Title
Comparative posttranslational proteomic approaches for understanding development and degeneration

Permalink
https://escholarship.org/uc/item/3zx520wk

Author
Crain, Andrew M.

Publication Date
2012

Peer reviewed|Thesis/dissertation
University of California, San Diego

Comparative Posttranslational proteomic approaches for understanding development and degeneration

A Dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy

in

Biomedical Sciences

by

Andrew M. Crain

Committee in Charge;
Professor Evan Y. Snyder, Chair
Professor Bruce A. Hamilton, Co-Chair
Professor Wendy Campana
Professor Pieter Dorrestein
Professor Joseph Gleeson
Professor Binhai Zheng

2012
The Dissertation of Andrew Michael Crain is approved and is acceptable in quality and form for publication on microfilm and electronically:

__________________________________________

__________________________________________

__________________________________________

__________________________________________

__________________________________________

Co-Chair

__________________________________________

Chair

University of California, San Diego

2012
# Table of Contents

Signature Page.......................................................................................................................... iii

Table of Contents...................................................................................................................... iv

List of Abbreviations................................................................................................................ vi

List of Figures and Tables.......................................................................................................... vii

Acknowledgements................................................................................................................... x

Curriculum Vitae....................................................................................................................... xi

Abstract of the dissertation....................................................................................................... xiii

Introduction to the dissertation................................................................................................. 1

Chapter 1: The Purkinje Cell Degeneration (pcd) mouse......................................................... 3

  Preface............................................................................................................................... 3

  Introduction to the chapter................................................................................................. 3

  1.1 In vitro analysis of Purkinje Cell degeneration and dendritic arborization in the pcd mouse............................................................................................................ 5

  1.2 Comparative proteomics of mitochondrial fractions identifies a predicted HMG box containing protein.......................................................... 9

Methods.................................................................................................................................. 17

Figures.................................................................................................................................. 29

Chapter 2: Phosphoproteomics approaches to understanding hES cells......................... 43

  Preface............................................................................................................................... 43
2.1 Phosphoproteomic analysis: An emerging role in deciphering cellular signaling in human embryonic stem cells and their differentiated derivatives .................................................................43

2.2 Phosphoproteomic analysis of human embryonic stem cells........77

Methods..................................................................................................................................96

Figures......................................................................................................................................99

References................................................................................................................................104
List of Abbreviations

PC-Purkinje Cell

pcd-Purkinje cell degeneration

Nna1- nervous system nuclear protein induced by axotomy

CP- carboxypeptidase

CCP- cytosolic carboxypeptidase

mtDNA- mitochondrial deoxyribonucleic acid

8-oxodG - 8’ oxo-deoxyguanosine

NSE-neuron specific enolase

2D-DiGE-two dimensional differential gel electrophoresis

HMGB1-high mobility group box 1

MS-mass spectrometry

ChAT-choline acetyltransferase

AGTPBP1 – ATP/GTP Binding Protein 1
List of Figures and Tables

Figure 1. Similar extent of degeneration of spinal cord motor neurons after sciatic axotomy in pcd3J+/+ and pcd3J-- spinal cord.................................................................29

Figure 2. Degeneration and dysmorphic dendritic profiles in PC’s of cerebellar organotypic slice cultures.................................................................30

Figure 3. Calbindin immunofluorescence image of PC from pcd3J+/+ cerebellar slice culture at DIV14 shows abnormal dendritic morphology.......................31

Figure 4. Novel putative phosphorylation of human Nna1 at T1127 and/or S1128 identified using SILAC phosphoproteomics........................................32

Figure 5. Comparative Mitochondria proteome analysis workflow......................33

Figure 6. 2D-DiGE comparative proteomics of P14 pcd3J+/+ and pcd3J-- cerebellar mitochondria identifies a predicted HMG box containing protein with differential migration and focusing patterns..............................................34

Figure 7. Spot picking/MS/MS analysis reveals two differentially present spots to be identical proteins.................................................................35

Figure 8. Analysis of spot shifts shows pcd3J-- LOC100045876 slightly increased in size and acidic residues.................................................................36

Figure 9. LOC100045876 homology to HMGB1 and domain map....................37

Figure 10. LOC100045876 and Nna1 adult tissue gene expression profile......38
Figure 11. LOC100045876 co-localizes with the mitochondrial marker OxPhos.

Figure 12. Increase in 8oxo-deoxyguanosine immunoreactivity in pcd P14 cerebellum.

Figure 13. Nuclear levels of 8-oxodG are unchanged in the pcd3J/1 mouse.

Figure 14. Evaluation of protein oxidation shows similar extent of derivatized carbonyl levels in subcellular compartments of pcd cerebella.

Figure 15. LOC1000458876 interacts specifically with Nna1.

Figure 16. Undifferentiated hESCs Expressed Markers of Pluripotency, whereas the Markers Were Downregulated upon Differentiation.

Figure 17. Number of Protein Phosphorylation Sites and Phosphoproteins Identified in hESCs and Their Differentiated Derivatives, Prominence of Tyrosine Phosphorylation.

Figure 18. Protein Kinase Inhibitors Resulted in Differentiation of hESCs.

Figure 19. PDGF and a Subthreshold Concentration of bFGF Sustained Long-Term Culture of hESCs.

Figure 20. Semi-Quantitative Western Blot Analysis, using Antibodies that Recognize Specific Protein Phosphorylation Sites, Supported Phosphoproteomic Results using MDLC-MS/MS.
Table 1. Selected concepts and terminology of relevance to (phospho)proteomics

.................................................................75

Table 2. Bioinformatic resources useful to (phospho)proteomics....................76
Acknowledgements

A supportive environment was especially important for the completion of this dissertation due to the fact that the bulk of the thesis work was initiated with no pre-existing data, vectors, animals, established specialized techniques, or procedures. First I would like to acknowledge the outstanding support by Dr. Evan Snyder, MD, PhD during my graduate career. I would also like to acknowledge colleagues whose scientific interaction enhanced and contributed to the development of this work. Intense discussions of data, ideas and concepts with Drs. Brian Tobe, Larry Brill and Gustavo Gutierrez, along with collaborations on various projects, were instrumental in my training and progression as a scientist.

Curriculum Vitae

Education:
San Francisco State University, B.S. Exercise Physiology 2001
California State University, Northridge B.S. Cellular and Molecular Biology 2004
California State University, Northridge M.S. Exercise Physiology 2004
University of California, San Diego, Ph.D. Biomedical Sciences 2012

Publications:

Reviews:
Crain AM, Gutierrez G, Snyder EY. Cell cycle proteins in the post mitotic neuron and their role in cytoskeletal elements. In preparation


Wakeman DR, Crain AM, Snyder EY. Large animal models are critical for rationally advancing regenerative therapies. Regenerative Medicine. 2006 July;1(4), 405-413.

Research Papers:
Crain AM, Gutierrez G, Snyder EY. Mitochondrial oxidative damage and dysregulation of a novel HMG box protein in the pcd mouse. submitted


**Awards & Honors:**

2007-2008 Howard Hughes Medical Institute Med-into-Grad Fellowship
2003-2004 Graduate Equity Fellowship Award, CSUN
2003-2004 Thesis Award, CSUN
2003-2004 Graduate Tuition Grant, CSUN
1999-2001 Deans High Honors, SFSU
ABSTRACT OF THE DISSERTATION

Comparative Posttranslational proteomic approaches for understanding development and degeneration

by

Andrew M. Crain

Doctor of Philosophy in Biomedical Sciences

University of California, San Diego, 2012

Evan Snyder, MD, Chair
Bruce Hamilton, Co-Chair

Cellular function is largely determined by posttranslational modification (PTM) of proteins that govern protein-protein interactions, protein-nucleic acid interaction, transcription, enzymatic activities, etc. Yet, identifying and determining the role of various PTM’s in development and degeneration remains a daunting task.

Among the many PTM’s, glutamylation/deglutamylation was identified more than 30 years ago and involves the addition of glutamic acid residues branching from an encoded glutamic acid residue in an acidic region (deglutamylation is the removal of C-terminal primary sequence Glu residues or PTM Glu). Alpha tubulin was the first protein to be identified as being posttranslationally glutamylated which has since been found to regulate microtubule stability and ciliary function. However the identification of other proteins undergoing deglutamylation remains poorly defined. Using a mouse model of neurodegeneration that is deficient for a single deglutamylase, we utilized a direct comprehensive comparative proteomics approach and identified...
a predicted HMG box containing protein at the mitochondrion and show specific interaction of this protein with the deglutamylase. Furthermore, the affected cell type in this model shows an increase in oxidative damage of mtDNA indicating a potential role for de glutamylation in maintaining mtDNA integrity.

Unlike glutamylation, phosphorylation is the most well studied PTM, however, the roles of most predicted phosphosites across the proteome have yet to be determined. Recent technological advances have dramatically improved phosphoproteomic approaches, yet the identification of critical phosphosites involved in maintaining pluripotent human ES cells has largely been underexplored. The developmental pathways controlling the maintenance of the undifferentiated state and the initiation of the earliest neur ectodermal lineage were investigated using comparative total cellular phosphoproteomics. As one of the first reports of comprehensive phosphoproteomic profiling and analysis of hESCs, we have revealed a previously unknown role for JNK activity in maintaining the pluripotent state.
Introduction to the dissertation

Since the initial indication of phosphorylation of amino acids over 100 years ago, the list of posttranslational modifications has grown to include glutamylation, sumoylation, and many others. The biological significance of these modifications continues to be heavily investigated in a wide range of contexts. Many signaling pathways incorporate protein phosphorylation as the means for executing a wide range of cellular events in both development and physiology. These kinase pathways play important roles in the response to infection and inflammation yet abnormal signaling has itself been found to result in pathology such as is the case with type II diabetes mellitus. From the earliest pluripotent cell to a fully differentiated post mitotic neuron, the role of posttranslational modifications remains an area of great interest for better understanding processes controlling development, normal physiology, and disease.

The use of comparative proteomic strategies to elucidate the posttranslational modifications contributing to a normal or disease process has recently gained momentum through cutting edge technological advances and represents a powerful approach for interrogating complex biology. By focusing comparisons specifically on the posttranslational modification of proteins one can quickly identify differential pathways or processes giving considerable insight and direction into the key biological events involved in the system studied. Whereas, in some cases, gene expression analysis is sometimes referred to as fishing expeditions, the focusing of the search specifically to posttranslational
modifications may be more analogous to a spearfishing endeavor in that the process specifically targets a result within a particular category. This dissertation focused on using comparative proteomic approaches to identify important pathways in development (kinase pathways involved in the maintenance of the pluripotent state of human embryonic stem cells and their neurectodermal derivatives) and degeneration (the effects of the lack of a known deglutamylase in neurodegeneration).
Chapter 1: The Purkinje Cell Degeneration (pcd) Mouse

Preface

Chapter 1.1 contains the initial work on the pcd mouse by the dissertation author which was shared in collaboration with Jianxue Li and Richard Sidman in 2007 and formed part of the experimental approaches and methods used in *Nna1 mediates Purkinje cell dendritic development via lysyl oxidase propeptide and NF-κB signaling* published in Neuron Oct 6th 2010 in which the dissertation author is a co-author.

Chapter 1.2 contains the main body of work designed and executed by the dissertation author. A portion of this chapter is being condensed into a brief communication for submission to Journal of Neuroscience with the dissertation author as the first author.

Introduction to the Chapter

The Purkinje cell degeneration (pcd) mouse was originally identified and described more than 30 years ago as a spontaneous autosomal recessive mutation in the Jackson Laboratory mouse colony. The identifying phenotype, ataxia, was observed in animals at approximately 3 weeks of age and despite substantial investigative efforts, the causative genetic mutation was only recently discovered. In the last three decades several other pcd strains have been identified and numerous investigations have shed light on the pathology of Purkinje cells (PC) in these mice yet a clear molecular mechanism causing the PC degeneration has yet to be identified. The mutation in the pcd2J, pcd3J, and
pcd5J^4 strains was found to occur within the AGTPBP1 gene locus and a study using chimeric mice suggested that the death of PC's in pcd is cell autonomous^5. The AGTPBP1 gene was originally discovered in spinal cord motor neurons and dorsal root ganglion neurons in a differential screen from sciatic injured mice. In both development and in the adult it was found to be expressed widely in post mitotic neurons in the cortex and cerebellum^6. Interestingly, expression was also dramatically increased in response to sciatic injury and returned to normal levels after motor neuron regeneration. Therefore, when it was determined as the gene responsible for the degeneration of PC's, it became the first neuroregeneration associated gene that when disrupted, causes neurodegeneration.

The mouse AGTPBP1 gene encodes a predicted zinc carboxypeptidase (CP), commonly referred to as Nna1 (nervous system nuclear protein induced by axotomy) or CCP1 (cytosolic carboxypeptidase 1) based on sequence homology of the putative enzymatic site with previously known CP's. A conserved glutamic acid residue in this region, E1094, is responsible for the coordination of a water molecule for nucleophilic attack on the carbonyl of the peptide adjacent to the one ultimately removed by the CP activity. A zinc ion at the catalytic site functions to stabilize the intermediate during the reaction. Investigation into the role of the CP in the pathology of the pcd mouse revealed that the ability to rescue the pcd phenotype with wild type Nna1 is dependent on an intact CP domain^7 further suggesting that the function of Nna1 is that of a true CP and that deficiency in the Nna1 CP activity results in the degeneration of PC's. The
characteristic feature of the M14 family of CP’s, to which Nna1 has been recently added, is the relatively open catalytic site suggesting that Nna1 may have a rather wide range of substrates. The cytosolic location of Nna1 is different than other know CP’s such as CPA & CPE, CPN and CPD which are specifically localized to the secretory granules, plasma membrane and the transgolgi network, respectively. Therefore, due to the lack of known CP’s containing a similar feature profile as Nna1, it has been difficult to reasonably predict the biological function.

1.1 In vitro analysis of Purkinje Cell degeneration and dendritic arborization in the pcd mouse

Introduction

The identification of Nna1 in spinal motor neurons, and the specific upregulation in response to sciatic injury, suggested a possible role in regeneration of this cell type. Upon regeneration of motor neurons after sciatic crush injury, Nna1 transcripts return to pre-injury levels. Interestingly, spinal motor neurons do not undergo degeneration in the pcd mouse suggesting that Nna1 is not required for normal survival of this cell type. However, it was not clear whether during the regeneration process, Nna1 may be playing a functional role and loss of Nna1 could result in an increase in degeneration in response to such injury. Surprisingly, experimental investigation into the potential increased
sensitivity to axonal injury in the pcd mouse has not been reported. Additionally, the establishment of a reliable in vitro experimental approach for studying the PC degeneration and Nna1 function had not been established. Therefore, the two initial goals were to determine whether 1) lack of Nna1 results in increased susceptibility to motor neuron degeneration following sciatic injury and 2) organotypic slice culture faithfully maintains the in vivo degeneration of PC's and is a viable platform for experimentation.

**Results**

*Extent of sciatic axotomy induced spinal cord motor neuron degeneration is unchanged in the pcd mouse.* Motor neuron axonal injury in the pcd mouse had not been reported and we hypothesized that the absence of functional Nna1 might result in an increase in motor neuron degeneration in response to sciatic axotomy. To determine whether axonal injury induced motor neuron degeneration is altered in pcd3J^-/- mice, P7 to P8 pups were subjected to sciatic axotomy and allowed to recover for 28 days. Spinal cord samples were sectioned and anti-NSE/anti-ChAT immunohistochemistry was used to compare motor neuron survival in the dorsolateral quadrants. ChAT positive neuron numbers were similar in both WT and mutant samples indicating that the sensitivity to axotomy induced motor neuron degeneration was similar (*Figure 1*).

*Maintenance of Purkinje Cell Degeneration phenotype in organotypic cerebellar slice culture.* An in vitro system for observing the degeneration of PC's
in the pcd mouse was not previously established. Dissociated cerebellar cell preparations result in very low PC numbers and combined with the relatively low frequency of obtaining mutant mice, we chose to evaluate whether organotypic slice culture of cerebellum might be a feasible in vitro platform for both experimentation and observation of the PC degeneration. This method provides a consistent and reliable system that maintains tissue architecture. Postnatal day 7 cerebellar slice cultures maintain the cell-cell contact and PC survival is maintained for 28 days and beyond. Although the organotypic slice culture conditions could alter the dynamics/influence the degeneration of PC’s, we hypothesized that the preparation of cerebellar slices, which results in the axotomy of nearly all but the most sagitally oriented PC soma and axons in slices containing deep cerebellar nuclei, would have little impact on the PC degeneration. Since the degeneration of PC’s in vivo in the pcd3J<sup>−/−</sup> mouse occur at approximately 21 days, we evaluated the presence of PC’s in slice cultures after 14 or 21 DIV using immunohistochemistry. At both time points there was a dramatic reduction of PC’s in the pcd3J<sup>−/−</sup> slices when compared to pcd3J<sup>+/+</sup> slices (Figure 2).

*Altered dendritic morphology of pcd3J<sup>−/−</sup> PC’s in organotypic slice culture*

An immediate observation in the organotypic slice culture experiments was the altered dendritic morphology of PC’s from mutant samples. At 7 and 14 DIV the PC’s in the pcd3J<sup>−/−</sup> cerebellar slice culture exhibited abnormal dendritic arbors which were not observed in pcd3J<sup>+/+</sup> samples at all time points analyzed (up to 28
DIV) (Figure 3). The PC dendritic tree was bi-directional and under developed with shorter branches similar to what was described in our recent report$^{13}$.

**Conclusion and future directions**

The unaltered axotomy induced spinal motor neuron degeneration in pcd mice suggests the possibility that the role of Nna1 is different in this cell type than in populations of neurons that show degeneration in this model. Strikingly, Chakrabarti et al. found that Nna1 has differential subcellular localization depending upon the type of cell investigated. Nna1 was shown to localize to the outer mitochondrial membrane in samples prepared from rat brain, whereas spinal cord samples did not show this localization. This evidence further supports a prominent, degeneration associated role of Nna1 at the mitochondrion that has yet to be defined.

In vitro experimental accessibility is of great importance when pursuing to understand a complex in vivo phenotype. Organotypic slice culture has been a robust technique, predominantly in the electrophysiology field, that allows a section of tissue to be maintained largely intact and preserves cell integrity. Cerebellar slice culture has been reported for other mouse models such as nr/nr, but it was unknown whether the pcd phenotype could be studied using this system. After 14DIV, slice cultures from pcd3J$^{1/1}$ showed massive loss of PC's as determined via Calbindin$^+/\text{NSE}^-$ immunoreactivity. The time point corresponds to P21 and at this stage few PC's remain in the cerebellum of pcd3J$^{1/1}$ mice,
therefore, the slice culture parallels the degeneration dynamics displayed *in vivo*. Genetic manipulation of slice culture is best achieved via viral means followed by biollistic’s however these strategies are not highly efficient. However, pcd cerebellar slice culture remains a useful experimental approach for use of small molecules and environmental manipulation.

The postnatal development of the dendritic arbor of PC’s is quite extensive and results in a rather elaborate profile set to receive input from stellate, basket, and granule cells as well as neurons from the inferior olivary nucleus. The pcd mouse has long been known as a CNS degeneration model, therefore, when we first observed the dendritic morphology of PC’s in slice culture it was immediately realized that the pcd mouse may also be a model of postnatal neuronal maturation as well. In collaboration with Jianxue Li, we found increased intranuclear localization of lysyl oxidase propeptide causing the attenuated postnatal dendritic development in the pcd mouse.  

1.2 Comparative proteomics of mitochondrial fractions identifies a predicted HMG box containing protein

Introduction

Within the last few years, three independent groups have reported entirely different functional roles for Nna1. Berezniuk et al. have suggested that Nna1
carboxypeptidase activity is required for protein degradation pathways, namely the degradation of small oligopeptides, and that lack of Nna1 results in a decrease in the free pool of amino acids with concomitant increased di- and tripeptides\textsuperscript{14}. An exhaustive in vitro investigation into the role of Nna1 by Rogowski et al. proposed that the carboxypeptidase activity of Nna1 removes C-terminal translational and posttranslationally added glutamic acid residues of proteins and that the degeneration of PC’s is due to cytoskeletal dysfunction resulting from defective deglutamylation of alpha tubulin\textsuperscript{15}. Using both the pcd mouse and a recently described drosophila melanogaster equivalent where the NnaD gene is disrupted with a P element insertion, Chakrabarti et al. revealed a metabolic phenotype in these models. They suggest that Nna1 may play a prominent role in the regulation of glycolysis as well as mitochondrial function wherein Nna1 appears to localize to the outer mitochondrial membrane\textsuperscript{16}. Interestingly, mitochondrial abnormalities in pcd PC’s were noted in some of the earliest reports where ultrastructural analysis revealed rather enlarged and swollen mitochondrial morphologies\textsuperscript{17}.

Despite these recent efforts to determine the molecular aberrations resulting from deficiency of Nna1, mechanisms causing the degeneration have not yet been reported with sufficient evidence to clearly define the role in PC death. The \textit{in vitro} enzymatic evidence for the alpha tubulin deglutamylase activity, when also considering the open catalytic site of Nna1, yields a reasonable likelihood of other substrates more directly tied to the \textit{in vivo}
metabolic observations, particularly the mitochondrial dysfunction. Therefore, we employed a focused mitochondrial proteomic approach to identify pre-degenerative posttranslational defects in the cerebellum of pcd3J<sup>−/−</sup> mice.

**Results**

*Novel putative phosphorylation site(s) identified on hNna1 corresponding to 1159 Threonine/1160 Serine of mNna1.* Global phosphoproteomic analysis of human ES cells and the retinoic acid induced differentiated derivatives was completed as described in Chapter 2.2. Querying this data for human AGTPBP1 revealed a putative phosphorylation site at either T1159 or S1160 in the human ES cell sample. Given the proximity to the carboxypeptidase catalytic site it is possible that phosphorylation in the region may regulate activity. Interestingly, in neurectoderm differentiated hES derivatives, phosphorylation at this site was not identified perhaps indicating negative regulation of Nna1 by phosphorylation at this site. When the corresponding sequence of mouse Nna1 sites, T1127 and S1128, were analyzed with NetPhosK, a list of possible kinases predicted to be capable of phosphorylation at these residues was generated (*Figure 4*).

*Comparative Proteomic analysis of Mitochondria reveals changes in alpha tubulin and a novel HMG box protein.* Recent evidence suggesting a mitochondrial pathology in the pcd mouse led us to investigate whether perturbation of the mitochondrial proteome preceded the PC neurodegeneration. Cerebellar mitochondria isolates were analyzed from 2 pcd3J<sup>−/−</sup> and 2 pcd3J<sup>+/−</sup>
littermates at age P14 (Figure 5) using 2D DiGE to identify alterations in proteins at or within the mitochondrion. Image quantification based comparison of the relative abundance of protein spots yielded few differences above the 1.5 fold threshold (Figure 6). These spots were extracted and subjected to MS/MS for protein identification. Alpha Tubulin was identified in spots 1 and 2 from pcd3J+/− and pcd3J+/+ respectively (Figure 6). The acidic shift of pcd3J+/− alpha tubulin is consistent with recent findings by Rogowski et al. which suggested levels of polyglutamylated alpha tubulin are elevated in pcd3J+/− cerebellum15.

Independent protein identification for spots 4 and 5 revealed both as a predicted protein similar to HMGB1 (LOC100045876). Spots 4 and 5 from the 2D-DiGE gel were paired but almost entirely exclusive to the respective sample. These spots were picked and subjected to MS/MS analysis for protein identification. Spectral counts were queried against the Mascot database of the *mus musculus* proteome. The list of fragments obtained from digestion of each spot is presented in Figure 7. Percent coverage of spots 4 and 5 was 40.4% and 42.9%, respectively, and the confidence intervals were both above 99%.

Qualitative comparison of migration distances for spots 4 and 5 yielded an acidic shift in the isoelectric point of approximately 0.2 pH and a slight increase in molecular weight of the pcd3J+/− sample (Figure 8). These observations reflect similarity to the observed shift in alpha tubulin but more importantly, the predicted acidic shift of LOC10045876 when the C-terminal encoded glutamic acid residue
is removed from the primary sequence is also 0.2 pH (predicted pl's were calculated using the Scripps Protein Calculator).

**LOC100045876 sequence analysis revealed a distinct region between the DNA binding domains and regulatory acidic C-terminus.** Primary sequence analysis showed that the majority of LOC100045876 is identical to HMGB1 with approximately 40 residues downstream of the second HMG box being slightly different. The C-terminal acidic tail was shorter in LOC100045876 but nearly identical in the Glu and Asp sequence (Figure 9).

**LOC100045876 RT-PCR reveals widespread expression pattern in adult tissues.** To examine the tissue distribution of LOC100045876 expression, adult mouse cDNA was used for PCR analysis to determine expression levels relative to GAPDH. LOC100045876 is expressed throughout tissue in the adult mouse (Figure 10) and it is important to note that CCP2-CCP6, recently reported cytosolic carboxypeptidases in the same family as Nna1, also show widespread tissue distribution.

**Increased oxidative damage of Mitochondrial DNA in the pcd mouse.** Tandem HMG box domain containing proteins are known to bind DNA and given that much of the N-terminal sequence of LOC10045876 is identical to the DNA binding protein HMGB1 (Figure 9), we decided to look at basic mitochondrial DNA (mtDNA) parameters. Total DNA was isolated from pcd3J+/− and pcd3J−/− cerebella using a PureGene kit (Qiagen) which isolates both nuclear and mtDNA,
and qPCR was used to determine relative mtDNA levels. No difference in mtDNA was observed (data not shown) so we next sought to evaluate the mtDNA integrity.

Mitochondrial DNA damage was evaluated using a highly specific antibody to 8-oxodeoxyguanosine (monoclonal 8-oxodG, Trevigen), which arises through oxidative damage. Cerebella were isolated from P14 pcd3J\textsuperscript{+/−} and pcd3J\textsuperscript{−/−}, sectioned and probed for 8-oxodG. PC’s from pcd3J\textsuperscript{−/−} showed increased immunoreactivity when compared to pcd3J\textsuperscript{+/−}. The signal appeared to be weakest within the nucleus so to confirm that nuclear DNA damage was not contributing to the signal, an 8-oxodG ELISA was used with total nuclear DNA. The levels of 8-oxodG in the nuclear compartment in pcd3J\textsuperscript{−/−} were unchanged (Figure 13). These data suggest that in PC’s in the pcd3J\textsuperscript{−/−} mouse either 1) the mtDNA is more susceptible to oxidative damage or 2) the mitochondrial matrix environment is more oxidative than pcd3J\textsuperscript{+/−}. If the latter were true, the increased oxidative environment in the mitochondrion might result in protein oxidative damage. To evaluate the level of protein oxidative damage an Oxyblot was used and the subcellular fractions of cerebella isolates analyzed. No difference in protein oxidative damage was noted (Figure 14). These data suggest that the mtDNA specifically, is more prone to oxidative damage in PC’s of the pcd3J\textsuperscript{−/−} mouse.

In-vitro interaction of Nna1 and LOC100045876. Since the shift in pcd3J\textsuperscript{−/−} LOC100048576 resembled that of alpha tubulin, and at least one function of Nna1 has been reported to be deglutamylation of both primary and
posttranslational polyglutamylate chains of alpha tubulin, we sought to determine whether this shift might be due to lack of deglutamylation by Nna1. Protein-protein interaction was investigated using GST-LOC100045876 and S\textsuperscript{35}-Methionine labeled Nna1, and a specific interaction was identified (Figure 15). The GST-LOC100045876 was sufficient to pull down Nna1 and this interaction was specific to Nna1 as this fusion protein did not interact with Cdt1 (a protein involved in nuclear DNA replication). Additionally, Nna1 was not pulled down using GST, GST-VHR, or GST-CDK1. Although, no direct evidence is presented to prove the shift of LOC100045876 in pcd3J\textsuperscript{1/1} mitochondria is due to hyperglutamylation or lack of primary deglutamylation, 1) the highly acidic tail in LOC100045876 and the predicted preference of acidic C-terminal sequences for the catalytic activity of Nna1 and 2) the only known property of Nna1 is it’s deglutamylase activity, suggest that this is the most likely explanation of the shift.

**Conclusion & Future Directions**

Early ultrastructural analysis of PC’s in the pcd mouse indicated the presence of abnormal mitochondria prior to degeneration\textsuperscript{17}. More recently, Charbartaki et al. reported decreased oxidative capacity in pcd5J\textsuperscript{1/1} mitochondria isolated from the cerebellum and similar results were also observed in the NnaD fly\textsuperscript{16}. These data provided the rational for investigating the mitochondrial proteome in the pcd mouse. We chose to utilize a simplified comparative proteomics technique that is sensitive to posttranslation modifications due to the fact that Nna1 is a functional carboxypeptidase and hypothesized that one or
more putative substrates would be revealed in the mutant sample via differential migration. Using 2D-DiGE on mitochondria isolates from a pre-neurodegeneration timepoint, we found differential migration of alpha tubulin and a predicted HMG box containing protein, LOC100045876, containing high identity to HMGB1. Additionally, we provide evidence of interaction with Nna1.

Due to the high sequence identity to HMGB1, particularly in the DNA binding HMG boxes and the acidic tail, it is likely that LOC100045876 is a DNA binding protein. Recently, the acidic tail of HMGB1, specifically, the terminal DDDDE sequence, was reported to regulate the DNA binding activity of dimerized HMGB1. When these residues were replaced with an EEEED sequence, the DNA binding affinity was reduced over 2 fold\textsuperscript{19}. Therefore, it is possible that DNA binding of LOC100045876 may be regulated through Nna1 mediated posttranslational modification the C-terminal acidic tail. This regulation may be disrupted in pcd3J\textsuperscript{1/1} cerebellar mitochondria leading to increased susceptibility to oxidative damage. Strikingly, La Spada’s group has reported an extension of lifespan in NnaD flies that were fed Vitamin E and SOD1 which may indicate the contribution of oxidative damage to development of the phenotype. Whether or not this effect was mediated by mitigating oxidative damage at the level of mtDNA is currently unknown.

The main biological impact of the 8-oxodG:C base pair is a G:C->T:A transversion during replication. Therefore, mitochondrial dysfunction in the pcd mouse may be due to a higher rate of mutation in mtDNA. However, a
confounding factor to this hypothesis is that deletion of the excision repair enzyme Oxoguanine glycosylase (OGG1) in mice has been reported to result in a 20 fold increase in 8-oxodG in mtDNA, yet no phenotype has been observed in this model\textsuperscript{20}. These data have been recently debated due to a report by Trapp et al. in which an increased mutation rate was not found and using an alternate method for determining 8-oxodG they showed that the levels are not different in the OGG1 knockout\textsuperscript{21}. The authors suggest that while base excision repair is disrupted in this model, differences in mtDNA replication may account for the lack of increased 8-oxodG\textsuperscript{21}. Furthermore, Stuart et al. showed that the OGG1 mouse does not display a reduction in mitochondrial function\textsuperscript{22}. Although mitochondrial autophagy (mitophagy) has been observed in the pcd mouse, the possible connection to mtDNA has yet to be determined.

Within the known mitochondrial proteome, very few HMG box containing proteins have been identified and the most intensely investigated has been mitochondrial transcription factor A (TFAM). TFAM, which regulates both transcription and replication of mtDNA, has been used to create a mouse model of some symptoms of Kearns-Sayre Syndrome (KSS), a human mitochondrial disease. Furthermore, some KSS patients display cerebellar ataxia in addition to retinopathy, showing slight similarity to some aspects of the pcd phenotype\textsuperscript{23}.

In summary, our findings indicate oxidative damage of mtDNA is an early aberration PC's of the pcd3J\textsuperscript{1/1} mouse and a predicted HMG box containing protein at the mitochondrion that interacts with Nna1 may lack normal
posttranslational modification. Taken together, these data suggest that the
dysfunction of mitochondrial respiration previously reported, may be related to
mtDNA damage. Future considerations will include a mutagenesis analysis of
LOC100045876 to better define the posttranslation modification (or lack of)
causing the shift observed in the pcd PC mitochondrion. Additionally, determining
the pre-degenerative mtDNA mutational load might reveal the significance of the
8-oxodG damage.

Methods

Animals and Genotyping

Heterozygous breeding pairs of BALB/cByJ-Nna1<sup>pcd-3j</sup> mice were
purchased from Jackson Labs and housed in the onsite vivarium at Sanford Burnham
Medical Research Institute. This particular strain of pcd mice has a
12.2 kb deletion in the Nna1 gene from intron 5 to exon 8. At P1 pups were
marked via toe clipping and toes used for the preparation of genomic DNA for
PCR. Genotyping was carried out using genomic PCR with the following primers;
Intron 5 For 5’-CTGTAGTAGGTCCGCCTGCCTGC-3’, Exon 8 For 5’-
CAGTCTTGTTGAGACCACCG-3’, Exon 8 Rev 5’-GCTGCAGTA
CAAGGTTCATT-3’. Toe samples were dissolved in 40µL 25 mM NaOH/ 2mM
EDTA at 99 C for 3 hours. Samples were then neutralized with 40µL 40 mM Tris-
HCL and briefly vortexed. Samples were cleared by brief centrifugation and 5µL
was used for each genotyping PCR at a final reaction volume of 25µL. Standard Taq polymerase with Thermopol buffer was used with the following cycle parameters; \(95^\circ C\) 2 min, \(95^\circ C\) 30 sec, \(55^\circ C\) 20 sec, \(72^\circ C\) 30 sec, 35 cycles. Reactions were loaded onto a 1% agarose gel and visualized with ethidium bromide staining. The mutant allele results in the production of a PCR product at approximately 490 bp and the wild-type allele results in a 790 bp band.

All protocols for animal use were approved by the animal use committee at the Sanford|Burnham Medical Research Institute.

**Organotypic Cerebellar and Spinal Cord Slice Culture**

Methods used for preparation and culture of cerebellar slice cultures have been described previously\(^2^4\). Postnatal 7 day old pups were ice anesthetized, decapitated and the cerebellum removed using sterile technique. Isolated cerebellum was quickly transferred to a 35mm dish containing ice cooled dissection buffer (Earles Balanced Salt Solution (EBSS) supplemented with 4.5g/L D-glucose) and menninges removed. Unwanted structures such as midbrain and flocculi were removed and the cerebellum was quickly transferred to a McIlwain tissue chopper. 350 µm thick slices were prepared from each cerebellum and were transferred to a new 35mm dish with ice cold dissection buffer. With the aid of a stereo dissecting microscope samples were gently separated into individual slices using microspatulas. Culture inserts were placed into 6 well culture plates containing 1 mL slice culture media (25% Horse Serum,
50% MEM without phenol red, 25% Earles Balanced Salt Solution, 1x Normycin antibiotic (Invivogen) and 2 mM L-Glutamine) and 3 to 4 slices were gently placed on top of each culture insert starting at the vermis and moving laterally so that each well comprised similar regions of cerebellum between animals. Media was changed every other day and cultures were maintained for up to 4 weeks at 37 C, 95% humidity, 5% CO2.

**Sciatic Axotomy**

Sciatic axotomy was performed similar to previously described methods. Pups, at postnatal day 7 to 8, were ice anesthetized and the incision site was disinfected with 70% ethanol. Using a scalpel, a short incision parallel and posterior to the femur was made and the sciatic nerve was exposed. Approximately 1-2mm of sciatic nerve was removed and the incision was sutured, disinfected, and returned to the cage. Twenty-one days after sciatic axotomy, animals were sacrificed and lumbar spinal cord isolated and placed into 4% Paraformaldehyde at 4 C overnight. Samples were cryoprotected in 30% sucrose/PBS overnight and cryosectioned at 18µm for immunohistochemistry described below.

**Immunofluorescence**

Organotypic slice cultures were cut from the insert membrane and washed once with PBS before fixing with 4% paraformaldehyde (PFA) for 30 min at room temperature. Slices were blocked in 3% BSA, 0.3% Triton-X 100, 0.1 % sodium
azide in PBS for 45 min at room temperature. Primary antibodies were used in antibody dilution buffer (1% BSA, .3% Triton-X 100, 0.1 % sodium azide in PBS) and incubated overnight at room temperature with gentle shaking. After washing slices with PBS 3 times for 10 minutes each, secondary antibody was added at a 1:500 dilution in antibody dilution buffer and incubated for 2 hours. After washing with PBS 3 times for a total of 3 hrs slices were then placed on glass slides, mounted with vectashield containing DAPI and coverslipped before imaging using standard fluorescence microscopy.

Spinal cords were extracted from PFA perfused mice and placed in 30% sucrose overnight at 4°C for cryoprotection. Samples were then placed in molds and embedded in OCT, frozen at -20°C and stored prior to cryosectioning. Samples were sectioned at 12 µm at -17°C and transferred to precoated slides. Samples were then probed with antibodies recognizing Cholineacyltransferase and Neuron specific enolase using a Shandon sequenza coverslip system using the same parameters as described above for slice cultures. Grayscale images were used to create pseudo colored composites using NIH Image J.

**RT-PCR & Cloning**

Adult WT BALB/C tissues and C17.2 mouse neural stem cell total RNA were isolated using an RNeasy Mini kit (Qiagen). RNA samples were DNase treated prior to cDNA synthesis. One microgram of total RNA was reverse transcribed using the Superscript Reverse Transcriptase III (Invitrogen) with oligo
dT primers for generation of cDNA. The cDNA was used for PCR amplification of the Nna1 transcript with the following primers; For 5'- GGA AAT GAC AAC CAA AGG TTC TAC -3' and Rev 5'- CTT GTC CCT CTC AAA CTT CCT AC -3'. Phusion polymerase (New England Biolabs), a Pfu polymerase, was used to amplify Nna1 and a standard Taq polymerase was added at the end of cycling to add 3’ adenosine overhangs required for topo TA cloning. The pcDNA4 Hismax Topo kit (Invitrogen) was used to clone Nna1 to yield an N-terminal HA/FLAG tag fusion. LOC100045876 cDNA was ordered from Genscript in the pDream 2.1 vector with an N-terminal FLAG tag for subsequent immunofluorescence and subcloned for GST fusion protein production (described below).

Expression of LOC100045876 across adult tissues was determined via RT-PCR using total mRNA isolated from adult wildtype BALB/C tissues. Adult tissue samples were quickly isolated and snap frozen on dry ice prior to lysis. Total RNA was isolated and cDNA prepared as described above. The following primers and cycling profile were used for evaluation of LOC100045876 expression patterns; pLOCF 5' CAGCCATTGCAGTACATTGAG and pLOCR 5' AAAAACTGGCTAGAACCAC.

**Cell Culture and Transfection**

C17.2 cells were grown as described previously and subjected to transfection using a pDream 2.1-LOC100045876 with Fugene 6 (Roche) according to the manufacturer’s instructions. At 48 hrs post transfection wells
were washed 3 times with PBS and fixed with 4% PFA for 20 min at RT. Level of transgene expression was determined via immunofluorescence using anti-FLAG (Cell Signaling Technologies) at 1:200 dilution. Mitochondria were visualized with anti-OxPhos (Life Technologies) and images overlaid.

2D-DiGE

Sample Preparation

To identify protein and posttranslational differences in the mitochondria of cerebellum prior to PC degeneration in the pcd mouse, two dimensional differential gel electrophoresis (2D DiGE) was performed using P14 cerebella. This time point was chosen due to the fact that it is just prior to the time at which Fluorojade C (Chemicon) positive PC’s begin to be detected and thus represents the pre-degeneration stage. Cerebella were extracted and homogenized on ice with 6 strokes of a loose pestle and 6 strokes of a tight pestle in a 2mL dounce homogenizer. Mitochondria isolates were prepared using a qProteome Mitochondria Isolation kit according to the manufacturers instructions (Qiagen). Mitochondria pellets were solubilized in 2-D lysis buffer (30 mM Tris-HCl, pH 8.8, containing 7 M urea, 2 M thiourea and 4% CHAPS) and incubated at room temperature for 30 min on a shaker and centrifuged for 30 min at 16 000 x g to remove insoluble proteins. A portion of this lysate was saved for qualitative verification of mitochondrial enrichment via OxPhos immunoblotting.

CyDye Labeling
For protein labeling 30 µg of each protein sample was incubated with 0.7 µl of CyDye (1:5 diluted in DMF from 1 nmol/µl stock, GE Healthcare, Piscataway, NJ) at 4 °C for 30 min. Labeling was stopped by the addition of 0.7 µl of 10 mM L-Lysine and incubating for 15 min at 4 °C. The labeled samples were mixed, and equal volumes of 2X 2-D sample buffer (8 M urea, 4% CHAPS, 20 mg/ml DTT, 2% pharmalytes and trace amount of bromophenol blue) and 100 µl of destreak solution (GE Healthcare) were added. The total sample volumes were adjusted to 260 µL with rehydration buffer (7 M urea, 2 M thiourea, 4% CHAPS, 20 mg/ml DTT, 1% pharmalytes and trace amount of bromophenol blue). The samples were then incubated at room temperature for 10 min on a shaker and centrifuged for 10 min at 16000 x g. The supernatants were then subjected to isoelectric focusing (IEF). An additional 600 µg of unlabeled protein was run in parallel with the labeled samples to provide sufficient quantities of protein species for identification via mass spectrometry as described below.

**IEF, SDS-PAGE and Image Analysis**

Immobilized pH gradient (IPG) gel strips (pH 3-10) were placed onto the loaded samples and mineral oil added to cover the top of the strip. IEF were run following the protocol provided by the manufacturer (GE Healthcare). After the IEF, IPG gel strips were equilibrated in buffer 1 (50 mM Tris-HCl, pH 8.8, containing 6 M urea, 30% glycerol, 2% SDS, 0.01% Bromophenol Blue and 63 mM DTT) for 15 minutes and buffer 2 (50 mM Tris-HCl, pH 8.8, containing 6 M
urea, 30% glycerol, 2% SDS, 0.01% bromophenol blue and 300mM DTT) for 10 minutes with gentle agitation. IPG strips were then rinsed once in the SDS-gel running buffer and transferred to an SDS-Gel (10.5% SDS-gel prepared using low fluorescent glass plates). SDS-gels were run a 150V for approximately 6 hours or until the dye front reached the bottom of the gel. Gels were scanned using Typhoon TRIO (Amersham BioSciences) following the manufacturer's instructions and images processed by Image Quant software (version 5.0, Amersham BioScience). Quantitative analysis of protein spots were performed using DeCyder software (version 6.5) and potential pick lists were generated by using a protein fold change cutoff value of 1.5. Spots of interest were picked using an Ettan spot picker for mass spectrometry analysis.

**Mass Spectrometry**

Spot picked samples were subjected to in-gel trypsin digestion, peptide extraction, and desalting prior to MALDI-TOF/TOF (ABI 4700, Applied Biosystems). A GPS Explorer workstation was used to search the NCBI database to match both MS and MS/MS data to proteins in the database using the MASCOT search engine.

**Nuclear gDNA 8-oxodG ELISA**

The level of 8-oxodG in nuclear DNA was determined using an ELISA based approach with the monoclonal 8-oxodG (Trevigen). Total nuclear DNA was isolated from whole cerebella using a column based isolation kit (Qiagen).
Samples containing 50 μg of genomic DNA were digested with DNase I for 1 hour at 37°C. An 8-oxodG standard was used to create a serial dilution from which the regression equation was calculated. All samples were run in triplicate and levels of 8-oxodG were calculated. A simple T-test was used to determine whether a significant difference between the means existed.

**Subcellular Comparison of Protein Oxidative Damage**

Relative protein oxidation was determined using the DNP labeling method commercially known as the Oxyblot kit (Millipore). Cerebella samples were isolated and subjected to subcellular fractionation as described in the 2D-DiGE methods. Fractions were split into two tubes to be labeled or unlabeled (negative control) reactions for derivatizing of carbonyl groups. Protein content of samples was determined using the BCA assay kit according to the manufacturers instructions (Pierce). All samples were electrophoresed under reducing conditions on 4-12% gradient bis-tris gels and separated at 200V for 30 min. Gels were equilibrated in transfer buffer and transferred to PVDF membranes for 80 min at 30V. Membranes were blocked in 5% nonfat dry milk in tris buffered saline (TBS) for 1 hour. Anti-DNP antibody was prepared at 1:500 dilution in 5% BSA in Tween-TBS and incubated for 2 hours at RT. Membranes were then washed with Tween-TBS 3 times for 5 min each and infrared secondary was diluted 1:5000 in 5% BSA Tween-TBS and incubated for 1 hour at RT. Membranes were imaged using a Licor Odyssey infrared scanner set on 700 nm,
high quality, 69 µm resolution. Images were converted to grayscale and saved as tiff files.

Bands within each lane were defined and quantitated using the Odyssey imager software package. Integrated intensity values for each band were compared.

**Recombinant Protein Purification, Co-immunoprecipitation, SDS-PAGE**

Nna1 protein was prepared using 2 methods; the bacculovirus SF9 protein expression system and in vitro transcription translation (IVTT). The pFastBac vector and N terminal HA/FLAG Nna1 vector were used to generate the final bacculovirus. SF9 insect cells were transduced with bacculovirus and 72 hr later cells were harvested and lysate prepared. Lysates were subjected to nickel resin columns for crude purification of the HIS tagged Nna1 and resultant elutions were pooled and aliquots run on SDS-PAGE gells for verification of recombinant protein expression.

IVTT was used to produce Nna1 protein for use in protein-protein interaction. In some cases the $^{35}$S -Methionine was used to label IVTT derived proteins for use in interaction experiments (see below). The pCMV-TnT vector was used to generate a T7-Nna1 vector for use in the TnT Quick Couple IVTT kit from Promega. The pGEX4T-2 vector was used for GST fusion protein production of LOC100045876 in BL21DE E. coli. One liter cultures were used to produce usable quantities of GST-LOC100045876. Glutathione columns were
used to isolated and purify recombinant protein and protein yield was determined via spectrophotometry and coomassie staining of SDS-PAGE separated aliquots.

Equal amounts of GST, GST-LOC100045876, or GST-JNK2 were combined with $^{35}$S–Methionine labeled Nna1 from IVTT (or $^{35}$S–Methionine labeled Cdt1 for negative and positive controls). In a separate experiment GST-VHR was used as a negative control to rule out the possibility of non-specific binding of Nna1 yielding false positive results. Samples were incubated at 4°C for 2 hours with rotation. Glutathione beads were then added to each sample and incubated for 2 hours at 4°C with rotation. Beads were washed 5 times, 5 min each, in wash buffer (25 mM Tris-HCl, pH 7.6, 150 mM NaCl, 0.2% TritonX-100, protease and phosphatase inhibitors). Samples were then boiled in sample buffer and run on a 10% polyacrylamide gel. Gels were then coomassie stained, dried in a gel drier, and scanned in a BioRad Gel Doc imager for comparison of total GST fusion protein loaded. After scanning, gels were placed into a cassette and exposed to a phosphor screen for 72 hours. Screens were scanned with a Fuji FLA-5100 phosphorimager and tiff files saved. The predicted lack of interaction between LOC100045876 and Cdk2 was used to verify that evidence of interaction in the experimental results was not due unusually high non-specific binding of LOC10045876.
Figure 1. Similar extent of degeneration of spinal cord motor neurons after sciatic axotomy in pcd3J+/- and pcd3J-/- spinal cord. P7 pups from both pcd3J+/- and pcd3J-/- were subjected to sciatic axotomy and allowed to recover for 21 days. L4-L5 spinal cord was removed, sectioned and probed for ChAT (red) and NeuN (green). ChAT immunofluorescence revealed degeneration of spinal cord motor neurons in both pcd3J+/- and pcd3J-/- mice to similar degrees.
Figure 2. Degeneration and dysmorphic dendritic profiles in PC’s of cerebellar organotypic slice cultures. Immunofluorescence imaging of organotypic cerebellar slice cultures from WT (A,B,C) and pcd3J -/- (D,E,F) mice at P7-P8 and grown for either 14 or 21 days in vitro (DIV). A,C,D,F NeuN (green), Calbindin (red). B,E Confocal imaging of Calbindin showing PC dendritic arbors in P8DIV14 organotypic cerebellar slice cultures. A,C,D,F Scale bars = 250μm; B,E = 70μm.
Figure 3. Calbindin immunofluorescence image of PC from pcd3J/− cerebellar slice culture at DIV14 shows abnormal dendritic morphology. Representative image of dendrite pattern observed in slices from pcd3J/− cerebellum.
Figure 4. Novel putative phosphorylation of human Nna1 at T1127 and/or S1128 identified using SILAC phosphoproteomics. Phosphoproteomic analysis of human embryonic stem cells and the retinoic acid induced differentiated derivatives revealed that in the pluripotent state, human Nna1 appears to be phosphorylated at a C-terminally located Threonine or Serine residue. This phosphorylation was not observed in the RA differentiated samples.
Figure 5. Comparative Mitochondria proteome analysis workflow. Two pcd3J+/− and two pcd3J−/− (all littermates; P14) were used for a global mitochondria proteomic analysis to identify perturbations in protein content and/or posttranslational modifications. Western blot anti-OxPhos; crude homogenate and enriched mitochondria lysate.
Figure 6. 2D-DiGE comparative proteomics of P14 pcd3J+/− and pcd3J−/− cerebellar mitochondria identifies a predicted HMG box containing protein with differential migration and focusing patterns. Pcd3J+/− (green) and pcd3J−/− (red) mitochondria lysates were prepared from littermate samples (2 cerebella from each group were pooled and used for mitochondria isolation). Lysates were independently fluorescently labeled and combined to run on a 2D gel. Red indicates a greater presence in the pcd3J−/− sample; green from pcd3J+/−; and yellow indicates a similar presence in each sample. Circled spots were picked and subjected to MS/MS analysis.
Figure 7. Spot picking/MS/MS analysis reveals two differentially present spots to be identical proteins. Higher magnification images of differentially present spots and the corresponding MS results indicated each spot to be from a predicted protein similar to high-mobility group box 1 (LOC100045876).
Figure 8. Analysis of 2D-DiGE spot shifts shows pcd3J/- LOC100045876 slightly increased in size and acidity. In the pcd3J/- mitochondrial proteome, LOC100045876 protein has a decreased isoelectric point (estimated to be approximately 0.2 < WT pI) and decreased electrophoretic mobility compared to WT LOC0045876.
Figure 9. LOC100045876 homology to HMGB1 and domain map. The DNA binding site is flanked by HMG boxes and the sequence of these domains are identical in both HMGB1 and LOC100045876. The acidic tail is also identical to HMGB1. The major difference in sequences occurs between the second HMG box and acidic tail.
Figure 10. LOC100045876 and Nna1 adult tissue gene expression profile. LOC100045876 is expressed in most tissues whereas Nna1 expression is most abundant in CNS.
Figure 11. LOC100045876 localizes to the mitochondrion. Using anti-FLAG immunofluorescence C17.2 mNSC’s transfected with FLAG-LOC100045876 showed localization to the mitochondrial as marked by anti-OxPhos.
Figure 13. Nuclear gDNA 8-oxodG levels are unchanged in the pcd3J-/- mouse. ELISA based quantification of total genomic DNA isolated from cerebellar samples of pcd mice showed no significant difference in oxidative damage in the nuclear compartment. n=6 (WT, pcd3J+/−), n=5 (mut, pcd3J−/-)
Figure 14. Evaluation of protein oxidation shows similar extent of derivatized carbonyl levels in subcellular compartments of pcd cerebella. An oxyblot kit was used to label oxidized residues in protein samples from the nuclear, cytoplasmic, and mitochondrial compartments obtained from WT (pcd3J$^+$) or Mut (pcd3J$^-$) cerebella.
Figure 15. LOC100048576 interacts specifically with Nna1. GST-LOC100048576 was mixed with $^{35}$S-labeled Nna1 from IVTT (left panel). The specificity of the interaction of Nna1 with LOC100048576 was shown by the lack of interaction of $^{35}$S-Nna1 with either GST-VHR or GST-CDK1 (middle panel). The specificity of the interaction with LOC100048576 was shown by lack of binding of myc-Cdt1 to LOC100045876 (right panel). JNK2 has been reported as a Cdt1 interacting protein.
Chapter 2: Phosphoproteomic approaches to understanding hES cells

Preface

Chapter 2.1 is a reprint of; *Phosphoproteomic analysis: An emerging role in deciphering cellular signaling in human embryonic stem cells and their differentiated derivatives* published in Stem Cell Reviews and Reports October 19\textsuperscript{th} 2011. The dissertation author is a co-author on this publication. Chapter 2.2 is a reprint of *Phosphoproteomic analysis of human embryonic stem cells* published in Cell Stem Cell August 7\textsuperscript{th} 2009. The dissertation author was a co-author on this publication.

2.1 Phosphoproteomic analysis: An emerging role in deciphering cellular signaling in human embryonic stem cells and their differentiated derivatives

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 combined
total proteome plus total phosphoproteome (termed (phospho)proteome) 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 intelligent 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.

**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. Knowledge of the molecular mechanisms of self-renewal, pluripotency and differentiation has
consistently expanded with the increasing depth of stem cell biology research. Very briefly, self-renewal of human pluripotent stem cells relies on a relatively well-characterized network of transcriptional and chromatin regulation\textsuperscript{28-30}. Less well characterized, especially at the systems-biology/proteomic level, cellular signaling pathways involving TGFβ/Activin/Nodal, WNT, FGF, IGF, EGF, PDGF ERK, and JNK, are also critical to the control of self-renewal, pluripotency and differentiation\textsuperscript{31-42}. Identification of components contributing to maintenance of self-renewing hESCs has provided the means to identify pluripotent stem cells by immunochemistry, including immunoblotting 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\textsuperscript{43}, a revision of the initial estimate of 518\textsuperscript{44}. The majority of them are serine/ (S) threonine (T) kinases and about 90 are tyrosine (Y) kinases\textsuperscript{43}. 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\textsuperscript{33,36,39,41,42}. 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 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 activate SMAD2/3 whereas BMP activates SMAD1/5/8. In turn, NANOG transcription is activated by SMAD2/3 and inhibited by SMAD1/5/842,45,46. 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 genes47. 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 unclear41,45.

Reactivation of only a few transcription factor proteins, including OCT4 (POU5F1), SOX2, KLF4, MYC and/or NANOG are sufficient, depending on the cell type, for reprogramming of differentiated human cells to induced pluripotent stem cells (iPSCs)48-50. A growing body of evidence links these factors to regulatory signaling components important to self-renewal. KLF4 is a direct target of the TGFβ pathway33, and SOX2 and MYC may also be targets of TGFβ signaling51. Similarly, MYC is a downstream transcriptional target of canonical WNT signaling52. 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 well characterized53,54. 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. Although some reports estimate that, quantitatively, as few as ca. 50% of the mRNA transcripts 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. These observations and challenges make it clear that proteins, the final products of the vast majority of the genes, require direct analysis.

The quality of the cells used is critical to success of (phospho)proteomic studies

Although unclear at this time, due to 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. 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. Perhaps as a result of some of these variables, gene expression profiles correlated strongly with lab environment in which the cells were grown. Of the four studies published to date reporting
large-scale phosphoproteomic profiling of hESCs, two used WA-01 (H1) hESCs\textsuperscript{41,58}, a the third examined HUES-7\textsuperscript{36} while the fourth used HUES-9 and Odense-3\textsuperscript{42}. Between WA-01 and HUES-7\textsuperscript{36,41}, 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\%\textsuperscript{59}. Between a recent pair of (phospho)proteomic analyses of hESCs and their non-specifically differentiated derivatives, there was overlap of ca. 76\% of all identified proteins\textsuperscript{36,42}. With technical improvements, we identified much larger numbers of proteins and phosphorylation sites from WA-09 (H9)-hESCs and their primordial neural stem cell (NSC) derivatives, which included 86\% of the phosphoproteins identified in WA-01 hESCs and their retinoic acid- (RA) differentiated derivatives\textsuperscript{41} and ca. 98\% of the (phospho)proteins identified by each of two other groups from hESCs and their differentiated derivatives\textsuperscript{36,42}. However, more data is needed, including the use of more cell lines. Improving (phospho)proteomic workflows could have increased chances of detecting proteins identified in earlier datasets, due to increasing sensitivity and potentially greater (phospho)proteome coverage. Current observations suggest high reproducibility of the hESC (phospho)proteome.

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\textsuperscript{60}, high quality cell cultures
with the maximum possible homogeneity are indispensable for successful 
(phospho)proteomic results. Additionally, it is important to examine markers of 
pluripotency, differentiation and the nuclear to cytoplasmic ratio as measures of 
the quality of the cell populations used for phosphoproteomic analyses\(^\text{41}\). Due to 
the large requirement for resources and time, total (phospho)proteomic analysis 
generally precludes extensive replication of results. Thus, a careful choice of 
experimental conditions, 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 
immunoblots, for the presence, at expected quantitative levels, and 
phosphorylation of known phosphoproteins. Maximizing homogeneity of the cell 
population is important, as even small sub-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 
multipotency, and should allow more rapid discovery of (phospho)proteins 
helping to drive specific differentiation.

Delineating changing molecular profiles that occur during differentiation is 
a major impetus for application of (phospho)proteomics, to examine effects of 
differentiation on the cellular (phospho)proteome\(^\text{36,41,42}\). Recent studies have 
utilized retinoic acid\(^\text{41}\), modulation of BMP signaling\(^\text{36}\), and use of un-conditioned
medium or phorbol 12-myristate 13-acetate\textsuperscript{42}, which resulted in non-specifically differentiated cells.

**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\textsuperscript{61,62}, consistent with observations discussed above. Understanding pluripotency and germ line-specific differentiation has benefited from identification of protein 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\textsuperscript{63}.

There is increasing evidence that proteins and their activity may be regulated more extensively by changes in phosphorylation than by changes in protein abundance\textsuperscript{36,42,64}, and that extensive regulation of protein phosphorylation occurs during hESC differentiation\textsuperscript{36,41,42}. Many proteins detectable in both pluripotent and multipotent cell populations were detectably phosphorylated in one cell population or the other, but not both, contributing to a low overlap of proteins without detectable phosphorylation. Whether others
observe similar trends remains to be seen. Although preliminary, 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. In addition, phosphorylation can inhibit or facilitate additional PTMs including SUMOylation, acetylation, methylation and ubiquitination in adjacent portions of the protein\textsuperscript{65,66}.

Analogous to identification of interacting transcription factors by interrogating a gene expression dataset, (phospho)proteomic analysis can identify the regulation of signaling networks\textsuperscript{36,41,42,58}, 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\textsuperscript{67}. However, the size of datasets from identified (phospho)proteomes has increased rapidly, likely representing improving degrees of comprehensiveness of the analyses. Larger (phospho)proteomic datasets, although more difficult to analyze, should further improve the value of the analyses for systems biology and to support studies focused on smaller numbers of (phospho)proteins. It is unknown how many protein phosphorylation events are in proteomes. 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\textsuperscript{64}, 66.5% in hESCs and their non-specifically differentiated derivatives\textsuperscript{42}, 70% in HeLa cells\textsuperscript{68}. 
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 assays of cell signaling, based on Western blotting and immunofluorescence, although critical for hypothesis-based investigation of discreet components in pathway function, are likely to be insufficient for characterization of cellular signaling networks. Even targeted kinase disruption may modulate many components of robust signaling networks, most of which would not be identified without unbiased, large-scale (phospho)proteomics. Similarly, RNAi-based knockdown of a single phosphatase causes a comparatively large percentage of significant changes in protein phosphorylation relative to protein abundance. 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, large-scale (phospho)proteomics is an essential and powerful tool for understanding control of pluripotency and differentiation.

**The workflow of phosphoproteomics: 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 (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.
in contrast with “top-down” proteomics, which involve MS and MS/MS analyses of intact polypeptides or proteins\textsuperscript{74-76}. The application of ever improving peptide separation and phosphopeptide enrichment techniques has facilitated improvement of the sensitivity of phosphoproteomic analyses. For analysis of protein phosphorylation, effective phosphopeptide enrichment is essential. Commonly employed techniques for phosphopeptide enrichment include IMAC\textsuperscript{32,69,71,73,77-80}, TiO\textsubscript{2}-based phosphopeptide enrichment\textsuperscript{81-84}, and soluble polymer-based phosphopeptide enrichment\textsuperscript{81,85,86}.

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 more innovative analytical platforms be broadly adopted by others\textsuperscript{87}. However,
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 raw 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.

**Superb protein sample preparation is indispensable**

Following culture of high quality cells, yielding at least 1 mg of total protein per sample, cells are lysed\(^\text{32}\). Buffers include salts, detergent, protease and phosphatase inhibitors and compatibility with the MDLC-MS/MS methods. Protein kinase inhibitors and protease inhibitors are valuable in preserving the fidelity of the (phospho)proteome during and after cell lysis. 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. Other groups have also reported successful procedures. One
lysis buffer contained 8 M urea, and the protein phosphatase inhibitors sodium
fluoride and sodium orthovanadate, while another included these same
ingredients and β-glycerophosphate (a protein phosphatase inhibitor) plus
protease inhibitor tablets. A fourth lysis buffer contained 8 M urea, sodium
pyrophosphate, sodium orthovanadate, complete mini ETDA-free protease
inhibitor and combined phosSTOP phosphatase inhibitor. A different study
demonstrated the importance of phosphatase inhibitors for increased
identification of protein phosphorylation sites.

Commonly used separation and phosphopeptide enrichment strategies
include SCX, IMAC and/or TiO₂, 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 strategies 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\(^8\). Hydrophilic interaction chromatography (HILIC) has been used as a successful alternative to SCX\(^9\,10\). 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\(^{3+}\)\(\,32,58,71,73,77,78,92,93\)). A complement to, and/or substitute for IMAC, is the use of TiO\(_2\) to enrich phosphopeptides\(^81\,84\). 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. Less commonly used phosphopeptide enrichment uses soluble polymers, variously termed dendrimers or PolyMAC\(^81,85,86\). Among the reports of hESC (phospho)proteomes, SCX and IMAC separation/enrichment strategies were used in two of the studies\(^32,58,94\), whereas TiO\(_2\) rather than IMAC was used in the two others\(^36,42\).

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\textsuperscript{32,36,42,58,68,71,73,77,80,93}, although we currently use higher flow RP-LC with robust operation and sensitivity similar to nanoflow LC\textsuperscript{41}. 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\textsuperscript{+}-6\textsuperscript{+}, with most ions in the lower charge states.

**MS/MS of peptides and phosphopeptides is used in large-scale (phospho)proteomic analyses**

These 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, their fragmentation, and re-scanning the product ions from one precursor ion at a time. Product ion scans are termed MS/MS scans. In some cases, a dominant product ion, resulting from prominent neutral loss of H\textsubscript{3}PO\textsubscript{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 sequence the phosphopeptide\textsuperscript{95}. 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 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 mass analyzer for an Orbitrap\textsuperscript{95}. Recently, MS and MS/MS scans were both performed in the Orbitrap, using newer instruments\textsuperscript{96}.

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 seconds, 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 analyze 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 scans.

Collision-induced dissociation (CID) and electron-transfer dissociation (ETD) are commonly used methods of peptide fragmentation. CID is most frequently employed in phosphoproteomic analyses, often using an ion trap mass spectrometer, which is used to deduce peptide and phosphopeptide sequences. Activation by CID results primarily in b- and y-product ions whose \textit{m/z} is measured in the MS/MS scans, in order to deduce the peptide sequence from the product ions. Several search engines are available to 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
engine, used for the examples presented is SEQUEST\textsuperscript{97}, although a variety of other search algorithms are available. 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. 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. Each peptide has unique fragmentation characteristics, and it is difficult to predict how specific peptides will fragment\textsuperscript{98}.

Some peptides and phosphopeptides are more effectively identified using fragmentation by ETD\textsuperscript{99} than CID, especially those precursor ions from phosphopeptides with lower $m/z$ ratios and higher charge states\textsuperscript{100}. One challenge with ETD MS/MS spectra is effectively searching them against protein databases. Recent advances in processing of the raw data, including the removal of un-fragmented precursor ions plus charge-reduced precursor ions, neutral loss products from these precursor ions, and potential clusters of un-fragmented ions in the MS/MS spectra have led to increased numbers of high quality IDs made from ETD-MS/MS spectra\textsuperscript{101-104}. This results in an improved distribution of Xcorr scores and hence more abundant, higher confidence peptide IDs.

Among the first reports of hESC phosphoproteomes, three\textsuperscript{32,36,42} exclusively used CID, and one\textsuperscript{58} 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. The instrument control software of some mass spectrometers includes the decision tree method, and this is our current MS/MS method of choice. However, there are additional MS/MS methods available.

Fragmentation of phosphopeptides by CID sometimes results in prominent neutral loss of $\text{H}_3\text{PO}_4$ 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. Proline is also a favored site of fragmentation by CID. Potential phosphorylation site rearrangement by CID was suggested, but others reported that it is negligible. In contrast, ETD tends to result in an absence of neutral loss of $\text{H}_3\text{PO}_4$ from phosphopeptides, and yields primarily c- and z product ions, although a lesser quantity of b- and y-ions result from ETD. Moreover, fragmentation of peptides containing multiple Arg and Lys residues by ETD tends to be more robust than by CID.

Use of CID and ETD peptide fragmentation was compared in a (phospho)proteomic analysis of hESCs. 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, similar to our results (J. Hou and L. M. Brill, unpublished data). ETD was twice as likely to produce backbone cleavage in vicinity of phosphoserine (pS) residues than CID, although there was no consistent difference in efficiency between the fragmentation methods at phosphotyrosines.
In total, CID identified 5773 unique phosphopeptides and ETD identified 8603 unique phosphopeptides with 2421 identified by both methods. Sequence coverage, defined as at least one fragmentation event per amino acid residue along the peptide backbone, was 86% for ETD and 74% for CID, which is important for phosphorylation site localization, which was 49.8% for ETD and 26.9% for CID.

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, but we have had better results without the use of lock mass.

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

Although it is tedious, time consuming and requires expert personnel, manual validation of PSMs of phosphopeptides can be very important\textsuperscript{32,116}. There are defined criteria for accuracy of manual validations\textsuperscript{116}. In addition, software to analyze the reliability of phosphopeptide identifications has been reported\textsuperscript{117}. In anecdotal discussions, there is general agreement that phosphopeptides are often not accurately scored by database searching and filtering algorithms. Although phosphopeptide identifications with high scores are reliable in our experience, some PSMs scoring below stringent FDR thresholds are, in spite of the lower score, accurate, whereas others are incorrect. One reliable strategy involves recovery of accurate, but lower scoring MS/MS spectra that are matched to peptides previously and confidently identified\textsuperscript{118}. In addition, manual use of chromatographic alignment, in up to 3 dimensions of separation (e.g. SCX, phosphopeptide enrichment, and RPLC) improves confidence in manually validated phosphopeptide identifications made two or more times. It would be valuable, but computationally demanding, to automate the use of such a "3D retention time alignment". Protein phosphorylation IDs that form the basis for extensive biological follow-up experiments should be especially carefully validated\textsuperscript{116}.

Quantitative (phospho)proteomics used on hESCs include SILAC and label-free
One quantitative proteomics method which was reported in mouse ESCs, but not hESCs, is termed iTRAQ (isobaric tagging for relative and absolute quantification\textsuperscript{119}). Stable isotope labeling with amino acids in cell culture (SILAC\textsuperscript{120}) 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 carbon and nitrogen atoms of Arg and/or Lys contain stable, “heavy” isotopes of carbon (\textsuperscript{13}C) and nitrogen (\textsuperscript{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 in 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 by 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\textsubscript{6}, $^{15}$N\textsubscript{4} Arg and $^{13}$C\textsubscript{6}$^{15}$N\textsubscript{2} Lys labeled, undifferentiated hESCs to unlabeled hESCs that were treated by adding BMP4 and removing FGF2 for 30, 60 or 240 min\textsuperscript{36}. SCX separation, phosphopeptide enrichment with TiO\textsubscript{2} and LC-MS/MS was employed. About 2/3rds of the identified peptides were quantifiable, and of these, half of the phosphopeptides showed changes in relative abundance.
during treatment. 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. The relative abundance of ca. 50% of the phosphopeptides also changed during differentiation.

Label-free quantification, frequently using spectral counts, which can include extracted ion chromatograms for further quantification, is also a reliable method to estimate the relative abundance of proteins and protein phosphorylation events. 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. A large portion of the peptides is derived from only one known protein (non-degenerate peptides). The protein from which degenerate peptides were 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. In the only direct comparison of SILAC- and label-free (spectral counting) quantification we are aware of, the two methods performed similarly. 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, a similar percentage as the studies using SILAC, in spite of the differing protocols 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 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 to be important in undifferentiated hESCs, on the basis of phosphoproteomic and pathway analyses, and these are important to the undifferentiated state of the cells\textsuperscript{32}. The demonstration 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\textsuperscript{32}, was consistent with earlier results, under un-defined culture conditions, that PDGF together with sphingosine-1 phosphate facilitated maintenance of undifferentiated hESCs\textsuperscript{34}. Thus, phosphoproteomic analyses of hESCs and their differentiated derivatives support an accelerated understanding of the biology of cell states.

**Effective bioinformatics analysis of (phospho)proteomic datasets is essential**

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\textsuperscript{124,125}. 


Delineating connections of identified and quantified proteins to functional and structural protein categories, disease markers, therapeutic targets and signaling pathways can cumulatively establish informative trends among datasets and highlight critical components otherwise masked by data overload\textsuperscript{126}. 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\textsuperscript{127} and MS-GF\textsuperscript{110} may be increasingly used. Other tools with relevance to (phospho)proteomics are available (\textbf{Table 2}). MaxQuant software has been reported for peptide identification and quantification of SILAC data\textsuperscript{128}. Spectral counting/label-free quantification can include the analysis of statistical significance of spectral counts of proteins between or among samples\textsuperscript{123,129}. Several proteomic and phosphoproteomic databases are available, such as Phosphopep\textsuperscript{130}, PHOSIDA\textsuperscript{131}, Phosphosite\textsuperscript{132} and PRIDE\textsuperscript{133}. 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 structural and functional protein classes, 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\textsuperscript{134}. 
Pictorial representations can be constructed indicating the possible known and inferred connections between kinases and substrates. NetworKIN was used to identify the potential kinases phosphorylating the proteins identified in hESCs\textsuperscript{36}.

**Biological insights from unbiased phosphoproteomic studies of hESCs**

The results of large-scale (phospho)proteomic analyses are unbiased\textsuperscript{32,135}, meaning that the proteins that are present and detectable are what is discovered, in contrast with specifically targeting the detection of a given (phospho)protein, such as with FACS or Western immunoblots. Although specific peptides and phosphopeptides can be specifically targeted for precise quantification using an MS method entitled multiple reaction monitoring (MRM, also termed selective reaction monitoring, SRM)\textsuperscript{119,136}, 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 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\textsuperscript{32}. Using pY immunoprecipitation followed by mass spectrometry, it was proposed that the surprisingly prominent pY in hESCs could be facilitated by utilization of unique pY target sequences\textsuperscript{137}. In addition, pS and phosphothreonine (pT) can have similar or differing biological effects as pY\textsuperscript{138} and pS/pT can oppose the action of pY\textsuperscript{139}. For instance, Cortactin, though a Src
substrate activated by pY, showed six detectable pS sites but no detectable pY$^{140}$.

A protein implicated in epigenetic regulation of pluripotency, DNMT3B, exhibits multiple phosphorylated sites in hESCs$^{32,36,42}$. Phosphorylation of DNMT3B was identified more often in undifferentiated than differentiated cells$^{41}$ and this protein is more heavily phosphorylated and abundant in undifferentiated hESCs than in their pure, primordial NSC derivatives. 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$^{36,41,42}$.

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$^{58}$, phosphorylation of Smad 2/3, GSK3 and β-catenin was identified$^{32,36,42,58}$, and additional phosphoproteins in the TGFβ and WNT pathways contained identified phosphorylation in hESCs. In response to BMP4, the abundance of phosphorylated transcription factors and nucleic acid binding factors decreased, whereas total phosphopeptides increased as ES cells began to differentiate$^{36}$. In response to RA, many proteins showed increased or decreased phosphorylation during differentiation, though cumulatively an increase in phosphorylation during differentiation was also observed$^{32}$. It is
unclear whether effects of exogenous RA or BMP4 are directly downstream of the ligand or common, non-specific effects of differentiation\textsuperscript{32,36}. In addition, signaling cascades not previously known in hESCs, or whose importance was under-appreciated 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\textsuperscript{32}. Similarly, phosphorylation of the JNK target JUN increased after BMP4 addition during differentiation\textsuperscript{36}.

Oxygen tension is a critical factor in the biology of stem cells\textsuperscript{141}. For example, mild hypoxia improves the efficiency of generation of both mouse and human iPSCs\textsuperscript{142}. FRAP1 (mTOR) was phosphorylated in hESCs\textsuperscript{32,42} and activates HIF-1 (Hypoxia-inducible factor-1, a heterodimer of HIF1A and ARNT (HIF-1\(\beta\))\textsuperscript{143}. ARNT was also phosphorylated in undifferentiated hESCs\textsuperscript{32}. The HIF pathway is the primary mediator of hypoxic adaptation\textsuperscript{141}, and in mouse ESCs, Arnt directly activates expression of Pou5f1 (Oct3/4)\textsuperscript{144}, 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\textsuperscript{145}.

\textbf{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. 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\textsuperscript{32,36,42}. 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\textsuperscript{32}. However, biological follow-up experiments have been increasingly pursued in association with proteomic analyses\textsuperscript{31,37,38,146,147}. 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, but there is often a poor correlation
between mRNA and protein expression\textsuperscript{119}. Similarly, preliminary examinations of the correlation between protein phosphorylation and mRNA expression also suggested that it could be relatively low\textsuperscript{32}. In addition, histone methylation patterns indicative of inactive, poised and actively transcribed chromatin regions, which are important to the biology of pluripotency\textsuperscript{148}. 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\textsuperscript{36}.

**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, and must be recognized during the interpretation of the data. 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\textsuperscript{32}. 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. Some phosphorylation sites are ambiguous, but the confidence in the proposed site localization can be estimated\textsuperscript{42,149}, 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\textsuperscript{32}.

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 indicate whether a given phosphorylation site is activating, inhibitory, involved in altering protein-protein interactions, a modulator of protein stability, or other potential effects. However, it is becoming increasingly clear that (phospho)proteins participate in regulation of the cellular state\textsuperscript{32,36,42,58}. 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 in HUES-7 cells, but no biological effect on these hESCs was mentioned\textsuperscript{36}.

\textbf{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 traditional hypothesis-based approaches such as antibody-based strategies. The application of MDLC-MS/MS-based (phospho)proteomics to stem cell biology will likely increase, and requires multi-disciplinary 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. Similar approaches may 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\textsuperscript{150}, and 6367 sites of N-glycosylation were mapped to 2352 proteins from 4 mouse tissues\textsuperscript{151}.

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\textsuperscript{56,152}. Similar trends appear to 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\textsuperscript{63}. As detection of PTMs becomes more
sensitive, thousands of new phosphorylated residues in larger datasets have been detected. The proportion of a given protein that is phosphorylated at specific residues also varies widely\textsuperscript{87,153}. 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.

Chapter 2.1 is a reprint as it appears in “Phosphoproteomic analysis: An emerging role in deciphering cellular signaling in human embryonic stem cells and their differentiated derivatives” by Brian Tobe, Junjie Hou, Andrew Crain, Ilyas Singec, Evan Snyder, Larry Brill published in Stem Cell Reviews and Reports Oct 19\textsuperscript{th} 2011.
Table 1. Selected concepts and terminology of relevance to (phospho)proteomics (adapted and expanded from\textsuperscript{154,155})

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-dimensional liquid chromatography (MDLC)</td>
<td>Separation strategy 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\textsubscript{2} (dimension 2), and reversed-phase (RP; dimension 3).</td>
</tr>
<tr>
<td>Strong cation exchange chromatography (SCX)</td>
<td>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</td>
</tr>
<tr>
<td>Immobilized metal affinity chromatography (IMAC)</td>
<td>Phosphopeptide enrichment method using trivalent metal ions (usually Fe\textsuperscript{3+} or Ga\textsuperscript{3+}) bound to a stationary phase to selectively chelate negatively charged phosphate groups of phosphoproteins or more commonly phosphopeptides.</td>
</tr>
<tr>
<td>TiO\textsubscript{2}-based phosphopeptide enrichment</td>
<td>Mode of chromatography to selectively enrich phosphopeptides, in which TiO\textsubscript{2} particles selectively bind phosphate groups of phosphopeptides through Lewis acid base interactions.</td>
</tr>
<tr>
<td>Tandem mass spectrometry (MS/MS)</td>
<td>Measurement of mass-charge (m/z) ratio and intensity of precursor ions followed by isolation of individual precursor ions, their fragmentation and scanning the mass-to-charge (m/z) ratios of the resulting product ions to deduce peptide sequence</td>
</tr>
<tr>
<td>Collision induced/activated dissociation (CID/CAD)</td>
<td>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.</td>
</tr>
<tr>
<td>Electron transfer dissociation (ETD)</td>
<td>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</td>
</tr>
<tr>
<td>Stable isotope labeling by amino acids in cell culture (SILAC)</td>
<td>Method for differential quantification of protein expression between or among two or three samples that relies on incorporation of \textsuperscript{13}C and \textsuperscript{15}N labeled (&quot;heavy&quot;) 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 masses</td>
</tr>
<tr>
<td>False discovery rate (FDR)</td>
<td>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\textsuperscript{114,115}</td>
</tr>
<tr>
<td>Orbitrap</td>
<td>Widely used mass analyzer with high mass accuracy and resolution that measures the oscillation frequency of ions along the Orbitrap axis to enable precise determination of the m/z ratio</td>
</tr>
<tr>
<td>Biological replicate</td>
<td>Repeat of the biological aspect of a given experiment, such as use of two separate cell samples, or independently repeating a cellular treatment.</td>
</tr>
<tr>
<td>Technical Replicate</td>
<td>Repeat of the proteomics portion of the experiment, thus analyzing the same sample again with the proteomic workflow.</td>
</tr>
</tbody>
</table>
### Table 2. Bioinformatic resources useful to (phospho)proteomics

<table>
<thead>
<tr>
<th>Resource</th>
<th>Purpose</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>NetworKIN (<a href="http://networkin.info/search.php">http://networkin.info/search.php</a>)</td>
<td>Kinase-substrate and phosphorylation site interactions</td>
<td>134</td>
</tr>
<tr>
<td>GeneGo Metacore Pathway Analysis</td>
<td>Systems biology/pathway analysis</td>
<td>156</td>
</tr>
<tr>
<td>(<a href="http://www.genego.com/">http://www.genego.com/</a>)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ingenuity Pathway Analysis</td>
<td>Systems biology/pathway analysis</td>
<td></td>
</tr>
<tr>
<td>(<a href="http://www.ingenuity.com">http://www.ingenuity.com</a>)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KEGG</td>
<td>Systems biology</td>
<td>157</td>
</tr>
<tr>
<td>(<a href="http://www.genome.jp/kegg/">http://www.genome.jp/kegg/</a>)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STRING</td>
<td>Systems biology</td>
<td>158</td>
</tr>
<tr>
<td>(<a href="http://string-db.org/">http://string-db.org/</a>)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NetPhosK (<a href="http://www.cbs.dtu.dk/services/NetPhosK/">http://www.cbs.dtu.dk/services/NetPhosK/</a>)</td>
<td>Kinase-substrate and phosphorylation site interactions</td>
<td>159</td>
</tr>
<tr>
<td>PHOSIDA (<a href="http://www.phosida.com">http://www.phosida.com</a>)</td>
<td>Interactive database of phosphopeptides</td>
<td>160</td>
</tr>
<tr>
<td>Phospho.ELM (<a href="http://phospho.elm.eu.org">http://phospho.elm.eu.org</a>)</td>
<td>Kinase-substrate and phosphorylation site interactions</td>
<td>161</td>
</tr>
<tr>
<td>PhosphoSite (<a href="http://www.phosphosite.org">http://www.phosphosite.org</a>)</td>
<td>Kinase-substrate and phosphorylation site interactions</td>
<td>132</td>
</tr>
<tr>
<td>MaxQuant</td>
<td>Precursor mass correction, peptide identification and quantification of SILAC data</td>
<td>138</td>
</tr>
<tr>
<td>Abacus, QSpec</td>
<td>Spectral count analysis</td>
<td>129,162</td>
</tr>
<tr>
<td>Phosphopep</td>
<td>Phosphorylation site database, analysis of the sites, data integration</td>
<td>130</td>
</tr>
<tr>
<td>PRIDE</td>
<td>Proteomic database, contains PTMs, support manuscripts, papers</td>
<td>133</td>
</tr>
<tr>
<td>(<a href="http://www.ebi.ac.uk/pride">http://www.ebi.ac.uk/pride</a>)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trans-Proteomic Pipeline (TPP)</td>
<td>Software tools for FDR filtering, quantification, others</td>
<td>114,115</td>
</tr>
<tr>
<td>(<a href="http://tools.proteomecenter.org/">http://tools.proteomecenter.org/</a>)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.2 Phosphoproteomics of human embryonic stem cells and neuroectodermal derivatives

Abstract

Protein phosphorylation, while critical to cellular behavior, has been under-characterized in pluripotent cells. Therefore, we performed phosphoproteomic analyses of human embryonic stem cells (hESCs) and their differentiated derivatives. 2546 phosphorylation sites were identified on 1602 phosphoproteins; 389 proteins contained more phosphorylation site identifications in undifferentiated hESCs, whereas 540 contained more such identifications in differentiated derivatives. Phosphoproteins in receptor tyrosine kinase (RTK) signaling pathways were numerous in undifferentiated hESCs. Cellular assays corroborated this observation by showing that multiple RTKs cooperatively supported undifferentiated hESCs. In addition to bFGF, EGFR, VEGFR and PDGFR activation was critical to the undifferentiated state of hESCs. PDGF-AA complemented a sub-threshold bFGF concentration to maintain undifferentiated hESCs. Also consistent with phosphoproteomics, JNK activity participated in maintenance of undifferentiated hESCs. These results support the utility of phosphoproteomic data, provide guidance for investigating known and novel proteins in hESCs, and complement transcriptomics/epigenetics for broadening our understanding of hESC fate determination.
Introduction

Human embryonic stem cells (hESCs) are a model developmental system that may have potential clinical value for mitigating diseases. Mechanisms of hESC fate determination are not well defined, although there has been progress in elucidating molecular circuitry of self-renewing ESCs. Transcriptional profiles of hESCs and more limited ChIP-on-chip and proteomic analyses suggest mechanisms underlying hESC self-renewal and differentiation. In addition to transcriptional and translational regulation, cell fate determination is controlled by protein phosphorylation, a critical determinant of cell signaling. Recent phosphoproteomic analyses of human mesenchymal stem cells identified 716 and 703 protein phosphorylation sites. However, protein phosphorylation has not been well characterized in pluripotent cells. Therefore, we performed a large-scale multidimensional liquid chromatography-tandem mass spectrometry-based phosphoproteomic analysis of undifferentiated hESCs and their differentiated derivatives for identification of protein phosphorylation sites in these cells.

Undifferentiated hESCs were cultured under feeder-free conditions with bFGF. Comparable differentiated derivatives were obtained by removal of bFGF and treatment with retinoic acid (RA), which induces nearly complete albeit non-specific differentiation to a heterogeneous population of cells. Removal of bFGF alone does not result in complete differentiation, whereas concurrent RA treatment causes virtually complete loss of the undifferentiated population in 4
days (required for this type of analysis). Our data provide a freely available
resource of protein phosphorylation sites in hESCs and differentiated derivatives
(http://www.ebi.ac.uk/pride/). These data have begun to prove informative and
predictive. For example, as proof-of-concept, pathway analyses of the
phosphoproteins suggested potential responses of hESCs to perturbations of
receptor tyrosine kinase (RTK) signaling pathways. To test some RTK pathways
for a role in the maintenance of undifferentiated hESCs, we treated hESC
cultures with selected agonists or antagonists of these pathways. Their effects
were consistent with predictions of the phosphoproteomic analyses. Furthermore,
the data suggested a role of novel proteins in hESC self-renewal or
differentiation, thus providing extensive guidance for future research.

Results

Because phosphoproteomic analysis is challenging\textsuperscript{466} and has not been
reported in hESCs, we chose to analyze the well-characterized hESC line H1
(WiCell; WA01)\textsuperscript{27}, which has been used in molecular studies of hESCs e.g.\textsuperscript{29,31,37}.
59 hESC lines, including H1, showed remarkable conservation of hESC
markers\textsuperscript{169}, which provided confidence that our findings would be representative.
Before analyzing protein phosphorylation, the undifferentiated hESC markers
OCT4\textsuperscript{27} and SSEA-4\textsuperscript{170} were examined to assess whether the hESCs were truly
undifferentiated under our culture conditions, and whether differentiation was
complete. Undifferentiated hESCs were cultured on matrigel-coated plates in
feeder-free cultures using conditioned media (CM) that contained 8 ng/ml of
added bFGF. A heterogeneous population of differentiated derivatives of the hESCs were obtained by removal of bFGF and treatment with 5 µM RA for 4 days. OCT4 was detected in ca. 97% of the hESCs under the feeder-free conditions, whereas it was nearly undetectable in differentiated derivatives (Figure 16). Similarly, SSEA-4 was positive in the undifferentiated hESCs and nearly absent in differentiated derivatives. Moreover, the nucleus-to-cytoplasm ratio, also monitored as an indicator of whether hESCs are undifferentiated or differentiated, was consistent with OCT4 and SSEA-4 expression (Figure 16).

These observations suggested that our cells represented two distinct populations -- “undifferentiated” or “differentiated” hESC derivatives -- that might then be reliably subjected to phosphoproteomic analysis, using MDLC-MS/MS technology, that can result in unbiased discovery of protein phosphorylation sites.\textsuperscript{135}

Phosphoproteomic analyses of hESCs and their differentiated derivatives were performed using automated MDLC, a linear ion trap mass spectrometer and readily available bioinformatics algorithms. Phosphorylated peptides from total proteins from undifferentiated hESCs or their differentiated derivatives were separated, enriched, and analyzed using MDLC comprised of strong cation exchange chromatography (SCX), reversed-phase (RP) desalt-Fe\textsuperscript{3+}-immobilized metal affinity chromatography (desalt-IMAC) and RP HPLC coupled to nano-electrospray ionization-tandem mass. IMAC, for phosphopeptide enrichment, coupled to RP HPLC-ESI-MS/MS is a robust technique for phosphoproteomic
analyses\textsuperscript{71,81,171} and automation improves reliability and reproducibility\textsuperscript{79}. Because phosphorylated proteins are frequently at low abundance, sub-stoichiometrically phosphorylated and difficult to identify\textsuperscript{166}, replicate analyses were performed to increase phosphoproteome coverage. Replicates increase proteome coverage, especially of lower abundance proteins\textsuperscript{121}, and the impact of experimental variation in LC-MS/MS can be minimized by replicates\textsuperscript{172}. Phosphopeptides were identified with high confidence.

To complement identification, extracted ion chromatograms (XICs) were used to quantify the relative abundance of phosphopeptides. The normalized abundance of randomly selected phosphopeptides identified in all 4 phosphoproteomic analyses (2 biological replicates, \textit{i.e.} phosphopeptides from two pairs of independent cultures of undifferentiated hESCs or their differentiated derivatives) demonstrated relatively low variability. This degree of consistency agrees with previous findings in which proteomic data can be reliably compared among experiments\textsuperscript{172}.

In contrast, differential phosphopeptide identification implies differential phosphopeptide abundance. We used data-dependent MS/MS, and peptide abundance and identification correlate in data-dependent MS/MS\textsuperscript{121}. Selected phosphopeptides identified in undifferentiated hESC or differentiated derivative cell populations were also quantified using XICs. Furthermore, signal from each of the selected phosphopeptides was manually sought in the MS/MS data from analyses in which it had not been identified by SEQUEST searches, in order to
test whether the phosphopeptide was detectable, and if so, its relative abundance among the phosphoproteomic analyses. Only a fraction of the phosphopeptides not identified in SEQUEST searches was detectable (via a poor quality MS/MS spectrum) when searching the raw data. However, every phosphopeptide that was examined demonstrated a higher normalized abundance in analyses in which it was identified than in analyses in which it was not identified by SEQUEST searches. Although lack of identification of a phosphopeptide is not evidence for its absence, identification vs. lack of identification implies that the phosphopeptide is likely to be more abundant in the cell population in which it was identified, consistent with our results and those of others121.

Western blots were performed on proteins from undifferentiated hESCs and differentiated derivatives, using antibodies recognizing phosphorylation sites previously identified by MDLC-MS/MS. All 9 antibodies that were used recognized bands with the expected mobility on Western blots, providing confidence in phosphorylation site identifications.

Representative Western blots, including normalized integrated intensities of phosphoprotein bands, are shown in Figure 20. Phosphorylation of mTOR on Ser2448 was apparently more abundant in undifferentiated than differentiated cells (Figure 20A), and mTOR Ser2448 phosphorylation was identified in undifferentiated but not differentiated cells using MDLC-MS/MS. PAK1 phosphorylation on Ser144 was identified twice in undifferentiated cells and once
in differentiated cells by MDLC-MS/MS and Western blots suggested that PAK1 phosphoserine144 was more abundant in undifferentiated than differentiated cells (Figure 20). Antibodies recognizing PTK2 phosphotyrosine576/577 suggested that phosphorylation of this site was more abundant in differentiated derivatives than undifferentiated hESCs consistent with identification of PTK2 phosphorylated on Tyr576, using MDLC-MS/MS, only in differentiated derivatives. Phosphorylation of CDK1/2/3/5 on Thr14 and Tyr15 (2 conserved residues in all 4 CDK proteins) was more abundant in undifferentiated cells and XIC peak areas suggested that phosphorylation of CDK1/2/3 on Thr14 and Tyr15 was more abundant in undifferentiated cells. CDK1/2/3/5 phosphorylated on Thr14 and Tyr15 are recognized in Western blots, and the corresponding phosphopeptides identified by MDLC-MS/MS (IGEGT*YGVVY and IEGTY*GVVY; for brevity, designated as originating from CDK2) are identical among CDK1/2/3, whereas the corresponding peptide from CDK5 differs at 2 amino acid residues (IGEGT*Y*GTVF), which is easily distinguishable by MS/MS. The relative abundance of JUN phosphorylated on Ser63, and HSP27 phosphorylated on Ser82 was similar in undifferentiated and differentiated cells on Western blots (not shown), and phosphorylated JUN Ser63 as well as phosphorylated HSP27 Ser82 were both identified the same number of times in undifferentiated and differentiated cells, demonstrating further agreement between Western blots and MDLC-MS/MS.
If subsequent studies focus on one or a few especially critical sites of protein phosphorylation, it is advisable to examine the phosphorylation site using an independent technique. However, MDLC-MS/MS is reliable for phosphoproteome analysis, and can yield unbiased, large-scale discovery of protein phosphorylation and our findings support its accuracy. Together, these results suggest that application of MDLC-MS/MS for identification of phosphopeptides was suitable for phosphoproteomic analysis of undifferentiated hESCs and their differentiated derivatives.

Each phosphoprotein, from which phosphopeptides were derived, was classified as either: (1) containing more phosphorylation site identifications in undifferentiated hESCs, (2) containing more phosphorylation site identifications in differentiated hESC derivatives, and (3) containing a similar number of phosphorylation site identifications in both cell populations. A protein is conservatively defined to contain more phosphorylation site identifications in a cell population if its phosphorylation was identified exclusively in this population or at least 3-fold more frequently than in the other population, otherwise the protein is considered to contain a similar number of phosphorylation site identifications in populations from both cell states. Although identification of protein phosphorylation sites was unlikely to be comprehensive, as implied by studies using different cell types, among the 2546 non-redundant phosphorylation sites, 472 were on proteins containing more phosphorylation site identifications in undifferentiated hESCs, whereas 726 were on proteins
containing more phosphorylation site identifications in differentiated hESC-derivatives (Figure 17A). 94% of the peptides were singly phosphorylated, whereas the rest were doubly phosphorylated, similar to other studies using IMAC for phosphopeptide enrichment\textsuperscript{81}. Serine, Thr and Tyr phosphorylation comprised ca. 82%, 14% and 4% of the sites, respectively, and Tyr phosphorylation was relatively prominent in undifferentiated hESCs (Figure 17C). Among the 1602 proteins, 389 contained more phosphorylation site identifications in undifferentiated hESCs, whereas 540 contained more phosphorylation site identifications in differentiated hESC-derivatives (Figure 17B).

Transcription factors can reprogram differentiated cell types to ESC-like cells when ectopically expressed, and were the most abundant known phosphoprotein category (Figures 17G–I). This observation, not typical of proteomic analyses, could reflect the growing consensus that many transcription regulators are important in control of ESC cell state. Among the 158 phosphorylated transcription regulators, 41 contained more phosphorylation site identifications in undifferentiated hESCs, 46 contained more phosphorylation site identifications in differentiated hESC-derivatives, and 71 contained a similar number of phosphorylation site identifications in both cell populations.

Most of the transmembrane receptors and predicted extracellular proteins contained more phosphorylation site identifications in either undifferentiated or differentiated hESCs, whereas fewer of these proteins contained a similar
number of phosphorylation site identifications in both cell populations (Figure 17D–I), implying that growth factors, cytokines, their receptors and corresponding signaling pathways could participate in controlling hESC fate. Furthermore, kinases, which are key players in cell signaling, represented the second largest category of known phosphoproteins (Figures 17G–I). Phosphorylation of cytoplasmic, cytoskeletal, and cell adhesion proteins was identified relatively frequently in differentiated derivatives (Figures 17D–I).

**Phosphorylated Transcription Regulators in Undifferentiated hESCs**

The transcription regulator ESG1 (official symbol TLE1) is expressed only in pre-implantation embryos, ESCs, and primordial germ cells. ESG1 is co-expressed with OCT4 and SOX2 in both mouse and human ESCs, suggesting it is a potential pluripotency marker. In addition, SUPT16H and SSRP1 were phosphorylated in undifferentiated hESCs, and are the two subunits of FACT (facilitates chromatin transcription). FACT destabilizes nucleosomes to allow transcription without disruption of the epigenetic state and promotes initiation of DNA replication in the S-phase of the cell cycle. CREBBP has histone acetyltransferase activity. Its mRNA is enriched in undifferentiated hESCs. AKT phosphorylates CREBBP, increasing CREBBP acetyltransferase activity, promoting NF-kB-mediated transcription and enhanced cell survival. Furthermore, CREBBP increases ERK1 expression. ERK1 activity contributes to hESC self-renewal in the presence of bFGF. 
At least 18 phosphorylated transcription regulators identified in undifferentiated hESCs can modify chromatin structure via histone methylation or acetylation and may contribute to the epigenetic pattern that is likely to be important to hESCs\textsuperscript{148,178,179}. We identified phosphorylation of DNMT3B, MBD3 and EZH2 in undifferentiated hESCs. \textit{DNMT3B} encodes a DNA-methyltransferase, which was expressed in all 59 hESC lines tested\textsuperscript{169}, was enriched in undifferentiated hESCs\textsuperscript{29} and was phosphorylated in undifferentiated hESCs. Differential phosphorylation could modulate EZH2 activity. Phosphorylation at S21 by AKT inhibits the histone H3 Lys27 methyltransferase activity of EZH2, and we identified a novel phosphorylation site of EZH2 in undifferentiated hESCs (S371 or T372), a site whose phosphorylation was also identified in undifferentiated mouse ESCs (unpublished data).

Phosphorylated transcription regulators in undifferentiated hESCs can participate in transcriptional activation or repression, histone modification and more. These and other functions may be integrated to favor the undifferentiated state of hESCs, as implied by the complexity of the phosphoproteome (Figure 17). Although some of these transcriptional and epigenetic regulators were previously reported to influence hESCs, the mechanisms are unclear. The identified phosphorylation sites provide focused information for future studies of the function of these factors in hESCs. Furthermore, we also identified hundreds of novel phosphoproteins whose presence in hESCs was unknown, providing a
rich resource for further investigation. For instance, TNRC6A, a factor for gene silencing via RNA interference\textsuperscript{180} was phosphorylated in undifferentiated hESCs.

**Growth Factor-Mediated Signaling Pathways in Undifferentiated hESCs**

Tyrosine phosphorylation, which plays a dominant role in growth factor/RTK signaling pathways\textsuperscript{167}, was relatively prominent in undifferentiated hESCs (Figure 17C). Signaling pathways participating in self-renewal of hESCs include bFGF-, TGF-β/Activin-, Insulin/IGF-, EGFR family-, PDGF-, Wnt-, Neurotrophin-, Integrin- and Notch pathways\textsuperscript{31,33,34,38-40,181}. However, detailed understanding of the action of these pathways is lacking. The phosphoproteins were grouped into signaling pathways to further explore their functional potential. 41 canonical and metabolic pathways were suggested using the phosphoproteins as input for pathway analysis (not shown). Proteins in RTK pathways were phosphorylated in undifferentiated hESCs, including the adaptors GAB1, SHC1 and NCK2, the kinases LCK, NEK4, MAPK6, MAPK7, mTOR, PIK3C3 and PIK3R4, phospholipases PLC-γ1 and PLC-γ2 and the phosphatase PPAP2B. Some phosphoproteins are shared among pathways and some are more pathway-specific, such as APC in Wnt signaling and NUMB in Notch signaling. Figure 17C imply that a variety of signaling pathways are important in undifferentiated hESCs. For example, EGF pathway members ErbB2, AREG and EPS15L1 were phosphorylated in undifferentiated hESCs, complementing a report showing that the ErbB2/ErbB3 ligand heregulin-1β helps support
undifferentiated hESCs. KDR (VEGFR2, FLK1) was phosphorylated in undifferentiated hESCs, and stimulation of hESCs with CM elicits tyrosine phosphorylation (site(s) undefined) of PDGFRα. Components of the VEGF and PDGF pathways were phosphorylated in undifferentiated hESCs. We also identified phosphoproteins from signaling pathways whose presence in hESCs have not been reported, and a large number of novel phosphoproteins were identified.

Molecular profiling studies typically lack biological follow-up, e.g.\textsuperscript{28,29,71,79,81,88,164,165,168,171,178,179}. However, a few, including transcriptomic\textsuperscript{182} and proteomic\textsuperscript{31,37,38,146,147} studies demonstrated that cells responded to stimulation in manners consistent with molecular profiles. To test the cellular relevance of the phosphoproteomic and pathway analyses, we began by targeting EGF-, VEGF-and PDGF pathways in undifferentiated hESCs using inhibitors of their receptors. Although specificity of RTK inhibitors is imperfect, we used some of the widely accepted. Treatment of undifferentiated hESC cultures with an EGFR inhibitor at 10µM resulted in extensive apoptosis (not shown), similar to another report\textsuperscript{38}. The hESCs were also treated with 10µM KDR inhibitor II or 10µM Gleevec, a PDGFRα inhibitor\textsuperscript{183}. Undifferentiated control colonies were compact and expressed OCT4 and SSEA-4 (Figure \textsuperscript{18B} and not shown). In contrast, most cells differentiated in the presence of KDR or PDGFR inhibitor, shown by flattening of the colonies, altered cellular morphology and nearly undetectable OCT4 and SSEA-4 (Figure \textsuperscript{18C} and data not shown). Vehicle-only
controls lacked any noticeable effect on the cells (Figure 18B). The results were similar under feeder-free conditions in CM and feeder-free conditions in chemically defined media (CDM). Furthermore, KDR or PDGFR inhibitor, at 10 µM, resulted in decreased expression of NANOG and OCT4 (Figure 18A).

To further investigate the effect of RTK signaling pathways, we decreased bFGF to a sub-threshold 4 ng/ml (at least 20 ng/ml is required under feeder free conditions in CDM and systematically supplemented cultures with EGF, PDGF-AA or VEGF-AA at different concentrations to determine which trophic factor could complement bFGF deficiency. Although PDGF-AA without bFGF was unable to maintain long-term cultures of undifferentiated hESCs, PDGF-AA at 10 ng/ml and the sub-threshold concentration of 4 ng/ml of bFGF (subsequently abbreviated PDGF/bFGF) stably maintained undifferentiated hESCs under feeder-free conditions in CDM for >15 passages, and the hESCs remained undifferentiated throughout all 4 experiments (Figure 19D). The cells displayed undifferentiated morphology and robust expression of OCT4. In contrast, when undifferentiated hESCs, which had been stably maintained in CDM containing PDGF/bFGF for >15 passages were subsequently cultured for 4 days in CDM containing 4 ng/ml of bFGF but no PDGF, the cells differentiated (Figure 19B). FACS analyses demonstrated that ca. 89% of the hESCs in CDM containing PDGF/bFGF were positive for SSEA-4, comparable to cultures in CDM containing 20 ng/ml of bFGF (86%; Figure 19). Similar FACS results were obtained when cells were stained and sorted for the pluripotency marker Tra-1-
Moreover, PDGF/bFGF in CDM resulted in sustained expression of NANOG and OCT4 transcripts, whereas their abundance declined within 4 d in the absence of PDGF-AA or the presence of the PDGFR inhibitor Gleevec (Figure 19), further supporting the proposal that PDGF-AA facilitates maintenance of undifferentiated hESCs. Together, phosphoproteomic and pathway analyses suggested that PDGF should favor maintenance of undifferentiated hESCs. PDGFR inhibitor, and separate use of PDGF-AA, provided clear evidence that PDGF, when bFGF is at a sub-threshold concentration, can promote the undifferentiated state of hESCs in CDM under feeder-free conditions, insights that derived directly from the phosphoproteomic analysis.

Our data further suggested that ErbB and VEGFR activation participate in maintenance of undifferentiated hESCs, because disruption of these pathways caused either apoptosis (not shown) and/or differentiation (Figure 19) (although EGF and VEGF-AA demonstrated limited efficacy at complementing the deficiencies of 4 ng/ml bFGF). The ErbB2/ErbB3 ligand heregulin-1β contributes to maintenance of undifferentiated hESCs\(^{38}\). In addition, insulin/IGF pathway members\(^{31}\) were phosphorylated in hESCs (including proteins in the PI3K/AKT/mTOR pathway).

Phosphoproteomics, cellular assays and other reports\(^{31,38,40}\), suggest that multiple RTK pathways are required, although none of them alone is sufficient to support self-renewal in the absence of bFGF. Also consistent with our results,
although less clear due to the undefined media that was used, Sphingosine-1-phosphate plus PDGF contributes to maintenance of undifferentiated hESCs in the presence of mouse embryonic fibroblasts (MEFs) or MEF-conditioned media. It previously appeared that bFGF alone might sustain self-renewal of hESCs. However, as predicted by our phosphoproteomic analysis, several other factors that exist in serum and/or are secreted by feeders, acting through autocrine or paracrine effects or as culture additives, are also important for hESC self-renewal. Our phosphoproteomic and pathway analysis also imply that additional pathways could favor undifferentiated hESCs.

**Phosphorylated Signal Transduction Proteins in Undifferentiated hESCs**

PI3K signaling facilitates ESC self-renewal, and the PI3K pathway is activated by PDGF in mesenchymal stem cells, but the mechanism of action of the PI3K pathway has been unclear. PI3K/AKT/mTOR pathway members were phosphorylated in undifferentiated hESCs. For example, PIK3C3 is enriched in undifferentiated hESCs and PIK3C3 was phosphorylated in undifferentiated hESCs. mTOR plays a role in proliferation of undifferentiated hESCs and is phosphorylated at Ser2448 during mitogenic stimulation. mTOR, phosphorylated at Ser2448 and Ser2454 in undifferentiated hESCs is a protein that enhances cell survival. TSC1 was also phosphorylated in undifferentiated hESCs. TSC1 can limit cell size and its over-expression caused cells to form compact clusters with increased re-aggregation in vitro, similar to the small
size of undifferentiated hESCs and compact morphology of hESC colonies. Phosphorylated PI3K/AKT/mTOR pathway members in undifferentiated hESCs suggest which pathway members may regulate undifferentiated hESCs.

Phosphoproteins participating in MAPK signaling were identified. The ERK pathway contributes to hESC self-renewal under conditions that include bFGF\textsuperscript{177}, whereas JNK signaling in hESCs has not been reported. Some phosphoproteins downstream of RTK pathways are relatively specific to JNK signaling, such as TRAF4, MLK4, CRKL and MINK1. To test for JNK signaling in undifferentiated hESCs, we tested two JNK inhibitors in hESC cultures under feeder free conditions in CM. JNK inhibitor II, a small molecule (SP600125), widely used in JNK studies\textsuperscript{188-190} and JNK inhibitor III, a polypeptide\textsuperscript{191} were used. Each inhibitor, alone, resulted in cellular differentiation, demonstrated by colony morphology and decreased OCT4 expression. In contrast, controls lacking JNK inhibitors, including vehicle-only controls, remained undifferentiated. Induction of differentiation by JNK inhibitors was similar under feeder-free conditions in CDM. Furthermore, OCT4 and NANOG mRNA was depleted in the presence of JNK inhibitor II (Figure 19A). Thus, this phosphoproteomic analysis provides the first suggestion that JNK, an important signal transduction protein downstream of many RTKs, may facilitate maintenance of undifferentiated hESCs. Moreover, these experiments further demonstrate agreement between phosphoproteomic and cellular analyses in hESCs.
Analysis of molecular mechanisms underlying hESC properties is essential for optimal use of these cells. Complementing previous analyses of promoters, transcripts, and protein expression, our phosphoproteomic analysis suggests that multiple protein phosphorylation events participate in control of hESC fate. Application of MDLC-MS/MS-based phosphoproteomics to pluripotent cells may represent an important tool for stem cell biologists. While this study focused on its use for hESCs, one can envision its application to induced pluripotent somatic cells and other somatic stem cells.

Our phosphoproteomic analyses identified known and novel proteins potentially participating in self-renewal or differentiation of hESCs and focused attention on pathways heretofore under-appreciated and under-explored. Transcription regulators, including epigenetic and transcription factors, and kinases contained many phosphorylated members, suggesting that these proteins may be key determinants of hESC fate decisions. Although a variety of proteins have been implicated in hESC self-renewal, some of their functions have been unclear. The identified phosphorylation sites, some on central signaling proteins, expand the knowledge of protein phosphorylation in hESCs. We also identified many proteins whose potential functions in hESCs is novel. In other words, phosphoproteomic analyses may provide guidance for a systematic, rather than solely serendipitous or overly broad-based approaches in future studies of self-renewal and differentiation of pluripotent cells.
Phosphoproteomic analyses identified proteins favoring an undifferentiated or differentiated state of hESCs. For example, phosphorylation of proteins in the JNK pathway was identified, and our cellular follow-up experiments, which are atypical of molecular profiling studies, suggested that inhibition of JNK leads to differentiation of hESCs. A role of JNK in undifferentiated hESCs has not been reported. The VEGF and PDGF pathways are candidates to favor maintenance of undifferentiated hESCs because inhibitors of their receptors resulted in hESC differentiation. However, the growth factors that were added singly could not replace bFGF. Together, these results suggested that activation of these pathways are necessary but not sufficient to sustain self-renewal of hESCs, consistent with increasing evidence that multiple growth-factor driven pathways act together to maintain undifferentiated hESCs. For example, PDGF-AA complemented a sub-threshold concentration of bFGF, shown by long-term maintenance of undifferentiated cultures under feeder-free conditions in CDM. Use of CDM allowed improved knowledge of the composition of the media, rather than use of undefined media in the presence of, or conditioned by, feeder fibroblasts\textsuperscript{40}, so the pathways that were targeted in our cellular assays were more clearly defined. Together, our results expanded the repertoire of pathways that facilitate hESC culture and support the suggestion that multiple signaling inputs are needed to maintain undifferentiated hESCs\textsuperscript{38}. Moreover, phosphoproteomic analyses complement epigenetics, gene
expression profiles, and total protein MS to facilitate an improved understanding of hESC fate determination.

The functions of most phosphorylated proteins are unknown in pluripotent cells, including the plethora of novel ones, and should be evaluated for their influence on stem cell behavior. Application of further advances in proteomic and allied technologies should enhance future studies through improved analysis of protein phosphorylation. As phosphoproteins controlling pluripotent behavior are understood better, methods for developing model systems with stem cells, and potential therapeutic applications, may become increasingly clear.

Methods

Cell Culture, Phosphoproteomic Analysis

Feeder free cultures were in Matrigel coated plates in CM containing 8 ng/ml bFGF\textsuperscript{192}. Differentiation was with 5 µM RA and no added bFGF. In CDM, hESCs were cultured in Matrigel-coated plates in N2/B27-CDM\textsuperscript{40}. Phosphoproteomic analyses used cells from CM. Cells were rinsed with PBS, lysed, centrifuged, proteins precipitated with (NH\textsubscript{4})\textsubscript{2}SO\textsubscript{4} and pelleted by centrifugation.

Proteins were re-suspended in 100 mM NH\textsubscript{4}HCO\textsubscript{3}, 8 M urea containing phosphatase inhibitors, reduced, alkylated, digested with trypsin, and peptides desalted. Peptides were separated by SCX, phosphopeptides enriched by desalt-IMAC\textsuperscript{71,79}, separated by nanoflow HPLC and analyzed by ESI-MS/MS. MS/MS
spectra were matched to amino acid sequences using SEQUEST. All reported phosphopeptide identifications were manually verified\textsuperscript{71,79,148}.

Normalized XIC peak areas of some phosphopeptides were quantified. For analyses lacking the identification, MS/MS data was exhaustively searched for the phosphopeptide, which was rarely found via a poor quality MS/MS spectrum, and its XIC peak area was quantified.

Phosphoproteins were classified as containing more phosphorylation site identifications in undifferentiated hESCs or differentiated derivatives, or as containing a similar number of phosphorylation site identifications in the 2 cell populations, as described in the Results section.

**Western Blot Analysis**

Proteins were run on Bis-Tris gels, transferred to PVDF membranes, blocked, and incubated with antibodies recognizing phosphorylation sites identified by MDLC-MS/MS. Anti-GAPDH was the loading control. Membranes were washed, incubated with fluorophore-conjugated secondary antibodies, washed, imaged, and bands quantified according to the manufacturer (LiCor).

**Phosphoprotein category, subcellular location and pathway Analysis**

Ingenuity Pathway Analysis, Metacore, NCBI, Gene Ontology and peer-reviewed literature were used to identify phosphoprotein subcellular location, category, and signaling pathways.
**Cellular Assays, RT-PCR**

EGFR-, JNK- or PDGFR inhibitors were used. Untreated and vehicle-only controls were included for each experiment. PDGF-AA/bFGF were used in cultures for >15 passages.

For immunostaining and DAPI staining, monoclonal mouse anti-OCT4 and anti-SSEA-4 were used. Secondary antibodies were Cy2-conjugated rabbit anti-mouse IgM and Cy3-conjugated rabbit anti-mouse IgG. For RT-PCR, mRNA was isolated, cDNA was synthesized; OCT4, NANOG and GAPDH were amplified. For FACS, cells were incubated with mouse monoclonal anti-SSEA-4 or anti-TRA-1-60 antibodies, washed with PBS and incubated with Cy3-conjugated rabbit anti-mouse IgG.

Chapter 2.2, in part, is a reprint as it appears in “Phosphoproteomic Analysis of Human Embryonic Stem Cells” by Laurence M. Brill, Wen Xiong, Ki-Bum Lee, Scott B. Ficarro, Andrew Crain, Yue Xu, Alexey Terskikh, Evan Y. Snyder, and Sheng Ding published in Cell Stem Cell Aug 7th 2009.
Figure 16. Undifferentiated hESCs Expressed Markers of Pluripotency, whereas the Markers Were Downregulated upon Differentiation. Cells were cultured to yield undifferentiated hESCs (hESCs), or differentiated hESC derivatives (derivs) under feeder-free conditions by withdrawing bFGF and including 5 µM RA in the media for the final 4 days of culture. Nuclei were stained with DAPI (A and B; left column).(A) Cells were stained with antibodies detecting OCT4 (center column), and OCT4 and DAPI images were merged (right column).(B) Cells were stained with antibodies detecting SSEA-4 (center column), and SSEA-4 and DAPI images were merged (right column). All photomicrographs were at the same magnification. The scale bar represents 50 µM.
Figure 17. Number of Protein Phosphorylation Sites and Phosphoproteins Identified in hESCs and Their Differentiated Derivatives, Prominence of Tyrosine Phosphorylation, Predicted Subcellular Location of the Phosphoproteins, and Phosphoprotein Categories

(A) Total number of nonredundant phosphorylation sites and (B) number of proteins with more phosphorylation site identifications in undifferentiated hESCs (line H1/WA01) (represented in red), RA differentiated, H1-hESC derivatives (represented in gold), or with a similar number of phosphorylation site identifications in the two cell populations (represented in gray). The percentage of the phosphorylation sites and phosphoproteins in each of the three groups of proteins is shown in parentheses.

(C) Percentage of nonredundant tyrosine phosphorylation sites, among the sites for which the phosphorylated residue could be defined as serine, threonine, or tyrosine (94% of all sites), that were on proteins containing more identified sites in undifferentiated hESCs, differentiated hESC derivatives, or that were on proteins with a similar number of identified sites between undifferentiated and differentiated cells.

(D–F) The subcellular localization of the phosphoproteins is shown; those widely associated with more than one subcellular location are designated as variable.

(G–I) Phosphoprotein categories, among those whose functions are known, are shown.
Figure 18. Protein Kinase Inhibitors Resulted in Differentiation of hESCs (A) Expression of NANO\(G\) (Chambers et al., 2003) and OCT4 mRNAs was assessed by RT-PCR, in the presence of protein kinase inhibitors that resulted in differentiation of hESCs. Cells were cultured with 20 ng/ml of bFGF, and inhibitors (10 \(\mu\)M) were included in the cultures for the final 4 days. Inhibitor identities are indicated in the figure. Slower decline of OCT4 than NANO\(G\) was typically observed during hESC differentiation. GAPDH was an internal control. (B and C) Undifferentiated, vehicle-only control (B) and differentiated, KDR inhibitor-treated (C) cells are shown under imaging conditions indicated above the columns. All photomicrographs were at the same magnification, and the scale bar (bottom right) represents 50 \(\mu\)M. Abbreviations include the following: i, inhibitor; uhESCs, undifferentiated hESCs.
Figure 19. PDGF and a Subthreshold Concentration of bFGF Sustained Long-Term Culture of hESCs

(A) RT-PCR to amplify *NANOG* and *OCT4* transcripts in long-term hESC cultures (>15 passages) in CDM containing 10 ng/ml of PDGF-AA and 4 ng/ml of bFGF (lane PDGF, bFGF4). Lanes bFGF20 or bFGF4 refer to 20 or 4 ng/ml of bFGF in the CDM for 4 days, respectively, in the absence of PDGF, following culture in 10 ng/ml of PDGF-AA and 4 ng/ml of bFGF for >15 passages.

(B–D) Colony morphology, *OCT4* staining, and fluorescence-activated cell sorting (FACS) demonstrated that PDGF/bFGF in CDM maintained undifferentiated hESCs passaged >15 times. Imaging conditions or FACS analyses of SSEA-4 expression, detected via Cy3-conjugated secondary antibodies, is indicated above the columns, and the culture additives that were varied are indicated beside the rows. In FACS plots, dotted lines delineate boundaries of fluorescence intensity approximately indicative of cellular identity as undifferentiated hESCs (uhESC) and differentiated hESC derivatives (deriv). Decline of SSEA-4 is incomplete in differentiated hESCs after 4 days (Figure 16). Following maintenance of the hESCs in CDM containing bFGF at 4 ng/ml and PDGF-AA at 10 ng/ml for >15 passages, cells were cultured for 4 days in CDM lacking PDGF and containing bFGF at 4 ng/ml (B) or 20 ng/ml (C), or in the continued presence of bFGF at 4 ng/ml and PDGF-AA at 10 ng/ml (D). All photomicrographs were at the same magnification, and the scale bar (bottom center panel) represents 100 µM (B–D).
FIGURE 20. Semi-Quantitative Western Blot Analysis, using Antibodies that Recognize Specific Protein Phosphorylation Sites, Supported Phosphoproteomic Results using MDLC-MS/MS. Proteins were prepared from undifferentiated hESCs (“uhESC”) or their differentiated derivatives (“deriv”), and Western blot analysis, including normalized, integrated intensities of the bands, were performed. Protein molecular mass standards, in kilodaltons (kDa) are shown (A-D). Triplicate lanes contained proteins from either undifferentiated hESCs or derivative cell populations. Each experimental antibody recognizes protein phosphorylation site(s) identified in the MDLC-MS/MS analysis reported in this paper. Integrated intensities of experimental bands were normalized by dividing their integrated intensities by the integrated intensity of the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) loading control (*). Normalized integrated mean intensities (percentage of the GAPDH control), +/- the standard deviation is shown above the blots. Arrowheads mark specific experimental bands recognized by the antibodies and their normalized, integrated intensity (A-D). The arrow (B) marks a band putatively containing phospho-PAK2, also recognized by the antibody recognizing phospho-PAK1, and its normalized, integrated intensity. The target protein phosphorylation site of each antibody is indicated at the top of each panel, and were mTOR (official symbol FRAP1) phosphorylated on Ser2448 (A); PAK1 phosphorylated on Ser144 (B); PTK2 phosphorylated on Tyr576/Tyr577 (C); and CDK1/2/3/5 phosphorylated on Thr14 and Tyr15 (D).
References


21. Trapp C, McCullough AK, Epe B. The basal levels of 8-oxoG and other oxidative modifications in intact mitochondrial DNA are low even in repair-deficient (Ogg1(-/-)/Csb(-/-)) mice. Mutat Res 2007;625:155-63.


119. O'Brien RN, Shen Z, Tachikawa K, Lee PA, Briggs SP. Quantitative proteome analysis of pluripotent cells by iTRAQ mass tagging reveals post-


