Stem Cell Rev and Rep (2012) 8:16–31
DOI 10.1007/s12015-011-9317-8

Phosphoproteomic Analysis: An Emerging Role in Deciphering Cellular Signaling in Human Embryonic Stem Cells and Their Differentiated Derivatives

Brian T. D. Tobe · Junjie Hou · Andrew M. Crain · Ilyas Singec · Evan Y. Snyder · Laurence M. Brill

Published online: 19 October 2011

Abstract Cellular signaling is largely controlled by protein phosphorylation. This post-translational modification (PTM) has been extensively analyzed when examining one or a few protein phosphorylation events that effect cell signaling. However, protein kinase-driven signaling networks, comprising total (phospho)proteomes, largely control cell fate. Therefore, large-scale analysis of differentially regulated protein phosphorylation is central to elucidating complex cellular events, including maintenance of pluripotency and differentiation of embryonic stem cells (ESCs). The current technology of choice for total phosphoproteome and combined total proteome plus total phosphoproteome (termed (phospho)proteome) [1] analyses is multidimensional liquid chromatography- (MDLC) tandem mass spectrometry (MS/MS). Advances in the use of MDLC for separation of peptides comprising total (phospho)proteomes, phosphopeptide enrichment, separation of enriched fractions, and quantitative peptide identification by MS/MS have been rapid in recent years, as have improvements in the sensitivity, speed, and accuracy of mass spectrometers. Increasingly deep coverage of (phospho)proteomes is allowing an improved understanding of changes in protein phosphorylation networks as cells respond to stimuli and progress from one undifferentiated or differentiated state to another. Although MDLC-MS/MS studies are powerful, understanding the interpretation of the data is important, and targeted experimental pursuit of biological predictions provided by total (phospho)proteome analyses is needed. (Phospho)proteomic analyses of pluripotent stem cells are in their infancy at this time. However, such studies have already begun to contribute to an improved and accelerated understanding of basic pluripotent stem cell signaling and fate control, especially at the systems-biology level.

Keywords Embryonic stem cells · Pluripotency · Differentiation · Proteomics · Phosphoproteomics · Mass spectrometry · Multidimensional liquid chromatography · Bioinformatics

B. T. D. Tobe · J. Hou · A. M. Crain · I. Singec · E. Y. Snyder · L. M. Brill
The Sanford-Burnham Medical Research Institute,
La Jolla, CA, USA

L. M. Brill (✉)
Proteomics,
Sanford Burnham Medical Research Institute,
10901 North Torrey Pines Rd.,
La Jolla, CA 92037, USA
e-mail: lbrill@sanfordburnham.org

E. Y. Snyder (✉)
Program for Stem Cell and Regenerative Biology,
Sanford Burnham Medical Research Institute,
10901 North Torrey Pines Rd.,
La Jolla, CA 92037, USA
e-mail: esnyder@sanfordburnham.org

Introduction

Pluripotent stem cells, including human embryonic stem cells (hESCs) and human induced-pluripotent stem cells (hiPSCs) have the potential to self-renew indefinitely and differentiate into >200 cell types in the body [2, 3]. Knowledge of the molecular mechanisms of self-renewal, pluripotency and differentiation has consistently expanded with the increasing depth of stem cell biology research. Briefly, self-renewal of human pluripotent stem cells relies on a relatively well-characterized network of transcription factors and epigenetic regulators [4–8]. Less well characterized, especially at the systems-biology/proteomic level,

are cellular signaling pathways (e.g. TGFβ/Activin/Nodal, WNT, PI3K, FGF, IGF, EGF, PDGF and JNK) with potential roles in controlling self-renewal, pluripotency and differentiation [1, 9–17]. Identification of components contributing to maintenance of self-renewing hESCs has provided the means to identify pluripotent stem cells by immunocytochemistry, including Western blotting and flow cytometry, to maintain pluripotent stem cells in culture, and to induce pluripotency via reprogramming of differentiated cell types with exogenous factors.

Phosphorylation is one of the most common and well-characterized PTMs. Human cells are thought to have about 480 protein kinases [18], a revision of the initial estimate of 518 [19]. The majority of them are serine/(S) threonine (T) kinases and about 90 are tyrosine (Y) kinases [18]. As with (perhaps all) other biological processes, dynamic regulation of reversible, site-specific protein phosphorylation is critical to the signaling networks that regulate self-renewal and differentiation [1, 10, 11, 13, 16]. Extra-cellular signals and intracellular regulatory events that activate pluripotency factors, inhibit differentiation pathways, promote growth and cell division, and inhibit cell death may contribute to the control of stem cell fate. Though much of this network was initially described in mouse models, it has become clear that there are differences in the regulation of pluripotency in mouse and human ESCs. In human ESCs (hESCs), TGFβ super-family members, including Activin, Nodal and BMP, modulate self-renewal through receptor-mediated phosphorylation of pathway-specific SMAD proteins. Nodal and Activin activates SMAD2/3 whereas BMP activates SMAD1/5/8. In turn, NANOG transcription is activated by SMAD2/3 and inhibited by SMAD1/5/8 [11, 20, 21]. Activation of the canonical WNT pathway likely regulates self-renewal through de-phosphorylation of β-catenin, allowing its nuclear localization and assembly with the TCF/LEF complex to enable transcriptional activation of target genes [22]. Conversely, the phosphatidylinositol-3-kinase (PI3K) pathway may inhibit differentiation of endoderm-derived cell lineages, but mechanisms by which other signaling pathways participate in self-renewal are relatively unclear [10, 20].

Reactivation of only a few transcription factor proteins, including OCT4 (POU5F1), SOX2, KLF4, MYC, NANOG and/or GLIS1 are sufficient, depending on the cell type, for reprogramming of differentiated human cells to induced pluripotent stem cells (iPSCs) [2, 23–25]. A growing body of evidence links these factors to regulatory signaling components important to self-renewal. KLF4 is a direct target of the TGFβ pathway [26], and SOX2 and MYC may also be targets of TGFβ signaling [27]. Similarly, MYC is a downstream transcriptional target of canonical WNT signaling [28]. Identification of downstream targets of these factors is in the early stages, and the environmental influences of extra-cellular ligands, cellular growth density, and oxygen concentration on this transcriptional network is also not characterized well [29–31]. Given the pivotal role of core transcription regulators, extensive efforts have been undertaken to describe the transcriptome of pluripotent cells. Analyses of mRNA microarray data suggest that protocols specific to individual laboratories in which the cells were cultured and analyzed are the most influential determinants of heterogeneous expression profiles [32]. Although some reports estimate that as few as ca. 50% of the mRNA transcripts quantitatively correlate with relative abundance of the encoded protein, 75% of protein-coding transcripts may be expressed in most human tissues, thus making it difficult to identify physiologically relevant genes [33]. These observations and challenges make it clear that proteins, the final products of the vast majority of the genes, require direct analysis.

In this review, we discuss current analytical platforms that have been applied in published (phospho)proteomic analyses of hESCs and their differentiated derivatives, and include closely allied technologies. Important experimental parameters, key findings of these pioneering studies, including biological implications and follow-up experiments, are also described. Efforts are not made to comprehensively review both phosphoproteomics and stem cell biology, but to help create a nexus between the two disparate fields. We strive to facilitate an improved understanding of the first instances of successful application of phosphoproteomic technologies to yield an improved understanding of hESC biology.

High Quality Cellular Material from Which the Proteins are Derived is Critical to the Success of (phospho) proteomic Studies

Although unclear at this time, due to the low number of published studies, multiple biological factors could influence (phospho)proteomic results from pluripotent stem cells. ESC lines have varying genotypes, passage number, and gender, which likely result in variation of growth and differentiation [34]. Furthermore, a given ESC line may have been passaged by mechanical or enzymatic techniques, cultured under feeder-free conditions, with mouse or human feeders, or exposed to varying conditions of oxygen tension, exogenous factors and media preparations [29, 35]. Perhaps as a result of some of these variables, gene expression profiles correlated strongly with the laboratory environment in which the cells were grown [32]. Of the four studies published to date reporting large-scale phosphoproteomic profiling of hESCs, two used WA-01 (H1) hESCs [10, 36], a the third examined HUES-7 [1] while the fourth used HUES-9 and Odense-3 [13]. Between WA-01 and HUES-7 [1, 10], there was 26% overlap in the phosphoproteins identified as being more prominent in

undifferentiated hESCs, whereas the overlap in the differentiated derivatives, under divergent differentiation conditions and times, was 6.9% [37]. Between a recent pair of (phospho)proteomic analyses of hESCs and their non-specifically differentiated derivatives, there was a relatively high overlap of ca. 76% of all identified proteins [1, 13]. Although current observations suggest good reproducibility of the hESC (phospho)proteome, more data is needed, including the use of more cell lines.

Demonstration of the condition/quality of the cells cultured for proteomics experiments has not been routine or standardized. Although optimization of stem cell culture is ongoing and the protocols vary widely [38], high quality cell cultures with the maximum possible homogeneity are indispensable for successful (phospho)proteomic results. Additionally, it is important to examine the morphology (e.g. nuclear-to-cytoplasmic ratio), markers of pluripotency and the cell surface, as well as differentiation potential (e.g. embryoid bodies, teratoma formation) as measures of the quality of the cell populations used for phosphoproteomic analyses [10]. Due to the large requirement for resources and time, total (phospho)proteomic analysis generally precludes extensive replication of results. Thus, careful choices must be made for experimental conditions to be used. Additional characterization of the cells and protein preparations from them is advisable, including karyotype analysis, immunostaining, flow cytometry, and directed biochemical assays (e.g. Western blots), for the presence, at expected quantitative levels, including phosphorylation of known phosphoproteins. Maximizing homogeneity of the cell population is important, because small populations of differentiated cells could contribute "noise" to (phospho) proteomic profiles. When comparative analyses between pluripotent and differentiated cell populations are performed, the highest possible differentiation specificity should help clarify (phospho)proteomic changes during transitions from pluripotency to specific lineages, and should allow more rapid discovery of (phospho)proteins helping to uncover mechanisms of differentiation.

Delineating dynamic molecular profiles that occur during cell fate choice is a major impetus for application of (phospho)proteomics, to examine effects of differentiation on the cellular (phospho)proteome [1, 10, 13]. Recent studies have utilized retinoic acid [10], modulation of BMP signaling [1], and use of un-conditioned medium or phorbol 12-myristate 13-acetate [13], which resulted in non-specific differentiation and hence, heterogeneous cell types.

(Phospho)proteomic Analysis Facilitates an Improved Understanding of Cell Signaling

Proteomics provides a snapshot of the detectable proteins present in cells at a given time point. Despite, or perhaps because of the increased complexity resulting from regulation at the transcriptional, translational, post-translational, and protein stability levels, proteomic signatures may be relatively stable in spite of changes in DNA copy number and cellular aging [39, 40]. Understanding pluripotency and germ line-specific differentiation has benefited from identification of proteins and other markers specific to particular lineages. However, many such markers identified show multiple cell type associations. Tissue specific proteins could be low abundance or localized to the cell surface, rendering them difficult to detect [41].

There is increasing evidence that proteins and their activity may be regulated extensively by changes in protein abundance and especially phosphorylation [1, 13, 42], and that extensive regulation of protein phosphorylation occurs during hESC differentiation [1, 10, 13]. Current data is consistent with an initial model in which phosphorylation is an important regulator of cell state, through interaction with protein activity and stability, as well as influencing transcription and translation, similar to regulation of these processes by protein phosphorylation in many biological systems (Fig. 1). In addition, phosphorylation can inhibit or facilitate additional PTMs including SUMOylation, acetylation, methylation and ubiquitination in adjacent portions of the protein [43–45].

Analogous to identification of interacting transcription factors by interrogating a gene expression dataset, (phospho)proteomic analysis can identify the regulation of signaling networks [1, 10, 13, 36], via the phosphorylation state of the kinases and kinase/phosphatase targets within a given network. One challenge is the difficulty of detecting low-abundance proteins, which frequently includes kinases [46]. However, the size of datasets from identified (phospho)proteomes has increased rapidly, likely representing improving degrees of comprehensiveness of the analyses (Fig. 2). Larger (phospho)proteomic datasets, although more difficult to analyze, should further improve the value of the analyses for systems biology and

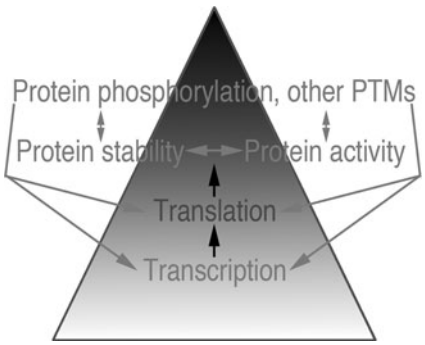

Fig. 1 Preliminary simplified model of the broad categories of regulatory molecules controlling pluripotency and differentiation of stem cells, which is consistent with general cell biology

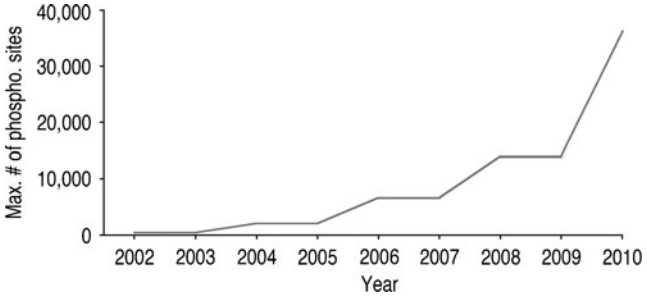

Fig. 2 Plot of the phosphoproteomic datasets with the largest number of identified phosphorylation sites as a function of year. The references from which the numbers were derived: [52, 67, 73, 137, 42]

to support studies focused on smaller numbers of (phospho)proteins. In several different biological systems, the percentage of identified proteins that were phosphorylated has been found to be high, with values of 52% in mouse tissues and organs [42], 66.5% in hESCs and their non-specifically differentiated derivatives [13] and 70% in HeLa cells [47].

Multiple kinase pathways may function to maintain pluripotency and differentiation. However, the regulation of protein-phosphorylation-based signaling networks is only beginning to be cataloged in hESCs and their derivatives. Traditional antibody-based assays of cell signaling (e.g. Western blotting, immunofluorescence), although critical for hypothesis-based investigation of distinct components in pathway function, are likely to be insufficient for characterization of global cellular signaling networks. Even targeted kinase or phosphatase deletion may modulate many components of interactive signaling networks, most of which would not be identified without unbiased, large-scale (phospho)proteomics [48]. Similarly, RNAi-based knockdown of a single phosphatase causes a comparatively large percentage of significant changes in protein phosphorylation relative to protein abundance [49]. Given that approximately 2–4% of eukaryotic genes encode kinases or phosphatases and regulation of protein phosphorylation networks is a likely determinant of cell- and tissue-specific function [42], large-scale (phospho)proteomics is an essential and powerful tool for understanding pluripotency and differentiation [1, 10, 13].

The Workflow of (phospho)proteomics: An Overview

Interdisciplinary, large-scale (phospho)proteomic analyses involve cell biology, multidimensional liquid chromatography (MDLC), tandem mass spectrometry (MS/MS) and bioinformatics analysis of the data, components of which are introduced in Tables 1 and 2. These analyses typically employ a "bottom-up" workflow, in which the total (phospho)proteome is digested by a protease, most commonly trypsin [1, 10, 13, 50–52], in contrast with "top-down" proteomics, which involve MS and MS/MS analyses of intact polypeptides or proteins [53–55]. The application of ever improving peptide separation and phosphopeptide enrichment techniques has facilitated improvement of the sensitivity of phosphoproteomic analyses. Separation of complex total (phospho)proteomes is required prior to phosphopeptide enrichment [13]. For analysis of protein phosphorylation, effective phosphopeptide enrichment is essential. Commonly employed techniques for phosphopeptide enrichment include IMAC [10, 50, 52, 56–59], TiO_2-based phosphopeptide enrichment [56, 60–62], and soluble polymer-based phosphopeptide enrichment [56, 63, 64].

Faster mass spectrometers with improved sensitivity and mass accuracy have enabled detection of thousands of phosphoproteins in a single sample. One commonly used instrument is the LTQ Orbitrap series, which are hybrid instruments comprised of a linear ion trap for high-sensitivity and speed, and an Orbitrap mass analyzer for high mass accuracy and resolution. Complementary use of these two mass analyzers is briefly described below. Another successful hybrid mass spectrometer design, among several, is a quadrupole-time-of-flight (qTOF or qqTOF) instrument.

The specific procedures of (phospho)proteomic analyses vary, including lysis buffer composition, phosphopeptide enrichment methods, peptide fragmentation mode, mass spectrometer and bioinformatics algorithms. Standardization efforts have been proposed, including the suggestion that the more innovative analytical platforms should be broadly adopted by others [65]. However, early suggestions of consistency among hESC (phospho)proteomes, reported by different groups, is emerging, as described above.

Total (phospho)proteome analyses provide a daunting amount of data. The data is searched against a protein database, to identify proteins, on the basis of peptides derived from the proteins, via the typical bottom-up workflows. Database searches, and subsequent analyses, such as combining multiple data files from a single sample, differential quantification, pathway analyses, and other tasks requires powerful computational resources that deliver results which are amenable to understanding the data, the differences among samples, and gleaning the important trends and specific results provided by the data. Ongoing analyses of the data, termed data mining, are important to increase the value of (phospho)proteomic data.

Protein Sample Preparation is a Critical Factor

Following culture of high quality cells, yielding at least 1 mg of total protein per sample, cells are lysed [10]. Buffers include salts, detergent, protease and phosphatase inhibitors and compatibility with the MDLC-MS/MS

Table 1 Selected concepts and terminology of relevance to (phospho)proteomics (adapted and expanded from [65, 132])

Term	Description
Multi-dimensional liquid chromatography (MDLC)	Separation scheme for total (phospho)proteome analyses using three or more dimensions of separation of the complex mixtures of peptides derived from total (phospho)proteomes. The most common example would be SCX (dimension 1), IMAC and/or TiO_2 (dimension 2), and reversed-phase (RP; dimension 3).
Strong cation exchange chromatography (SCX)	Common mode of separation, typically of complex mixtures of peptides, in which there can be a partial enrichment of phosphopeptides based on the lower solution charge state of peptides due to phosphorylation, at a pH of ca. 2.7, which can result in earlier elution of phosphopeptides than cognate non-phosphopeptides, to obtain some phosphopeptide enrichment. SCX results in separation of the total proteome into simpler fractions.
Immobilized metal affinity chromatography (IMAC)	Phosphopeptide enrichment method using trivalent metal ions (usually Fe^{3+} or Ga^{3+}) bound to a stationary phase to selectively chelate negatively charged phosphate groups of phosphoproteins, or more commonly, phosphopeptides.
TiO_2-based phosphopeptide enrichment	Mode of chromatography to selectively enrich phosphopeptides, in which TiO_2 particles selectively bind phosphate groups of phosphopeptides through Lewis acid-base interactions.
Tandem mass spectrometry (MS/MS)	Measurement of mass-charge (*m/z*) ratio and intensity of precursor ions followed by isolation of individual precursor ions, their fragmentation and scanning the *m/z* ratios of the resulting product ions to deduce peptide sequence.
Collision-induced/activated dissociation (CID/CAD)	Peptide fragmentation mode based on collision in the presence of a low pressure of inert gas and resonant excitation. CID and CAD are similar terms for the same fragmentation (activation) mode.
Electron transfer dissociation (ETD)	Peptide fragmentation mode that can be more suitable to preserving PTMs, notably phosphorylation and glycosylation, than CID. ETD results in fragmentation of peptides by reactions initiated by the transfer of electrons to the peptides.
Stable isotope labeling by amino acids in cell culture (SILAC)	Method for relative quantification of protein expression between or among two or three samples that relies on incorporation of ^{13}C and ^{15}N labeled ("heavy") Arg and Lys residues in one or two of the samples, so that peptides with the same sequence, from the different samples, are distinguishable by their *m/z* ratios.
False discovery rate (FDR)	Predicted rate of false positive identification, calculated by comparing peptide and/or protein identifications from forward database searches with reversed database searches, or by statistical models based on the expectation maximization algorithm [90, 91].
Orbitrap	Widely used mass analyzer with high mass accuracy and resolution, which measures the oscillation frequency of ions along the Orbitrap axis, to enable precise determination of the *m/z* ratio.
Biological replicate	Repeat of the biological aspect of a given experiment, such as use of two separate cell samples, or independently repeating a cellular treatment.
Technical Replicate	Repeat of the proteomics portion of the experiment, thus analyzing the same sample again with the proteomic workflow.

methods is important. Protein kinase inhibitors and protease inhibitors are valuable in preserving the fidelity of the (phospho)proteome during and after cell lysis. The importance of phosphatase inhibitors for increased identification of protein phosphorylation sites was demonstrated [66]. We have had success with a specific lysis buffer, protein precipitation from clarified lysates using ammonium sulfate, re-suspension in the presence of phosphatase inhibitors and 8 M urea, gel filtration chromatography, digestion with modified trypsin, including standard reduction and alkylation reactions, desalting and drying the peptides prior to subsequent separation and phosphopeptide enrichment [10]. Other groups have also reported successful procedures [1, 13, 36]. One lysis buffer contained 8 M urea (to enhance lysate component solubility and inhibit many protein-based biochemical reactions), and the protein phosphatase inhibitors sodium fluoride (serine/threonine/acid phosphatase inhibitor) and sodium orthovanadate (tyrosine phosphatase inhibitor) [1], while another included these ingredients and β-glycerophosphate (a protein serine/threonine phosphatase inhibitor) plus protease inhibitor tablets [13]. A fourth lysis buffer contained 8 M urea, sodium pyrophosphate (buffering agent and emulsifier), sodium orthovanadate, complete mini ETDA-free protease inhibitor and combined phosphatase inhibitor [36].

Table 2 Bioinformatics resources useful to (phospho)proteomics

Resource	Purpose	Reference
NetworKIN (http://networkin.info/search.php)	Kinase-substrate and phosphorylation site interactions	111]
GeneGo Metacore Pathway Analysis (http://www.genego.com/)	Systems biology/pathway analysis	[103]
Ingenuity Pathway Analysis (http://www.ingenuity.com)	Systems biology/pathway analysis	
KEGG (http://www.genome.jp/kegg/)	Systems biology	[133]
STRING (http://string-db.org/)	Systems biology	[134]
NetPhosK (http://www.cbs.dtu.dk/services/NetPhosK/)	Kinase-substrate and phosphorylation site interactions	[135]
PHOSIDA (http://www.phosida.com)	Interactive database of phosphopeptides	[108]
Phospho.ELM http://phospho.elm.eu.org	Kinase-substrate and phosphorylation site interactions	[136]
PhosphoSite (http://www.phosphosite.org)	Kinase-substrate and phosphorylation site interactions	[109]
MaxQuant	Precursor mass correction, peptide identification and quantification of SILAC data	[105]
Abacus, QSpec	Spectral count analysis	[100, 106]
Phosphopep	Phosphorylation site database, analysis of the sites, data integration	[107]
PRIDE http://www.ebi.ac.uk/pride	Proteomic database, contains PTMs, support manuscripts, papers	[110]
Trans-Proteomic Pipeline (TPP) http://tools.proteomecenter.org/	Software tools for FDR filtering, quantification, others	[90, 91]

Commonly Used Separation and Phosphopeptide Enrichment Strategies Include SCX, IMAC, TiO_2, and Reversed-Phase

Because a total (phospho)proteome is exceedingly complex, the range of protein abundance is large, many phosphoproteins are low abundance, phosphorylation site occupancy is often sub-stoichiometric and some phosphopeptides ionize and fragment inefficiently, chromatographic fractionation of the (phospho)proteome, including phosphopeptide enrichment, is essential to successful (phospho)proteomic analyses. Several separation and phosphopeptide enrichment methods have been developed (Table 1). SCX is a commonly used mode of separation, primarily to simplify the (phospho) proteome, and SCX can result in partial enrichment of phosphopeptides, especially in fractions eluting early in the gradient [67]; (J. Hou and L. M. Brill, unpublished data). Hydrophilic interaction chromatography (HILIC) has been used as a successful alternative to SCX [68, 69]. Following simplification of the (phospho)proteome by SCX or HILIC, phosphopeptide enrichment, most commonly by IMAC or TiO_2, is performed. IMAC is thought to selectively enrich phosphopeptides by chelation of their negatively charged phosphate group(s) to metal cations (often Fe^{3+} or Ga^{3+}) [10, 36, 50, 52, 57, 58, 70, 71]. A complement to, and/or substitute for IMAC, is the use of TiO_2 to enrich phosphopeptides [56, 60–62]. Peptides containing an abundance of negatively charged aspartate and/or glutamate residues are also enriched by IMAC and TiO_2, so this is a common reason why, in practice, pure populations of phosphopeptides are not obtained. However, use of methyl-esterification reactions of these amino acid side chains and peptide C-termini can improve selectivity of IMAC for phosphopeptides [50, 52]. Less commonly used phosphopeptide enrichment uses soluble polymers, variously termed dendrimers or PolyMAC, depending on the polymer preparation [56, 63, 64]. Among the reports of large-scale hESC (phospho)proteomes, SCX and IMAC separation/enrichment strategies were used in two of the studies [10, 36], whereas TiO_2 was used in the two others [1, 13].

Immediately before introduction into the mass spectrometer, reversed-phase (RP) liquid chromatography (LC) is used to separate the peptide mixtures in the elution fractions (or flow-through/wash fractions, if the experimental goals include their analysis) from phosphopeptide enrichments. RP-LC is coupled directly to electrospray ionization- (ESI) MS/MS. Nanoflow (flow rates of ca. 10–300 nanoliters/min) RP-LC is commonly used for high sensitivity ESI-MS/MS studies [1, 10, 13, 36, 47, 50, 52, 57, 59, 71], although we currently use higher flow RP-LC with robust operation and sensitivity similar to nanoflow LC [72]. The result of LC-ESI is introduction of ionized peptides and phosphopeptides into the mass spectrometer for MS/MS analyses. The peptide ions are typically positively charged, predominantly with charges of ca. 2^+–6^+, with most ions in the lower portion of this charge state range. Another common ionization technique, termed matrix assisted, laser desorption ionization (MALDI) has not been used in published descriptions of total (phospho)proteome analyses of hESCs.

MS/MS of Peptides and Phosphopeptides is Used in Large-Scale (Phospho)Proteomic Analyses

The MS/MS methods are typically “data-dependent”, meaning that a scan of the precursor ions (MS scan) is performed first, in which the mass to charge ratio (m/z) is

measured, followed by isolation of the most abundant precursors, one at a time, their fragmentation, and re-scanning the product ions. Product ion scans are termed MS/MS scans. In some cases, a dominant product ion, resulting from prominent neutral loss of H_3PO_4 (phosphoric acid) from the phosphopeptide precursor ion, is re-isolated, re-fragmented, and the products re-scanned as an MS/MS/MS scan, which can result in an improved ability to identify the phosphopeptide [73]. When a high-resolution, high mass-accuracy mass analyzer is used, e.g. an Orbitrap on the "back" of a hybrid linear ion trap/Orbitrap instrument, precursor ions are typically scanned in the Orbitrap, whereas MS/MS scans are usually performed in the fast, highly sensitive linear ion trap in the "front" of the instrument (which has lower mass accuracy and resolution than the Orbitrap). An alternative hybrid mass spectrometer substitutes a Fourier-transform-ion cyclotron resonance (FT-ICR) mass analyzer for an Orbitrap [73], and this instrument's availability preceded that of those with Orbitraps. Recently, MS and MS/MS scans were both performed in the Orbitrap, using a newer generation of Orbitrap instrument [74].

A "top 20, data-dependent MS/MS" method refers to a method in which each of the repeating instrument cycles consists of one MS scan followed by isolation, fragmentation and MS/MS scans of the 20 most abundant precursor ions, one precursor ion at a time. In modern mass spectrometers these cycles are fast, taking only ca. 2–4 s, which enables thousands of ionized peptides to be analyzed in single LC-MS/MS runs. Following their MS/MS analysis, precursor ions are placed on an exclusion list, to enable analysis of as many precursor ions as possible, rather than analyzing only the most abundant ions over and over. This then enables the mass spectrometer to subject other, lower abundance precursor ions to MS/MS scanning.

Collision-induced dissociation (CID) and electron-transfer dissociation (ETD) are commonly used methods of peptide fragmentation in (phospho)proteomic analyses. Electron capture dissociation (ECD) is believed to have similar peptide fragmentation mechanisms as ETD, is used in conjunction with differing mass spectrometers (typically FT-ICR instruments lacking ion traps) [75] and has not been used in published studies on hESC (phospho) proteomics. CID is most frequently employed in phosphoproteomic analyses, often using an ion trap mass spectrometer, which is used to deduce peptide and phosphopeptide sequences (Fig. 3a). Activation by CID results primarily in b- and y-product ions whose *m/z* is measured in the MS/MS scans, in order to deduce the peptide sequence using the product ions. Several search algorithms are available to

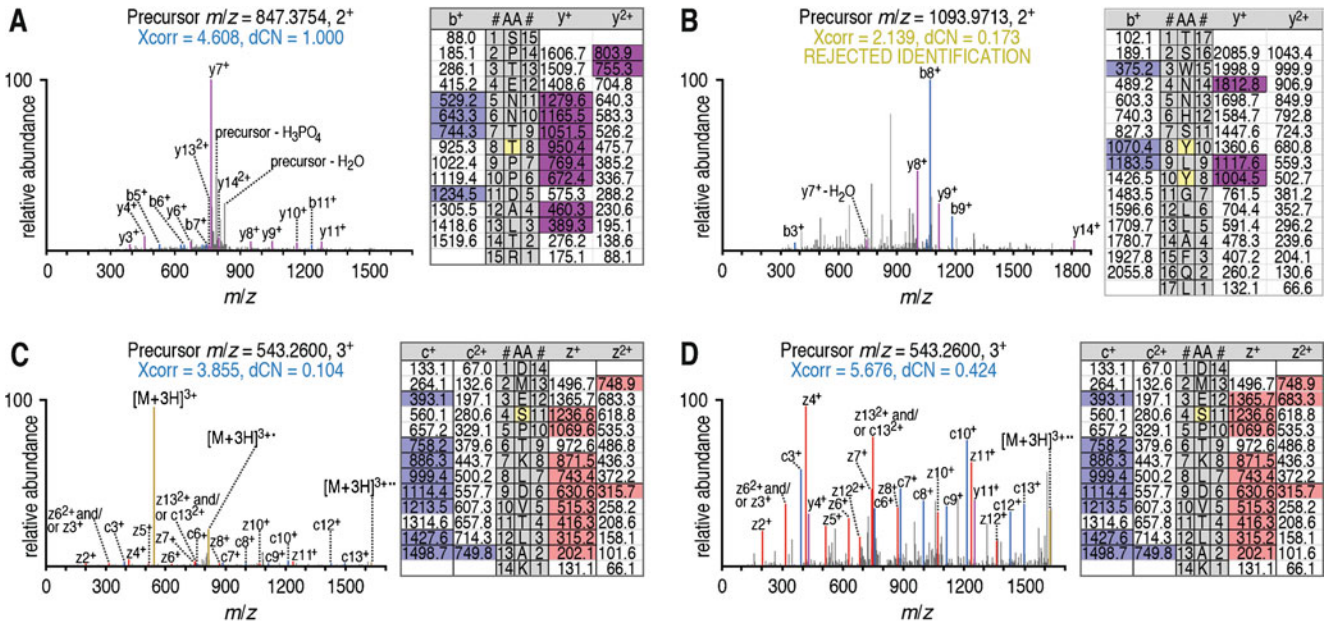

Fig. 3 Examples of annotated MS/MS spectra used to identify phosphopeptides, and evaluation of data quality. MS/MS spectra were modified from their COMET Spectrum View (by J. Eng © Institute for Systems Biology (ISB) 2001), and the SEQUEST Xcorr plus dCN scores [76] are also presented as important indicators of the quality of the peptide-spectrum match (PSM) (panels A–D). A. High-quality MS/MS spectrum using activation by CID to identify a phosphopeptide containing phosphothreonine (pT) residue 230 from the Serine/Threonine Protein Kinase PAK1 in pluripotent hESCs. B. Low-quality MS/MS spectrum, using activation by CID and the proposed phosphopeptide identification, which was rejected. C. High-quality MS/MS spectrum, using activation by ETD, to identify a phosphopeptide containing phosphoserine (pS) residue 280 from Microtubule-Associated Protein 4 (MAP4) in pluripotent hESCs. D. The same spectrum as in panel C, but following processing to remove un-fragmented precursor and charge-reduced, un-fragmented precursor ions and to remove neutral loss product ions from un-fragmented precursor ions

match the experimental MS/MS spectra to theoretical MS/MS spectra, which are computed from protein databases, resulting in a peptide-spectrum match (PSM). A common search algorithm, used for the examples presented in Fig. 3, is SEQUEST [76], which we focus on for clarity and describe briefly. However, a variety of other search algorithms are available (see below). The Xcorr score from SEQUEST is a measure of the goodness of fit of the experimental and theoretical MS/MS spectra, and the dCN score is a measure of the difference between the best fitting peptide sequence and the next-best fitting peptide sequence [76]. Relatively high Xcorr and dCN scores are generally necessary for high confidence peptide IDs, whereas low scores typically lead to rejection of the proposed identification (Fig. 3b). Each peptide has unique fragmentation characteristics, and it is difficult to predict how specific peptides will fragment [77].

Some peptides and phosphopeptides are more effectively identified using fragmentation by ETD [78] (Fig. 3c–d) than CID, especially those precursor ions from phosphopeptides with lower *m/z* ratios and higher charge states [79]. One challenge with ETD MS/MS spectra is effectively searching them against protein databases, largely because search algorithms have been predominantly developed for searching CID MS/MS spectra. Recent advances in processing of the raw data, including the removal of un-fragmented precursor ions plus charge-reduced precursor ions and neutral loss products from these precursor ions have led to increased numbers of high quality IDs made from ETD-MS/MS spectra [75, 80–82] (Fig. 3d).

Among the first reports of hESC phosphoproteomes, three [1, 10, 13] exclusively used CID, and one [36] used a MS/MS method termed "decision tree", in which CID and ETD were applied in combination, depending on the *m/z* ratio and charge state of the precursor ion to be fragmented [79]. The instrument control software of some mass spectrometers includes the decision tree method [79], and this is our current MS/MS method of choice. However, there are additional MS/MS methods available, which are beyond the scope of this review.

Fragmentation of phosphopeptides by CID (Fig. 3a) sometimes results in prominent neutral loss of H_3PO_4 from the precursor, at the expense of peptide backbone fragmentation, resulting in fewer b- and y-ions and thus a decreased ability to deduce the peptide sequence [46]. Proline is also a favored site of fragmentation by CID (e.g. Fig. 3a). Potential phosphorylation site rearrangement by CID was proposed [83], but others reported that it is negligible [84, 85]. In contrast, ETD tends to result in an absence of neutral loss of H_3PO_4 from phosphopeptides, and yields primarily c- and z product ions [46, 78], although a lesser quantity of b- and y-ions result from ETD [86]; (Fig. 3d). Moreover, fragmentation of peptides containing multiple Arg and Lys residues by ETD tends to be more robust than by CID [78].

Use of CID and ETD peptide fragmentation was compared in a (phospho)proteomic analysis of hESCs and yeast [36, 79]. ETD identified more phosphopeptides in later SCX fractions, whereas CID identified proportionally more phosphopeptides in early SCX fractions, in which many of these peptides had a lower charge [79], similar to our results (J. Hou and L. M. Brill, unpublished data). ETD was twice as likely to produce backbone cleavage in the vicinity of phosphoserine (pS) residues than CID, although there was no consistent difference in efficiency between the fragmentation methods at phosphotyrosines (pY) [79]. In total, CID identified 5773 unique phosphopeptides and ETD identified 8603 unique phosphopeptides with 2421 identified by both methods [79]. Sequence coverage was better for ETD than for CID, which is important for phosphorylation site localization, which was 49.8% for ETD and 26.9% for CID [36].

Filtering to achieve a low false discovery rate (FDR) are important to yield reliable MDLC-MS/MS-based (phospho) proteomic analyses. High mass accuracy, high resolution MS data improves the sensitivity and accuracy of peptide and phosphopeptide identifications. "Lock mass" MS/MS methods can further improve the mass accuracy of Orbitrap analyzers and can result in more (phospho)protein IDs [87, 88], but we have had more identifications without the use of lock mass (J. Hou and L. M. Brill, unpublished data).

The FDR can be estimated on the basis of decoy database searches in conjunction with "forward" database searches [89], and by the use of statistical tools, including PeptideProphet and ProteinProphet [90, 91] from the transproteomic pipeline (TPP) at the Institute for Systems Biology. Although there is no generally accepted FDR, and the FDR used can vary depending on the experimental goals, we suggest a FDR range of 0.005–0.009 with the large-scale datasets collected during total (phospho)proteome analyses, to yield datasets with good (phospho)proteome coverage and minimal false discoveries. Poor quality, less reliable spectra, with poor fragment ion coverage and numerous unexplained, relatively large fragment ions, tend to score low (Fig. 3b), and are filtered out of datasets by application of strict FDRs.

Although it is tedious, time consuming and requires expert personnel, manual validation of PSMs of phosphopeptides can be very important [10, 92]. There are defined criteria for accuracy of manual validations [92]. In addition, software to analyze the reliability of phosphopeptide identifications has been reported [93]. In anecdotal discussions, there is general agreement that phosphopeptides are often not accurately scored by database searching and filtering algorithms. Note that protein phosphorylation IDs that form the basis for extensive biological follow-up experiments should be especially carefully validated [92].

Quantitative (phospho)proteomics Used on hESCs Include SILAC and Label-Free

One quantitative proteomics method that was reported in mouse ESCs, but not hESCs, is termed iTRAQ (isobaric tagging for relative and absolute quantification [94]), and, although a good technique, will not be described in detail here, for clarity. Another method, which has been applied to hESC (phospho)proteomics, is termed stable isotope labeling with amino acids in cell culture (SILAC) [95]. SILAC uses amino acids containing either heavy (labeled) or light (normal/unlabeled) stable isotopes to compare two or three samples by MS, providing relative quantification of peptides and phosphopeptides. One of two or three cell populations is grown in culture media in which Arg and/or Lys contain stable, "heavy" isotopes of carbon (^{13}C) and nitrogen (^{15}N). This causes an increase in the mass of the "heavy" peptides derived from labeled proteins compared to the unlabeled peptides from unlabeled proteins. The labeled and unlabeled peptides are thus distinguishable by the mass spectrometer. With SILAC, two or three samples can be analyzed simultaneously, eliminating the need to compare separate analyses, thus reducing variability. In "triple SILAC", one sample is unlabeled, a second sample is labeled with intermediate mass labels and a third contains proteins with the heaviest mass labels. However, SILAC media is expensive, and the complexity of the peptide mixtures are approximately doubled or tripled. SILAC was used to compare $^{13}C_6$, $^{15}N_4$ Arg and $^{13}C_6{}^{15}N_2$ Lys-labeled, undifferentiated hESCs to unlabeled hESCs that were treated by adding BMP4 and removing FGF2 for 30, 60 or 240 min [1]. SCX separation, phosphopeptide enrichment with TiO_2 and LC-MS/MS was employed. About 2/3rds of the identified peptides were quantifiable, and of those, half of the phosphopeptides showed changes in relative abundance during treatment. Rapid, temporal dynamics of protein phosphorylation has also been examined in HeLa cells [73]. Similarly, three isotopically distinct versions of Lys and Arg ("triple SILAC") were used to label hESCs in another study using SILAC to examine quantitative changes during two different treatments, each inducing non-specific differentiation of HUES-9 and Odense-3 hESCs [13]. The relative abundance of ca. 50% of the phosphopeptides also changed during 24 h of differentiation [13].

Label-free quantification, frequently using spectral counts, and which can also include extracted ion chromatograms for further quantification, is also a reliable method to estimate the relative abundance of proteins and protein phosphorylation events [10, 42, 59, 96–99]. The spectral count of a protein is the number of times that the protein is identified by the MS/MS spectra that yield identification of peptides derived from the protein [96]. A large portion of the peptides is derived from only one known protein (non-degenerate peptides). The protein from which degenerate peptides are derived is ambiguous, but effective methods to assign the relative number of spectral counts to specific proteins, for these degenerate peptides, have been reported recently [98, 100]. In the only direct comparison of SILAC- and label-free (spectral counting) quantification we are aware of, the two methods performed similarly [51]. Using label-free quantification, over 50% of the phosphoproteins (929 of 1602) contained more phosphorylation site identifications in either undifferentiated hESCs or their RA-differentiated derivatives [10], a similar percentage (58%) as the studies using SILAC [1, 13], despite the use of very different methods to induce differentiation of the cells. Moreover, prominent groups of phosphoproteins with extensive changes during differentiation, in each of the three studies, included transcription regulators, protein kinase-driven signaling cascades, and the networks of phosphoproteins identified and quantified were highly complex. Although not proof of a causative relationship, these correlations suggest a prominent role for protein phosphorylation in the regulation of pluripotency and differentiation. Several pathways, including the VEGF, JNK, EGF, IGF and PDGF pathways were predicted, and found, through follow-up experimentation, to be important in undifferentiated hESCs, on the basis of phosphoproteomic and pathway analyses [10, 15]. Recent work demonstrating that PDGF-AA (10 ng/ml) was capable of complementing a sub-threshold bFGF concentration (4 ng/ml) to stably maintain undifferentiated hESCs under chemically defined conditions, as predicted by phosphoproteomics and pathway analyses [10], was consistent with earlier results, under undefined culture conditions, that PDGF together with sphingosine-1 phosphate facilitated maintenance of undifferentiated hESCs [12]. Thus, phosphoproteomic analyses of hESCs and their differentiated derivatives support an accelerated understanding of the biology of cell states.

Bioinformatics Analysis of (phospho)proteomic Datasets

As stem cell (phospho)proteomic applications expand, robust bioinformatics procedures for data analysis, including cross-referencing of genotypic and clinical parameters with proteomics, will also be necessary [101, 102]. Delineating identified and quantified proteins to families, disease markers, therapeutic targets and signaling pathways can cumulatively establish informative trends among datasets and highlight critical components otherwise masked by data overload [103]. No universal data analysis workflows have been established, and studies typically employ different bioinformatics algorithms with varying foci. SEQUEST, Mascot, Spectrum Mill, X!Tandem and OMSSA are commonly used database search algorithms, and Inspect [104] and MS-GF [86] may be increasingly

used. Other tools with relevance to (phospho)proteomics are available (Table 2). MaxQuant software has been reported for peptide identification and quantification of SILAC data [105]. Spectral counting/label-free quantification can include the analysis of statistical significance of spectral counts of proteins between or among samples, which requires extensive computation [100, 106]. Several proteomic and phosphoproteomic databases are available, such as Phosphopep [107], PHOSIDA [108], Phosphosite [109] and PRIDE [110]. This allows researchers to query known protein phosphorylation events, among other functionalities. Moreover, TPP contains a suite of useful bioinformatics tools.

MetaCore Pathway Analysis (GeneGo, Inc.) and Ingenuity Pathway Analysis (Ingenuity, Inc.) can bring increased organization and biological meaning to (phospho)proteomic data. The datasets may be queried for known protein families, cellular processes, signaling pathways, molecular complexes, disease-related protein networks and pharmaceutical targets. Pathway analyses can provide user-friendly graphical interfaces. NetworKIN allows bioinformatics prediction of kinases and their substrates [111], and was used to identify the potential kinases phosphorylating the proteins identified in hESCs [1]. Pictorial representations can be constructed indicating the possible known and inferred connections between kinases and substrates.

Biological Insights from Unbiased Phosphoproteomic Studies of hESCs

The results of large-scale (phospho)proteomic analyses are unbiased [10, 112], meaning that the proteins that are present and detectable are what is discovered, in contrast with techniques specifically targeting the detection of a given (phospho)protein, such as with FACS, immunohistochemistry, immunofluorescence, or Western blots. Although specific peptides and phosphopeptides can be targeted for precise quantification using an MS method entitled multiple reaction-monitoring (MRM, also termed selective reaction monitoring, SRM) [94, 113], MRM-based analyses of hESCs have not been reported. Examination of proteomics data for biological insight has both addressed and stimulated hypotheses in stem cell biology. One unanticipated observation is a disproportionately large percentage of pY on proteins with more phosphorylation site identifications in undifferentiated hESCs, consistent with the observation that receptor-tyrosine kinase signaling is critical to the cellular state [10]. Using pY immunoprecipitation followed by MS/MS, it was proposed that the surprisingly prominent pY in hESCs could be facilitated by utilization of unique pY target sequences [114]. In addition, pS and phosphothreonine (pT) can have similar or differing biological effects as pY [115] and pS/pT can oppose the action of pY [116]. For instance, Cortactin, though a Src substrate activated by pY, showed six detectable pS sites but no detectable pY [117].

A protein implicated in epigenetic regulation of pluripotency, DNMT3B, is consistently phosphorylated in hESCs [1, 10, 13] and its phosphorylation was identified more often in undifferentiated than differentiated cells [10]. Several phosphorylation sites were also identified on LIN28, an RNA binding protein conducive to cellular reprogramming, and some sites changed in relative abundance during differentiation of hESCs [1, 10, 13].

These studies identified numerous phosphorylation sites on proteins involved in signaling pathways that regulate self-renewal, such as WNT and TGFβ. For instance, 17 non-redundant sites on proteins known to participate in TGFβ signaling were identified but not disclosed [36] and phosphorylation of SMAD 2/3, GSK3 and β-catenin was identified [1, 10, 13, 36]. In response to addition of BMP4 and removal of bFGF, the abundance of phosphorylated transcription factors and nucleic acid binding factors decreased, whereas total phosphopeptides increased [1]. In response to RA, many proteins showed increased or decreased phosphorylation during differentiation, though cumulatively an increase in phosphorylation during differentiation was again observed [10]. It is unclear whether the effects of exogenous RA or BMP4 are directly downstream of the ligand or common, non-specific effects of differentiation [1, 10], but a set of phosphorylation sites that are generally responsive to non-specific differentiation was proposed [13]. In addition, signaling cascades not previously known in hESCs, or whose importance was underappreciated in pluripotency were also implicated, via phosphorylation of some of their members, as participating in maintenance of undifferentiated hESCs. The JNK and PDGF pathways were important for maintenance of undifferentiated hESCs, shown in biological follow-up experiments [10]. Similarly, phosphorylation of the JNK target JUN increased after BMP4 addition and removal of bFGF to induce differentiation [1], novel results revealed by (phospho)proteomic analyses.

Oxygen tension is a critical factor in the biology of stem cells [118]. For example, mild hypoxia improves the efficiency of generation of both mouse and human iPSCs [119]. FRAP1 (mTOR) was phosphorylated in hESCs [10, 13] and this protein activates HIF-1 (Hypoxia-inducible factor-1, a heterodimer of HIF1A and ARNT (HIF-1β) [120]. ARNT was also phosphorylated in undifferentiated hESCs [10]. The HIF pathway is the primary mediator of hypoxic adaptation [118], and in mouse ESCs, Arnt directly activates expression of Pou5f1 (Oct3/4) [121], also suggesting that Arnt participates in regulation of undifferentiated ESCs. These findings suggest a potential role of phosphoprotein networks in facilitating growth and decreased spontaneous differentiation of hESCs in mild hypoxia [122].

Repeat Analysis and Experimental Validation of (phospho) proteomic Data

Although challenging from the perspective of workload and data analysis, it is important to complete a given workflow two or more times on each of two or more identically cultured and/or treated, independent samples. This approach provides technical replicates of biological replicates. It is not possible to identify all the same (phospho)proteins in separate experiments, but substantial overlap is often obtained, and replicate analyses help to increase the coverage of the (phospho)proteome. Because proteomics technologies are rapidly developing it is not clear how many phosphoproteins will be identified or how many residues in a proteome are phosphorylated (Fig. 2). It is also likely that different cell types will have different numbers of (phospho)proteins.

It is important to validate (phospho)proteomic results with independent experiments. One common approach is to select a portion of the proteins for confirmatory studies examining their presence, phosphorylation and relative abundance, often by Western blotting [1, 10, 13]. Agonists or antagonists of key signaling proteins in pathways implicated by (phospho)proteomics can also be used to test if cellular responses are as predicted, but this approach was only reported once as a component of a large-scale hESC (phospho)proteomic study [10]. However, biological follow-up experiments have been increasingly pursued in association with proteomic analyses [9, 14, 15, 123, 124]. Another potential approach is to perform large-scale epigenetic or genomic studies to enable bioinformatics comparisons of proteomic datasets to genomic studies, which was done with mouse ESCs but not hESCs, although there is often a poor correlation between mRNA and protein abundance [94, 125]. Similarly, preliminary examinations of a correlation between relative protein phosphorylation and mRNA abundance also suggested that it could be relatively low (ca. 30%; Supplementary Table 8 [10]). In addition, histone methylation patterns that indicate inactive, poised or actively transcribed chromatin regions contribute to the maintenance of pluripotency [126]. Comparison of histone methylation patterns, and the implied active or inactive transcriptional state associated with the genes encoding the proteins identified by (phospho)proteomics, showed that the vast majority of the methylation patterns and ability to identify the encoded protein were in agreement [1].

Limitations of Proteomics Approaches

Each step of the MDLC-MS/MS workflow can cause variation in the results, which should be minimized with attention to detail and strict consistency of the procedures from one sample to the next. If a phosphopeptide is identified in one sample and not another, it is likely to be more abundant in the sample in which it was identified, but failure to identify the phosphopeptide does not necessarily mean it is absent [10]. Also, highly annotated and reliable protein databases can contain errors, so peptide/protein identifications can be missed, and database search algorithms are imperfect, thus also missing some IDs. However, combined use of different database search algorithms (mentioned above) is a well-known method to increase (phospho)proteome coverage. Some phosphorylation sites are ambiguous, but the confidence in the proposed site localization can be estimated [13, 127], and when warranted, manually examined. However, confidently identified phosphopeptides with ambiguous phosphorylation sites still limit the possible sites to the sequence of the peptide and its S, T or Y residues, and suggest the possibility of targeting the phosphoprotein, from which they are derived, for follow-up [10].

Low abundance proteins can be difficult to detect, but improvements in instrumentation and separation protocols are leading to advances in their analysis. Larger, more comprehensive total (phospho)proteome datasets could mitigate the problem of important (phospho)proteins evading detection. Moreover, improvements in database searching algorithms are leading to rapid advances in confident peptide identification, and likely to decreased false negative IDs.

Identification of phosphorylation sites does not directly indicate the effect on the proteins (structure, stability, interactions, catalytic activity) or cellular pathways (i.e. activating or inhibiting). However, it is becoming increasingly clear that (phospho)proteins play an important role in regulating cellular identities [1, 10, 13, 36]. More detailed follow-up is subsequently needed, but care is required to select the most important phosphorylation sites to investigate in detail, such as with site-specific mutants. Site-specific follow-up was described for SOX2 in HeLa cells over-expressing SUMO2, and nuclear localization of SOX2 was not affected by these site-specific mutations in HUES-7 cells [1].

Conclusions and Perspectives

Use of MS for study of PTMs provides an unbiased, systems-level view of molecular states of biological systems that is unobtainable by smaller-scale approaches using antibodies. Proteins and phosphoproteins, the final products of most genes, are also directly analyzed, rather than analyzing intermediate products, the mRNAs. The application of MDLC-MS/MS-based (phospho)proteomics to stem cell biology will likely increase, and requires multidisciplinary teams of scientists. Understanding proteomics

data, and how it was derived, is critical to its interpretation and evaluation. Similarly, those more focused on proteomics need to understand the stem cell biology and implications of the data. (Phospho)proteomics studies of hESCs are some of the first studies to profile post-translational regulation of proteins during self-renewal and differentiation, but a large-scale analysis of the hESC proteome had been reported previously [128]. Proteomic analyses could be applied to other PTMs in hESCs. For example, glycopeptide enrichment followed by LC-MS/MS was used to identify 180 glycoproteins from murine ESCs and their embryoid body derivatives [129], and 6367 sites of N-glycosylation were mapped to 2352 proteins from 4 mouse tissues [130].

Deciphering the relationships of the genome, transcriptome and proteome is challenging. A likely obstacle to understanding proteome/transcriptome relationships has been insufficient sensitivity of mRNA and protein detection assays. As the sensitivity of mRNA detection has improved, it appears that many more genes are actively transcribed in diverse cell types than previously thought [33, 131]. Similar trends could be emerging with proteomes as MDLC-MS/MS technology improves. In addition, the least abundant proteins may follow a tendency to be the most tissue-specific [41]. As detection of PTMs becomes more sensitive, thousands of new phosphorylated residues in larger datasets have been detected (Fig. 2). The proportion of a given protein that is phosphorylated at specific residues also varies widely [65, 132]. Moreover, it is important to clarify the biological relevance of the protein phosphorylation sites. Long lists of (phospho)proteins are not the only intended experimental endpoints. Thoughtful biochemical, bioinformatics, and biological follow-up experiments are needed as well.

Phosphoproteomics may be capable of identifying abnormally regulated proteins and pathways in disease models, including the use of iPSCs, as it has in cancer and some other biological model systems. Future bioinformatics studies could strive to model and predict potential effects of genetic variation on protein expression, PTMs, pluripotency, multipotency, health and disease. Finally, further improvements in MDLC-MS/MS and supporting technologies and application of MRM-MS methods should help to discriminate between false negatives and true negatives, as well as to make the technologies increasingly quantitative and useful as biological analysis and prediction tools.

Acknowledgements We apologize to those whose work was not cited due to space constraints.

Support was provided by the SBMRI-NCI Cancer Center Support Grant 5 P30 CA30199-28, the La Jolla Interdisciplinary Neuroscience Center Cores Grant 5 P30 NS057096 from NINDS, RC2 MH090011 from NIMH and by the UCSD Dept. of Psychiatry NIMH T32.

Conflicts of Interest The authors declare no potential conflicts of interest.

References

1. Van Hoof, D., Munoz, J., Braam, S. R., et al. (2009). Phosphorylation dynamics during early differentiation of human embryonic stem cells. *Cell Stem Cell, 5*, 214–226.
2. Takahashi, K., Tanabe, K., Ohnuki, M., et al. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell, 131*, 861–872.
3. Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., et al. (1998). Embryonic stem cell lines derived from human blastocysts. *Science, 282*, 1145–1147.
4. Boyer, L. A., Lee, T. I., Cole, M. F., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell, 122*, 947–956.
5. Brandenberger, R., Wei, H., Zhang, S., et al. (2004). Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation. *Nature Biotechnology, 22*, 707–716.
6. Grskovic, M., & Ramalho-Santos, M. (2008). The pluripotent transcriptome. In: StemBook. Cambridge (MA): Harvard Stem Cell Institute.
7. Sato, N., Sanjuan, I. M., Heke, M., Uchida, M., Naef, F., & Brivanlou, A. H. (2003). Molecular signature of human embryonic stem cells and its comparison with the mouse. *Developmental Biology, 260*, 404–413.
8. Sperger, J. M., Chen, X., Draper, J. S., et al. (2003). Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. *Proceedings of the National Academy of Sciences of the United States of America, 100*, 13350–13355.
9. Bendall, S. C., Stewart, M. H., Menendez, P., et al. (2007). IGF and FGF cooperatively establish the regulatory stem cell niche of pluripotent human cells in vitro. *Nature, 448*, 1015–1021.
10. Brill, L. M., Xiong, W., Lee, K. B., et al. (2009). Phosphoproteomic analysis of human embryonic stem cells. *Cell Stem Cell, 5*, 204–213.
11. James, D., Levine, A. J., Besser, D., & Hemmati-Brivanlou, A. (2005). TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development, 132*, 1273–1282.
12. Pebay, A., Wong, R. C., Pitson, S. M., et al. (2005). Essential roles of sphingosine-1-phosphate and platelet-derived growth factor in the maintenance of human embryonic stem cells. *Stem Cells, 23*, 1541–1548.
13. Rigbolt, K. T., Prokhorova, T. A., Akimov, V., et al. (2011). System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. *Science Signal, 4, rs3.*
14. Wang, J., Rao, S., Chu, J., et al. (2006). A protein interaction network for pluripotency of embryonic stem cells. *Nature, 444*, 364–368.
15. Wang, L., Schulz, T. C., Sherrer, E. S., et al. (2007). Self-renewal of human embryonic stem cells requires insulin-like growth factor-1 receptor and ERBB2 receptor signaling. *Blood, 110*, 4111–4119.
16. Xu, R. H., Peck, R. M., Li, D. S., Feng, X., Ludwig, T., & Thomson, J. A. (2005). Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. *Nature Methods, 2*, 185–190.
17. Yao, S., Chen, S., Clark, J., et al. (2006). Long-term self-renewal and directed differentiation of human embryonic stem cells in

chemically defined conditions. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 6907–6912.
18. Braconi Quintaje, S., & Orchard, S. (2008). The annotation of both human and mouse kinomes in UniProtKB/Swiss-Prot: one small step in manual annotation, one giant leap for full comprehension of genomes. *Molecular & Cellular Proteomics: MCP, 7*, 1409–1419.
19. Manning, G., Whyte, D. B., Martinez, R., Hunter, T., & Sudarsanam, S. (2002). The protein kinase complement of the human genome. *Science, 298*, 1912–1934.
20. Pera, M. F., & Tam, P. P. (2010). Extrinsic regulation of pluripotent stem cells. *Nature, 465*, 713–720.
21. Watabe, T., & Miyazono, K. (2009). Roles of TGF-beta family signaling in stem cell renewal and differentiation. *Cell Research, 19*, 103–115.
22. Katoh, M. (2007). Networking of WNT, FGF, Notch, BMP, and Hedgehog signaling pathways during carcinogenesis. *Stem Cell Reviews, 3*, 30–38.
23. Kim, J. B., Zaehres, H., Arauzo-Bravo, M. J., & Scholer, H. R. (2009). Generation of induced pluripotent stem cells from neural stem cells. *Nature Protocols, 4*, 1464–1470.
24. Maekawa, M., Yamaguchi, K., Nakamura, T., et al. (2011). Direct reprogramming of somatic cells is promoted by maternal transcription factor Glis1. *Nature, 474*, 225–229.
25. Wang, Y., Mah, N., Prigione, A., Wolfrum, K., Andrade-Navarro, M. A., & Adjaye, J. (2010). A transcriptional roadmap to the induction of pluripotency in somatic cells. *Stem Cell Reviews, 6*, 282–296.
26. Hu, D., & Wan, Y. (2011). Regulation of Kruppel-like factor 4 (KLF4) by APC pathway is involved in TGF-{beta} signaling. *The Journal of biological chemistry, 286*, 6890–6901.
27. Maherali, N., & Hochedlinger, K. (2009). Tgfbeta signal inhibition cooperates in the induction of iPSCs and replaces Sox2 and cMyc. *Current Biology: CB, 19*, 1718–1723.
28. Wend, P., Holland, J. D., Ziebold, U., & Birchmeier, W. (2010). Wnt signaling in stem and cancer stem cells. *Seminars in Cell & Developmental Biology, 21*, 855–863.
29. Millman, J. R., Tan, J. H., & Colton, C. K. (2009). The effects of low oxygen on self-renewal and differentiation of embryonic stem cells. *Current Opinion in Organ Transplantation, 14*, 694–700.
30. Page, R. L., Ambady, S., Holmes, W. F., et al. (2009). Induction of stem cell gene expression in adult human fibroblasts without transgenes. *Cloning and Stem Cells, 11*, 417–426.
31. Rodin, S., Domogatskaya, A., Strom, S., et al. (2010). Long-term self-renewal of human pluripotent stem cells on human recombinant laminin-511. *Nature Biotechnology, 28*, 611–615.
32. Newman, A. M., & Cooper, J. B. (2010). Lab-specific gene expression signatures in pluripotent stem cells. *Cell Stem Cell, 7*, 258–262.
33. Ramskold, D., Wang, E. T., Burge, C. B., & Sandberg, R. (2009). An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Computational Biology, 5*, e1000598.
34. Tavakoli, T., Xu, X., Derby, E., et al. (2009). Self-renewal and differentiation capabilities are variable between human embryonic stem cell lines I3, I6 and BG01V. *BMC Cell Biology, 10*, 44.
35. Takahashi, K., Narita, M., Yokura, M., Ichisaka, T., & Yamanaka, S. (2009). Human induced pluripotent stem cells on autologous feeders. *PloS One, 4*, e8067.
36. Swaney, D. L., Wenger, C. D., Thomson, J. A., & Coon, J. J. (2009). Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 995–1000.
37. Hutchins, A. P., & Robson, P. (2009). Unraveling the human embryonic stem cell phosphoproteome. *Cell Stem Cell, 5*, 126–128.
38. Chen, G., Gulbranson, D. R., Hou, Z., et al. (2011). Chemically defined conditions for human iPSC derivation and culture. *Nature Methods, 8*, 424–429.
39. Geiger, T., Cox, J., & Mann, M. (2010). Proteomic changes resulting from gene copy number variations in cancer cells. PLoS Genetics, 6, pii: e1001090.
40. Walther, D. M., & Mann, M. (2011). Accurate quantification of more than 4000 mouse tissue proteins reveals minimal proteome changes during aging. *Molecular & Cellular Proteomics: MCP, 10*, M110 004523.
41. Lundberg, E., Fagerberg, L., Klevebring, D., et al. (2010). Defining the transcriptome and proteome in three functionally different human cell lines. *Molecular Systems Biology, 6*, 450.
42. Huttlin, E. L., Jedrychowski, M. P., Elias, J. E., et al. (2010). A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell, 143*, 1174–1189.
43. Esteve, P. O., Chang, Y., Samaranayake, M., et al. (2011). A methylation and phosphorylation switch between an adjacent lysine and serine determines human DNMT1 stability. *Nature Structural and Molecular Biology, 18*, 42–48.
44. Guo, B., & Sharrocks, A. D. (2009). Extracellular signal-regulated kinase mitogen-activated protein kinase signaling initiates a dynamic interplay between sumoylation and ubiquitination to regulate the activity of the transcriptional activator PEA3. *Molecular and Cellular Biology, 29*, 3204–3218.
45. Simboeck, E., Sawicka, A., Zupkovitz, G., et al. (2010). A phosphorylation switch regulates the transcriptional activation of cell cycle regulator p21 by histone deacetylase inhibitors. *The Journal of Biological Chemistry, 285*, 41062–41073.
46. Grimsrud, P. A., Swaney, D. L., Wenger, C. D., Beauchene, N. A., & Coon, J. J. (2010). Phosphoproteomics for the masses. *ACS Chemical Biology, 5*, 105–119.
47. Olsen, J. V., Vermeulen, M., Santamaria, A., et al. (2010). Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Science Signaling, 3*, ra3.
48. Bodenmiller, B., Wanka, S., Kraft, C., et al. (2010). Phosphoproteomic analysis reveals interconnected system-wide responses to perturbations of kinases and phosphatases in yeast. *Science Signaling, 3*, rs4.
49. Hilger, M., Bonaldi, T., Gnad, F., & Mann, M. (2009). Systems-wide analysis of a phosphatase knock-down by quantitative proteomics and phosphoproteomics. *Molecular & cellular proteomics: MCP, 8*, 1908–1920.
50. Brill, L. M., Salomon, A. R., Ficarro, S. B., Mukherji, M., Stettler-Gill, M., & Peters, E. C. (2004). Robust phosphoproteomic profiling of tyrosine phosphorylation sites from human T cells using immobilized metal affinity chromatography and tandem mass spectrometry. *Analytical Chemistry, 76*, 2763–2772.
51. Collier, T. S., Sarkar, P., Franck, W. L., Rao, B. M., Dean, R. A., & Muddiman, D. C. (2010) Direct comparison of stable isotope labeling by amino acids in cell culture and spectral counting for quantitative proteomics. *Analytical Chemistry, 82*, 8696–8702.
52. Ficarro, S. B., McCleland, M. L., Stukenberg, P. T., et al. (2002). Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nature Biotechnology, 20*, 301–305.
53. Breuker, K., Jin, M., Han, X., Jiang, H., & McLafferty, F. W. (2008). Top-down identification and characterization of biomo-

lecules by mass spectrometry. *Journal of the American Society for Mass Spectrometry, 19*, 1045–1053.
54. Collier, T. S., Sarkar, P., Rao, B., & Muddiman, D. C. (2010) Quantitative top-down proteomics of SILAC labeled human embryonic stem cells. *Journal of the American Society for Mass Spectrometry, 21*, 879–889.
55. Durbin, K. R., Tran, J. C., Zamdborg, L., et al. (2010) Intact mass detection, interpretation, and visualization to automate Top-Down proteomics on a large scale. *Proteomics, 10*, 3589–3597.
56. Bodenmiller, B., Mueller, L. N., Mueller, M., Domon, B., & Aebersold, R. (2007). Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nature Methods, 4*, 231–237.
57. Ficarro, S. B., Adelmant, G., Tomar, M. N., Zhang, Y., Cheng, V. J., & Marto, J. A. (2009). Magnetic bead processor for rapid evaluation and optimization of parameters for phosphopeptide enrichment. *Analytical Chemistry, 81*, 4566–4575.
58. Ficarro, S. B., Parikh, J. R., Blank, N. C., & Marto, J. A. (2008). Niobium(V) oxide (Nb2O5): application to phosphoproteomics. *Analytical Chemistry, 80*, 4606–4613.
59. Nguyen, V., Cao, L., Lin, J. T., et al. (2009). A new approach for quantitative phosphoproteomic dissection of signaling pathways applied to T cell receptor activation. *Molecular & Cellular Proteomics, 8*, 2418–2431.
60. Hou, J., Cui, Z., Xie, Z., et al. (2010). Phosphoproteome analysis of rat L6 myotubes using reversed-phase C18 prefractionation and titanium dioxide enrichment. *Journal of Proteome Research, 9*, 777–788.
61. Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P., & Jorgensen, T. J. (2005). Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Molecular & Cellular Proteomics: MCP, 4*, 873–886.
62. Pinkse, M. W., Mohammed, S., Gouw, J. W., van Breukelen, B., Vos, H. R., & Heck, A. J. (2008). Highly robust, automated, and sensitive online TiO2-based phosphoproteomics applied to study endogenous phosphorylation in Drosophila melanogaster. *Journal of Proteome Research, 7*, 687–697.
63. Iliuk, A. B., Martin, V. A., Alicie, B. M., Geahlen, R. L., & Tao, W. A. (2010). In-depth analyses of kinase-dependent tyrosine phosphoproteomes based on metal ion-functionalized soluble nanopolymers. *Molecular & Cellular Proteomics: MCP, 9*, 2162–2172.
64. Tao, W. A., Wollscheid, B., O'Brien, R., et al. (2005). Quantitative phosphoproteome analysis using a dendrimer conjugation chemistry and tandem mass spectrometry. *Nature Methods, 2*, 591–598.
65. Choudhary, C., & Mann, M. (2010). Decoding signalling networks by mass spectrometry-based proteomics. *Nature Reviews. Molecular Cell Biology, 11*, 427–439.
66. Thingholm, T. E., Larsen, M. R., Ingrell, C. R., Kassem, M., & Jensen, O. N. (2008). TiO(2)-based phosphoproteomic analysis of the plasma membrane and the effects of phosphatase inhibitor treatment. *Journal of Proteome Research, 7*, 3304–3313.
67. Beausoleil, S. A., Jedrychowski, M., Schwartz, D., et al. (2004). Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proceedings of the National Academy of Sciences of the United States of America, 101*, 12130–12135.
68. McNulty, D. E., & Annan, R. S. (2008). Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. *Molecular & Cellular Proteomics, 7*, 971–980.
69. Saleem, R. A., Rogers, R. S., Ratushny, A. V., et al. (2010). Integrated phosphoproteomics analysis of a signaling network governing nutrient response and peroxisome induction. *Molecular & Cellular Proteomics, 9*, 2076–2088.
70. Posewitz, M. C., & Tempst, P. (1999). Immobilized gallium (III) affinity chromatography of phosphopeptides. *Analytical Chemistry, 71*, 2883–2892.
71. Villen, J., & Gygi, S. P. (2008). The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nature Protocols, 3*, 1630–1638.
72. Brill, L. M., Motamedchaboki, K., Wu, S., & Wolf, D. A. (2009). Comprehensive proteomic analysis of Schizosaccharomyces pombe by two-dimensional HPLC-tandem mass spectrometry. *Methods, 48*, 311–319.
73. Olsen, J. V., Blagoev, B., Gnad, F., et al. (2006). Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell, 127*, 635–648.
74. Nagaraj, N., D'Souza, R. C., Cox, J., Olsen, J. V., & Mann, M. (2010). Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation. *Journal of Proteome Research, 9*, 6786–6794.
75. Xia, Q., Lee, M. V., Rose, C. M., et al. (2011) Characterization and diagnostic value of amino Acid side chain neutral losses following electron-transfer dissociation. *Journal of the American Society for Mass Spectrometry, 22*, 255–264.
76. Eng, J., McCormack, A., & Yates, J. R. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry, 5*, 976–989.
77. Huang, Y., Triscari, J. M., Tseng, G. C., et al. (2005). Statistical characterization of the charge state and residue dependence of low-energy CID peptide dissociation patterns. *Analytical Chemistry, 77*, 5800–5813.
78. Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., & Hunt, D. F. (2004). Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America, 101*, 9528–9533.
79. Swaney, D. L., McAlister, G. C., & Coon, J. J. (2008). Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nature Methods, 5*, 959–964.
80. Baker, P. R., Medzihradszky, K. F., & Chalkley, R. J. (2010) Improving software performance for peptide electron transfer dissociation data analysis by implementation of charge state- and sequence-dependent scoring. *Molecular & Cellular Proteomics, 9*, 1795–1803.
81. Good, D. M., Wenger, C. D., McAlister, G. C., Bai, D. L., Hunt, D. F., & Coon, J. J. (2009). Post-acquisition ETD spectral processing for increased peptide identifications. *Journal of the American Society for Mass Spectrometry, 20*, 1435–1440.
82. Sun, R. X., Dong, M. Q., Song, C. Q., et al. (2010). Improved peptide identification for proteomic analysis based on comprehensive characterization of electron transfer dissociation spectra. *Journal of Proteome Research, 9*, 6354–6367.
83. Palumbo, A. M., & Reid, G. E. (2008). Evaluation of gas-phase rearrangement and competing fragmentation reactions on protein phosphorylation site assignment using collision induced dissociation-MS/MS and MS3. *Analytical Chemistry, 80*, 9735–9747.
84. Mischerikow, N., Altelaar, A. F., Navarro, J. D., Mohammed, S., & Heck, A. J. (2010). Comparative assessment of site assignments in CID and electron transfer dissociation spectra of phosphopeptides discloses limited relocation of phosphate groups. *Molecular & Cellular Proteomics, 9*, 2140–2148.
85. Aguiar, M., Haas, W., Beausoleil, S. A., Rush, J., & Gygi, S. P. (2010). Gas-phase rearrangements do not affect site localization reliability in phosphoproteomics data sets. *Journal of Proteome Research, 9*, 3103–3107.

86. Kim, S., Mischerikow, N., Bandeira, N., et al. (2010). The generating function of CID, ETD, and CID/ETD pairs of tandem mass spectra: applications to database search. *Molecular & Cellular Proteomics: MCP, 9*, 2840–2852.
87. Lee, K. A., Farnsworth, C., Yu, W., & Bonilla, L. E. (2011). 24-hour lock mass protection. *Journal of Proteome Research, 10*, 880–885.
88. Olsen, J. V., de Godoy, L. M., Li, G., et al. (2005). Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Molecular & Cellular Proteomics: MCP, 4*, 2010–2021.
89. Elias, J. E., & Gygi, S. P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods, 4*, 207–214.
90. Keller, A., Nesvizhskii, A. I., Kolker, E., & Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Analytical Chemistry, 74*, 5383–5392.
91. Nesvizhskii, A. I., Keller, A., Kolker, E., & Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. *Analytical Chemistry, 75*, 4646–4658.
92. White, F. M. (2011). The potential cost of high-throughput proteomics. *Science Signal, 4*, pe8.
93. Lu, B., Ruse, C., Xu, T., Park, S. K., & Yates, J., 3rd. (2007). Automatic validation of phosphopeptide identifications from tandem mass spectra. *Analytical Chemistry, 79*, 1301–1310.
94. O'Brien, R. N., Shen, Z., Tachikawa, K., Lee, P. A., & Briggs, S. P. (2010). Quantitative proteome analysis of pluripotent cells by iTRAQ mass tagging reveals post-transcriptional regulation of proteins required for ES cell self-renewal. *Molecular & Cellular Proteomics, 9*, 2238–2251.
95. Ong, S. E., Blagoev, B., Kratchmarova, I., et al. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics, 1*, 376–386.
96. Liu, H., Sadygov, R. G., & Yates, J. R., 3rd. (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Analytical Chemistry, 76*, 4193–4201.
97. Park, S. K., Venable, J. D., Xu, T., & Yates, J. R., 3rd. (2008). A quantitative analysis software tool for mass spectrometry-based proteomics. *Nature Methods, 5*, 319–322.
98. Zhang, Y., Wen, Z., Washburn, M. P., & Florens, L. (2010). Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. *Analytical Chemistry, 82*, 2272–2281.
99. Zhou, J. Y., Schepmoes, A. A., Zhang, X., et al. (2010). Improved LC-MS/MS spectral counting statistics by recovering low-scoring spectra matched to confidently identified peptide sequences. *Journal of Proteome Research, 9*, 5698–5704.
100. Fermin, D., Basrur, V., Yocum, A. K., Nesvizhskii, A. I. (2011). Abacus: a computational tool for extracting and preprocessing spectral count data for label-free quantitative proteomic analysis. *Proteomics, 11*, 1340–1345.
101. Ostrowski, J., & Wyrwicz, L. S. (2009). Integrating genomics, proteomics and bioinformatics in translational studies of molecular medicine. *Expert Review of Molecular Diagnostics, 9*, 623–630.
102. Kumar, C., & Mann, M. (2009). Bioinformatics analysis of mass spectrometry-based proteomics data sets. *FEBS Letters, 583*, 1703–1712.
103. Nikolsky, Y., Kirillov, E., Zuev, R., Rakhmatulin, E., & Nikolskaya, T. (2009). Functional analysis of OMICs data and small molecule compounds in an integrated "knowledge-based" platform. *Methods in Molecular Biology, 563*, 177–196.
104. Payne, S. H., Yau, M., Smolka, M. B., Tanner, S., Zhou, H., & Bafna, V. (2008). Phosphorylation-specific MS/MS scoring for rapid and accurate phosphoproteome analysis. *Journal of Proteome Research, 7*, 3373–3381.
105. Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology, 26*, 1367–1372.
106. Choi, H., Fermin, D., & Nesvizhskii, A. I. (2008). Significance analysis of spectral count data in label-free shotgun proteomics. *Molecular & Cellular Proteomics, 7*, 2373–2385.
107. Bodenmiller, B., Malmstrom, J., Gerrits, B., et al. (2007). PhosphoPep–a phosphoproteome resource for systems biology research in Drosophila Kc167 cells. *Molecular Systems Biology, 3*, 139.
108. Gnad, F., Ren, S., Cox, J., et al. (2007). PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biology, 8*, R250.
109. Hornbeck, P. V., Chabra, I., Kornhauser, J. M., Skrzypek, E., & Zhang, B. (2004). PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics, 4*, 1551–1561.
110. Jones, P., Cote, R. G., Martens, L., et al. (2006). PRIDE: a public repository of protein and peptide identifications for the proteomics community. *Nucleic Acids Research, 34*, D659–D663.
111. Linding, R., Jensen, L. J., Ostheimer, G. J., et al. (2007). Systematic discovery of in vivo phosphorylation networks. *Cell, 129*, 1415–1426.
112. Kruger, M., Kratchmarova, I., Blagoev, B., Tseng, Y. H., Kahn, C. R., & Mann, M. (2008). Dissection of the insulin signaling pathway via quantitative phosphoproteomics. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 2451–2456.
113. Jovanovic, M., Reiter, L., Picotti, P., et al. (2010). A quantitative targeted proteomics approach to validate predicted microRNA targets in C. elegans. *Nature Methods, 7*, 837–842.
114. Ding, V. M., Boersema, P. J., Foong, L. Y., et al. (2011). Tyrosine phosphorylation profiling in fgf-2 stimulated human embryonic stem cells. *PloS One, 6*, e17538.
115. Gual, P., Le Marchand-Brustel, Y., & Tanti, J. F. (2005). Positive and negative regulation of insulin signaling through IRS-1 phosphorylation. *Biochimie, 87*, 99–109.
116. Liu, Y. F., Herschkovitz, A., Boura-Halfon, S., et al. (2004). Serine phosphorylation proximal to its phosphotyrosine binding domain inhibits insulin receptor substrate 1 function and promotes insulin resistance. *Molecular and Cellular Biology, 24*, 9668–9681.
117. Old, W. M., Shabb, J. B., Houel, S., et al. (2009). Functional proteomics identifies targets of phosphorylation by B-Raf signaling in melanoma. *Molecular Cell, 34*, 115–131.
118. Mohyeldin, A., Garzon-Muvdi, T., & Quinones-Hinojosa, A. (2010). Oxygen in stem cell biology: a critical component of the stem cell niche. *Cell Stem Cell, 7*, 150–161.
119. Yoshida, Y., Takahashi, K., Okita, K., Ichisaka, T., & Yamanaka, S. (2009). Hypoxia enhances the generation of induced pluripotent stem cells. *Cell Stem Cell, 5*, 237–241.
120. Hudson, C. C., Liu, M., Chiang, G. G., et al. (2002). Regulation of hypoxia-inducible factor 1alpha expression and function by the mammalian target of rapamycin. *Molecular and Cellular Biology, 22*, 7004–7014.
121. Covello, K. L., Kehler, J., Yu, H., et al. (2006). HIF-2alpha regulates Oct-4: effects of hypoxia on stem cell function, embryonic development, and tumor growth. *Genes & Development, 20*, 557–570.
122. Ezashi, T., Das, P., & Roberts, R. M. (2005). Low O2 tensions and the prevention of differentiation of hES cells. *Proceedings of*

the National Academy of Sciences of the United States of America, 102, 4783–4788.
123. Kratchmarova, I., Blagoev, B., Haack-Sorensen, M., Kassem, M., & Mann, M. (2005). Mechanism of divergent growth factor effects in mesenchymal stem cell differentiation. *Science, 308*, 1472–1477.
124. Mukherji, M., Brill, L. M., Ficarro, S. B., Hampton, G. M., & Schultz, P. G. (2006). A phosphoproteomic analysis of the ErbB2 receptor tyrosine kinase signaling pathways. *Biochemistry, 45*, 15529–15540.
125. Schwanhausser, B., Busse, D., Li, N., et al. (2011). Global quantification of mammalian gene expression control. *Nature, 473*, 337–342.
126. Bernstein, B. E., Mikkelsen, T. S., Xie, X., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell, 125*, 315–326.
127. Beausoleil, S. A., Villen, J., Gerber, S. A., Rush, J., & Gygi, S. P. (2006). A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nature Biotechnology, 24*, 1285–1292.
128. Van Hoof, D., Passier, R., Ward-Van Oostwaard, D., et al. (2006). A quest for human and mouse embryonic stem cell-specific proteins. *Molecular & Cellular Proteomics, 5*, 1261–1273.
129. Alvarez-Manilla, G., Warren, N. L., Atwood, J., 3rd, Orlando, R., Dalton, S., & Pierce, M. (2010). Glycoproteomic analysis of embryonic stem cells: identification of potential glycobiomarkers using lectin affinity chromatography of glycopeptides. *Journal of Proteome Research, 9*, 2062–2075.
130. Zielinska, D. F., Gnad, F., Wisniewski, J. R., Mann, M. (2010). Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell, 141*, 897–907.
131. Sultan, M., Schulz, M. H., Richard, H., et al. (2008). A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science, 321*, 956–960.
132. Macek, B., Mann, M., & Olsen, J. V. (2009). Global and site-specific quantitative phosphoproteomics: principles and applications. *Annual Review of Pharmacology and Toxicology, 49*, 199–221.
133. Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., & Hirakawa, M. (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Research, 38*, D355–D360.
134. Jensen, L. J., Kuhn, M., Stark, M., et al. (2009). STRING 8–a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Research, 37*, D412–D416.
135. Miller, M. L., & Blom, N. (2009). Kinase-specific prediction of protein phosphorylation sites. *Methods in Molecular Biology, 527*, 299–310. x.
136. Dinkel, H., Chica, C., Via, A., et al. (2011). Phospho.ELM: a database of phosphorylation sites–update 2011. *Nucleic Acids Research, 39*, D261–D267.
137. Dephoure, N., Zhou, C., Villen, J., et al. (2008). A quantitative atlas of mitotic phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 10762–10767.