Published in final edited form as:

Proteomics. 2015 January ; 15(0): 554–566. doi:10.1002/pmic.201400132.

The human embryonic stem cell proteome revealed by multidimensional fractionation followed by tandem mass spectrometry

Peng Zhao¹, Thomas C. Schulz², Eric S. Sherrer², D. Brent Weatherly¹, Allan J. Robins², and Lance Wells¹

¹Complex Carbohydrate Research Center, University of Georgia, Athens, GA, USA

²ViaCyte Inc., Athens, GA, USA

Abstract

Human embryonic stem cells (hESCs) have received considerable attention due to their therapeutic potential and usefulness in understanding early development and cell fate commitment. In order to appreciate the unique properties of these pluripotent, self-renewing cells, we have performed an in-depth multidimensional fractionation followed by LC-MS/MS analysis of the hESCs harvested from defined media to elucidate expressed, phosphorylated, *O*-linked β -*N*-acetylglucosamine (O-GlcNAc) modified, and secreted proteins. From the triplicate analysis, we were able to assign more than 3000 proteins with less than 1% false-discovery rate. This analysis also allowed us to identify nearly 500 phosphorylation sites and 68 sites of O-GlcNAc modification with the same high confidence. Investigation of the phosphorylation sites allowed us to deduce the set of kinases that are likely active in these cells. We also identified more than 100 secreted proteins of hESCs that likely play a role in extracellular matrix formation and remodeling, as well as autocrine signaling for self-renewal and maintenance of the undifferentiated state. Finally, by performing in-depth analysis in triplicate, spectral counts were obtained for these proteins and posttranslationally modified peptides, which will allow us to perform relative quantitative analysis between these cells and any derived cell type in the future.

Keywords

Cell biology; Embryonic stem cells; O-GlcNAc; Phosphorylation; Secretome; Shotgun proteomics

1 Introduction

Embryonic stem cells (ESCs), which are derived from blastocysts, are pluripotent, having the theoretical potential to become most specialized cell types. Studies using ESCs not only provide basic information about early development and cell fate decisions, but may also serve as a potential source for the creation of large populations of a defined cell type. The

Correspondence: Dr. Lance Wells, Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Road, Athens, GA 30602, USA. lwells@ccrc.uga.edu, Fax: +1-706-542-7806.

We declare no conflicts of interest with the publication of this manuscript.

Additional supporting information may be found in the online version of this article at the publisher's web-site

latter subject raises possibilities of cell-based therapies to treat disease, often referred to as regenerative medicine. Data now show that human ESCs (hESCs) can initiate lineage-specific differentiation programs of various types of tissue and cells in vitro [1, 2]. For example, the creation of primitive endoderm and mesoderm under defined media raises the realistic possibility of producing cardiac progenitors for heart tissue regeneration in cardiovascular disease therapies [3], and the recent development in generating functional pancreatic progenitors from hESCs on a scalable system for the therapy of type 1 diabetes [4].

Given the likely importance of these cells in basic and translational science, several genome and transcriptome studies have been performed. Likewise, several proteomic efforts have elucidated some of the proteins expressed by these cells and their derivatives [5–12].

Multidimensional chromatography coupled with MS-based protein assignment technology has dramatically increased resolution and loading capacity for identifying large number of proteins with a wide dynamic range of abundances, especially for proteins with PTMs such as phosphorylation and glycosylation [13–16]. Here we describe a large-scale proteomic study of hESCs grown under defined media without serum or feeder layers using multidimensional chromatography coupled with MS/MS. We applied strong cation exchange (SCX) and RP LC to provide in-depth peptide separation, and LC coupled with MS/MS (LC-MS/MS) techniques for protein assignment. We also investigated the PTMs of the hESC proteome, particularly the modification by phosphate and *O*-linked β -*N*-acetylglucosamine (*O*-GlcNAc). Furthermore, since the cells were cultured in defined media, we were able to explore the secreted proteome of these cells. Functional classification analysis of this large, high-confidence dataset obtained from hESC proteome elucidates many of the biological features of this pluripotent cell type.

2 Materials and methods

2.1 Culture of hESCs and collection of secreted proteins

Karyotypically normal cultures of the NIH-registered, formerly approved BG02 hES cell line were used throughout the experiments. BG02 cells were grown in DC-HAIF defined media [17], which consisted of DMEM/F12 (Invitrogen), 2% fatty acid free Cohn's fraction V BSA (Serologicals), $1\times$ nonessential amino acids, 50 U/mL penicillin, 50 μ g/mL streptomycin, 50 μ g/mL ascorbic acid, 10 μ g/mL transferrin, 0.1 mM β -mercaptoethanol (all from Invitrogen), $1\times$ trace elements A, B, and C (Mediatech), 10 ng/mL HRG1 β (Peprotech), 10 ng/mL ActA (R&D Systems), 200 ng/mL LR³-IGF1 (JRH Biosciences), and 8 ng/mL FGF2 (Sigma or R&D Systems). Cultures were expanded by passaging with Accutase (Innovative Cell Technologies) and plating on tissue culture flasks coated with growth factor reduced matrigel (BD Biosciences) diluted 1:200, as described [18]. To create samples for proteomics, near-confluent cultures were washed extensively with PBS and grown for 24 h in DC-HAIF without BSA. Cultures were harvested with Accutase, washed with PBS, pelleted, and snap-frozen. Initial experiments were performed on the undifferentiated population, with an estimated 5×10^7 cells provided for each sample run (multidimensional protein identification technology (MudPIT) and single-dimensional LC-MS experiments). The media were harvested with extreme care and then centrifuged once at

1800 rpm at 4°C for 7 min. The supernatants were filtered using 1 µm syringe filters (PALL). The secreted protein sample was then collected in centrifuge tubes and centrifuged again at 30 000 G at 4°C for 30 min.

2.2 Subcellular fractionation, precipitation, and digestion of proteins

Cells were fractionated into four subcellular proteomes using ProteoExtract® Subcellular Proteome Extraction Kit (Calbiochem): membrane/organelle, nucleus, cytosol, and cytoskeleton. Specifically, the pellet was resuspended and rotated at 4°C for 10 min in ice-cold Extraction Buffer I (Calbiochem) with protease and phosphatase inhibitors (Sigma). The suspension was then centrifuged for 10 min at 800 G at 4°C and the supernatant was collected as cytosol fraction. The remaining pellet was resuspended and rotated at 4°C for 30 min in ice-cold Extraction Buffer II (Calbiochem) with protease and phosphatase inhibitors (Sigma). The suspension was then centrifuged for 10 min at 5500 G at 4°C and the supernatant was collected as membrane/organelle fraction. The remaining pellet was resuspended and rotated at 4°C for 10 min in ice-cold Extraction Buffer III and benzonase nuclease (Calbiochem) with protease and phosphatase inhibitors (Sigma). The suspension was then centrifuged for 10 min at 7000 G at 4°C and the supernatant was collected as nucleus fraction. The remaining pellet was resuspended in ice-cold Extraction Buffer IV (Calbiochem) with protease and phosphatase inhibitors (Sigma) and then collected as cytoskeleton fraction.

Proteins from the four subproteomes were then precipitated using ProteoExtract® Protein Precipitation Kit (Calbiochem). Each fraction was incubated in Precipitation Agent (Calbiochem) at -20°C for 1 h and then centrifuged for 10 min at 10 000 G at 4°C. The resulting pellet was washed three times with ice-cold Wash Solution (Calbiochem) and centrifuged for 2 min at 10 000 G at 4°C. The pellet from each fraction was collected, dried at room temperature for 15 min, and resuspended in the mixture of 8 M urea, 10 mM DTT, 40 mM NH₄CO₃, and incubated at 37°C for 1 h. After cooling to room temperature, the suspension was diluted to 2 M of urea, and then incubated in 40 mM NH₄CO₃ with 55 mM iodoacetamide in dark for 45 min. After that, each fraction was digested overnight at 37°C using trypsin (Sequencing grade modified trypsin, Promega).

2.3 SCX LC

Peptides generated from digestion were separated by SCX-LC using an Agilent 1100 series HPLC system (Agilent Technologies). Solvent A (5 mM KH₂PO₄/30% ACN, pH 2.7), solvent B (solvent A with 350 mM KCl), and solvent C (0.1 M Tris/0.5 M KCl, pH 7.0) were used to develop a salt gradient consisting of 5 min at 100% solvent A, 48 min gradient at variable slope to 100% solvent B, 12 min at 100% solvent B, 15 min to 100% solvent C, and 10 min to 100% solvent A. Dried peptides of each subproteome were resuspended in 50 µL of SCX solvent A and separated on a 2.1 × 100 mm Polysulfoethyl A column (PolyLC) at a flow rate of 150 µL/min. Fractions were collected every 2 min and then combined into five fractions, desalted, and dried.

2.4 RP LC

Each fraction generated by SCX-LC (5×4 fractions in total) was separated by RP-LC using the same Agilent 1100 series HPLC system (Agilent Technologies) as described in [19]. Briefly, solvent A (0.1% TFA) and solvent B (0.085% TFA/80% ACN) were used to develop a linear gradient consisting of 5 min at 95% solvent A, 60 min gradient at variable slope to 95% solvent B, 3 min at 95% solvent B, 1.5 min to 95% solvent A, and 4.5 min at 95% solvent A. Dried peptides of each SCX fractions were dissolved in 20 μ L of RP solvent A and separated on a 2.1×250 mm silica-based C18 column (VYDAC) at a flow rate of 100 μ L/min. Fractions were collected every 4 min and then combined into five fractions (F1, 15–32%; F2, 32–40%; F3, 40–45%; F4, 45–55%; and F5, 55–85%), desalted, and dried.

2.5 RP nanoLC-MS/MS analysis

As described in [19], dried peptides from each fraction generated by RP-LC (20×5 fractions in total) were resuspended in 0.5 μ L of solvent B (0.1% formic acid/80% ACN) and 19.5 μ L of solvent A (0.1% formic acid) and loaded on a $75 \mu\text{m} \times 85$ mm C18 RP column (packed in house, YMC GEL ODS-AQ120 $\text{\AA}$ S-5, Waters) by nitrogen bomb. Peptides collected from each offline RP-LC separation (F1–F5) were eluted directly into the nanospray source of a linear ion trap (Thermo Finnigan LTQTM) with five respective 140-min linear gradients consisting of 5–100% solvent B over 90–95 min at a flow rate of ~ 250 nL/min. In order to optimize the separation of peptides eluted into the mass spectrometer, gradients were expanded over a 70-min period in the appropriate region corresponding to each fraction collected from the previous RP-LC separation (F1, 4–30%; F2, 9–35%; F3, 15–42%; F4, 20–55%; and F5, 28–85%). The spray voltage was set to 2.0 kV, and the temperature of the heated capillary was set to 210°C. Full-scan MS spectra were acquired from m/z 300–2000 followed by eight MS/MS events of the most intense ions. Subsequent MS³ events were triggered by the detection of neutral losses of phosphoric acid (98 Da) or GlcNAc (203.08 Da) between precursor ions and performed on the resulting three most intense product ions. All fragmentation was performed under CID. A dynamic exclusion window was applied to prevent the same m/z value from being selected for 12 s after its acquisition. All 100 subfractions were analyzed in three technical replicates. Data were automatically acquired using Xcalibur[®] (ver. 2.0.7, Thermo Fisher Scientific). For protein identification in mixture by MS/MS, it has been shown that repeated experiments are required in order to reach a reasonable completeness (all predicted proteins in the mixture being identified). A statistical evaluation performed by Yates' group found that triplicate analysis can discover approximately 95% of all predicted proteins in samples with relatively high complexity in an LC-MS/MS experiment [20].

2.6 Single-dimensional LC-MS/MS experiment

In order to compare the depths of protein identification at an equal level of material loaded for each LS-MS/MS run between single-dimensional and multidimensional experiments, approximately 1% of the peptides generated from previous tryptic digestions described above were directly analyzed by online nanospray LC coupled with linear ion trap (Thermo Finnigan LTQTM) without prior chromatography separation (SCX or RP-LC). The online RP-LC gradient used for peptides of each subcellular fraction was a 160-min linear gradient

consisting of 5–100% solvent B over 100 min at a flow rate of ~250 nL/min. The parameters of the instrument setup and analysis method were as described above.

2.7 Secreted proteome of hESCs

The secreted proteins were digested and the resulting peptides were separated by SCX chromatography as described above. As described in [21], dried peptides from each fraction (five fractions in total) were resuspended in 0.5 µL of solvent B (0.1% formic acid/80% ACN) and 19.5 µL of solvent A (0.1% formic acid) and loaded on a 75 µm × 105 mm C18 RP column (packed in house, YMC GEL ODS-AQ120ÅS-5, Waters) by nitrogen bomb. Peptides were eluted directly into the nanospray source of an LTQ Orbitrap XL™ (Thermo Fisher Scientific) with a 160-min linear gradient consisting of 5–100% solvent B over 100 min at a flow rate of ~250 nL/min. The spray voltage was set to 2.0 kV and the temperature of the heated capillary was set to 200°C. Full-scan MS spectra were acquired from m/z 300–2000 with a resolution of 60 000 at m/z 400 after accumulation of 10^6 ions (mass accuracy < 2 ppm). MS/MS events under CID were triggered by the six most intense ions from the preview of full scan and a dynamic exclusion window was applied, which prevents the same m/z value from being selected for 6 s after its acquisition. All five subfractions were analyzed in technical triplicates. Data were acquired using Xcalibur® (ver. 2.0.7, Thermo Fisher Scientific).

2.8 Data analysis

For protein identification, MSⁿ data were searched against UniProt human proteome database (32 876 entries, August 13, 2007) using SEQUEST (Bioworks 3.3, Thermo Fisher Scientific) with the following settings [19]: 1000 ppm (10 ppm for data acquired using LTQ Orbitrap XL™) tolerance was set for precursor masses and 0.5 Da for fragment masses; trypsin was specified as the enzyme and only fully tryptic peptide identifications were retained; a maximum of three missed cleavage sites, three differential amino acids per modification, and three differential modifications per peptide were allowed; oxidation of methionine (+15.99 Da), carbamidomethylation of cysteine (+57.02 Da), phosphorylation of serine/threonine/tyrosine (+79.97 Da), and O-GlcNAc modification of serine/threonine (+203.08 Da) were set as differential modifications. All the raw spectra were searched against both normal (forward) and reverse databases under the same parameters.

SEQUEST search results were then submitted to ProteoIQ (www.nusep.com) for proteome validation and analysis. All of the output files from SEQUEST search were first filtered to achieve a 1% false-discovery rate at protein level using the ProValT algorithm [22], as deployed in PROTEOIQ, using a starting peptide coverage of 3 and the peptide discriminant score as the metric for the calculations [23, 24].

Where applicable, proteins were submitted to the following external resources for biological interpretation: (i) Ingenuity (www.ingenuity.com) for function and pathway classification, (ii) PSORT (wolfpsort.org) for subcellular localization, (iii) Human Protein Reference Database (HPRD, www.hprd.org), (iv) GO (www.geneontology.org), (v) UniProt Protein Knowledgebase (www.uniprot.org), (vi) Scansite (www.scansite.mit.edu) for kinase prediction (63 motifs available).

2.9 Validation of phosphorylated and O-GlcNAc-modified peptides of hESCs

In addition to the statistical validation described above, all peptides assigned with phosphate (79.97 Da) or O-GlcNAc (203.08 Da) modification were further filtered using the following criteria: (i) candidate peptides could not have both phosphate and O-GlcNAc on the same sequence, (ii) candidate peptides had to be identified by SEQUEST with an Xcorr above 1.8 (single-charged), 2.2 (double-charged), or 2.5 (triple-charged), (iii) candidate proteins were required to have at least one unmodified peptide assignment, and (iv) the MS/MS spectra of the candidate peptides were manually verified; in the case of phosphorylated serines and threonines, candidate peptides were required to show a neutral loss of phosphoric acids (98 Da) in the spectra.

3 Results

3.1 Assignment of hESC proteome

As described above and presented in Fig. 1, the hESC proteomic analysis resulted in more than 300 LC-MS/MS experiments in total. Database searches were performed by SEQUEST and the resulting assignments were filtered at 1% protein-level false-discovery rate using ProteoIQ™. After eliminating duplicates and contaminants, the combined dataset consists of 3184 proteins with approximately 24,000 peptides, 78% of which (2469/3184) were identified by more than one peptide (Table 1, Supporting Information Tables 1 and 11). Thus, our MS/MS analysis combined with stringent cutoff filters in ProteoIQ™ resulted in an in-depth, high-confidence proteome.

Additionally, 73% of the proteins (2328/3184) were identified in at least two of the technical triplicates, including 1721 (54.1%) that were detected in all three replicates (Fig. 2A, Supporting Information Table 2). The lack of 100% or close-to-100% reproducibility can be explained by the incomplete sampling of data-dependent acquisition by tandem mass spectrometer [20]. Among the 1721 common proteins observed in technical triplicates, 464 (27.0%) were detected in all of the four subcellular fractions (cytosol, membrane/organelle, nucleus, and cytoskeleton), 506 (29.4%) were detected in three of the four fractions, and 474 (27.5%) proteins were detected in two fractions (Fig. 2B). Two hundred seventy-seven (16.1%) proteins were detected in only one of the four subcellular fractions, and further analysis showed that 51% of those proteins were uniquely identified in the nucleus subcellular fraction and 22% were in the cytosol fraction (Fig. 2B), which is consistent with the result from biological function analysis, presented in Fig. 2C, observing that the majority of the proteins were identified as localized in cytoplasm and nucleus. A substantial amount of proteins detected in the membrane/organelle subcellular fraction were identified mostly as cytosolic or nuclear proteins, indicating a certain level of cross-contamination during subcellular protein extraction. Interestingly, 1.6% of the dataset was identified as extracellular proteins, likely reflecting proteins that were being processed through the secretory pathway, being recycled, or that are associated with the plasma membrane (Fig. 2C). Further interrogation on the entire proteomic dataset via Ingenuity Pathway Analysis discovered that the proteins identified in hESCs were involved in a variety of biological functions related to transcription, synthesis, molecular transport, DNA repair, and cell signaling as was to be expected (Supporting Information Table 3).

3.2 Assignment of the posttranslationally modified proteome in hESCs

After manual validation of the peptides assigned with phosphate and O-GlcNAc modification, we report 491 sites of phosphorylation on 287 proteins (Table 1, Supporting Information Tables 4 and 5) and 68 sites of O-GlcNAc modification on 55 proteins (Table 1, Supporting Information Tables 6 and 7). Among all the validated phosphorylated peptides carrying 491 sites (Supporting Information Table 5), 71.8% (310/431) contained multiple potential phosphorylation sites. Because the gas-phase rearrangements of phosphate groups during CID may hinder site confirmation on such peptides [25], 52.4% (258/491) of the validated sites may be ambiguous. However, it has also been argued that the gas-phase rearrangements should not affect phosphorylation site localization in any statistically significant way [26].

In order to investigate which specific kinases are likely active in hESCs, the validated dataset of phosphorylated proteins were then submitted to Scansite (<http://www.scansite.mit.edu>, 63 motifs available) for kinase prediction. With high-stringency scanning, we were able to discover 11 different motifs (four motif families) interacting with at least 55 sites on 45 proteins from our dataset (Table 2).

3.3 Assignment of hESCs secreted proteome

After eliminating duplicates and contaminants from cell-culture medium and basement membrane matrix (Matrigel™), the combined dataset consists of 123 proteins with approximately 500 peptides (data not shown). By searching against the HPRD, we were able to determine that 30 of the 123 proteins are supported by literature as secreted proteins. We suspected that the other 93 proteins might have originated as intracellular proteins but being inadvertently detected in the extracellular matrix (ECM) due to cell apoptosis. We compared the 93 proteins with the 1000 most abundant proteins (ranked by spectral counts) in our entire proteomic dataset and were able to classify 57 of them into various subgroups of intracellular proteins (cytoplasm, nucleus, etc.), while the rest of them remained undetermined. In addition, we had identified extracellular proteins in the cell proteomic analysis (2%, 74 proteins) and the combination of those two datasets resulted in a total of 102 secreted proteins from hESCs (Table 1, Supporting Information Table 8).

4 Discussion

In this study, we combined various protein and peptide separation techniques with MS to reduce the sample complexity and increase protein assignment and coverage of hESCs grown under the defined media. This strategy resulted in the identification of 3184 proteins in hESCs including most of the well-established stem cell markers, such as Oct4 and Sox2. In our hESC sample, without intentional enrichment during experiments, we discovered 491 sites of phosphorylation and 68 sites of O-GlcNAc modification that have each been stringently filtered and manually examined. Furthermore, because of the unique feeder-free condition under which the cells were cultured, we were able to identify 102 proteins that were secreted from hESCs.

4.1 The proteome of hESCs

In order to validate our results and explore the scope of the hESC proteome, we compared the identified proteins in our dataset to previous hESC proteomic studies that used similar techniques [27–29], instead of the ones that have employed quantitative methods, different and/or more advanced instrumentation [5–12]. Baharvand et al. [27] identified 685 proteins from three different hES cell lines (Royan H2, Royan H3, and Royan H5) by employing 2DE combined with MALDI TOF-TOF-MS. By comparing with our dataset, only 75 (11%) proteins were found in common. Van Hoof et al. [29] identified 1775 proteins from undifferentiated hES cell line HES-2 by performing 2DE separation and FT-ICR-MS/MS. After comparison, 1065 (60%) proteins were found in common with our dataset, which indicates certain consistency in the results both obtained from MS-based proteomic experiments. Finally, Schulz et al. [28] identified 589 proteins from undifferentiated hES cell line BG01 by using 934 antibodies in a large-scale Western blotting system. By comparison, 134 (23%) proteins were found in common, and most of them were identified as relatively low-abundance transcription and translation regulators, receptors, and transporters, which can be explained by the specificity of antibodies used in the experiment. The discrepancy between the two datasets not only relates to the biological variations between cell lines and differences in the analyzing techniques but also reflects upon the complexity of the hES cell proteome. It is noteworthy that the two RP LC separations in our experiments were both operated under acidic condition, which could be improved to provide better orthogonality by employing two RP-LC systems under different pHs or pressures [30, 31].

Several classes of proteins with particular relevance to hES cell biology were observed in our experiment. For instances, nuclear factors that are known to regulate self-renewal and others that mark the pluripotent state were identified including Oct4, Sox2, UTF1, DPPA4, and FoxD3 [32–35].

The POU transcription factor Oct4 (encoded by POU5F1) plays a critical role in maintaining the undifferentiated state of ES cells [36–39]. It has been demonstrated that the level of Oct4 protein activity defines differentiation, dedifferentiation, and self-renewal of ES cells, which suggests that Oct4 expression is regulated at the optimal level to ensure pluripotency in ES cells [37]. Sox2 is another transcription factor that is essential to the maintenance of self-renewal in pluripotent ES cells. Sox2 often works with Oct4 to regulate gene expression [40, 41]. Recent studies showed that Sox2 is necessary for regulating multiple transcription factors that affect Oct4 expression, thus maintaining the requisite level of Oct4 for pluripotent state in ES cells [42]. Oct4 and Sox2 are also key participants in reprogramming somatic cells back to undifferentiated states (induced pluripotent stem cells) [43, 44].

Mammalian pluripotent cells including the zygote, blastomeres, inner cell mass, primitive ectoderm, primordial germ cells, gonadal germ cells, and gametes form a developmental life cycle that ensures continued propagation of the species. Consequently, it is expected that these cells and their in vitro derivatives will have stringent mechanisms to ensure genomic integrity. While aneuploidies and genomic alterations can arise sporadically in hES cell cultures [45, 46], hESCs have also been shown to be remarkably stable over prolonged culture periods [47]. We identified a network of DNA repair and replication proteins

expressed by hESCs (Supporting Information Fig. 1), consisting of DNA mismatch and double-strand break repair factors, repair recruitment factors, chromosome cohesion and mitotic chromosome associated factors, as well as replication factors and s-phase checkpoint regulators. Our data lend mechanistic support to recent findings demonstrating that hESCs exhibit enhanced repair of multiple forms of DNA damage and resistance to ionizing radiation [48,49], highlighting systems that maintain genetic integrity in pluripotent cells.

4.2 Phosphorylated proteome of hESCs

Given that a significant number of phosphorylated peptides and phosphorylation sites detected, it was not surprising that several networks consisting entirely of phosphorylated proteins were assembled. β -Arrestin is an adaptor protein involved in the desensitization of G-protein coupled receptors, receptor internalization, and recycling/degradation, as well as multiple other signaling functions in development and cellular movement [50]. Phosphorylation of β -arrestin at serine 412 is known to limit activity to basal levels [51]. β -Arrestin (phosphorylated S412) was a focal point in a network of more than 30 phosphorylated proteins spanning the plasma membrane, cytoplasm, and nucleus (Supporting Information Fig. 2). Unmodified serine 412 was not detected in our experiment, indicating that β -arrestin may be held in the basal state in undifferentiated cells. An alternate network of β -arrestin and nonphosphorylated proteins highlighted interactions with tubulins and staufer RNA localization factors (Supporting Information Fig. 3). A network of phosphorylated proteins involved in chromatin epigenetics was also generated (Supporting Information Fig. 4), including SWI/SNF chromatin regulators, members of the histone deacetylase complex, topoisomerases, DNA-interacting proteins, DNA-dependent protein kinase, splicing factors, and a cohesin complex protein. These networks contain numerous phosphorylation sites and point to complex regulatory pathways that can be uncovered by deep proteomic mining.

The sequences of phosphorylated peptides were interrogated for consensus kinase recognition via Scansite. Eleven different motifs, representing four motif families, were detected at 55 sites in 45 proteins using high-stringency filtering. Of those, expression of the kinases CK2, GSK3 β , and a splice variant of CDK5 was detected in our dataset, and others have been implicated as important for hESCs self-renewal or differentiation. The significance of protein kinase B (Akt) in hESCs self-renewal has been demonstrated in several ways. Akt signals downstream of PI3 kinase, and direct inhibition of either of these proteins causes differentiation of hESCs [52, 53], as does inhibition of FKBP12-rapamycin-associated protein/mechanistic target of rapamycin (kinase), which signals downstream of Akt [17]. Furthermore, we have shown that growth factors and receptors that activate this pathway, namely insulin growth factor/insulin and heregulin, support long-term self-renewal of hESCs in defined media [17]. Similarly, a role for ERK1/2 signaling has been suggested by MEK1/2-dependant activation, which occurs following FGF2 stimulation [54] and increases in cell death and differentiation that are observed when MEK1/2 is inhibited [52]. Inhibition of Src-family kinases results in elevated differentiation of hESCs [55], and roles in self-renewal and differentiation of mouse ES cells have also been demonstrated [56]. Previously, activation of WNT signaling via inhibition of GSK3 β was suggested to promote self-renewal in short-term assays [57]. However, subsequent reports have indicated that

WNT/ β -catenin signaling is not sufficient to maintain hESCs [58] and prolonged inhibition of GSK3 β leads to mesendodermal differentiation [59]. Our analysis of protein phosphorylation within hESCs provides candidates that may be regulated by these kinases and a framework for continued dissection of the signaling pathways that control self-renewal and differentiation.

CK2 is a ubiquitous serine/threonine kinase involved with numerous cellular processes such as DNA replication, protein translation, the cell cycle, and proliferation [60, 61]. Mutation of the CK2 regulatory beta subunit leads to early embryonic lethality in the mouse, failure to expand ES cells from explanted blastocysts, or generation of null ES cells in vitro [62]. In contrast, inhibition of CK2 with two different small-molecule antagonists had no overt effects on hESCs in our short-term assays. Fundamental differences between the signaling strategies of mouse and human pluripotent cells are suspected, for example, hESCs are not dependent on the LIF/gp130 pathway [63]. The lack of obvious effects of CK2 inhibition on hESCs suggests a further point of species-specific signaling. Protein kinase C (PKC) is a family of multifunctional serine/threonine kinases encompassing more than ten different isoenzymes, with different cofactor and activation requirements [64]. The detection of a PKC epsilon kinase recognition site with high-stringency filtering, and additional PKC sites with medium-stringency filtering, suggests functional activity for these kinases in hESCs or in differentiation. Increased proliferation of mouse ESCs in response to ATP has been suggested to involve PKC signaling [65], and S719 phosphorylation of PKC epsilon has been correlated with the proliferation rate of differentiating cells [66]. It has been further demonstrated that PKC signaling is critical to balancing self-renewal and lineage commitment in mouse ESCs [67]. Recent study in human ESCs confirmed that the activation of PKC pathway, PKC- δ in particular, hESCs undergo an epithelial mesenchymal transition and differentiate into the lineage of extraembryonic endoderm [68]. Treatment of proliferating hESCs with a range of small molecules that impact PKC signaling either as inhibitors, modulators, or activators, elicited complex response profiles. Therefore, it is possible that PKC family members have multiple important roles in hESCs, and knockdown of individual components in combination with agonist/antagonist probing will be required to unravel these potentially complex pathways.

4.3 O-GlcNAc-modified proteome of hESCs

In addition to phosphorylation, we also took interest in glycosylation in hESCs, specifically the O-GlcNAc modification. Glycosylation through serine and threonine by a single O-GlcNAc moiety is commonly seen among cytosolic and nuclear proteins in metazoans. O-GlcNAc modification is a nutrient/stress-sensitive modification that regulates proteins involved in a wide array of biological processes, including transcription, signal transduction, and metabolism [69, 70]. Particularly, it has been shown that a functional O-GlcNAc transferase, the enzyme that adds O-GlcNAc to serine or threonine residues, is essential for ESC viability and mouse ontogeny [71]. In our dataset, we identified 68 O-GlcNAc modification sites on 55 proteins expanding a diverse range of functional classes including enzymes (UDP-glucose 6-dehydrogenase, fatty acid synthase), kinases (obscurin, TAO kinase), actin-based cytoskeletal proteins (talin), phosphatases (PPP1R12A), transcription regulators (general transcription factor II-I, elongation factor 2), transporters (Nup155,

Nup160), and other undefined classes (histone, proliferating cell nuclear antigen). Studies have shown that there is a complex and dynamic interplay between O-GlcNAc and O-phosphate modification [72–74], two important protein modifications that share several common traits. We also discovered that 16 proteins in our dataset are modified by both O-phosphate and O-GlcNAc (Table 3), including lamin A/C, and nuclear pore complex proteins such as Nup358 and Nup160, etc. Lamins are structural protein components of the nuclear lamina, a protein network underlying the inner nuclear membrane that determines nuclear shape and size. It has been proven that lamina modulates the extent of deformation of nucleus and therefore makes nuclei in undifferentiated stem cells physically plastic and more pliable than nuclei in differentiated cells [75]. Furthermore, it has been shown that the expression of lamin A/C is activated during human ES cell differentiation before the downregulation of the pluripotency marker Oct4 but not before the downregulation of the pluripotency markers Tra-1–60, Tra-1–81, and SSEA-4. Those findings identified the absence of lamin A/C expression as a novel marker for undifferentiated ES cells and further support a role for nuclear lamins in cell maintenance and differentiation [76]. Besides nuclear lamins, nuclear pore complexes, which are responsible for the protected exchange of components between the nucleus and cytoplasm and for preventing the transport of material not destined to cross the nuclear envelope, have also been demonstrated to be involved in stem cell cardiovascular differentiation. Nuclear pore complex proteins exhibit a significantly decreased density and depth in stem cell derived cardiomyocytes, indicating that stem cell derived cardiomyocytes undergo structural adaptation and mobilize nuclear transport regulators in support of nucleocytoplasmic communication during commitment to mature cardiac lineage [77].

4.4 Comparison between MudPIT and single-dimension LC-MS/MS experiment

In eukaryotes, protein PTM, such as phosphorylation and glycosylation, often plays important roles in regulation of biological processes. Detecting and identifying posttranslationally modified proteins facilitates the elucidation of the mechanisms in cellular signal transduction and other crucial biological activities. Despite the recent advancement of MS-based proteomics, identifying posttranslationally modified proteins, especially characterizing the modification sites, still remains a challenge since most posttranslationally modified proteins are naturally expressed at low abundance. In order to increase the coverage of posttranslationally modified proteins in the context of a complex cell proteome, certain enrichment and fractionation techniques, such as chromatography, become necessary prior to protein/peptide sequencing by mass spectrometer.

In our study, we compared the data of identified cellular proteome, particularly the posttranslationally modified proteome, obtained, respectively, from protein subfractionation followed by 1D LC-MS/MS and the MudPIT experiments described above. By comparing the number of identified proteins in two respective analyses, we observed an overall 2.26-fold increase (2414/1067) in identified proteins with a 3.7-fold increase (21523/5821) in total peptides (Fig. 3A, Supporting Information Table 9) in the MudPIT experiments. Particularly, both phosphate and O-GlcNAc-modified proteomes showed more than an order of magnitude (at least a 12-fold increase) in identified proteins and corresponding peptides in the MudPIT compared to 1D LC-MS/MS experiments (Fig. 3B, Supporting Information

Table 9). For instance, the two proteins that are identified as being modified by both phosphate and O-GlcNAc, fatty acid synthase (FASN) and nucleolin (NCL), both showed approximately twofold increase in the number of unique peptides assigned to them (Supporting Information Table 10). Furthermore, 94.4% (1007/1067) of identified proteins in the single LC-MS/MS experiment were also present in the MudPIT analyses, and 65.7% (701/1067) of them were observed in the 1000 most abundant proteins in the MudPIT experiment (ranked by spectral count). These data indicate that multidimensional separation of samples prior to MS/MS sequencing not only increases the number of identified proteins but also the peptide coverage of identified proteins leading to more confident protein identification. The less than threefold increase in overall proteome coverage may not completely justify the labor-intensive and time-consuming experiments involved in MudPIT (75 times more instrument time in this comparison), however, the drastic increase of over an order of magnitude in the number of protein PTMs discovered by MudPIT demonstrates that this technique improves the depth of proteomic analysis of complex mixture significantly by enriching low-abundance, posttranslationally modified proteins.

4.5 Secreted proteome of hESCs

Previously, the ECM proteins that mediate hES cell adherence and matrix-integrin signaling have been inferred by the combination of integrins expressed by hESCs and plating experiments with purified proteins. Laminin and fibronectin have therefore been used to support serial attachment and expansion of hESCs [78,79]. Our analysis of the secreted proteome identified a diverse array of human ECM proteins: tenascin X, fibulin, laminin alpha 2 and alpha 5, and collagen alpha I, IV, and XVIII. It is likely that hESCs deposit their own ECM after attachment and combinations of these proteins may provide more effective substrates for adhesion.

Transcription factor LIN28 was identified in our ECM dataset, which modulates cell growth and associates with a subset of cell cycle regulators in ESCs and therefore can be used as markers for cell pluripotency [80, 81]. We also observed several growth factors that are important in signal transduction in ESCs, such as Jagged-1, GDF3, and LEFTY. Jagged-1 is a Notch ligand that functions as WNT-dependent Notch signaling activator in ESCs and is the key molecule maintaining the homeostasis of the cells [82]. In previous studies, three secreted factors of the TGF β family, LEFTY1, LEFTY2, and GDF3, are expressed at high levels during pluripotency and decline dramatically when the cells are allowed to differentiate, suggesting a role during this process [83]. It was further proven that GDF3 regulates both of the two major characteristics of ESCs in a species-specific manner: the ability to maintain the undifferentiated state and the ability to differentiate into the full spectrum of cell types [84]. LEFTY was shown to be regulated by both pluripotency-supporting pathways (Smad, WNT) and transcription factor Oct4, and also induced upon the exit of pluripotent state [85–88]. Lastly, the observation of hESCs forming polarized epithelia [28, 89] was confirmed with the detection of a substantial number of tight-junction signaling complex proteins in our dataset.

4.6 Concluding remarks

These experiments serve to investigate the proteome of hESCs in order to discover proteomic candidates that are responsible for the maintenance of the undifferentiated cell state and the initial loss of pluripotency. By performing a multidimensional chromatography separation coupled with MS/MS-based protein identification experiment, we were able to assign the cell proteome with a certain depth and specific PTMs within the cell that are relevant to various cellular processes. To reveal the molecular mechanism and signal transduction pathways involved in hESC differentiation, proteomic analysis will need to be performed on both hESCs (undifferentiated) and hESC-derived cell types (differentiated) in a quantitative manner.

Acknowledgments

We would like to thank all the members of the Wells' laboratory as well as James Atwood (NuSep) and Ron Orlando (UGA) for helpful discussions. This work was supported in part by NIGMS/NIH (P41GM103490, L.W. senior investigator).

Abbreviations

ECM	extracellular matrix
ESCs	embryonic stem cells
hESCs	human embryonic stem cells
MudPIT	multidimensional protein identification technology
O-GlcNAc	O-linked β -N-acetylglucosamine
SCX	strong cation exchange

References

1. Metallo CM, Azarin SM, Ji L, de Pablo JJ, Palecek SP. Engineering tissue from human embryonic stem cells. *J Cell Mol Med*. 2008; 12:709–729. [PubMed: 18194458]
2. Abuljadayel IS. Harnessing pluripotency from differentiated cells: a regenerative source for tissue-specific stem cell therapies. *Curr Stem Cell Res Ther*. 2006; 1:325–331. [PubMed: 18220877]
3. Srivastava D, Ivey KN. Potential of stem-cell-based therapies for heart disease. *Nature*. 2006; 441:1097–1099. [PubMed: 16810246]
4. Schulz TC, Young HY, Agulnick AD, Babin MJ, et al. A scalable system for production of functional pancreatic progenitors from human embryonic stem cells. *PLoS One*. 2012; 7:e37004. [PubMed: 22623968]
5. Jadalila M, Lee HJ, Pakzad M, Fathi A, et al. Quantitative proteomic analysis of human embryonic stem cell differentiation by 8-plex iTRAQ labelling. *PLoS One*. 2012; 7:e38532. [PubMed: 22723866]
6. Phanstiel DH, Brumbaugh J, Wenger CD, Tian S, et al. Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nat Methods*. 2011; 8:821–827. [PubMed: 21983960]
7. Chaerkady R, Letzen B, Renuse S, Sahasrabudhe NA, et al. Quantitative temporal proteomic analysis of human embryonic stem cell differentiation into oligodendrocyte progenitor cells. *Proteomics*. 2011; 11:4007–4020. [PubMed: 21770034]
8. Rigbolt KT, Prokhorova TA, Akimov V, Henningsen J, et al. System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. *Sci Signal*. 2011; 4:rs3. [PubMed: 21406692]

9. Tian R, Wang S, Elisma F, Li L, et al. Rare cell proteomic reactor applied to stable isotope labeling by amino acids in cell culture (SILAC)-based quantitative proteomics study of human embryonic stem cell differentiation. *Mol Cell Proteomics*. 2011; 10:M110.000679.
10. Bendall SC, Hughes C, Campbell JL, Stewart MH, et al. An enhanced mass spectrometry approach reveals human embryonic stem cell growth factors in culture. *Mol Cell Proteomics*. 2009; 8:421–432. [PubMed: 18936058]
11. Prowse AB, McQuade LR, Bryant KJ, Marcal H, Gray PP. Identification of potential pluripotency determinants for human embryonic stem cells following proteomic analysis of human and mouse fibroblast conditioned media. *J Proteome Res*. 2007; 6:3796–3807. [PubMed: 17655345]
12. Swaney DL, Wenger CD, Thomson JA, Coon JJ. Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci USA*. 2009; 106:995–1000. [PubMed: 19144917]
13. Link AJ, Eng J, Schieltz DM, Carmack E, et al. Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol*. 1999; 17:676–682. [PubMed: 10404161]
14. Moore AW Jr, Larmann JP Jr, Lemmo AV, Jorgenson JW. Two-dimensional liquid chromatography-capillary electrophoresis techniques for analysis of proteins and peptides. *Methods Enzymol*. 1996; 270:401–419. [PubMed: 8803979]
15. Opiteck GJ, Jorgenson JW. Two-dimensional SEC/RPLC coupled to mass spectrometry for the analysis of peptides. *Anal Chem*. 1997; 69:2283–2291. [PubMed: 9212702]
16. Washburn MP, Wolters D, Yates JR. 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol*. 2001; 19:242–247. [PubMed: 11231557]
17. Wang L, Schulz TC, Sherrer ES, Dauphin DS, et al. Self-renewal of human embryonic stem cells requires insulin-like growth factor-1 receptor and ERBB2 receptor signaling. *Blood*. 2007; 110:4111–4119. [PubMed: 17761519]
18. Robins, AJ.; Schulz, TC. *Emerging Technology Platforms for Stem Cells*. John Wiley & Sons; Hoboken, NJ: 2008.
19. Porterfield M, Zhao P, Han H, Cunningham J, et al. Discrimination between adenocarcinoma and normal pancreatic ductal fluid by proteomic and glycomic analysis. *J Proteome Res*. 2014; 13:395–407. [PubMed: 24328148]
20. Liu H, Sadygov RG, Yates JR. 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem*. 2004; 76:4193–4201. [PubMed: 15253663]
21. Zhao P, Viner R, Teo CF, Boons GJ, et al. Combining high-energy C-trap dissociation and electron transfer dissociation for protein O-GlcNAc modification site assignment. *J Proteome Res*. 2011; 10:4088–4104. [PubMed: 21740066]
22. Weatherly DB, Atwood JA 3rd, Minning TA, Cavola C, et al. A heuristic method for assigning a false-discovery rate for protein identifications from Mascot database search results. *Mol Cell Proteomics*. 2005; 4:762–772. [PubMed: 15703444]
23. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem*. 2002; 74:5383–5392. [PubMed: 12403597]
24. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem*. 2003; 75:4646–4658. [PubMed: 14632076]
25. Palumbo AM, Reid GE. Evaluation of gas-phase rearrangement and competing fragmentation reactions on protein phosphorylation site assignment using collision induced dissociation-MS/MS and MS(3). *Anal Chem*. 2008; 80:9735–9747. [PubMed: 19012417]
26. Aguiar M, Haas W, Beausoleil SA, Rush J, Gygi SP. Gas-phase rearrangements do not affect site localization reliability in phosphoproteomics data sets. *J Proteome Res*. 2010; 9:3103–3107. [PubMed: 20377248]
27. Baharvand H, Hajheidari M, Ashtiani SK, Salekdeh GH. Proteomic signature of human embryonic stem cells. *Proteomics*. 2006; 6:3544–3549. [PubMed: 16758447]
28. Schulz TC, Swistowska AM, Liu Y, Swistowski A, et al. A large-scale proteomic analysis of human embryonic stem cells. *BMC Genomics*. 2007; 8:478. [PubMed: 18162134]

29. Van Hoof D, Passier R, Ward-Van Oostwaard D, Pinkse MW, et al. A quest for human and mouse embryonic stem cell-specific proteins. *Mol Cell Proteomics*. 2006; 5:1261–1273. [PubMed: 16600995]
30. Zhou F, Cardoza JD, Ficarro SB, Adelmant GO, et al. Online nanoflow RP-RP-MS reveals dynamics of multicomponent Ku complex in response to DNA damage. *J Proteome Res*. 2010; 9:6242–6255. [PubMed: 20873769]
31. Rogatsky E, Stein DT. Two-dimensional reverse phase-reverse phase chromatography: a simple and robust platform for sensitive quantitative analysis of peptides by LC/MS. *Hardware design. J Sep Sci*. 2006; 29:538–546. [PubMed: 16583692]
32. van den Boom V, Kooistra SM, Boesjes M, Geverts B, et al. UTF1 is a chromatin-associated protein involved in ES cell differentiation. *J Cell Biol*. 2007; 178:913–924. [PubMed: 17785516]
33. Liu Y, Labosky PA. Regulation of embryonic stem cell self-renewal and pluripotency by Foxd3. *Stem Cells*. 2008; 26:2475–2484. [PubMed: 18653770]
34. Boyer LA, Lee TI, Cole MF, Johnstone SE, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*. 2005; 122:947–956. [PubMed: 16153702]
35. Babaie Y, Herwig R, Greber B, Brink TC, et al. Analysis of Oct4-dependent transcriptional networks regulating self-renewal and pluripotency in human embryonic stem cells. *Stem Cells*. 2007; 25:500–510. [PubMed: 17068183]
36. Chambers I, Colby D, Robertson M, Nichols J, et al. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell*. 2003; 113:643–655. [PubMed: 12787505]
37. Nichols J, Zevnik B, Anastasiadis K, Niwa H, et al. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell*. 1998; 95:379–391. [PubMed: 9814708]
38. Niwa H, Miyazaki J, Smith AG. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet*. 2000; 24:372–376. [PubMed: 10742100]
39. Mitsui K, Tokuzawa Y, Itoh H, Segawa K, et al. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell*. 2003; 113:631–642. [PubMed: 12787504]
40. Rodda DJ, Chew JL, Lim LH, Loh YH, et al. Transcriptional regulation of nanog by OCT4 and SOX2. *J Biol Chem*. 2005; 280:24731–24737. [PubMed: 15860457]
41. Ambrosetti DC, Basilico C, Dailey L. Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. *Mol Cell Biol*. 1997; 17:6321–6329. [PubMed: 9343393]
42. Masui S, Nakatake Y, Toyooka Y, Shimosato D, et al. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol*. 2007; 9:625–635. [PubMed: 17515932]
43. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. 2006; 126:663–676. [PubMed: 16904174]
44. Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science*. 2007; 318:1917–1920. [PubMed: 18029452]
45. Draper JS, Smith K, Gokhale P, Moore HD, et al. Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. *Nat Biotechnol*. 2004; 22:53–54. [PubMed: 14661028]
46. Maitra A, Arking DE, Shivapurkar N, Ikeda M, et al. Genomic alterations in cultured human embryonic stem cells. *Nat Genet*. 2005; 37:1099–1103. [PubMed: 16142235]
47. Buzzard JJ, Gough NM, Crook JM, Colman A. Karyotype of human ES cells during extended culture. *Nat Biotechnol*. 2004; 22:381–382. author reply 382. [PubMed: 15060545]
48. Maynard S, Swistikowa AM, Lee JW, Liu Y, et al. Human embryonic stem cells have enhanced repair of multiple forms of DNA damage. *Stem Cells*. 2008; 26:2266–2274. [PubMed: 18566332]
49. Banuelos CA, Banath JP, Macphail SH, Zhao J, et al. Mouse but not human embryonic stem cells are deficient in rejoining of ionizing radiation-induced DNA double-strand breaks. *DNA Repair (Amst)*. 2008; 7:1471–1483. [PubMed: 18602349]

50. Buchanan FG, DuBois RN. Emerging roles of beta-arrestins. *Cell Cycle*. 2006; 5:2060–2063. [PubMed: 16969081]
51. Luttrell LM, Ferguson SS, Daaka Y, Miller WE, et al. Beta-arrestin-dependent formation of beta2 adrenergic receptor-Src protein kinase complexes. *Science*. 1999; 283:655–661. [PubMed: 9924018]
52. Armstrong L, Hughes O, Yung S, Hyslop L, et al. The role of PI3K/AKT, MAPK/ERK and NFkappabeta signalling in the maintenance of human embryonic stem cell pluripotency and viability highlighted by transcriptional profiling and functional analysis. *Hum Mol Genet*. 2006; 15:1894–1913. [PubMed: 16644866]
53. McLean AB, D'Amour KA, Jones KL, Krishnamoorthy M, et al. Activin a efficiently specifies definitive endoderm from human embryonic stem cells only when phosphatidylinositol 3-kinase signaling is suppressed. *Stem Cells*. 2007; 25:29–38. [PubMed: 17204604]
54. Kang HB, Kim JS, Kwon HJ, Nam KH, et al. Basic fibroblast growth factor activates ERK and induces c-fos in human embryonic stem cell line MizhES1. *Stem Cells Dev*. 2005; 14:395–401. [PubMed: 16137228]
55. Anneren C, Cowan CA, Melton DA. The Src family of tyrosine kinases is important for embryonic stem cell self-renewal. *J Biol Chem*. 2004; 279:31590–31598. [PubMed: 15148312]
56. Meyn MA 3rd, Schreiner SJ, Dumitrescu TP, Nau GJ, Smithgall TE. SRC family kinase activity is required for murine embryonic stem cell growth and differentiation. *Mol Pharmacol*. 2005; 68:1320–1330. [PubMed: 15985613]
57. Sato N, Meijer L, Skaltsounis L, Greengard P, Brivanlou AH. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nat Med*. 2004; 10:55–63. [PubMed: 14702635]
58. Dravid G, Ye Z, Hammond H, Chen G, et al. Defining the role of Wnt/beta-catenin signaling in the survival, proliferation, and self-renewal of human embryonic stem cells. *Stem Cells*. 2005; 23:1489–1501. [PubMed: 16002782]
59. Bakre MM, Hoi A, Mong JC, Koh YY, et al. Generation of multipotential mesendodermal progenitors from mouse embryonic stem cells via sustained Wnt pathway activation. *J Biol Chem*. 2007; 282:31703–31712. [PubMed: 17711862]
60. Pinna LA, Meggio F. Protein kinase CK2 (“casein kinase-2”) and its implication in cell division and proliferation. *Prog Cell Cycle Res*. 1997; 3:77–97. [PubMed: 9552408]
61. Guerra B, Issinger OG. Protein kinase CK2 and its role in cellular proliferation, development and pathology. *Electrophoresis*. 1999; 20:391–408. [PubMed: 10197447]
62. Buchou T, Vernet M, Blond O, Jensen HH, et al. Disruption of the regulatory beta subunit of protein kinase CK2 in mice leads to a cell-autonomous defect and early embryonic lethality. *Mol Cell Biol*. 2003; 23:908–915. [PubMed: 12529396]
63. Okita K, Yamanaka S. Intracellular signaling pathways regulating pluripotency of embryonic stem cells. *Curr Stem Cell Res Ther*. 2006; 1:103–111. [PubMed: 18220859]
64. Breitkreutz D, Braiman-Wiksmann L, Daum N, Denning MF, Tennenbaum T. Protein kinase C family: on the crossroads of cell signaling in skin and tumor epithelium. *J Cancer Res Clin Oncol*. 2007; 133:793–808. [PubMed: 17661083]
65. Heo JS, Han HJ. ATP stimulates mouse embryonic stem cell proliferation via protein kinase C, phosphatidylinositol 3-kinase/Akt, and mitogen-activated protein kinase signaling pathways. *Stem Cells*. 2006; 24:2637–2648. [PubMed: 16916926]
66. Prudhomme W, Daley GQ, Zandstra P, Lauffenburger DA. Multivariate proteomic analysis of murine embryonic stem cell self-renewal versus differentiation signaling. *Proc Natl Acad Sci USA*. 2004; 101:2900–2905. [PubMed: 14978270]
67. Wang S, Huang X, Sun D, Xin X, et al. Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates Akt signaling. *PLoS One*. 2012; 7:e37427. [PubMed: 22629392]
68. Li T, Li Z, Li J, Wang J, et al. Elevation of cellular O-GlcNAcylation level by a potent and selective O-GlcNAcase inhibitor based on tetrahydroimidazopyridine scaffold. *Bioorg Med Chem Lett*. 2012; 22:6854–6857. [PubMed: 23058883]
69. Love DC, Hanover JA. The hexosamine signaling pathway: deciphering the “O-GlcNAc code”. *Sci STKE*. 2005; 2005:re13. [PubMed: 16317114]

70. Zachara NE, Hart GW. O-GlcNAc a sensor of cellular state: the role of nucleocytoplasmic glycosylation in modulating cellular function in response to nutrition and stress. *Biochim Biophys Acta*. 2004; 1673:13–28. [PubMed: 15238246]
71. Shafi R, Iyer SP, Ellies LG, O'Donnell N, et al. The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. *Proc Natl Acad Sci USA*. 2000; 97:5735–5739. [PubMed: 10801981]
72. Slawson C, Hart GW. Dynamic interplay between O-GlcNAc and O-phosphate: the sweet side of protein regulation. *Curr Opin Struct Biol*. 2003; 13:631–636. [PubMed: 14568619]
73. Wang Z, Udeshi ND, Slawson C, Compton PD, et al. Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates cytokinesis. *Sci Signal*. 2010; 3:ra2. [PubMed: 20068230]
74. Feng X, Zhang J, Smuga-Otto K, Tian S, et al. Protein kinase C mediated extraembryonic endoderm differentiation of human embryonic stem cells. *Stem Cells*. 2012; 30:461–470. [PubMed: 22213079]
75. Pajeroski JD, Dahl KN, Zhong FL, Sammak PJ, Discher DE. Physical plasticity of the nucleus in stem cell differentiation. *Proc Natl Acad Sci USA*. 2007; 104:15619–15624. [PubMed: 17893336]
76. Constantinescu D, Gray HL, Sammak PJ, Schatten GP, Csoka AB. Lamin A/C expression is a marker of mouse and human embryonic stem cell differentiation. *Stem Cells*. 2006; 24:177–185. [PubMed: 16179429]
77. Perez-Terzic C, Behfar A, Mery A, van Deursen JM, et al. Structural adaptation of the nuclear pore complex in stem cell-derived cardiomyocytes. *Circ Res*. 2003; 92:444–452. [PubMed: 12600892]
78. Xu C, Inokuma MS, Denham J, Golds K, et al. Feeder-free growth of undifferentiated human embryonic stem cells. *Nat Biotechnol*. 2001; 19:971–974. [PubMed: 11581665]
79. Brimble SN, Zeng X, Weiler DA, Luo Y, et al. Karyotypic stability, genotyping, differentiation, feeder-free maintenance, and gene expression sampling in three human embryonic stem cell lines derived prior to August 9, 2001. *Stem Cells Dev*. 2004; 13:585–597. [PubMed: 15684826]
80. Richards M, Tan SP, Tan JH, Chan WK, Bongso A. The transcriptome profile of human embryonic stem cells as defined by SAGE. *Stem Cells*. 2004; 22:51–64. [PubMed: 14688391]
81. Assou S, Cerecedo D, Tondeur S, Pantesco V, et al. A gene expression signature shared by human mature oocytes and embryonic stem cells. *BMC Genomics*. 2009; 10:10. [PubMed: 19128516]
82. Katoh M. Notch ligand, JAG1, is evolutionarily conserved target of canonical WNT signaling pathway in progenitor cells. *Int J Mol Med*. 2006; 17:681–685. [PubMed: 16525728]
83. Sato N, Sanjuan IM, Heke M, Uchida M, et al. Molecular signature of human embryonic stem cells and its comparison with the mouse. *Dev Biol*. 2003; 260:404–413. [PubMed: 12921741]
84. Levine AJ, Brivanlou AH. GDF3, a BMP inhibitor, regulates cell fate in stem cells and early embryos. *Development*. 2006; 133:209–216. [PubMed: 16339188]
85. Katoh M. WNT/PCP signaling pathway and human cancer (review). *Oncol Rep*. 2005; 14:1583–1588. [PubMed: 16273260]
86. Meijer L, Skaltsounis AL, Magiatis P, Polychronopoulos P, et al. GSK-3-selective inhibitors derived from Tyrian purple indirubins. *Chem Biol*. 2003; 10:1255–1266. [PubMed: 14700633]
87. Cserjesi P, Brown D, Lyons GE, Olson EN. Expression of the novel basic helix-loop-helix gene eHAND in neural crest derivatives and extraembryonic membranes during mouse development. *Dev Biol*. 1995; 170:664–678. [PubMed: 7649392]
88. Beck F, Erler T, Russell A, James R. Expression of Cdx-2 in the mouse embryo and placenta: possible role in patterning of the extra-embryonic membranes. *Dev Dyn*. 1995; 204:219–227. [PubMed: 8573715]
89. Krtolica A, Genbacev O, Escobedo C, Zdravkovic T, et al. Disruption of apical-basal polarity of human embryonic stem cells enhances hematoendothelial differentiation. *Stem Cells*. 2007; 25:2215–2223. [PubMed: 17569786]

Figure 1.

Experiment workflow. Proteins extracted from hESCs were analyzed with strong cation exchange chromatography (SCX-LC), RP-LC, and nanoLC-MS/MS to establish a three-dimension separation prior to MS/MS peptide sequencing.

Figure 2.

Statistics across LC-MS/MS technical triplicate experiments. (A) A total of 1721 proteins were commonly detected in technical triplicates; respectively, 195, 265, 147 proteins were detected in both and only replicates 1 and 2, replicates 1 and 3, and replicates 2 and 3; respectively, 420, 180, 256 proteins were detected uniquely in replicate 1, replicate 2, and replicate 3. (B) Among the 1721 common proteins, 464 of them were detected in all four subcellular fractions, 506 of them were in three of the four fractions, 474 of them were in two of the four fractions, and 277 were uniquely detected in one of the four fractions; specifically, 60 proteins were detected uniquely in the cytosol fraction, 68 proteins were unique to the membrane/organelle fraction, 141 proteins were unique to the nucleus fraction, and eight proteins were unique to the cytoskeleton fraction. (C) Subcellular localization of the 1721 proteins commonly observed in all technical triplicates.

Figure 3.

Comparison between single LC-MS/MS and MudPIT experiments. (A) Comparison of the total number of identified proteins and peptides; (B) comparison of the total number of the modified proteins and peptides. An overall 2.26-fold increase (2414/1067) in total number of the identified proteins was observed with a 3.7-fold increase (21523/5821) in the number of peptides in MudPIT compared to the single LC-MS/MS experiment. Both phosphate and O-GlcNAc-modified proteomes showed an increase for over an order of magnitude in the number of identified proteins and corresponding peptides in the MudPIT experiments.

Table 1

Protein and PTM assignment in hESCs

(A) Assignment of hES cell and secreted proteome						
	Membrane/organelle	Nucleus	Cytosol	Cytoskeleton	Combined cell proteome	Secreted proteome
Protein	1843	2037	1676	848	3184	102
Peptide	18 268	20 902	17 960	7439	24 824	445
Spectra	68 190	79 573	84 732	29 101	261 626	4032

(B) Assignment of PTMs in hESCs		
	Protein	Peptide
Phosphate	287	431
O-GlcNAc	55	63

Table 2

Implicated kinases predicted by Scansite

Motif found	Motif family	Number of sites found
Akt_Kin	Baso_ST_kin	9
Casn_Kin2	Acid_ST_kin	22
Cdc2_Kin	Pro_ST_kin	14
Cdk5_Kin	Pro_ST_kin	25
Clk2_Kin	Baso_ST_kin	2
Erk1_Kin	Pro_ST_kin	3
GSK3_Kin	Acid_ST_kin	10
PKA_Kin	Baso_ST_kin	3
PKC_epsilon	Baso_ST_kin	1

Table 3

Proteins modified by both phosphate and O-GlcNAc

Uniprot accession	Protein name	Modification site	
		Phosphate	O-GlcNAc
P02545	Lamin-A/C	S390 S392	T91
P13639	Elongation factor 2	T57	T657
P19338	Nucleolin	T59 S60 S67 T69 S563	T301 T438
P25205	DNA replication licensing factor MCM3	S672 T674 S711 T722	T154 T617
P49327	Fatty acid synthase	T2204	T315
P49792	E3 SUMO-protein ligase RanBP2	T2128 S2280	T2192
P50991	T-complex protein 1 subunit delta	Y24	T217
Q08AD3	Nuclear pore complex protein Nup160	T1080	T306
Q5SWY0	Translocated promoter region	S1108	S1318
Q8WZ42	Titin	S1465 Y2602 T2603 T2611 Y13346 T13690 Y17874 Y20867 S20868 Y23644 Y32869	S1571 T14674 S28157
Q96T37	Putative RNA-binding protein 15	S656 S670 S674	T401
Q9BQG0	Myb-binding protein 1A	S1267	S1308
Q9NR30	Nucleolar RNA helicase 2	S89 S121	S759
Q9UHD8	Septin-9	S85	T142