution of the quantified peptide ratios by LTQ-Orbitrap before and after Lowess normalization. While for the SYNAPT G2 data, a global median-center method might be sufficient.

CONCLUSIONS

In summary, we have developed a new program addition to our UNiquant software pipeline for SIL-based quantitative proteomics using the DIA approach. The LC-MS^E technology, coupled with our new program addition to our UNiquant software pipeline, yielded high quantitation accuracy in the analysis of complex proteome mixtures and is a viable alternative for SIL-based quantitative proteomics applications. Our study herein also provides valuable insight into the performance of the two cutting-edge MS platforms (LTQ-Orbitrap & SYNAPT G2) for SIL-based quantitative proteomic analysis today.

ASSOCIATED CONTENT

Supporting Information. Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Address: Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, NE 68198-5900. Phone: +1-402-559-4183. Fax: +1-402-559-4651. E-mail: dings@unmc.edu.

Notes

The authors have declared no conflict of interest.

ACKNOWLEDGMENT

We thank Dr. Lawrence Schopfer for the editing of this manuscript, Dr. Aleksey Tolmachev for the helpful discussion of the results, and Dr. Stephen Slahck for the arrangement of the demo experiments on SYNAPT G2MS. This work was financially supported by NEHHS LB606 (S.J.D.), and X.H. was supported by a scholarship from China Scholarship Council.

REFERENCES

(1) Nilsson, T.; Mann, M.; Aebersold, R.; Yates, J. R., 3rd; Bairoch, A.; Bergeron, J. J. *Nat. Methods* **2010**, *7*, 681–685.
(2) Mallick, P.; Kuster, B. *Nat. Biotechnol.* **2010**, *28*, 695–709.
(3) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198–207.
(4) Gstaiger, M.; Aebersold, R. *Nat. Rev. Genet.* **2009**, *10*, 617–627.
(5) Choudhary, C.; Mann, M. *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 427–439.
(6) Mann, M. *Nat. Biotechnol.* **1999**, *17*, 954–955.
(7) Wilm, M. *Proteomics* **2009**, *9*, 4590–4605.
(8) Afkarian, M.; Bhasin, M.; Dillon, S. T.; Guerrero, M. C.; Nelson, R. G.; Knowler, W. C.; Thadhani, R.; Libermann, T. A. *Mol. Cell. Proteomics* **2010**, *9*, 2195–2204.
(9) Mann, M. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 952–958.
(10) Pan, S.; Aebersold, R. *Methods Mol. Biol.* **2007**, *367*, 209–218.
(11) Tao, W. A.; Aebersold, R. *Curr. Opin. Biotechnol.* **2003**, *14*, 110–118.
(12) Qian, W. J.; Petritis, B. O.; Kaushal, A.; Finnerty, C. C.; Jeschke, M. G.; Monroe, M. E.; Moore, R. J.; Schepmoes, A. A.; Xiao, W.;

Moldawer, L. L.; Davis, R. W.; Tompkins, R. G.; Herndon, D. N.; Camp, D. G.; Smith, R. D. *J. Proteome Res.* **2010**, *9*, 4779–4789.
(13) Domon, B.; Aebersold, R. *Nat. Biotechnol.* **2010**, *28*, 710–721.
(14) Mortz, E.; O'Connor, P. B.; Roepstorff, P.; Kelleher, N. L.; Wood, T. D.; McLafferty, F. W.; Mann, M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8264–8267.
(15) Jaffe, J. D.; Mani, D. R.; Leptos, K. C.; Church, G. M.; Gillette, M. A.; Carr, S. A. *Mol. Cell. Proteomics* **2006**, *5*, 1927–1941.
(16) Li, G. Z.; Vissers, J. P.; Silva, J. C.; Golick, D.; Gorenstein, M. V.; Geromanos, S. J. *Proteomics* **2009**, *9*, 1696–1719.
(17) Venable, J. D.; Dong, M. Q.; Wohlschlegel, J.; Dillin, A.; Yates, J. R. *Nat. Methods* **2004**, *1*, 39–45.
(18) Purvine, S.; Eppel, J. T.; Yi, E. C.; Goodlett, D. R. *Proteomics* **2003**, *3*, 847–850.
(19) Ramos, A. A.; Yang, H.; Rosen, L. E.; Yao, X. *Anal. Chem.* **2006**, *78*, 6391–6397.
(20) Williams, J. D.; Flanagan, M.; Lopez, L.; Fischer, S.; Miller, L. A. *J. Chromatogr., A* **2003**, *1020*, 11–26.
(21) Geiger, T.; Cox, J.; Mann, M. *Mol. Cell. Proteomics* **2010**, *9*, 2252–2261.
(22) Silva, J. C.; Gorenstein, M. V.; Li, G. Z.; Vissers, J. P.; Geromanos, S. J. *Mol. Cell. Proteomics* **2006**, *5*, 144–156.
(23) Hughes, M. A.; Silva, J. C.; Geromanos, S. J.; Townsend, C. A. *J. Proteome Res.* **2006**, *5*, 54–63.
(24) Vissers, J. P.; Pons, S.; Hulin, A.; Tissier, R.; Berdeaux, A.; Connolly, J. B.; Langridge, J. I.; Geromanos, S. J.; Ghaleb, B. *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.* **2009**, *877*, 1317–1326.
(25) Li, X. J.; Zhang, H.; Ranish, J. A.; Aebersold, R. *Anal. Chem.* **2003**, *75*, 6648–6657.
(26) Park, S. K.; Venable, J. D.; Xu, T.; Yates, J. R., 3rd *Nat. Methods* **2008**, *5*, 319–322.
(27) Cox, J.; Mann, M. *Nat. Biotechnol.* **2008**, *26*, 1367–1372.
(28) Bakalarski, C. E.; Elias, J. E.; Villen, J.; Haas, W.; Gerber, S. A.; Everley, P. A.; Gygi, S. P. *J. Proteome Res.* **2008**, *7*, 4756–4765.
(29) Huang, X.; Tolmachev, A. V.; Shen, Y.; Liu, M.; Huang, L.; Zhang, Z.; Anderson, G. A.; Smith, R. D.; Chan, W. C.; Hinrichs, S. H.; Fu, K.; Ding, S. J. *J. Proteome Res.* **2011**, *10*, 1228–1237.
(30) Kao, W. C.; Chen, Y. R.; Yi, E. C.; Lee, H.; Tian, Q.; Wu, K. M.; Tsai, S. F.; Yu, S. S.; Chen, Y. J.; Aebersold, R.; Chan, S. I. *J. Biol. Chem.* **2004**, *279*, 51554–51560.
(31) Petritis, B. O.; Qian, W. J.; Camp, D. G., 2nd; Smith, R. D. *J. Proteome Res.* **2009**, *8*, 2157–2163.
(32) Gilar, M.; Olivova, P.; Chakraborty, A. B.; Jaworski, A.; Geromanos, S. J.; Gebler, J. C. *Electrophoresis* **2009**, *30*, 1157–1167.
(33) Silva, J. C.; Denny, R.; Dorschel, C. A.; Gorenstein, M.; Kass, I. J.; Li, G. Z.; McKenna, T.; Nold, M. J.; Richardson, K.; Young, P.; Geromanos, S. *Anal. Chem.* **2005**, *77*, 2187–2200.
(34) Hou, J.; Dong, J.; Sun, L.; Geng, L.; Wang, J.; Zheng, J.; Li, Y.; Bridge, J.; Hinrichs, S. H.; Ding, S. J. *J. Transl. Med.* **2011**, *9*, 64.
(35) Avigan, M. I.; Strober, B.; Levens, D. *J. Biol. Chem.* **1990**, *265*, 18538–18545.
(36) Chen, Y.; Reese, D. H. *J. Cell. Biochem.* **2011**, DOI: 10.1002/jcb.23200.
(37) Gupta, N.; Wollscheid, B.; Watts, J. D.; Scheer, B.; Aebersold, R.; DeFranco, A. L. *Nat. Immunol.* **2006**, *7*, 625–633.
(38) Du, J. X.; McConnell, B. B.; Yang, V. W. *J. Biol. Chem.* **2010**, *285*, 28298–28308.
(39) Frehlick, L. J.; Eirin-Lopez, J. M.; Ausio, J. *Bioessays* **2007**, *29*, 49–59.
(40) Morris, S. M.; Albrecht, U.; Reiner, O.; Eichele, G.; Yu-Lee, L. Y. *Curr. Biol.* **1998**, *8*, 603–606.
(41) Pan, C.; Kora, G.; McDonald, W. H.; Tabb, D. L.; VerBerkmoes, N. C.; Hurst, G. B.; Pelletier, D. A.; Samatova, N. F.; Hettich, R. L. *Anal. Chem.* **2006**, *78*, 7121–7131.
(42) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R. *J. Mass Spectrom.* **2005**, *40*, 430–443.
(43) Callister, S. J.; Barry, R. C.; Adkins, J. N.; Johnson, E. T.; Qian, W. J.; Webb-Robertson, B. J.; Smith, R. D.; Lipton, M. S. *J. Proteome Res.* **2006**, *5*, 277–286.