Journal of proteome research

Article

pubs.acs.org/jpr

Quantitative Proteomic Analysis of Mouse Embryonic Fibroblasts and Induced Pluripotent Stem Cells Using $^{16}O/^{18}O$ Labeling

Xin Huang,[†] Changhai Tian,[‡] Miao Liu,[†] Yongxiang Wang,[‡] Aleksey V. Tolmachev,[∥] Seema Sharma,[⊥] Fang Yu,[#] Kai Fu,[†] Jialin Zheng,[‡] and Shi-Jian Ding*[,†,§]

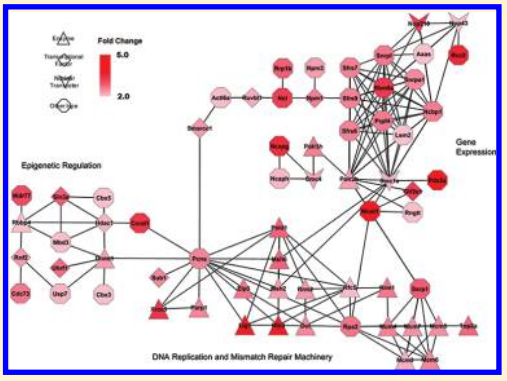
[†]Department of Pathology and Microbiology, [‡]Department of Pharmacology and Experimental Neuroscience, [#]Department of Biostatistics, and [§]Mass Spectrometry and Proteomics Core Facility, University of Nebraska Medical Center, Omaha, Nebraska 68198, United States

[∥]Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99352, United States

[⊥]Thermo Fisher Scientific, San Jose, California 95134, United States

Supporting Information

ABSTRACT: Induced pluripotent stem cells (iPSC) hold great promise for regenerative medicine as well as for investigations into the pathogenesis and treatment of various diseases. Understanding of key intracellular signaling pathways and protein targets that control development of iPSC from somatic cells is essential for designing new approaches to improve reprogramming efficiency. Here, we report the development and application of an integrated quantitative proteomics platform for investigating differences in protein expressions between mouse embryonic fibroblasts (MEF) and MEF-derived iPSC. This platform consists of $^{16}O/^{18}O$ labeling, multidimensional peptide separation coupled with tandem mass spectrometry, and data analysis with UNiquant software. With this platform, a total of 2481 proteins were identified and quantified from the $^{16}O/^{18}O$-labeled MEF-iPSC proteome mixtures with a false discovery rate of 0.01. Among them, 218 proteins were significantly upregulated, while 247 proteins were significantly downregulated in iPSC compared to MEF. Many nuclear proteins, including Hdac1, Dnmt1, Pcna, Ccnd1, Smarcc1, and subunits in DNA replication and RNA polymerase II complex, were found to be enhanced in iPSC. Protein network analysis revealed that Pcna functions as a hub orchestrating complicated mechanisms including DNA replication, epigenetic inheritance (Dnmt1), and chromatin remodeling (Smarcc1) to reprogram MEF and maintain stemness of iPSC.



KEYWORDS: *quantitative proteomics, $^{16}O/^{18}O$ labeling, stem cell proteomics, reprogramming, UNiquant, Hdac1, Pcna*

INTRODUCTION

Recent creation of induced pluripotent stem cells (iPSC) by direct reprogramming of somatic cells was accomplished by introducing a defined set of factors into mouse embryonic fibroblasts (MEF) and human cells.[1−3] In addition to reprogram embryonic or adult fibroblasts, hepatocytes, neural progenitor cells, adipose stem cells, and B-lymphocytes were also successfully reprogrammed into pluripotent cells.[4] Remarkably, iPSC offer the option of autologous transplantation which could revolutionize the field of regenerative medicine by eliminating the technical issue of host rejection and the ethical issues surrounding human embryonic stem cells (ESC).[5] Nevertheless, gaps in the gene expression profiles between iPSC and ESC still exist.[2] In addition, several challenges still hamper the translational development of iPSC such as the tumorigenicity,[6] and retention and dependency of epigenetic memory from the original cell type.[7]

Although the occurrence of iPSC using a defined set of transcription factors is groundbreaking,[2] the four retroviral factor-mediated (Oct3/4, Sox2, Klf4, and c-Myc) reprogramming process is elaborate, labor intensive, and inefficient.[8] Recent studies indicate that reprogramming efficiency can be improved by small molecules, that is, DNA methyltransferase (DNMT) inhibitors and histone deacetylase (HDAC) inhibitors, which alter chromatin structures, and proteins that facilitate chromatin-remodeling.[9−11] To discover differences between iPSC and their original somatic cells at a system level, analyses of genome-wide regulation of the programming process have been made.[9,12] However, understanding of key intracellular signaling pathways and protein targets that control development of iPSC from somatic cells at proteome level is still lacking.

Mass spectrometry-based proteomics technology has been increasingly utilized for the study of stem cell biology.[13] In particular, the integration of liquid chromatography coupled with tandem mass spectrometry (LC−MS/MS), for protein identification and quantitation, has significantly expanded the

Received: June 29, 2011
Published: February 29, 2012

ACS Publications © 2012 American Chemical Society

dx.doi.org/10.1021/pr300155r | J. Proteome Res. 2012, 11, 2091−2102

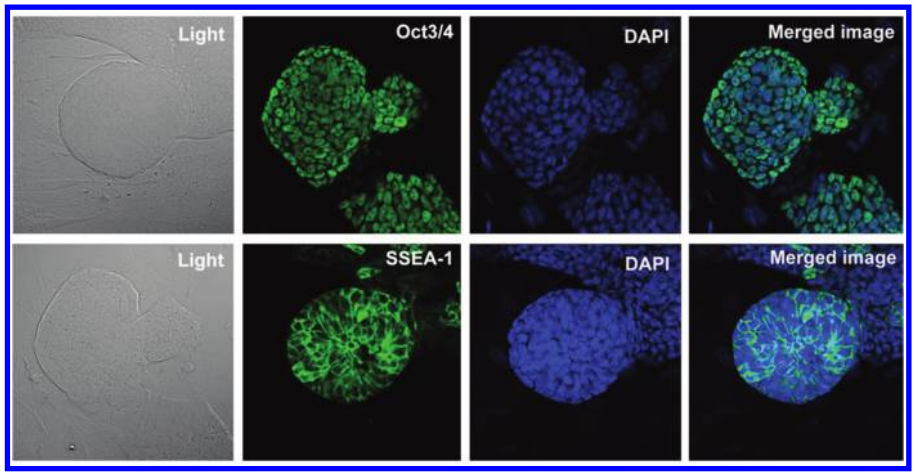


Figure 1. The generation of ES-like cells by overexpression of oct3/4, sox2, klf4, and c-myc in MEF derived from mouse embryos. Two separate colonies are shown, one in each set of panels (top and bottom). The second image in the top panel shows the expression of the stem cell transcription factor, Oct4 (green), following retroviral overexpression; the bottom panel shows the expression of SSEA-1 (green). The third image in each panel shows DAPI nuclear staining (blue) of a fucosylated derivative of type-2 polylactosamine that appears during late cleavage stages of mouse embryos. Merging of DAPI with Oct4 and SSEA-1 is displayed in the fourth image of each panel.

scale of proteomics studies.[14] In combination with quantitative stable isotope labeling (SIL), mass spectrometry can now be used to quantify and compare thousands of proteins from multiple samples. With the use of this technique, dynamic processes in the development and reprogramming of iPSC can be investigated.

In SIL methodologies, stable isotopes are usually introduced into selected proteins or peptides during sample preparation either chemically, enzymatically, or metabolically.[15,16] For instance, stable isotope labeling by amino acids in cell culture (SILAC) has been used to identify factors that mediate the efficiency of reprogramming and to study the effect of feeder cells during the co-culture of MEF with the membrane proteomes of self-renewing and differentiating human ESC.[11,17,18] Another SILAC-based proteomic study revealed that a small group of proteins related to metabolism, antigen processing and cell adhesion were differentially expressed between human iPSC and ESC.[19] In addition, the iTRAQ technique (a chemistry labeling method) has been applied to study human ESC during induced differentiation, and to study the proteome differences between human ESC and embryonal carcinoma cells.[20,21] Of these SIL methods, the trypsin-catalyzed $^{16}O/^{18}O$ labeling approach is exceptional because of its simplicity, broad applicability, and capacity to be used for either small-scale biological studies or for large-scale clinical proteomics.[22,23] In the current study, we coupled $^{16}O/^{18}O$ labeling with multidimensional (MD) LC−MS/MS analysis to investigate protein expression profiles of MEF and MEF-derived iPSC by the four-factor-mediated nuclear reprogramming. UNiquant,[24] a new software tool which was recently introduced for SIL-based quantitation, was modified and applied to analyze $^{16}O/^{18}O$-based quantitative proteomics data obtained on LTQ-Orbitrap and Q-Exactive MS platforms. The locally weighted scatterplot smoothing (LOWESS) method was applied to correct the heavy/light (H/L) ratios that deviated from ideality. Our quantitative proteomic study reveals that some of key molecular events such as expression of DNA replication machinery, activation of Hdac1, and interaction between Ccnd1 and p21/p27 are involved in the reprogramming process.

MATERIALS AND METHODS

Development of MEF-Derived iPSC

MEF were isolated from C57BL/6J mice and maintained in DMEM containing 10% FBS supplemented with glutamine (2 mg/mL) and MEF nonessential amino acids. Cells grown in culture for 4 days were transfected with equal volumes of supernatants from retroviruses containing Oct4, Sox2, Klf4, and c-Myc expression vectors. Retrovirus addition was repeated twice (12 and 72 h) after the first infection. Cells were then transferred to plates preseeded with mitomycin C-treated MEF. The medium was changed to mouse ES culture medium 5 days after transfection. Embryoid bodies (iPSC) were collected and examined using Oct3/4 and SSEA-1 immunostaining (Figure 1).

Sample Preparation and $^{16}O/^{18}O$ Labeling

MEF and MEF-derived iPSC were harvested separately and lysed in 7 M urea, 2 M thiourea, and 50 mM ammonium bicarbonate, pH 8.3. The lysates were centrifuged at 13 000 rpm for 20 min and supernatants were collected. These supernatants were reduced with 10 mM dithiothreitol at 56 °C for 45 min then alkylated with 55 mM iodoacetamide in the dark at room temperature for 1 h. Then, samples were diluted 8-fold using 50 mM ammonium bicarbonate and digested with sequencing grade trypsin (1 μg/μL) at a trypsin/protein ratio of 1:50 (w/w) overnight at 37 °C. The resulting peptides were desalted with Sep-Pak Cartridges (C18 Plus from Waters, Milford, MA) and dried in a vacuum centrifuge. We followed a previously published protocol for $^{16}O/^{18}O$ labeling that minimized the "back-exchange" reaction.[25] Briefly, about 200 μg of dried peptides were resuspended in 100 μL of ^{18}O-water (purity >98%, Shanghai Research Institute of Chemical Industry, China) containing 50 mM ammonium bicarbonate, 10 mM calcium chloride, and trypsin (trypsin/peptides = 1:50, w/w). The sample was incubated at 37 °C for 5 h. Then, labeling reaction was terminated by boiling the sample for 10 min, adding 5 μL of formic acid to further inhibit any residual trypsin activity. With ^{18}O-water, this procedure introduced two atoms of ^{18}O into the C-terminus of essentially all of the heavy peptides. The labeling procedure was also performed in parallel using ^{16}O-water for the light peptides. We performed three biological replicates for the comparative proteomic study of MEF and

dx.doi.org/10.1021/pr300155r | J. Proteome Res. 2012, 11, 2091−2102

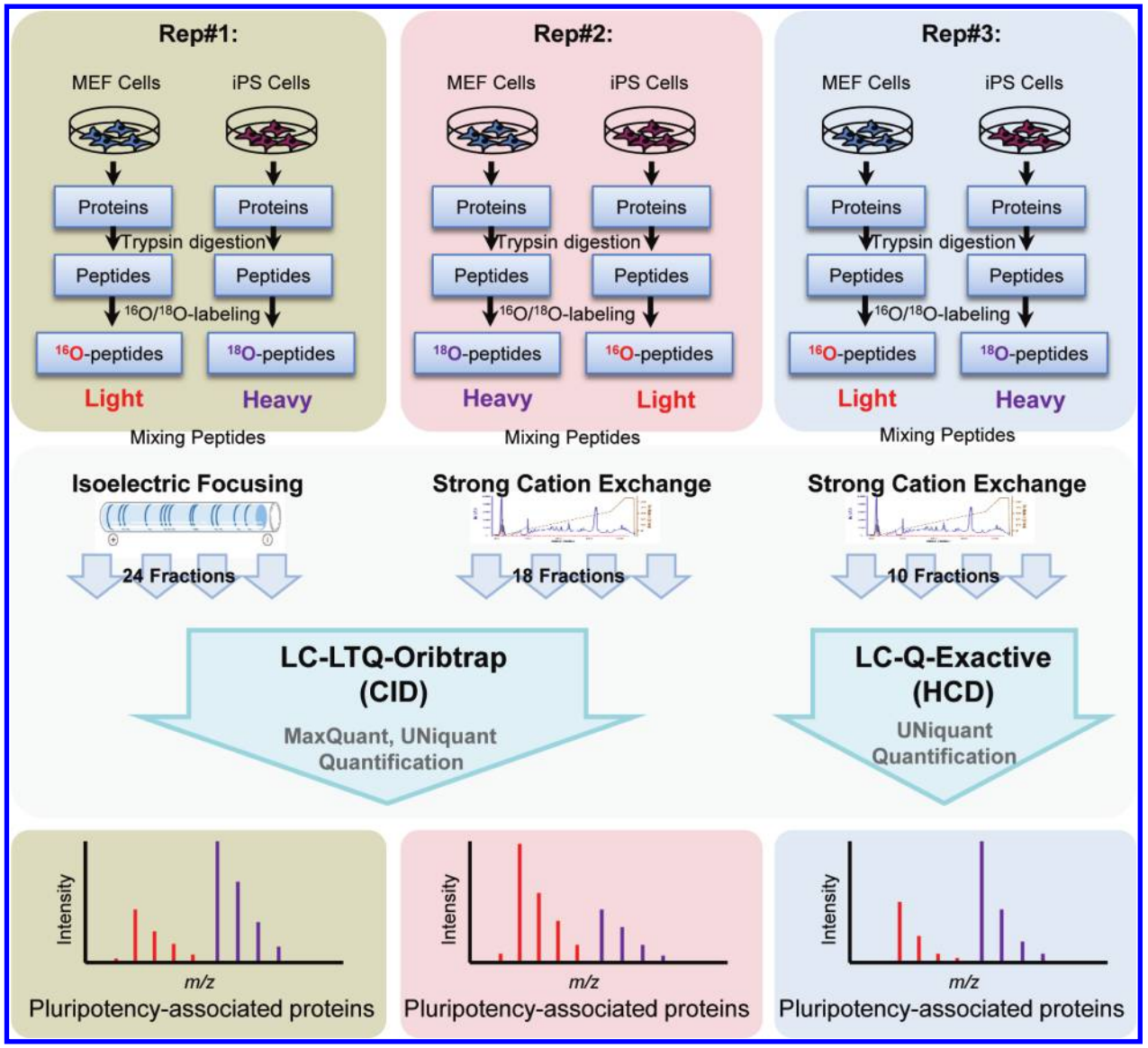


Figure 2. A diagram showing the proposed proteomics analysis of MEF and MEF-derived iPSC, using $^{16}O/^{18}O$-labeling, MDLC−MS/MS analysis and protein quantification. Samples extracted from three biological replicates of MEF and MEF-derived iPSC lysates are digested with trypsin, and followed by different strategies of $^{16}O/^{18}O$-labeling and MDLC−MS/MS analysis. Rep#1 and Rep#3: iPSC peptides were labeled with ^{18}O form to create heavy peptides; MEF peptides were processed in parallel with ^{16}O-water to create light peptides. Reverse $^{16}O/^{18}O$ labeling was done for Rep#2. Rep#1 and Rep#2 samples were analyzed by LC-LTQ-Orbitrap MS while Rep#3 was analyzed by LC-Q-Exactive MS. The MS spectra at bottom panels indicate that the pluripotency-associated proteins should increase in the heavy form of Rep#1 and Rep#3, and increase in the light form of Rep#2.

iPSC (Figure 2). In the first (Rep#1) and third (Rep#3) replicates, ^{16}O-water was used for labeling MEF peptides while ^{18}O-water was used for labeling iPSC peptides. In the second replicate (Rep#2), ^{16}O-water was used for labeling iPSC peptides while ^{18}O-water was used for labeling MEF peptides (reverse labeling). Finally, equal amounts of the ^{16}O- and ^{18}O-labeled peptides in each replicate were mixed together for MDLC−MS/MS analysis.

MDLC−MS/MS Analysis

Isoelectric focusing (IEF) was used as first-dimensional approach for separating peptides in Rep#1. The IEF was set up on a 24-well gel on the 3100 OFFGEL fractionator using an OFFGEL Room Temp Kit, High Resolution (Agilent Technology, Santa Clara, CA), following the manufacture's protocol. Peptides were separated into 24 fractions with a pH range of 3−10. Strong cation exchange (SCX) chromatography was used as first-dimensional approach to fractionate peptides in Rep#2 and Pep#3 as described previously.[24] SCX fractions were combined into 18 fractions for Rep#2, and 10 fractions for Rep#3. Each fraction from Rep1#1 and Rep#2 was analyzed by a nanoLC−MS/MS on a hybrid LTQ-Orbitrap-XL MS (Thermo Scientific, San Jose, CA). The LC system consisted of a IntegraFrit trap column (100 μm × 1.5 cm, New Objective, Woburn, MA) packed with Magic C18AQ resin (5 μm, 200 Å pore size; Michrom Bioresources, Auburn, CA), followed by a PicoFrit analytical column (75 μm × 27 cm, New Objective) packed with Magic C18AQ resin (5 μm, 100 Å pore size). The columns were connected to an Eksigent 2D nano-HPLC (Eksigent Technologies, Dublin, CA) with a vented system.[26] Sample was loaded onto the trap column and desalted at a flow rate of 3 μL/min. After sample desalting, the flow was directed

dx.doi.org/10.1021/pr300155r | J. Proteome Res. 2012, 11, 2091−2102

to the analytical column at 300 nL/min. A 150-min gradient was used for LC separation, starting from 5% acetonitrile (ACN)/0.1% trifluoroacetic acid (TFA) and progressing linearly to 35% ACN/0.1% TFA. In MS configuration, each MS cycle in the Orbitrap analyzer starts with a full MS scan (from m/z 400 to 1800). The five most intense ions from each full MS scan were selected for MS/MS fragmentation by collision-induced dissociation in the linear ion trap, with normalized collision energy of 35% and a minimum target intensity value of 10 000 counts. Ions selected for MS/MS were dynamically excluded from reanalysis for 75 s (excluded mass window width of −0.55 to ∼−1.55 m/z). Each peptide fraction from Rep#3 was analyzed by Easy-nanoLC coupled with Q-Exactive mass spectrometer (Thermo Scientific) equipped with a 75 μm × 15 cm, Magic C18 AQ LC column. The LC gradient was 5−35% ACN (0.1% formic acid) over 70 min at 400 nL/min. Q-Exactive mass spectrometer was operated in the data-dependent mode to automatically switch between MS and MS/MS acquisition. The top 12 most intense ions were sequentially isolated for high energy collisional dissociation (HCD). For full MS scans, resolution was set to 75 000 and target value to 1 000 000 with a maximum ion inject time of 60 ms and a scan range of 400−1800 m/z. For MS/MS scans, resolution was set at 17 500 and target value to 50 000 with a maximum ion inject time of 120 ms. The dynamic exclusion window is set at 20 s.

Peptide Identification and Quantification

DeconMSn[27] was used to determine the monoisotopic mass and correct the charge state of precursor ions from the MS raw data. DtaRefinery[28] was used to improve mass measurement accuracy for the precursor ions by modeling systematic errors based on preliminary peptide identifications. The mgf files were submitted to the Mascot search engine (Matrix Science, London, U.K.). MS data were searched against a mouse IPI database (version 3.52) with decoy sequences. Carbamidomethylation was set as the fixed modification. Variable modifications included modifications induced by sample preparation: oxidation of methionine, single and double ^{18}O replacement at the peptide C-terminal carboxyl group. The initial mass deviation tolerance for the precursor ion was set to 20 ppm and fragment ion tolerance was set to 0.5 Da for the LTQ-Orbirrap data, and 0.05 Da for the Q-Exactive data. A maximum of 2 missed cleavages was allowed in peptide identification.

UNiquant program was applied for protein quantitation for all three replicates. Briefly, intensities of the monoisotopic peaks of light (^{16}O-labeled) and heavy (single and double ^{18}O-labeled) peptide precursor ions were determined from their parent full mass spectra by UNiquant program using a MS/MS-directed quantitation strategy.[24] Peptide peak detection criteria are the same as previously described:[24] a signal-to-noise (S/N) cutoff of 2, and a mass accuracy cutoff of ±20 ppm between the calculated and observed precursor masses. A quality of peptide identification (QPI) was calculated for each identified peptide, based on the Mascot identification score of the peptide and the mass accuracy of its precursor ion. The single QPI was used to filter peptides and calculate the false discovery rate (FDR) of peptide identification using decoy database approach. For those cases where the precursor ion appeared more than once in different full mass spectra, intensities were summed. The isotopic ratios for such peptides are given by the sum of the heavy intensities divided by the sum of the light intensities. Peptides exclusively expressed in only one population (i.e., either MEF or iPSC) were exported in a separate table. To complement the quantitation results obtained by UNiquant, MaxQuant[29] (version 1.2.0.18, with Andromeda engine) program was also used for quantitation analysis for the MS data obtained from LTQ-Orbitrap MS. Parameters used for database search were the same for both MaxQuant and UNiquant. Other parameters followed the default setting of MaxQuant.

Postmeasurement Normalization of Quantified Peptides

LOWESS method was used to correct the H/L ratios of quantified peptides.[30] Specifically, this method was based on Minus-Add (M-A) plot of the peptide intensities for the heavy and light species:

$$M = \log_2(\mathrm{Int}_{\mathrm{heavy}}/\mathrm{Int}_{\mathrm{light}}) \quad (1)$$

$$A = \frac{1}{2}\log_{10}(\mathrm{Int}_{\mathrm{heavy}} \times \mathrm{Int}_{\mathrm{light}}) \quad (2)$$

where $\mathrm{Int}_{\mathrm{heavy}}$ is the intensity of the heavy species from a quantified peptide; $\mathrm{Int}_{\mathrm{light}}$ is the intensity of the corresponding light species of this peptide; M is the $\log_2$ H/L intensity ratio, and A is half of the $\log_{10}$ H × L intensity product of each quantified peptide. These M-A points were equally divided into 20 groups, based on their A-values. A linear regression line was obtained from the points in each group. Then, a fitted regression curve was obtained by connecting all these regression lines. Normalization was performed by subtracting the fitted LOWESS curve from the measured $\log_2$ H/L ratio in the M-A plot:

$$M' = \log_2(\mathrm{Int}_{\mathrm{heavy}}/\mathrm{Int}_{\mathrm{light}}) - c(A) = \log_2(\mathrm{Int}_{\mathrm{heavy}}/(k \times \mathrm{Int}_{\mathrm{light}})) \quad (3)$$

where $c(A)$ is the fitted LOWESS curve, which is a function of A. M' is the normalized log ratio of quantified peptides, which is obtained by subtracting the value of LOWESS fitted function from the measured log ratio at each value of A.

After LOWESS correction, relative abundance of each protein was calculated as the sum of corrected intensities for all the heavy peptides from this protein divided by the sum of corrected intensities for all the light peptides from the same protein.

Differentially Expressed Proteins between MEF and iPSC

Proteins that were identified and quantified in at least two out of the three replicates were retained for further analysis. For each protein, the geometric means of H/L ratios (average of $\log_2$-ratio) quantified by both UNiquant and MaxQuant were used. LIMMA (linear model for microarray data) approach with default setting was applied to calculate the statistical significance of having no-zero mean $\log_2$-ratio between MEF and iPSC for all identified and quantified proteins.[31] A protein ratio with p-value less than 0.05 and more than 2 fold-change (iPSC/MEF ratio >2 or <0.5) was considered as differentially expressed proteins with significance between MEF and iPSC.

Gene Ontology (GO), Pathway, and Protein−Protein Interaction Analysis

GO terms in cellular component and molecular function for the identified proteins were obtained from Ingenuity Pathway Analysis software. Pathway analysis was performed based on GeneBrower2 software search against the KEGG pathway database.[32] The significance score threshold for the enriched pathways was set to $p < 0.05$. The differentially expressed proteins were uploaded

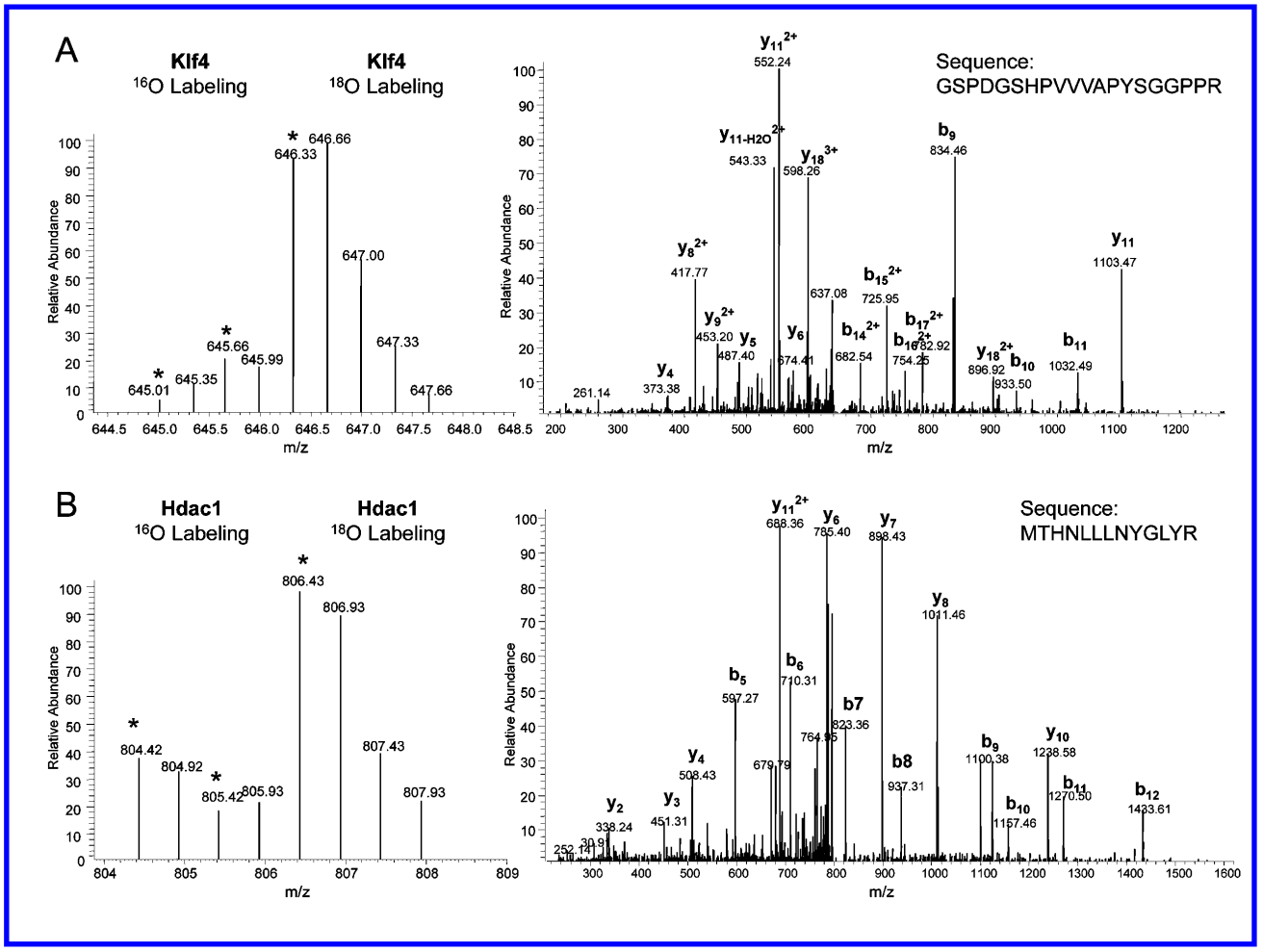


Figure 3. Two examples showing the identification and quantitation of proteins by LC-LTQ-Orbitrap analysis (Rep#1). The asterisk (*) labels the monoisotopic peak for a peptide without ^{18}O labeling, the peak with one ^{18}O atom, and the peak with two ^{18}O atoms. (A) Klf4 precursors were detected in the full MS scan. The monoisotopic peak of the heavy form (m/z = 646.33, charge = 3) is 1.32 m/z units higher than the monoisotopic peak for the light form (m/z = 645.01, charge = 3) as is expected for triply charged peptides with a mass difference of 4 Da. The precursor ion of heavy peptide was selected for collision-induced dissociation in the linear ion trap. From the MS/MS spectrum, the precursor ion was identified as "GSPDGSHPVVVAPYSGGPPR". (B) A pair of Hdac1 precursors was detected on the full MS scan. The monoisotopic peak of the light form (m/z = 804.42, charge = 2) is 2.01 m/z units lower than the monoisotopic peak for the heavy form (m/z = 806.43, charge = 2) as is expected for doubly charged peptides with a mass difference of 4 Da. The precursor ion of light peptide was selected for collision induced dissociation and it was identified as "MTHNLLLNYGLYR" from the MS/MS spectrum.

to the STRING databases of protein−protein interaction.[33] The interaction network was generated using the experimental knowledge component of the STRING database and visualized with Cytoscape software.

RESULTS

Strategy Overview

The objective of this study is to develop and apply proteomic approaches to investigate proteome differences between MEF and MEF-derived iPSC. To accomplish this goal, an integrated strategy that combines trypsin-catalyzed $^{16}O/^{18}O$ labeling, peptide fractionation by either off-gel IEF or SCX, and nanoLC-LTQ-Orbitrap/Q-Exactive analysis was designed (Figure 2). A total of 3 biological replicates were performed and an inverse labeling strategy was applied to one of the biological replicates (Rep#2). Both UNiquant and MaxQuant software were used for protein quantitation. A LOWESS method was used to normalize the peptide ratios obtained by UNiquant. Figure 3 shows two examples of protein identification and quantitation. Klf4 is a spiked protein induced by retrovirus infection in nuclear reprogramming. Figure 3A depicts the identification and quantification of Klf4 by MDLC−MS/MS analysis with $^{16}O/^{18}O$ labeling. Since the somatic cell such as MEF does not express Klf4, endogenous expression of this protein is very low in MEF. But during nuclear reprogramming, Klf4 expression dramatically increased in iPSC. Relative abundance of Klf4 in iPSC was 0.97 (monoisotopic peak), while the relative abundance of Klf4 in MEF was less than 0.03 (S/N ratio <2). We found that another important transcriptional regulator, histone deacetylase 1 (Hdac1), was also up-regulated during the reprogramming process. Expression of Hdac1 was observed in both MEF and iPSC, but with different quantity. Relative abundance of Hdac1 in iPSC was 1.00 (monoisotopic peak), and the relative abundance of Hdac1 in MEF was 0.40 (Figure 3B).

LOWESS Normalization

To address the issue of incomplete $^{16}O/^{18}O$ labeling due to the use of impure ^{18}O-water and to potential ^{18}O back-exchange, the

dx.doi.org/10.1021/pr300155r | J. Proteome Res. 2012, 11, 2091−2102

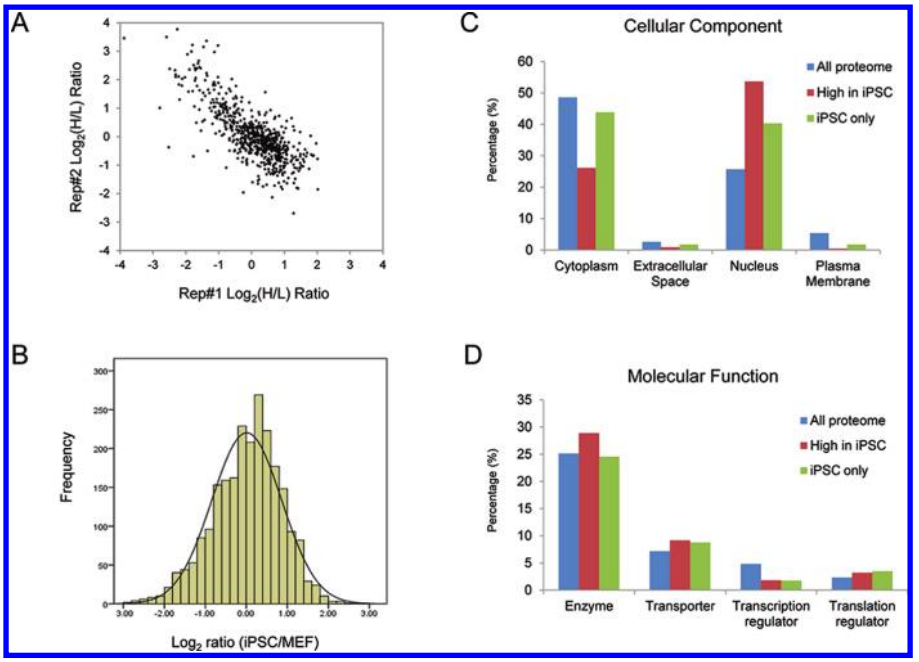


Figure 4. Proteomic analysis of MEF and MEF-derived iPSC. (A) Distribution of protein $\log_2$ heavy-versus-light (H/L) ratios for the quantified proteins from Rep#1 and Rep#2. Because of reverse labeling, protein $\log_2$ (H/L) ratios from these replicates negatively correlated with each other (Pearson correlation, $P < 0.001$, $R^2 = 0.59$). (B) Distribution of $\log_2$ ratios for the proteins quantified from more than two (out of three) predicates. Ratios are displayed as the average of relative iPSC/MEF intensities, with a fitted Gaussian curve. (C) Cellular component comparison for three categories of proteins: all identified proteins (designated as All proteome, $n = 2481$), proteins highly expressed in iPSC (designated as High in iPSC, iPSC/MEF > 2, $P < 0.05$, $n = 218$), and proteins only detected in iPSC (designated as iPSC only, $n = 57$). Cellular components are classified as cytoplasm, extracellular, nucleus, and plasma membrane. (D) Molecular function comparison for three categories of proteins: all identified proteins (designated as All proteome, $n = 2481$), proteins highly expressed in iPSC (designated as High in iPSC, iPSC/MEF > 2, $P < 0.05$, $n = 218$), and proteins only detected in iPSC (designated as iPSC only, $n = 57$). Molecular functions are classified as enzyme, transcription regulator, translation regulator, and transporter.

intensity of single ^{18}O-labeled (+2 Da) peaks was first corrected using a previously described equation,[34] and then the LOWESS method was adopted to normalize the peptide H/L ratios based on peptide intensities. An M-A plot of the peptide $\log_2$ H/L intensity ratios versus $\log_{10}$ (H × L)/2 intensity product, before correction, is shown in Supporting Information Figure 1A. Peptide H/L ratio is normally distributed, but the average H/L ratio is 0.763. The red line indicates the LOWESS curve, as a function of *A*, which is below the true $\log_2$-ratio of zero. Applying the LOWESS correction straightens the LOWESS curve into a straight line (Supporting Information Figure 1B). The corrected peptide H/L ratio is still normally distributed, and the LOWESS curve is independent of value of *A*. Here, the LOWESS method corrects the H/L ratios not only for insufficient ^{18}O labeling, but also for unequal mixing of heavy and light peptides during sample preparation.

Quantitative Proteomic Profiling of MEF and MEF-Derived iPSC

Using UNiquant and MaxQuant, we identified and quantified 2354 proteins from two or more replicates with FDR < 0.01 (Supporting Information Table 1). Protein H/L ratios from Rep#1 and Rep#2 (reverse $^{16}O/^{18}O$ labeling) were significantly negatively correlated (Pearson correlation, $P < 0.001$, $R^2 = 0.59$, Figure 4A). And proteins H/L ratios from Rep#1 and Rep#3 (forward $^{16}O/^{18}O$ labeling) were significantly positively correlated ($P < 0.001$, $R^2 = 0.39$). The results demonstrated a reasonably good reproducibility of our quantitative proteomics analysis. The protein $\log_2$-ratios (iPSC/MEF) displayed a Gaussian-like distribution (Figure 4B), centered at zero (indicating an iPSC/MEF ratio of 1.0) with a standard deviation of 0.866. Using LIMMA approach, we found 247 proteins were highly expressed in MEF (iPSC/MEF ratio < 0.5, *p*-value < 0.05), while 218 proteins were highly expressed in iPSC (iPSC/MEF ratio > 2, *p*-value <0.05). UNiquant also identified 70 proteins which were exclusively expressed in MEF, and 57 proteins which were only expressed in iPSC (FDR < 0.01, Supporting Information Table 1). The cellular components and molecular functions, associated with all identified proteins, were annotated by gene ontology (GO) analysis. In Figure 4C, proteins were categorized based on their primary localization such as in the extracellular space, plasma membrane, cytoplasm, or nucleus. Protein percentage in each category was designated for all identified proteins (All proteome, $n = 2481$), differentially expressed proteins in iPSC (High in iPSC, $n = 218$), and the proteins exclusively expressed in iPSC (iPSC only, $n = 57$). Strikingly, the proportion of nuclear proteins in the "High in iPSC" group (53.7%) and in the "iPSC only" group (40.4%) was higher than the proportion of nuclear proteins in the "All proteome" group (25.7%). But the percentage of cytoplasmic proteins in the "High in iPSC" group (23.1%) was lower than in the "All proteome" group (40.3%). Next, we compared the GO categories in molecular function for these proteins. As shown in Figure 4D, the percentage of enzyme and transcription regulators in the "High in iPSC" group is higher than the other two groups. However, nearly no difference in categories of molecular function was observed between "All proteome" group and "iPSC only" group.

dx.doi.org/10.1021/pr300155r | *J. Proteome Res.* 2012, 11, 2091−2102

Table 1. List of Nuclear Proteins Highly Expressed in iPSC or Only Expressed in iPSC

accession	gene symbol	gene name	ratio (iPSC/MEF)[a] Rep1	Rep2	Rep3	F.C.	*p*-value
		Proteins Highly Expressed in iPSC:					
IPI00314370	Mnat1	Menage a trois homologue 1, cyclin H assembly factor (*Xenopus laevis*)	2.41		9.96	4.90	0.010
IPI00473314	Lig1	Ligase I, DNA, ATP-dependent	3.10	4.92	5.36	4.34	0.000
IPI00669709	Pds5a	PDS5, regulator of cohesion maintenance, homologue A (*S. cerevisiae*)	3.62		3.69	3.66	0.003
IPI00665571	Rfc3	Replication factor C (activator 1) 3, 38 kDa	4.41		2.93	3.59	0.004
IPI00222509	Rcc2	Regulator of chromosome condensation 2	4.04	1.71	6.32	3.52	0.004
IPI00132957	Ddx56	DEAD (Asp-Glu-Ala-Asp) box polypeptide 56	2.65		4.43	3.42	0.006
IPI00122202	Ncapg	Non-SMC condensin I complex, subunit G	4.72		2.45	3.40	0.007
IPI00896002	Ccnd1	Cyclin D1	4.21		2.57	3.29	0.002
IPI00321154	Xrcc5	X-ray repair complementing defective repair in Chinese hamster cells 5 (double-strand-break rejoining)	3.37	2.19	4.77	3.28	0.002
IPI00462291	Hmgb2	High-mobility group box 2	3.66	1.59	5.80	3.23	0.006
IPI00109860	Rbm8a	RNA binding motif protein 8A		1.96	5.32	3.23	0.016
IPI00132417	Gtf2e1	General transcription factor IIE, polypeptide 1, alpha 56 kDa	3.67	2.45	3.57	3.18	0.001
IPI00342158	Nup210	Nucleoporin 210 kDa	1.69		5.94	3.17	0.026
IPI00317794	Ncl	Nucleolin	2.78	2.18	5.19	3.16	0.003
IPI00153743	Srsf7	Serine/arginine-rich splicing factor 7		3.62	2.52	3.02	0.007
IPI00114819	Wdr77	WD repeat domain 77	2.31		3.96	3.02	0.009
IPI00117932	Sin3a	SIN3 homologue A, transcription regulator (yeast)	4.38		1.83	2.83	0.020
IPI00407571	Ssrp1	Structure specific recognition protein 1	2.79	1.91	4.18	2.81	0.003
IPI00130246	Rrp1b	Ribosomal RNA processing 1 homologue B (*S. cerevisiae*)	2.54	1.94	4.42	2.80	0.004
IPI00110050	Cdk11a	Cyclin-dependent kinase 11A	1.46		5.31	2.79	0.040
IPI00310173	Msh6	Muts homologue 6 (*E. coli*)	1.37	3.22	4.76	2.76	0.010
IPI00313515	Pold1	Polymerase (DNA directed), delta 1, catalytic subunit 125 kDa	2.22	2.91	3.25	2.76	0.006
IPI00122223	Top2a	Topoisomerase (DNA) II alpha 170 kDa	2.00	6.20	1.62	2.72	0.015
IPI00130200	Uhrf1	Ubiquitin-like with PHD and ring finger domains 1	3.20	2.25		2.68	0.011
IPI00224682	Elp3	Elongation protein 3 homologue (*S. cerevisiae*)	1.86	1.88	5.12	2.62	0.011
IPI00311131	Rpa2	Replication protein A2, 32 kDa	2.05	2.82	3.09	2.62	0.002
IPI00315127	Rrm1	Ribonucleotide reductase M1	3.53	1.55	3.17	2.59	0.007
IPI00170345	Cdc73	Cell division cycle 73, Paf1/RNA polymerase II complex component, homologue (*S. cerevisiae*)	2.83	2.93	2.07	2.58	0.003
IPI00117371	Snrpf	Small nuclear ribonucleoprotein polypeptide F		2.72	2.45	2.58	0.011
IPI00458908	Prpf4	PRP4 premrna processing factor 4 homologue (yeast)	2.10	3.00	2.65	2.56	0.003
IPI00458056	Ncbp1	Nuclear cap binding protein subunit 1, 80 kDa	3.01		2.16	2.55	0.014
IPI00122400	Ddx20	DEAD (Asp-Glu-Ala-Asp) box polypeptide 20	3.14	2.35	2.23	2.54	0.003
IPI00317298	Snw1	SNW domain containing 1	1.72		3.75	2.54	0.026
IPI00187529	Polr3h	Polymerase (RNA) III (DNA directed) polypeptide H (22.9 kDa)		3.43	1.87	2.53	0.020
IPI00127415	Npm1	Nucleophosmin 1	3.22	1.48	3.38	2.53	0.008
IPI00117016	Mcm4	Minichromosome maintenance complex component 4	2.17	2.73	2.68	2.51	0.002
IPI00551454	Pdcd11	Programmed cell death 11	2.73	1.83	3.18	2.51	0.004
IPI00229787	Rnaseh2a	Ribonuclease H2, subunit A	2.34		2.69	2.51	0.013
IPI00127554	Emg1	EMG1 nucleolar protein homologue (*S. cerevisiae*)	2.13	4.20	1.73	2.49	0.008
IPI00113870	Pcna	Proliferating cell nuclear antigen	2.42	1.66	3.82	2.48	0.009
IPI00112473	Parp1	Poly (ADP-ribose) polymerase 1		1.87	3.28	2.47	0.021
IPI00123881	Mcm6	Minichromosome maintenance complex component 6	2.87	2.03	2.55	2.46	0.003
IPI00187434	Dut	Deoxyuridine triphosphatase		2.09	2.89	2.46	0.016
IPI00225633	Sub1	SUB1 homologue (*S. cerevisiae*)	1.81		3.31	2.45	0.023
IPI00170008	Snrpa1	Small nuclear ribonucleoprotein polypeptide A′	1.88	3.62	2.14	2.44	0.006
IPI00309398	Mcm5	Minichromosome maintenance complex component 5	2.79	1.61	3.06	2.40	0.006
IPI00121135	Srsf2	Serine/arginine-rich splicing factor 2		1.51	3.79	2.39	0.040
IPI00126396	Mcm7	Minichromosome maintenance complex component 7	3.72	1.58	2.27	2.37	0.009
IPI00225912	Mki67ip	MKI67 (FHA domain) interacting nucleolar phosphoprotein	2.71	2.05	2.39	2.37	0.003
IPI00115650	Cacybp	Calcyclin binding protein	2.65	1.76	2.84	2.37	0.005
IPI00120691	Ddx21	DEAD (Asp-Glu-Ala-Asp) box polypeptide 21	2.60	3.49	1.38	2.32	0.012
IPI00125662	Smarcc1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily c, member 1	3.09	1.36	2.92	2.31	0.012
IPI00112645	Rrm2	Ribonucleotide reductase M2	2.51		2.11	2.30	0.019
IPI00118158	Msh2	Muts homologue 2, colon cancer, nonpolyposis type 1 (*E. coli*)	3.46		1.52	2.29	0.041
IPI00133880	Rnf2	Ring finger protein 2	1.85		2.80	2.28	0.024

dx.doi.org/10.1021/pr300155r | *J. Proteome Res.* 2012, 11, 2091−2102

Table 1. continued

accession	gene symbol	gene name	ratio (iPSC/MEF)[a] Rep1	Rep2	Rep3	F.C.	*p*-value
		Proteins Highly Expressed in iPSC:					
IPI00131725	Npm3	Nucleophosmin/nucleoplasmin 3	3.63	2.41	1.35	2.28	0.015
IPI00124284	Polr2h	Polymerase (RNA) II (DNA directed) polypeptide H	2.32		2.23	2.27	0.019
IPI00469323	Dnmt1	DNA (cytosine-5-)-methyltransferase 1	2.77	2.13	1.99	2.27	0.005
IPI00229397	Smc4	Structural maintenance of chromosomes 4	1.78		2.78	2.23	0.028
IPI00127071	Ddx41	DEAD (Asp-Glu-Ala-Asp) box polypeptide 41	2.49		1.96	2.21	0.024
IPI00133985	Ruvbl1	Ruvb-like 1 (*E. coli*)	2.30	1.85	2.38	2.16	0.006
IPI00108338	Mcm3	Minichromosome maintenance complex component 3	3.49	0.98	2.86	2.14	0.039
IPI00463367	Usp7	Ubiquitin specific peptidase 7 (herpes virus-associated)	2.02	2.38	1.95	2.11	0.007
IPI00114232	Hdac1	Histone deacetylase 1	1.82	1.78	2.85	2.10	0.020
IPI00123870	Smc1a	Structural maintenance of chromosomes 1A	1.47		2.97	2.09	0.050
IPI00224053	Ncaph	Non-SMC condensin I complex, subunit H	2.04	2.12		2.08	0.029
IPI00132340	Srsf9	Serine/arginine-rich splicing factor 9	1.02	3.02	2.83	2.06	0.037
IPI00132481	Rfc5	Replication factor C (activator 1) 5, 36.5 kDa	3.91	1.18	1.87	2.05	0.037
IPI00126716	Eif4a3	Eukaryotic translation initiation factor 4A3	1.84	2.22	2.08	2.04	0.008
IPI00323660	Actl6a	Actin-like 6A	2.21	1.90	2.00	2.03	0.008
IPI00116711	Rngtt	RNA guanylyltransferase and 5′-phosphatase	2.49		1.64	2.02	0.041
IPI00131067	Mbd3	Methyl-cpg binding domain protein 3	2.44	1.22	2.77	2.02	0.023
IPI00177202	Nup43	Nucleoporin 43 kDa	2.53	1.79	1.82	2.02	0.010
IPI00129468	Cbx3	Chromobox homologue 3	1.41	1.78	3.27	2.01	0.023
IPI00122696	Rbbp4	Retinoblastoma binding protein 4		2.34	1.72	2.01	0.040
IPI00136054	Lsm2	LSM2 homologue, U6 small nuclear RNA associated (*S. cerevisiae*)	2.36	1.29	2.65	2.01	0.020
IPI00123755	Cbx5	Chromobox homologue 5	1.92	2.43	1.72	2.00	0.011
IPI00311509	Aaas	Achalasia, adrenocortical insufficiency, alacrimia	1.86	2.16		2.00	0.035
		Proteins Exclusively Expressed in iPSC:[b]					
IPI00115987	Nasp	Nuclear autoantigenic sperm protein (histone-binding)	+	+	+		
IPI00119201	Ngdn	Neuroguidin, EIF4E binding protein	+		+		
IPI00719871	Nolc1	Nucleolar and coiled-body phosphoprotein 1	+		+		
IPI00459636	Sf3b1	Splicing factor 3b, subunit 1, 155 kDa		+	+		
IPI00120384	Klf4	Kruppel-like factor 4 (gut)	+		+		
IPI00221543	Pou2f1	POU class 2 homeobox 1	+		+		
IPI00122714	Tcf3	Transcription factor 3	+		+		

[a]Ratios are relative protein expression of iPSC/MEF after LOWESS correction in each replicates. The fold-change (F.C.) is the average ratio of iPSC/MEF (geometric mean) of the replicates. [b]For the proteins only expressed in iPSC proteome, a "+" indicates that this protein was detected in the corresponding replicate.

Proteins Upregulated in iPSC during Nuclear Reprogramming

In Supporting Information Table 1, the identified and quantified proteins from more than two biological replicates were listed. Different regulatory proteins expressed in either MEF or in iPSC, especially the transcriptional factors, have different cellular consequences during stem cell reprogramming. Table 1 tabulated a subset of the nuclear proteins which were highly expressed in iPSC compared to MEF, or exclusively expressed in iPSC. Mnat1 is a subunit of TFIIH kinase that is required for RNA Polymerase II (Pol II) transcription and for nucleotide excision repair.[35] Mnat1 and RNA Polymerase II subunit Polr2h were upregulated by 4.9- and 2.28-fold in iPSC compared to MEF, respectively. Cyclin D1 (Ccnd1), which functions as a promoter of G1-to-S progression in the cell cycle, was upregulated in iPSC by 3.29-fold. Smarcc1 and Actl6a both subunits of SWI/SNF complex which regulate transcription of certain genes by remodeling the chromatin structure around those genes, were upregulated in iPSC, by 2.31- and 2.03-fold, respectively. Expression of DNA methyltransferase 1 (Dnmt1) was enhanced by 2.27-fold in iPSC. Interestingly, we found that the nuclear proteins in several protein families were also upregulated in iPSC (ratio in parentheses), such as DEAD (Asp-Glu-Ala-Asp) box containing proteins Ddx56 (3.42), Ddx20 (2.54), Ddx21 (2.32), and Ddx41 (2.21); proteins involved in spliceosomes Srsf7 (3.02), Srsf6 (2.18), Srsf9 (2.06), Snrpa1 (2.44), and Snrpf (2.58); and proteins involved in the mini-chromosome maintenance complex (MCM) proteins Mcm4 (2.52), Mcm6 (2.46), Mcm5 (2.40), Mcm7 (2.37), and Mcm3 (2.14).

KEGG pathway analysis was performed on the upregulated nuclear proteins. The enriched pathways are listed in Table 2. As expected, proteins involved in DNA replication and proteins from the mismatch repair machinery were significantly overexpressed in iPSC. In the eukaryotic replication complex, the mini chromosome maintenance (MCM) proteins (Mcm3/4/5/6/7) function as the DNA helicase, Pcna (proliferating cell nuclear antigen) works as the DNA clamp, and replication factor C (RFC) proteins (Rfc3/5) work as the DNA clamp loader. All these proteins facilitate the function of the DNA polymerase complex (Pold1) during DNA replication. Furthermore, the increased expression of RNA polymerase associated genes (Polr2h, Polr3h) indicated that gene transcription was enhanced in iPSC. We also observed genes involved in Notch (Hdac1, Snw1) and Wnt (Cacybp, Ccnd1, Ruvbl1) signaling pathways were upregulated in iPSC (Table 2).

Table 2. KEGG Pathway Analysis for the Upregulated Nuclear Proteins in iPSC

KEGG ID	no. genes.	pathway name	*p*-value	included genes
mmu03430	8	Mismatch repair	1.80×10^{-13}	Lig1, Msh2, Msh6, Pcna, Pold1, Rfc3, Rfc5, Rpa2
mmu03030	12	DNA replication	3.47×10^{-13}	Lig1, Mcm3, Mcm4, Mcm5, Mcm6, Mcm7, Pcna, Pold1, Rfc3, Rfc5, Rnaseh2a, Rpa2
mmu03420	7	Nucleotide excision repair	8.12×10^{-13}	Lig1 ,Mnat1, Pcna, Pold1, Rfc3, Rfc5, Rpa2
mmu04110	9	Cell cycle	1.10×10^{-11}	Ccnd1, Hdac1, Mcm3, Mcm4, Mcm5, Mcm6, Mcm7, Pcna, Smc1a
mmu03410	5	Base excision repair	3.92×10^{-6}	Lig1, Parp1, Parp3, Pcna, Pold1
mmu00240	7	Pyrimidine metabolism	5.78×10^{-4}	Dut, Pold1, Polr2h, Polr3h, Rpa2, Rrm1, Rrm2
mmu03020	4	RNA polymerase	8.04×10^{-3}	Polr2h, Polr3h, Rpa2, Mnat1
mmu00790	2	Folate biosynthesis	8.64×10^{-3}	Ddx41,Ddx56
mmu00230	6	Purine metabolism	1.48×10^{-2}	Pold1, Polr2h, Polr3h, Rpa2, Rrm1, Rrm2
mmu04330	2	Notch signaling pathway	3.70×10^{-2}	Hdac1, Snw1
mmu04310	3	Wnt signaling pathway	5.08×10^{-2}	CacyBP, Ccnd1, Ruvbl1

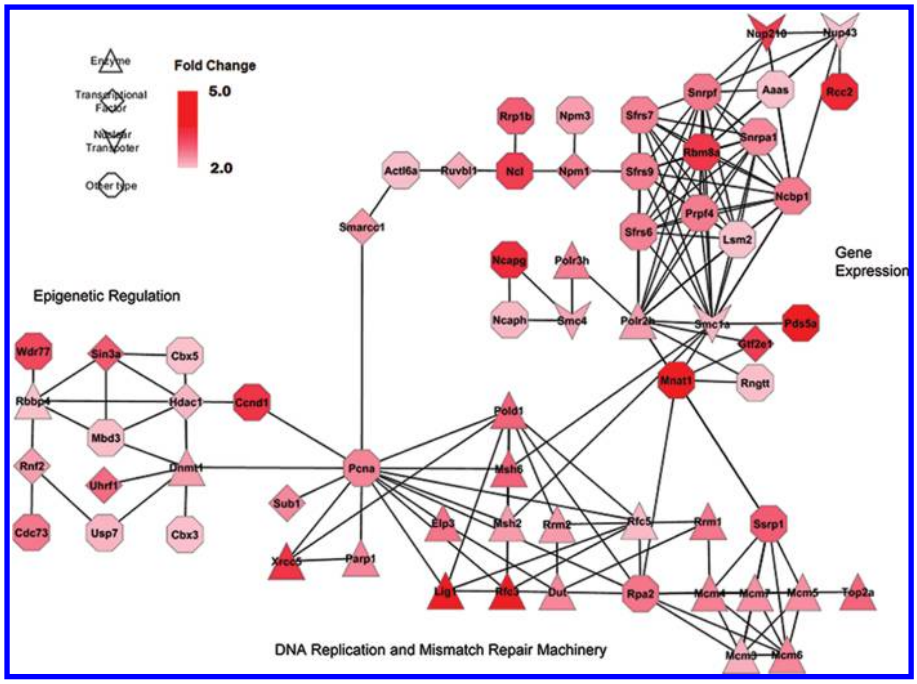


Figure 5. Protein−protein interaction network of upregulated nuclear proteins in iPCSs. All interactions were based on the experimental knowledge in STRING database. Relative expression (iPSC versus MEF) of each protein is displayed according to the color scale shown.

Analysis of protein−protein interaction was performed based on the evidence from experimental knowledge for the upregulated nuclear proteins (Figure 5). Interaction analysis revealed Pcna occupied a pivot position in this network, and that 12 proteins (Dnmt1, Smarcc1, Ccnd1, Sub1, Pold1, Msh6, Msh2, Rfc3, Elp3, Lig1, Parp1, and Xrcc5) have experimentally validated interactions with Pcna. Most of these proteins function in the DNA replication and mismatch repair machinery (Figure 5), which is consistent with the result of pathway analysis. Another important regulation revealed by protein interaction analysis was increased expression of RNA Pol II. It was reported that Mnat1 activates several cyclin-associated kinases (CAKs) associated with TFIIH to activate transcription by RNA Pol II.[35] Upregulation of Polr2h in iPSC is associated with upregulation of maintenance of chromosomes (SMC) proteins (Smc1/4), spliceosome complex (Srsf6/7/9, Snrpa1, Snrpf), non-SMC condensin I complex (Ncapg, Ncaph), nuclear chaperones (Npm1/3), and nucleotransporters (Nup210, Nup43) in iPSC compared to MEF. Altogether, our data delineated a mechanism of enhanced gene transcription by RNA Pol II in iPSC.

We also found that proteins in cell cycle control were significantly upregulated in iPSC, i.e., Ccnd1 (3.29) was upregulated in iPSC. It was reported that Hdac1 promotes the G1-to-S progression of the cell cycle by repressing the expression of Kip family proteins: p21 (Cdkn1a) and p27 (Cdkn1b).[36] Therefore, we checked the protein expression of p21, p27, and Ccnd1 by immunobloting. As shown in Figure 6, expression of Oct3/4 (positive control) and Ccnd1 was upregulated, while expression of p21 and p27 disappeared in iPSC during reprogramming. This is consistent with the upregulation of Hdac1 that we observed.

DISCUSSION

Significant progress has recently been made in stem cell proteomics in terms of evaluating the proteome-wide regulation of proteins during formation of the ESC-like state from somatic cells and during differentiation to specific progenitors in different tissues.[37,38] Fundamentals of these experiments are based on the comprehensive and accurate quantification of proteins from complex proteome mixtures using mass spectrometry. Here, we described an integrated and alternative approach for quantitative proteomic analysis of MEF and MEF-derived

dx.doi.org/10.1021/pr300155r | J. Proteome Res. 2012, 11, 2091−2102

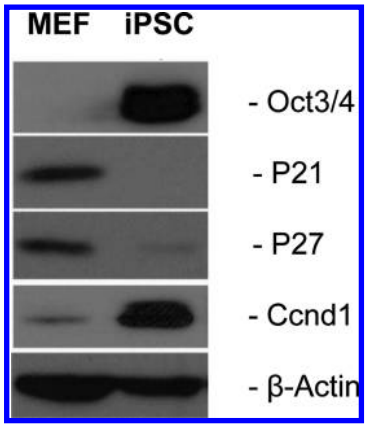


Figure 6. Protein expression of Oct3/4, p21, p27, and Ccnd1 in MEF and MEF-derived iPSC by Western blot analysis. Expression of β-Actin is used as the loading control.

iPSC. Our approach provides two significant advantages. First, compared to SILAC and any other metabolic labeling methods as well as chemical methods such as iTRAQ, ^{18}O-labeling is simple and broadly applicable for any digestible proteome samples on a broad range of MS instruments. Second, our UNiquant pipeline for quantitative proteomic analysis is versatile and can work with a wide spectrum of search engines and stable isotope labeling approaches.[24]

To achieve accurate quantitation and to correct for the ^{16}O/^{18}O ratio for incomplete labeling, an equation described by Yao et al. was applied.[34] Previously, three different equations for ^{16}O/^{18}O ratio correction were developed (Yao et al.,[34] Johnson et al.,[39] and Qian et al.[40]). We applied and compared these three different correction methods in our study and found that Johnson's and Qian's equations were quite similar and that both could correct the ^{16}O/^{18}O ratios without the need of ratio normalization. However, after correction, the $\log_2$-ratio distribution changes from Gaussian to a slightly distorted Gaussian-like shape (Supporting Information Figure 2). While Yao's equation did not change the shape of $\log_2$-ratio distribution, the median of the distribution was not zero after the correction. We chose to use Yao's method because the distribution remained Gaussian, and added a nonlinear normalization method, LOWESS, to normalize the H/L ratios (assuming that the $\log_2$-ratio distribution was centered on zero). The LOWESS method is robust, simple-to-use, and can correct the H/L ratios for deviations caused by other factors, such as unequal mixing of the heavy and light peptides. Both MaxQuant and UNiquant were used to analyze MS data generated from the LTQ-Orbitrap platform because these two programs provided complementary coverage of the quantified proteome as we demonstrated previously.[24]

Although reprogramming from somatic cell to pluripotent cells is initiated by a few transcriptional factors, regulation of this process affects thousands of genes and proteins. For the nuclear proteins that were significantly upregulated in iPSC versus MEF, 71 out of 117 (60.7%) proteins were also confirmed by another comparative proteomic study of human iPSC and their precursor fibroblast cells.[19] A key regulation of "stemness" in iPSC is cell-renewal, which requires enhancement of DNA replication. In our study, we observed that the DNA replication components were upregulated in iPSC. The reprogramming process is also regulated by gene transcription (RNA Pol II), cell cycle activation, and epigenetic regulators. Moreover, Smarcc1/Baf155 chromatin remodeling complex plays an important role in these processes. It was reported that overexpression of this complex during four factor-mediated reprogramming increases the kinetics of Oct4, Nanog, and Rex1 promoter demethylation.[11]

HDACs regulate gene expression by deacetylating histones and also modulate the acetylation of a number of nonhistone proteins, thus, impinging on various cellular processes. Hdac1 was found to be increased 2.10-fold in iPSC compared to MEF. Hdac1 act in concert to promote the G1/S progression of the cell cycle by repressing the expression of *Cip/Kip* family proteins (p21 and p57), which is an important process for cell proliferation and self-renewal of stem cells.[36] It was also reported that histone H3/H4 in the p21 and p16 promoter regions can be directly deacetylated by Hdac1 thereby repressing transcription.[41,42] Overexpression of Hdac1 can also repress the major inhibitors of Ccnd1, which can arrest the cell progress at the G1/S check-point. The primary function of Ccnd1 is to start the G1 phase of the cell cycle by binding to Cdk4/6. Ccnd1 can also bind to the upstream regulatory region of the Notch1 gene, where it serves to recruit CREB binding protein (CBP) associated histone acetylation for gene transcription.[43,44]

Another important issue for generating iPSC is to improve its low reprogramming efficiency. For instance, four factor-mediated generation of iPSC can be achieved in ~2−3 weeks, and <5% of the clones undergo reprogramming.[8] Recent studies revealed that targeting p53/p21 using small interfering RNA, or genetic inhibition of the *Ink4/Arf* locus (encoding p15, p16, and p19) has a profound positive effect on iPSC generation.[45,46] Interestingly, small molecular chemicals that inhibit DNMTs or HDACs also appear to improve reprogramming efficiency.[9,10] In our study, we found that both Dnmt1 and Hdac1 were upregulated in iPSC. Inhibition of these proteins will change the chromatin modification and probably start the transcription of essential genes of differentiation. However, because Hdac1 is also necessary to inhibit p21, we suggest that inhibition of HDACs should be combined with inhibition of p21, to keep the stemness and enhance the reprogramming efficiency of iPSC. Finally, protein network analysis reveals that Pcna functions as a hub orchestrating complicated mechanisms including DNA replication, epigenetic inheritance (Dnmt1), and chromatin remodeling (Smarcc1, Actl6a) to reprogram MEF and maintain stemness of iPSC.[47]

■ CONCLUSION

An integrated quantitative proteomics platform was developed and applied to understanding the cellular pathways associated with the reprogramming of MEF into iPSC. This method consisted of ^{16}O/^{18}O labeling, MDLC−MS/MS, and UNiquant pipeline for data analysis. A snapshot of the expression change of key target proteins during the reprogramming and self-renewal process was captured in our study.

■ 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

*Dr. Shi-Jian Ding, 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 declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Dr. Lawrence Schopfer for the editing of this manuscript and Jim Keagy from Thermo Scientific for arranging demo experiments on the Q-Exactive mass spectrometer. Part of the research is supported by the Environmental Molecular Sciences Laboratory (EMSL) user framework. This work was financially supported by NEHHS LB606 (S.-J.D). X.H. was supported by a scholarship from the Chinese Scholarship Council.

ABBREVIATIONS

iPSC, induced pluripotent stem cells; MEF, mouse embryonic fibroblasts; ESC, embryonic stem cells; DNMT, DNA methyltransferase; HDAC, histone deacetylase; LC−MS/MS, liquid chromatography coupled with tandem mass spectrometry; SIL, stable isotope labeling; SILAC, stable isotope labeling by amino acids in cell culture; MD, multidimensional; H/L, heavy/light; IEF, isoelectric focusing; S/N, signal-to-noise; QPI, quality of peptide identification; FDR, false discovery rate; LIMMA, linear model for microarray data; LOWESS, locally weighted scatterplot smoothing; M-A, minus-add; DE, differentially expressed; GO, gene ontology; MCM, mini chromosome maintenance; RFC, replication factor C

REFERENCES

(1) Takahashi, K.; Tanabe, K.; Ohnuki, M.; Narita, M.; Ichisaka, T.; Tomoda, K.; Yamanaka, S. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **2007**, *131* (5), 861−72.

(2) Takahashi, K.; Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **2006**, *126* (4), 663−76.

(3) Yu, J.; Vodyanik, M. A.; Smuga-Otto, K.; Antosiewicz-Bourget, J.; Frane, J. L.; Tian, S.; Nie, J.; Jonsdottir, G. A.; Ruotti, V.; Stewart, R.; Slukvin, I. I.; Thomson, J. A. Induced pluripotent stem cell lines derived from human somatic cells. *Science* **2007**, *318* (5858), 1917−20.

(4) Selvaraj, V.; Plane, J. M.; Williams, A. J.; Deng, W. Switching cell fate: the remarkable rise of induced pluripotent stem cells and lineage reprogramming technologies. *Trends Biotechnol.* **2010**, *28* (4), 214−23.

(5) Amabile, G.; Meissner, A. Induced pluripotent stem cells: current progress and potential for regenerative medicine. *Trends Mol. Med.* **2009**, *15* (2), 59−68.

(6) Aoi, T.; Yae, K.; Nakagawa, M.; Ichisaka, T.; Okita, K.; Takahashi, K.; Chiba, T.; Yamanaka, S. Generation of pluripotent stem cells from adult mouse liver and stomach cells. *Science* **2008**, *321* (5889), 699−702.

(7) Kim, K.; Doi, A.; Wen, B.; Ng, K.; Zhao, R.; Cahan, P.; Kim, J.; Aryee, M. J.; Ji, H.; Ehrlich, L. I.; Yabuuchi, A.; Takeuchi, A.; Cunniff, K. C.; Hongguang, H.; McKinney-Freeman, S.; Naveiras, O.; Yoon, T. J.; Irizarry, R. A.; Jung, N.; Seita, J.; Hanna, J.; Murakami, P.; Jaenisch, R.; Weissleder, R.; Orkin, S. H.; Weissman, I. L.; Feinberg, A. P.; Daley, G. Q. Epigenetic memory in induced pluripotent stem cells. *Nature* **2010**, *467* (7313), 285−90.

(8) Wernig, M.; Lengner, C. J.; Hanna, J.; Lodato, M. A.; Steine, E.; Foreman, R.; Staerk, J.; Markoulaki, S.; Jaenisch, R. A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nat. Biotechnol.* **2008**, *26* (8), 916−24.

(9) Mikkelsen, T. S.; Hanna, J.; Zhang, X.; Ku, M.; Wernig, M.; Schorderet, P.; Bernstein, B. E.; Jaenisch, R.; Lander, E. S.; Meissner, A. Dissecting direct reprogramming through integrative genomic analysis. *Nature* **2008**, *454* (7200), 49−55.

(10) Huangfu, D.; Maehr, R.; Guo, W.; Eijkelenboom, A.; Snitow, M.; Chen, A. E.; Melton, D. A. Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nat. Biotechnol.* **2008**, *26* (7), 795−7.

(11) Singhal, N.; Graumann, J.; Wu, G.; Arauzo-Bravo, M. J.; Han, D. W.; Greber, B.; Gentile, L.; Mann, M.; Scholer, H. R. Chromatin-remodeling components of the BAF complex facilitate reprogramming. *Cell* **2010**, *141* (6), 943−55.

(12) Lu, R.; Markowetz, F.; Unwin, R. D.; Leek, J. T.; Airoldi, E. M.; MacArthur, B. D.; Lachmann, A.; Rozov, R.; Ma'ayan, A.; Boyer, L. A.; Troyanskaya, O. G.; Whetton, A. D.; Lemischka, I. R. Systems-level dynamic analyses of fate change in murine embryonic stem cells. *Nature* **2009**, *462* (7271), 358−62.

(13) Baharvand, H.; Fathi, A.; van Hoof, D.; Salekdeh, G. H. Concise review: trends in stem cell proteomics. *Stem Cells* **2007**, *25* (8), 1888−903.

(14) Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422* (6928), 198−207.

(15) Afkarian, M.; Bhasin, M.; Dillon, S. T.; Guerrero, M. C.; Nelson, R. G.; Knowler, W. C.; Thadhani, R.; Libermann, T. A. Optimizing a proteomics platform for urine biomarker discovery. *Mol. Cell. Proteomics* **2010**, *9* (10), 2195−204.

(16) Mann, M. Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell Biol.* **2006**, *7* (12), 952−8.

(17) Graumann, J.; Hubner, N. C.; Kim, J. B.; Ko, K.; Moser, M.; Kumar, C.; Cox, J.; Scholer, H.; Mann, M. Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins. *Mol. Cell. Proteomics* **2008**, *7* (4), 672−83.

(18) Prokhorova, T. A.; Rigbolt, K. T.; Johansen, P. T.; Henningsen, J.; Kratchmarova, I.; Kassem, M.; Blagoev, B. Stable isotope labeling by amino acids in cell culture (SILAC) and quantitative comparison of the membrane proteomes of self-renewing and differentiating human embryonic stem cells. *Mol. Cell. Proteomics* **2009**, *8* (5), 959−70.

(19) Munoz, J.; Low, T. Y.; Kok, Y. J.; Chin, A.; Frese, C. K.; Ding, V.; Choo, A.; Heck, A. J. The quantitative proteomes of human-induced pluripotent stem cells and embryonic stem cells. *Mol. Syst. Biol.* **2011**, *7*, 550.

(20) Yocum, A. K.; Gratsch, T. E.; Leff, N.; Strahler, J. R.; Hunter, C. L.; Walker, A. K.; Michailidis, G.; Omenn, G. S.; O'Shea, K. S.; Andrews, P. C. Coupled global and targeted proteomics of human embryonic stem cells during induced differentiation. *Mol. Cell. Proteomics* **2008**, *7* (4), 750−67.

(21) Chaerkady, R.; Kerr, C. L.; Kandasamy, K.; Marimuthu, A.; Gearhart, J. D.; Pandey, A. Comparative proteomics of human embryonic stem cells and embryonal carcinoma cells. *Proteomics* **2010**, *10* (7), 1359−73.

(22) Qian, W. J.; Liu, T.; Petyuk, V. A.; Gritsenko, M. A.; Petritis, B. O.; Polpitiya, A. D.; Kaushal, A.; Xiao, W.; Finnerty, C. C.; Jeschke, M. G.; Jaitly, N.; Monroe, M. E.; Moore, R. J.; Moldawer, L. L.; Davis, R. W.; Tompkins, R. G.; Herndon, D. N.; Camp, D. G.; Smith, R. D. Large-scale multiplexed quantitative discovery proteomics enabled by the use of an ^{18}O-labeled "universal" reference sample. *J. Proteome Res.* **2009**, *8* (1), 290−9.

(23) Ding, S. J.; Wang, Y.; Jacobs, J. M.; Qian, W. J.; Yang, F.; Tolmachev, A. V.; Du, X.; Wang, W.; Moore, R. J.; Monroe, M. E.; Purvine, S. O.; Waters, K.; Heibeck, T. H.; Adkins, J. N.; Camp, D. G. II; Klemke, R. L.; Smith, R. D. Quantitative phosphoproteome analysis of lysophosphatidic acid induced chemotaxis applying dual-step ^{18}O labeling coupled with immobilized metal-ion affinity chromatography. *J. Proteome Res.* **2008**, *7* (10), 4215−24.

(24) 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. UNiquant, a program for quantitative proteomics

analysis using stable isotope labeling. *J. Proteome Res.* **2011**, *10* (3), 1228−37.
(25) Petritis, B. O.; Qian, W. J.; Camp, D. G. II; Smith, R. D. A simple procedure for effective quenching of trypsin activity and prevention of ^{18}O-labeling back-exchange. *J. Proteome Res.* **2009**, *8* (5), 2157−63.
(26) Licklider, L. J.; Thoreen, C. C.; Peng, J.; Gygi, S. P. Automation of nanoscale microcapillary liquid chromatography-tandem mass spectrometry with a vented column. *Anal. Chem.* **2002**, *74* (13), 3076−83.
(27) Mayampurath, A. M.; Jaitly, N.; Purvine, S. O.; Monroe, M. E.; Auberry, K. J.; Adkins, J. N.; Smith, R. D. DeconMSn: a software tool for accurate parent ion monoisotopic mass determination for tandem mass spectra. *Bioinformatics* **2008**, *24* (7), 1021−3.
(28) Petyuk, V. A.; Jaitly, N.; Moore, R. J.; Ding, J.; Metz, T. O.; Tang, K.; Monroe, M. E.; Tolmachev, A. V.; Adkins, J. N.; Belov, M. E.; Dabney, A. R.; Qian, W. J.; Camp, D. G. II; Smith, R. D. Elimination of systematic mass measurement errors in liquid chromatography-mass spectrometry based proteomics using regression models and a priori partial knowledge of the sample content. *Anal. Chem.* **2008**, *80* (3), 693−706.
(29) Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26* (12), 1367−72.
(30) Cleveland, W. S. Robust locally weighted regression and smoothing scatterplots. *J. Am. Statist. Assoc.* **1979**, *74* (368), 829−36.
(31) Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **2004**, *3*, No. Article3.
(32) Arrais, J. P.; Fernandes, J.; Pereira, J.; Oliveira, J. L. GeneBrowser 2: an application to explore and identify common biological traits in a set of genes. *BMC Bioinf.* **2010**, *11*, 389.
(33) Szklarczyk, D.; Franceschini, A.; Kuhn, M.; Simonovic, M.; Roth, A.; Minguez, P.; Doerks, T.; Stark, M.; Muller, J.; Bork, P.; Jensen, L. J.; von Mering, C. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **2011**, *39* (Database issue), D561−8.
(34) Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. Proteolytic ^{18}O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal. Chem.* **2001**, *73* (13), 2836−42.
(35) Helenius, K.; Yang, Y.; Tselykh, T. V.; Pessa, H. K.; Frilander, M. J.; Makela, T. P. Requirement of TFIIH kinase subunit Mat1 for RNA Pol II C-terminal domain Ser5 phosphorylation, transcription and mRNA turnover. *Nucleic Acids Res.* **2011**, *39* (12), 5025−35.
(36) Yamaguchi, T.; Cubizolles, F.; Zhang, Y.; Reichert, N.; Kohler, H.; Seiser, C.; Matthias, P. Histone deacetylases 1 and 2 act in concert to promote the G1-to-S progression. *Genes Dev.* **2010**, *24* (5), 455−69.
(37) Vermeulen, M.; Selbach, M. Quantitative proteomics: a tool to assess cell differentiation. *Curr. Opin. Cell Biol.* **2009**, *21* (6), 761−6.
(38) Reiland, S.; Salekdeh, G. H.; Krijgsveld, J. Defining pluripotent stem cells through quantitative proteomic analysis. *Expert Rev. Proteomics* **2011**, *8* (1), 29−42.
(39) Johnson, K. L.; Muddiman, D. C. A method for calculating 16O/18O peptide ion ratios for the relative quantification of proteomes. *J. Am. Soc. Mass Spectrom.* **2004**, *15* (4), 437−45.
(40) Qian, W. J.; Monroe, M. E.; Liu, T.; Jacobs, J. M.; Anderson, G. A.; Shen, Y.; Moore, R. J.; Anderson, D. J.; Zhang, R.; Calvano, S. E.; Lowry, S. F.; Xiao, W.; Moldawer, L. L.; Davis, R. W.; Tompkins, R. G.; Camp, D. G. II; Smith, R. D. Quantitative proteome analysis of human plasma following in vivo lipopolysaccharide administration using 16O/18O labeling and the accurate mass and time tag approach. *Mol. Cell. Proteomics* **2005**, *4* (5), 700−9.
(41) Zhou, R.; Han, L.; Li, G.; Tong, T. Senescence delay and repression of p16INK4a by Lsh via recruitment of histone deacetylases in human diploid fibroblasts. *Nucleic Acids Res.* **2009**, *37* (15), 5183−96.
(42) Lee, S.; Park, J. R.; Seo, M. S.; Roh, K. H.; Park, S. B.; Hwang, J. W.; Sun, B.; Seo, K.; Lee, Y. S.; Kang, S. K.; Jung, J. W.; Kang, K. S. Histone deacetylase inhibitors decrease proliferation potential and multilineage differentiation capability of human mesenchymal stem cells. *Cell Proliferation* **2009**, *42* (6), 711−20.
(43) Fu, M.; Rao, M.; Bouras, T.; Wang, C.; Wu, K.; Zhang, X.; Li, Z.; Yao, T. P.; Pestell, R. G. Cyclin D1 inhibits peroxisome proliferator-activated receptor gamma-mediated adipogenesis through histone deacetylase recruitment. *J. Biol. Chem.* **2005**, *280* (17), 16934−41.
(44) Bienvenu, F.; Jirawatnotai, S.; Elias, J. E.; Meyer, C. A.; Mizeracka, K.; Marson, A.; Frampton, G. M.; Cole, M. F.; Odom, D. T.; Odajima, J.; Geng, Y.; Zagozdzon, A.; Jecrois, M.; Young, R. A.; Liu, X. S.; Cepko, C. L.; Gygi, S. P.; Sicinski, P. Transcriptional role of cyclin D1 in development revealed by a genetic-proteomic screen. *Nature* **2010**, *463* (7279), 374−8.
(45) Hong, H.; Takahashi, K.; Ichisaka, T.; Aoi, T.; Kanagawa, O.; Nakagawa, M.; Okita, K.; Yamanaka, S. Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. *Nature* **2009**, *460* (7259), 1132−5.
(46) Li, H.; Collado, M.; Villasante, A.; Strati, K.; Ortega, S.; Canamero, M.; Blasco, M. A.; Serrano, M. The Ink4/Arf locus is a barrier for iPS cell reprogramming. *Nature* **2009**, *460* (7259), 1136−9.
(47) Moldovan, G. L.; Pfander, B.; Jentsch, S. PCNA, the maestro of the replication fork. *Cell* **2007**, *129* (4), 665−79.