Quantitative Proteomics of Protein Complexes and Their Implications for Cell Reprograming and Pluripotency

ARTICLE *in* JOURNAL OF PROTEOME RESEARCH · NOVEMBER 2013

Impact Factor: 4.25 · DOI: 10.1021/pr4008877 · Source: PubMed

7 AUTHORS, INCLUDING:

Chein-Hung Chen
Academia Sinica
24 PUBLICATIONS **400** CITATIONS
SEE PROFILE

Hung-Chih Kuo
Academia Sinica
46 PUBLICATIONS **1,141** CITATIONS
SEE PROFILE

Journal of proteome research

Article

pubs.acs.org/jpr

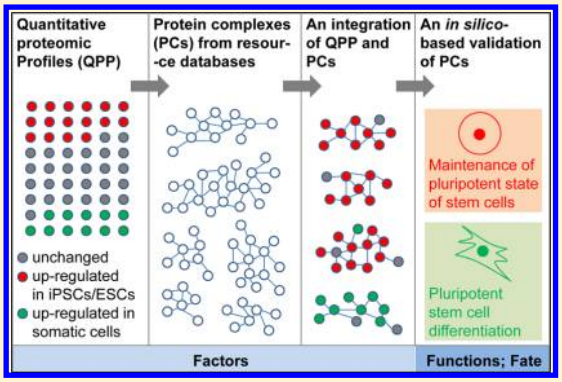
Quantitative Proteomics of Protein Complexes and Their Implications for Cell Reprograming and Pluripotency

Putty-Reddy Sudhir,[†] Madireddy Pavana Kumari,[†] Wei-Ting Hsu,[‡] Julien Massiot,[†] Chein-Hung Chen,[†] Hung-Chih Kuo,*[,†,‡] and Chung-Hsuan Chen*[,†]

[†]Genomics Research Center and [‡]Institute of Cellular and Organismic Biology, Academia Sinica, Taipei 11529, Taiwan

Supporting Information

ABSTRACT: Pluripotency of embryonic stem cells (ESCs)/induced pluripotent stem cells (iPSCs) and reprograming of somatic cells (SCs) to pluripotency are governed by known and unknown factors. These factors, including protein complexes, are poorly described at the proteome level. Here, we established the quantitative proteomic profiles across three types of cells (iPSCs, ESCs, and SCs) using OFFGEL fractionation coupled with LTQ-Orbitrp analysis. Additionally, we utilized the previously published proteomic profiles of iPSCs, ESCs, and SCs. By integrating these proteomic profiles with protein−protein interaction resources, we identified numerous protein complexes in iPSCs and/or ESCs, which include known and novel chromatin remodeling complexes that facilitate cell reprograming. The identified protein complexes also include the previously unreported ones that are associated with the imperfect aspects of iPSCs or cell reprograming process. Further, we performed a comparison between our study and previously published studies and highlighted a partial conservation of the identified protein complexes across the iPSCs generated by different laboratories and iPS cell-type specific protein complexes. The identified protein complexes were validated by integrated *in silico* analysis of microarray repository data related to ESCs differentiation into embryoid bodies. A majority of the protein complexes exhibited significant ($p < 0.005$) co-regulation of their components upon ESC differentiation, suggesting their role in the maintenance of the pluripotent state. Finally, we showed a link between the components of the protein complexes and embryonic development using the existing loss-of-function phenotype data. Together, our integrated approach provides the first comprehensive view of the protein complexes that may have implications for cell reprograming and pluripotency.



KEYWORDS: *pluripotent stem cell proteomics, integrated proteomics approach, protein complexes, cell reprograming, pluripotency*

INTRODUCTION

Embryonic stem cells (ESCs) are derived from the inner cell mass of an early embryo, and they possess the ability to self-renew and differentiate into multiple lineages (i.e., pluripotency). Induced pluripotent cells (iPSCs) are ESC-like cells derived from somatic cells by a process known as cell reprogramming. Cell reprogramming has been successfully demonstrated using different combinations of transcription factors, including Oct4, Sox2, Klf4, and c-Myc.[1] Recently, researchers focused on determining the significance of pluripotent stem cells or PSCs (ESCs/iPSCs) in regenerative medicine. However, the level of understanding of key pluripotency regulators, including the proteins implicated in processes such as chromatin remodeling, DNA replication, RNA splicing, gene expression, and histone modifications (e.g., acetylation, methylation), is very limited.

Protein−protein interaction networks or protein complexes are essential for most cellular functions and processes. Analyses of protein complexes provide insight into key players involved in the regulation of cellular functions and processes. A targeted approach based on affinity purification (AP) coupled with mass spectrometry (MS) analysis has been extensively used to characterize protein complexes. Several AP-MS studies have identified the protein complexes of transcription factors such as Oct4, Sox2, and Nanog in mouse ESCs.[2−6] Some of the identified transcription factor-interacting proteins are components of chromatin remodeling complexes and polycomb group protein complexes that are known to regulate transcription and gene expression processes in ESCs.[4,5] Additionally, it is well-known that some of the interacting proteins of Oct4 are master regulators of pluripotency. However, these AP-MS studies limit the dynamic range of protein complexes to only transcription factor and protein interactions in PSCs, which is insufficient to address all biologically important protein−protein interactions.

In contrast to previous studies, we set out to address the protein complexes on a proteome-wide scale. First, the proteomes of three types of human cells, iGra (iPSCs), H9 (ESCs), and granulosa (Gra), were analyzed by high resolution mass spectrometry and were compared using a MaxQuant-based label-free quantitative approach. Furthermore, by integrating the

Received: August 29, 2013
Published: November 13, 2013

quantitative proteomic profiles with the resources of protein−protein interactions, we identified 87 protein complexes in three types of human cells (iGra, H9, and Gra). Of the 87 complexes, 71 were identified as differentially regulated between iGra and Gra or between iGra and H9. When we used a similar type of integrated approach, the analysis of previously published proteomic profiles of nine human cells (three iPSCs, three ESCs, and three fibroblast cells)[7,8] revealed that some of those 71 complexes were conserved among the iPSCs generated by other laboratories. The comparison of protein complexes identified in our study and previously published studies highlighted the iPS cell type-specific protein complexes. The identified protein complexes were linked with diverse cellular roles such as protein acetylation/methylation, nucleic acid/protein binding, and chromosome organization/chromatin remodeling. Further, an informatics-assisted validation of the protein complexes revealed their role in the maintenance of the pluripotent state. The analysis of existing phenotype data from genetic manipulation (e.g., gene knockout) studies established a link between the identified protein complexes and embryonic development. This study provides a resource of protein complexes that expands our view of protein−protein interaction networks in PSCs.

EXPERIMENTAL PROCEDURES

Cell Culture

Human granulosa cells (Gra) obtained from the In vitro Fertilization Center at National Taiwan University were cultured in RPMI supplemented with 10% fetal bovine serum (FBS), 1% nonessential amino acid (Invitrogen), 2 mM L-glutamine (Invitrogen), and 1% penicillin/streptomycin (Invitrogen). The iPSCs (iGra cells) are derived from Gra cells by means of cell reprograming.[9] Human ESCs (H9) and iPSCs (iGra) were grown on MEF feeders (2×10^4 cells/cm^2) in DMEM/F-12 medium with 20% knockout Serum Replacement (Invitrogen) and 4 ng/mL of bFGF (Sigma).

Cell Lysis, Trypsin Digestion, and OFFGEL Fractionation

Cells were lysed in modified RIPA buffer containing 1% NP40 (Igepal CA-630), 300 mM NaCl, 1 mM EDTA, 0.5 mM DTT, 1 mM $NaVO_3$, 10 mM NaF, and a cocktail of protease inhibitors (Roche, Indianapolis, IN). Supernatants were collected from cell lysates after centrifugation at 15000*g* for 5 min at 4 °C. The in-solution trypsin digestion was performed using a standard protocol as described previously.[10] Briefly, proteins were denatured using 6 M urea, reduced with 10 mM DTT, alkylated with 40 mM IAA, and digested with 20 μg of trypsin (Promega)/mg protein overnight. Tryptic peptide samples were desalted using C18 cartridge (Waters) and fractionated by using ImmobilinTM DryStrip, pH 3−10, 13 cm (GE Healthcare) and 3100 OFFGEL fractionator (Agilent Technologies), according to the manufacturer's protocol.[10,11] The fractions were desalted using C18, dried, reconstituted in 0.1% TFA, and subjected to MS analysis.

Mass Spectrometry Analysis

LC−MS/MS analysis of peptides was conducted on an LTQ-Orbitrap XL ETD mass spectrometer (Thermo Fisher Scientific, San Jose, CA) equipped with a nanoelectrospray ion source (New Objective, Inc.), and accela LC system (Thermo Fisher Scientific, San Jose, CA) as described previously.[10] The sample was injected (5 μL) at 10 μL/min flow rate on to a self-packed precolumn (150 μm i.d. × 30 mm, 5 μm, 200 Å). Chromatographic separation was performed on a self-packed reversed phase C18 nanocolumn (75 μm i.d. × 200 mm, 3 μm, 200 Å) using 0.1% formic acid in water as mobile phase A and 0.1% formic acid in 80% acetonitrile as mobile phase B with split flow rate of 300 nL/min. The full-scan mass range was set from m/z 320−2000 with resolution 60 000 at m/z 400. The 10 most intense ions were sequentially isolated for MS2 by LTQ. Tandem-MS in iontrap was carried out using CAD mode and using normalized collision energy (NCE) of 35%. Dynamic exclusion was enabled (repeat count, 1; repeat duration, 90; exclusion list size, 500; exclusion duration, 180; exclusion mass width, ±15 ppm). The electrospray voltage was maintained at 1.8 kV and the capillary temperature was set at 200 °C.

Proteomic Data Analysis

Raw MS files obtained from LTQ-Orbitrap were analyzed using MaxQuant (version 1.3.0.5), which consists of an integrated Andromeda search engine. MS/MS spectra were searched against the decoy ipi_HUMAN_3.87 database (91 464 entries). MaxQuant analysis was performed using a precursor mass tolerance of 20 ppm (first search for mass calibration) and of 6 ppm (main search) and a fragment mass tolerance of 0.5 Da (main search). We set the search parameters, such as enzyme trypsin digestion, up to two missed cleavages, a peptide length of seven amino acids, a false discovery rate (FDR) of 0.01, variable modifications of methionine oxidation and N-terminal acetylation, and fixed modification of carbamidomethyl cysteine. MaxQuant-based label-free quantitation between cell lines was performed using normalized protein intensities as described previously.[12,13] Label-free quantitative analysis required at least two ratio counts to acquire the normalized protein intensity (LFQ intensity). Finally, the proteins identified from the reverse database and common contaminants were eliminated from the list of quantified proteins.

Microarray-Based Transcriptomic Analysis

RNA was isolated from three types of cells (iGra, H9, and Gra) that were used in the proteomics analysis and GeneChip Human Genome Affymetrix U133 Plus 2.0 Array analysis was performed according to manufacturer's protocol.

Microarray Data Analysis

Array data was normalized and quantified using Agilent's GeneSpring 11.5 software. We excluded the probes that exhibit lowest 20% of the range of intensities in all samples. The probes with highest summed intensities were considered for analysis to eliminate the probe sets redundancy. Unigene IDs of the affymetrix probes were used to compare the proteomics and microarray data sets.

Quantitative Profiling of Protein Complexes and Cellular Functions

Protein complexes were analyzed by following two approaches. In the first approach, significant regulations of protein complexes in one-dimension (proteome level) or in two-dimensions (both proteome and transcriptome levels) were analyzed by integrating the quantitative profiles of proteins and protein products encoded by transcripts with CORUM database (02/2012).[14,15] The integrated analysis calculates whether the protein ratios of a given protein complex are significantly different from the global distribution ratios of all proteins. The analysis also gives relative regulation or difference score between 1 and −1, a significance (p-value) value, and a false discovery rate (FDR) for each category. The scores of protein complexes indicates up-regulation or down-regulation when the value is close to 1 or −1, respectively, and significant terms are < −0.3 or >0.3. The

dx.doi.org/10.1021/pr4008877 | J. Proteome Res. 2013, 12, 5878−5890

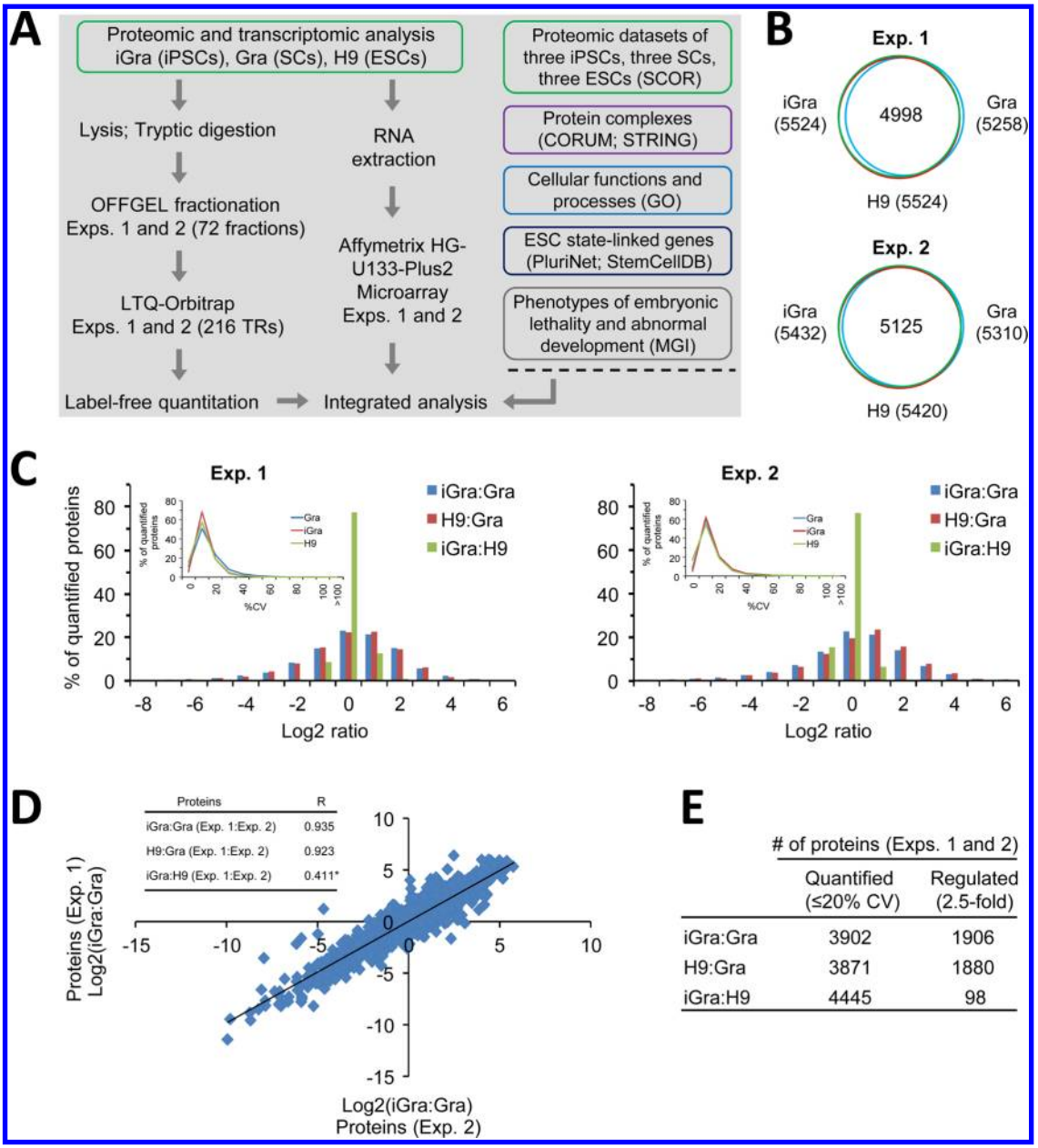


Figure 1. Overview of experimental design and quantitative proteomic profiling of PSCs and Gra cells. (A) Schematic diagram showing the workflow used for proteomic, transcriptomic, and integrated analyses. Comparative proteomic and transcriptomic analyses of iGra, H9, and Gra cells were performed in biological replicates (Exps. 1 and 2) and were integrated with various database resources (SCOR, CORUM, STRING, GO, PluriNet, StemCellDB, and MGI). SCs and TRs indicate somatic cells and technical replicates, respectively. (B) Venn diagrams representing the overlap of proteins identified in iGra, H9, and Gra cells in biological replicates. (C) Insets show the % of total proteins quantified with a coefficient of variation (CV) $\leq$ 20% in TRs. Distribution of the log2 ratios of proteins (CV $\leq$ 20%) observed between iGra, Gra and H9 cells in biological replicates. (D) The variability between two biological replicates is shown by comparing the log2 ratios of proteins (CV $\leq$ 20%) identified in iGra/H9:Gra, iGra:H9 analyses. Inset shows the Pearson correlation (r) values observed between two biological replicates. (E) Number of quantified proteins and differentially expressed proteins between iGra, Gra, and H9 cells.

significance of the deregulated category was estimated using the Wilcoxon-Mann–Whitney test and a Benjamini-Hochberg FDR of 2%. In addition, differential regulations of cellular function categories were analyzed by means of the above integrated approach and Gene Ontology database. We used Perseus software 1.3.0.4 for one- and two-dimensional enrichment analyses. In the second approach, differentially regulated (2.5-fold) candidates were identified from proteomic profiles and were subjected to STRING database to output the protein–protein interaction networks. The networks were generated using high confidence criterion and default parameters in STRING 9.0. Furthermore, tightly connected protein complexes were derived from the networks by MCODE 1.32 analysis using Cytoscape 2.8.3 platform.[16−18]

An Integrated *in Silico*-Based Validation of Protein Complexes

To identify the protein complexes associated with the pluripotent state of ESCs, we examined the regulation status of the components of the protein complexes upon human ES (hES) cell differentiation. To this end, we utilized the microarray-based

dx.doi.org/10.1021/pr4008877 | J. Proteome Res. 2013, 12, 5878−5890

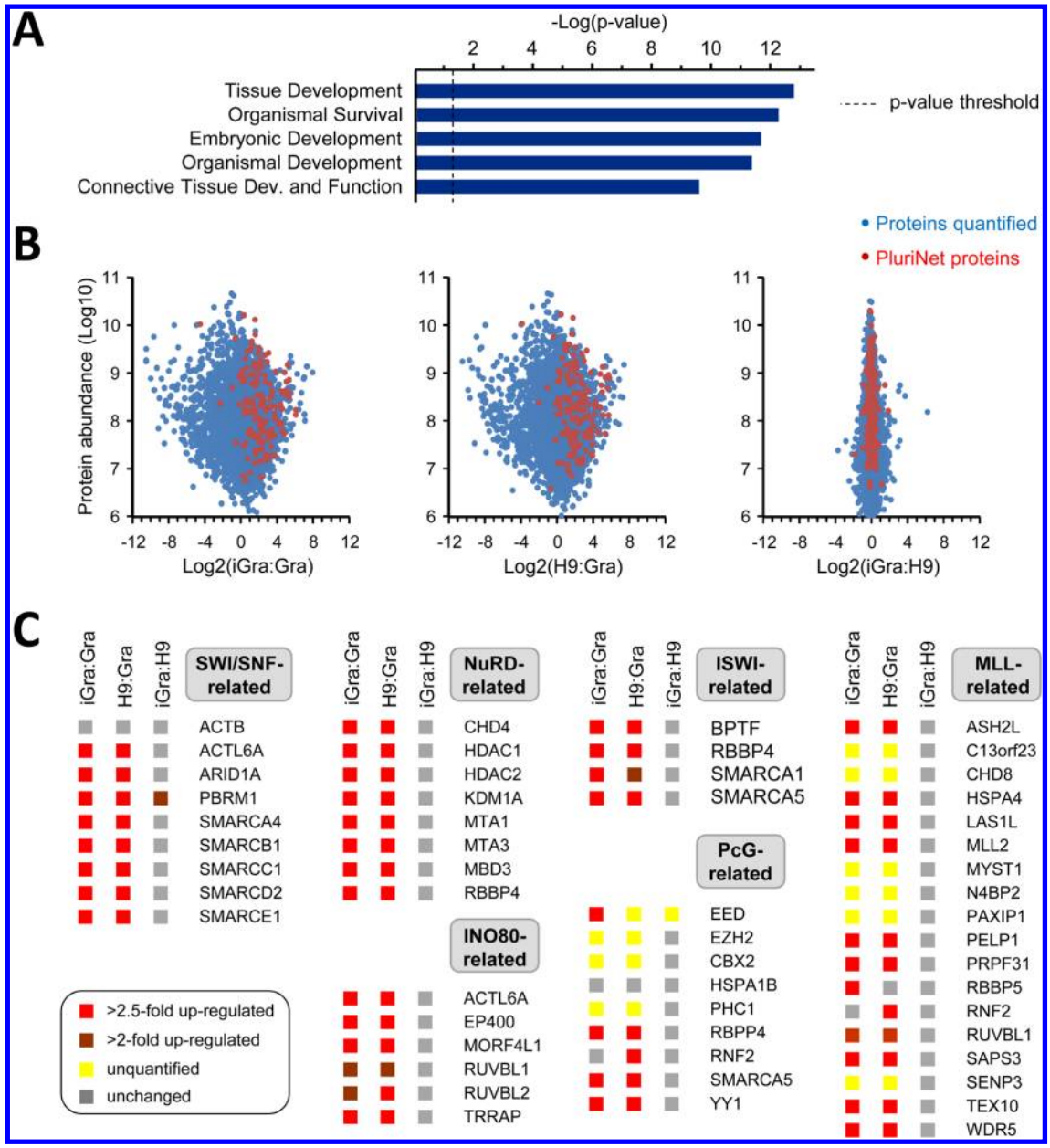


Figure 2. Evaluation of the quantitative proteomic profiles of PSCs and Gra cells. (A) The physiological function categories of quantified proteins in three comparisons (iGra:Gra, H9:Gra, and iGra:H9) were listed and the top five significant categories are shown. (B) The regulation status of PSC-associated proteins (PluriNet) in three comparisons is shown. (C) The regulation status of the components of PSC-associated complexes (chromatin remodeling group (SWI/SNF, ISWI, NuRD, and INO80), PcG group, and MLL group complexes) in three comparisons is shown.

gene expression profiles from a study[19] on the differentiation of hES cell lines (n = 21) into embryoid bodies (EBs). The microarray data is available on the Human Pluripotent Stem Cell Database (StemCellDB) at the National Institutes of Health (http://stemcelldb.nih.gov). We focused on the microarray data related to the hES differentiation into ectodermal (EB_ecto) and mesendodermal (EB_mesend) lineages. Entrez Gene IDs of the Agilent probes were used to compare the proteomics and microarray data sets. The mean intensities were considered to eliminate the probe sets redundancy.

To explore the significance of the protein complexes with reference to developmental processes (e.g., embryonic development), we examined the loss-of-function phenotypes from genetic manipulation studies (e.g., gene knockout) that were contained in the Mouse Genome Informatics (MGI) database (http://www.informatics.jax.org/). First, we listed the mouse orthologs for all of the components of the identified protein complexes. Further, the roles of the components of each complex were examined in the context of developmental processes by analyzing the phenotype data.

RESULTS

Quantitative Proteomic Profiling of PSCs and Granulosa Cells

Our previous study showed the derivation and characterization of iPSCs (iGra cells) from granulosa (Gra) cells.[9] In the present study, we demonstrated the proteomic and transcriptomic profiles of iGra (iPSCs), H9 (ESCs), and Gra cells. To establish the quantitative proteomic profiles of these three types of cells, we followed a standard approach, as shown in Figure 1A. In brief,

dx.doi.org/10.1021/pr4008877 | J. Proteome Res. 2013, 12, 5878−5890

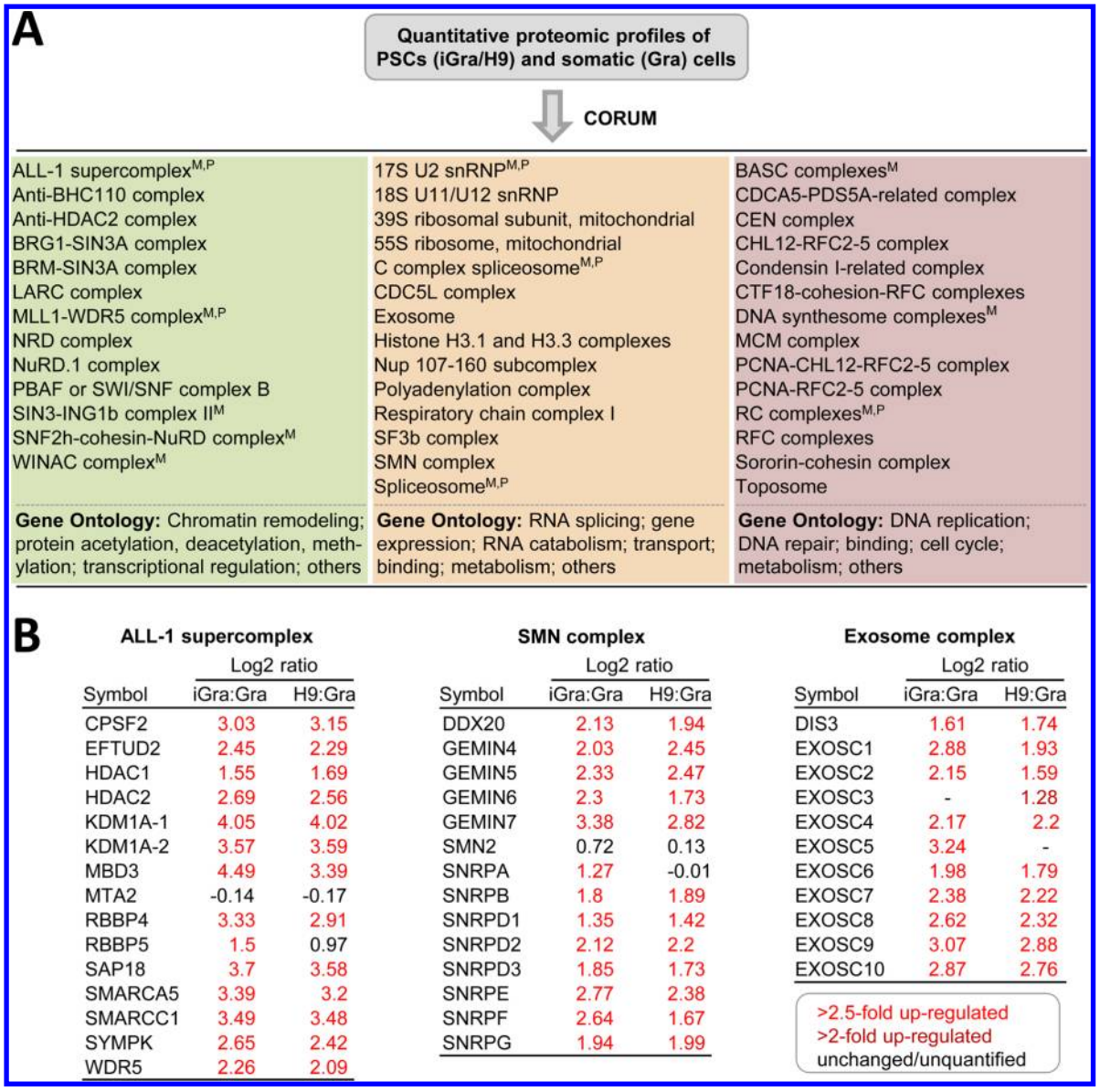


Figure 3. Identification of protein complexes using proteomic profiles and the CORUM Database. (A) Integrated analysis of quantitative proteomic profiles of iGra, H9, and Gra cells using the CORUM database identified statistically significant protein complexes in iGra and H9. Statistically significant (Wilcoxon−Mann−Whitney test and Benjamini−Hochberg FDR of 0.02) differences were observed for all of the CORUM populations between stem and somatic cells. The major cellular roles of the complexes are listed based on Gene Ontology. Some of the protein complexes were grouped and mentioned (e.g., four types of DNA synthesome complexes were identified but mentioned as DNA synthesome complexes). Protein complexes conserved in the Phanstiel et al.[8] (P) and Munoz et al.[7] (M) studies are noted with the designated superscripted letters. (B) Examples of protein complexes with the regulation status of their components are shown.

whole cell lysates were trypsin-digested and subjected to an OFFGEL fractionator, which not only reduces sample complexity but also ensures maximal peptide recovery.[10,11] A total of 36 fractionations (12 fractions per cell type) were collected, and each fraction was analyzed in three technical replicates (TRs) by MS, yielding a total of 108 raw data files. We obtained 216 raw data files from two biological replicate analyses (Exp. 1 and Exp. 2). Raw files were processed using MaxQuant software, which performs protein identification, normalization and label-free quantitation simultaneously.[12,13] MaxQuant analysis resulted in the identification of 6526 proteins (5745 in Exp. 1 and 5583 in Exp. 2) (Supporting Information Table S1). We observed an excellent correlation ($r = 0.972-0.985$) for the protein intensities between the three TRs in both Exps. 1 and 2 (Supporting Information Figure S1), indicating the reliability of the MS analysis. The identified proteins showed a large amount of overlap (87% of 5745 in Exp. 1 and 92% of 5583 in Exp. 2) between the three types of cells (Figure 1B).

To quantify the proteins, we first examined the variance of protein abundance in the three TRs by determining the coefficient of variation (CV).[10] More than 85% of the proteins were quantified with a CV $\leq$ 20% in the three types of cells (Exps. 1 and 2) and were used for all further analyses (Figure 1C). We observed a large number of differentially regulated proteins in the comparisons of PSCs and granulosa cells (iGra:Gra and H9:Gra); in contrast, a large number of unchanged proteins were observed in the comparison of iPSCs and ESCs (iGra:H9) (Figure 1C). We then examined the correlation for all the three comparisons between Exp. 1 and Exp. 2 and found good correlations in the iGra:Gra (r = 0.935) and H9:Gra (0.923)

dx.doi.org/10.1021/pr4008877 | J. Proteome Res. 2013, 12, 5878−5890

cases, and a poor correlation (r = 0.411) in the iGra:H9 case (Figure 1D). The latter was affected by a large number of unchanged protein ratios that were clustered around zero.[7] The average protein abundance ratios between the replicate experimental groups (Exps. 1 and 2) were calculated, and the proteins differentially regulated at least 2.5-fold were listed (Figure 1E and Supporting Information Table S2). The quantitative proteomic data obtained using a criterion of a CV $\leq$ 20% are in good agreement with the results observed in previous studies.[7,8]

The physiological functions of the 4812 unique proteins quantified in the three comparisons (Figure 1E) between iGra, H9, and Gra cells were defined using the IPA (Ingenuity Pathway Analysis) knowledge platform. IPA revealed that annotations, such as tissue development, embryonic development, organismal survival, and organismal development, were significantly associated with the quantified proteins (Figure 2A and Supporting Information Figure S2).

Expression Status of PSC-Associated Genes in the Quantitative Proteomic Profiles

The quantitative proteomic profiles of iGra, H9, and Gra cells were assessed by examining the expression status of PluriNet genes (n = 299). The PluriNet represents the PSC-associated genes identified by microarray-based gene expression analysis of ~150 PSC lines.[20] We found 144, 143, and 179 proteins encoded by PluriNet genes in the quantitative proteomic profiles of iGra:Gra, H9:Gra, and iGra:H9, respectively (Figure 2B and Supporting Information Table S3). Of the 179 PluriNet proteins identified in the iGra:H9 case, only 2.23% were differentially expressed between iGra and H9 cells, indicating that the expression pattern of PluriNet proteins between iPSCs and ESCs is similar. In contrast, of the PluriNet proteins identified in the iGra:Gra and H9:Gra cases, more than 79% showed 2-fold up-regulation in iGra and/or H9 compared to Gra cells while less than 2.1% showed 2-fold up-regulation in Gra compared to iGra and/or H9 cells, indicating a high abundance of PluriNet proteins in PSCs relative to granulosa cells. Furthermore, we found that core pluripotency-associated transcription factors (POU5F1, SOX2, and NANOG) exhibited label-free quantitative intensities in iGra and H9 but not in Gra cells, suggesting their lower expression levels in granulosa cells relative to PSCs (Supporting Information Table S1).

Identification of Protein Complexes Associated with PSCs and Granulosa Cells

To confirm the accuracy of the proteomic profiles, we examined whether the quantitative proteomic profiles of iGra, H9, and Gra cells included the components of chromatin remodeling complexes (SWI/SNF, ISWI, NuRD, and INO80), polycomb group (PcG) complexes, and trithorax group (Trx/MLL) complexes that are known to regulate gene expression in PSCs.[21,22] We found several components of these complexes in the quantitative proteomic profiles, and notably, all of them were up-regulated in PSCs (iGra and/or H9) relative to Gra cells (Figure 2C). Additionally, several components of these complexes were quantified only in the iGra:H9 case due to a lack of quantitative intensities of the components in Gra cells, suggesting their low abundance in granulosa cells relative to PSCs.

To identify protein complexes in an unbiased manner, we integrated all of the proteins quantified in the three comparisons (iGra:Gra, H9:Gra, and iGra:H9) with the CORUM (the Comprehensive Resource of Mammalian protein complexes) database. CORUM consists of a large number of protein complexes that have been experimentally characterized in mammals, including humans.[23] The quantitative profiles of protein complexes were established using one-dimensional enrichment analysis, as described in the Experimental Procedures. Analysis of the proteomic profiles of PSCs and granulosa cells (iGra:Gra and H9:Gra) identified 51 protein complexes that were significantly up-regulated in iGra and/or H9 relative to Gra cells (Figure 3 and Supporting Information Table S4). Among the 51 complexes, we found 48 in iGra cells, 36 in H9 cells, and 33 in both iGra and H9 cells (Supporting Information Figure S3A). Interestingly, none of the protein complexes were up-regulated in Gra cells compared to iGra/H9 cells. Analysis of the proteomic profiles of iPSCs and ESCs (iGra:H9) using a similar approach to that described above identified the up-regulation of two complexes (cytoplasmic ribosome and cytoplasmic 60S ribosomal subunit) in H9 cells compared to iGra cells (Supporting Information Tables S4 and S5). In conclusion, the integrated proteomic approach identified 53 CORUM complexes in PSCs.

Next, the CORUM complexes identified were subjected to Gene Ontology to uncover their cellular roles (Supporting Information Table S5). The identified protein complexes were found to be involved in the cellular roles of chromosome organization, transcription regulation, RNA processing, and gene expression. We also identified complexes responsible for the mitotic cell cycle, cell cycle check points, chromosome segregation, DNA replication, and DNA repair. Furthermore, we exposed several complexes involved in binding activity (nucleic acid, protein, or ATP binding) and protein modifications (acetylation, deacetylation, and methylation). Complexes responsible for RNA catabolism, phosphate metabolism, protein biosynthesis, protein complex assembly, aerobic respiration, transport of proteins or ions, nuclear transport, and hormone mediated signaling were also discovered. Finally, two complexes identified to be differentially regulated between iGra and H9 cells were found to be associated with protein biosynthesis. In summary, in addition to the known protein complexes, several novel complexes regulating various cellular functions in PSCs were highlighted by using an integrated proteomic approach.

Identification of Protein Complexes in Large Protein Interaction Networks of PSCs and Granulosa Cells

We used the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database[24] to further explore the physical and functional protein interaction networks associated with PSCs. STRING analysis was performed using the proteins identified with at least 2.5-fold differential expression in the three comparisons (iGra:Gra, H9:Gra, and iGra:H9). The analysis did not show any protein interaction networks in the case of iGra:H9 as we had identified only a small number of differentially expressed proteins between these two types of PSCs. However, the analysis generated four networks (two up-regulated and two down-regulated) in the cases of iGra:Gra and H9:Gra. The generated networks were further analyzed by MCODE (Molecular Complex Detection) to identify the tightly connected regions that represent protein complexes.[16,17] MCODE derives the complexes from seeds (the absence of a seed, which is the highest scoring node in a complex, leads to fragmentation of the complex). The MCODE analysis resulted in the identification of 34 complexes, 26 of which were up-regulated in iGra and/or H9 cells, and 8 were up-regulated in Gra cells (Supporting Information Figure S3B and Supporting Informa-

dx.doi.org/10.1021/pr4008877 | J. Proteome Res. 2013, 12, 5878−5890

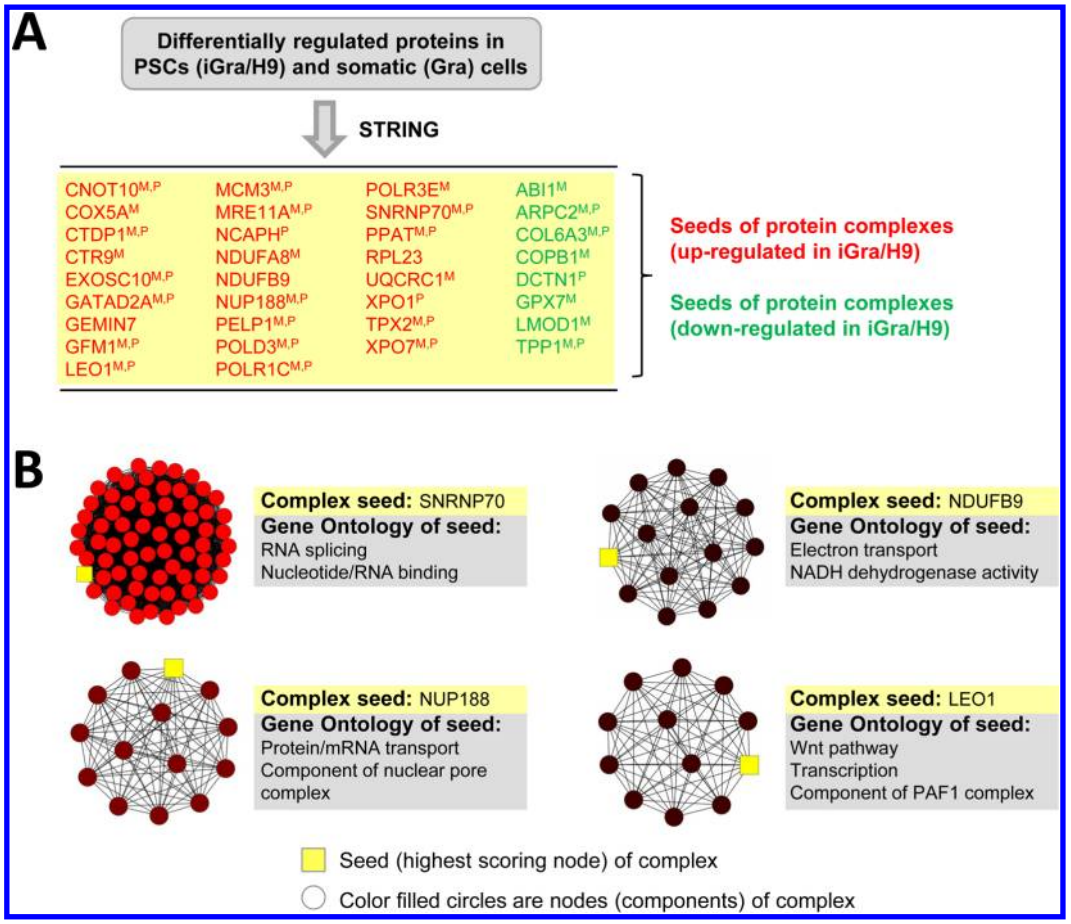


Figure 4. Identification of protein complexes using proteomic profiles and the STRING Database. (A) Integrated analysis of differentially expressed proteins associated with iGra/H9 and Gra cells using the STRING database identified protein interaction networks. Using MCODE analysis of the protein interaction networks, tightly connected or highly ranked complexes derived from seeds were identified. Seeds of the complexes up- and down-regulated in PSCs (iGra and/or H9), compared to Gra cells, are indicated by red and green colors, respectively. Complexes conserved in the Phanstiel et al.[8] (P) and Munoz et al.[7] (M) studies are noted with the designated superscripted letters. (B) Examples of a few protein complexes identified in iGra cells are shown with their seeds (yellow color) and cellular roles.

tion Table S6). The seeds of the STRING complexes are listed in Figure 4A, and examples of complexes are shown in Figure 4B.

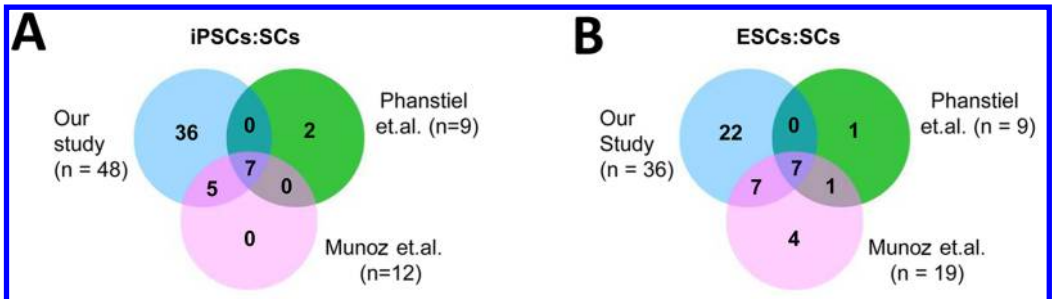
Next, we looked into the cellular roles of the identified seeds. Of the seeds identified, five (CTR9, COL6A3, DCTN1, LEO1, and TPP1) were known to be involved in developmental processes. Seeds involved in regulation of protein phosphorylation (DCTN1, MRE11A, TPP1, and TPX2), histone methylation (CTR9), glycosyl group transfer (PPAT), cell cycle processes (DCTN1, MCM3, MRE11A, NCAPH, POLD3, TPX2, and XPO1), chromosome organization (CTR9, LEO1, MRE11A, NCAPH, and POLD3) and RNA processing (CNOT10, EXOSC10, LEO1, and SNRNP70) were also discovered. These seeds could be critical in the governance of the stem cell fate, and their roles in PSC biology should be explored in future studies.

Comparison of Proteomic and Transcriptomic Profiles of PSCs and Granulosa Cells

To compare the protein complexes and their cellular roles at the proteome and transcriptome levels, transcriptomic profiles of iGra, H9, and Gra cells were analyzed in biological replicates (Exps. 1 and 2) using a standard approach, as shown in Figure 1A. Probe intensity ratios were well correlated (r = 0.983−0.875) between the biological replicates (Supporting Information Figure S4 and Supporting Information Table S7). Transcriptomic profiles were established by averaging the probe intensity ratios observed in biological replicates and were assessed by examining the regulation status of PluriNet genes (Supporting Information Table S7). We identified 275 PluriNet genes, 210 and 181 of which showed at least 1.5-fold differential regulation in iGra and H9, relative to Gra cells, respectively. However, only 23 of the 275 PluriNet genes showed at least 1.5-fold differential regulation between iGra and H9 cells. These results showed the high-abundance of pluripotency-associated genes in PSCs relative to granulosa cells, in agreement with the observed proteomics results.

We next examined the correlation between transcript and protein abundances for all three comparisons (iGra:Gra, H9:Gra, and iGra:H9). The iGa:Gra and H9:Gra cases showed good correlation (r = 0.77 and 0.75, respectively), and the iGra:H9 case showed a poor correlation, as expected (Supporting Information Figure S5A). The results were comparable with those of the previous stem cell-omics study that reported a correlation (r) of ∼0.7 between protein and transcript abundances.[7] Furthermore, we compared the proteomic and transcriptomic profiles of iGra:Gra, H9:Gra, and iGra:H9 using two-dimensional enrichment analysis as described in Experimental Procedures. The analysis listed the cellular roles and protein complexes that were significant at both the transcriptome and proteome levels (Supporting Information Figure S5B, C and

dx.doi.org/10.1021/pr4008877 | J. Proteome Res. 2013, 12, 5878−5890



C

# of complexes or seeds	# of complexes or seeds linked with: PluriNet genes	ESC state (ES_undiff to EB): EB_ecto	ESC state (ES_undiff to EB): EB_mesendo	Dev. processes (eg. embryonic dev.)
iGra or H9:Gra				
51 complexes	40	41	46	42
26 seeds	06	15	22	06
iGra:H9				
02 complexes	–	02	02	02

D

Complexes (**iGra or H9:Gra**)	Total # of components identified	# of components linked with: PluriNet genes	ESC State (EB_ ecto)	ESC State (EB_ mesendo)	Dev. processes (eg. embryonic dev.)
1. BASC complex	09	04	02	07	05
2. Exosome complex	11	04	07	09	–
3. MCM complex	05	04	05	05	04
4. MLL1-WDR5 complex	10	02	05	07	03
5. SMN complex	14	08	11	14	01

Figure 5. Protein complexes are conserved in iPSCs of other laboratories and are linked with pluripotency, development, and the ESC state. (A) Integrated analysis of previously published quantitative proteomic profiles (Munoz et al.[7] and Phanstiel et al.[8]) of iPSCs, ESCs, and somatic cells (SCs) identified statistically significant CORUM complexes (Wilcoxon–Mann–Whitney test and Benjamini–Hochberg FDR of 0.02). The Venn diagram represents the overlap of the complexes identified in comparisons of iPSCs and SCs from our study and from published studies. (B) The Venn diagram representing the overlap of the complexes identified in comparisons of ESCs and SCs from our study and from published studies is shown. (C) The total number of complexes identified in our study by comparing PSCs and granulosa cells (iGra/H9:Gra), as well as two types of PSCs (iGra:H9), is shown. In addition, the number of complexes linked with PluriNet, the ESC state, or developmental processes (e.g., embryonic development) is shown. In the case of the ESC state, the number of complexes linked with two types of embryoid bodies (EBs), ectodermal (EB_ecto) and mesendodermal (EB_mesendo) linages, derived from undifferentiated hES (ES_undiff) cell lines is shown. (D) Examples of protein complexes identified in iGra and/or H9 cells are shown. For each complex, the total number of components and the number of components linked with PluriNet, the ESC state, or developmental processes are shown.

Supporting Information Table S8). This included up-regulated (e.g., chromatin remodeling; DNA replication and repair; nucleic acid metabolism, transport, and binding; RNA processing; and telomere organization) and down-regulated (e.g., actin binding, cell migration, cell secretion, exocytosis, platelet activation, and vesicle organization) cellular roles in iGra and H9 relative to Gra cells. However, none of the cellular roles or protein complexes were differentially regulated in iGra relative to H9 cells. Furthermore, four up-regulated complexes (spliceosome, C complex spliceosome, 17S U2 snRNP, and ALL-1 super-complex) were identified in iGra and/or H9 relative to Gra cells. Although we found 51 complexes at the proteome level (Supporting Information Table S5), transcriptome analysis revealed only a few (n = 4) complexes in PSCs. These results suggest that post-transcriptional regulation of gene expression plays a key role in the existence of complexes at the protein level to govern their downstream cellular functions in PSCs.

Comparison of the Protein Complexes Identified in Our Study and Previously Published Studies

In our study, a total of 87 complexes were identified, 53 and 34 of which were derived from the CORUM and STRING databases, respectively. Of the 53 CORUM complexes, 51 were up-regulated in human PSCs (48 in iGra and 36 in H9) relative to Gra cells (Supporting Information Figure S3A) and 2 were identified as down-regulated in iGra cells relative to H9 cells. Of the 34 STRING complexes, 26 were up-regulated in human PSCs (iGra and/or H9) relative to Gra cells (Supporting Information Figure S3B).

Next, we examined whether the CORUM complexes identified in our study were also identified in previously published proteomic profiles of human PSCs. To this end,

dx.doi.org/10.1021/pr4008877 | J. Proteome Res. 2013, 12, 5878–5890

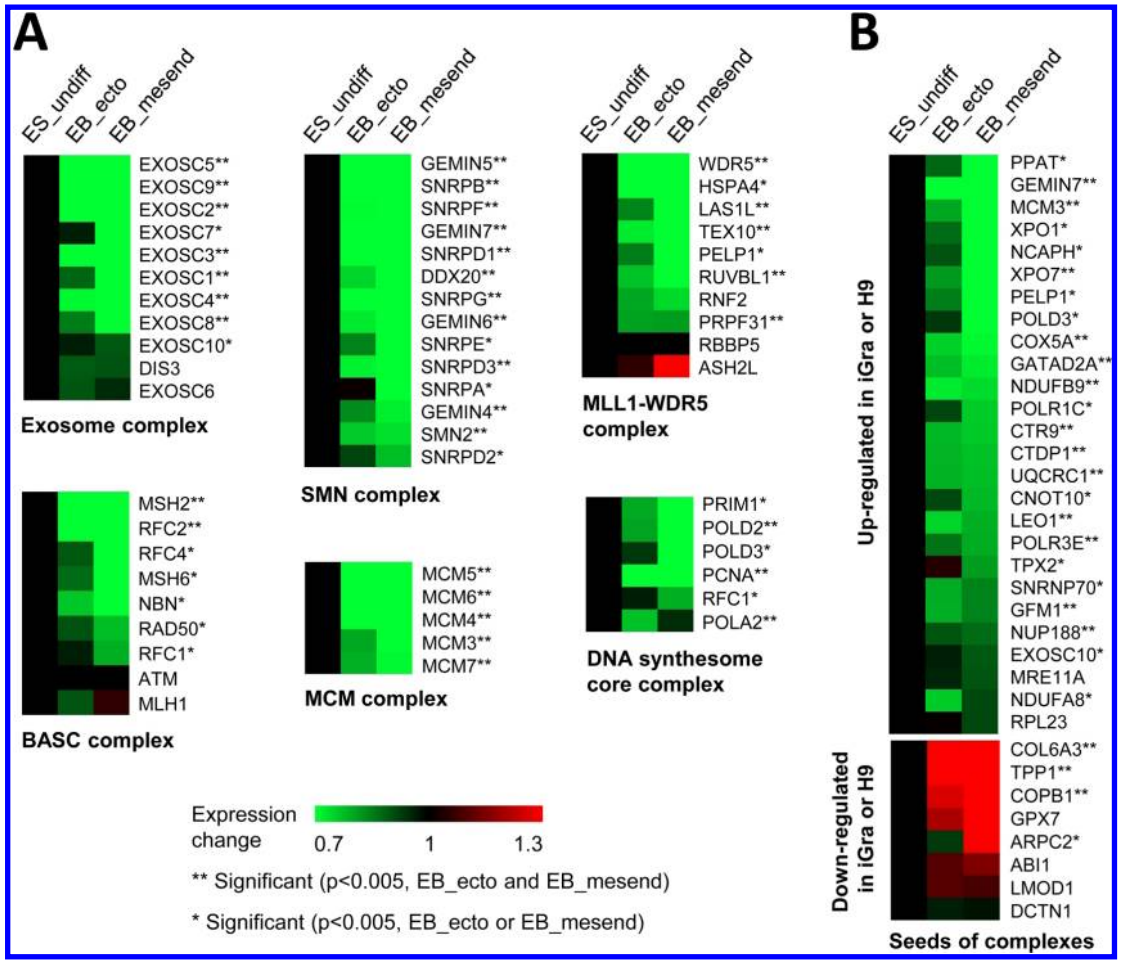


Figure 6. Protein complexes are co-regulated upon ESC differentiation. Microarray-based gene expression profiles of undifferentiated hES (ES_undiff) cell lines (n = 21) and their embryoid bodies (EBs), ectodermal (EB_ecto) and mesendodemal (EB_mesendo) lineages, show the co-regulated complexes upon hES cell differentiation. (A) Examples of CORUM complexes, which were identified in PSCs (iGra and/or H9), are shown. (B) Seeds of STRING complexes, which were identified in PSCs (iGra and/or H9), are shown.

published quantitative proteomic profiles of human PSCs and somatic cells[7,8] were integrated with the CORUM database, and a similar type of analysis as described in the Experimental Procedures was performed. In the comparisons of human iPSCs and somatic cells, we identified 9 and 12 complexes in the studies by Phanstiel et al.[8] and Munoz et al.,[7] respectively. Of those complexes, 7 (7 of 9) and 12 (12 of 12) overlapped with the 48 complexes identified in our study (Figure 5A and Supporting Information Figure S6A). In the comparisons of human ESCs and somatic cells, 14 of the 20 complexes identified in published studies exhibited overlap with 36 complexes identified in our study (Figure 5B and Supporting Information Figure S6B). In the comparisons of human iPSCs and ESCs, two down-regulated complexes identified in iGra relative to H9 cells were identified as up-regulated in human iPSCs relative to ESCs in one of the two published studies (Supporting Information Figure S7). Furthermore, we examined whether the regulation status of the seeds of the STRING complexes identified in our study was similar to previously published proteomic profiles of human PSCs. We analyzed 26 seeds of the up-regulated complexes identified in PSCs (iGra and/or H9) relative to Gra cells and found that only 11 of them had a similar pattern of regulation status (at least a 2.5-fold difference) in at least one of the two published studies[7,8] (Supporting Information Figure S8A). The published studies also displayed similar patterns of regulation status for six of the eight seeds of the down-regulated complexes identified in PSCs (iGra/H9) relative to Gra cells (Supporting Information Figure S8B).

The above results revealed a partial overlap of identified protein complexes across the iPSCs generated by us and other laboratories. The partial overlap highlights that the iPSCs generated from different types of somatic cells (4Skin fetal fibroblasts, IMR90 fetal fibroblasts, newborn foreskin fibroblast, and granulosa cells) and experimental procedures exhibit iPS cell-type specific protein complexes (Figures S3A and S6A). However, a large number of complexes, including known chromatin remodeling complexes (NuRD and PBAF), were identified in the present study but not in previously published data sets, suggesting that iPSCs derived from granulosa cells are more appropriate for the study of protein complexes than the iPSCs derived from fibroblasts used in other laboratories. Furthermore, the existence of differentially regulated protein complexes between ESCs and iPSCs in our study and in one of the published studies (Supporting Information Figure S7) not only suggests the imperfect aspects of the cell reprograming process but also indicates that iPSCs are a type of PSCs that differs from ESCs to some extent.

dx.doi.org/10.1021/pr4008877 | J. Proteome Res. 2013, 12, 5878−5890

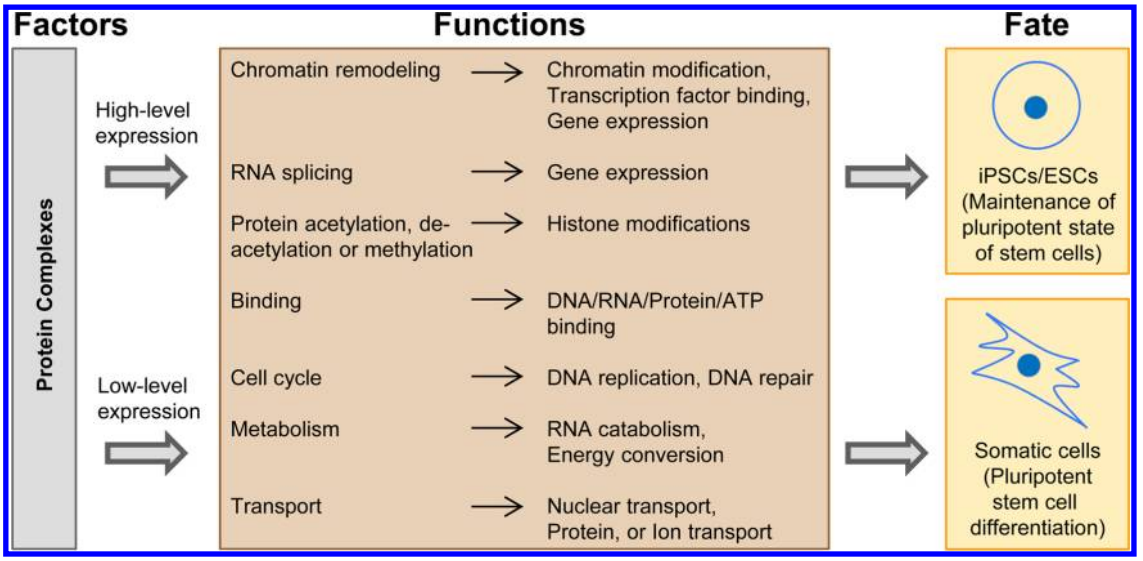


Figure 7. Hypothetical schematic of the role of protein complexes in maintenance of the pluripotent state of the ESCs/iPSCs. The known and novel components and seeds of the protein complexes, which mediate diverse cellular functions, are expressed at high levels to maintain the pluripotent state of the stem cells (iPSCs/ESCc), and alterations (e.g., down-regulation) in their expression levels cause pluripotent stem cell differentiation.

Associating the Identified Protein Complexes with PSC-Associated Genes

To find a link between protein complexes and pluripotency, identified complexes were analyzed in the context of PSC-associated proteins or proteins encoded by PluriNet genes.[20] In the present study, a total of 77 complexes (51 CORUM complexes and 26 STRING complexes) were identified in PSCs by comparing iGra/H9 and Gra cells. We detected the presence of PluriNet genes in 40 of the 51 CORUM complexes (Figure 5C and Supporting Information Table S5). Furthermore, six of the 26 seeds of the STRING complexes were identified as PluriNet genes (Figure 5C and Supporting Information Table S6). However, two CORUM complexes identified by comparing two types of PSCs (iGra:H9) lacked PluriNet genes (Figure 5C and Supporting Information Table S5), indicating their importance in metabolic processes (e.g., protein synthesis) rather than pluripotency. Most significantly, 12 PluriNet genes were identified as components of the spliceosome complex, suggesting the importance of mRNA splicing and post-transcriptional regulation of gene expression in PSCs.

Associating the Identified Protein Complexes with the ESC State

To identify the protein complexes associated with the ESC state, we examined the regulation status of the protein complex components upon human ES (hES) cell differentiation into embryoid bodies (EBs), ectodermal lineages and mesendodermal lineages, as described in the Experimental Procedures. In this study, a total of 77 complexes (51 CORUM complexes and 26 STRING complexes) were identified in PSCs by comparing iGra/H9 and Gra cells. The integrated analysis of microarray data revealed that 47 of the 51 CORUM complexes consisted of the components that displayed down-regulation upon hES cell differentiation. Of those 47 complexes, 41 and 46 were linked with hES cell differentiation into ectodermal and mesendodermal lineages, respectively (Figure 5C and Supporting Information Table S5). Additionally, those 47 complexes included ones that were known to play roles in pluripotency as well as several complexes that were not previously linked with the ESC state, examples showed in Figures 5D and 6A. In addition, two CORUM complexes identified by comparing two types of PSCs (iGra:H9) were also linked with hES cell differentiation (Figure 5C and Supporting Information Figure S9). Furthermore, 24 of the 26 seeds of the STRING complexes showed down-regulation upon hES cell differentiation (Figures 5C and 6B), 15 and 22 of which were linked with hES cell differentiation into ectodermal and mesendodermal lineages, respectively (Supporting Information Table S6). Interestingly, the seeds (n = 8) identified as down-regulated in pluripotent cells by comparing iGra and H9 cells to Gra cells showed up-regulated gene expression upon hES cell differentiation (Figure 6B and Supporting Information Table S6). The detailed information of co-regulated complexes is shown in Supporting Information Table S9. The results of the co-regulation analysis of the protein complexes suggest a link between the complexes and the ESC state.

Associating the Identified Protein Complexes with Cell Reprograming

To find a link between the identified protein complexes and cell reprograming, we looked into a recent proteomics study by Hansson and co-workers, which demonstrated the highly coordinated proteome dynamics during reprograming of mouse embryonic fibroblasts into iPS cells.[25] Hansson and co-workers have identified the expression levels of the components of several protein complexes at multiple stages (0−15 days) of reprograming process. Interestingly, eight protein complexes identified by Hansson et al. were also identified in the present study. They are associated with various functions such as chromatin remodeling (anti-HDAC2 complex), DNA repair (BASC complex), DNA replication (DNA synthesome complex (17 subunits) and MCM complex), mRNA surveillance (exosome), protein biosynthesis (mitochondrial 39S ribosomal subunit), RNA processing (spliceosome), and transport (Nup 107−160 subcomplex). Hansson et al. observed the significant expression of the components of the above complexes during reprograming process, suggesting their critical role in generation of pluripotent cells.

dx.doi.org/10.1021/pr4008877 | J. Proteome Res. 2013, 12, 5878−5890

Associating the Identified Protein Complexes with Embryonic Development

A large number of genetic manipulation studies (e.g., gene knockout) related to developmental processes (e.g., embryonic development) have been performed in mice but not in humans due to ethical issues. Therefore, to explore the significance of the protein complexes with reference to developmental processes, we examined the loss-of-function phenotypes from genetic manipulation studies as described in the Experimental Procedures. We found loss-of-function phenotypes for 42 of the 51 CORUM complexes (Figure 5C and Supporting Information Table S5). We also found loss-of-function phenotypes for 6 of the 26 seeds of the STRING complexes (Figure 5C and Supporting Information Table S6) as well as for the two CORUM complexes identified as differentially regulated between two types of PSCs (iGra:H9) (Figure 5C and Supporting Information Table S10). Loss-of-function phenotypes for 100 of the 332 components of the 51 CORUM complexes were identified. Of these, 83% account for embryonic, perinatal, prenatal, postnatal, preweaning, or neonatal lethality as shown (Supporting Information Table S10). Most significantly, 66% of the genes are linked with embryonic lethality, suggesting the significance of the protein complexes in embryonic developmental processes.

DISCUSSION

Intensive research based on genomic approaches has demonstrated a wealth of information on key components of chromatin organization and transcriptional regulatory networks that contribute to pluripotency and cell reprograming.[21,22,26−28] MS-based proteomic approaches have been proposed to determine the key players of cellular functions at the protein level.[29,30] A targeted approach based on AP-MS analysis has been used to identify the protein complexes of transcription factors such as Oct4, Sox2, and Nanog in PSCs.[2−6] Here, we demonstrated a global approach that revealed that the highly abundant proteins identified in human PSCs are the components of numerous protein complexes involved in diverse cellular processes (Figure 7). The majority of the identified complexes consisted not only of components known to be linked with pluripotency, the ESC state, and developmental processes but also of components that were not previously linked with human PSCs. Because the components of the complexes exhibit protein−protein interaction characteristics, future studies on the characterization of the novel components may indicate their direct and indirect roles in regulating pluripotency, maintaining the ESC state, and cell reprograming.

It is known that the components of ATP-dependent chromatin remodeling complexes are implicated in pluripotency and cell reprograming. In addition, the PcG- and MLL-related complexes play roles in maintaining the ESC state through opposite transcriptional regulation activities.[21,22] Depletion of EED, a PcG complex component, revealed the overexpression of developmental genes in ESCs.[31] WDR5, a member of the MLL complexes, mediates self-renewal via the core transcriptional network, and its expression is required for efficient cell reprogramming.[32] In line with above studies, our proteomic study revealed up-regulation of EED and WDR5 in PSCs (Figure 2C). In PSCs, we also identified the elevated levels of the components of several protein complexes (ALL-1, anti-BHC110, anti-HDAC2, cohesion, LARC, NuRD, PBAF (SWI/SNF), and SIN3) that are known to be linked with chromosome organization and transcriptional activation or repression by protein acetylation, deacetylation, or methylation (Supporting Information Table S5). Of these complexes, ALL-1 (a histone methyl transferase) is a supercomplex that constitutes the components of the SWI/SNF, NuRD, Sin3A, ISWI, TFIID, and Set1 complexes.[33] Here, several novel complexes (e.g., ALL-1, LARC, and anti-BHC110), in addition to the known chromatin remodeling complexes, were identified as being involved in chromatin remodeling, and, thus, may be implicated in cell reprograming.

Recent studies have been reported the proteomic and transcriptomic differences as well as the genetic and epigenetic discrepancies between iPSCs and ESCs.[34,35] One of the reasons for these differences is the fact that the iPSCs maintain a residual memory of their parental somatic cells, which could be due to incomplete reprograming process. Kim and co-workers have demonstrated the epigenetic memory of iPSCs could influence their differentiation capacity.[34,36,37] Therefore, incomplete epigenetic reprograming has been considered as one of the challenges in advancing personalized pluripotent stem cell-based therapeutics.[34] To overcome this challenge, a further improvement in the accuracy of current transcription factor-based reprograming approach has to be achieved. The components of several known and novel chromatin remodeling complexes identified in the present study may have the ability to open up the somatic heterochromatin and further augment the efficiency of somatic cell reprogramming. In this way, the identified complexes, which are associated with chromatin remodeling, may have implications for personalized iPSC-based therapeutics.

Our study not only explored the previous findings linked with chromatin remodeling and transcriptional regulation but also provided information on the regulation of other cellular functions by protein complexes. For example, rapid proliferation is one of the characteristics of PSCs[38] that depend on cell cycle-related processes such as DNA replication, DNA repair, biosynthesis of histone proteins, chromatin formation, and chromosome segregation. The cohesion complex is known to be associated with DNA repair and chromosome segregation, and its contribution to the ESC state has been demonstrated by a genome-wide RNAi study.[39] In the present study, several protein complexes involved in cell-cycle related processes were identified in PSCs (Supporting Information Table S5). Some of these complexes consist of known genes linked with pluripotency and development (Supporting Information Tables S5 and S10). We also observed the co-regulation of cell cycle-related complexes, including cohesion complexes, upon ESC differentiation, which suggests their association with the ESC state (Supporting Information Tables S5 and S9).

The spliceosome, a large protein complex composed of ∼145 distinct proteins,[40] is essential for RNA processing. Notably, a few genes regulating Oct4 expression in human ESCs are associated with the spliceosome complex.[16] In our study, the spliceosome was the most significantly up-regulated protein complex identified, with 106 components in PSCs (Supporting Information Table S5). We found that more than 50% of the spliceosome components showed reduced gene expression levels upon human ESC differentiation (Supporting Information Figure S10). In addition, we determined the roles of several components of the spliceosome in embryonic development (Supporting Information Table S10). These results suggest that the spliceosome is linked with the ESC state, and that RNA processing mediated by the spliceosome plays significant roles in pluripotency and development.

dx.doi.org/10.1021/pr4008877 | J. Proteome Res. 2013, 12, 5878−5890

The exosome complex is known to be involved in RNA catabolism and a recent study demonstrated that EXOSC9 inhibits differentiation and maintains self-renewal in human epidermal tissue.[41] We identified an exosome complex with 11 components including EXOSC9, in PSCs (Supporting Information Table S5). Nine of the 11 components showed reduced gene expression upon ESC differentiation (Figure 6A). Furthermore, a total of five seeds related to STRING complexes involved in developmental processes were identified in PSCs (Figure 4). Of these five seeds, pluripotency-associated seeds (CTR9 and LEO1) were up-regulated, while system or organ development-associated seeds (COL6A3, DCTN1, and TPP1) were down-regulated. These observations suggest that PSCs maintain high and low levels of the seeds that promote pluripotency and differentiation, respectively. The components including the seeds of protein complexes identified in this study may assist to extend our knowledge on the regulators of pluripotency and the ESC state.

CONCLUSIONS

Methods need to be developed for integrated analysis of various large-scale data sets that provides system-level knowledge of the pluripotent stem cell-associated functions. Our integrated approach has demonstrated the protein complexes on a proteome-wide scale to explore the regulators of a broad range of cellular roles, including chromatin remodeling, gene expression, RNA splicing, protein acetylation and deacetylation, pluripotency, and embryonic development. Notably, further characterization of the members of the chromatin remodeling complexes identified may provide information to promote the binding efficiency of transcription factors that have been used in cell reprograming. The data presented here not only connect genomic and proteomic studies but also serve as a basis for future investigations that may provide insight into stem cell biology.

ASSOCIATED CONTENT

Supporting Information

Additional figures and tables as described in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*Hung-Chih Kuo, Tel: +886-2-2789-9580; Fax: +886-2-2789-9587; E-mail: kuohuch@gate.sinica.edu.tw.
*Chung-Hsuan Chen, Tel: +886-2-2787-1202; Fax: +886-2-2789-9923; E-mail: winschen@gate.sinica.edu.tw.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The financial grant (NSC 99-2113-M-001-002-MY3) to CHC, and the postdoctoral fellowship (NSC 101-2811-M-001-066) to PRS from National Science Council, Taiwan are greatly appreciated. This work was partially supported by National Health Research Center (NHRI-EX101-9803EI) and GRC, Academia Sinica, Taiwan. Technical support by Tzu-Ning Ho in microarray core facility in GRC and Dr. Chun-Ying Yu, ICOB are also acknowledged.

ABBREVIATIONS:

ESCs, embryonic stem cells; iPSCs, induced pluripotent stem cells; PSCs, pluripotent stem cells; CV, coefficient of variation; TRs, technical replicates; EBs, embryoid bodies; CORUM, Comprehensive Resource of Mammalian protein complexes; STRING, Search Tool for the Retrieval of Interacting Genes/Proteins; GO, Gene Ontology

REFERENCES

(1) Stadtfeld, M.; Hochedlinger, K. Induced pluripotency: history, mechanisms, and applications. *Genes Dev.* **2010**, *24* (20), 2239−63.
(2) Liang, J.; Wan, M.; Zhang, Y.; Gu, P.; Xin, H.; Jung, S. Y.; Qin, J.; Wong, J.; Cooney, A. J.; Liu, D.; Songyang, Z. Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. *Nat. Cell Biol.* **2008**, *10* (6), 731−9.
(3) Mallanna, S. K.; Ormsbee, B. D.; Iacovino, M.; Gilmore, J. M.; Cox, J. L.; Kyba, M.; Washburn, M. P.; Rizzino, A. Proteomic analysis of Sox2-associated proteins during early stages of mouse embryonic stem cell differentiation identifies Sox21 as a novel regulator of stem cell fate. *Stem Cells* **2010**, *28* (10), 1715−27.
(4) Pardo, M.; Lang, B.; Yu, L.; Prosser, H.; Bradley, A.; Babu, M. M.; Choudhary, J. An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. *Cell Stem Cell* **2010**, *6* (4), 382−95.
(5) van den Berg, D. L.; Snoek, T.; Mullin, N. P.; Yates, A.; Bezstarosti, K.; Demmers, J.; Chambers, I.; Poot, R. A. An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell* **2010**, *6* (4), 369−81.
(6) Wang, J.; Rao, S.; Chu, J.; Shen, X.; Levasseur, D. N.; Theunissen, T. W.; Orkin, S. H. A protein interaction network for pluripotency of embryonic stem cells. *Nature* **2006**, *444* (7117), 364−8.
(7) 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.
(8) Phanstiel, D. H.; Brumbaugh, J.; Wenger, C. D.; Tian, S.; Probasco, M. D.; Bailey, D. J.; Swaney, D. L.; Tervo, M. A.; Bolin, J. M.; Ruotti, V.; Stewart, R.; Thomson, J. A.; Coon, J. J. Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nat. Methods* **2011**, *8* (10), 821−7.
(9) Huang, H. P.; Chen, P. H.; Yu, C. Y.; Chuang, C. Y.; Stone, L.; Hsiao, W. C.; Li, C. L.; Tsai, S. C.; Chen, K. Y.; Chen, H. F.; Ho, H. N.; Kuo, H. C. Epithelial cell adhesion molecule (EpCAM) complex proteins promote transcription factor-mediated pluripotency reprogramming. *J. Biol. Chem.* **2011**, *286* (38), 33520−32.
(10) Sudhir, P. R.; Chen, C. H.; Pavana Kumari, M.; Wang, M. J.; Tsou, C. C.; Sung, T. Y.; Chen, J. Y. Label-free quantitative proteomics and N-glycoproteomics analysis of KRAS-activated human bronchial epithelial cells. *Mol. Cell. Proteomics* **2012**, *11* (10), 901−15.
(11) Hubner, N. C.; Ren, S.; Mann, M. Peptide separation with immobilized pI strips is an attractive alternative to in-gel protein digestion for proteome analysis. *Proteomics* **2008**, *8* (23−24), 4862−72.
(12) Geiger, T.; Wehner, A.; Schaab, C.; Cox, J.; Mann, M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol. Cell. Proteomics* **2012**, *11* (3), No. M111 014050.
(13) Wisniewski, J. R.; Ostasiewicz, P.; Dus, K.; Zielinska, D. F.; Gnad, F.; Mann, M. Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma. *Mol. Syst. Biol.* **2012**, *8*, 611.
(14) Sharma, K.; Vabulas, R. M.; Macek, B.; Pinkert, S.; Cox, J.; Mann, M.; Hartl, F. U. Quantitative proteomics reveals that Hsp90 inhibition preferentially targets kinases and the DNA damage response. *Mol. Cell. Proteomics* **2012**, *11* (3), M111 014654.
(15) Stingele, S.; Stoehr, G.; Peplowska, K.; Cox, J.; Mann, M.; Storchova, Z. Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. *Mol. Syst. Biol.* **2012**, *8*, 608.

(16) Chia, N. Y.; Chan, Y. S.; Feng, B.; Lu, X.; Orlov, Y. L.; Moreau, D.; Kumar, P.; Yang, L.; Jiang, J.; Lau, M. S.; Huss, M.; Soh, B. S.; Kraus, P.; Li, P.; Lufkin, T.; Lim, B.; Clarke, N. D.; Bard, F.; Ng, H. H. A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* **2010**, *468* (7321), 316−20.
(17) Choudhary, C.; Kumar, C.; Gnad, F.; Nielsen, M. L.; Rehman, M.; Walther, T. C.; Olsen, J. V.; Mann, M. Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **2009**, *325* (5942), 834−40.
(18) Bader, G. D.; Hogue, C. W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinf.* **2003**, *4*, 2.
(19) Mallon, B. S.; Chenoweth, J. G.; Johnson, K. R.; Hamilton, R. S.; Tesar, P. J.; Yavatkar, A. S.; Tyson, L. J.; Park, K.; Chen, K. G.; Fann, Y. C.; McKay, R. D. StemCellDB: the human pluripotent stem cell database at the National Institutes of Health. *Stem Cell Res.* **2013**, *10* (1), 57−66.
(20) Muller, F. J.; Laurent, L. C.; Kostka, D.; Ulitsky, I.; Williams, R.; Lu, C.; Park, I. H.; Rao, M. S.; Shamir, R.; Schwartz, P. H.; Schmidt, N. O.; Loring, J. F. Regulatory networks define phenotypic classes of human stem cell lines. *Nature* **2008**, *455* (7211), 401−5.
(21) Orkin, S. H.; Hochedlinger, K. Chromatin connections to pluripotency and cellular reprogramming. *Cell* **2011**, *145* (6), 835−50.
(22) Sha, K.; Boyer, L. A. The Chromatin Signature of Pluripotent Cells *StemBook*; Girard , L., Ed.; Harvard Stem Cell Institute, Harvard University: Cambridge, MA, 2009; DOI: 10.3824/stembook.1.45.1.
(23) Ruepp, A.; Waegele, B.; Lechner, M.; Brauner, B.; Dunger-Kaltenbach, I.; Fobo, G.; Frishman, G.; Montrone, C.; Mewes, H. W. CORUM: the comprehensive resource of mammalian protein complexes—2009. *Nucleic Acids Res.* **2010**, *38* (Database issue), D497−501.
(24) 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.
(25) Hansson, J.; Rafiee, M. R.; Reiland, S.; Polo, J. M.; Gehring, J.; Okawa, S.; Huber, W.; Hochedlinger, K.; Krijgsveld, J. Highly coordinated proteome dynamics during reprogramming of somatic cells to pluripotency. *Cell Rep.* **2012**, *2* (6), 1579−92.
(26) Boyer, L. A.; Lee, T. I.; Cole, M. F.; Johnstone, S. E.; Levine, S. S.; Zucker, J. P.; Guenther, M. G.; Kumar, R. M.; Murray, H. L.; Jenner, R. G.; Gifford, D. K.; Melton, D. A.; Jaenisch, R.; Young, R. A. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **2005**, *122* (6), 947−56.
(27) Kim, J.; Chu, J.; Shen, X.; Wang, J.; Orkin, S. H. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **2008**, *132* (6), 1049−61.
(28) Loh, Y. H.; Wu, Q.; Chew, J. L.; Vega, V. B.; Zhang, W.; Chen, X.; Bourque, G.; George, J.; Leong, B.; Liu, J.; Wong, K. Y.; Sung, K. W.; Lee, C. W.; Zhao, X. D.; Chiu, K. P.; Lipovich, L.; Kuznetsov, V. A.; Robson, P.; Stanton, L. W.; Wei, C. L.; Ruan, Y.; Lim, B.; Ng, H. H. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* **2006**, *38* (4), 431−40.
(29) Brumbaugh, J.; Rose, C. M.; Phanstiel, D. H.; Thomson, J. A.; Coon, J. J. Proteomics and pluripotency. *Crit. Rev. Biochem. Mol. Biol.* **2011**, *46* (6), 493−506.
(30) Whetton, A. D.; Williamson, A. J.; Krijgsveld, J.; Lee, B. H.; Lemischka, I.; Oh, S.; Pera, M.; Mummery, C.; Heck, A. J. The time is right: proteome biology of stem cells. *Cell Stem Cell* **2008**, *2* (3), 215−7.
(31) Chamberlain, S. J.; Yee, D.; Magnuson, T. Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. *Stem Cells* **2008**, *26* (6), 1496−505.
(32) Ang, Y. S.; Tsai, S. Y.; Lee, D. F.; Monk, J.; Su, J.; Ratnakumar, K.; Ding, J.; Ge, Y.; Darr, H.; Chang, B.; Wang, J.; Rendl, M.; Bernstein, E.; Schaniel, C.; Lemischka, I. R. Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **2011**, *145* (2), 183−97.
(33) Nakamura, T.; Mori, T.; Tada, S.; Krajewski, W.; Rozovskaia, T.; Wassell, R.; Dubois, G.; Mazo, A.; Croce, C. M.; Canaani, E. ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. *Mol. Cell* **2002**, *10* (5), 1119−28.
(34) Byrne, J. A. Nuclear reprogramming and the current challenges in advancing personalized pluripotent stem cell-based therapies. *Gene Ther. Regul.* **2012**, *7* (1), No. 1230002.
(35) Wang, T.; Wu, H.; Li, Y.; Szulwach, K. E.; Lin, L.; Li, X.; Chen, I. P.; Goldlust, I. S.; Chamberlain, S. J.; Dodd, A.; Gong, H.; Ananiev, G.; Han, J. W.; Yoon, Y. S.; Rudd, M. K.; Yu, M.; Song, C. X.; He, C.; Chang, Q.; Warren, S. T.; Jin, P. Subtelomeric hotspots of aberrant 5-hydroxymethylcytosine-mediated epigenetic modifications during reprogramming to pluripotency. *Nat. Cell Biol.* **2013**, *15* (6), 700−11.
(36) 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.
(37) Kim, K.; Zhao, R.; Doi, A.; Ng, K.; Unternaehrer, J.; Cahan, P.; Huo, H.; Loh, Y. H.; Aryee, M. J.; Lensch, M. W.; Li, H.; Collins, J. J.; Feinberg, A. P.; Daley, G. Q. Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells. *Nat. Biotechnol.* **2011**, *29* (12), 1117−9.
(38) Becker, K. A.; Stein, J. L.; Lian, J. B.; van Wijnen, A. J.; Stein, G. S. Establishment of histone gene regulation and cell cycle checkpoint control in human embryonic stem cells. *J. Cell. Physiol.* **2007**, *210* (2), 517−26.
(39) Kagey, M. H.; Newman, J. J.; Bilodeau, S.; Zhan, Y.; Orlando, D. A.; van Berkum, N. L.; Ebmeier, C. C.; Goossens, J.; Rahl, P. B.; Levine, S. S.; Taatjes, D. J.; Dekker, J.; Young, R. A. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **2010**, *467* (7314), 430−5.
(40) Zhou, Z.; Licklider, L. J.; Gygi, S. P.; Reed, R. Comprehensive proteomic analysis of the human spliceosome. *Nature* **2002**, *419* (6903), 182−5.
(41) Mistry, D. S.; Chen, Y.; Sen, G. L. Progenitor function in self-renewing human epidermis is maintained by the exosome. *Cell Stem Cell* **2012**, *11* (1), 127−35.