analytical chemistry

pubs.acs.org/ac Article

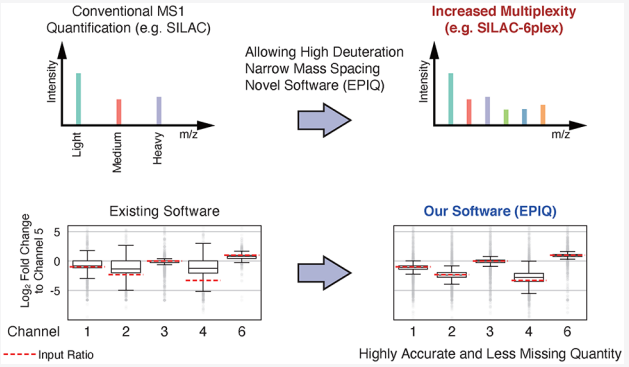
MS1-Level Proteome Quantification Platform Allowing Maximally Increased Multiplexity for SILAC and *In Vitro* Chemical Labeling

Yeon Choi,[⊥] Kyowon Jeong,[⊥] Sanghee Shin,[⊥] Joon Won Lee,[⊥] Young-suk Lee, Sangtae Kim, Sun Ah Kim, Jaehun Jung, Kwang Pyo Kim, V. Narry Kim,* and Jong-Seo Kim*

Cite This: *Anal. Chem.* 2020, 92, 4980−4989 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Quantitative proteomic platforms based on precursor intensity in mass spectrometry (MS1-level) uniquely support *in vivo* metabolic labeling with superior quantification accuracy but suffer from limited multiplexity (≤3-plex) and frequent missing quantities. Here we present a new MS1-level quantification platform that allows maximal multiplexing with high quantification accuracy and precision for the given labeling scheme. The platform currently comprises 6-plex *in vivo* SILAC or *in vitro* diethylation labeling with a dedicated algorithm and is also expandable to higher multiplexity (e.g., nine-plex for SILAC). For complex samples with broad dynamic ranges such as total cell lysates, our platform performs highly accurately and free of missing quantities. Furthermore, we successfully applied our method to measure protein synthesis rate under heat shock response in human cells by 6-plex pulsed SILAC experiments, demonstrating the unique biological merits of our *in vivo* platform to disclose translational regulations for cellular response to stress.



Among proteome quantification strategies based on liquid chromatography tandem mass spectrometry (LC-MS/MS), the precursor intensity (MS1)-based approach is a highly accurate and reliable method that allows *in vivo* metabolic labeling such as stable isotope labeling with amino acids in cell culture (SILAC).[1,2] However, the multiplexity of existing MS1-level quantification methods (e.g., triple SILAC)[3,4] does not meet the increasing demand for highly multiplexed experimental design.[5] High multiplexity can be achieved by the use of deuterium, which allows for the largest number of heavy isotope incorporation on labels, and narrow mass spacing (≤2 Da) between labeling channels (Figure S1A). However, deuterated labels can cause inconsistent retention time (termed "RT shift"), and the narrow mass spacing results in severe isotope cluster overlapping between channel signals.[6−8] These complications substantially hinder the exact determination of interchannel quantity ratios, as illustrated in Figure S1B-E. Thus, despite several attempts[9,10] to address the issues, MS1-level quantification is largely limited to a multiplexity of only three, keeping the conventional terms for labeling scheme (≥4 Da of mass spacing and minimum use of deuteriums).[11,12]

In addition to limited multiplexity, MS1-level quantification methods suffer from frequent missing quantities,[13,14] especially when the peptides are low abundant. Existing computational tools often fail to detect signals from such peptides. Signal extraction failure leads to missing quantities that are hard to recover by general imputation methods.

Here, we introduce a MS1-level proteome quantification platform comprising metabolic (*in vivo*)/chemical (*in vitro*) isotopic labeling and a dedicated algorithm called Epic Protein Integrative Quantification (EPIQ), enabling maximally increased multiplexity for any isotopic labeling scheme. To increase the multiplexity to the maximum level for a given chemical structure of labels, we allowed heavily differential deuteration and narrow mass spacing (2 Da) for both metabolic and chemical isotopic labeling. The necessarily accompanied complications such as retention time shift and isotopic overlapping are computationally addressed by EPIQ, achieving accurate quantification with few missing quantities.

Through the benchmark tests using HeLa lysates labeled with metabolic SILAC-6plex labeling or chemical dimethylation (DM)-5plex[9]/diethylation[7,12,15] (DE)-6plex labelings designed under the aforementioned labeling terms, we show that EPIQ exclusively achieves protein quantification over existing tools. Furthermore, we demonstrate that EPIQ together with our 6-plex labeling schemes enables sensitive

Received: November 12, 2019
Accepted: March 11, 2020
Published: March 13, 2020

https://dx.doi.org/10.1021/acs.analchem.9b05148
Anal. Chem. 2020, 92, 4980−4989

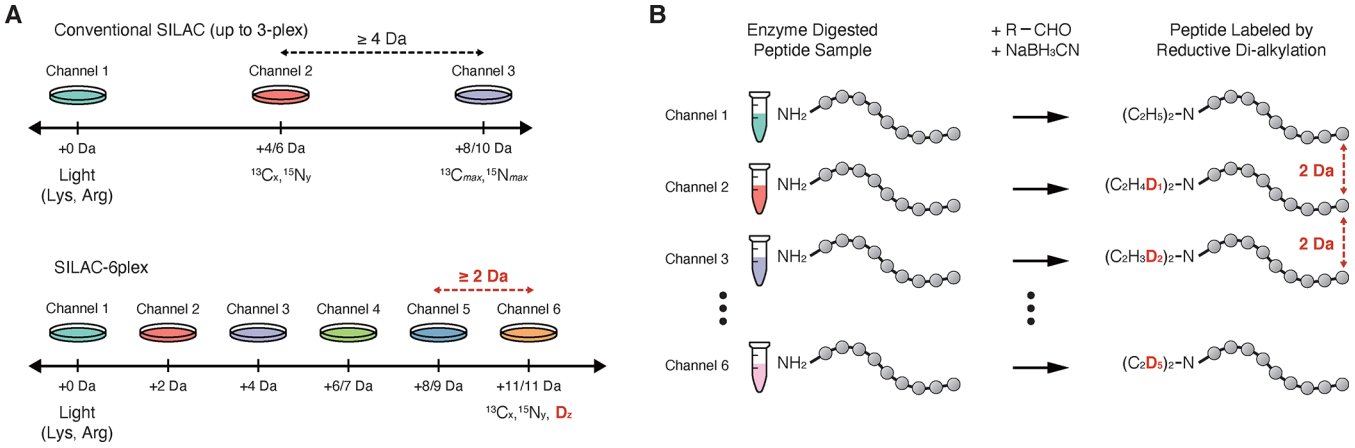


Figure 1. New *in vivo* and *in vitro* labeling schemes for maximally multiplexed MS1-level proteome quantification platform. (A) Schematic view of conventional triple SILAC (top) and SILAC-6plex (bottom) labeling experiments. (B) Schematic view of diethylation-6plex labeling experiments. For both multiplexed labeling schemes, a narrow mass spacing (2 Da) and heavily differential deuteration should be allowed to increase multiplexity to the maximum for the given labeling scheme.

and specific detection of differentially expressed proteins (DEPs). Lastly, by applying SILAC-6plex and EPIQ, we performed time-resolved profiling of global protein abundance and synthesis rate in HeLa cells undergoing the heat shock response, which is not fully understood at the system-wide level despite its importance in protein homeostasis. Our results clearly demonstrate that the benefit of increased multiplexity in SILAC for time-resolved profiling of cellular events and provide a resource for the system-wide understanding of heat shock response.

EXPERIMENTAL SECTION

Six-Plexed SILAC Media Preparation and HeLa Cell Culture. For six-plexed SILAC, lysine and arginine deficient DMEM (Dulbecco's modified Eagle's medium, Thermo Fisher) was used with supplement of isotopically distinct L-lysine:2HCl (K, 0.80 mM) and L-arginine:HCl (R, 0.40 mM) for 6 individual SILAC channels (Figure S2). The more detail ingredients and media preparation of each SILAC channel are described in the Supporting Information. HeLa cell line from ATCC was cultured without antibiotics in the customized DMEM media to >5 cell cycle passages. The normal DMEM (Welgene) containing 10% FBS was used for HeLa cell culture as well, which cells were used for other labeling experiments.

Peptide Sample Preparation. Harvested HeLa cells were lysed using 8 M urea in 50 mM ammonium bicarbonate (ABC) containing protease inhibitors (Pierce). The denatured samples in 8 M urea were reduced and alkylated with 10 mM dithiothreitol and 40 mM iodoacetamide, respectively. The samples were diluted with 50 mM ABC to be <1 M of urea for digestion by 2% (w/w) of trypsin (MS grade, Pierce) at 37 °C for overnight. The digested peptides were desalted using C18 SPE cartridge prior to further analysis.

Six-Plexed Diethylation of Tryptic Peptides. Desalted tryptic peptide samples were subject to reductive diethylation using acetaldehyde and sodium cyanoborohydride isotopologues (details in Supporting Information). The labeled samples were mixed with various desired input ratios and then desalted using C18 SPE cartridge for LC-MS/MS analysis.

Prefractionation and LC-MS/MS Analysis of Labeled HeLa Samples. Multidimensional LC-MS/MS analysis based on strong-cation exchange (SCX) fractionation and concatenated HILIC fractionation was carried out for DE-6plex and SILAC-6plex sample, respectively. LC-MS/MS analysis of the fractionated samples was carried out by Orbitrap Fusion Lumos Tribrid or Q-Exactive mass spectrometer coupled with ultimate 3000 RSLCnano liquid chromatography.

Protein Identification and Quantification. Generated raw files are searched against UniProt reference proteome UP000005640 (last modified on March 15, 2019) by EPIQ, MaxQuant (ver 1.5.3.30) and Proteome Discoverer (ver. 2.2.0.388). More detailed information on each search engine parameters can be found in the Supporting Information.

Heat Shock Experimental Procedures. For steady-state SILAC-6plex, fully labeled HeLa cells with six channels were incubated at 43 °C for the desired time. For pulsed SILAC experiment, culture media of unlabeled HeLa cells were exchanged with heavy channel SILAC media as described in the Supporting Information.

RESULTS

Maximally Multiplexed MS1-level *in Vivo* and *in Vitro* Labeling Schemes. To enable maximally multiplexed *in vivo* metabolic labeling using lysine and arginine amino acids, we designed a SILAC-6plex labeling scheme (expandable to 9plex; Figure S2E-D), by allowing narrow mass spacing (≥2 Da) and highly differential deuterium counts between channels (Figures 1A and S2). We also designed two *in vitro* chemical labeling schemes, DE-6plex (Figures 1B and S3) and DM-5plex[9] (Figure S4), by allowing the same mass spacing and differential deuterium counts on reductive dialkylation labeling.[12,16] To test the feasibility of those labeling schemes, we analyzed HeLa cell lysates labeled by either SILAC-6plex or DE-6plex by LC-MS/MS. After incubating HeLa cell with SILAC-6plex media for five doubling period, nearly all (~99%) peptides were fully labeled *in vivo* by SILAC-6plex arginine or lysine (Figure S2B,C). *In vitro* DE-6plex labeling was also successfully done; ~98% of all tryptic peptides were fully labeled in all channels (Figure S3B).

Our design of the labeling scheme was successful and validated. However, as expected, the data sets had signal complications from narrow mass spacing and RT shifts (Figure S5), which would make it impossible to analyze the data using

https://dx.doi.org/10.1021/acs.analchem.9b05148
Anal. Chem. 2020, 92, 4980−4989

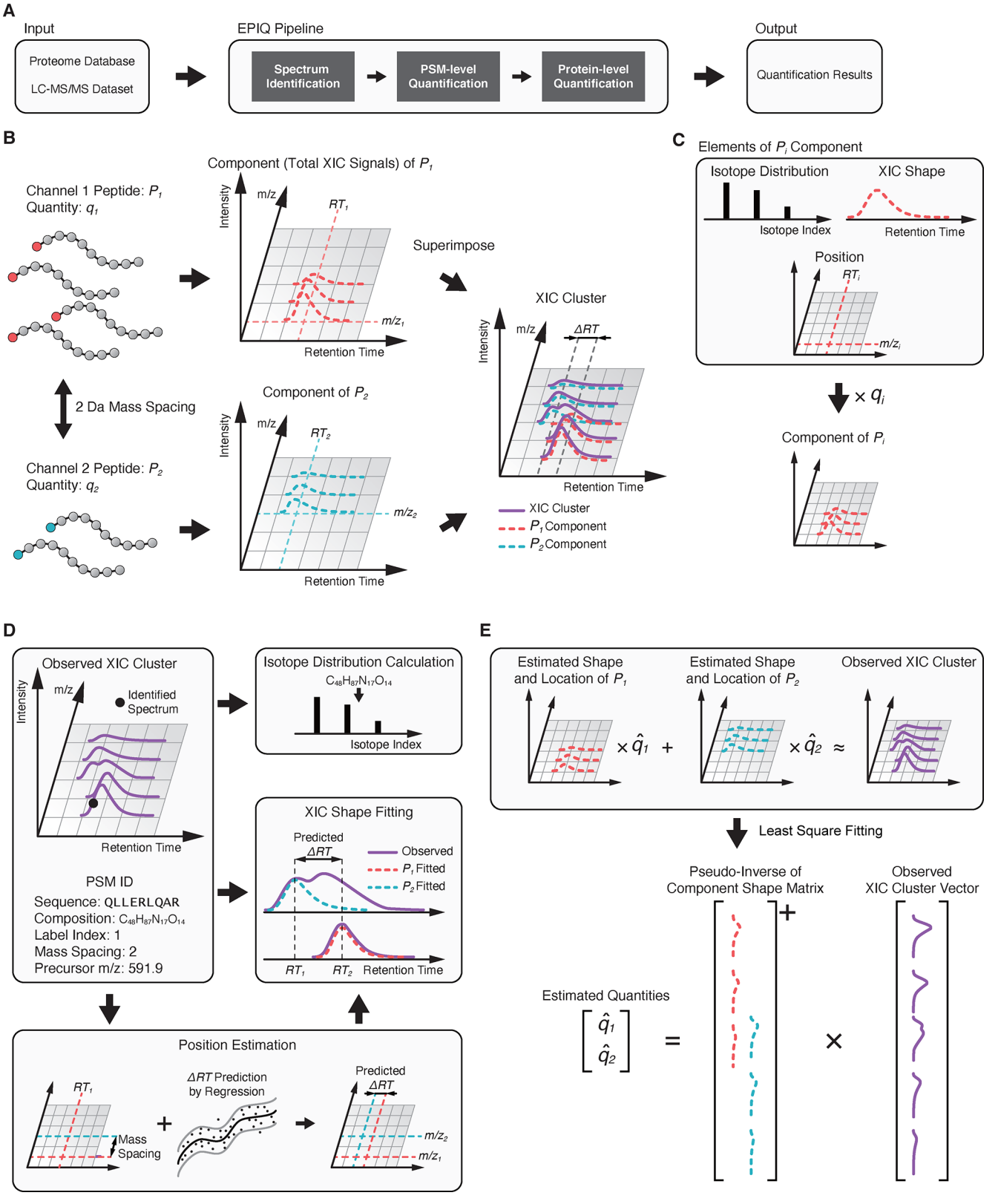


Figure 2. EPIQ pipeline and the model-based reconstruction algorithm for PSM-level quantification. The EPIQ algorithm first estimates the shape and (m/z and RT) position of each channel XIC signal (termed channel template) and then attempts to reconstruct the observed signal as a weighted summation of the templates. The weights minimizing the error between the reconstruction and the observed signal yield the quantities. In the template estimation step, the signal complication issues shown in Figure S1 preventing high-multiplexity are resolved; RT shift is accurately predicted by a regression method and (overlapping) channel signals are effectively separated. The missing quantity count is minimized in the following reconstruction step, where the estimated templates enable sensitive detection of noisy or even partial channel signals. This figure only provides a simple illustration of the EPIQ algorithm. A detailed description of each step is given in Supporting Information. (A) EPIQ pipeline. The inputs are LC-MS/MS data set and proteome database. The outputs are PSM-, peptide-, and protein-level quantification results (.tsv or.xlsx file).

https://dx.doi.org/10.1021/acs.analchem.9b05148
Anal. Chem. 2020, 92, 4980−4989

Figure 2. continued

EPIQ pipeline consists of three stages: PSM identification by MS-GF+, PSM quantification, and protein quantification. Panels B−E illustrate the model-based reconstruction algorithm used in PSM-level quantification. (B and C) The model used in EPIQ, describing the generation of the XIC cluster (purple lines) from the peptide forms P_1 and P_2 with known quantities q_1 and q_2. (B) The XIC cluster is the combined signal from P_1 and P_2. The total signal from P_1 (from P_2) is called component for P_1 (for P_2). By superimposing the components, the XIC cluster is generated. (C) The model assumes that three elements are given for P_i: (i) the position where its XICs are located in the m/z-RT plane, (ii) the shape of its XICs, and (iii) its isotope distribution. The isotope distribution and the XIC shape determine the shape of the component for P_i. By locating the component shape at the position and scaling it up using the quantity q_i, the component for P_i (red dashed lines) is defined. (D) Component shape and location estimation given a PSM (peptide-spectrum match) and its corresponding observed XIC cluster. The three elements constituting a component are estimated. The position is estimated by combining the position of the PSM, the mass spacing, and the predicted RT shift, where the RT shift is predicted using a machine learning based regression method. The XIC shapes are then obtained by fitting log-normal probability density function (pdf) to the observed XICs. The isotope distribution is efficiently calculated per PSM by using an approximation algorithm. With the estimated elements, the component shape and location are acquired. (E) The reconstruction of the observed XIC cluster and the quantification. We find the quantities q_1 and q_2 yielding the best-fit reconstruction of the observed XIC cluster, by solving a least-squares fitting problem. Briefly, we represent the observed XIC cluster as column vector by concatenating XICs in different m/z values sequentially. Each component shape is converted into a column vector in the same way. By appending these column vectors, all of the component shapes are represented by a single matrix (the component shape matrix). By multiplying Moore-Penrose pseudoinverse of the component shape matrix (denoted by $[\bullet]^+$) to the observed XIC cluster vector, the quantities are obtained.

the existing MS1-level quantification tools like MaxQuant[17] and Proteome Discoverer. To resolve these issues, we had to develop a new quantification algorithm, named EPIQ.

EPIQ Pipeline. EPIQ comprises three steps: spectrum identification, peptide spectrum match (PSM)-level quantification, and protein-level quantification (Figure 2A). First, in the spectrum identification step, reliable PSMs are harvested using the MS-GF+ search engine.[18] An identified PSM conveys the peptide sequence and the channel (the index specifying the peptide form from the lightest to the heaviest, e.g., for a SILAC pair, the light form corresponds to channel 1; heavy, channel 2). Second, in the PSM-level quantification step, the quantity is estimated per channel for each identified PSM, using a model-based reconstruction algorithm. Finally, PSM-level quantities are merged during protein-level quantification step. EPIQ is implemented in C# based on "The Informed Proteomics" project (https://github.com/PNNL-Comp-Mass-Spec/Informed-Proteomics).[19] A detailed description of the EPIQ algorithms is provided in Supplementary Algorithm Notes. Below we briefly describe the model-based reconstruction algorithm.

Model-Based Reconstruction Algorithm for PSM-Level Quantification. We focus on a specific peptide P and its peptide forms P_i of known quantities q_i for all channels i. Assume here that the peptide P has only a single identified spectrum for simplicity; a more realistic case in which multiple spectra are identified for P is discussed in Supplementary Algorithm Notes. The algorithm is based on a generative model that models the generation of the extracted ion chromatogram (*XIC*) *cluster* (the combined XIC signals from all the peptide forms of P; for example, a SILAC pair corresponds to one XIC cluster). Define a component for P_i as the total XIC signals from a given peptide form P_i. Then we model the XIC cluster is generated by superimposing the components for P_i for all i (Figure 2B).

The component for each P_i is generated as follows. Assume that the model contains information regarding the following three factors per P_i: exact position (m/z and RT), the shape of XIC, and natural isotope distribution (Figure 2C). The XIC shape only specifies the shape of XIC, not its quantity. The XIC shape and the isotope distribution together determine the shape of the component. By locating the component shape at the position and scaling it using the quantity q_i, the component for P_i is defined.

To apply the aforementioned model for quantification, the reconstruction algorithm first estimates the above three factors from the observed XIC cluster and available information extracted from the identified spectrum (Figure 2D). For the position, the m/z positions are readily calculated with the precursor m/z of the identified peptide form and the mass spacing. To determine the RT positions, the RT shifts from the identified peptide form to other forms are required. To accurately predict the RT shifts, we adopted the nu-SVR method[20] with empirically selected features including the number of deuteriums and the relative RT (on the whole LC plane) of the identified PSM. Thereafter, the XIC shape per channel is estimated by fitting log-normal probability density functions (pdf) to the observed XIC, based on our finding that an XIC shape fits well to a log-normal pdf. The observed XIC to be fitted is extracted at the m/z and RT position estimated above. Lastly, isotope distribution of the identified peptide is calculated using an efficient approximation algorithm (see Supplementary Algorithm Notes for details).

Per P_i, the three estimated factors determine the component shape in its position. The observed XIC cluster is reconstructed as a weighted summation of these component shapes such that the power of the discrepancy between the reconstruction and the observation is minimized. This is a well-defined least-squares fitting issue, which is readily solvable using Moore-Penrose pseudoinverse (see Figure 2E and Supplementary Algorithm Notes). The obtained weights are the estimates for quantities q_i. Of note, this reconstruction amounts to the deconvolution of the components from the observed XIC cluster.

Accurate Quantification of *in Vivo* and *in Vitro* Labeled HeLa Proteome Samples. To evaluate the quantification performance of our EPIQ tool in the existing labels first, we applied EPIQ analysis to the dimethylated data sets (dimethylation (DM)-3plex and DM-5plex) generated via single shot LC-MS/MS experiments, along with two other widely used algorithms, Proteome Discoverer (PD) and MaxQuant (MQ).[17] Here, a protein stands for a protein group for EPIQ and MQ, and a master protein stands for PD. A protein/PSM is called quantified if at least one channel quantity is positive, and all-channel-quantified if all channel quantities are positive. EPIQ quantification results were highly accurate and precise in those known labels, and outperformed PD and MQ in terms of number of all-channel-quantified

https://dx.doi.org/10.1021/acs.analchem.9b05148
Anal. Chem. 2020, 92, 4980−4989

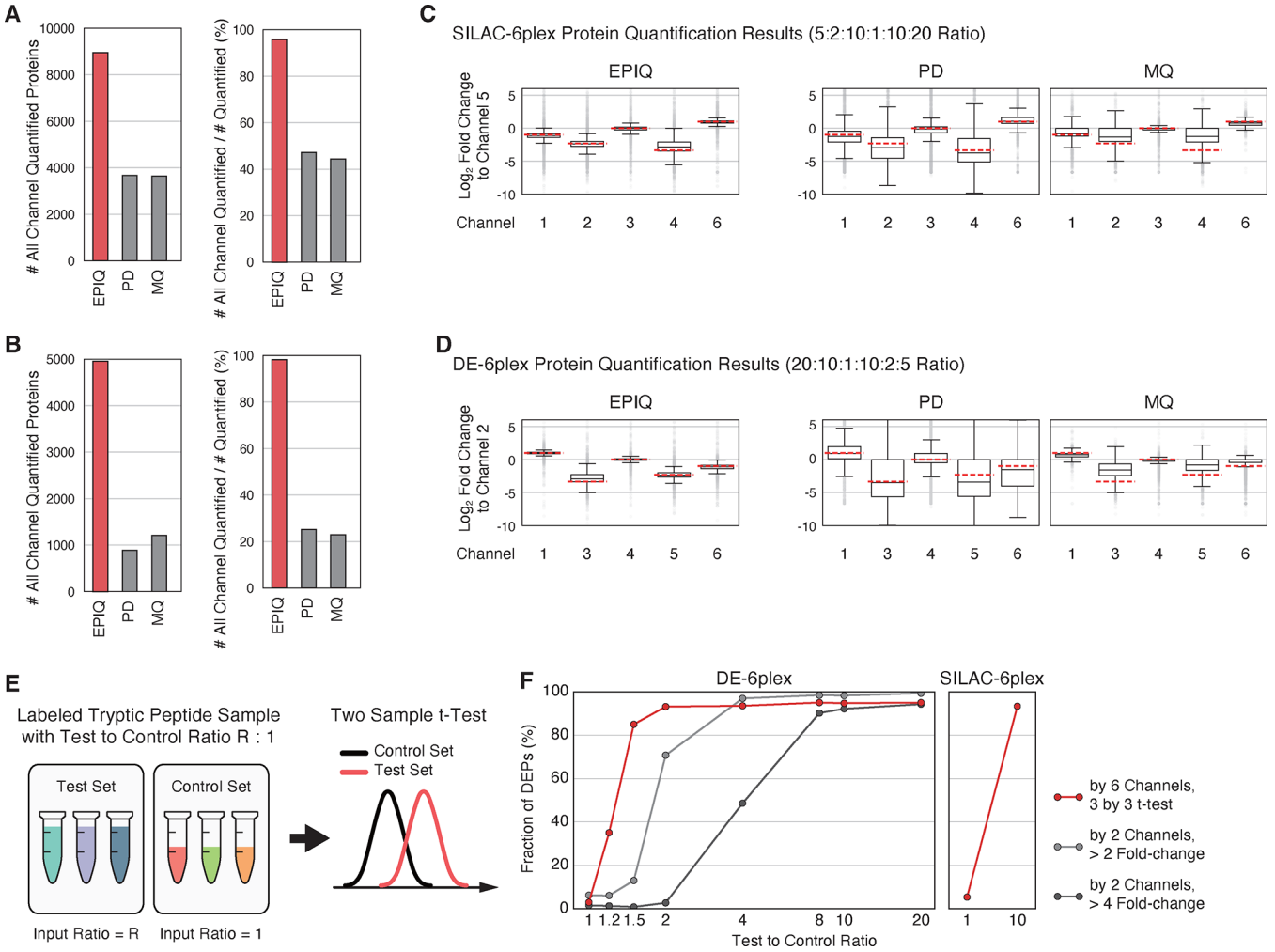


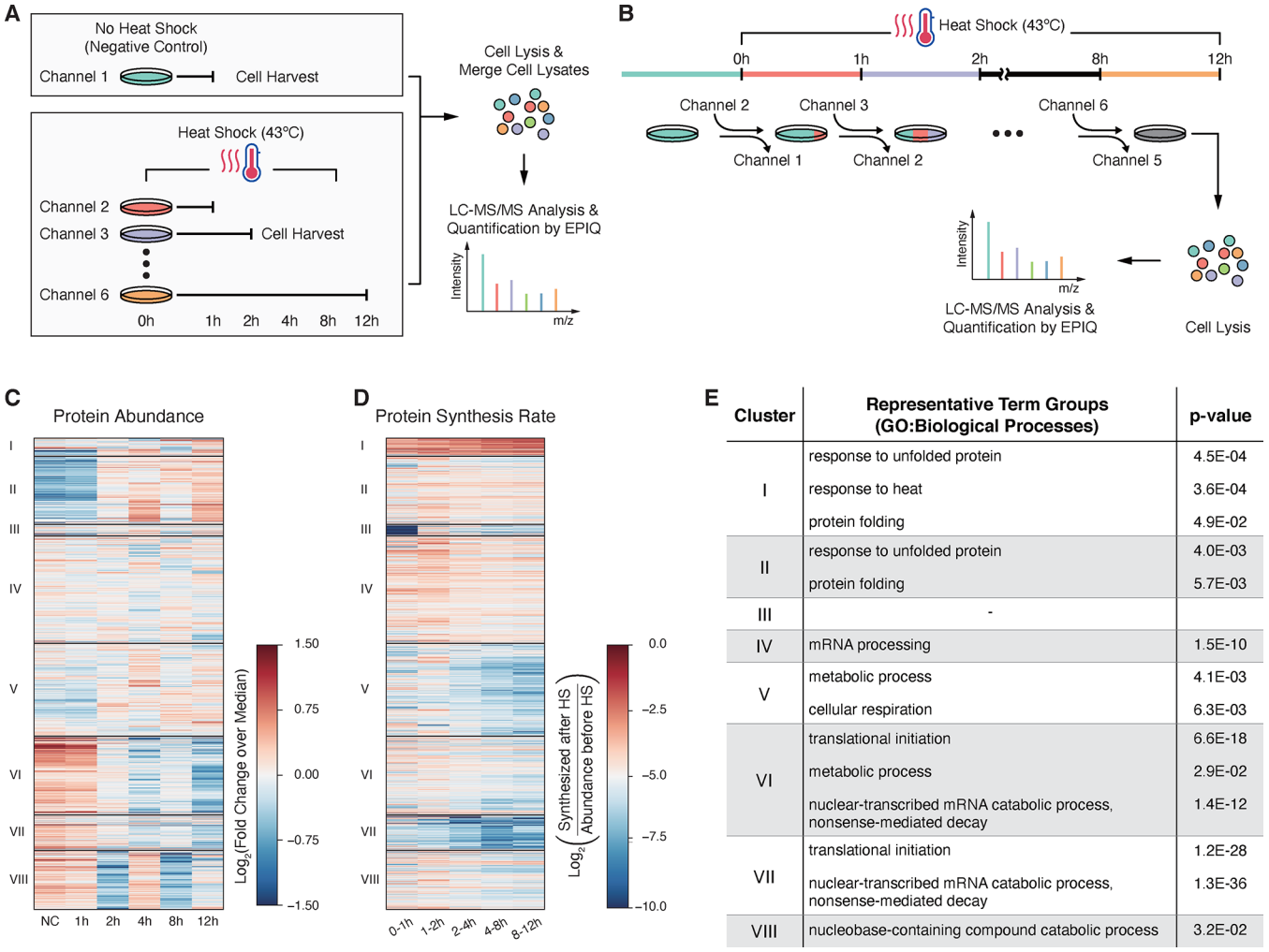
Figure 3. Benchmark tests with 6-plex *in vivo* or *in vitro* labeled HeLa lysates. (A) The barplots showing the number (left panel) and proportion (right panel) of all channel quantified proteins in SILAC-6plex-labeled HeLa lysate sample (multidimensional LC-MS/MS using HILIC). (B) Analogs of (A) for the DE-6plex-labeled HeLa lysate sample (multidimensional LC-MS/MS using SCX-stage-tip). (C) The boxplot for protein-level fold changes from the quantities of channel 2, reported by EPIQ (left) and other tools (right), for SILAC-6plex-labeled HeLa lysate sample. Red dashed lines specify the input ratio. The median (center line), first and third quartiles (lower and upper box limits, respectively), and 1.5 times the interquartile range (whiskers) are shown in boxplots. All quantified proteins were counted, where zero quantities were substituted by 1/100 of the maximum quantity per protein (see materials and methods). (D) Analogs of (C) for the DE-6plex-labeled HeLa lysate sample. (E) The preparation of HeLa lysate sample for detection of differentially expressed proteins (DEPs). Three channels were assigned to control and the others to test. The ratios between test and control were 1:*R*, where R = 1, 0.8, 4, 8, 10, and 20 for DE-6plex and R = 1 and 10 for SILAC-6plex-labeled samples. $R = 1$ (or $R \neq 1$) corresponds to the case in which no (or all) proteins are differentially expressed. A protein is declared as a DEP if the p-value from two sample *t* tests between control and test triplets is less than 0.05. (F) The fraction of DEPs determined by two sample *t* test (red line). The specificity can be estimated by one minus the portion of the DEPs when the test to control ratio equals 1. The sensitivity is given by the portion of the DEPs for each of other test to control ratios. For comparison, we also drew the fractions of proteins having fold changes larger than 2 (gray line) and 4 (black line) for DE6-plex samples. For those, only two channels (one from test and the other from control) were used to calculate the fold changes per protein, which simulates nonmultiplexed determination of DEPs (see the Experimental Section).

proteins in both DM-3plex and DM-5plex data sets at three different input mixture ratios (Figure S6).

To demonstrate the effectiveness of our platform in newly developed labels, we prepared HeLa cell lysates labeled by either *in vivo* (SILAC-6plex) or *in vitro* (DE-6plex) labeling. Input mixture ratio of 5:2:10:1:10:20 and 20:10:1:10:2:5 were used for SILAC-6plex and DE-6plex, respectively. These ratios have a large dynamic range of 20-fold and both subtle and large fold changes between two adjacent channels. Both samples were prefractionated via "deuterium effect"-free separation before reversed phase LC-MS/MS (Figure S7). For the SILAC-6plex-labeled sample, 32 fractions were prepared by hydrophilic interaction liquid chromatography (HILIC).[21,22] In the case of the DE-6plex-labeled sample, the sample was fractionated into eight fractions by the SCX-StageTip before LC-MS/MS. The generated data sets were analyzed using EPIQ along with PD and MQ as well.

Figure 3A shows the number and proportion (out of quantified) of all-channel-quantified proteins from each tool for the SILAC-6plex sample (see Table S1 for the protein lists). Figure 3B is the analog of Figure 3A for the DE-6plex sample (see Table S2 for the protein lists). EPIQ reported almost 9000 (from HILIC separated 32 fractions) and 5000 (from SCX-StageTip separated eight fractions) all-channel-quantified proteins. Of note, the portions of all-channel-quantified proteins were higher than 95%, demonstrating that

https://dx.doi.org/10.1021/acs.analchem.9b05148
Anal. Chem. 2020, 92, 4980−4989



E

Cluster	Representative Term Groups (GO:Biological Processes)	p-value
I	response to unfolded protein	4.5E-04
	response to heat	3.6E-04
	protein folding	4.9E-02
II	response to unfolded protein	4.0E-03
	protein folding	5.7E-03
III	-	
IV	mRNA processing	1.5E-10
V	metabolic process	4.1E-03
	cellular respiration	6.3E-03
VI	translational initiation	6.6E-18
	metabolic process	2.9E-02
	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	1.4E-12
VII	translational initiation	1.2E-28
	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	1.3E-36
VIII	nucleobase-containing compound catabolic process	3.2E-02

Figure 4. SILAC-6plex and EPIQ reveal protein abundance and synthesis rate dynamics during the heat shock response (HSR) of HeLa cells. (A) Schematic view of the protein abundance dynamics analysis by steady-state SILAC. For SILAC-6plex channels 1−6, fully labeled HeLa cells were incubated at 43 °C for 0 (negative control), 1, 2, 4, 8, and 12 h and harvested. Then, harvested cells were analyzed by LC-MS/MS together. Protein quantities were obtained from LC-MS/MS results by EPIQ. (B) Schematic view of the protein synthesis rate dynamics analysis by pulsed SILAC. A single cultured HeLa cell dish containing the SILAC channel 1 media was sequentially subject to media-exchange with SILAC channels 2, 3, 4, 5, and 6 media at the time points of 0, 1, 2, 4, and 8 h during the 43 °C heat shock treatment, respectively. After 12 h of heat shock, cells were lysed and analyzed by LC-MS/MS. Resulting LC-MS/MS data were analyzed by EPIQ to attain protein quantities. (C and D) Heat maps showing protein abundance dynamics (C) and protein synthesis rate dynamics (D) under HSR of HeLa cells. The 3894 proteins quantified with ≥2 unique peptide sequences in both steady-state SILAC and pulsed SILAC are included. In the heat map in panel C, each cell indicates a $\log_2$ fold ratio between median protein abundance and that at each time point. In the heat map in panel D, each cell shows the $\log_2$ fold ratio between steady state abundance before heat shock (HS) and pulse labeled abundance after HS time points. NC stands for negative control. (E) Enrichment p-values of several selected GO-terms in each cluster. p-values were calculated and adjusted with the default option of g:profiler.

EPIQ drastically reduces missing quantities. The other tools reported substantially smaller numbers (<2-fold) of proteins with higher missing quantities than EPIQ did. Figure 3C,D shows the boxplots for protein-level fold change for EPIQ (left panel), PD, and MQ (right; see Table S3 for medians and standard deviations of protein-level fold changes). Most intriguingly, EPIQ generates almost unbiased and precise quantities for all channels with a 20-fold linear dynamic range, unlike the other tools. Even in PSM-level quantification, EPIQ shows the unbiased results for the fully tryptic peptides with narrow mass spacing and the missed cleavage peptides with high RT shifts as well (Figure S8 and see Tables S4 and S5 for each PSM-level quantification results).

Sensitive and Specific Detection of Differentially Expressed Proteins. Next, we applied our platform for the detection of differentially expressed proteins (DEPs). To this end, the six channels in our labels were configured such that three channels are assigned to control and the others to test. The ratios between test and control triplets (denoted by R) were set to R:1, where R = 1, 1.2, 1.5, 2, 4, 8, 10, and 20 (in DE-6plex-labeled HeLa samples; Figure 3E). The prepared dilution series were subject to an LC-MS/MS run using Q-Exactive Classic followed by EPIQ analysis. A protein was classified as a DEP if the $p < 0.05$ in two-sided Welch's two-sample t test between test and control triplets.

Fractions of the DEPs among the all-channel-quantified proteins are shown in Figure 3F (red line in left panel; see Figure S9A for fold change box plots). Our platform showed a high specificity (1-DEP fraction for R = 1) of 97% and sensitivity (DEP fraction for $R \neq 1$) of approximately 95% for the ratios exceeding 2:1. Even at a ratio as subtle as 1.5:1, the sensitivity exceeded 80%. We also drew the fraction of proteins

https://dx.doi.org/10.1021/acs.analchem.9b05148
Anal. Chem. 2020, 92, 4980−4989

with fold changes higher than 2 or 4 (gray and black lines) in Figure 3F. The fold change per protein was determined using only two channel intensities (one from test and the other from control; see materials and methods). A large discrepancy between red and gray/black lines in Figure 3F, therefore, illustrates that multiplexed DEP detection via our platform can uncover substantially more DEPs particularly with modest fold changes than nonmultiplexed detection methods. The same multiplexed analysis was performed with SILAC-6plex-labeled HeLa samples for R = 1 and 10 (analyzed by Orbitrap Fusion Lumos). The specificity was 95%; sensitivity, 93% (Figure 3F right panel; see also Figure S9B). Intriguingly, we found that EPIQ well-quantifies the peptides with relatively uniform quantity between labeled channels (R = 1) across the whole precursor intensity range, while a little insensitive DEP detection was observed at low precursor intensity range in the highly differential expression (R = 10; Figure S10). DEP analysis results are provided in Table S6.

Application of SILAC-6plex and EPIQ Analysis: Highly Time-Resolved Profiling of Protein Synthesis Rate during Heat Shock Response. To demonstrate the unique benefits of our SILAC-6plex and EPIQ platform, we used the platform to dissect the dynamics of protein regulation during heat shock response (HSR) in HeLa cells across 6 time points (negative control (no heat shock; 0 h), 1, 2, 4, 8, and 12 h after 43 °C heat shock). To this end, we performed both steady-state SILAC (Figure 4A) and pulsed SILAC experiments[23] (Figure 4B) covering the same six time points, to analyze the protein synthesis rate as well as abundance changes. As a result, a total of 8159 and 7741 protein groups were quantified for the steady-state and pulsed SILAC data sets, respectively, via HILIC based multidimensional LC-MS/MS and EPIQ analysis (Table S7). For protein groups quantified in both experiments, clustering analysis was carried out by considering global protein abundance dynamics and protein synthesis rate dynamics together (Figure 4C,D and Table S7). The representative gene ontology (GO) terms in biological processes for the respective cluster were also presented with p-values (Figure 4E).

The quantified proteins showed typically three patterns of dynamics in protein abundance after heat shock (Figure 4C); up-regulated (clusters I and II), largely unchanged (clusters III, IV, and V), and down-regulated patterns (clusters VI, VII, and VIII). However, intriguingly, the protein synthesis rate (Figure 4D), which could be uniquely characterized by our 6-plexed pulsed SILAC and EPIQ platform, provided distinctive information on the translational regulation; this may be hardly revealed by only measuring the dynamics of global protein abundance. Compared to transcriptional regulation, translational regulation is known to be a much more immediate process against cellular stressors to adjust the protein concentration for cell survival.[24] Therefore, integrating both steady-state and pulsed SILAC experiments can provide unique information to get a deeper understanding of the cellular response under severe heat stress.

For example, the highly upregulated clusters I and II, which were related with heat stress response GO terms such as "response to unfolded protein", "protein folding", and "response to heat", showed largely constant synthesis rates but with different orders; substantially faster protein synthesis was observed for cluster I compared to that for cluster II. This suggests that the proteins in cluster I might be more active players in the HSR and thereby more actively degraded after fulfilling their responding roles (Figures 4C−E and S11A-B). For instance, the abundance trends of Hsp70 in cluster I, a known key player for heat shock response, were well matched with the previous study.[25] Also, continuous transcriptional induction of HSR genes reported in heat shocked mouse embryonic fibroblasts (MEF)[26] can support the stable up-regulation of Hsp70 translation in our data (Figure S11C).

For the down-regulated protein clusters VI and VII, "translation initiation" and "nonsense-mediated decay" in GO terms were highlighted, and the related proteins with these terms are known to be post-translationally repressed under cellular stress conditions.[27−29] These two clusters also showed decreasing patterns of synthesis rate, which are well matched with the observation of transcriptional down-regulation of metabolism and translation related genes in MEF upon heat shock.[26] This data suggests a possibility of conserved HSR-mediated down-regulation of cellular metabolism and protein synthesis in mammalian cells.

The decreasing dynamics of protein abundance and synthesis rate was quite different between the two clusters and the integration of two dynamics provides an interesting perspective regarding protein stability for the proteome in clusters VI and VII. The pattern of steeply decreasing protein abundance under gently decreasing protein synthesis rate in cluster VI and the opposite pattern in cluster VII (Figure 4C,D) can be interpreted as the proteins in cluster VI are more unstable or actively degraded than those in cluster VII (Figure S11A,B), although they share similar enriched-GO terms (Figure 4E).

The largely unchanged clusters (III, IV, and V) in global protein abundance could be differentiated by the distinctive patterns of synthesis rate (Figure 4C,D). They were largely related with housekeeping biological processes such as mRNA processing (cluster IV) and cellular respiration (cluster V), which partially explain their stable expression under HSR. Notably, both mRNA processing and metabolism-related genes are transcriptionally down-regulated in heat-shocked MEF, but the repression of metabolism-related genes precedes that of mRNA processing genes.[26] This is consistent with the higher protein synthesis rate in cluster IV compared to that for cluster V, especially at the 1−2 h after heat shock.

Like the differences between mRNA processing genes and cellular respiration genes, pulsed SILAC data would open the possibility to discover the hidden regulation regarding protein synthesis or degradation for these stably expressed proteins. For example, the level of β-actin (cluster IV protein) was almost unchanged upon heat shock stress (Figure S11C), as expected from previous report.[25] However, we observed a surge of protein synthesis at the early time point (1−2 h), indicating some role of β-actin in HSR in the early phase. A similar observation in heat shocked MEF, a rapid transient induction of cytoskeleton transcripts,[26] also supports the credibility of SILAC-6plex and EPIQ platform. All of these results not only reaffirmed many previously known regulations of HSR, indicating the robustness of the platform, but also provided rich proteome-wide resources for further HSR studies, covering both protein abundance and protein synthesis rate dynamics.

DISCUSSION

We introduced a highly multiplexed MS1-based proteome quantification platform applicable to both *in vitro* and *in vivo* labeling. We demonstrated that our platform achieves accurate

https://dx.doi.org/10.1021/acs.analchem.9b05148
Anal. Chem. 2020, 92, 4980−4989

and sensitive quantification with a wide dynamic range and minimal missing quantities. While the present study focused on the demonstration of 6-plexed labeling schemes, EPIQ can support higher multiplexed experiments. For instance, we are currently preparing 9-plexed SILAC (Figure S2D,E) and further multiplexed reductive peptide dialkylation (e.g., dipropylation or dibutylation) under the same labeling constraints (2 Da of mass spacing and heavily differential deuteration) to further extend our platform.

Highly multiplexed metabolic labeling has been reported by introducing various very tiny mass defects (within 36 mDa of mass window) in a single nominal mass of lysine isotopologues, termed as neutron-encoded or NeuCode SILAC.[10,30,31] This approach allowed for 6plex of multiplexity with a ≥6 mDa mass difference between labeling channels by virtue of extremely ultrahigh resolution mass spectrometry (near one million of resolving power at m/z 200), which is super expensive and limitedly accessible. With high-end instruments such as Orbitrap Fusion Lumos (500 k resolving power at m/z 200), practically ≥18 mDa of mass difference between channels should be secured to be distinctly resolved, thereby decreasing the multiplexity to 3plex. To generate neutron-encoded mass difference, it is indispensable to allow for differential number of deuterium atoms in NeuCode-lysine isotopologues. Thereby, the issue of RT shift and the corresponding reduction of quantification precision[30] (Figure S12) might be inevitable without computational solution. Furthermore, XICs from different NeuCode SILAC channels can be intermixed by the coalescence of the measured m/z values especially for high-abundance peptides even at the ultrahigh resolution of mass spectrometer, compromising the quantification accuracy.[31] To increase multiplexity of NeuCode SILAC, it is recommended to combine NeuCode approach with traditional SILAC, which uses different nominal masses (e.g., K4, K8, and K12).[31] Considering our platform is based on the nominal mass differences (≥2 Da) among labels with higher multiplexity than traditional SILAC, the expansion of our platform to neutron-encoded labeling approach in the future would allow for even higher multiplexity.

The increased MS signal complexity is an unavoidable complication due to increased multiplexity in our platform. Even if overlapped MS signals across channels can be resolved using our suggested method, MS/MS triggering would be more biased toward abundant precursor ions, decreasing the number and diversity of quantified proteins. For example, we identified less than half the number of protein groups from a single shot LC-MS/MS analysis of our 6-plexed samples compared with those of single channel only labeled sample (Figure S13A). The reduction in protein identification could not be resolved by introducing a long gradient LC-MS acquisition, only leading to slight increase in the number of identification (Figure S13B,C).

However, this hurdle could be addressed via "deuterium effect"-free HILIC prefractionation (Figure S7) based on the multiple fraction concatenation strategy.[32,33] When a concatenated HILIC fractionation method was applied, we were able to quantify ∼9k protein groups from 32 fractions (Figure 3A). Although this final number of quantified protein groups is comparable to that from the well-established RPLC fractionation based experiments, which generally allows for 9−10k protein groups from 24 fractions,[34−36] the increase of quantified protein groups in 6plexed samples was quite slower compared with the nonmultiplexed sample in the initial stage of fractions. This shows that a fairly large number of fractions should be analyzed to get a comparable number of proteins (Figure S14). In other words, a higher number of fractions and accordingly a longer total MS running time would be indispensable to offset the identification reduction induced by the MS1-level degeneracy, which negatively affects the overall analysis throughput.

The increased multiplexity could allow for statistically concrete experimental designs for diverse biological/medical studies that need to compare large numbers of experimental replicates/conditions. The *in vitro* dialkylation with EPIQ could be a potent alternative for the widely adopted isobaric labeling because our platform does not require high-end instruments capable of MS3 analysis such as synchronous precursor selection mode, which is essential for most isobaric methods to achieve acceptable quantification accuracy.[37]

We illustrated the potentially unique benefits of our highly multiplexed *in vivo* quantification platform by integrating both steady-state and pulsed SILAC experiments, which generate the protein dynamics regarding not only global abundance but also synthesis rate. This approach can be useful to discover hidden players, which are hardly disclosed with global proteomics platform alone, and lead to a deeper understanding of the cellular response against stressors such as severe heat shock. Collectively, the present approach offers a versatile and affordable proteomic platform for the biological/medical research community.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.9b05148.

Supplementary figures, supplementary table list, detailed materials and experimental procedures, and detailed description on bioinformatic algorithms used in this paper (supplementary algorithm notes) (PDF)

Table S1. SILAC-6plex 5:2:10:1:10:20 proteins (related to Figure 3A,C) (XLSX)

Table S2. DE-6plex 20:10:1:10:2:5 proteins (related to Figure 3B,D) (XLSX)

Table S3. Medians and standard deviations of protein-level fold changes (related to Figure 3C,D) (XLSX)

Table S4. SILAC-6plex 20:10:1:10:2:5 PSMs, by EPIQ (related to Figure S8A) (XLSX)

Table S5. DE-6plex 20:10:1:10:2:5 PSMs, by EPIQ (related to Figure S8B) (XLSX)

Table S6. DEP analysis on protein groups, by EPIQ (related to Figures 3E,F and S9) (XLSX)

Table S7. Analysis on protein dynamics of HeLa cells under HSR (related to Figure 4) (XLSX)

Accession Codes

The source codes of EPIQ and its usage guide are freely accessible though GitHub (https://github.com/YeonChoi/EPIQ). All of the original LC-MS/MS data sets and related identification files used to support this paper have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository[38] with the data set identifier <PXD009952>.

https://dx.doi.org/10.1021/acs.analchem.9b05148
Anal. Chem. 2020, 92, 4980−4989

AUTHOR INFORMATION

Corresponding Authors

V. Narry Kim − *Center for RNA Research, Institute for Basic Science, Seoul 08826, Korea; School of Biological Sciences, Seoul National University, Seoul 08826, Korea;* orcid.org/0000-0002-3213-4965; Email: narrykim@snu.ac.kr

Jong-Seo Kim − *Center for RNA Research, Institute for Basic Science, Seoul 08826, Korea; School of Biological Sciences, Seoul National University, Seoul 08826, Korea;* orcid.org/0000-0002-8909-8440; Email: jongseokim@snu.ac.kr

Authors

Yeon Choi − *Center for RNA Research, Institute for Basic Science, Seoul 08826, Korea; School of Biological Sciences, Seoul National University, Seoul 08826, Korea*

Kyowon Jeong − *Center for RNA Research, Institute for Basic Science, Seoul 08826, Korea; School of Biological Sciences, Seoul National University, Seoul 08826, Korea*

Sanghee Shin − *Center for RNA Research, Institute for Basic Science, Seoul 08826, Korea; School of Biological Sciences, Seoul National University, Seoul 08826, Korea*

Joon Won Lee − *Department of Applied Chemistry, Kyung Hee University, Yongin 17104, Korea*

Young-suk Lee − *Center for RNA Research, Institute for Basic Science, Seoul 08826, Korea; School of Biological Sciences, Seoul National University, Seoul 08826, Korea*

Sangtae Kim − *Illumina, Inc., San Diego, California 92122, United States*

Sun Ah Kim − *Center for RNA Research, Institute for Basic Science, Seoul 08826, Korea; School of Biological Sciences, Seoul National University, Seoul 08826, Korea*

Jaehun Jung − *Department of Applied Chemistry, Kyung Hee University, Yongin 17104, Korea*

Kwang Pyo Kim − *Department of Applied Chemistry, Kyung Hee University, Yongin 17104, Korea;* orcid.org/0000-0003-0095-3787

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.9b05148

Author Contributions

[⊥]Y.C., K.J., S.S., and J.W.L. contributed equally. J.-S.K. and K.J. conceived the idea of the quantification platform. J.-S.K. designed all *in vivo*/*in vitro* labeling schemes. K.J. and Y.C. developed and implemented EPIQ algorithm with S.K. and Y.S.L.'s assistance. Y.C. performed bioinformatic analysis. S.S. carried out SILAC-6plex experiments. J.W.L. performed DE-6plex experiments with S.A.K., J.J., and K.P.K.'s support. V.N.K. led the project and provided resources. Y.C., K.J., S.S., J.W.L., V.N.K., and J.-S.K. wrote the paper. All authors commented on and approved the paper.

Notes

The authors declare the following competing financial interest(s): Competing financial interests: K.J., Y.C., V.N.K., and J.-S.K. are co-inventors on patent applications filed by IBS (Institute for Basic Science) and SNU R&DB Foundation (application number: KR 10-2017-0085558) covering the method described herein. S.K. is an employee of Illumina, Inc. The other authors declare no competing financial interests.

ACKNOWLEDGMENTS

This work was supported by IBS-R008-D1 of Institute for Basic Science from the Ministry of Science and ICT of Korea (K.J., Y.C., S.S., S.A.K., Y.S.L, V.N.K., J.-S.K.) and by the BK21 Research Fellowships from the Ministry of Education of Korea (Y.C., S.S.). We are grateful to Jong-Eun Park, Hyeshik Chang, Ahyeon Son, and other members of our laboratory for discussion and technical help.

REFERENCES

(1) Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (12), 6591−6596.

(2) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. *Mol. Cell. Proteomics* **2002**, *1* (5), 376−386.

(3) Harsha, H. C.; Molina, H.; Pandey, A. *Nat. Protoc.* **2008**, *3* (3), 505−516.

(4) Blagoev, B.; Ong, S. E.; Kratchmarova, I.; Mann, M. *Nat. Biotechnol.* **2004**, *22* (9), 1139−1145.

(5) Savitski, M. M.; Zinn, N.; Faelth-Savitski, M.; Poeckel, D.; Gade, S.; Becher, I.; Muelbaier, M.; Wagner, A. J.; Strohmer, K.; Werner, T.; Melchert, S.; Petretich, M.; Rutkowska, A.; Vappiani, J.; Franken, H.; Steidel, M.; Sweetman, G. M.; Gilan, O.; Lam, E. Y. N.; Dawson, M. A.; et al. *Cell* **2018**, *173* (1), 260−274.

(6) Boutilier, J. M.; Warden, H.; Doucette, A. A.; Wentzell, P. D. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2012**, *908*, 59−66.

(7) Koehler, C. J.; Arntzen, M. O.; Thiede, B. *Rapid Commun. Mass Spectrom.* **2015**, *29* (9), 830−836.

(8) Hebert, A. S.; Merrill, A. E.; Bailey, D. J.; Still, A. J.; Westphall, M. S.; Strieter, E. R.; Pagliarini, D. J.; Coon, J. J. *Nat. Methods* **2013**, *10* (4), 332−334.

(9) Wu, Y.; Wang, F.; Liu, Z.; Qin, H.; Song, C.; Huang, J.; Bian, Y.; Wei, X.; Dong, J.; Zou, H. *Chem. Commun. (Cambridge, U. K.)* **2014**, *50* (14), 1708−1710.

(10) Rose, C. M.; Merrill, A. E.; Bailey, D. J.; Hebert, A. S.; Westphall, M. S.; Coon, J. J. *Anal. Chem.* **2013**, *85* (10), 5129−5137.

(11) Tzouros, M.; Golling, S.; Avila, D.; Lamerz, J.; Berrera, M.; Ebeling, M.; Langen, H.; Augustin, A. *Mol. Cell. Proteomics* **2013**, *12* (11), 3339−3349.

(12) Jung, J.; Jeong, K.; Choi, Y.; Kim, S. A.; Kim, H.; Lee, J. W.; Kim, V. N.; Kim, K. P.; Kim, J. S. *J. Proteome Res.* **2019**, *18* (3), 1078−1087.

(13) Altelaar, A. F.; Frese, C. K.; Preisinger, C.; Hennrich, M. L.; Schram, A. W.; Timmers, H. T.; Heck, A. J.; Mohammed, S. *J. Proteomics* **2013**, *88*, 14−26.

(14) Webb-Robertson, B. J.; Wiberg, H. K.; Matzke, M. M.; Brown, J. N.; Wang, J.; McDermott, J. E.; Smith, R. D.; Rodland, K. D.; Metz, T. O.; Pounds, J. G.; Waters, K. M. *J. Proteome Res.* **2015**, *14* (5), 1993−2001.

(15) Barrios-Llerena, M. E.; Pritchard, J. C.; Kerr, L. E.; Le Bihan, T. *J. Proteomics* **2011**, *74* (12), 2798−2809.

(16) Boersema, P. J.; Aye, T. T.; van Veen, T. A.; Heck, A. J.; Mohammed, S. *Proteomics* **2008**, *8* (22), 4624−4632.

(17) Cox, J.; Mann, M. *Nat. Biotechnol.* **2008**, *26* (12), 1367−1372.

(18) Kim, S.; Pevzner, P. A. *Nat. Commun.* **2014**, *5*, 5277.

(19) Park, J.; Piehowski, P. D.; Wilkins, C.; Zhou, M.; Mendoza, J.; Fujimoto, G. M.; Gibbons, B. C.; Shaw, J. B.; Shen, Y.; Shukla, A. K.; Moore, R. J.; Liu, T.; Petyuk, V. A.; Tolic, N.; Pasa-Tolic, L.; Smith, R. D.; Payne, S. H.; Kim, S. *Nat. Methods* **2017**, *14* (9), 909−914.

(20) Scholkopf, B.; Smola, A. J.; Williamson, R. C.; Bartlett, P. L. *Neural Comput* **2000**, *12* (5), 1207−1245.

(21) Guo, K.; Ji, C.; Li, L. *Anal. Chem.* **2007**, *79* (22), 8631−8638.

(22) Wu, C. J.; Chen, Y. W.; Tai, J. H.; Chen, S. H. *J. Proteome Res.* **2011**, *10* (3), 1088−1097.

(23) Schwanhausser, B.; Gossen, M.; Dittmar, G.; Selbach, M. *Proteomics* **2009**, *9* (1), 205−209.

(24) Hershey, J. W.; Sonenberg, N.; Mathews, M. B. *Cold Spring Harbor Perspect. Biol.* **2012**, *4* (12), a011528.

(25) Fortugno, P.; Beltrami, E.; Plescia, J.; Fontana, J.; Pradhan, D.; Marchisio, P. C.; Sessa, W. C.; Altieri, D. C. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (24), 13791−13796.

(26) Mahat, D. B.; Salamanca, H. H.; Duarte, F. M.; Danko, C. G.; Lis, J. T. *Mol. Cell* **2016**, *62* (1), 63−78.
(27) Goetz, A. E.; Wilkinson, M. *Cell. Mol. Life Sci.* **2017**, *74* (19), 3509−3531.
(28) Vries, R. G.; Flynn, A.; Patel, J. C.; Wang, X.; Denton, R. M.; Proud, C. G. *J. Biol. Chem.* **1997**, *272* (52), 32779−32784.
(29) Duncan, R.; Hershey, J. W. *J. Biol. Chem.* **1984**, *259* (19), 11882−11889.
(30) Merrill, A. E.; Hebert, A. S.; MacGilvray, M. E.; Rose, C. M.; Bailey, D. J.; Bradley, J. C.; Wood, W. W.; El Masri, M.; Westphall, M. S.; Gasch, A. P.; Coon, J. J. *Mol. Cell. Proteomics* **2014**, *13* (9), 2503−2512.
(31) Overmyer, K. A.; Tyanova, S.; Hebert, A. S.; Westphall, M. S.; Cox, J.; Coon, J. J. *Nat. Protoc.* **2018**, *13* (1), 293−306.
(32) Wang, Y.; Yang, F.; Gritsenko, M. A.; Wang, Y.; Clauss, T.; Liu, T.; Shen, Y.; Monroe, M. E.; Lopez-Ferrer, D.; Reno, T.; Moore, R. J.; Klemke, R. L.; Camp, D. G., 2nd; Smith, R. D. *Proteomics* **2011**, *11* (10), 2019−2026.
(33) Di Palma, S.; Raijmakers, R.; Heck, A. J.; Mohammed, S. *Anal. Chem.* **2011**, *83* (21), 8352−8356.
(34) Mun, D. G.; Bhin, J.; Kim, S.; Kim, H.; Jung, J. H.; Jung, Y.; Jang, Y. E.; Park, J. M.; Kim, H.; Jung, Y.; Lee, H.; Bae, J.; Back, S.; Kim, S. J.; Kim, J.; Park, H.; Li, H.; Hwang, K. B.; Park, Y. S.; Yook, J. H.; et al. *Cancer Cell* **2019**, *35* (1), 111−124.
(35) Lee, H.; Mun, D. G.; Bae, J.; Kim, H.; Oh, S. Y.; Park, Y. S.; Lee, J. H.; Lee, S. W. *Analyst* **2015**, *140* (16), 5700−5706.
(36) Beck, S.; Michalski, A.; Raether, O.; Lubeck, M.; Kaspar, S.; Goedecke, N.; Baessmann, C.; Hornburg, D.; Meier, F.; Paron, I.; Kulak, N. A.; Cox, J.; Mann, M. *Mol. Cell. Proteomics* **2015**, *14* (7), 2014−2029.
(37) Ting, L.; Rad, R.; Gygi, S. P.; Haas, W. *Nat. Methods* **2011**, *8* (11), 937−940.
(38) Vizcaino, J. A.; Cote, R. G.; Csordas, A.; Dianes, J. A.; Fabregat, A.; Foster, J. M.; Griss, J.; Alpi, E.; Birim, M.; Contell, J.; O'Kelly, G.; Schoenegger, A.; Ovelleiro, D.; Perez-Riverol, Y.; Reisinger, F.; Rios, D.; Wang, R.; Hermjakob, H. *Nucleic Acids Res.* **2012**, *41*, D1063−1069.

https://dx.doi.org/10.1021/acs.analchem.9b05148
Anal. Chem. 2020, 92, 4980−4989