Title
Analysis of pluripotent mouse stem cell proteomes : insights into post- transcriptional regulation of pluripotency and differentiation

Permalink
https://escholarship.org/uc/item/5wb983jr

Author
O'Brien, Robert Norman

Publication Date
2010

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Analysis of pluripotent mouse stem cell proteomes: insights into post-transcriptional regulation of pluripotency and differentiation

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

in

Biology

by

Robert Norman O'Brien

Committee in charge:

      Professor Steven P. Briggs, Chair
      Professor Lawrence Goldstein
      Professor Cornelius Murre
      Professor Karl Willert
      Professor Yang Xu

2010

The Dissertation of Robert Norman O'Brien is approved, and it is acceptable in

quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
                                                                Chair

University of California, San Diego

2010

DEDICATION

To all of my friends and family who made this possible. I can't thank all of you

enough.

I especially dedicate this to Chi, who has been my companion, friend and love

throughout my time in graduate school: your unwavering support was essential in

getting me to this point.

EPIGRAPH

My feet tug at the floor
And my head sways to my shoulder
Sometimes when I watch trees sway,
From the window or the door.
I shall set forth for somewhere,
I shall make the reckless choice
Some day when they are in voice
And tossing so as to scare
The white clouds over them on.
I shall have less to say,
But I shall be gone.

Robert Frost
The sound of the trees

TABLE OF CONTENTS

Chapter 3: Phosphoproteome analysis identifies phosphorylation of pluripotency associated proteins 73

Chapter 4: Analysis of the phosphoproteome of mouse ES cells identifies increases in phosphorylation of mRNA splicing factors during differentiation

LIST OF ABBREVIATIONS

Abbreviation	Definition
ATRA	All-trans retinoic acid
CHX	Cycloheximide
DBD	DNA binding domain
EC	Embryonal carcinoma
ECC	Embryonal carcinoma cell
ES	Embryonic stem
ESC	Embryonic stem cell
FDR	False discovery rate
hESC	Human Embryonic stem cell
ICM	Inner cell mass
IF	Immunofluorescence
LIF	Leukemia inhibitory factor
mESC	Mouse Embryonic stem cell
RT-PCR	Reverse transcriptase polymerase chain reaction
SR proteins	Serine/arginine rich splicing factors

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I want to thank my friends and co-workers who always had time for a beer, a coffee, a bike ride or a death-defying climbing trip, so thanks to Stephane Bellafiore, Zhouxin Shen, Chris van Schie, Kiyoshi Tachikawa, Amanda Mason, Kelly Lagor, Sarah and Anthony Carroll, Dan Anderson and Anne Ruminski, Siler Panowski, Jenni Duriex, Jose Maraches, Kendra Hogan, Aaron Parker, Xianquan Li, Kitty Cheung, Stevan Djachovic, Nicole Orazio, Samantha Serey, Rebecca Sanders, Jeff Keil and everyone else whose support and friendship has meant so much to me over the last six years.

I especially want to thank my mentor and dissertation advisor, Dr. Steve Briggs whose knowledge and insights were instrumental in helping me turn data into a story. None of this work could have happened without his guidance and contributions.

Likewise, none of this work would have been possible without the technical wizardry of Dr. Zhouxin Shen. Dr. Shen developed the protocols and ran all the samples for mass spectrometry. I am indebted to him for his friendship and cheerful assistance in understanding and interpreting the data presented here.

Chapter 2 includes data submitted for publication at Molecular and Cellular Proteomics (published by the American Society for Biochemistry and Molecular Biology), and was co-authored by Zhouxin Shen, Kiyoshi Tachikawa, Angel Lee and Steven Briggs. The dissertation author was the primary investigator and author of this material.

Chapter 3 includes work that is still in preparation for publication and includes contributions from Zhouxin Shen, Kiyoshi Tachikawa, Loes Drenthe, Susanne Kooistra, Bart Eggan and Steven Briggs. The dissertation author was the primary investigator and author of this material.

Chapter 4 includes contributions from Zhouxin Shen and Steven Briggs. The dissertation author was the primary investigator and author of this material.

2010	Doctor of Philosophy, Biological Sciences, University of California, San Diego
2004	Bachelor of Science, Biological Sciences, University of Vermont
2000	High school diploma, Rice memorial high school

PUBLICATIONS

O'Brien, R. N., Shen, Z., Tachikawa, K., Briggs, S.P. Quantitative proteome analysis of pluripotent cells by iTRAQ mass tagging identifies post-transcriptional regulation of proteins required for ES cell self-renewal. Submitted to Molecular Cellular Proteomics. January 2010, revised, May 2010.

O'Brien, R. N., Shen, Z., Drenthe, L, Kooistra S. L. Tachikawa, K., Eggan, B. E. Briggs, S. P. Quantitative phosphoproteome analysis of pluripotent cells identifies phosphorylation of pluripotency-associated factors. In preparation.

TEACHING EXPIRIENCE

BIBC 103 Biochemistry Techniques Teaching Assistant
Professor Aresa Touqdarian, University of California, San Diego, Spring 2006

BICD 134 Human Reproduction and Development Teaching Assistant
Professor Ramon Pinon, University of California, San Diego, Spring 2007

BIMM 100 Molecular Biology Teaching Assistant
Professor Aubrey Davis, University of California, San Diego, Fall 2007

ABSTRACT OF THE DISSERTATION

Analysis of pluripotent mouse stem cell proteomes: insights into post-transcriptional regulation of pluripotency and differentiation

by

Robert Norman O'Brien

Doctor of Philosophy in Biology

University of California, San Diego, 2010

Professor Steven P. Briggs, Chair

Placental mammals begin life as free-living embryos that implant into the uterus. Before implantation, the embryo differentiates into two lineages, the trophectoderm and the inner cell mass (ICM). The ICM contains undifferentiated cells that give rise the adult animal. This ability to produce all the cells of the adult animal is called pluripotency.

This dissertation consists of work to understand the molecular makeup of pluripotent cells. Previous work on mRNA content of pluripotent cells lead to

important discoveries of factors necessary and sufficient for pluripotency. I reasoned that a quantitative analysis of the proteome of pluripotent cells would be useful in confirming mRNA observations, identifying protein markers of the pluripotent and differentiated state and identifying post-transcriptional regulation. This thesis describes my work toward that goal.

I describe the quantitative proteomes of two systems also used for transcriptome studies of pluriopotent cells: mouse ES cells and mouse EC cells. I compared of the two protein datasets to each other to identify conserved proteins associated with pluripotency and to mRNA datasets to identify putative cases of post-transcriptional regulation. I confirmed putative cases of post-transcriptional regulation by western blot and RT-PCR, and confirmed the role of one of these proteins in ES cell colony formation by knockdown.

I describe my semi-quantitative analyses of the phosphoproteomes of the same cell types, allowing me to identify phoshoproteins associated with the undifferentiated and differentiated states. I tested the phosphorylation of the transcription factor UTF1 for changes in the protein's function, and found no difference between wild-type and phosphomutant proteins.

Finally, I describe my work to combine the two datasets in order to identify changes in phosphorylation that are independent of changes in underlying protein levels. I identified factors associated with alternative splicing that become phosphorylated after differentiation. I inhibited a candidate kinase, Clk1, that phosphorylates splicing factors and performed a phosphoproteome on the inhibited cells.

These studies represent a deep, quantitative representation of the protein makeup of pluripotent cells and result in new markers of pluripotency and differentiation as well as new phosphorylation events on factors known to be necessary and sufficient for pluripotency.

Chapter 1: Introduction

1.1 Introduction

Multicellular organisms develop from a single-cell, the zygote. The specifics of this growth and development are quite variable among different phyla, classes and orders of organisms. Placental mammals are unusual in that their development begins in a free-living state in the fallopian tube, but then becomes dependent on the supportive environment of the uterus after implantation[1]. Understanding pre-implantation and post-implantation development is an important focus of basic and clinical research in order to understand the process of normal development and the ways in which this process can go wrong.

1.2 Overview of early mammalian development

The overall question about mammalian development remains: how does a zygote organize and proliferate to give rise to a full organism? The general physiology and anatomy of this development is known: the zygote divides several times to form a morula, which differentiates into a blastocyst comprising an outer cell mass (trophectoderm, TE) and an inner cell mass (ICM), which will give rise to the organism proper[1]. After hatching from the zona pelucida the blastocyst implants and some of the cells within the ICM differentiate into the primitive endoderm while the remaining ICM cells will ultimately differentiate into three primordial germ layers that give rise to the three major tissue lineages: ectoderm

(skin and neural tissue), mesoderm (blood and muscle) and endoderm (vasculature and most organs)[1].

1.3 Stem cell biology

To understand the molecular basis of how the inner cell mass is able to give rise to the three germ layers, ICMs have been grown *in vitro* resulting in a "useful artifact" known as embryonic stem cells (ESCs)[2-4]. Human and mouse ESCs retain the pluripotent developmental potential of the ICM and also display remarkably extensive potential for self-renewal when grown in conditions permissive to self-renewal[5].

The definition of the term "stem cell" is generally accepted to cover a cell that divides (or has the potential to divide given the correct environmental signals) *and* has the potential to gives rise to either a daughter cell like itself (self-renewal) or gives rise to a cell different from itself in some way (differentiation). A self-renewing cell may not have the potential to differentiate (transformed cell lines for example) and a differentiating cell may not be capable of self-renewal (progenators), therefore a stem cell is characterized by possessing the twin properties of self-renewal and developmental potential[6, 7].

1.4 Therapeutic uses of adult stem cells

Stem cells have been used therapeutically since the 1970's to treat immunodeficiency and leukemia[8, 9], a disease that has been recognized as a stem cell-based disease for more than 40 years[10-12]. Stem cells from mature organisms have been identified and characterized in vitro and in vivo and include

hematopoetic stem cells (HSCs)[13], neural stem cells (NSCs)[14-18], and Mesenchymal stem cells (MSCs)[19, 20]. Derivation and *in vitro* culture and expansion of most adult stem cells (ASCs) has been problematic since adult stem cells reside in a tightly regulated niche that maintains AS cells and limits their proliferation to levels necessary to replace tissue over time[21-24]. Therefore therapeutic applications that ASCs might otherwise be suited for have, with the notable exception of HSCs (via bone marrow transplant), failed to materialize.

1.5 Cancerous stem cells: embryonal carcinoma

The concept of cancer stem cells is a relatively old one that has enjoyed a revival in the last few years[25]. Cancer stem cells remain a nebulous term, but generally refer to the tumorogenic subpopulation of cells "tumor initiating cells" within a tumor that are actively dividing and building tumor mass[26, 27]. In some cases, these tumorogenic stem cells are likely to be somatic cells that have misregulated protooncogenes/tumor suppressor genes[28, 29], but in some cases these are true stem cells that have acquired cancerous properties due to a loss of sensitivity to the local environment of the niche. These include some leukemic stem cells[30] and embryonal carcinoma cells (ECCs)[31]. EC cells are cancerous, malignant germ line tumor cells that arise in the testis and like ES cells, they are pluripotent or multipotent *in vitro* and *in vivo*[32-35] and before the derivation of mouse and human ES cells, they were used as models of preimplantation development, and remain in use today[10, 35-37].

1.6 Embryonic stem cells

Embryonic stem cells are unusual because unlike most ASCs, they proliferate rapidly[6] and do so with very low rates of mutation relative to somatic cells[38]. Despite the fact that ES cells are euploid and untransformed, they are tumorogenic in allogenic transplants[39-41] suggesting that they are not subject to the same tight environmental regulation that characterizes untransformed somatic cells and ASCs. Clinicians find ESCs to be exciting because they are in addition to being capable of extensive expansion *ex vivo*, they are relatively easily manipulated *in vitro* to differentiate into specific lineages[42-48] making them therapeutically exciting as 1) potential sources of replacement tissues[49] and 2) potential in vitro models of developmental and disease states[50].

The undifferentiated ES cell state itself is interesting because it is a non-cancerous, euploid cell type that proliferates rapidly, extensively and faithfully[6, 38]. How are ES cells able to do this while all other cells do not, while maintaining extensive developmental potential?

1.7 Molecular basis of pluripotency: transcription factors

Much is already known about the molecular basis of the pluripotent state, including the transcriptional and signaling programs that set up and maintain the pluripotent ES cell state in mouse and human ES cells. In both species, the transcription factors necessary to maintain the undifferentiated ES state include the Oct4[51-54], Sox2[55, 56] and Nanog[57-59]. These three proteins form transcriptional feedback network that regulate themselves and downstream transcription factors[60-63] that are also necessary for maintaining the ES cell state.

1.8 Molecular basis of pluripotency: chromatin remodeling

A functionally related class of proteins are chromatin modifying enzymes that are specific to the pluripotent state. These include the polycomb complex proteins including Suz12[64-67], the de novo DNA methyltransferases Dnmt3b and Dnmt3a[68-72] as well as the general repressors Ronin (Thap11)[73] and UTF1[74]. All of these proteins are expressed at high levels in pluripotent cells and are necessary for either ES cell self-renewal or pluripotency presumably by maintaining the chromatin in a configuration compatible with the transcription factors' targets and repressing factors that specify the various differentiated lineages.

1.9 Molecular basis of pluripotency: signaling pathways

The other broad category of molecular determinants of pluripotency are comprised of the signaling pathways that maintain the undifferentiated state as well as the pathways that promote exit from the undifferentiated state. Unlike transcription and chromatin modifying proteins, the signaling pathways that maintain human and mouse ES cells are slightly different. Both mouse and human ES cells can be maintained on a feeder layer comprised of mouse embryonic fibroblasts, however mouse ES cells grown in the absence of feeders are LIF[75-79] and Bmp4[80] dependent.

Human ES cells grown in the absence of feeders require an extracellular matrix such as matrigel or Laminin-511[81] as well as extrinsic bFGF (FGF2) that interacts with hES derived fibroblasts to support self-renewal.[82]

The undifferentiated state of both human and mouse ES cells is maintained by several other signaling pathways that are shared in common.

These include embryonic cadherin (eCadherin)-Integrin-Beta catenin[83] and the

PI3K-AKT[84] signaling pathways.

1.10 Reprogramming somatic cells

Entry of cells into a pluripotent state is usually a unidirectional

phenomenon; that is, an egg is fertilized and gives rise to Oct4 postitive,

pluripotent cells that organize into a blastocyst, implant and differentiate into the

three somatic germ layers and the germline cells which will start the process

again in the next generation. Lentiviral transduction of Oct4 and Sox2 transgenes

is sufficient to reprogram human and mouse somatic cells into ES-like induced

pluripotent stem cells (iPSCs)[85-87]. There have been no studies demonstrating

that signaling factors or environmental conditions alone are sufficient to set up

the induced pluripotent state, suggesting that the transcriptional networks are

central factors in maintaining the pluripotent state.

1.11 Neural differentiation

In vitro methods of recapitulating embryonic development in Embryonic

stem cells or embryonal carcimoma cells often times use aggregation and, in the

case of ES cells growth factor withdrawal as a method to begin the cells' exit

from the pluroptent state. The resulting cell masses are known as embryoid

bodies, EBs. EB formation results in a mixed population of cells that are sensitive

to growth factors, serum and other environmental factors. One of the most

common methods to push pluripotent cells into a neural lineage is the use ofAll-

Trans Retinoic Acid (ATRA)[42, 88-92], a metabolite necessary for many aspects of

embryonic development, including neural, limb bud, cardiac and retinal differentiation. Treatment of EBs derived from ES cells or P19 EC cells with ATRA results in a shift toward the neural lineage. Indeed, P19 cells are so predisposed to enter the neural lineage, 4-day old EBs grown in 1µM ATRA plated on a solid matrix will begin to make neural processes after 24 hours[93-95]. ES cells are less poised to enter the neural lineage and require more time and often supplemental growth factors in order to efficiently develop into cells of the neural lineage[96-99].

1.12 Systems biology approach to understanding stem cells

Embryonic Stem cells are ideal candidates for study by a systems biology approach. Systems Biology is the analysis of large datasets; ideally a comprehensive inventory and quantitation of a class of molecules found in a particular cell type (commonly appended with the suffix "–ome"). The systems biology approach does not rely on mutants or screens for biological insights, but rather on data from changes in levels and composition of molecules of interest. The best known and most widely used systems approach is the study of the transcriptome[100]; the study of messanger RNA levels in cells of interest. This approach has been instrumental in discovering genes associated with, and therefore good candidates to test for roles in reprogramming[85, 86, 101] and lineage specific differentiation[102] .

Transcript-level analysis of changes in gene expression have been extremely valuable to stem cell biologists, however the vast majority of enzymatic, structural and regulatory functions of the cell occur are performed not

by messenger RNA, but by proteins. Regulation of protein translation, stability and modifications are all invisible to biologists studying the transcriptome, therefore I set out to generate a deep, quantitative inventory of the proteins in a pluripotent cell and to identify those proteins and modifications that are specific to the undifferentiated pluripotent state.

1.13 Proteomics approach to systems biology

Proteomics is a quickly developing field, but is still not as mature as studies of the transcriptome. This is because there are many methods and instruments that can be used to identify and quantify proteins, and not all are suited to rapid, quantitative identification of many different proteins[103, 104].

The central technology at the heart of proteomics is mass-spectrometry, a technology that measures the mass-to-charge ratio (called m/z) of a molecule. Where a RNA-based analysis can use a microarray and specialized cameras to capture data relating to thousands of mRNAs in parallel, mass spec in a fundamentally linear analytical pipeline. For that reason, measurements of different proteins must occur in a linear, one-at-a-time fashion. Therefore three central issues in mass spectrometry are separation, ionization and detection.

Our lab's approach to these problems is an adaptation of the MudPIT (Multidimensional Protein Identification Technology) technique pioneered by the lab of John Yates[105]. In brief, this technology uses the protease trypsin to digest proteins (soublized in an acid cleavable detergent) into peptides, compatible with an electron spray LTQ Ion Trap mass-spectrometer. The peptides are then loaded onto a C18 (reverse phase) loading column, combined with a strong

cation exchange (SCX) column and separated in two dimensions by HPLC via increasing gradients of organic solvent (formic acid) and salt.

The tip of the column is attached to a positive voltage giving the peptides that exit the column a positive charge. Peptides enter the ion-trap, are selected and are scanned for m/z value (the MS1 scan), fragmented and the fragments re-scanned (MS2 scan) in order to accurately identify the sequence of the peptide.

After the MS/MS data are collected, they are processed by analytical software that matches collected spectra to a database of expected m/z scores for peptides and peptide fragments. As with most bioinformatics tools, this software can be adjusted to be more or less stringent, correspondingly, the rate of false discoveries will also vary. In order to assess the false discovery rate (FDR), a decoy database, composed of peptides from the original database, but whose order has been reversed.

Quantitation of proteins is another difficult problem facing the proteomics community. Measuring protein levels by mass spectrometry is confounded by the fact that different peptides have different properties when ionized, so they "fly" differently in the mass spectrometer. Some labs have chosen to ignore this problem and simply publish inventories of proteins in the cell, while others have begun to address the problem using isotope labeling of peptides from different samples in order to be able to pool the samples and run them in parallel. One method that has been used on ES cells is known a SILaC[104]. This method uses "heavy" media with components composed of stable isotopes of sulfur or nitrogen to grow cells in, with comparator cells grown in normal "light" media. This

approach is decent enough, but the resulting peptides have different m/z scores in the first MS analysis, and are therefore selected and analyzed separately. Also, complete labeling of cells with a stable isotope is next to impossible if cells are to be grown in media containing fetal bovine serum, since FBS contains "light" isotopes of these elements.

Our lab has therefore adopted the iTRAQ reporter system, which uses in vitro labeling of tryptic peptides with an isobaric reporter tag whose net mass is the same before fragmentation (in the MS1 scan), but is different after fragmentation, when the "balancer" fragment is removed from the reporter fragment.

The iTRAQ system represents a significant step forward in relative quantitation of proteins by mass spectrometry, but in order to take advantage of this system, I designed experiments that measured protein levels in the cells of interest (pluripotent ES and EC cells), as well as a comparator cell type. This is conceptually similar to the design of transcriptome studies, which often use comparator cell types in order to relatively quantify transcripts. For this reason, I looked to the literature from transcript-level studies of pluripotent cells for insights, and found a study from Aiba et. al. 2006 of the transcriptome of 129/SvEv ES and P19 EC cells[106] on which I chose to model my own proteome analyses.

Protein redundancy is another problem facing anyone trying to conduct a proteomic study. Tryptic peptides are not always unique to one protein, and can therefore be matched to several different proteins. This problem is especially

difficult when working with quantitative samples, since identical peptides from different proteins can introduce artifacts into quantitation. To address this problem, Dr. Shen adopted a system of binning groups of peptides that are matched the multiple proteins entries into a "protein group" which are represented by a "group leader", is the is the protein with the most peptides matched to it. When proteins are quantified by iTRAQ, the reporter intensities from the peptides that match only that protein group and are not shared by another protein group are summed and used for quantitation.

1.14 Focus of the thesis

My thesis research focuses on understanding the connection between the signaling pathways and transcriptional networks that maintain the ES cells in the undifferentiated state, and how protein expression changes as cells exit the pluripotent state, while looking more closely at regulation of gene products at the post-transcriptional level in order to identify:

- Protein markers of pluripotency

- Cases of post-transcriptional regulation of proteins associated with the pluripotent state

- Direct regulation of factors necessary for pluripotency by kinase signaling pathways via phosphorylation

- Changes that occur to the proteome and phosphoproteome as cells exit the pluripotent state

While much work has been done to describe pluripotency at the level of mRNA, very little work has looked at the global protein makeup of pluripotent

cells before they differentiate. The work that has been done has been of limited utility due to the fact that the authors compiled an inventory of proteins present rather than a robust, quantitative measurement of protein levels[104]. My thesis dissertation will describe my work to compile a robust, quantitative proteome and phosphoproteome analysis of ES cells and their cancerous counterparts, EC cells before and after differentiation into a neural lineage. My experimental design follows an experimental design published by and unaffiliated group which compiled a deep transcriptome analysis[106] of the same cell types differentiating into the same lineages, using the same protocols, allowing me to look at how proteins and phosphoproteins' levels compare to underlying transcripts in the same cell types. After I identified and verified genes and post-transcriptional modifications of interest, I performed functional experiments to better understand the role that these modifications and regulations played in pluripotent cells.

Chapter 2: Large-scale analysis of the proteome of pluripotent mouse stem cells identifies protein markers of the pluripotent state and cases of post-transcriptional regulation

2.1 Summary

Pluripotent cells such as embryonic stem cells (ESCs) and embryonal carcinoma cells (ECCs) possess two characteristics that make them exciting to both the basic science and translational medicine communities: pluripotency and extensive growth potential in an untransformed euploid state. Much has been done to identify and characterize transcription factors that are necessary and sufficient to maintain this state. Proteins like Oct4 and Nanog are necessary for maintaining the pluripotent state and are downregulated at the mRNA level during differentiation. I hypothesized that there may be other factors whose levels are unchanged at the mRNA level, but are destabilized or translationally inhibited, during differentiation of pluripotent cells. I generated deep, semi-quantitative profiles of ES and EC cells during differentiation by treatment with 1 μM all-trans retinoic acid (ATRA), replicating a microarray-based study by Aiba et al[106]. I identified several cases of protein levels decreasing during differentiation that were unchanged at the mRNA level. I confirmed several of these putative cases of post-transcriptional regulation by RT-PCR and western blot. One protein in particular, Racgap1 (also known as mgcRacgap) stood out as interesting since

it has been reported to be pre-implantation lethal[107] and necessary for HSC survival and multilineage differentiation[108]. I confirmed the observation from Aiba et al that Racgap1 is unchanged at the mRNA level during RA-mediated differentiation of ESCs by RT PCR. To confirm my observation of Racgap1 declining during RA-mediated differentiation, I employed a mass-spec based strategy, MRM and determined that indeed, Racgap1 levels decline by 2-fold during RA-mediated differentiation. I obtained a panel of 5 shRNAs to Racgap1 from Openbiosystems (www.openbiosystems.com) and knocked down Racgap1. Knockdown resulted in a statistically significant loss of ES colony forming potential that correlated with the level of knockdown. I conclude that Racgap1 is a post-transcriptionally regulated protein during mouse development that is necessary for ES cell self-renewal.

2.2 Introduction

Embryonic Stem cells (ES) have attracted significant interest from the public and scientific community in recent years. ES cells are interesting due to their pluripotent developmental potential and capability for extensive self-renewal. Pluripotency is the ability to differentiate into cells representing or derived from, any of the three primordial germ layers that make up a post-implantation embryo. This property is shared with a small class of cancers of the testis, embryonal carcinomas (ECs). The extensive self-renewal capability of ES cells is distinct from almost all other euploid, non-cancerous cell types. Many cancer cells are also able to divide extensively, but accumulate genetic mutations at a much greater rate than ES cells[109]. The developmental potential of ES cells

means that they have begun to be used to serve as models of, and potentially, treatments for a variety of human diseases.

Aggregation and treatment of mouse ES and EC cells with retinoic acid (RA) has been shown to increase the proportion of ES and EC cells that differentiate into the neuroectodermal lineage[109, 110], though the rate at which these cells differentiate is quite different[106]. Differentiation by RA treatment and aggregation is a standard first step in many mouse ESC differentiation protocols, and is sufficient to efficiently drive mouse EC cells of line P19 to neural differentiation[106].

The last few years have lead to many advances in the understanding of the undifferentiated state of ES cells. The transcription factor Oct4, also called Oct3/4, was considered to be the key determinant and marker of undifferentiated cells. Oct4 was later found to dimerize with another transcription factor, Sox2 [56, 60]. Around that time, another factor, Nanog was identified as being necessary and sufficient for the undifferentiated ES cell state. More recent work from the Yamanaka and Thompson labs have identified combinations of transcription factors including Oct4, Sox2, Nanog, Lin28, cMyc and Klf4 which are sufficient to reprogram somatic cells into ES-like induced pluripotent stem cells (iPS) cells. While the transcriptional networks that establish and maintain the pluripotent state have been extensively studied, there is relatively little work identifying how signaling from the environment affects cells as they exit the undifferentiated state. Recent publications have identified Caspase 3 as a negative regulator of Nanog and Ronin protein stability in early differentiation, but questions remain

about how ES and EC cells behave at the levels of proteins, that is post-transcription, level during differentiation. Recent publications have suggested that mRNA transcripts in ES cells are subject to extensive post-transcriptional control[111] affecting loading of ribosomes in undifferentiated ES cells resulting in changes in protein levels during differentiation that are not observed at the mRNA levels. I set out to analyze the proteome of ES and EC cells before and after differentiation to identify proteins enriched in each cell type and cases of post-transcriptional regulation.

There have been several other semi-quantitative and non-quantitative proteomic analyses of human and mouse ES cells, One well known paper from Graumann et al. used SILAC to identify and semi-quantify subcellular localization of proteins in undifferentiated cells[104]. More recently Swaney et al. followed the phosphoproteome of human ES cells during differentiation[112, 113]. My study complements and extends these studies by 1) adding a comparator, that is differentiated cells to allow semi-quantitation before and after differentiation 2) including embryonal carcinoma cells, which are pluripotent and not sensitive to the same environmental requirements that ES cells require to maintain self-renewal. 3) Replicates a previous study of the transcriptome to identify putative cases of post-transcriptional regulation in differentiating pluripotent cells.

I used a tandem mass spectrometry (MS/MS) approach to develop a deep, semi-quantitative analysis of the mouse ES and EC proteome before and after differentiation. I designed my experiments to replicate the cell lines and differentiation protocols published by Aiba et al. in 2006, in order to compare my

measurements of the proteome with their measurements of the transcriptome[106].

To achieve the depth and robust quantitation required to compare proteome data

with a microarray dataset I utilized two complementary techniques: MUDPIT on-

line two-dimensional separation of tryptic peptides and *in vitro* iTRAQ mass tag

labeling.

Dr Shen and I employed the MuDPIT technique developed by the Yates

lab[105, 114-119]. The advantages of this technique versus offline techniques include

the ability to separate tryptic peptides in one, two or three dimensions. I used a

two dimensional nanoflow separation method to reproducibly produce deep

proteome surveys of ES cells.

While the MuDPIT method allowed us to identify many proteins in

undifferentiated and differentiated ES cells, I also required a method to quantify

the relative abundance of those proteins in cells before and after differentiation.

To achieve this Dr. Shen and I considered four different mass labeling

techniques: SILAC, iCAT, O^{18}, or iTRAQ labeling. Dr. Shen and I rejected SILAC,

iCAT and O^{18} labeling because all three techniques perform quantifications at the

MS1 stage, meaning that the initial identification of each mass labeled peptide

will result from a different ms1 scan. We also rejected SILAC because of the

limited flexibility of the technique. SILAC requires samples to be grown in labeled

media. This make the technique unsuitable for clinical plant or whole tissue

samples, therefore I chose the iTRAQ reagent. The MS1 and MS2/CID

identifications are the same and only at the MS2/PQD stage will reporter ions be

measured. Another advantage of iTRAQ is that the labeling occurs during sample preparation and is an irreversible chemical reaction.

Here I report a set of markers that I propose as a protein fingerprint for undifferentiated pluripotent cells that agree with predictions made by analysis of the transcriptome. I also report cases or protein markers of pluripotency that transcriptional profiles of cells fail to identify and are likely subject to post-transcriptional regulation. Further investigation confirmed several cases of post-transcriptional regulation of protein levels during differentiation of genes essential for the self-renewal of undifferentiated cells. I identified one protein, Racgap1 that is enriched at the protein level, but not the mRNA level and is necessary for mouse ES cell colony formation.

2.3 Results

2.3.1 Differentiation of ES and P19 cells by aggregation and retinoic acid treatment leads to exit from the pluripotent state

In order to understand the composition of the proteome of pluripotent mouse stem cells, I cultured two pluripotent stem cell types: 129/SvEv mouse ES cells and P19 EC cells. Both cell types are self-renewing and pluripotent, however the two cell types differ markedly in their requirements for maintenance of the pluripotent state. Mouse ES cells require exogenous LIF in order to stay undifferentiated, while P19 EC cells are able to stay undifferentiated in the absence of exogenous growth factors.

In order to understand the proteins unique to the pluripotent state in these two cell types, I chose to differentiate the cells into a neural lineage using 1 μM

All-trans retinoic acid (ATRA). This experimental design was adapted from a study by an unaffiliated group whose studies on the transcriptomes of 129/SvEv ES cells and P19 EC cells[106] serve as a useful resource in my understanding of the proteome of how the proteome of pluripotent cells reflects as well as differs from the transcriptome.

Following the Aiba et al. protocol, I aggregated ES in the absence of LIF, changing media every 2 days and supplementing the media with 1µM ATRA on day 4 to initiate neural differentiation. Undifferentiated and differentiated cells were collected and processed for either mass spectrometry or for immunofluorescence (IF) in order to confirm that the cells expressed known markers of pluripotency before differentiation and lose them upon differentiation.
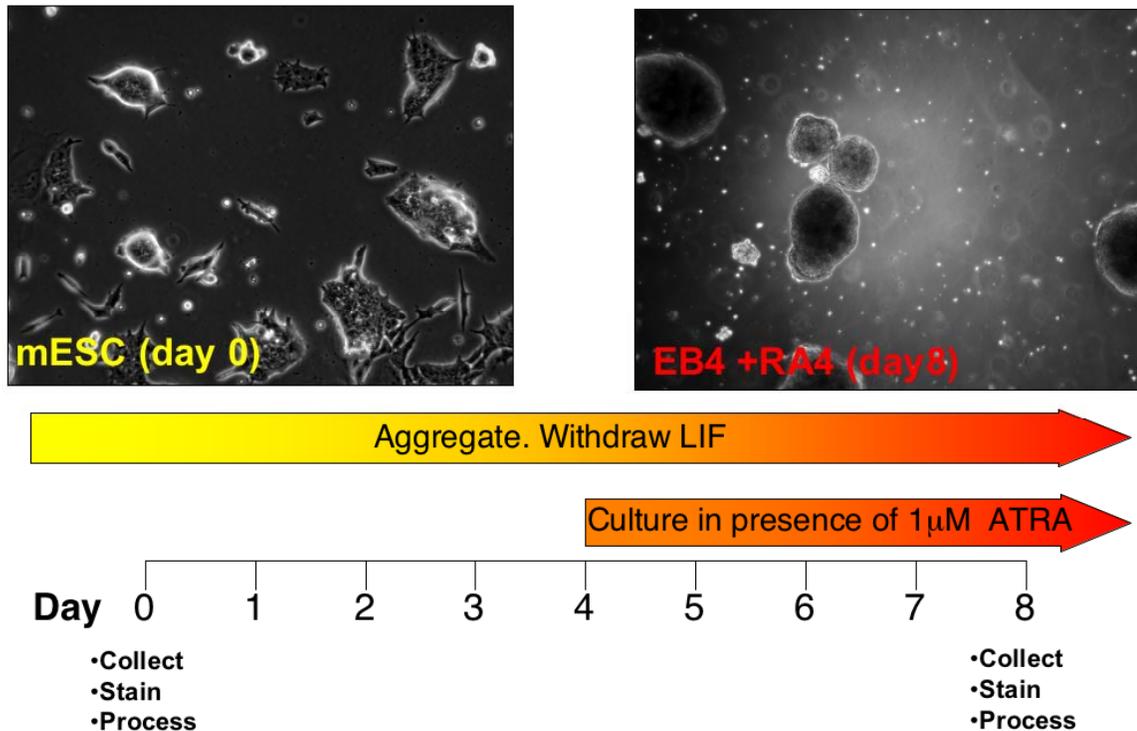
Figure 2.1: RA mediated differentiation of ES cells. Experimental design of RA-mediated differentiation of ES cells. ES cells of line 129/SvEv from ATCC cells were grown in undifferentiated conditions in the presence of LIF as described in the publication from Aiba et al. or differentiated by aggregation in bacteriological plates in the absence of 1 μM ATRA for 4 days, and then in the presence of 1 μM ATRA for 4 days as described in material and methods. Media was refreshed at differentiation day 2 and cells were collected at the indicated time points by treatment with versene and washed in 10 mM Hepes buffered saline (pH 7.4) and frozen at –80c. Cells were dissolved in 2% rapigest, reduced, trypsinized and labeled with the iTRAQ mass tagging reagent as described in the materials and methods.
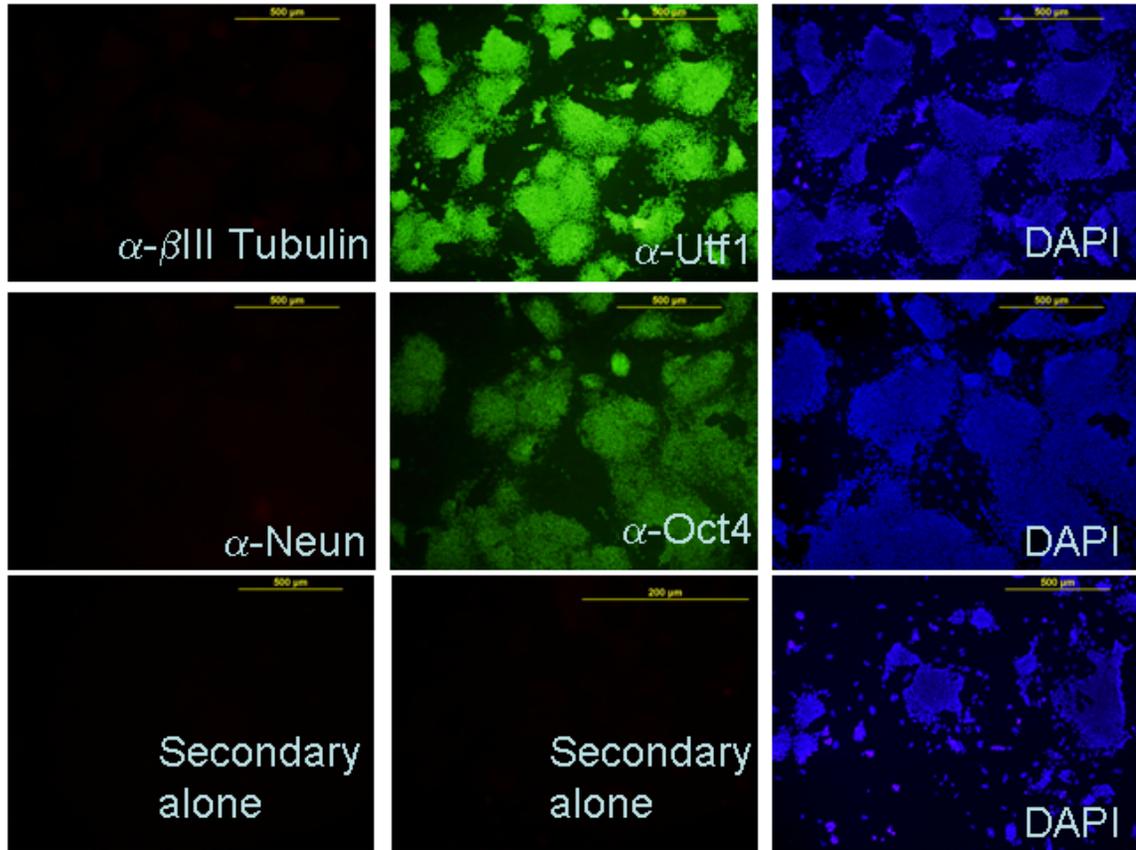
Figure 2.2: Immunofluorescence of ES cells before differentiation. Undifferentiated 129/SvEv cells express markers of pluripotency but not of mature neural lineages. Immunofluorescence of ES cells before differentiation confirmed their undifferentiated state. Cells were positive for the pluripotency markers UTF1 and Oct4 and negative for both βIII Tubulin and Neun.
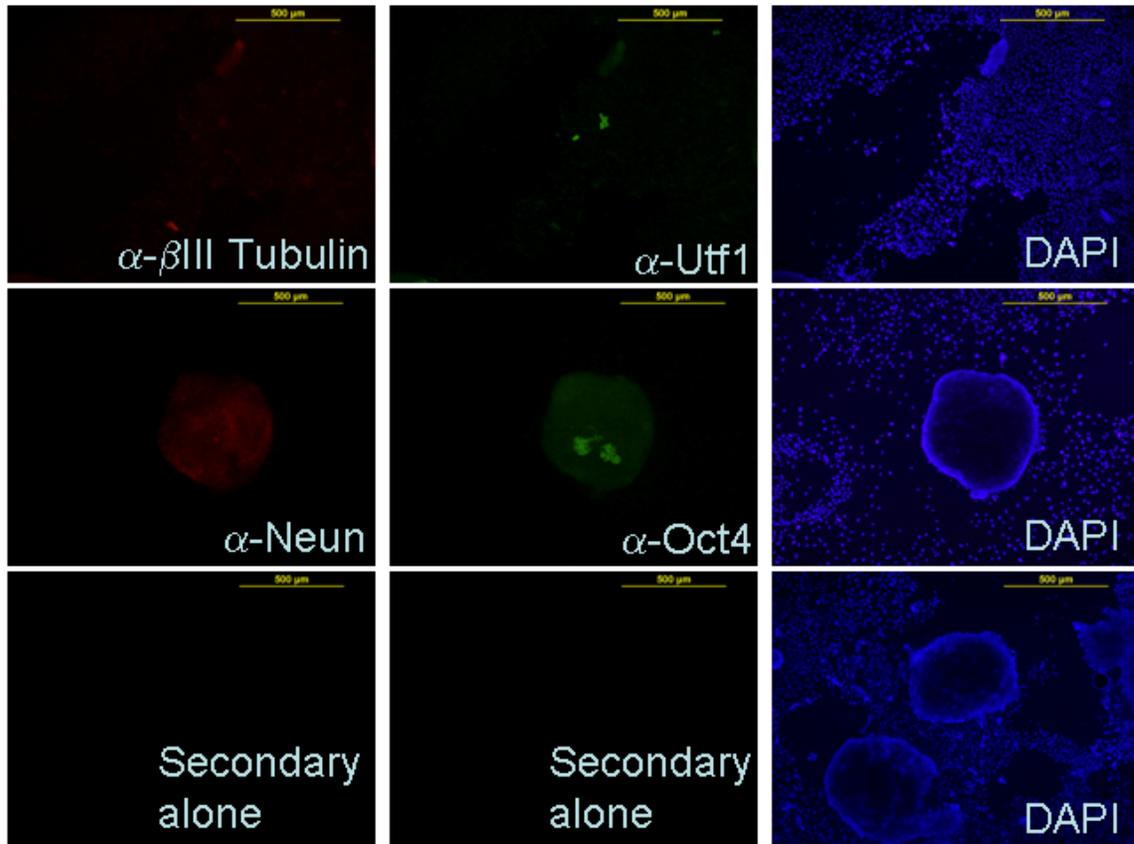
Figure 2.3: Immunofluorescence of ES cells after differentiation. Differentiated ES cells have lost most expression of pluripotency-associated markers and have induced expression of neural markers Immunofluorescence of ES cells after differentiation confirmed that the cells had largely differentiated. Cells were mainly negative for the pluripotency markers UTF1 and Oct4, however unlike in the case of the P19 cells, a subset of the cells in the embryoid bodies maintained expression of Oct4 and UTF1.
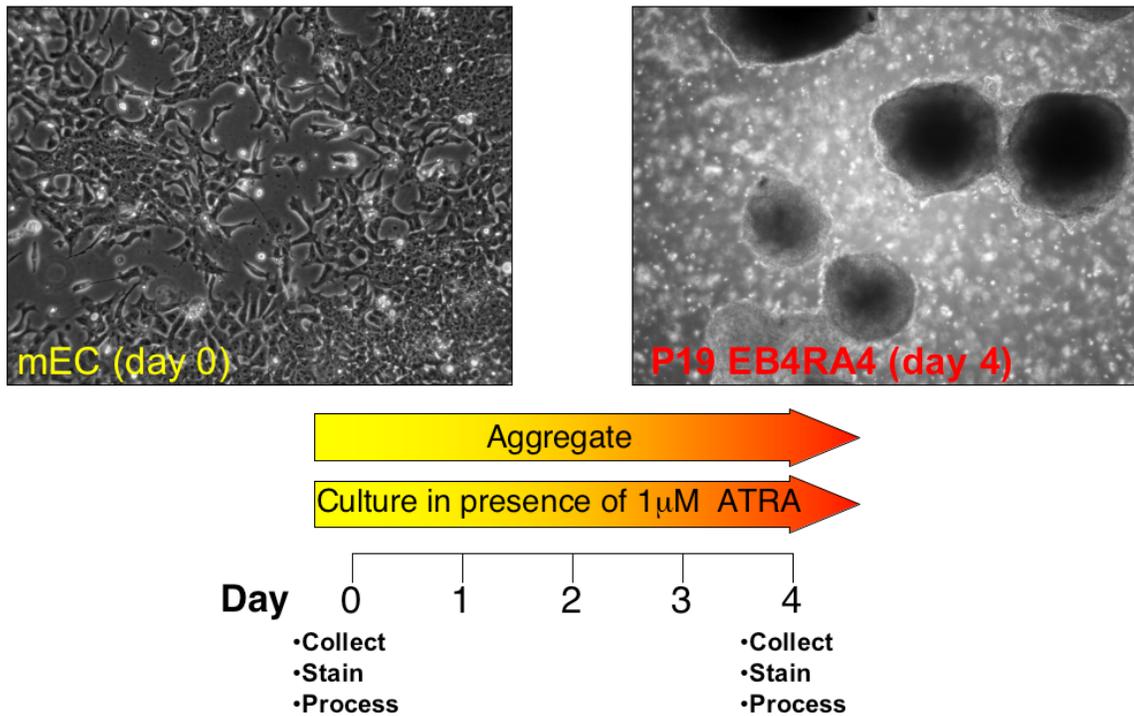
Figure 2.4: RA mediated differentiation of EC cells. Experimental design of RA-mediated differentiation of P19 EC cells. P19 cells from ATCC cells were grown in undifferentiated conditions as described in the publication from Aiba et al. or differentiated by aggregation in bacteriological plates in the presence of 1 µM all-trans retinoic acid (ATRA) for 4 days as described in material and methods. Media was refreshed at differentiation day 2 and cells were collected at the indicated timepoints by treatment with versene and washed in 10 mM Hepes buffered saline (pH 7.4) and frozen at –80c. Cells were dissolved in 2% rapigest, reduced, trypsinized and labeled with the iTRAQ mass tagging reagent.
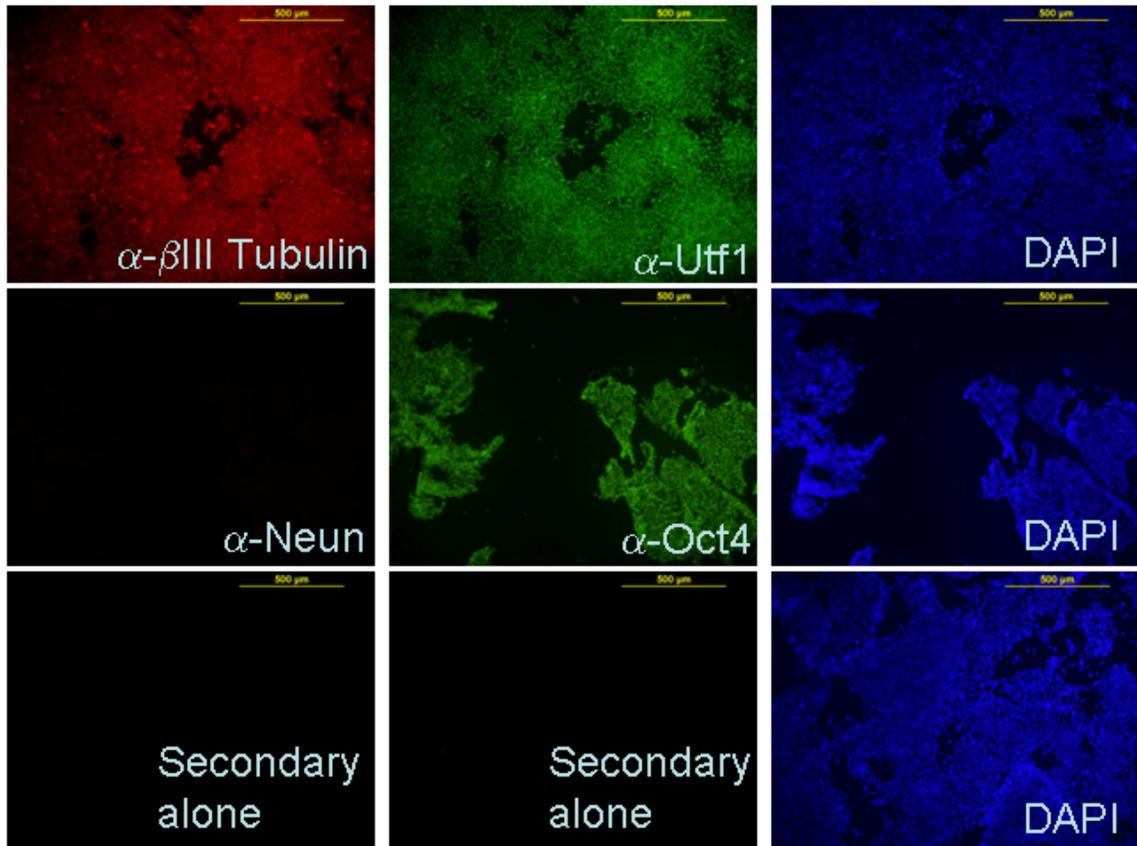
Figure 2.5: Immunofluorescence of EC cells before differentiation. Immunofluorescence of P19 cells before differentiation confirmed their undifferentiated state. Undifferentiated EC cells were positive for the pluripotency markers UTF1 and Oct4. P19 cells were also positive for the neuroectodermal marker βIII Tubulin presumably reflecting their predisposition to differentiate into a neuroectodermal lineage.
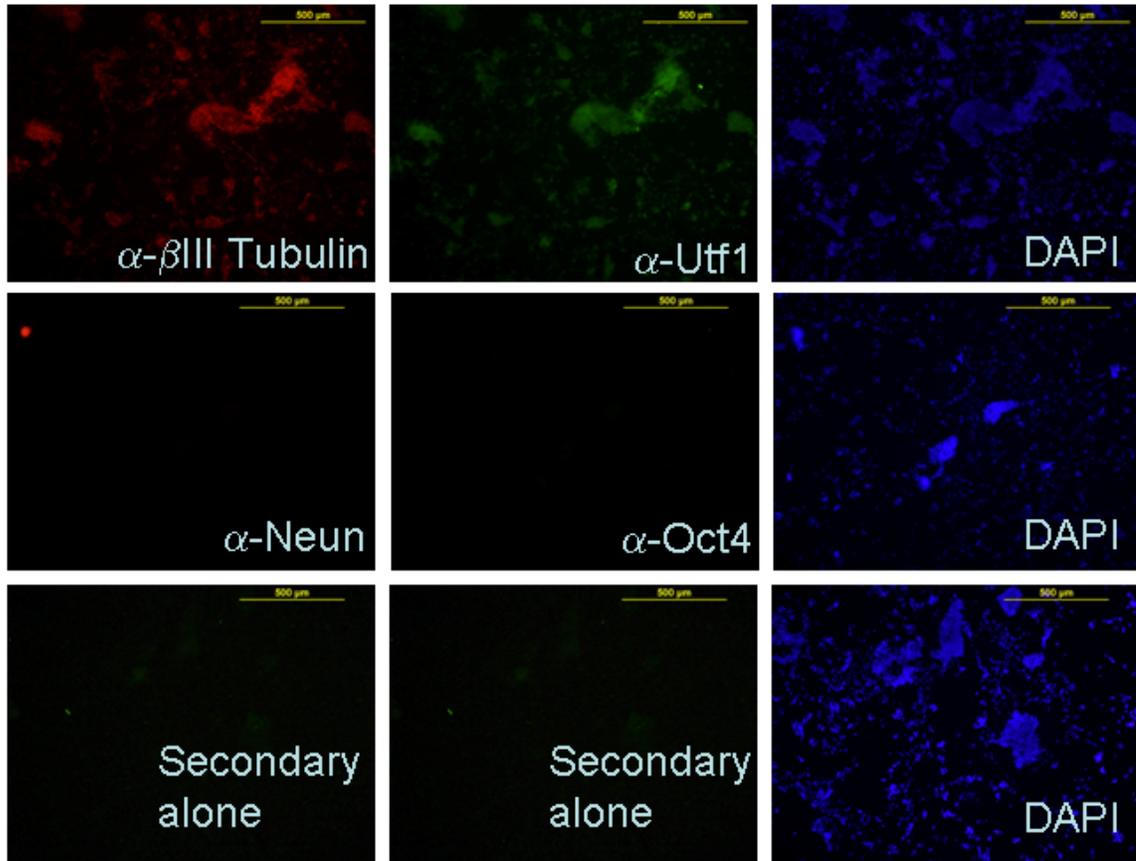
Figure 2.6: Immunofluorescence of ES cells after differentiation. Immunofluorescence of P19 cells after differentiation confirmed that the cells have exited the pluripotent state. By day 4 of aggregation and retinoic acid treatment, cells have lost most expression of pluripotency markers UTF1 and Oct4, but not yet terminally differentiated into mature neurons as evidenced by the lack of Neun staining.

2.3.2 Proteome analysis results in deep datasets with low measured false discovery rates

MS/MS analysis of three biological replicates of 129/SvEv ESCs before and after Retinoic acid mediated differentiation identified a total of 4053 proteins with a measured FDR of 0.56% (23 hits from a decoy database out of 4053 proteins identified.

MS/MS analysis of two biological replicates of P19 ECCs before and after Retinoic acid mediated differentiation resulted in identification of 4501 protein groups with a measured FDR of 0.2% (9 hits from a decoy database).

To increase the depth of this proteome analysis, nuclei were purified from P19 EC cells before and after differentiation, using a fractionation kit from Pierce. The same fractionations on ES cells and ES derived EBs failed due to the tough, fibrous nature of the ES derived EBs. MS/MS analysis of three biological replicates of Nuclei from P19 ECCs before and after Retinoic acid mediated differentiation allowed me to identify 4046 protein groups with a measured FDR of 0.39% (16 hits from a decoy database).

Table 2.1: Summary of whole proteome data sets.

	Biol. reps	Total protein hits	Reverse database hits	Forward database hits	FDR	iTRAQ Labeling efficiency (N-term, lysine or both)	Proteins with iTRAQ reporter intensity >100
129/Sv Ev ESCs	3	4053	23	4030	0.56%	98.5%	3566
Whole P19 cells	2	4501	9	4492	0.20%	92%	3801
P19 Nuclei	3	4046	16	4030	0.39%	97.7%	3569

2.3.3 Single peptide hits make up the majority of false discoveries

To understand where false discoveries were coming from, and to lower the FDR of these three datasets, I categorized proteins by number of unique peptides identified. In all three datasets, proteins identified with only one unique peptide contributed the vast majority (>80%) of false discoveries (**Figure 2.7**), while making up <15% of proteins identified (**Figure 2.8, 2.9** and **2.10**). Removing all single peptide hits from the data sets reduced the FDRs by at least 5-fold in all cases (**Figure 2.11**) lowering the measured FDR of identified proteins in the three datasets to <0.1%.
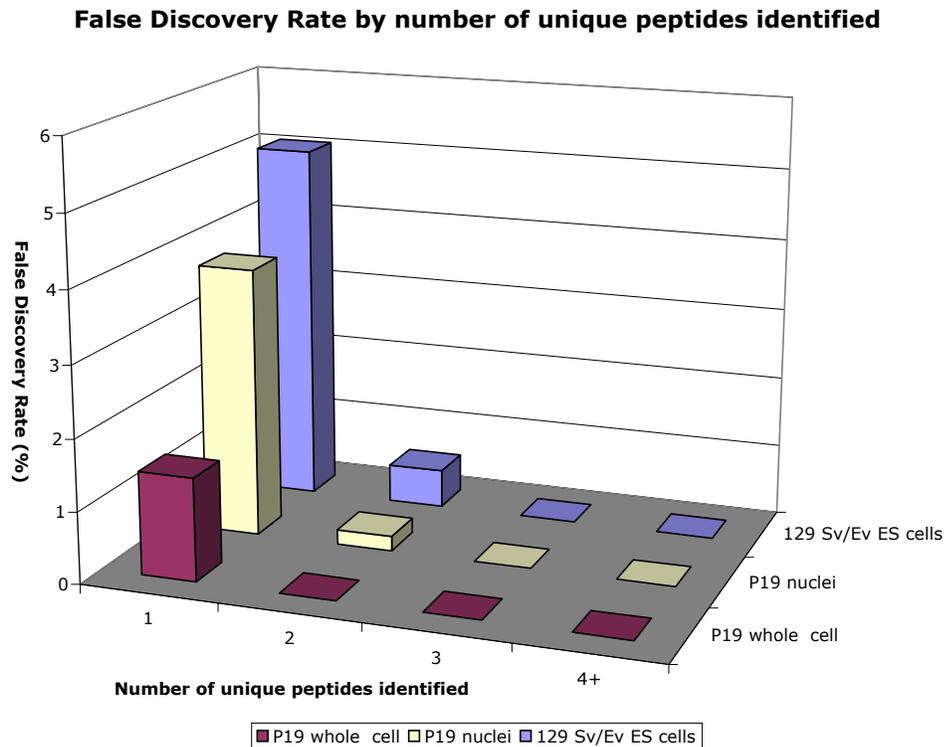
Figure 2.7: False discovery rate of proteins from MS analysis. False discovery rate of proteins identified with one peptide is higher than FDR of proteins identified with multiple peptides. Proteins were binned into four categories and false discovery rate calculated for each category. FDRs are presented as percentages.

Unique peptides identified per protein group 129 SV/EV ES cells
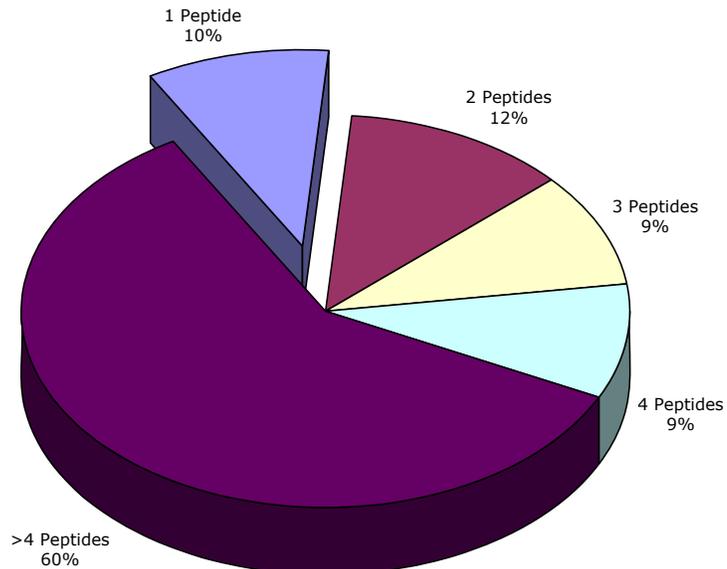
Figure 2.8: Peptides identified per protein group in ES cells. Distribution of unique peptides identified per protein among proteins identified in ES data Proteins identified by 1 peptide make up 11% of whole mES cell data.

Unique peptides identified per protein group P19 Whole cell
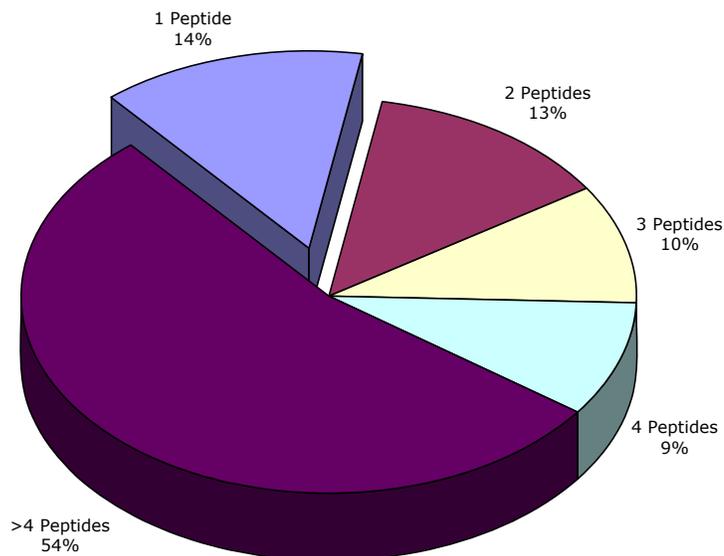
Figure 2.9: Peptides identified per protein group in EC cells. Distribution of unique peptides identified per protein among proteins identified in P19 whole cell data. Proteins identified by 1 peptide make up 14% of whole P19 cell data.

Unique peptides identified per protein group P19 Nuclei
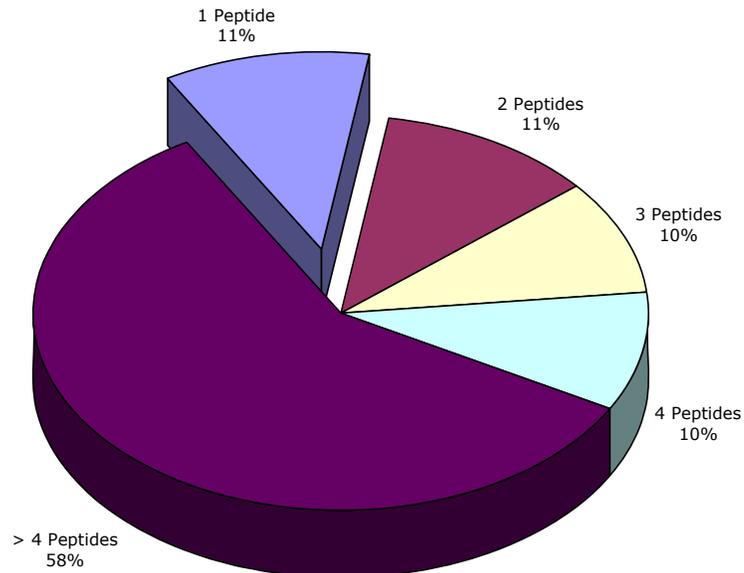
Figure 2.10: Peptides identified per protein group in ES nuclei. Distribution of unique peptides identified per protein among proteins identified in P19 nuclear data Proteins identified by 1 peptide make up 11% of P19 nuclear data.
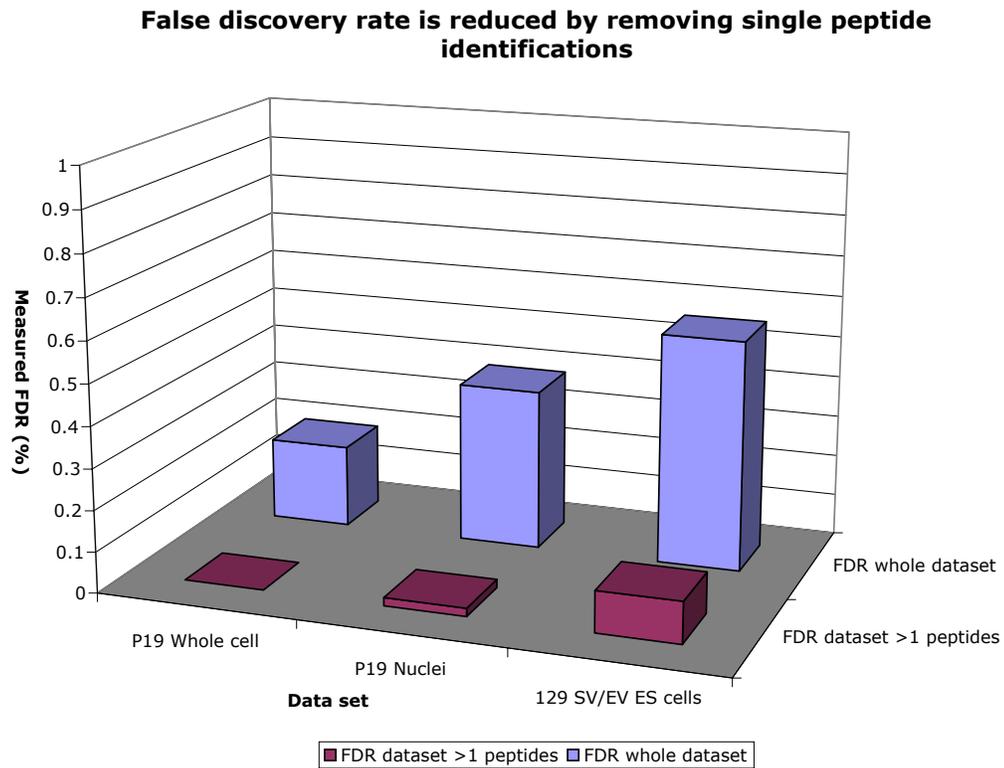
Figure 2.11: Removal of single peptide hits reduces FDR. Removal of proteins identified with single peptides reduces FDR of all datasets.

2.3.4 Incorporation of iTRAQ reagent allows for relative quantification of proteins enriched in undifferentiated EC and ES cells

During preparation of proteins for mass spectrometry analysis, we modified N-termini and lysine side chains of tryptic peptides using the iTRAQ reporter (Applied Biosystems). Cells were collected and processed as described in the methods after reduction and trypsinization, samples were split into two. Peptides from undifferentiated samples were modified with iTRAQ reporter ion mass tags of 114 or 115, while peptides from differentiated cells were tagged with iTRAQ reporter ion mass tags of 116 or 117. For this reason each protein identified in the three datasets had four values from the iTRAQ mass tagging reporters associated with them.

After removing proteins with low total iTRAQ reporter ion intensities (below 100), I generated a list of 3566 proteins identified and quantified with the iTRAQ reporter tagged in 129/SvEv ES cells before and after differentiation, 3801 proteins identified and relatively quantified with the iTRAQ reporter tagged in P19 cells before and after differentiation and 3569 proteins identified and relatively quantified with the iTRAQ reporter tagged in P19 nuclei before and after differentiation (**Table 2.1**). These three datasets thus represent deep, semi-quantitative analyses of the proteomes of two different pluripotent stem cell types before and after differentiation.
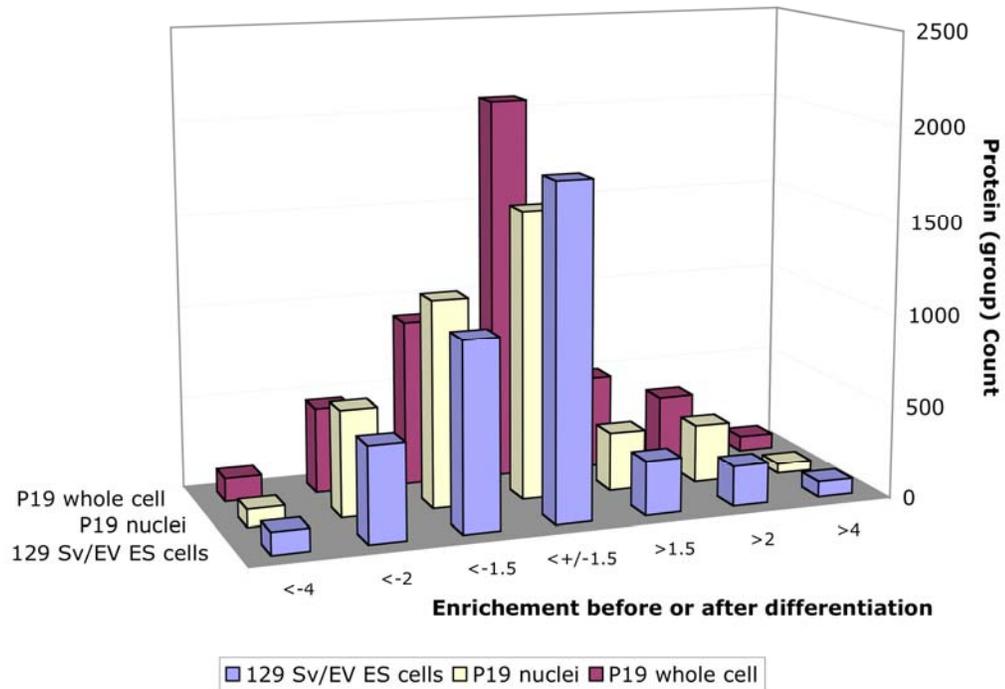
Figure 2.12: Distribution of iTRAQ ratios in three datasets. Distribution of ratios of iTRAQ reporter (reporter intensity undifferentiated / reporter intensity of differentiated) before and after differentiation follows a normal distribution with slightly more proteins enriched after differentiation in all three datasets.

2.3.5 Quantitative proteome profiles identify conserved protein markers of pluripotency and differentiation

To identify those proteins that are enriched in pluripotent cells before and after RA mediated differentiation, I employed a two-tailed paired students T-test with a p value cutoff of 0.05 and a fold-change cutoff of +/-1.5 fold. **Figure 2.3** summarizes the results of these analyses. Analysis of the overlap of these proteins enriched in all three datasets results in a relatively short list of proteins specific to the undifferentiated pluripotent state (4) and a slightly longer list of proteins (21) enriched in differentiated cells of al three datasets. The proteins present in undifferentiated pluripotent cells included the adhesion protein Cdh1 (E-Cadherin) and the nuclear importin Kpn2. Of the 21 proteins enriched across all three RA-differentiated cells, 6 are metabolic enzymes.
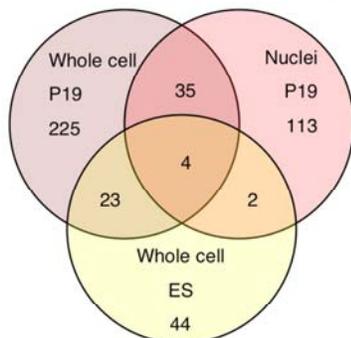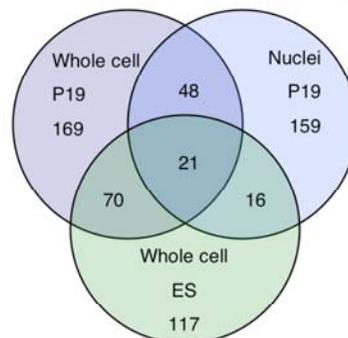
Figure 2.13: Overlap of proteins enriched in all datasets using t-test. Venn diagrams to illustrate the overlap of proteins considered enriched (p<0.05, >50% fold change, >1 peptide identified) before (**A**) and after (**B**) differentiation in three proteome datasets.

Table 2.2: Proteins enriched in all three datasets meeting p-value threshold of <0.05. Proteins enriched in all three cell types, determined by p-value and fold change, before and after differentiation.

Gene Symbol	Accession	Percent Coverage	Num Peps Unique	Spectrum num	Score Unique	Protein	P19 NER undiff-average	P19 NER undiff-p-value	P19 undiff-average	P19 undiff-p-value	mES undiff-average	mES undiff-p-value
Cdh1	IPI00318626	17	8	123	165.56	Cdh1 Epithelial-cadherin precursor	6.74	0.04	7.15	0.04	1.76	0.04
Mki67ip	IPI00225912	38	8	105	161.67	Mki67ip Isoform 1 of MKI67 FHA dom	3.36	0.01	2.13	0.04	1.82	0.04
LOC100043906	IPI00124973	52	20	296	385.35	Kpna2,LOC100039592,LOC10004390	2.28	0.03	2.15	0.008	1.85	0.02
Mdn1	IPI00458039	15	53	219	1027.48	Mdn1 Midasin homolog	1.54	0.01	1.6	0.04	1.68	0.02
Atp2b1	IPI00556827	14	11	60	228.91	Atp2b1 plasma membrane calcium AT	0.45	0.04	0.6	0.03	0.56	0.04
Spag9	IPI00462746	7	7	31	132.53	Spag9 Isoform 2 of C-jun-amino-termi	0.43	0.03	0.4	0.03	0.47	0.02
Vat1	IPI00126072	19	7	53	130.23	Vat1 Synaptic vesicle membrane prot	0.49	0.02	0.64	0.03	0.45	0.02
Prdx2	IPI00117910	68	10	229	196.16	Prdx2 Peroxiredoxin-2	0.53	0.03	0.59	0.002	0.46	0.03
Glod4	IPI00110721	58	14	110	264.15	Glod4 Isoform 1 of Glyoxalase domair	0.56	0.03	0.57	0.015	0.64	0.01
Ldhb	IPI00229510	38	12	128	231.12	Ldhb L-lactate dehydrogenase B chai	0.44	0.04	0.41	0.0008	0.36	0.003
2610301G19Rik	IPI00123624	40	24	138	442.54	2610301G19Rik Protein KIAA1967 ho	0.51	0.04	0.49	0.05	0.63	0.03
Vim	IPI00227299	63	30	934	589.07	Vim Vimentin	0.46	0.02	0.23	0.005	0.27	0.02
Etfa	IPI00116753	55	14	142	273.11	Etfa Electron transfer flavoprotein sub	0.44	0.04	0.6	0.0005	0.62	0.03
LOC100046081	IPI00154004	48	9	117	184.56	Otub1,LOC100046081 Ubiquitin thioe	0.43	0.04	0.57	0.03	0.64	0.05
Clic1	IPI00130344	61	9	221	184.03	Clic1 Chloride intracellular channel pr	0.34	0.009	0.49	0.03	0.44	0.02
Decr1	IPI00387379	23	6	49	108.81	Decr1 2,4-dienoyl-CoA reductase, mit	0.55	0.04	0.33	0.005	0.36	0.0005
Ndufs2	IPI00128023	24	8	48	148.6	Ndufs2 NADH dehydrogenase [ubiqui	0.47	0.02	0.57	0.008	0.56	0.03
Ddah2	IPI00336881	51	10	149	200.46	Ddah2 NG,NG-dimethylarginine dime	0.27	0.03	0.3	0.005	0.24	0.006
Ero1l	IPI00754386	21	7	55	136.56	Ero1l ERO1-like protein alpha precurs	0.33	0.02	0.47	0.01	0.39	0.007
Dpysl3	IPI00122349	48	19	181	367.28	Dpysl3 Dihydropyrimidinase-related p	0.26	0.004	0.24	0.0002	0.32	0.04
Msi1	IPI00121300	27	7	85	135.02	Msi1 Isoform 1 of RNA-binding protein	0.2	0.02	0.21	7.35E-05	0.34	0.047
Dach1	IPI00135978	10	6	10	99.86	Dach1 Isoform 1 of Dachshund homol	0.14	0.04	0.13	0.03	0.45	0.02
Mdk	IPI00114392	35	4	45	73.2	Mdk Midkine precursor	0.15	0.002	0.15	0.01	0.11	0.04
Cotl1	IPI00132575	53	7	97	122.81	Cotl1 Coactosin-like protein	0.16	0.04	0.29	0.02	0.27	0.04
Lamp2	IPI00134549	7	4	25	72.07	Lamp2 Isoform LAMP-2A of Lysosom	0.31	0.02	0.18	0.001	0.29	0.04

This list of proteins fails to include a large number of proteins detected in my data that have previously reported to be abundant before differentiation of pluripotent cells, such as Oct4[51-54, 120-123], UTF1[63, 124-128] and Dnmt3b[71, 129-131]. These proteins and other known pluripotency-specific proteins that I identified in my profiles reach the threshold of >1.5 fold enriched before differentiation in all replicates that they are identified in across all three datasets. The high variability in how much greater than 1.5 fold enriched they are, as well as the fact that they are not detected in some data sets results in very few of these proteins meeting the p value threshold. For this reason, I argue that the T-test is not an informative way to identify proteins enriched using iTRAQ quantitation in global proteome studies. I propose using a fold change cutoff to identify proteins with two or more unique peptides identified, that reproducibly fall into the category of >50% enriched before or after differentiation across replicates and experiments in to identify the reproducibly pluripotency associated proteins.

TABLE 2.3: Examples of known markers of pluripotency enriched in all datasets without meeting p-value threshold. Examples of Proteins reproducibly enriched in pluripotent cells before and after differentiation that fail to make the p-value cutoff of 0.05.

Gene Symbol	Accession	Percent Coverage	Num Peps Unique	Score Unique	Protein	P19 NER un/diff-1	P19 NER un/diff-2	P19 NER un/diff-3	P19 NER un/diff-average	P19 NER un/diff-p-value	mES un/diff-1	mES un/diff-2	mES un/diff-3	mES un/diff-average	mES un/diff-p-value	P19 un/diff-1	P19 un/diff-2	P19 un/diff-average	P19 un/diff-p-value	Total iTRAQ reporter intensity
Utf1	IPI00454162	66	13	233.52	Utf1 Isoform 1 of Undifferentiated embryonic cell transcription factor 1	16.4	15.697	11.314	14.28	0.01	4.315	6.082	5.673	5.3	6.61E-05	6.883	5.166	5.96	0.074	760333
Pou5f1	IPI00117218	35	11	200.42	Pou5f1 POU domain, class 5 transcription factor 1	5.647	8.695	8.683	7.53	0.058	1.811		3.742	2.6	0.074	14.43	7.764	10.58	0.1	179111.79
Dnmt3b	IPI00757806	51	34	651.48	Dnmt3b DNA methyltransferase 3B	6.723	2.022	13.929	5.74	0.128	2.678	0.784	3.195	1.89	0.117	4.323	1.264	2.34	0.14	11222.778
Crabp2	IPI00133687	66	9	174.78	Crabp2 Cellular retinoic acid-binding protein 2	0.062	0.0649	0.042	0.06	0.096		0.079	0.087	0.08	0.019	0.0671	0.035	0.05	0.001	2293135.8
Crabp1	IPI00230721	78	10	193.74	Crabp1 Cellular retinoic acid-binding protein 1	0.109	0.179	0.141	0.14	0.062	0.281	0.086	0.035	0.09	0.056	0.154	0.079	0.11	1.65E-04	941173.3
Cdh2	IPI00323134	24	11	221.15	Cdh2 Cadherin-2 precursor	0.238	0.245	0.25	0.24	0.132	0.173	0.254	0.565	0.29	0.132	0.213	0.252	0.23	0.009	29370.309
Hoxb6	IPI00132375	43	7	128.68	Hoxb6 Homeobox protein Hox-B6	0.507	0.135	0.26	0.26	0.089		0.087	0.168	0.12	0.163	0.189	0.376	0.27	0.08	8680.349
Mapk12	IPI00117172	18	4	78.6	Mapk12 Mitogen-activated protein kinase 12			0.53	0.53	0.066			0.07	0.07	0.271	0.296	0.638	0.43	0.11	2814.664

Using fold change alone, I identified 101 proteins >1.5 enriched before differentiation in both ES and EC cells in all three datasets. These include known pluripotency associated proteins Oct4, UTF1, Dnmt3b, Cadherin 1[83, 132-139] (embryonic cadherin) and Tcf3[140]. Fold change also identifies 175 proteins enriched after retinoic acid mediated differentiation in all three experiments. These proteins include the retinoic acid response proteins Rbp1, Crabp1 and Crabp2[141-148] as well as Cadherin 2 (neural cadherin)[149-154], the embryonic patterning gene Hox gene HoxB6[155-159] as well as many metabolism proteins Thus, identification of proteins with >1.5 fold change during differentiation of two pluripotent cell types across three experiments results in a list of protein that includes proteins already known to be associated with pluripotency and differentiation, as well as proteins not previously described as pluripotency associated.

A

Protein groups, >1 peptide identified
>50% enriched before differentiation
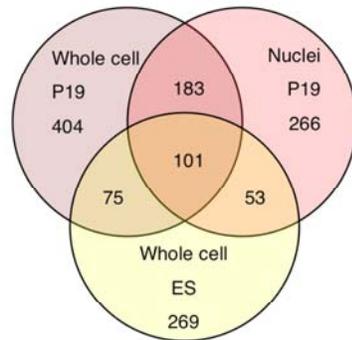
B

Protein groups, >1 peptide identified
>50% enriched after differentiation
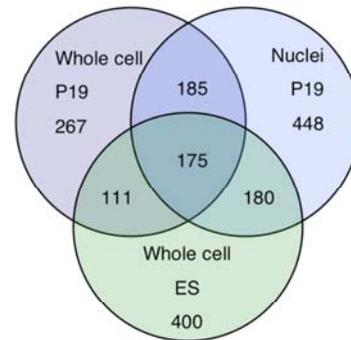
Figure 2.14: Overlap of proteins enriched in all datasets using fold change. Venn diagrams to illustrate the overlap of proteins considered enriched without using p-value cutoff (>50% fold change, >1 peptide identified) before (**A**) and after (**B**) differentiation in three proteome datasets.

Table 2.4: Proteins enriched before differentiation in all three datasets by fold-change. Proteins >50% enriched before differentiation in all three proteome datasets (after removal of single peptide hits and proteins with total iTRAQ reporter intensity of <100).

| Gene Symbol	Accession	Percent Coverage	Num Peps Unique	Score Unique	Protein	P19 NER un/diff-average	P19 NER un/diff p-value	P19 whole cell un/diff-average	P19 whole cell un/diff p-value	mES un/diff-average	mES un/diff p-value
L1td1	IPI00807952	45	30	570.09	L1td1 hypothetical protein LOC381591	9.16	0.084965	13.84	0.0026491	5.78	0.0067201
Utf1	IPI00454162	66	13	233.52	Utf1 Isoform 1 of Undifferentiated embryonic cell transcription factor 1	14.28	0.0098165	5.96	0.0739715	5.30	6.606E-05
Tcf3	IPI00135957	16	5	97.51	Tcf3 transcription factor 3 isoform 2	1.81	0.8547618	11.88	0.1304924	7.10	0.2377052
Pou5f1	IPI00117218	35	11	200.42	Pou5f1 POU domain, class 5, transcription factor 1	7.53	0.0575179	10.58	0.1002931	2.60	0.0735243
Gprasp1	IPI00229642	3	3	52.25	Gprasp1 Isoform 1 of G-protein coupled receptor-associated sorting protein 1	11.21	0.4423944	3.08	0.3041877	2.43	0.237913
Cdh1	IPI00318626	17	8	165.56	Cdh1 Epithelial-cadherin precursor	6.74	0.0428693	7.15	0.0368834	1.76	0.0371649
A030007L17Rik	IPI00111004	26	4	67.85	A030007L17Rik Uncharacterized protein C7orf24 homolog	1.85	0.054622	9.69	0.3695167	1.91	0.1702572
Pdxk	IPI00283511	12	2	38.45	Pdxk Pyridoxal kinase	9.18	0.3023532	2.16	0.2094897	2.08	0.4897989
Enpp3	IPI00458003	9	6	103.87	Enpp3 Ectonucleotide pyrophosphatase/phosphodiesterase family member 3	6.04	0.1298608	4.14	0.0592703	2.81	0.0505687
Foxo1	IPI00311101	6	2	39.83	Foxo1 Forkhead protein FKHR	2.33	0.1381533	3.08	0.0553611	7.37	0.5
Bub1	IPI00117147	6	3	56.38	Bub1 Budding uninhibited by benzimidazoles 1 homolog	3.43	0.2976869	2.67	0.043422	5.99	0.4298397
Zscan10	IPI00756815	7	4	79.61	Zscan10 Zinc finger protein 206 variant 1	7.45	0.1064909	2.48	0.4835676	1.68	0.0216565
Kif22	IPI00116757	18	8	142.32	Kif22 Kinesin-like protein KIF22	2.42	0.1468803	5.49	0.0009643	3.51	0.1705057
Sall4	IPI00475164	41	24	456.25	Sall4 Isoform 1 of Sal-like protein 4	4.35	0.0774645	4.98	0.0162983	1.96	0.1173906
Topbp1	IPI00453855	6	4	67.35	Topbp1 DNA topoisomerase II-binding protein 1	5.92	0.105959	3.48	0.1033487	1.89	0.1689394
Phc1	IPI00230256	15	6	115.15	Phc1 Isoform 2 of Polyhomeotic-like protein 1	1.68	0.0873214	1.69	0.0292453	7.66	0.1351906
Isg20l1	IPI00467786	18	3	54.61	Isg20l1 Adult male medulla oblongata cDNA, RIKEN full-length enriched library, clone:6330517B13 product:hypothetical Exonuclease containing protein, full insert sequence	3.85	0.2165312	3.29	0.0722625	3.70	0.380856
Mrps2	IPI00126011	10	2	34.89	Mrps2 Mitochondrial 28S ribosomal protein S2	5.62	0.2252093	2.48	0.0681721	2.61	0.0805303
Cth	IPI00122344	33	10	188.44	Cth Cystathionine gamma-lyase	2.14	0.1341341	3.22	0.0902969	4.84	0.1299461
Noc3l	IPI00128918	17	9	174.07	Noc3l Nucleolar complex protein 3 homolog	2.34	0.1076763	5.92	0.2361788	1.73	0.8559903
Dnmt3b	IPI00757806	51	34	651.48	Dnmt3b DNA methyltransferase 3B	5.74	0.1282206	2.34	0.1402339	1.89	0.1170031
Aurka	IPI00125590	10	2	36.09	Aurka Isoform 1 of Serine/threonine-protein kinase 6	4.50	0.0670543	2.64	0.1909256	2.82	0.1256913
Ddx10	IPI00336329	20	13	248.72	Ddx10 DEAD (Asp-Glu-Ala-Asp) box polypeptide 10	2.99	0.137951	4.52	0.0848457	2.17	0.1958073
Coro1a	IPI00323600	11	3	53.18	Coro1a Coronin-1A	3.45	0.0009932	4.32	0.5169017	1.62	0.1950294
Tuba4a	IPI00117350	59	21	416.54	Tuba4a Tubulin alpha-4A chain	2.20	0.1424376	2.94	0.0742415	3.92	0.0010784
Npm3	IPI00753838	36	5	92.86	Npm3 similar to Nucleoplasmin-3	4.61	0.1671706	2.75	0.0055101	1.61	0.0315462
G3bp2	IPI00124245	28	10	185.76	G3bp2 Isoform A of Ras GTPase-activating protein-binding protein 2	1.54	0.1949599	2.07	0.2450366	5.26	0.295512
Pdlim7	IPI00130607	22	6	112.14	Pdlim7 Isoform 1 of PDZ and LIM domain protein 7	2.01	0.1312603	4.67	0.0183264	2.15	0.0574952
Tcfe3	IPI00272774	15	5	90.2	Tcfe3 Isoform 4 of Transcription factor E3	2.73	0.1352201	3.50	0.1330427	2.53	0.4791699
Tjp2	IPI00323349	29	23	423.61	Tjp2 Tight junction protein ZO-2	2.36	0.1347266	4.13	0.1044093	2.06	0.1104097
Capg	IPI00277930	20	5	85.57	Capg 10 days neonate olfactory brain cDNA, RIKEN full-length enriched library, clone:E530115O13 product:capping protein (actin filament), gelsolin-like, full insert sequence	1.89	0.3452906	4.35	0.0840018	2.12	0.1693508
Rbpms	IPI00122127	46	8	144.9	Rbpms RNA binding protein gene with multiple splicing isoform 2	3.50	0.1225056	3.20	0.191568	1.61	0.7007377
Rif1	IPI00408125	25	43	803.92	Rif1 Isoform 2 of Telomere-associated protein RIF1	3.26	0.1050751	2.97	0.0009166	1.96	0.0277288
Usp36	IPI00346327	13	11	195.49	Usp36 ubiquitin specific protease 36 isoform 1	3.20	0.1106344	1.68	0.0316361	3.30	0.2131234
Lin28	IPI00170180	69	12	224.27	Lin28 Lin-28 homolog A	2.37	0.1016922	3.77	0.0079677	2.04	0.0176319
Mkrn1	IPI00459256	20	2	34.54	Mkrn1 Adult male cecum cDNA, RIKEN full-length enriched library, clone:9130412D24 product:makorin, ring finger protein, 1, full insert sequence	1.53	0.4279024	2.87	0.3084549	3.75	0.2040904
Ddx21	IPI00652987	67	51	977.4	Ddx21 DEAD (Asp-Glu-Ala-Asp) box polypeptide 21	3.45	0.0548644	2.80	0.0001693	1.80	0.0395709
Hspbap1	IPI00222501	26	8	151.61	Hspbap1 HSPB1-associated protein 1	3.24	0.5018068	2.36	0.0512058	2.43	0.0477619
Aurkb	IPI00268655	16	4	66.73	Aurkb Serine/threonine-protein kinase 12	1.88	0.107303	2.11	0.0314252	3.95	0.5
Abt1	IPI00471226	26	5	89.18	Abt1 Activator of basal transcription 1	3.14	0.0951022	2.49	0.0976508	2.22	0.1679071
Jmjd1c	IPI00670418	5	8	140.86	Jmjd1c similar to jumonji domain containing 1C isoform 3	2.38	0.0449766	3.54	0.0644999	1.88	0.0948829
Ddx24	IPI00113576	23	14	258.52	Ddx24 ATP-dependent RNA helicase DDX24	3.17	0.0154638	2.80	0.0131825	1.70	0.1533312
D19Bwg1357e	IPI00222675	29	17	299.41	D19Bwg1357e D19Bwg1357e protein	2.69	0.0807975	3.10	0.0244988	1.78	0.0653405
Jup	IPI00229475	14	8	149.64	Jup Junction plakoglobin	1.75	0.1215732	1.57	0.1593575	4.24	0.0073559
Arid3b	IPI00277032	26	7	137.48	Arid3b Isoform 1 of AT-rich interactive domain-containing protein 3B	1.98	0.0291409	2.44	0.0016091	3.06	0.1090453
Sdad1	IPI00387439	9	5	83.97	Sdad1 Isoform 1 of Protein SDA1 homolog	1.94	0.0098051	2.05	0.1458318	3.36	0.6785408
Wdr43	IPI00462015	38	18	331.67	Wdr43 Bone marrow macrophage cDNA, RIKEN full-length enriched library, clone:I830061N24 product:hypothetical G-protein beta WD-40 repeat containing protein, full insert sequence	3.29	0.0823067	2.31	0.0256284	1.71	0.0586789
Mki67ip	IPI00225912	38	8	161.67	Mki67ip Isoform 1 of MKI67 FHA domain-interacting nucleolar phosphoprotein	3.36	0.0108332	2.13	0.0387555	1.82	0.038492

Table 2.4 continued

Name	IPI				Description						
Bcat1	IPI00653423	20	7	138.03	Bcat1 Blastocyst blastocyst cDNA, RIKEN full-length enriched library, clone:I1C0025P20 product:branched chain aminotransferase 1, cytosolic, full insert sequence	2.13	0.0236284	2.53	0.0101436	2.61	0.1090195
8030462N17Rik	IPI00466384	11	3	55.31	8030462N17Rik RIKEN cDNA 8030462N17 gene	3.86	0.0968482	1.69	0.3082521	1.71	0.1798589
Ddx27	IPI00122038	24	17	315.92	Ddx27 Isoform 1 of Probable ATP-dependent RNA helicase DDX27	3.04	0.0648435	2.51	0.0082047	1.66	0.0782354
Nol10	IPI00119760	24	8	154.27	Nol10 Nucleolar protein 10	2.29	0.0145036	2.63	0.0997731	2.19	0.3794751
Rrp15	IPI00458958	14	5	84.26	Rrp15 RRP15-like protein	3.13	0.086389	2.29	0.1173712	1.57	0.2128934
Ddx18	IPI00459381	46	24	470.11	Ddx18 ATP-dependent RNA helicase DDX18	2.99	0.1018269	2.14	0.0149973	1.83	0.0350904
Nolc1	IPI00719871	38	28	484.96	Nolc1 nucleolar and coiled-body phosphoprotein 1 isoform D	3.03	0.0726665	2.06	0.0012386	1.86	0.046554
Plk1	IPI00120767	16	7	121.13	Plk1 Serine/threonine-protein kinase PLK1	2.00	0.0062241	3.20	0.0411606	1.67	0.6672482
Rrp12	IPI00420344	25	25	463.11	Rrp12 RRP12-like protein	2.70	0.1321032	2.51	0.0004868	1.65	0.030148
Rbm34	IPI00225915	21	8	148.8	Rbm34 RNA binding motif protein 34	2.15	0.0078377	3.05	0.0290355	1.66	0.2293701
Ltv1	IPI00132479	7	3	58.82	Ltv1 LTV1 homolog	2.65	0.136924	2.16	0.0059289	2.04	0.1347148
Wdsof1	IPI00129701	37	12	231.28	Wdsof1 WD repeats and SOF domain containing 1	2.76	0.0954986	2.29	0.0037951	1.78	0.3118498
Cwf19l2	IPI00221510	4	4	70.74	Cwf19l2,LOC100044213 Adult male corpora quadrigemina cDNA, RIKEN full-length enriched library, clone:B230117G05 product:hypothetical Lysine-rich region containing protein, full insert sequence	2.27	0.0630271	2.11	0.0293286	2.44	0.1070907
Wdr74	IPI00122120	19	5	94.13	Wdr74 WD repeat protein 74	2.40	0.0939858	2.67	0.018526	1.56	0.9684531
Nfkbil2	IPI00467906	2	2	34.28	Nfkbil2 Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor-like 2	1.58	0.6296566	1.55	0.0330156	3.49	0.147342
Nola3	IPI00133008	43	3	56.57	Nola3 H/ACA ribonucleoprotein complex subunit 3	2.10	0.0703402	1.65	0.3146346	2.86	0.3112713
Ftsj3	IPI00119632	36	23	432.73	Ftsj3 Putative rRNA methyltransferase 3	2.78	0.0778309	2.21	0.0191815	1.61	0.0935661
Jarid2	IPI00124575	12	11	202.33	Jarid2 Protein Jumonji	2.24	0.0079739	2.69	0.4369401	1.62	0.0687644
Cdca2	IPI00228728	17	9	160.12	Cdca2 Isoform 1 of Cell division cycle-associated protein 2	2.66	0.0527861	1.94	0.2348364	1.93	0.1081391
Imp3	IPI00122383	29	4	76.13	Imp3 U3 small nucleolar ribonucleoprotein protein IMP3	2.46	0.0074388	2.08	0.2580965	1.88	0.138858
Heatr1	IPI00411022	28	52	914.3	Heatr1 BAP28 protein	2.45	0.0659204	2.21	0.0537795	1.66	0.0912751
Cebpz	IPI00752710	19	15	279.92	Cebpz Isoform CBF1 of CCAAT/enhancer-binding protein zeta	2.30	0.0522641	2.43	0.028801	1.57	0.2568679
LOC100043906	IPI00124973	52	20	385.35	Kpna2,LOC100039592,LOC100043906 Importin subunit alpha-2	2.28	0.0288776	2.15	0.0082342	1.85	0.0217318
Nol1	IPI00311453	42	26	481.77	Nol1 Putative RNA methyltransferase NOL1	2.48	0.0730623	2.04	0.0002893	1.74	0.006978
Rcl1	IPI00122220	18	6	105.52	Rcl1 RNA 3'-terminal phosphate cyclase-like protein	2.76	0.1657709	1.97	0.0189545	1.50	0.0667522
Rbm27	IPI00828892	16	11	208.76	Rbm27 Isoform 1 of RNA-binding protein 27	1.79	0.0155251	1.76	0.0055276	2.63	0.1066549
Terf1	IPI00108266	20	6	109.07	Terf1 Telomeric repeat-binding factor 1	2.40	0.0909806	2.23	0.0346752	1.51	0.2895418
Rps19bp1	IPI00226227	37	4	78.98	Rps19bp1 40S ribosomal protein S19-binding protein 1	1.97	0.0432351	1.52	0.3288047	2.58	0.1478491
LOC671392	IPI00111560	49	13	257.29	Set,LOC671392 Isoform 1 of Protein SET	1.85	0.0748865	1.81	0.034654	2.42	0.0273172
Dtd1	IPI00133713	7	2	43.91	Dtd1 Isoform 1 of Probable D-tyrosyl-tRNA(Tyr) deacylase 1	1.50	0.3257092	2.99	0.0684543	1.54	0.1774372
Bop1	IPI00387354	38	18	332.5	Bop1 Ribosome biogenesis protein BOP1	2.37	0.1071995	1.97	0.0143856	1.68	0.0993812
Car2	IPI00121534	50	9	179.59	Car2 Carbonic anhydrase 2	1.91	0.1821303	1.81	0.0050836	2.27	0.0388472
Orc1l	IPI00130290	14	10	173.69	Orc1l Origin recognition complex subunit 1	1.76	0.0700562	2.22	0.0106619	1.98	0.8105306
Rexo4	IPI00407108	18	6	104.09	Rexo4 RNA exonuclease 4	1.83	0.0505305	2.36	0.032359	1.75	0.3034335
Hprt1	IPI00284806	50	8	154.32	Hprt1 Hypoxanthine-guanine phosphoribosyltransferase	1.87	0.0014514	2.04	0.0069548	2.01	0.0727534
Mina	IPI00131778	12	4	72.33	Mina 13 days embryo male testis cDNA, RIKEN full-length enriched library, clone:6030481A05 product:hypothetical RmlC-like structure containing protein, full insert sequence	2.38	0.012041	1.79	0.0078203	1.72	0.1881196
Utp20	IPI00330349	17	35	644.38	Utp20 similar to Down-regulated in metastasis homolog isoform 3	2.19	0.0634517	1.96	0.014653	1.74	0.0829494
Hells	IPI00121431	27	19	365.96	Hells Isoform 1 of Lymphocyte-specific helicase	2.12	0.0848939	1.99	0.0393205	1.75	0.0198195
Orc5l	IPI00125261	15	5	93.68	Orc5l Origin recognition complex subunit 5	2.18	0.101344	1.84	0.0215465	1.82	0.212788
Pml	IPI00229072	27	14	267.61	Pml Isoform 1 of Probable transcription factor PML	2.61	0.058409	1.63	0.0157735	1.57	0.1519568
Csda	IPI00330591	56	12	239.14	Csda Isoform 1 of DNA-binding protein A	1.74	0.0420395	2.11	0.0342698	1.95	0.1142239
Gnl3	IPI00222461	49	22	407.06	Gnl3 Isoform 1 of Guanine nucleotide-binding protein-like 3	2.29	0.0551017	1.79	0.0073918	1.67	0.0250393
Pdcd11	IPI00551454	34	50	946.07	Pdcd11 programmed cell death protein 11	2.14	0.0723326	1.93	0.0666671	1.58	0.0323705
Ncl	IPI00317794	60	49	893.53	Ncl Nucleolin	2.26	0.0830948	1.71	0.0049617	1.64	0.0108092
Utp14a	IPI00462762	37	20	380.71	Utp14a Isoform 1 of U3 small nucleolar RNA-associated protein 14 homolog A	2.21	0.0021275	1.64	0.02533	1.73	0.0805328
Nup50	IPI00420243	49	16	295.74	Nup50 17 days pregnant adult female amnion cDNA, RIKEN full-length enriched library, clone:I920062K22 product:nucleoporin 50, full insert sequence	1.87	0.1283407	2.02	0.0662509	1.56	0.0522654
Crip2	IPI00121319	50	6	115.26	Crip2 Cysteine-rich protein 2	2.17	0.0326938	1.59	0.1098105	1.67	0.1707507
Thoc7	IPI00377930	25	4	81.81	Thoc7 Ngg1 interacting factor 3 like 1 binding protein 1 isoform 1	1.79	0.0001419	1.88	0.1800281	1.59	0.7594608
Slc7a5	IPI00331577	8	3	61.42	Slc7a5 Solute carrier family 7 (Cationic amino acid transporter, y+ system), member 5	1.65	0.300097	1.68	0.5505977	1.80	0.1057967
C330023M02Rik	IPI00226576	10	9	157.91	C330023M02Rik Isoform 1 of TPR repeat-containing protein C12orf30 homolog	1.50	0.0670947	1.87	0.0516837	1.67	0.0570246
Eif2a	IPI00119806	40	17	324.99	Eif2a Isoform 1 of Eukaryotic translation initiation factor 2A	1.67	0.0011257	1.65	0.0150967	1.54	0.0579557
Mdn1	IPI00458039	15	53	1027.48	Mdn1 Midasin homolog	1.54	0.0131144	1.60	0.0405352	1.68	0.0163727
Wdr33	IPI00170242	14	10	194.18	Wdr33 Putative WDC146	1.59	0.0883342	1.68	0.1404682	1.52	0.004006

Table 2.5: Proteins enriched after differentiation in all three datasets by fold-change. Proteins >50% enriched after differentiation in all three proteome datasets (after removal of single peptide hits and proteins with total iTRAQ reporter intensity of <100).

Gene Symbol	Accession	Percent Coverage	Num Peps Unique	Score Unique	Protein	P19 NER un/diff-average	P19 NER un/diff p-value	P19 whole cell un/diff-average	P19 whole cell un/diff p-value	mES un/diff-average	mES un/diff p-value
Crabp2	IPI00133687	66	9	174.78	Crabp2 Cellular retinoic acid-binding protein 2	0.06	0.0958349	0.05	0.0006777	0.08	0.0189999
Crabp1	IPI00230721	78	10	193.74	Crabp1 Cellular retinoic acid-binding protein 1	0.14	0.0614642	0.11	0.0001646	0.09	0.055846
Tppp3	IPI00133557	30	5	96.62	Tppp3 Tubulin polymerization-promoting protein family member 3	0.21	0.0449177	0.04	0.0062344	0.17	0.068652
Mdk	IPI00114392	35	4	73.2	Mdk Midkine precursor	0.15	0.001746	0.15	0.0117169	0.11	0.0399454
Mxra7	IPI00134206	37	2	39.38	Mxra7 0 day neonate eyeball cDNA, RIKEN full-length enriched library, clone:E130302J09 product:hypothetical protein, full insert sequence	0.07	0.0285094	0.26	0.2056837	0.13	0.3725709
LOC100045055	IPI00225337	77	10	193.9	Rbp1,LOC100045055 Retinol-binding protein I, cellular	0.15	0.0654379	0.21	0.0008935	0.12	0.0280595
Trp53i11	IPI00342407	13	2	42.03	RIKEN full-length enriched library, clone:D330045A08 product:tumor protein p53 inducible protein 11, full insert sequence	0.22	0.0174964	0.10	0.1019938	0.24	0.1620252
Hoxb6	IPI00132375	43	7	128.68	Hoxb6 Homeobox protein Hox-B6	0.26	0.0891938	0.27	0.0794067	0.12	0.1633698
H1f0	IPI00404590	22	4	76.13	H1f0 Bone marrow stroma cell CRL-2028 SR-4987 cDNA, RIKEN full-length enriched library, clone:G430067B09 product:histone H1.0, full insert sequence	0.29	0.0047328	0.19	0.0041779	0.18	0.129987
Ptms	IPI00471441	28	9	164.22	Ptms Ptms protein	0.14	0.1146283	0.20	0.0434998	0.35	0.1085811
Tagln3	IPI00128986	38	7	120.89	Tagln3 Transgelin-3	0.26	0.0495007	0.18	0.1219121	0.27	0.1275559
Dach1	IPI00135978	10	6	99.86	Dach1 Isoform 1 of Dachshund homolog 1	0.14	0.0385166	0.13	0.0306614	0.45	0.0156971
Homer3	IPI00262616	5	2	39.35	Homer3 Isoform 1 of Homer protein homolog 3	0.21	0.0052242	0.14	0.0343836	0.36	0.1511102
Cotl1	IPI00132575	53	7	122.81	Cotl1 Coactosin-like protein	0.16	0.0356469	0.29	0.0168903	0.27	0.0345182
Msi1	IPI00121300	27	7	135.02	Msi1 Isoform 1 of RNA-binding protein Musashi homolog 1	0.20	0.0162783	0.21	7.35E-05	0.34	0.0477604
Dpysl4	IPI00313151	33	13	248.31	Dpysl4 Isoform 1 of Dihydropyrimidinase-related protein 4	0.26	0.1378843	0.21	0.0593414	0.28	0.0962137
Cdh2	IPI00323134	24	11	221.15	Cdh2 Cadherin-2 precursor	0.24	0.1319562	0.23	0.0086759	0.29	0.1318864
Idh2	IPI00318614	47	21	378.82	Idh2 Isoform 2 of Isocitrate dehydrogenase [NADP], mitochondrial precursor	0.24	0.0976478	0.24	0.0030836	0.28	0.0289157
Lamp2	IPI00134549	7	4	72.07	Lamp2 Isoform LAMP-2A of Lysosome-associated membrane glycoprotein 2 precursor	0.31	0.0161229	0.18	0.0011264	0.29	0.0436369
Ddah2	IPI00336881	51	10	200.46	Ddah2 NG,NG-dimethylarginine dimethylaminohydrolase 2	0.27	0.0302987	0.30	0.0047419	0.24	0.0057628
Tubb2b	IPI00109061	69	23	485.81	Tubb2b Tubulin beta-2B chain	0.18	0.1192562	0.13	0.0001398	0.49	0.0359771
Mtap2	IPI00118075	7	10	181.83	Mtap2 Microtubule-associated protein 2	0.30	0.1723494	0.21	0.1222361	0.30	0.1455731
Dpysl3	IPI00122349	48	19	367.28	Dpysl3 Dihydropyrimidinase-related protein 3	0.26	0.0035773	0.24	0.0002099	0.32	0.0356444
Rcn1	IPI00137831	28	6	112.58	Rcn1 Reticulocalbin-1 precursor	0.18	0.1465165	0.38	0.1329573	0.28	0.0959371
Syne2	IPI00668076	5	21	382.07	Syne2 synaptic nuclear envelope 2	0.35	0.002968	0.27	0.0161807	0.25	0.1563633
Dpysl2	IPI00114375	61	24	476.45	Dpysl2 Dihydropyrimidinase-related protein 2	0.32	0.0624112	0.27	0.0016298	0.29	0.0168245
Ppic	IPI00130240	15	2	44.09	Ppic Peptidyl-prolyl cis-trans isomerase C	0.10	0.0785176	0.55	0.0310896	0.23	0.0937371
Hp1bp3	IPI00342766	28	11	203	Hp1bp3 2 days neonate thymus thymic cells cDNA, RIKEN full-length enriched library, clone:E430016E04 product:heterochromatin protein 1, binding protein 3, full insert sequence	0.32	0.0482798	0.12	0.000135	0.44	0.1910641
Stxbp1	IPI00415402	10	5	90	Stxbp1 Isoform 1 of Syntaxin-binding protein 1	0.32	0.2529647	0.18	0.0290595	0.40	0.1041586
Apoa1bp	IPI00170307	32	6	107.48	Apoa1bp Apolipoprotein A-I-binding protein precursor	0.24	0.0648923	0.31	0.0354565	0.37	0.0561421
Marcks	IPI00229534	69	17	375.33	Marcks Myristoylated alanine-rich C-kinase substrate	0.26	0.0560325	0.39	0.0237038	0.27	0.0257725
Prkar2b	IPI00224570	16	5	99.64	Prkar2b cAMP-dependent protein kinase type II-beta regulatory subunit	0.46	0.0881482	0.32	0.0081899	0.17	0.1342375
Cfl2	IPI00266188	36	6	109.45	Cfl2 Cofilin-2	0.30	0.069761	0.28	0.0078358	0.38	0.1923305
Vim	IPI00227299	63	30	589.07	Vim Vimentin	0.46	0.0218033	0.23	0.0050817	0.27	0.0213442
H2afy	IPI00378480	67	17	319.57	H2afy Isoform 2 of Core histone macro-H2A.1	0.32	0.0535244	0.24	0.0065083	0.40	0.0275017
Etfdh	IPI00121322	10	4	73.55	Etfdh Electron transfer flavoprotein-ubiquinone oxidoreductase, mitochondrial precursor	0.17	0.04828	0.38	0.0492227	0.42	0.3985938
Tmed4	IPI00153468	12	2	46	Tmed4 Transmembrane emp24 domain-containing protein 4 precursor	0.22	0.1214775	0.38	0.0187418	0.38	0.0822627
H1fx	IPI00118590	42	8	146.51	H1fx H1 histone family, member X	0.30	0.1051358	0.28	0.0068444	0.41	0.0793937
Sh3bgrl	IPI00122265	85	7	134.96	Sh3bgrl SH3 domain-binding glutamic acid-rich-like protein	0.29	0.0347713	0.33	0.0024306	0.38	0.1140859
App	IPI00114389	5	3	54.23	App Isoform APP770 of Amyloid beta A4 protein precursor (Fragment)	0.24	0.0920576	0.23	0.0045469	0.53	0.1315332
Hmgb3	IPI00228879	42	8	163.9	Hmgb3 High mobility group protein B3	0.38	0.0659864	0.29	0.0033884	0.36	0.0426866
Mapk12	IPI00117172	18	4	78.6	Mapk12 Mitogen-activated protein kinase 12	0.53	0.0658995	0.43	0.1104085	0.07	0.2709521
Tmed1	IPI00355961	12	2	36.84	Tmed1 Isoform 1 of Transmembrane emp24 domain-containing protein 1 precursor	0.34	0.0497795	0.37	0.2119882	0.33	0.1940786
Mosc2	IPI00123276	11	3	57.12	Mosc2 MOSC domain-containing protein 2, mitochondrial precursor	0.34	0.0606351	0.22	0.0941986	0.49	0.2304577
Gm340	IPI00664739	5	4	75.23	Gm340 similar to Salivary glue protein Sgs-4 precursor	0.51	0.6264248	0.06	0.279895	0.52	0.1460987

Table 2.5 continued

Arl3	IPI00124787	58	8	155.13	Arl3 ADP-ribosylation factor-like protein 3	0.28	0.1181356	0.44	0.0061238	0.38	0.0561561
Epb4.1	IPI00395157	27	15	297.64	Epb4.1 Isoform 1 of Protein 4.1	0.47	0.1671545	0.43	0.0121227	0.21	0.2409676
Bcl7a	IPI00458752	50	6	101.96	Bcl7a Isoform 1 of B-cell CLL/lymphoma 7 protein family member A	0.41	0.0177238	0.30	0.0875928	0.40	0.4927139
Dynll2	IPI00132734	57	4	76.76	Dynll2 Dynein light chain 2, cytoplasmic	0.45	0.1763798	0.32	0.0055384	0.34	0.1162378
Mtap1b	IPI00130920	42	69	1367.39	Mtap1b Microtubule-associated protein 1B	0.34	0.0664519	0.39	0.0002726	0.39	0.0404066
Dpysl5	IPI00624192	36	14	243.25	Dpysl5 Dihydropyrimidinase-related protein 5	0.36	0.0848033	0.36	0.0051954	0.40	0.0461457
Dync1li2	IPI00420806	18	6	117.82	Dync1li2 Cytoplasmic dynein 1 light intermediate chain 2	0.27	0.0925867	0.35	0.1304696	0.51	0.0976497
Slc25a4	IPI00115564	60	19	349.53	Slc25a4 ADP/ATP translocase 1	0.28	0.0854542	0.43	0.0012696	0.42	0.0046022
Rcn2	IPI00474959	35	10	195.08	Rcn2 2 days neonate thymus thymic cells cDNA, RIKEN full-length enriched library, clone:E430020L13 product:reticulocalbin 2, full insert sequence	0.31	0.1074635	0.51	0.0566943	0.31	0.1207481
Senp7	IPI00315952	13	7	122.92	Senp7 SUMO1/sentrin specific protease 7 isoform 2	0.43	0.0601176	0.22	0.1919376	0.49	0.1641947
H2afy2	IPI00652934	52	16	304.01	H2afy2 Core histone macro-H2A.2	0.44	0.0757738	0.32	0.0002681	0.39	0.1265862
Bnip1	IPI00278462	21	4	67.08	Bnip1 Vesicle transport protein SEC20	0.47	0.2782876	0.64	0.0276288	0.04	0.0954777
Scp2	IPI00134131	15	11	182.54	Scp2 Isoform SCPx of Nonspecific lipid-transfer protein	0.30	0.0228739	0.32	7.771E-05	0.54	0.1677895
Camk2g	IPI00124695	17	5	88.12	Camk2g Isoform 1 of Calcium/calmodulin-dependent protein kinase type II gamma chain	0.47	0.026614	0.17	0.1709615	0.53	0.5536681
Acat1	IPI00154054	32	8	153.93	Acat1 Acetyl-CoA acetyltransferase, mitochondrial precursor	0.22	0.1649189	0.44	0.0716915	0.52	0.021277
Hspa4l	IPI00317710	14	9	158.18	Hspa4l Isoform 1 of Heat shock 70 kDa protein 4L	0.38	0.030965	0.45	0.0052674	0.36	0.081082
P4ha1	IPI00272381	17	8	157.04	P4ha1 2 days neonate sympathetic ganglion cDNA, RIKEN full-length enriched library, clone:7120485O14 product:procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), alpha 1 polypeptide, full insert sequence	0.38	0.0281736	0.62	0.0800956	0.19	0.006371
Lzic	IPI00173156	13	2	39.27	Lzic Protein LZIC	0.17	0.2401364	0.45	0.5424509	0.57	0.4187965
Txndc1	IPI00121341	19	4	82.73	Txndc1 Thioredoxin domain-containing protein 1 precursor	0.35	0.1199417	0.50	0.0695818	0.34	0.0955215
Marcksl1	IPI00281011	77	9	186.48	Marcksl1 MARCKS-related protein	0.32	0.1365377	0.31	0.0003163	0.56	0.0059621
Ero1l	IPI00754386	21	7	136.56	Ero1l ERO1-like protein alpha precursor	0.33	0.0223799	0.47	0.0102494	0.39	0.0069021
LOC100047170	IPI00459353	72	4	67.31	100048410 Guanine nucleotide-binding protein G(I)/G(S)/G(O) gamma-5 subunit precursor	0.41	0.095	0.46	0.0571639	0.33	0.0131196
Ldhb	IPI00229510	38	12	231.12	Ldhb L-lactate dehydrogenase B chain	0.44	0.0440762	0.41	0.0007562	0.36	0.0034397
Dis3l2	IPI00229771	7	4	67.66	Dis3l2 RIKEN cDNA 4930429A22 gene	0.47	0.0894481	0.32	0.1721725	0.42	0.584847
Phpt1	IPI00118757	37	5	84.28	Phpt1 14 kDa phosphohistidine phosphatase	0.37	0.2165074	0.46	0.0266185	0.39	0.1396078
Pafah1b2	IPI00118821	24	4	79.23	Pafah1b2 Platelet-activating factor acetylhydrolase IB subunit beta	0.37	0.1132585	0.51	0.0228384	0.35	0.0775603
Decr1	IPI00387379	23	6	108.81	Decr1 2,4-dienoyl-CoA reductase, mitochondrial precursor	0.55	0.0365704	0.33	0.0051023	0.36	0.0005591
Idh1	IPI00135231	48	20	374.86	Idh1 0 day neonate lung cDNA, RIKEN full-length enriched library, clone:E030024J03 product:isocitrate dehydrogenase 1 (NADP+), soluble, full insert sequence	0.45	0.0871798	0.34	0.0004449	0.45	0.0428307
2310056P07Rik	IPI00109611	37	7	114.45	2310056P07Rik E2-induced gene 5 protein homolog	0.33	0.1898419	0.34	0.1354199	0.59	0.1475801
Pafah1b3	IPI00118819	26	6	99.74	Pafah1b3 Platelet-activating factor acetylhydrolase IB subunit gamma	0.28	0.0373716	0.51	0.21173	0.47	0.0375587
Clic1	IPI00130344	61	9	184.03	Clic1 Chloride intracellular channel protein 1	0.34	0.0087767	0.49	0.0318507	0.44	0.0204959
Nes	IPI00453692	33	39	755.28	Nes Isoform 1 of Nestin	0.40	0.0740366	0.30	0.0018945	0.58	0.0654143
Msi2	IPI00120924	16	5	93.41	Msi2 Isoform 1 of RNA-binding protein Musashi homolog 2	0.41	0.1196061	0.32	0.1515036	0.57	0.0080389
Ckap4	IPI00223047	41	20	375.13	Ckap4 Cytoskeleton-associated protein 4	0.43	0.0525878	0.51	0.0984378	0.36	0.0176978
Spag9	IPI00462746	7	7	132.53	Spag9 Isoform 2 of C-jun-amino-terminal kinase-interacting protein 4	0.43	0.0327532	0.40	0.0259142	0.47	0.0201847
Cep170	IPI00662263	6	7	129.28	Cep170 similar to centrosomal protein 170kDa isoform 2	0.34	0.1633159	0.42	0.0040434	0.55	0.1882747
Papss1	IPI00311335	8	4	67.38	Papss1 Bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthetase 1	0.41	0.4219398	0.34	0.0247177	0.56	0.086838
Acaa2	IPI00226430	39	9	169.32	Acaa2 3-ketoacyl-CoA thiolase, mitochondrial	0.46	0.1594862	0.35	0.0059579	0.51	0.170079
Top2b	IPI00135443	33	47	866.43	Top2b DNA topoisomerase 2-beta	0.54	0.1215736	0.30	3.507E-05	0.49	0.0122927
Golph4	IPI00269029	12	5	98.73	Golph4 Golgi phosphoprotein 4	0.66	0.3062326	0.31	0.0889624	0.36	0.1124874
Erlin2	IPI00221540	26	7	119.71	Erlin2 Erlin-2 precursor	0.37	0.0044655	0.51	0.0850262	0.47	0.1083823
Tpmt	IPI00115215	25	4	82.53	Tpmt Thiopurine S-methyltransferase	0.46	0.1515692	0.39	0.0344459	0.50	0.0144996
Rps6ka3	IPI00114333	27	14	258.91	Rps6ka3 Ribosomal protein S6 kinase alpha-3	0.50	0.0578425	0.38	0.0180603	0.47	0.4178709
Ap1s1	IPI00118026	15	2	39.97	Ap1s1 AP-1 complex subunit sigma-1A	0.34	0.4081533	0.50	0.1335728	0.53	0.0656652
Alad	IPI00112719	31	5	97.19	Alad Delta-aminolevulinic acid dehydratase	0.47	0.2699111	0.26	0.0070074	0.63	0.0216674
Dak	IPI00310669	23	7	129.73	Dak Dihydroxyacetone kinase	0.50	0.4965609	0.30	0.1070809	0.58	0.2692494
Pcyox1	IPI00460063	14	4	75.24	Pcyox1 Prenylcysteine oxidase precursor	0.16	0.2246337	0.64	0.0362902	0.58	0.2799018
Snx6	IPI00111827	17	6	112.6	Snx6 Sorting nexin-6	0.43	0.1506494	0.57	0.1421019	0.39	0.0338477
Set	IPI00410883	46	12	246.15	Set Isoform 2 of Protein SET	0.49	0.0891252	0.35	0.0199435	0.54	0.0136402
H2-Ke6	IPI00115598	13	2	35.00	H2-Ke6 Isoform Short of Estradiol 17-beta-dehydrogenase 8	0.52	0.5	0.33	0.2618715	0.50	0.2981669
Hmgcl	IPI00127625	13	4	70.84	Hmgcl Hydroxymethylglutaryl-CoA lyase, mitochondrial precursor	0.27	0.4191899	0.62	0.0263175	0.50	0.143564
IPI00351472	IPI00351472	14	6	111.49	68 kDa protein	0.52	0.0014723	0.41	0.0095158	0.47	0.1667738
Prkra	IPI00471256	30	7	133.44	Prkra Interferon-inducible double stranded RNA-dependent protein kinase activator A	0.49	0.1419015	0.54	0.0309887	0.38	0.0547024
Gbe1	IPI00109823	5	3	53.65	Gbe1 1,4-alpha-glucan-branching enzyme	0.38	0.1445085	0.39	0.1507246	0.64	0.8191335
Akap2	IPI00119855	11	5	96.52	Akap2 Isoform KL1B of A-kinase anchor protein 2	0.57	0.0196028	0.49	0.1451148	0.35	0.1962616
Gkap1	IPI00626289	9	2	34.28	Gkap1 Mus musculus	0.61	0.655484	0.47	0.1848041	0.35	0.0332032
Aldoc	IPI00119458	47	15	289	Aldoc Fructose-bisphosphate aldolase C	0.49	0.0668752	0.40	0.0130834	0.54	0.1730368
9-Sep	IPI00457611	43	20	387.45	Sept9 Isoform 1 of Septin-9	0.34	0.0556928	0.47	0.0036509	0.63	0.0353227
Git1	IPI00470095	20	8	148.89	Git1 ARF GTPase-activating protein GIT1	0.37	0.1531904	0.63	0.1387439	0.45	0.9228509

Table 2.5 continued

Gene	IPI				Description						
Mecp2	IPI00131063	16	5	94.42	Mecp2 Isoform A of Methyl-CpG-binding protein 2	0.56	0.0259079	0.42	0.0313136	0.47	0.2922809
Dhrs1	IPI00331549	17	4	81.55	Dhrs1 Dehydrogenase/reductase SDR family member 1	0.32	0.0970799	0.49	0.0093403	0.64	0.1411868
Gnaq	IPI00228618	7	2	41.42	Gnaq CDNA, RIKEN full-length enriched library, clone:M5C1028J20 product:guanine nucleotide binding protein, alpha q polypeptide, full insert sequence	0.43	0.2304291	0.45	0.0374135	0.58	0.1737397
Acat2	IPI00228253	36	8	164.14	Acat2 Acetyl-CoA acetyltransferase, cytosolic	0.50	0.0517923	0.49	0.0086611	0.47	0.0467282
Dcakd	IPI00221828	40	8	146.32	Dcakd 9 days embryo whole body cDNA, RIKEN full-length enriched library, clone:D030055P10 product:weakly similar to DEPHOSPHO-COA KINASE	0.33	0.0852742	0.56	0.0013652	0.57	0.0626656
Klc1	IPI00395126	11	5	97.42	Klc1 64 kDa protein	0.35	0.0198198	0.45	0.1057506	0.66	0.2157202
Pgls	IPI00132080	48	9	172.29	Pgls 6-phosphogluconolactonase	0.39	0.1035807	0.47	0.0145971	0.61	0.005074
Rabgap1	IPI00378156	9	5	88.7	Rabgap1 RAB GTPase activating protein 1 isoform a	0.32	0.0725865	0.56	0.0328049	0.59	0.1623827
Mtpn	IPI00228583	49	6	108.09	Mtpn Myotrophin	0.47	0.0595003	0.53	0.0939098	0.47	0.0644716
Sdhb	IPI00338536	35	8	151.1	Sdhb Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial precursor	0.53	0.1925979	0.47	0.0169622	0.48	0.0699596
Nucb1	IPI00132314	34	14	240.94	Nucb1 Nucleobindin-1 precursor	0.40	0.0847445	0.57	0.0936826	0.51	0.0783014
Tbcd	IPI00461857	13	12	223.81	Tbcd Tubulin-specific chaperone D	0.41	0.2046492	0.55	0.0010068	0.53	0.0497197
Dlg3	IPI00776223	18	6	113.1	Dlg3 Discs, large homolog 3	0.61	0.1350286	0.41	0.007597	0.47	0.1212295
Sar1b	IPI00132397	31	5	95.17	Sar1b GTP-binding protein SAR1b	0.56	0.0647091	0.49	0.0216631	0.45	0.1097552
Garnl1	IPI00454161	2	3	49.4	Garnl1 Isoform 1 of GTPase-activating RapGAP domain-like 1	0.33	0.2249163	0.56	0.1517658	0.61	0.1520314
Gsk3b	IPI00125319	27	7	127.77	Gsk3b Glycogen synthase kinase-3 beta	0.55	0.104053	0.59	0.124056	0.37	0.1782878
Gnb1	IPI00120716	38	10	189.9	Gnb1 Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta 1	0.49	0.1469969	0.52	0.0063217	0.50	0.0157154
Rala	IPI00124282	23	5	97.95	Rala Ras-related protein Ral-A precursor	0.39	0.0629633	0.58	0.0744539	0.54	0.1131758
Nt5dc2	IPI00625954	25	9	155.31	Nt5dc2 5'-nucleotidase domain containing 2	0.45	0.0458915	0.46	0.0609596	0.61	0.0666682
Rab8b	IPI00411115	36	8	146.05	Rab8b Ras-related protein Rab-8B	0.51	0.1275606	0.66	0.0708245	0.35	0.126791
Akr7a5	IPI00331490	23	6	116.72	Akr7a5 Aflatoxin B1 aldehyde reductase member 2	0.48	0.1658549	0.52	0.0283769	0.53	0.0727459
Ctsz	IPI00125220	18	4	70.39	Ctsz Cathepsin Z	0.47	0.0260621	0.66	0.2605337	0.40	0.0834403
Atg4b	IPI00387185	14	3	62.22	Atg4b Cysteine protease ATG4B	0.56	0.0418419	0.64	0.2942117	0.33	0.1053412
Gstp1	IPI00555023	38	5	104.4	Gstp1 Glutathione S-transferase P 1	0.49	0.0612263	0.51	0.0084915	0.55	0.1022447
Sdhc	IPI00319111	17	2	36.96	Sdhc Succinate dehydrogenase cytochrome b560 subunit, mitochondrial precursor	0.47	0.1100242	0.49	0.1471004	0.60	0.9342183
Asph	IPI00474904	23	9	165.57	Asph Adult male urinary bladder cDNA, RIKEN full-length enriched library, clone:9530097J16 product:aspartate-beta-hydroxylase, full insert sequence	0.58	0.2658142	0.50	0.092875	0.50	0.1478451
Vat1	IPI00126072	19	7	130.23	Vat1 Synaptic vesicle membrane protein VAT-1 homolog	0.49	0.0225609	0.64	0.0303	0.45	0.0219153
Prdx4	IPI00116254	54	12	209.07	Prdx4 Peroxiredoxin-4	0.45	0.0072353	0.64	0.1164846	0.49	0.1564118
Dnajc9	IPI00128268	45	11	211.18	Dnajc9 DnaJ homolog subfamily C member 9	0.58	0.0963404	0.34	0.0057781	0.66	0.0309745
Oat	IPI00129178	42	12	235.32	Oat Ornithine aminotransferase, mitochondrial precursor	0.60	0.1472619	0.61	0.0004345	0.38	0.0236862
Prdx2	IPI00117910	68	10	196.16	Prdx2 Peroxiredoxin-2	0.53	0.0258362	0.59	0.0015199	0.46	0.0284696
Rnf214	IPI00221464	16	7	128.32	Rnf214 Isoform 1 of RING finger protein 214	0.59	0.1699547	0.52	0.1942293	0.49	0.0153965
Ndufs2	IPI00128023	24	8	148.6	Ndufs2 NADH dehydrogenase [ubiquinone] iron-sulfur protein 2, mitochondrial precursor	0.47	0.0163586	0.57	0.0080291	0.56	0.028184
Atp2b1	IPI00556827	14	11	228.91	Atp2b1 plasma membrane calcium ATPase 1	0.45	0.0409811	0.60	0.0261741	0.56	0.0388979
Wnk2	IPI00355244	3	4	67.21	Wnk2 Isoform 6 of Serine/threonine-protein kinase WNK2	0.31	0.3644208	0.65	0.215118	0.64	0.669271
Tmem43	IPI00120083	16	4	77.17	Tmem43 Transmembrane protein 43	0.41	0.152564	0.54	0.2520961	0.66	0.1678416
Cryz	IPI00134704	30	7	127.58	Cryz Quinone oxidoreductase	0.44	0.2596219	0.57	0.0114948	0.61	0.1576879
mt-Co2	IPI00131176	27	5	88.46	mt-Co2 Cytochrome c oxidase subunit 2	0.37	0.0414907	0.60	3.989E-05	0.65	0.0684685
Ctsb	IPI00113517	23	6	116.63	Ctsb Cathepsin B precursor	0.43	0.0107995	0.58	0.0116725	0.62	0.1208768
Myh10	IPI00515398	53	100	1950.32	Myh10 Myosin-10	0.62	0.0576441	0.64	0.000979	0.37	0.0450585
k	IPI00123624	40	24	442.54	2610301G19Rik Protein KIAA1967 homolog	0.51	0.0367472	0.49	0.0468724	0.63	0.0344518
Pdcd4	IPI00323064	35	10	184.29	Pdcd4 Programmed cell death protein 4	0.46	0.0215947	0.57	0.047551	0.60	0.2980612
Ccdc88a	IPI00461244	7	8	157.53	Ccdc88a Isoform 2 of Girdin	0.45	0.0796627	0.58	0.0569523	0.62	0.3992778
LOC100046081	IPI00154004	48	9	184.56	Otub1,LOC100046081 Ubiquitin thioesterase OTUB1	0.43	0.0401689	0.57	0.0265696	0.64	0.0474253
Zcd2	IPI00132350	25	3	46.65	Zcd2 Adult male cerebellum cDNA, RIKEN full-length enriched library, clone:1500009M05 product:hypothetical protein, full insert sequence	0.53	0.1847404	0.57	0.0139968	0.55	0.3845327
Vapb	IPI00135655	31	6	113.06	Vapb Vesicle-associated membrane protein-associated protein B	0.66	0.1915222	0.56	0.025179	0.43	0.1157551
Nmral1	IPI00169979	22	6	107.04	Nmral1 Isoform 1 of NmrA-like family domain-containing protein 1	0.59	0.1551621	0.65	0.0219049	0.41	0.0631382
Etfa	IPI00116753	55	14	273.11	Etfa Electron transfer flavoprotein subunit alpha, mitochondrial precursor	0.44	0.0424342	0.60	0.0005379	0.62	0.0269481
Erp29	IPI00118832	39	8	151.17	Erp29 Endoplasmic reticulum protein ERp29 precursor	0.50	0.1473205	0.52	0.0044695	0.63	0.0545961
Zfp292	IPI00131408	4	8	145.72	Zfp292 similar to zinc finger protein 292 isoform 4	0.64	0.4204706	0.51	0.1696014	0.52	0.4392028
Pafah1b1	IPI00309207	32	11	196.36	Pafah1b1 Isoform 1 of Platelet-activating factor acetylhydrolase IB subunit alpha	0.66	0.0318811	0.50	0.0376504	0.51	0.1787991
Fh1	IPI00129928	39	13	245.69	Fh1 Isoform Mitochondrial of Fumarate hydratase, mitochondrial precursor	0.64	0.1534344	0.46	0.0080587	0.59	0.0102751
Tpi1	IPI00467833	86	16	334.39	Tpi1 Triosephosphate isomerase	0.64	0.1128043	0.52	0.0092793	0.53	0.0090794
Hk1	IPI00283611	9	7	117.91	Hk1 Isoform HK1-SA of Hexokinase-1	0.64	0.482109	0.52	0.1097811	0.54	0.3088338
1810073G14Rik	IPI00336509	8	3	53.14	1810073G14Rik Isoform 1 of Protein FAM114A2	0.57	0.1107767	0.56	0.0546078	0.57	0.2158518
Slc25a11	IPI00230754	16	6	95.54	Slc25a11,LOC100048808 Mitochondrial 2-oxoglutarate/malate carrier protein	0.61	0.0662673	0.62	0.1328636	0.48	0.0596803
Aco2	IPI00116074	38	25	466.76	Aco2 Aconitate hydratase, mitochondrial precursor	0.51	0.0574281	0.64	0.0169634	0.56	0.0131811

Table 2.5 continued

Cox6a1	IPI00121443	40	2	37.73	Cox6a1 Adult male kidney cDNA, RIKEN full-length enriched library, clone:0610009H24 product:cytochrome c oxidase, subunit VI a, polypeptide 1, full insert sequence	0.51	0.4762361	0.57	0.1491512	0.65	0.4018022
Acadvl	IPI00119203	24	9	174.67	Acadvl Very-long-chain specific acyl-CoA dehydrogenase, mitochondrial precursor	0.64	0.3510202	0.56	0.0666042	0.55	0.1035042
Atp5f1	IPI00341282	24	7	126.84	Atp5f1 ATP synthase B chain, mitochondrial precursor	0.61	0.068199	0.51	0.1157062	0.64	0.0050303
Ndufa8	IPI00120984	25	4	69.58	Ndufa8 NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 8	0.56	0.0688111	0.65	0.0417841	0.55	0.1593421
Ywhae	IPI00118384	75	21	410.07	Ywhae 14-3-3 protein epsilon	0.60	0.0533901	0.59	0.0019586	0.57	0.0066372
Glod4	IPI00110721	58	14	264.15	Glod4 Isoform 1 of Glyoxalase domain-containing protein 4	0.56	0.0337272	0.57	0.0147119	0.64	0.0135017
Ywhaz	IPI00116498	68	17	337.38	Ywhaz 14-3-3 protein zeta/delta	0.61	0.0511756	0.56	0.0091572	0.62	9.342E-05
Sar1a	IPI00115644	26	6	108.67	Sar1a Osteoclast-like cell cDNA, RIKEN full-length enriched library, clone:I420011G14 product:SAR1a gene homolog 1 (S. cerevisiae), full insert sequence	0.47	0.0299358	0.65	0.0470482	0.65	0.109844
Tpm4	IPI00421223	63	16	301.75	Tpm4 Tropomyosin alpha-4 chain	0.61	0.0842442	0.61	0.0176879	0.57	0.1293682
Cttnbp2nl	IPI00453812	13	6	121.46	Cttnbp2nl CTTNBP2 N-terminal-like protein	0.62	0.0347276	0.64	0.0152714	0.54	0.6813778
Rnf20	IPI00380766	10	8	154.68	Rnf20 Isoform 1 of E3 ubiquitin-protein ligase BRE1A	0.66	0.057864	0.53	0.0460857	0.63	0.0285098
Nck1	IPI00453999	19	4	77.51	Nck1 non-catalytic region of tyrosine kinase adaptor protein 1	0.61	0.2269094	0.58	0.3811338	0.63	0.5171119
Ywhab	IPI00230682	70	15	298.36	Ywhab Isoform Long of 14-3-3 protein beta/alpha	0.63	0.1030145	0.62	0.0180768	0.57	0.0154638
Ugcgl1	IPI00420357	18	19	370.95	Ugcgl1 UDP-glucose:glycoprotein glucosyltransferase 1 precursor	0.64	0.1091005	0.63	0.0455268	0.56	0.0339344
Cat	IPI00312058	45	17	316.75	Cat Catalase	0.61	0.0943096	0.64	0.0078664	0.63	0.0361622

2.3.6 GO annotation identifies biological processes enriched before and after differentiation in all three datasets

To better understand the proteins enriched in cells before and after differentiation, I used the DAVID gene annotation database (http://david.abcc.ncifcrf.gov/home.jsp)[160, 161] to identify gene ontologies (GO terms) associated with the undifferentiated and differentiated states in all three datasets. The GO terms relating to biological processes enriched in before differentiation all three datasets included ribosome biogenesis, one carbon metabolism and amine biosynthesis (Figure). The GO terms relating to biological processes enriched after RA and aggregation-mediated differentiation all related to oxidative phosphorylation of glucose via the TCA cycle.
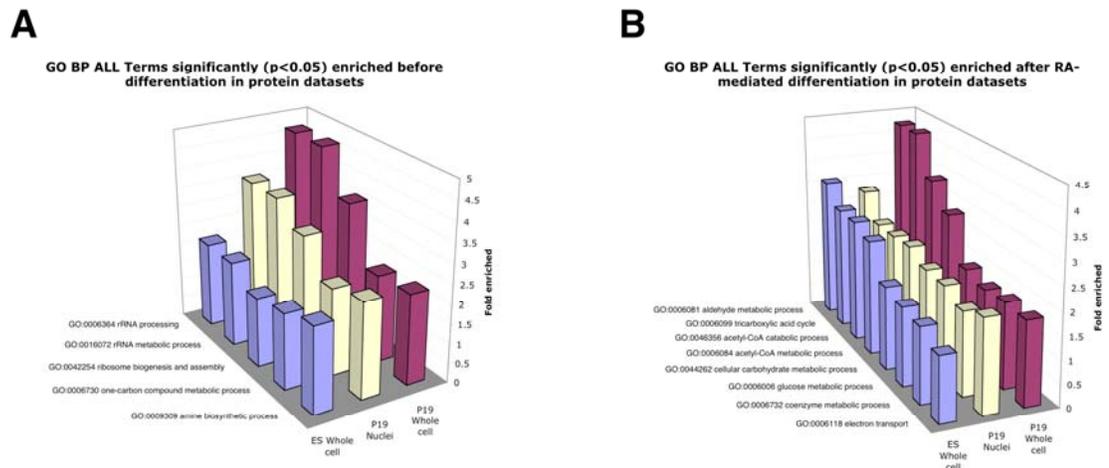
Figure 2.15: GO terms of proteins enriched all datasets. Gene ontologies of genes enriched in all three proteomes before (**A**) and after (**B**) differentiation. **A**) Proteins enriched before differentiation in all three proteome datasets were analyzed in the DAVID database using the GO_BP_all function. Proteins fell into five categories, three of them involved in rRNA and ribosome biogenesis. **B**) Proteins enriched after differentiation in all three proteome datasets were analyzed in the DAVID database using the GO_BP_all function. Proteins fell into 8 GO categories, all involved in oxidative metabolism (TCA cycle).

2.3.7 Comparison of GO terms enriched in proteome and transcriptome data

To get a better overview of the difference between mRNA and protein data, I used the DAVID database (http://david.abcc.ncifcrf.gov/home.jsp)[160, 161] to identify GO terms associated with each dataset. In al cases mRNA data produced more GO terms than the protein data, almost certainly due to the superior depth of a microarray. GO terms annotations of proteins were very similar to the annotations of mRNAs enriched, especially in undifferentiated ES cells. There were more differences between the mRNA and protein data of differentiated ES cells as well as undifferentiated and differentiated EC cells, however in all cases, the annotations assigned to the proteins enriched in each condition were >50% shared with the annotations assigned to the transcripts enriched in each condition suggesting that most pathways enriched in the various cell types are not subject to extensive regulation at the post-transcriptional level.
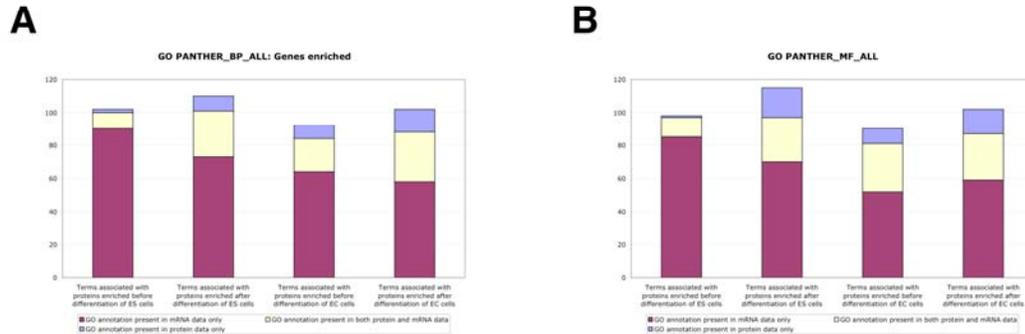
Figure 2.16: Comparison of GO terms enriched in mRNA and protein data. Comparison of gene ontologies of proteins identified as enriched before and after differentiation in proteome and transcriptome data from Aiba et al. 2006 that I reanalyzed. Proteins enriched before differentiation in ES and whole P19 data were analyzed by the DAVID go annotation database using the biological process (**A**) or molecular function (**B**) options. Almost all gene ontologies identified as enriched in the transcriptome of undifferentiated ES cells were also identified as enriched in the proteome of ES cells. Interestingly, this was not the case with ES cells after differentiation or EC cells in either state. I conclude that the ES transcriptomes and proteomes are more similar before differentiation than they are after differentiation. One possible explanation for this would be a change in post-transcriptional regulation as cells differentiate.

2.3.8 Identification and confirmation of cases of post-transcriptional regulation in pluripotent cells

My analyses of the proteome of pluripotent cells was intentionally modeled on a global transcriptome study by Aiba et al, using cell lines and differentiation conditions to replicate the study. I reanalyzed the data from Aiba et al and compared the resulting mRNA measurements of cells before and after differentiation to my proteome measurements. 196 of the 276 proteins enriched before or after differentiation in all three proteome experiments were also present on the microarray used in the Aiba experiments. Of the 196 proteins present in my proteome data before or after differentiation and Aiba's mRNA data, 97 of 196 (49.49%) are also enriched in both mRNA datasets, and therefore agree with my protein data. These gene products are therefore reproducible markers of pluripotency that work at the mRNA and protein level. On the other hand, 43 of 196 (21.94%) genes are enriched at the protein level but not the mRNA level. I propose these 43 proteins as putative targets of post-transcriptional regulation during RA mediated differentiation of pluripotent cells. Mechanisms such as translational repression by miRNAs, destabilization by ubiquination or other mechanisms of proteolysis could all explain the mechanism by which measurement of the mRNAs levels, do not reflect the protein levels in the cells during differentiation.

**Comparison of mRNA and protein measurements of 129/SvEV ES
cells before and after differentiation**

mRNA ratio before and after differentiation (Log2 scale)
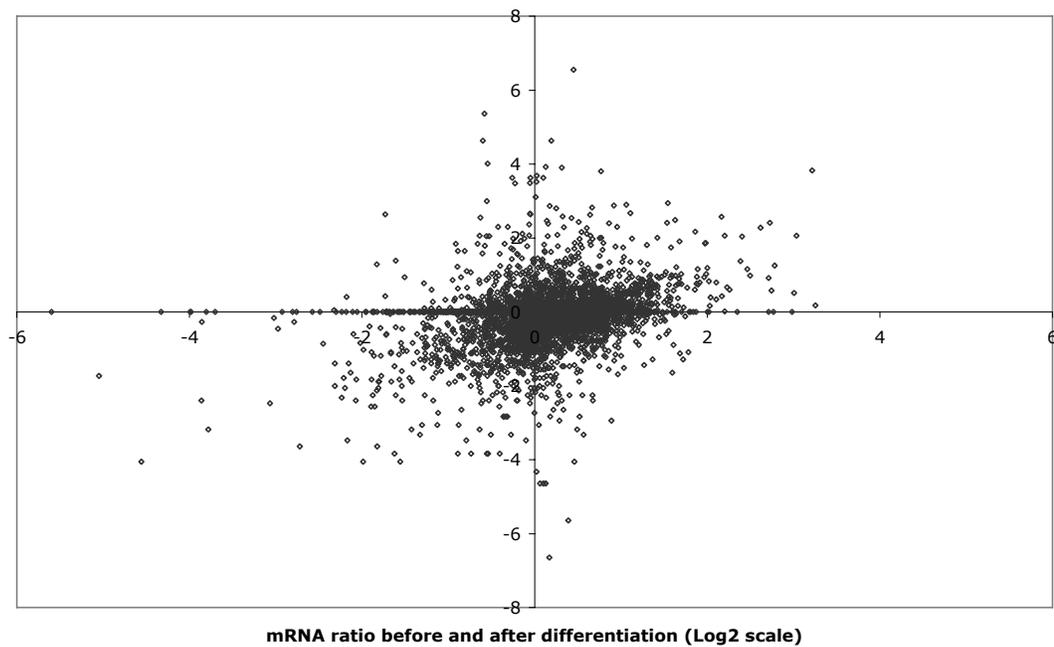
Figure 2.17: Comparison of mRNA and protein gene ratios in ES cell data.
Comparison of ratios of gene products before and after differentiation of mouse
ES cells. Gene products were measured at the protein (Y axis) and mRNA (X
axis) levels. Ratios are graphed in Log_2 scale. Positive values indicate gene
products enriched before differentiation, negative values indicate gene products
enriched after differentiation. A value of 0 indicates that the gene product is
unchanged during differentiation. This dataset includes gene products identified
in both datasets.

**Comparison of mRNA and protein measurements of whole P19 cells
before and after differentiation**

mRNA ratio before and after differentiation (Log2 scale)
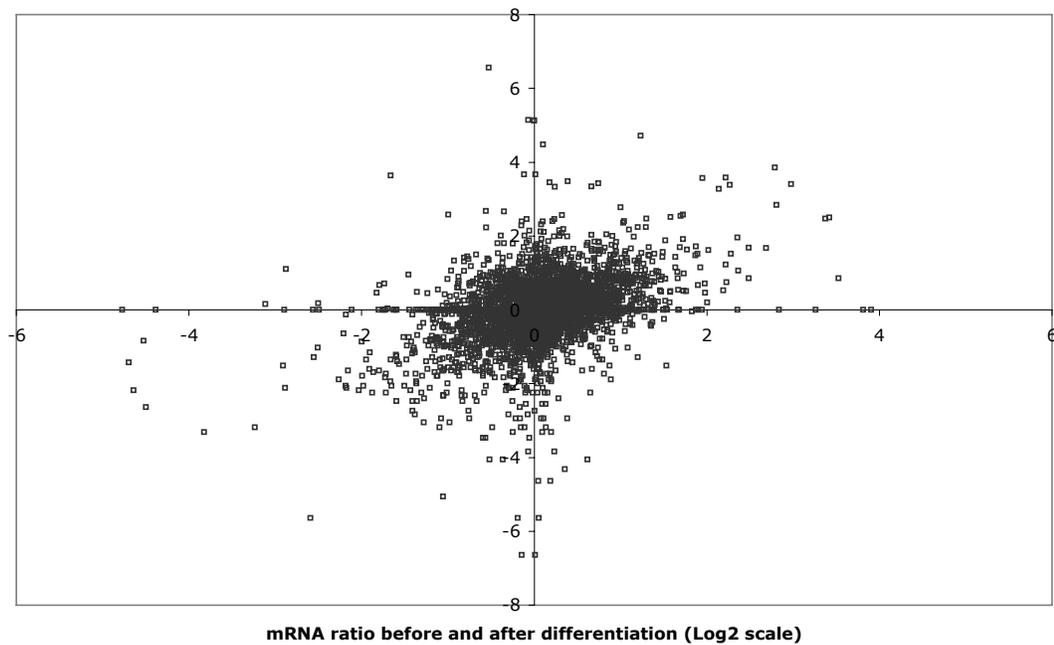
Figure 2.18: Comparison of mRNA and protein gene ratios in EC cell data.
Comparison of ratios of gene products before and after differentiation of P19 EC
cells. Gene products were measured at the protein (Y axis) and mRNA (X axis)
levels. Ratios are graphed in Log_2 scale. Positive values indicate gene products
enriched before differentiation, negative values indicate gene products enriched
after differentiation. A value of 0 indicates that the gene product is unchanged
during differentiation. This dataset includes gene products identified in both
datasets.

I validated select cases of putative post-transcriptional regulation by confirming the observations from microarray and mass spec by measuring protein and mRNA levels by RT-PCR and western blot. I chose Pdlim7, H2afy, Dpysl2, and Sall4. In all cases, the results of the western blots agreed with my iTRAQ data: Sall4 and Pdlim7 were enriched before differentiation, while Dpysl2, H2afy were enriched after differentiation. The RT-PCRs did not, in all cases confirm the observations made in the microarray studies: Sall4, Dpysl2 and H2afy all changed at the mRNA level in correspondence with the observed protein results, while Pdlim7 mRNA was unchanged before and after differentiation, confirming that Pdlim7 is subject to post-transcriptional regulation during RA mediated differentiation.
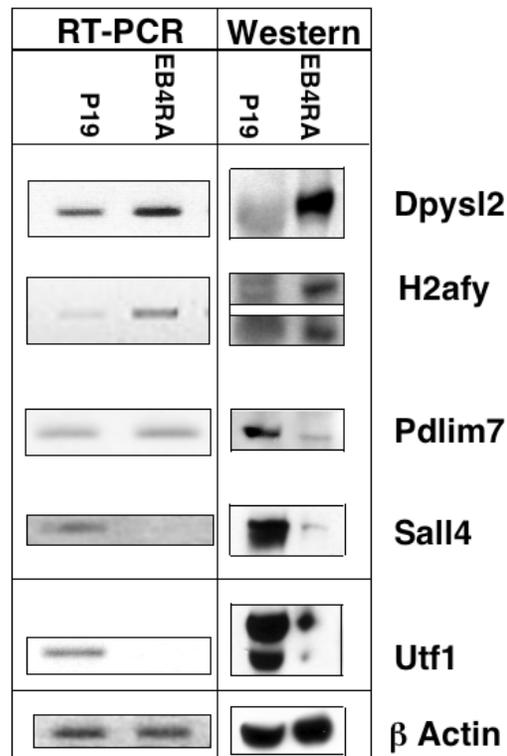
Figure 2.19: Confirmation of cases of post-transcriptional regulation. Validation of cases of post-transcriptional regulation of proteins enriched in P19 cells. P19 cells were differentiated as described above and cells were collected before (P19) and after differentiation (EBRA4) for either RT-PCR or western blot analysis. Of the putative cases of post-transcriptional regulation tested, all proteins' levels agreed with my observations from the whole proteome analysis, while two disagreed with the measurements of mRNA levels from the Aiba et. al. data. Pdlim7 is highly enriched before differentiation of P19 cells, with no change in its' underlying mRNA level. Dpysl2 protein is highly enriched after differentiation, while its mRNA is only slightly increased. In contrast, Sall4 and H2afy were predicted to be subject to post-transcriptional regulation, however their underlying mRNAs behaved like the protein.

Table 2.6: Putative cases of post-transcriptional regulation Putative cases of post-transcriptional regulation identified in both ES and EC cells.

Gene Symbol	Protein	ES/EB average mRNA ratio	ES/EB mRNA ratio ttest	P19/P19RA4 average mRNA ratio	P19/P19RA4 mRNA ratio Ttest	P19 NER un/diff-average	P19 NER un/diff-p-value	P19 un/diff-average	P19 un/diff-p-value	mE un/diff-average	mES un/diff-p-value
Bcat1	"Bcat1 Blastocyst blastocyst cDNA, RIKEN full-length enriched library, clone:I1C0025P20 product:branched chain aminotransferase 1, cytosolic, full insert sequence"	0.33	0.1206697	0.7	0.13310282	2.13	0.023628	2.53	0.01014	2.61	0.109019
Pdlim7	Pdlim7 Isoform 1 of PDZ and LIM domain protein 7	0.58	0.0755759	0.68	0.04243214	2.01	0.13126	4.67	0.01833	2.15	0.057495
Usp36	Usp36 ubiquitin specific protease 36 isoform 1	0.83	0.3934395	0.95	0.24865924	3.2	0.110634	1.68	0.03164	3.3	0.213123
Aurkb	Aurkb Serine/threonine-protein kinase 12	0.92	0.3278959	1.01	0.88339408	1.88	0.107303	2.11	0.03143	3.95	0.5
Plk1	Plk1 Serine/threonine-protein kinase PLK1	1.05	0.1206617	1.02	0.71203257	2	0.006224	3.2	0.04116	1.67	0.667248
Wdr33	Wdr33 Putative WDC146	1.07	0.5726907	1.06	0.23603713	1.59	0.088334	1.68	0.14047	1.52	0.004006
Sall4	Sall4 Isoform 1 of Sal-like protein 4	1.15	0.3454373	1.06	0.52238851	4.35	0.077465	4.98	0.0163	1.96	0.117391
Coro1a	Coro1a Coronin-1A	0.85	0.1155116	1.08	0.38640128	3.45	0.000993	4.32	0.5169	1.62	0.195029
Rbpms	Rbpms RNA binding protein gene with multiple splicing isoform 2	1.08	0.2320028	1.09	0.30604677	3.5	0.122506	3.2	0.19157	1.61	0.700738
Jup	Jup Junction plakoglobin	1.2	0.0118729	1.12	0.03935139	1.75	0.121573	1.57	0.15936	4.24	0.007356
Prdx2	Prdx2 Peroxiredoxin-2	0.8	0.3227094	0.96	0.6569543	0.53	0.025836	0.59	0.00152	0.46	0.02847
Dpysl5	Dpysl5 Dihydropyrimidinase-related protein 5	0.81	0.1559589	1.04	0.48683657	0.36	0.084803	0.36	0.0052	0.4	0.046146
Alad	Alad Delta-aminolevulinic acid dehydratase	0.83	0.1117185	0.89	0.23229857	0.47	0.269911	0.26	0.00701	0.63	0.021667
Etfa	"Etfa Electron transfer flavoprotein subunit alpha, mitochondrial precursor"	0.84	0.2586448	0.91	0.08523864	0.44	0.042434	0.6	0.00054	0.62	0.026948
Syne2	Syne2 synaptic nuclear envelope 2	0.84	0.23113	1.01	0.84967953	0.35	0.002968	0.27	0.01618	0.25	0.156363
Akr7a5	Akr7a5 Aflatoxin B1 aldehyde reductase member 2	0.86	0.3490434	0.81	0.07486269	0.48	0.165855	0.52	0.02838	0.53	0.072746
Vamp4	Vamp4 Vesicle-associated membrane protein 4	0.86	0.2402478	0.85	0.11387194	0.62	0.277845	0.63	0.48879	0.61	0.048969
Atp5f1	"Atp5f1 ATP synthase B chain, mitochondrial precursor"	0.88	0.3594421	1.08	0.42765883	0.61	0.068199	0.51	0.11571	0.64	0.00503
Nucb1	Nucb1 Nucleobindin-1 precursor	0.89	0.2835611	1.21	0.2642621	0.4	0.084745	0.57	0.09368	0.51	0.078301
Sh3bgrl	Sh3bgrl SH3 domain-binding glutamic acid-rich-like protein	0.89	0.1744672	1.04	0.15529893	0.29	0.034771	0.33	0.00243	0.38	0.114086
Ywhab	Ywhab Isoform Long of 14-3-3 protein beta/alpha	0.9	0.3462345	1.08	0.1333175	0.63	0.103015	0.62	0.01808	0.57	0.015464
Aco2	"Aco2 Aconitate hydratase, mitochondrial precursor"	0.91	0.4437561	1.24	0.01654988	0.51	0.057428	0.64	0.01696	0.56	0.013181
Decr1	"Decr1 2,4-dienoyl-CoA reductase, mitochondrial precursor"	0.92	0.0915834	1.03	0.13149011	0.55	0.03657	0.33	0.0051	0.36	0.000559
H2afy	H2afy Isoform 2 of Core histone macro-H2A.1	0.93	0.2774577	0.93	0.19307652	0.32	0.053524	0.24	0.00651	0.4	0.027502
Asph	"Asph Adult male urinary bladder cDNA, RIKEN full-length enriched library, clone:9530097J16 product:aspartate-beta-hydroxylase, full insert sequence"	0.94	0.0863566	1.01	0.77511995	0.58	0.265814	0.5	0.09288	0.5	0.147845
Gnb1	Gnb1 Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta 1	0.95	0.2761227	0.86	0.18067315	0.49	0.146997	0.52	0.00632	0.5	0.015715
Git1	GIT1	0.96	0.6085855	1.05	0.31211718	0.37	0.15319	0.63	0.13874	0.45	0.922851
Ndufs2	"Ndufs2 NADH dehydrogenase [ubiquinone] iron-sulfur protein 2, mitochondrial precursor"	0.97	0.4481206	0.95	0.26582667	0.47	0.016359	0.57	0.00803	0.56	0.028184
Ugcgl1	Ugcgl1 UDP-glucose:glycoprotein glucosyltransferase 1 precursor	1	0.9752455	1.08	0.33435986	0.64	0.109101	0.63	0.04553	0.56	0.033934
Fh1	"Fh1 Isoform Mitochondrial of Fumarate hydratase, mitochondrial precursor"	1.01	0.3358507	0.9	0.11287625	0.64	0.153434	0.46	0.00806	0.59	0.010275
Hk1	Hk1 Isoform HK1-SA of Hexokinase-1	1.01	0.7702729	1.15	0.00372279	0.64	0.482109	0.52	0.10978	0.54	0.308834
Ndufa8	Ndufa8 NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 8	1.02	0.8813331	0.91	0.06614536	0.56	0.068811	0.65	0.04178	0.55	0.159342
Trp53i11	"Trp53i11 13 days embryo heart cDNA, RIKEN full-length enriched library, clone:D330045A08 product:tumor protein p53 inducible protein 11, full insert sequence"	1.02	0.7199139	1.15	0.02592206	0.22	0.017496	0.1	0.10199	0.24	0.162025
Dpysl2	Dpysl2 Dihydropyrimidinase-related protein 2	1.04	0.0841299	1.13	0.1093435	0.32	0.062411	0.27	0.00163	0.29	0.016824
Bnip1	Bnip1 Vesicle transport protein SEC20	1.04	0.3808887	0.89	0.07485113	0.47	0.278288	0.64	0.02763	0.04	0.095478
Atp5f1	"Atp5f1 ATP synthase B chain, mitochondrial precursor"	1.05	0.5232486	0.91	0.00619146	0.61	0.068199	0.51	0.11571	0.64	0.00503
Cox7c	"LOC100048613,Cox7c Cytochrome c oxidase subunit 7C, mitochondrial precursor"	1.09	0.3417473	0.93	0.36289073	0.43	0.005235	0.45	0.17444	0.43	0.014092
Sdhc	"Sdhc Succinate dehydrogenase cytochrome b560 subunit, mitochondrial precursor"	1.11	0.1327288	1.03	0.67468681	0.47	0.110024	0.49	0.1471	0.6	0.934218
Mtap2	Mtap2 Microtubule-associated protein 2	1.14	0.197232	1.03	0.58229393	0.3	0.172349	0.21	0.12224	0.3	0.145573
Tbcd	Tbcd Tubulin-specific chaperone D	1.21	0.022053	0.98	0.49782665	0.41	0.204649	0.55	0.00101	0.53	0.04972
Set	Set Isoform 2 of Protein SET	3.06	0.011027	2.9	0.00035992	0.49	0.089125	0.35	0.01994	0.54	0.01364

2.3.9 Confirmation and functional analysis of a protein displaying evidence of post-transcriptional regulation during RA mediated ES differentiation

One case of putative post-transcriptional regulation occurred on a protein with rather marginal identification quality but that was unchanged in both Aiba et al. microarray datasets.

We chose this gene, Racgap1 to test my absolute quantification via MRM method. I first confirmed that Racgap1 mRNA is unchanged during differentiation by RT-PCR (Figure 2.20), and then used MRM to measure several different transitions from several peptides already identified in the initial mass spec analysis. MRM confirmed that Racgap1 is 2-fold enriched at the protein level before the differentiation of mES cells (Figure 2.22). I then asked if Racgap1 is necessary for ES cell self-renewal. I employed a library of 5 shRNAs from Open biosystems (www.openbiosystems.com) and assessed the effectiveness of each shRNA by cotransfection of HEK293 cells (Figure 2.23), (which contain none of the 5 shRNA target sequences) with a Flag-tagged Racgap1 containing expression vector and blotted for Flag. I determined that the shRNA sequence #2 was most effective in the knockdown, while 3 and 5 were somewhat effective. ShRNA #4 was marginal and Sh1 did not affect levels of recombinant transfected ES cells with these or an unrelated control vector. I then transfected three replicates of mES cells with the vectors, selected for transfected cells overnight and passaged the cells onto 6 well plates and allowed them to grow for three days (Figure 2.22). After three days plates were immunostained rabbit anti Oct4 (1:100, Santa Cruz), and DAB stained (Pierce cat #36000) to quantify Oct4+

colonies. Cells transfected with the most effective vectors (shRNAs #2, 3 and 5) resulted in the fewest colonies, significantly less than either the control (p=0.00077, 0.00037 and 0.00305 respectively), or the ineffective shRNA (sequence number 1) (p=0.022, 0.024 and 0.048 respectively). A separate experiment selecting for HEK293 cells, (with no mouse Racgap1 target mRNA sequence) confirmed that the plasmids were equally able to confer puromycin resistance on the cells (data not shown). I therefore propose that the reduction in ES colony number is due to loss of Racgap1 and potential for self-renewal of the mouse ES cells presumably due to failure of the cell cycle.
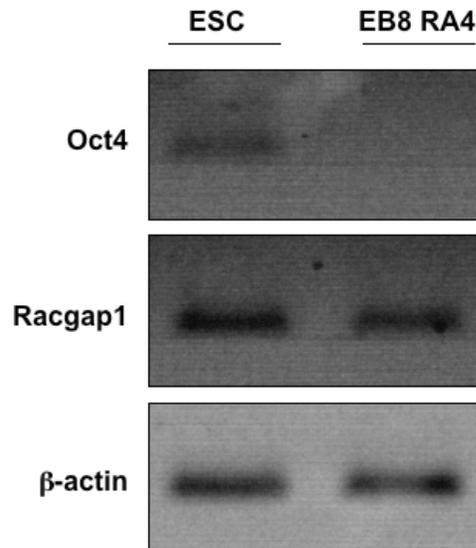
Figure 2.20: Racgap1 mRNA is unchanged during differentiation. RT-PCR of Racgap1, Oct4 and beta actin levels before and after differentiation. PCR reactions were performed at 25, 27 and 29 cycles to prevent saturation. Absence of genomic contamination was ensured by PCR with GAPDH primers on –RT samples (not shown). In addition, Racgap1 primers were designed with the 3' bases spanning an exon-exon junction.
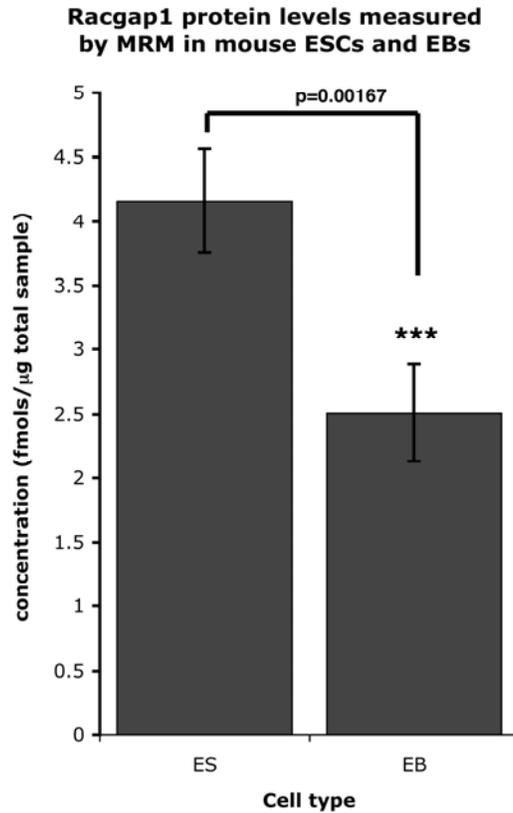
Figure 2.21: Racgap1 protein is enriched before differentiation. MRM analysis of Racgap1 levels in ES cells and EBs. Racgap1 protein levels significantly decrease upon differentiation of ES cells. Synthetic standard peptides from Racgap1 and Actin were used for absolute quantitation of both proteins' levels. Racgap1 levels were normalized to Actin levels. Racgap1 levels decline from 4 femtomoles per μg of total cellular protein to 2.5 femtomoles per μg of total cellular protein.

Colony formation of ES cells after Racgap1 knockdown
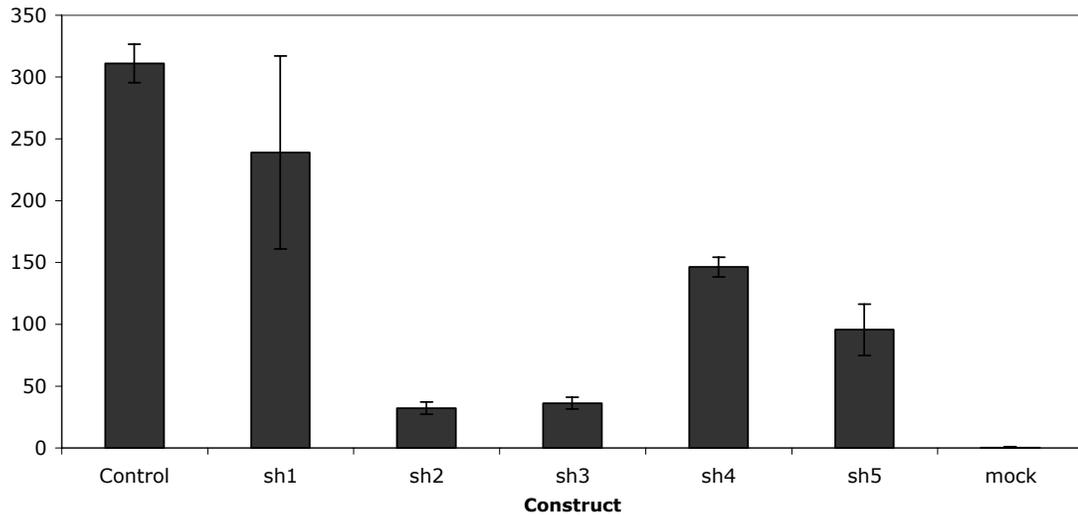
	Control	sh1	sh2	sh3	sh4	sh5	mock
T-test against Control	NA	0.1547	0.0007	0.0004	0.0032	0.0030	0.0004
T-test against shRNA1	0.1547	NA	0.0217	0.0238	0.0742	0.0484	0.0168

Figure 2.22: Knockdown confirms Racgap1 has a role in colony formation. Knockdown of Racgap1 via transient transfection with shRNA decreases ES cell colony formation. 10^5 ES cells were transfected with Lipofectamine 2000, shRNA vector containing a puro^r marker gene or an unrelated control plasmid containing a puro resistance gene. After 24 hours, cells were cultured in the presence of 1µg/ml puromycin for 48 hours and then cultured for an additional 4 days in the absence of puromycin. Colonies were stained with anti-Oct4 and DAB and counted. The experiment was performed in triplicate.
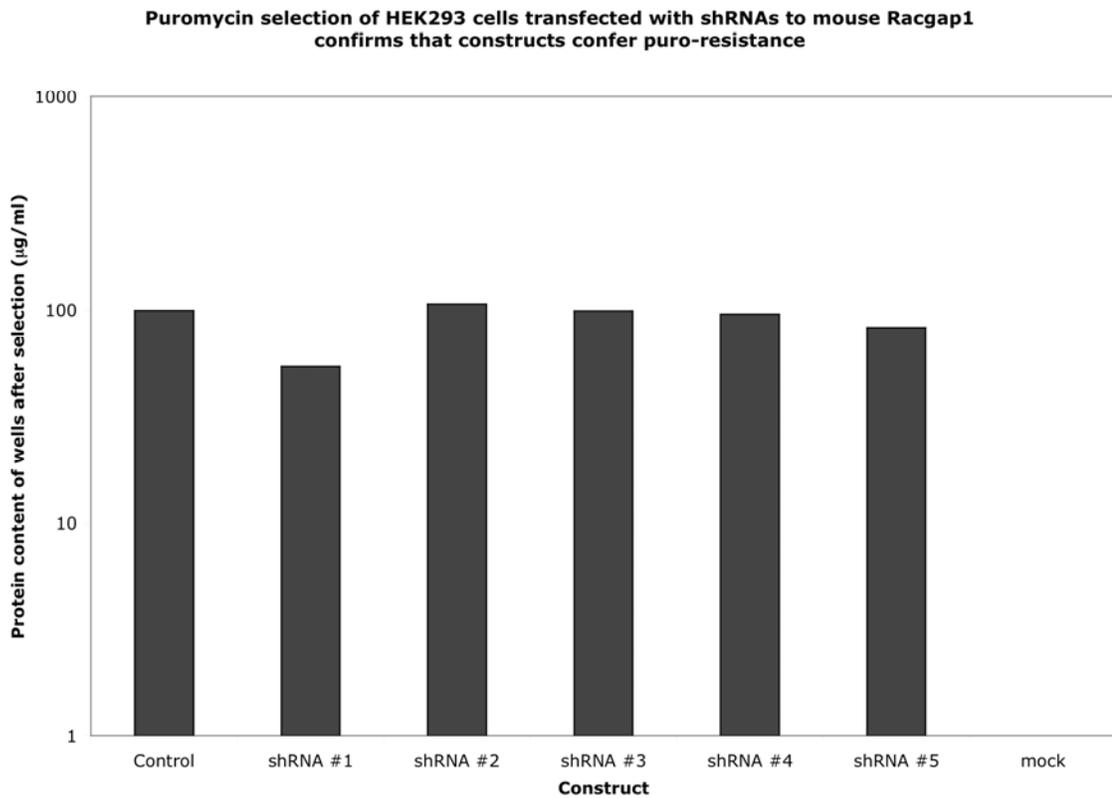
Figure 2.23: Transfection of shRNA plasmids confers similar puromycin resistance to HEK293 cells Transfection of HEK293 cells with shRNA plasmids and selection with Puromycin confirms that plasmids confer puromycin resistance. Cells were transfected with the indicated plasmid, and after 24 hours, were selected with 1μg/ml puromycin for 48 hours. Cells were collected in RIPA buffer and protein concentrations measured by Bradford assay. Readings were blanked with the mock samples, which contained no cells by 48 hours.

2.4 Discussion

2.4.1 Peptide identification and relative quantitation by iTRAQ

The iTRAQ method has several advantages over other peptide and protein labeling methods. First, my method of labeling samples occurs in vitro, after harvesting, for that reason I require no special media as in SILAC. A benefit of this method is that it can be used on samples from patients or collected from full organisms.

2.4.2 Proteins associated with the undifferentiated state

Proteins enriched in undifferentiated cells in all three datasets included proteins already known to be associated with pluripotency, including Oct4, UTF1, Tcf3 and Dnmt3b. I do not detect other known factors such as Nanog and Ronin (Thap11) likely due to pattern of peptides that result from trypsinization. To better understand and categorize the types of proteins enriched before differentiation in all three datasets, I used the Gene Functional Classification tool from the Database for Annotation, Visualization and Integrated Discovery (DAVID)[160, 161] (found at http://david.abcc.ncifcrf.gov/home.jsp) to characterize types of proteins enriched before differentiation. The proteins fell into general classes: rRNA processing and ribosome assembly, general RNA binding, transcriptional regulators, ATP dependant helicases and serine/threonine kinases (all involved in the cell cycle).

2.4.3 Proteins associated with RA-mediated differentiation

In addition to RA-responsive genes such as Crabp1 and 2, I used the Gene Functional Classification tool from DAVID to identify classes of proteins

enriched after RA-mediated differentiation in all three datasets (Figure 2.15). These clusters included proteins involved in the TCA cycle and general catabolism/metabolism, a class of dihydropyrimadinase-like proteins (dhpyl 2,3,4 and 5) involved in axon guidance and dendrite projection, cytoskeletal proteins, Ca^{2+} binding proteins, chromatin structural proteins of the H1/H5 family and small GTPases involved in signal transduction.

2.4.4 Putative cases of post-transcriptional regulation

An advantage of global proteome measurements is the ability to identify post-transcriptional regulation of gene products during differentiation. I replicated an experimental design from Aiba et al. 2006, and by comparing my quantitative iTRAQ data with their quantitative Agilent microarray data, I was able to identify cases where my datasets disagreed. I chose four of these genes to confirm by RT-PCR and western blot. In all cases, western blots confirmed my proteome data, however in three of the cases, changes in protein levels were accompanied by changes in mRNA levels, contradicting the microarray-based measurements. The one exception was Pdlim7, which is indeed unchanged at the mRNA level during the differentiation of P19 cells, but is downregulated at the protein level during differentiation of P19 cells. Pdlim7 is thought to be a protein scaffold that assembles assorted signaling molecules including Protein Kinase C (PKC) on the actin cytoskeleton to mediate signaling during various types of development.

One protein on my list large confirms an observation from Sampath et al. 2008[111], which reported several classes of mRNAs undergoing post-transcriptional regulation in mouse ES cells. They observed proteins that were

unchanged The case of best agreement is the 14-3-3 protein Ywhab, which in my data and the Sampath data does not change at the mRNA level, but is upregulated upon differentiation. Other cases do not agree, however the supplied list of cases from the publication is quite short. Base upon my protein data and the Sampath Ywhab is very likely to be subject to negative post-transcriptional (translational) regulation that is relieved during differentiation. The consequences of this regulation have not been investigated.

2.4.5 Role of Racgap1 in ES cell self-renewal

Previous studies have determined that Racgap1 is necessary for cell division; it is phosphorylated by aurora kinase B and becomes a RhoA GAP rather than a Rac1 or cdc42 GAP[162-165]. There seem to be several other mechanisms to limit its activity toward cdc42 and rac[166]. During mouse development, Racgap1 is expressed in the brain/ectoderm; Racgap1 is necessary for normal development of the pre-implantation embryo. Embryos that lack Racgap1 do develop past the zygote stage, but fail to form an ICM. Conditional knockout of the gene in HSCs and B cells leads to failure to proliferate or differentiate and leads to apoptosis, but apparently independent of the GAP activity of the protein[108, 167].

We observed that Racgap1 mRNA is expressed at the same level in undifferentiated and differentiated cells ES cells, yet MRM analysis determined that the protein is twofold enriched in undifferentiated cells. Knockdown of the gene results in a significant loss of Oct4+ colony formation that correlates with the level of knockdown observed in a heterologus system.

The phenotype associated with Racgap1 may be due to the fact that the gene has been reported to be necessary for cell division; I conclude that Racgap1 is a post-transcriptionally regulated protein necessary for ES cell self-renewal. The mechanism by which Racgap1 is post-transcriptionally regulated is still an open question. A search of the micro-RNA target database (found at www.microRNA.org)[168] identified 57 putative miRNA binding sites, however none of the 10 top predicted miRNAs are expressed in the neural lineage, therefore it is unlikely that the predicted miRNAs are responsible for the post-transcriptional regulation of Racgap1.

2.4.6 Protein level changes in pathways during differentiation of pluripotent cells

2.4.6.1 Ribosome assembly complexes

One of the functional annotation clusters from undifferentiated cells are ribosome biogenesis and rRNA binding proteins. These proteins are enriched in undifferentiated ES and EC cells. Initially, I hypothesized that the cells are rapidly proliferating and thus need to produce many ribosomes, however the Sampah et al study determined that mES cells actually contain fewer ribosomes per cell than cells from EBs[111] (as measured by rRNA content normalized to genomic DNA content) and the ribosomes that are present in ES cells are less loaded with mRNA than cells from EBs. Why, then do are proteins associated with rRNA synthesis and regulation enriched before differentiation?

Many of these proteins also regulation of cellular proliferation independent of ribosome synthesis either via inhibition of p53 mediated growth arrest

(Rps19bp1[169], Gnl3[170, 171], Nolc1[171]), cell cycle (Bop1[172]: PeBoW complex member, Mki67ip[173]) or regulation of chromatin structure/replication (Npm3[174], Noc3l[175-177]) and are associated with several rapidly proliferating cell types. An outlier is Ftsj3, an rRNA methyltransferase thought to be important in embryogenesis. Nola3 (Nop10) is also a member of the telomerase complex[178]. These diverse functions suggest that many of these ribosome synthesis proteins' enrichment is due to their recruitment to perform other tasks in proliferating cells rather than increased ribosome synthesis activity.

2.4.6.2 Metabolism

The most striking biological process upregulated as ES and EC cells differentiate in response to aggregation and ATRA is and increase in pathways responsible for glucose metabolism. This response might be counterintuitive, however ES cells', and many rapidly dividing cancer cell types', main mechanism of energy generation is glycolysis[179-181] (Warburg effect), while oxidative phosphorylation of glucose via the TCA cycle is increasingly recognized to be associated with quiescent cell types such as neurons[182-184]. Thus I observe that the switch from glycolysis to oxidative phosphorylation as pluripotent cells differentiate is consistent between untransformed ES cells and transformed P19 cells.

2.4.6.3 Retinoic acid signaling

ES and EC cells were treated with retinoic acid and aggregation, this treatment lead to an increase in RA response proteins. Both ES and EC cells upregulated Cellular retinoic acid proteins 1 and 2 (Crabp2, Crabp1) and retinol

binding protein 1 (Rbp1). RA signaling leads to increases in Tgfb1 signaling[110, 185, 186], although I do not see an increase in Tgfb1, I do see changes in the receptor (enriched in differentiated EC cells) and in downstream targets of Tgfb1 such as Smad5[187] (enriched in the nucleus of differentiated EC cells). I also see an increase in Smad5 in the nuclear fraction of differentiated cells, suggesting that the cells are indeed subject to increased Tgf beta signaling in response to retinoic acid.

2.4.6.4 Signaling pathways activated during differentiation

There are several signaling molecules enriched before differentiation of all three datasets, these include Aurora kinase A, Aurora kinase B, Bub1 and Polo-like kinase 1. These proteins are all involved in cell division[188-196], and likely reflect the rapid proliferation associated with pluripotent cells.

Differentiated cells, on the other hand, have several types of signaling proteins enriched. A striking class of signaling proteins are the 14-3-3 proteins Ywhab, Ywhaz and Ywhae. These proteins mediate signaling by binding phosphoproteins and mediating translocation, activity or degradation of proteins involved in development, apoptosis and metabolism[197, 198]. Interestingly, these proteins have been shown to interact with two proteins that are enriched before differentiation in all three datasets: Foxo1 and Tjp2, suggesting that these 14-3-3 proteins may be regulating these proteins during differentiation.

Differentiated cells also upregulated Gsk3b, a kinase that is known to be involved in neuronal differentiation[199] as well as a negative regulator of the pro-pluripotency AKT signaling pathway[84, 200].

2.4.6.5 Adhesion

Another group of proteins that change during differentiation are proteins involved in cell adhesion. Undifferentiated ES and EC cells express higher levels of the Embryonic Cadherin (Cdh1) as well as tight junction associated protein Tjp 2 and junction plackoglobin Jup. During differentiation, cells begin to express N-cadherin (Cdh2) and downregulate Jup and Tjp2.

Interestingly, undifferentiated P19 cells, but not mES cells express three Laminin subunits that make up the Laminin-511 complex. Recent publications have identified exogenous Laminin-511 as a potent positive regulator of pluripotency in mouse and human ES cells[81]. Mouse ES cells grown on a matrix of Laminin-511 no longer require external LIF in order to remain undifferentiated. P19 EC cells may maintain their undifferentiated state in the absence of external signals such as LIF due to autocrine Laminin-511 signaling.

2.4.6.6 Transcriptional regulators and chromatin remodeling

Aggregation and RA treatment of ES and EC cells results in downregulation of several transcription factors and chromatin modifying proteins.

These include the well-known ES related transcription factors Oct4, Sall4, UTF1 and Tcf3 as well as the chromatin modifying protein Dnmt3b. Aggregation and RA treatment also results in upregulation of histone family proteins H1/H5 and of transcription factors Hmgb3, Dach1 and Hoxb6.

Both UTF1 and H1/H5 have been implicated in epigenetic transcriptional regulation by remodeling and stabilizing condensed chromatin[201-204]. Changes in these factors likely reflect the changes in chromatin state during differentiation,

resulting in loss of expression of undifferentiated state-specific proteins such as Oct4, Sall4 and Utf1 and induction of neural differentiation-specific proteins such as Crabp1, Crabp2, Nestin, Hoxb6 and N-cadherin.

Chapter 2 includes data submitted for publication at Molecular and Cellular Proteomics (published by the American Society for Biochemistry and Molecular Biology), and was co-authored by Zhouxin Shen, Kiyoshi Tachikawa, Angel Lee and Steven Briggs. The dissertation author was the primary investigator and author of this material.

Chapter 3: Phosphoproteome analysis identifies phosphorylation of pluripotency associated proteins

3.1 Summary

Phosphorylation is a key mechanism of regulating protein activity, stability and localization. Phosphoproteome profiling of pluripotent cells allowed me to identify phosphoproteins associated with pluripotency and to identify sites of phosphorylation of proteins already known to be essential for maintaining the pluripotent state.

I chose phosphorylation of the ES cell specific transcription factor, UTF1, to test for a role in the protein's function. Missense mutation of sites of UTF1 phosphorylation resulted in phosphomimetic (serine to glutamic acid) or phosphonull (serine to alanine) species of the protein, which I used in assays previously developed by the lab of Bart Eggen.

I used strip-FRAP, subnuclear fractionation and confocal microscopy to analyze mutant protein localization. In all cases, both phosphomutant UTF1 proteins localized to the chromatin in a manner identical to wild type protein. To test phosphorylation's role in repression of gene expression, I fused the Gal4 DNA-binding domain (DBD) to UTF1 wild-type or phosphomutant genes and performed luciferase assays in HepG2 (expressing no endogenous UTF1) and ES cells. In both cell types, both mutant proteins maintained similar repressive effect on the UAS-TK-luciferase construct, there were statistically significant

differences, however it remains unclear how physiologically relevant a reproducible 15% difference in repressive activity actually is.

3.2 Introduction

The field of embryonic stem cell research has grown in the last few years, with many significant advances in our understanding of the of the pluripotent state, which has two key features: extensive capacity for self-renewal and the potential to give rise to all somatic lineages.

Our understanding of how this state is maintained is really divided into two sets of information and study: the first focuses on the environmental and signaling factors that maintain pluripotency: These pathways include the LIF/JAK/Stat[77] pathway in the mouse, bFGF in human[205], PI3/AKT[84, 206-208] and ECM/integrin signaling[81, 209].

The other, arguably deeper set of information relates to the transcription factors and chromatin modifiers necessary, and sufficient to establish and maintain a cell in the pluripotent, self-renewing state. The list of transcription factors found to be necessary for pluripotency is extensive, but a short list of proteins necessary for pluripotency include: Oct4[51, 53], Sox2[210], Nanog[57-59], Ronin[73] (Thap11), UTF1[74], Sall4[211-213], Tcf3[140], Suz12[66, 67], Dnmt3b[71], Stat3[77], and β-catenin[214]. Recent work from Shinyo Yamanaka[85, 86] and James Thompson has identified the minimal set of proteins necessary to establish the pluripotent state, the two indispensable proteins are Oct4 and Sox2[215], which form heterodimers and regulate themselves and many other pluripotency related genes, including Nanog, Fgf4 and UTF1[56, 215-217].

In this chapter, I set out to better understand how changes in signaling affects the proteins associated with and necessary for the pluripotent state in the undifferentiated state and during exit from pluripotency.

To accomplish this, I took a global mass-spectrometry approach tailored to enrich and analyze the phosphoproteomes of pluripotent cells before and after aggregation and retinoic acid treatment leading to differentiation into the neuroectodermal lineage. Enrichment on TiO2 columns and spectral counting of phosphopeptides were employed to provide semi-quantitative data to compare cells before and after differentiation. I identified phosphorylation of proteins known to be necessary for maintaining the pluripotent state including UTF1, Dnmt3b, Sox2, Stat3, Lin28, Rex1 and Dlg7. The most abundant of these phosphoproteins was UTF1, to better understand the effect that phosphorylation of UTF1 on protein activity and localization, I used assays previously published in van den Boom et al. 2007[74] to test the effect of abolition and mimicry of phosphorylation on the protein by generation of missense mutant UTF1 proteins with all 7 observed sites of serine phosphorylation mutated to alanine (null) or glutamic acid (mimetic).

3.3 Results

3.3.1 TiO$_2$ columns enrich phosphopeptides for deep phosphoproteome profiling on pluripotent cells before and after differentiation

I set out to identify phosphorylation events specific to undifferentiated ES and EC cells, by comparison of the phosphoproteome profiles of ES and EC cells before and after differentiation. Tryptic phosphopeptides were enriched by affinity

purification on elemental TiO2 columns, and analyzed by liquid chromatography and MS/MS analysis. Unlike unmodified proteome profiles, phosphoproteins were quantified by comparison of spectral counts of phosphopeptides from individual proteins. In the P19 EC cells I saw a total of 4621 phosphorylation events with a measured FDR of 0.76% at the peptide level. Of the identified phosphorylation events, 84.66% of events occurred on serines, 12.57% on threonines and 2.77% on tyrosines (**Figure 3.1**). In the 129/SvEv mES cells, I saw a total of 6654 phosphorylation events on 2037 proteins with a measured FDR of 1% at the peptide level. Of the identified phosphorylation events, 83.8% of events occurred on serines, 13.6% on threonines and 2.6% on tyrosines.
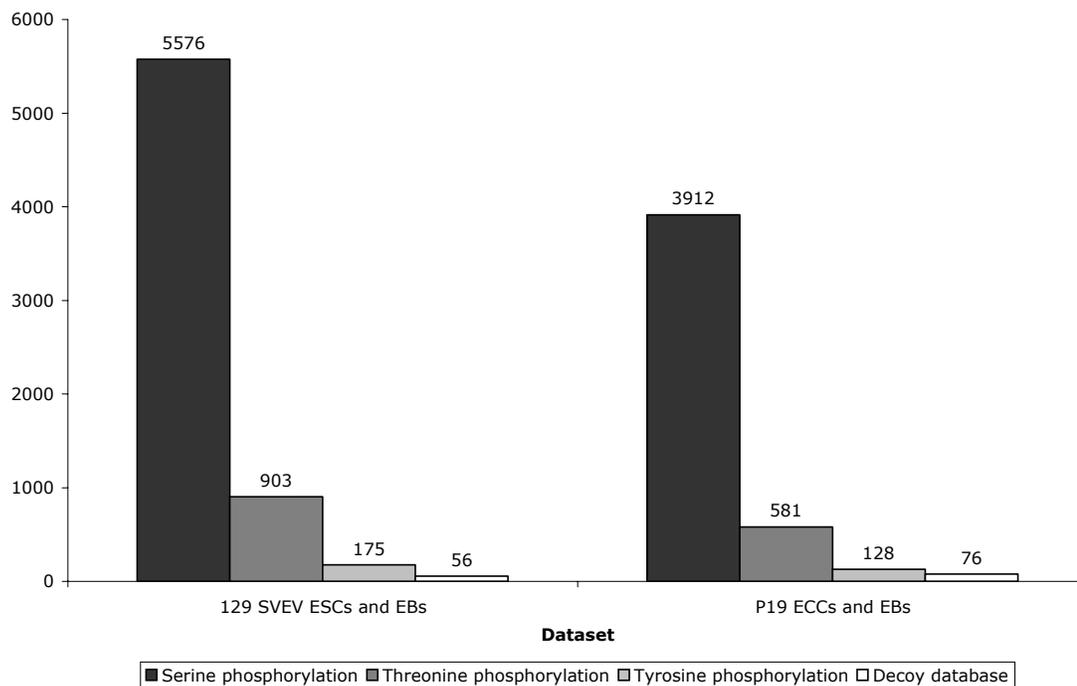
Figure 3.1: Distribution of sites of phosphorylation identified in profiling. Distribution of residues of phosphorylation in ES and EC datasets. Numbers on top of the bars refer to the number of spectra from the dataset that are phosphorylated at the indicated residue. False positive hits from the decoy database are also included.

3.3.2   Phosphorylation of Protein markers of pluripotency and differentiation

Many of the most abundant phosphopeptides in undifferentiated ES and EC cells are also the most abundant of the unmodified peptides in the ES and EC cells (Table 5). These proteins include Utf1, Dnmt3b, Sall4, Rif1 and Kpn2a. In the case of these proteins, whether the phosphorylation events are enriched in the undifferentiated state due to differential activities of kinases acting on the proteins or due to the proteins' enrichment in undifferentiated cells is unclear.

Table 3.1: Phosphorylation of pluripotency associated proteins identified in phosphoproteome data.

Protein name	Marker of pluripotency?	Necessary for pluripotency?	ESC iTRAQ	ESC spectra undiff / diff	P19 iTRAQ	P19 Peptides undiff / diff
Oct4	Y	Y	2.60	**0 / 0**	11.97	**0 / 0**
Thap11 (Ronin)	Y	Y	ND	**0 / 0**	ND	**0 / 0**
Nanog	Y	Y	ND	**0 / 0**	ND	**0 / 0**
Sox2	N	Y	1.20	**0 / 0**	0.71	**3 / 14**
Utf1	Y	Y	5.30	**187 / 28**	5.96	**161 / 1**
Dnmt3b	Y	N	1.89	**21 / 0**	11.98	**166 / 0**
Sall4	Y	Y	2.47	**25 / 5**	4.98	**11 / 0**
Lin28	Y	Y	2.04	**0 / 1**	3.77	**2 / 0**
Kpna2	Y	unknown	1.85	**39 / 9**	2.15	**5 / 4**
Rif1	Y	Y	1.96	78 / 14	2.97	**43 / 6**
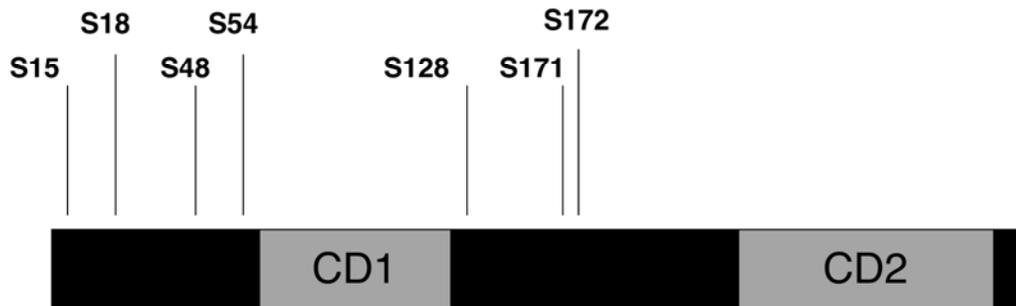
UTF1 phosphorylated at 7 serines flanking conserved domain 1
(putative cmyb/sant DNA binding domain)

Predicted PI of unmodified UTF1:   10.05
Predicted PI of phosphorylated UTF1:   6.36
pl predictions: scansite.mit.edu

Figure 3.2: Sites of phosphorylation identified on mouse UTF1. Sites of phosphorylation of mouse UTF1 protein: all sites of phosphorylation occur at serines toward the N-terminus of the protein, flanking the conserved DNA-binding domain. I used the scansite resource from MIT (*scansite.mit.edu*) to predict the isoelectric point of phosphorylated and non-phosphorylated UTF1. UTF1 is strongly positively charged at physiological pH, like the core histone proteins, but when phosphorylated, it is predicted to be slightly negatively charged.
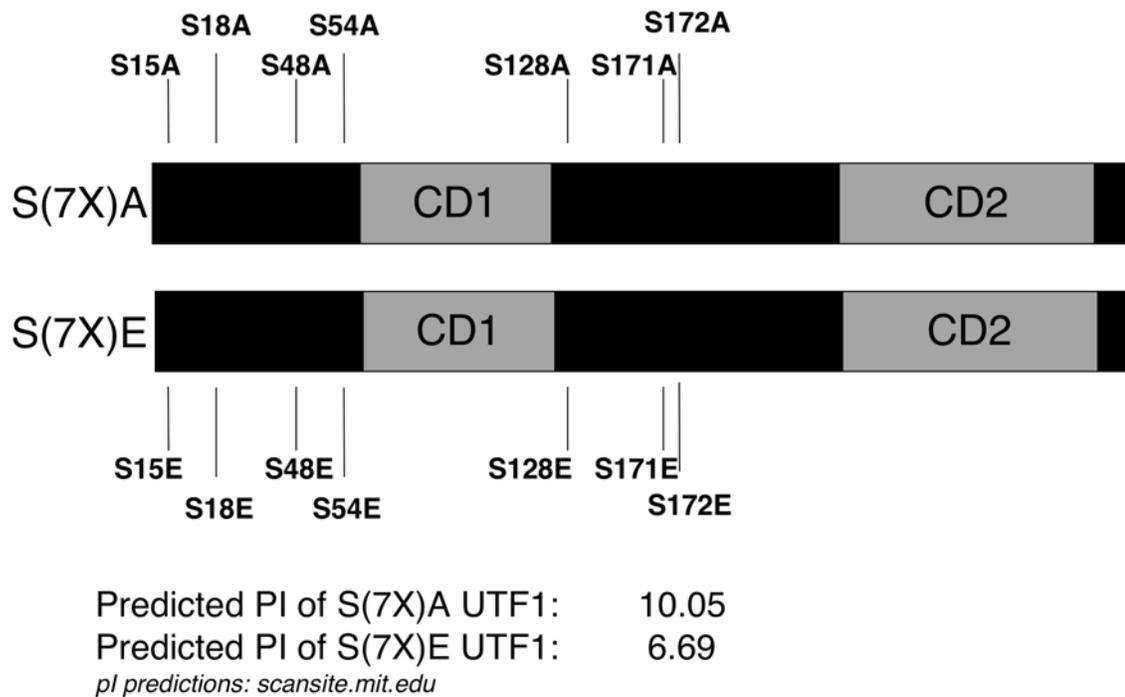
Figure 3.3: UTF1 mis-sense mutant proteins. Phosphomutant UTF1 constructs mimic unmodified UTF1 (S(7X)A) and fully phosphorylated UTF1 (S(7X)E). I created a phosphomimetic missense mutation of UTF1 and a phospho-null missense mutation of UTF1 to test the effect of phosphorylation of UTF1 on the protein's function. Serine to alanine mutation has no effect on the predicted isoelectric point of the protein, while serine to glutamic acid mutation dropped the isoelectric point to 6.69, close to the predicted isoelectric point of the fully phosphorylated protein (pI=6.39).

3.3.3 Mimicry or ablation of phosphorylation sites of UTF1 or mimicry of phosphorylation at those sites does not affect protein binding to the chromatin

I set out to test the importance of phosphorylation of the known pluripotency factor UTF1. UTF1 is a potent repressor of transcription in mouse and human ES cells, and is unusual in that unlike many transcription factors that are only loosely and transiently associated with the chromatin, UTF1 is tightly bound, similar to the core histones[74]. The role of phosphorylation of this protein has not yet been described. I identified 7 sites of serine phosphorylation on UTF1 and mutated these sites to abolish phosphorylation (serine to alanine abbreviated S(7X)A) or to mimic phosphorylation (serine to glutamic acid abbreviated S(7X)E).

UTF1 has been shown to be strongly chromatin interacting, similar to the core histones. UTF1 shares with the core histones a very high pI (figure 3.2), so I hypothesized that mimicry of phosphorylation by mutation of 7 serines to glutamic acids would lower the proteins' affinity for DNA due to the high reduction of pI (the S(7X)E mutant protein has a predicted pI of <7). To test this hypothesis, I generated multiple stable P19CL6 cell lines expressing GFP-UTF1 wt, phospho-mimetic and phospho-null cell lines. With these cell lines I performed 1) Strip-FRAP assays and 2) subnuclear fractionations. In both experiments, the S(7X)A and S(7X)E GFP-UTF1 constructs behaved like the wild-type constructs.

Strip Frap: GFP::Utf1 phosphomutant mobility assay
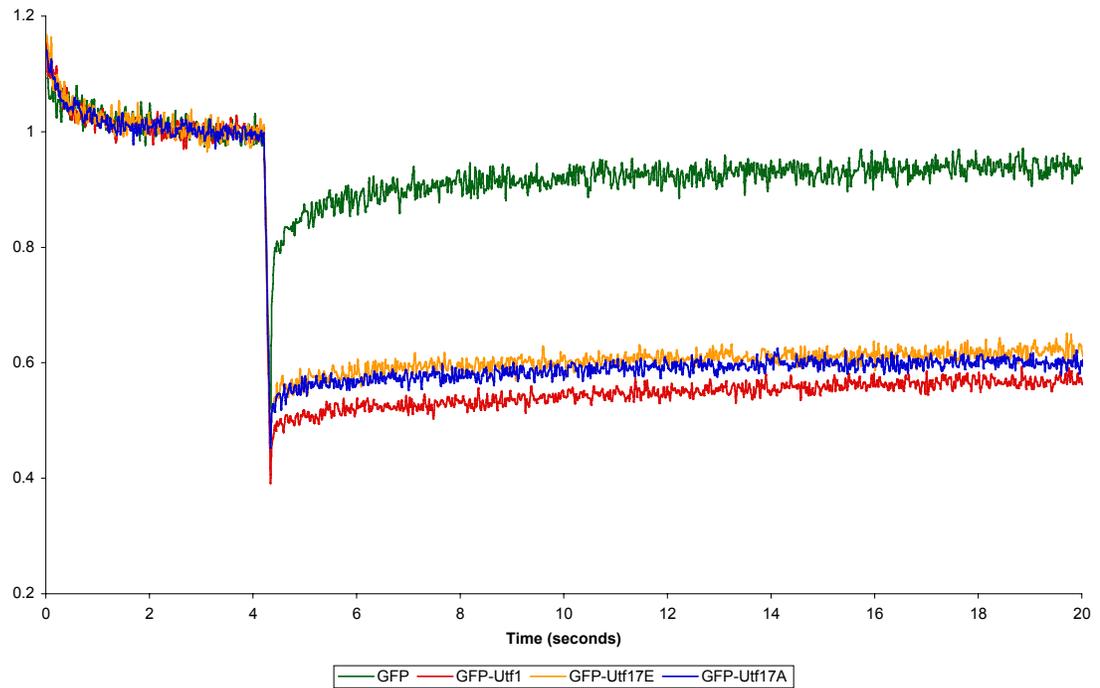
Figure 3.4: Strip-FRAP analysis of UTF1 phosphomutant DNA binding. Strip-FRAP analysis of UTF1 wild-type and phosphomutants reveals no change in the proteins' mobilities. Transgenic P19CL6 cells expressing GFP-UTF1 wt or phosphomutant proteins.

Subnuclear fractionation confirms stable chromatin association
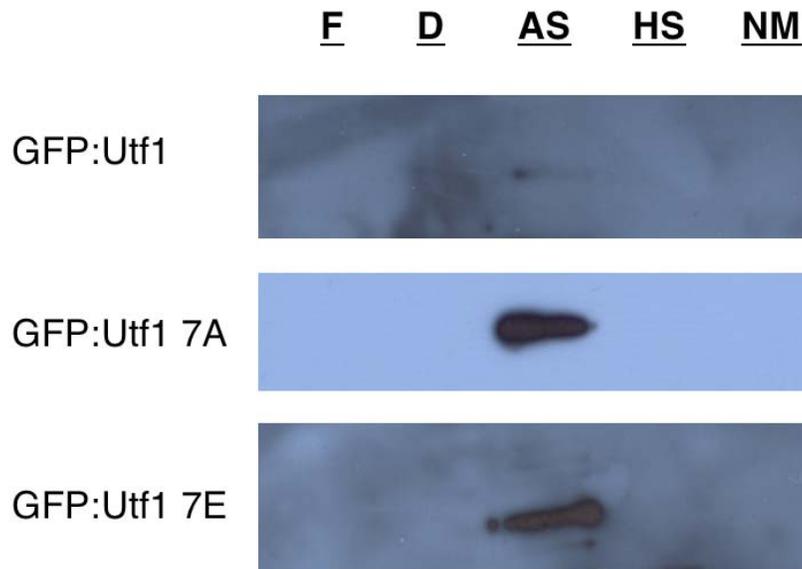
Figure 3.5: Subnuclear fractionation: phosphomutant chromatin binding. Subnuclear fractionation confirms strip frap results. Two transgenic P19CL6 lines for each condition were constructed and subjected to subnuclear fractionation with samples collected at 5 steps: F: Free diffusing and cytosolic fraction. D: Freed by DNAseI treatment, this faction included transcription factors like Oct4. AS: Ammonium Sulfate solublized fraction, includes proteins tightly bound to DNA in chromatin including the four core histones and wild-type UTF1. HS: High Salt fraction, includes proteins weakly bound to the nuclear matrix. NM: Nuclear matrix, includes proteins integral and tightly bound to the nuclear matrix. All three mutants remain tightly bound to the chromatin and elute in the AS fraction.
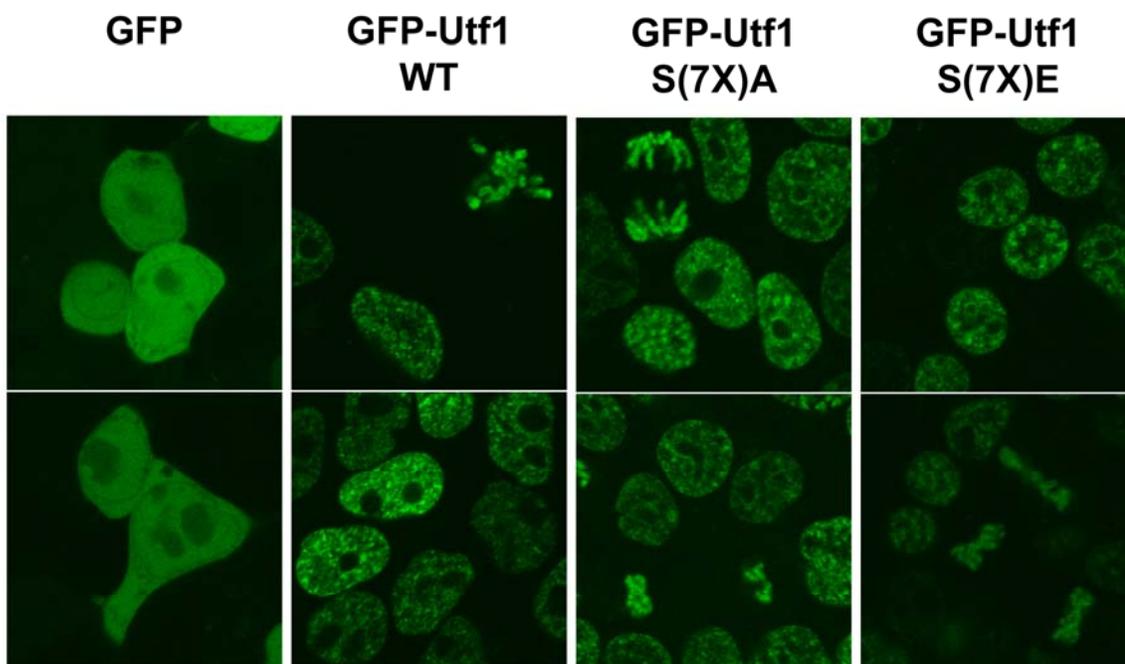
Figure 3.6: Subcellular localization of GFP-UTF1 phosphomutants. Confocal microscopy confirms that subnuclear localization and stable association with chromatin during mitosis is unchanged in UTF1 phosphomutant proteins.

3.3.4 Ablation of phosphorylation sites of UTF1 or mimicry of phosphorylation at those sites weakly alters UTF1 mediated repression of a TK-luciferase construct in ES, EC and HepG2 cells

UTF1 is a potent repressor of gene expression in pluripotent cells. I used a Gal4 reporter assay developed in van den Boom et al.[74] in HepG2 and ES cells. The TK-luciferase reporter constructs were similarly repressed by the three constructs within each experiement (Figures 3.7 and 3.8); despite the general similarity of the repression, the slight differences were statistically significant (Tables 3.2 and 3.3). physiological relevance of these small but significant differences is unclear.

Repression of TK-Luciferase by Utf1 phosphorylation mutants in HepG2 cells
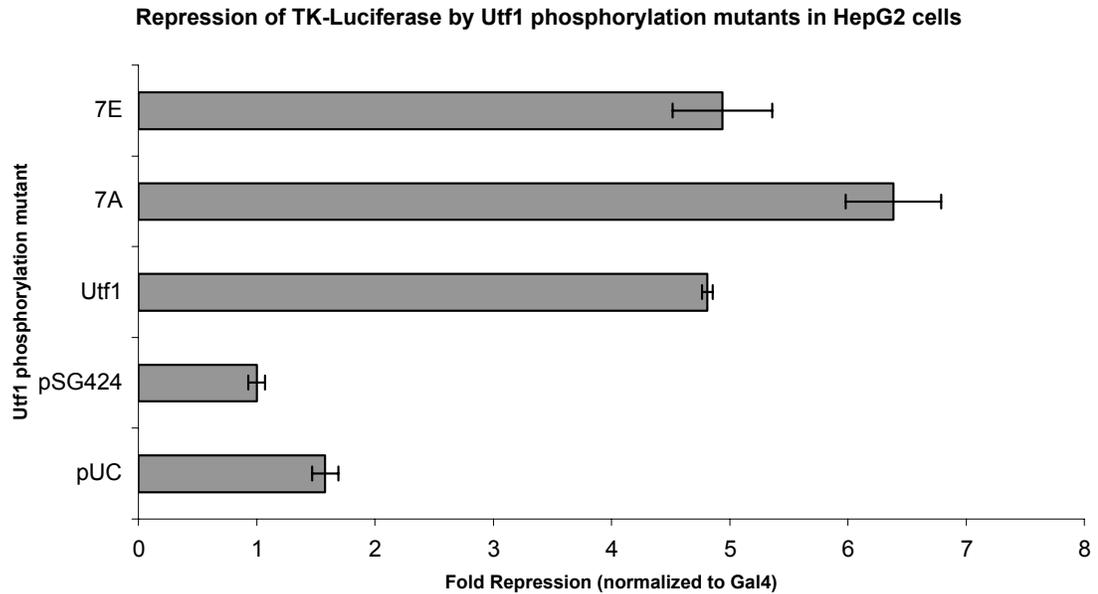
Figure 3.7: Repression of Luciferase reporter by UTF1 phosphomutants. Gal4DBD-UTF1 phosphomutants retain repressive activity when co-expressed with a UAS-TK::Luciferase reporter construct in Hepg2 cells. Luciferase activity was normalized to a LacZ reporter. Repression of the TK::Luciferase activity was normalized to the empty vector, pSG424. Experiments were performed in triplicate and repeated three times. These data are representative of the three experiments.

Table 3.2: T test of TK-luciferase assays in HepG2 cells. Student's t-test of TK-luciferase assays performed in HepG2 cells.

	Puc18	PSG424	WT UTF1	S(7X)A UTF1	S(7X)E UTF1
ttest vs pSG424	0.03978 *	N/A	0.00004 ***	0.00127 **	0.00495 ***
ttest vs WT UTF1	0.00069 ***	0.00004 ***	N/A	0.01787 *	0.66034
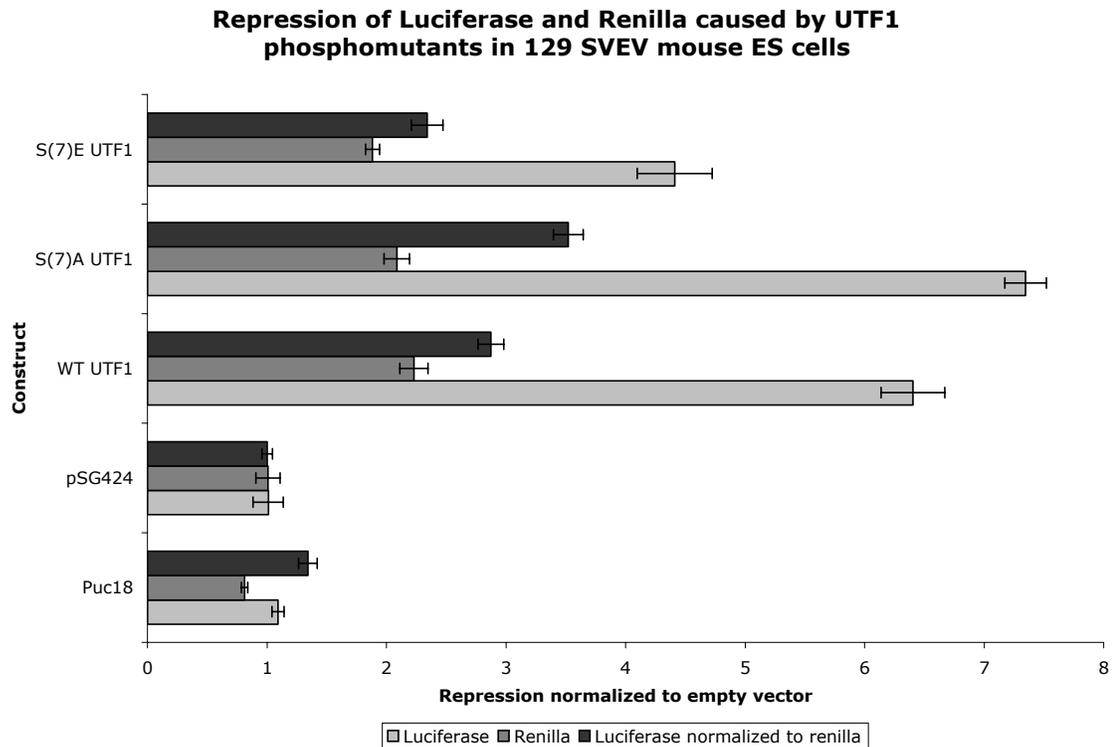
Figure 3.8: Repression of Luciferase reporter by UTF1 phosphomutants. Gal4DBD-UTF1 phosphomutants retain repressive activity when co-expressed with a UAS-TK::Luciferase reporter construct in mouse ES cells. Luciferase activity was normalized to a *renilla* luciferase reporter. Repression of the TK::Luciferase activity was normalized to the empty vector, pSG424. Experiments were performed in triplicate and repeated three times. These data are representative of the three experiments.

Table 3.3: T test of TK-luciferase assays in ES cells. Student's t-test of TK-luciferase assays performed in ES cells.

	Puc18	pSG424	WT UTF1	S(7X)A UTF1	S(7X)E UTF1
ttest vs pSG424	0.00581 **	N/A	0.00052 ***	0.00040 ***	0.00373 ***
ttest vs WT UTF1	0.00012 ***	0.00052 ***	N/A	0.00024 ***	0.01764 *

3.3.5 Wild-type and phosphomutant UTF1 constructs fail to repress Oct4 in ES cells

UTF1 is a potent repressor of the TK promoter, and as demonstrated in Figure 3.8, is able to repress expression of the TK-driven *renilla* construct that lacks the UAS sequence. I wanted to know if UTF1 would have the same repressive activity toward genes highly expressed in ES cells. To test that I co-transfected wt and phosphomutant UTF1 proteins with a 2.5 kb Oct4::Luciferase construct in ES cells.

Perhaps unsurprisingly, all three UTF1 constructs totally failed to repress Oct4-driven luciferase activity in ES cells. I also performed a western blot on material from the same luciferase assay and saw no reduction of native Oct4 protein levels.

WT and phosphomutant UTF1 constructs fail to repress a 2.5 kb Oct4 reporter in ES cells
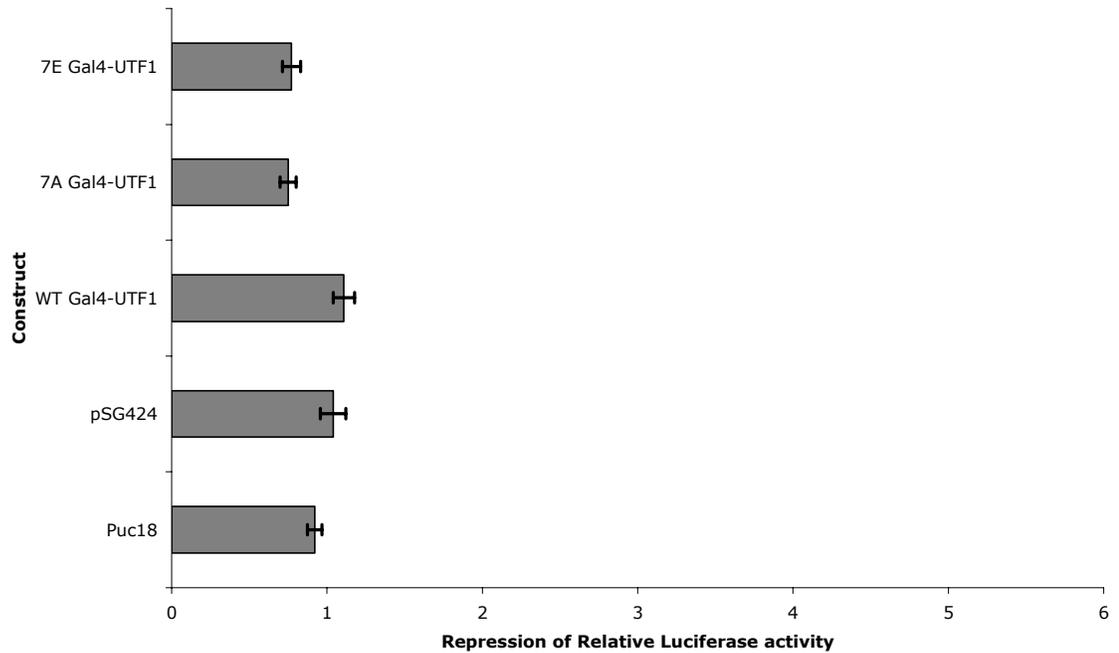
Figure 3.9: UTF1 phosphomutants do not repress Oct4 luciferase. Gal4DBD-UTF1 phosphomutant and wt protein fail to repress luciferase activity when co-expressed with a 2.5KB Oct4::Luciferase reporter construct in mouse ES cells. Luciferase activity was normalized to a *renilla* luciferase reporter. Repression of the 2.5kbOct4::Luciferase activity was normalized to the empty vector, pSG424. Experiments were performed in triplicate and repeated three times. These data are representative of the three experiments. Repression of the firefly luciferase constructs and renilla luciferase constructs are included to demonstrate that wild-type and mutant UTF1 constructs are also capable of repressing the *renilla* control constructs.

Table 3.4: T test of Oct4-luciferase assays in ES cells. Student's t-test of Oct4-luciferase assays performed in ES cells.

	Puc18	pSG424	WT UTF1	S(7X)A UTF1	S(7X)E UTF1
ttest vs pSG424	0.03783 *	N/A	0.04930 *	0.00024 ***	0.00019 ***
ttest vs WT UTF1	0.00033 ***	0.04930 *	N/A	5.4×10^{-7} ***	1.9×10^{-7} ***
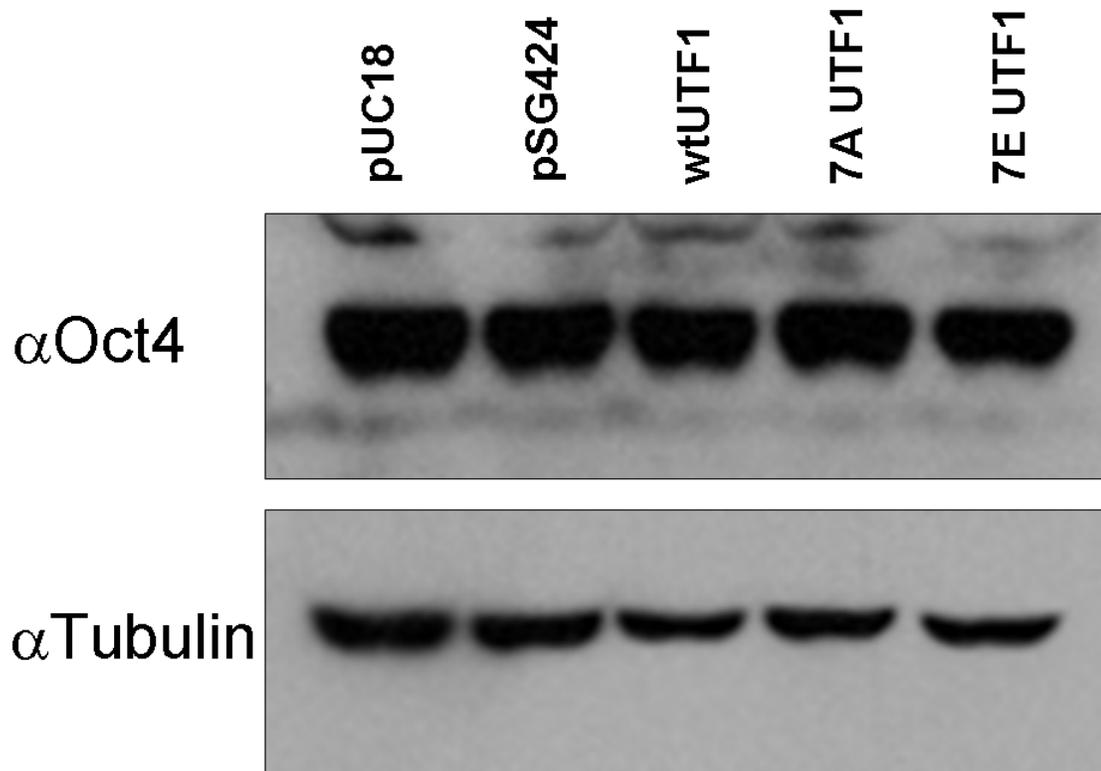
Figure 3.10: UTF1 phosphomutants do not repress endogenous Oct4. Western blot confirms that wild type and phosphomutant UTF1 constructs do not repress endogenous Oct4 levels in ES cells. Material for western blot was collected from the same experiment as figure 3.8.

3.3.6 Phosphomutant and wild-type UTF1 proteins all posses half-lives of greater than 24 hours

A possible explanation for the slight differences between the repressive activity of the UTF1 wild-type and phosphomutant proteins is a difference in protein expression level or stability. To test this, I attempted to directly measure the levels of Gal4-UTF1 in cells transfected for luciferase assays. Unfortunately, expression of the Gal4-UTF1 protein was not sufficiently high to see on western blot. As an alternative, I used my GFP fusion UTF1 protein to measure the half-life of wild-type and phosphomutant UTF1 proteins.

P19 cells expressing UTF1 constructs were treated with 10 μg/ml cycloheximide (CHX) and sampled at 0, 6, 12 and 24 hours after CHX treatment. 10 μg of protein was loaded per well and probed with anti-UTF1, anti-tubulin (a stable protein: loading control) see figure 3.10. In all cases, UTF1 phosphomutant and wild-type proteins' levels fail to decrease after 24 hours. Therefore the half-live of all proteins are greater than 24 hours, so there is no evidence to suggest that phosphorylation has a gross effect on protein stability. The increasing intensity of the UTF1 bands is likely due to the fact that as total cellular protein decreases during CHX treatment, the proportion of stable proteins such as tubulin and UTF1 increases within the sample. To rule out the possibility that phosphorylation may regulate the destabilization of UTF1 during differentiation of pluripotent cells, I also performed the experiment on EC cells that were grown in the presence of 1μM ATRA for 36 hours. The fact that UTF1

protein is stable, yet mostly disappears from EC cells only 4 days after differentiation, suggested to me that there is a process to remove UTF1 protein from the chromatin as the cells differentiate. Yet although endogenous UTF1 is nearly absent from cells after 36 hours on ATRA, the small amount that remains does not seem to turn over either. My only hypothesis that might explain these results is that there may be multiple populations of cells within the well, some of which turn over UTF1 protein quickly, and some of which maintain UTF1 stability.
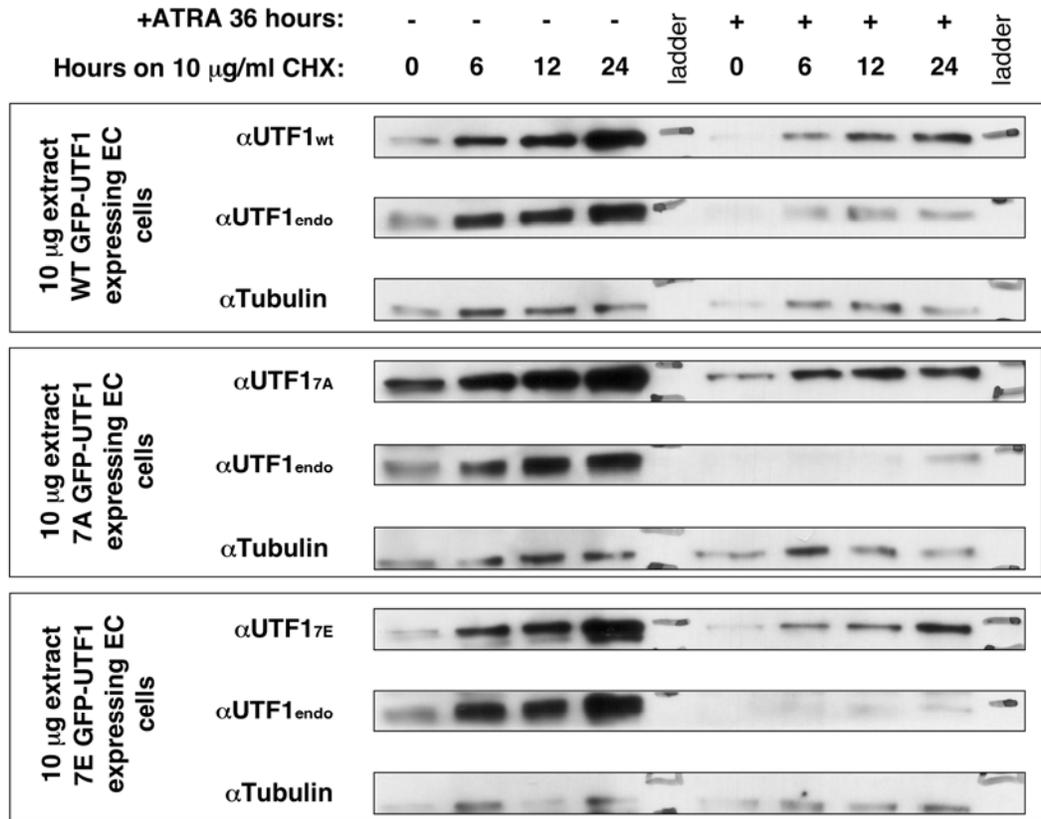
Figure 3.11: Stability of UTF1 phosphomutant proteins. Phosphomutant and wild-type UTF1 proteins posses half-lives of greater than 24 hours. EC cells transiently expressing GFP-UTF1 wild-type and phosphomutant constructs were treated with 10 μg/ml cycloheximide (CHX). Protein samples were taken at 0, 6, 12 and 24 hours post-CHX treatment. 10 μg of Western blot confirmed endogenous and transgenic UTF1 protein levels remained stable across all timepoints, as was the loading control, Tubulin (reported to have a half-life of >50 hours).
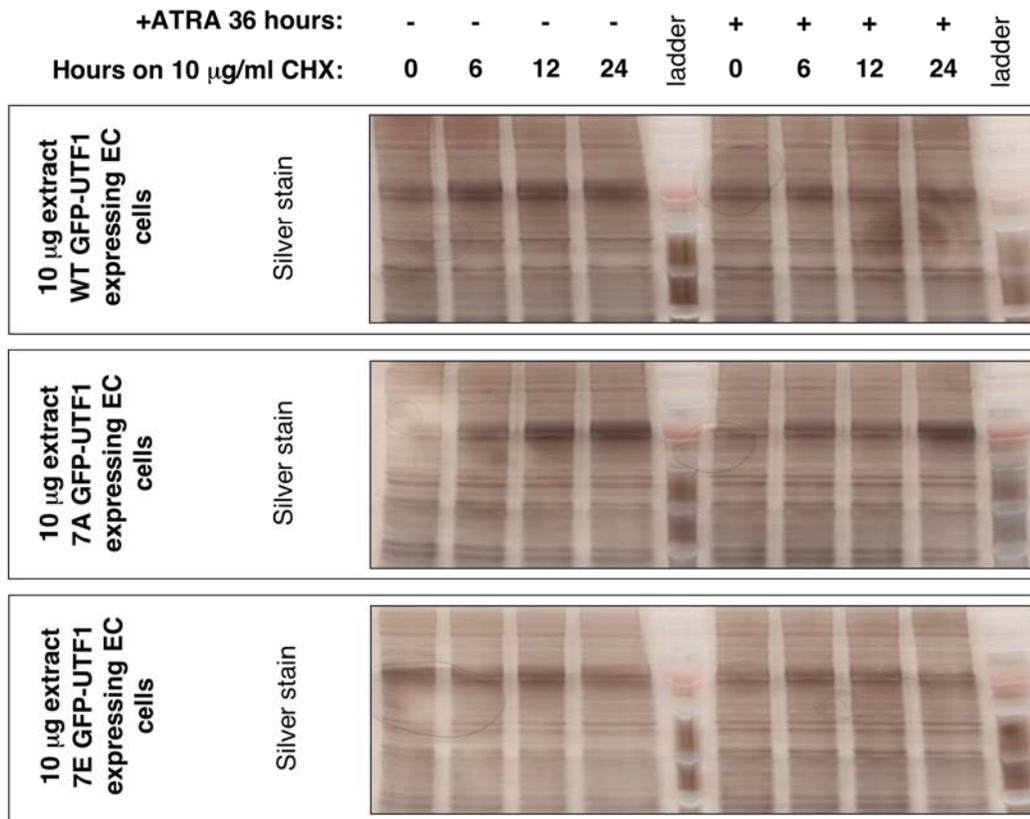
Figure 3.12: Silver stain loading control for figure 3.11. Phosphomutant and wild-type UTF1 protein loading control #2. Increasing Tubulin and UTF1 levels in figure 3.11 are not due to uneven loading, therefore uneven loading does not explain the increase in apparent levels of tubulin and UTF1 over the course of cycloheximide treatment.

3.4 Discussion

3.4.1 Use of comparators for semi-quantitative phosphoproteomes

A deep, semi-quantitative dataset of phosphoproteins and sites of phosphorylation present in ES cells is a valuable resource on its own, but I chose to include comparator cells as well, in order to identify phosphoproteins that are specific to undifferentiated cells.

3.4.2 Phosphorylation of ES specific proteins

The first use of these studies is further refinement of the markers used to characterize pluripotent cells. I previously compared two pluripotent cell types; ES and EC cells, and identified proteins that are enriched in both cell types before differentiation. In this study, I asked whether any of these factors are post-translationally modified by phosphorylation, and indeed several were. Surprisingly, I did not identify phosphorylation of Oct4 or Nanog in ES or EC cells, although it has been reported. This may be due to technical reasons, (i.e. the mass spec simply failed to detect phospho-species of these proteins that are present in undifferentiated cells) or due to the fact that my experimental design only sampled two timepoints during differentiation. A higher-resolution dataset could quite possibly identify phosphorylation of ES specific proteins as the cells turn on signaling pathways that lead to differentiation.

I did identify phosphorylation of several ES cell-specific proteins including known downstream targets of Nanog and the Oct4/Sox2 dimer including UTF1, Sall4 and Dnmt3b. This phosphorylation of downstream targets of Oct4/Sox2 and Nanog could mean that signaling pathways active in undifferentiated cells act on

proteins upstream and downstream of Oct4/Sox2 and Nanog. I suspect that inhibition and activation of signaling pathways that lead regulate ES cells' differentiation, coupled with phosphoproteomics will result in identification of new connections between signaling pathways and phosphorylated sites on proteins important to the pluripotent state.

3.4.3 UTF1 phosphorylation

Phosphorylation and phosphomimetic mutation of the 7 phosphorylated UTF1 residues that I identified in our proteome analysis significantly alter the overall charge of the protein from the basic, positively charged at physiological pH to neutral and slightly negatively charged at physiological pH. This lead us to hypothesize that phosphorylation is disrupting the strong chromatin-binding activity observed in the wild-type protein, and thus reduce the protein's ability to repress target proteins.

Strip-FRAP assays on stable cell lines expression WT, phospho null (S(7X)A) and phospho-mimetic (S(7X)E) caused us to reject this hypothesis. Abolition or mimicry of phosphorylation did not alter GFP-UTF1 chromatin interaction. Small differences in the level of photobleaching reflect differences in GFP-UTF1 expression rather than differences in the protein's binding to the chromatin. These results were confirmed by subnuclear fractionation as described previously. Consistent with the FRAP data and previous work, all three UTF1 proteins remain tightly bound to the chromatin after DNAse I treatment and are only released after treatment Ammonium Sulfate, similar to the core histones.

I tested the role of phosphorylation on UTF1 protein's effectiveness at repressing a UAS-TK Luciferase reporter in ES, EC and HepG2 cells. In all cases, abolition of sites of phosphorylation lead to small, but statistically significant increases in repression while mimicry of phosphorylation lead to small but statistically significant differences in repressive activities. These differences could either be due to differences in the stability of UTF1 mutant proteins or differences in the protein's repressive activity. I attempted to quantify Gal4-UTF1, however the fusion proteins not expressed at levels detectable by western blot.

Testing the repressive activity of wild-type and phosphomutant protein UTF1 on the Oct4 promoter construct produced results different than those from the TK promoter. Rather than acting as a repressor, as it does on the UAS-TK promoter, as well as on the TK promoter without the UAS sequence; UTF1 had no effect on Oct4 reporter expression, nor are endogenous Oct4 protein levels altered after transfection with WT and phosphomutant UTF1 proteins. My early experiments resulted in apparent weak activation of the Oct4 reporter by UTF1. After several separate experiments, I finally concluded that this apparent activation is due to repression of the TK-driven *renilla* reporter (similar to the repression displayed in figure 3.8), resulting in an artifactual increase in normalized Oct4 driven luciferase levels.

I propose that UTF1 is working as a general repressor of non-ES specific promoters such as Oct4, Dnmt3b, Rex1 and other highly expressed proteins that I describe in chapter 2. I propose that during some time in development, expression of the Oct4/Sox2 dimer up-regulates UTF1 gene expression, as

reported[56, 63, 215, 218] and UTF1 in turn represses the large swaths of the genome not expressed in the blastocyst, allowing only ES-associated genes and housekeeping genes to be expressed.

This hypothesis is supported by the recent report that use of UTF1 as a reprogramming factor increases the "quality "of iPS clones generated,[219] presumably by repressing genes detrimental to pluripotency (Thomas Zwaka refers to this as "reducing biological noise" in his work on the repressor Ronin[73]). A pet hypothesis of mine, arising from the repression of the TK-*renilla* control plasmid, is that UTF1 and the ES-specific DNA methyltransferases are responsible for the commonly observed silencing of the CMV promoter in ES cells before differentiation. I observe very strong repression of TK-*renilla* in ES cells, presumably due to the high levels of endogenous UTF1 and weaker repression of TK-*renilla* in HepG2 cells, where only CMV-driven UTF1 cDNA is present.

Chapter 3 includes work that is still in preparation for publication and includes contributions from Zhouxin Shen, Kiyoshi Tachikawa, Loes Drenthe, Suzanne Kooistra, Bart Eggan and Steven Briggs. The dissertation author was the primary investigator and author of this material.

Chapter 4: Global analysis of the phosphoproteome of pluripotent mouse stem cells during differentiation identifies increases in phosphorylation of serine-arginine splicing factors

4.1 Summary

Analysis of the phosphoproteome of pluripotent cells is a valuable tool to identify new sites of phosphorylation of proteins already known to be important for pluripotency, however these proteins are only a small fraction of the changes that occur in the phosphoproteome of pluripotent cells as they differentiate. A key question when looking at changes in levels of phospho proteins during differentiation is are these differences due to changes in underlying protein levels, or changes in kinase/phosphatase activity.

To answer this question, I took advantage of the fact that my whole and phosphoproteome analyses were performed on the same material. This gave me information about the underlying protein, allowing me to filter out proteins whose levels change during differentiation, and look only at those phosphoproteins whose underlying protein levels are unchanged.

Using this method, I identified a group of proteins involved in alternative splicing, belonging to the serine/arginine rich (SR) protein family that become phosphorylated after differentiation.

4.2 Introduction

Global changes in phosphorylation during differentiation of ES cells are important to understand because phosphorylation is a key mechanism by which cells signal changes in their environment. Global, unbiased analysis of the phosphoproteome provides a method by which to do this. Several studies have made inventories of phosphoproteins identified in undifferentiated cells or compared undifferentiated cells with cells induced to differentiate. One great limitation of these studies has been the difficulty in deconvoluting changes in abundance of a phosphopeptide from two possible explanations: 1) changes in activity of kinases/phosphatases that act in the phosphopeptide or 2) changes in levels of the underlying protein. Other phosphoproteome analyses have accepted this limitation, simply offering an inventory of phosphopeptides identified in undifferentiated and differentiated cells.

I reasoned that unless I also knew the levels of the underlying protein, I would run into the same limitation in attributing changes in peptide counts to changes in phosphorylation/dephosphorylation by kinases/phosphatases. For this reason, I replicated the experimental design of the whole proteome experimental design as described in chapter 2 for the phosphoproteome analysis of pluripotent stem cells described in chapter 3. In this chapter, I describe my analysis of these two complimentary datasets together in order to find cases of changes in phosphopeptide or phosphoprotein levels due to putative changes in kinase/phosphatase activity rather than changes in whole protein levels.

I accomplished this by looking for proteins whose levels in the whole proteome data were unchanged after differentiation but whose phosphorylation state were significantly changed.

In this chapter I report enrichment of phosphorylation, but not whole protein levels, of mRNA splicing factors after differentiation and test a putative kinase responsible. Inhibition of the candidate kinase, Clk1 with a small molecule (Tg003) during differentiation of ES cells did not inhibit phosphorylation of these SR proteins, indicating that this kinase is not responsible for the increased phosphorylation of these mRNA splicing factors. Phosphoproteome and whole proteome analysis of kinase-inhibited or control cells did, however reveal putative sites of phosphorylation caused by this kinase and changes in protein expression caused by inhibition of the kinase, including an unexpected induction of three hemoglobin subunits in cells grown in the presence of the inhibitor.

4.3 Results

4.3.1 Identification of phosphorylation events enriched before and after differentiation: Comparators and whole proteome data quantitation from iTRAQ allow us to identify changes in phosphoproteome

I used spectral counts to semi-quantify levels of phosphorylation of proteins in ES and EC cells before and after differentiation. Since spectral counting is a semi-quantitative method, I used cutoff of 5-fold enrichment to identify phosphopeptides enriched in pluripotent cells before and after differentiation.

Using these stringent cutoffs, I identified 41 phosphoproteins enriched in EC cells and 73 phosphoproteins enriched in ES cells before differentiation. I identified 105 phosphoproteins enriched in differentiated ES cells and 24 phosphoproteins enriched in differentiated EC cells.

These quantifications of phosphoproteins during differentiation are confounded by the fact that many of the underlying proteins change during differentiation, therefore changes in phosphopeptides do not always reflect changes in signaling, kinase, or phosphatase activity, but rather changes in transcription, translation and stability of the underlying protein.

To identify those proteins that are differently phosphorylated before and after differentiation, I used whole proteome data to identify proteins whose overall levels do not change, but whose phosphorylation levels increase or decrease during differentiation. For these analyses, I chose a cutoff value of +/- 20% (1.25 fold to 0.8 fold) to represent proteins whose levels are unchanged. I also included phosphoproteins enriched before or after differentiation whose underlying protein level was contrary to the phosphoprotein levels. Using these criteria, I identified 3 phosphorylation events enriched in undifferentiated EC cells, 9 in differentiated EC cells (Table 4.1). I found 31 phosphorylation events enriched in undifferentiated ES cells (Table 4.2) and 52 phosphorylation events enriched in differentiated ES cells (Table 4.3).

Table 4.1: Phosphoproteins enriched in EC cells with no change in underlying protein levels. Phosphoproteins enriched in P19 EC cells before (pink) and after (green) differentiation, but with no change in underlying protein levels. Phosphoproteins with >5 spectra and >5 fold difference after normalization were considered enriched, while total proteins with iTRAQ ratios between 0.8 and 1.25 were considered unchanged. Proteins whose total levels were changed, but disagree with phosphorylation data were also included.

Gene Symbol	accession number	num phospho Peps Unique	score Unique	Protein name	Phospho Spectra # Undifferentiated P19 cells	Phospho Spectra # Differentiated P19 cells	Phospho P19 Undiff/Diff normalized	iTRAQ P19 un/diff-average	iTRAQ P19 un/diff-p-value
Hnrpu	IPI00458583	2	37.81	Osteoclast-like cell cDNA, RIKEN full-length enriched library, clone:I420039N16 product:heterogeneous nuclear ribonucleoprotein U, full insert sequence	14	2	5.7882	1.0576	0.1183
Hspd1	IPI00461249	1	18.13	PREDICTED: similar to 60 kDa heat shock protein, mitochondrial precursor (Hsp60) (60 kDa chaperonin) (CPN60) (Heat shock protein 60) (HSP-60) (Mitochondrial matrix protein P1) (HSP-65)	6	1	6.7762	0.8477	0.1076
Snrp70	IPI00625105	1	14.69	Splice Isoform 1 of U1 small nuclear ribonucleoprotein 70 kDa	5	1	6.3738	1.0089	0.9871
Sfrs2	IPI00266871	2	28.25	30 kDa protein	5	54	0.1072	1.0492	0.7128
Mtap4	IPI00473748	5	84.36	Mtap4 protein	2	10	0.1564	1.0994	0.8803
Irf2bp2	IPI00357145	3	44	PREDICTED: similar to interferon regulatory factor 2 binding protein 2 isoform 1	1	7	0.1725	1.0208	0.8482
Sall1	IPI00342267	4	70.55	Sal-like 1	0	7	0.0001	1.0248	0.5922
Akp2	IPI00649302	1	19.23	Tissue-nonspecific alkaline phosphatase	1	6	0.1480	0.8440	0.0534
Ints3	IPI00380394	3	46.24	Expressed sequence C77668	1	5	0.1918	1.1917	0.4910
Bptf	IPI00649138	4	61.16	333 kDa protein	0	5	0.0001	1.0257	0.7015
Ruvbl2	IPI00123557	1	18.71	RuvB-like 2	0	5	0.0001	1.4665	0.0361
Etl4	IPI00222026	3	49.02	Sickle tail-a	1	4	0.1427	0.9430	0.8754

Table 4.2: Phosphoproteins enriched in ES cells after differentiation with no change in underlying protein levels. Phosphoproteins enriched in ES cells before differentiation but with no change in underlying protein levels. Phosphoproteins with >5 spectra and >5 fold difference after normalization were considered enriched, while total proteins with iTRAQ ratios between 0.8 and 1.25 were considered unchanged. Proteins whose total levels were changed, but disagree with phosphorylation data were also included.

Gene Symbol	accession number	num Phospho Peps Unique	score Unique	Protein name	Phospho Spectra # Undifferentiated ES cells	Phospho Spectra # Differentiated ES cells	Phospho ES Undiff/Diff normalized	iTRAQ ES un/diff-average	iTRAQ ES un/diff-p-value
Cd3eap	IPI00169700	2	31.44	Gene_Symbol=Cd3eap DNA-directed RNA polymerase I subunit RPA34	6	0	6001.00	0.91	0.2417
Lgtn	IPI00622438	1	18.26	Lgtn 61 kDa protein	6	0	6001.00	1.05	0.7988
Phf3	IPI00377615	3	49.97	Phf3 PHD finger protein 3	6	0	6001.00	1.08	0.2241
Ubr5	IPI00677913	3	41.47	Ubr5 similar to Ubiquitin-protein ligase EDD1 (Hyperplastic discs protein homolog) isoform 10	5	0	5001.00	0.95	0.3686
Mark2	IPI00554855	3	41.59	Mark2 Isoform 1 of Serine/threonine-protein kinase MARK2	5	0	5001.00	1.08	0.4902
E130014J05Rik	IPI00169923	3	46.22	E130014J05Rik Ayu17-449	5	0	5001.00	1.20	0.0771
Ttk	IPI00122018	3	45.93	Ttk Isoform 2 of Dual specificity protein kinase TTK	5	0	5001.00	0.61	0.0908
Cdk2	IPI00124240	1	18.7	Cdk2 Isoform CDK2-beta of Cell division protein kinase 2	66	4	21.16	1.12	0.3230
2310057J16Rik	IPI00225761	4	60.21	2310057J16Rik Isoform 1 of Uncharacterized protein KIAA1543	11	1	14.09	1.19	0.0765
Cgn	IPI00757790	3	44.07	Cgn cingulin	10	1	12.81	0.91	0.1958
Kti12	IPI00134106	1	15.51	Kti12 Protein KTI12 homolog	10	1	12.81	1.13	0.1856
Esrrb	IPI00752694	2	32.51	Esrrb Esrrb protein	10	1	12.81	0.60	0.0062
Helz	IPI00453654	2	32.54	Helz Helicase with zinc finger domain	9	1	11.53	0.94	0.9804
Lrrfip2	IPI00659860	1	16.41	Lrrfip2 Adult male olfactory brain cDNA, RIKEN full-length enriched library, clone:6430406O17 product:leucine rich repeat (in FLII) interacting protein 2, full insert sequence	8	1	10.25	1.09	0.9414
Rrp9	IPI00128256	2	30.55	Rrp9 U3 small nucleolar RNA-interacting protein 2	71	9	10.12	1.03	0.4300
Tlk1	IPI00273851	2	34.29	Tlk1 Serine/threonine-protein kinase tousled-like 1	7	1	8.97	1.02	0.6562
Pabpn1	IPI00136169	1	18.45	Pabpn1 Isoform 1 of Polyadenylate-binding protein 2	7	1	8.97	1.23	0.0577
2810004N23Rik	IPI00136186	3	45.32	2810004N23Rik Uncharacterized protein C1orf131 homolog	7	1	8.97	0.75	0.6797
Hspd1	IPI00461249	1	19.68	Hspd1 similar to 60 kDa heat shock protein, mitochondrial precursor	24	4	7.69	1.05	0.1103
Khsrp	IPI00462934	2	30.54	Khsrp Activated spleen cDNA, RIKEN full-length enriched library, clone:F830029A18 product:KH-type splicing regulatory protein, full insert sequence	18	3	7.69	0.99	0.7571
Dhx9	IPI00607980	1	21.77	Dhx9 Isoform 2 of ATP-dependent RNA helicase A	22	4	7.05	0.88	0.4697
Nup35	IPI00469331	4	54.79	Nup35 Nucleoporin NUP53	11	2	7.05	1.07	0.5096
Brwd1	IPI00654074	3	43.5	Brwd1 Adult female vagina cDNA, RIKEN full-length enriched library, clone:9930116L03 product:WDR protein, form B homolog	10	2	6.41	0.96	0.8777
Tfdp1	IPI00122992	1	17.98	Tfdp1 Transcription factor Dp-1	10	2	6.41	0.58	0.3797
Pdlim5	IPI00653381	1	19.16	Pdlim5 ENH1	5	1	6.41	1.06	0.8965
1500003O22Rik	IPI00653834	3	50.69	1500003O22Rik NOD-derived CD11c +ve dendritic cells cDNA, RIKEN full-length enriched library, clone:F630021J24 product:similar to Hypothetical protein KIAA0409	5	1	6.41	1.07	0.9799
Zc3h11a	IPI00421162	3	48.59	Zc3h11a Zinc finger CCCH domain-containing protein 11A	5	1	6.41	1.08	0.2799
Mga	IPI00135072	11	157.38	Mga MAX-interacting protein	17	4	5.45	1.13	0.4428
Wdr46	IPI00129141	2	28.48	Wdr46 WD repeat protein 46	8	2	5.13	0.88	0.4724
Usp28	IPI00649320	2	30.83	Usp28 Isoform 2 of Ubiquitin carboxyl-terminal hydrolase 28	4	1	5.13	1.11	0.9719
Nes	IPI00453692	4	49.72	Gene_Symbol=Nes Isoform 1 of Nestin	4	1	5.13	0.65	0.5720
LOC100039215	IPI00848519	1	18.46	LOC100039215 similar to translin associated protein X isoform 1	4	1	5.13	0.74	0.5083

Table 4.3: Phosphoproteins enriched in ES cells before differentiation with no change in underlying protein levels. Phosphoproteins enriched in ES cells after differentiation but with no change in underlying protein levels. Phosphoproteins with >5 spectra and >5 fold difference after normalization were considered enriched, while total proteins with iTRAQ ratios between 0.8 and 1.25 were considered unchanged. Proteins whose total levels were changed, but disagree with phosphorylation data were also included.

Gene Symbol	accession number	num Phospho Peps Unique	score Unique	Protein name	Phospho Spectra # Undifferentiated ES cells	Phospho Spectra # Differentiated ES cells	Phospho ES Undiff/Diff normalized	iTRAQ ES un/diff-average	iTRAQ ES un/diff-p-value
Bin1	IPI00114352	3	48.05	Bin1 Isoform 1 of Myc box-dependent-interacting protein 1	8	52	0.20	0.84	0.7764
Chd7	IPI00749535	4	64.27	Chd7 similar to chromodomain helicase DNA binding protein 7 isoform 1	5	34	0.19	0.81	0.2155
Hnrpd	IPI00330958	2	28.32	Hnrpd Isoform 1 of Heterogeneous nuclear ribonucleoprotein D0	1	7	0.18	0.82	0.2952
Lemd3	IPI00853982	2	31.45	Lemd3 Isoform 2 of Inner nuclear membrane protein Man1	1	7	0.18	0.98	0.8835
Rrp12	IPI00420344	2	30.91	Rrp12 RRP12-like protein	1	7	0.18	1.65	0.0301
Birc6	IPI00134095	3	37.01	Birc6 baculoviral IAP repeat-containing 6	1	8	0.16	0.85	0.0494
Dcp1a	IPI00130114	3	41.81	Dcp1a mRNA-decapping enzyme 1A	1	8	0.16	1.03	0.2807
Ctnna1	IPI00112963	1	20.25	Ctnna1 Catenin alpha-1	9	72	0.16	1.66	0.0594
Igf2bp1	IPI00131056	1	13.66	Gene_Symbol=Igf2bp1 Insulin-like growth factor 2 mRNA-binding protein 1	6	52	0.15	1.12	0.1838
Pdxdc1	IPI00336503	2	29.48	Pdxdc1 2 days pregnant adult female ovary cDNA, RIKEN full-length enriched library, clone:E330016A09 product:hypothetical Pyridoxal-dependent decarboxylase family containing protein, full insert sequence	7	61	0.15	1.04	0.7331
1200011I18Rik	IPI00172219	2	33.38	1200011I18Rik Uncharacterized protein KIAA1704	1	9	0.14	0.81	0.2941
Ints1	IPI00469067	3	42.22	Gene_Symbol=Ints1 Isoform 1 of Integrator complex subunit 1	1	9	0.14	0.91	0.8598
Hn1	IPI00314755	2	34.91	Hn1 Hematological and neurological expressed 1 protein	4	37	0.14	0.94	0.8504
Usp14	IPI00270877	1	16.5	Usp14 Ubiquitin carboxyl-terminal hydrolase 14	3	29	0.13	1.06	0.3758
Rbm25	IPI00421119	3	41.86	Rbm25 12 days embryo eyeball cDNA, RIKEN full-length enriched library, clone:D230046N10 product:S164 homolog	1	10	0.13	0.93	0.6219
Hsp90ab1	IPI00554929	7	99.44	Hsp90ab1 Heat shock protein HSP 90-beta	4	43	0.12	1.10	0.1404
Ncaph	IPI00224053	4	67.75	Ncaph Condensin complex subunit 2	1	11	0.12	0.94	0.5630
Oxsr1	IPI00223738	2	30.69	Oxsr1 Serine/threonine-protein kinase OSR1	1	11	0.12	1.07	0.1897
Eif5b	IPI00756424	5	78.46	Eif5b eukaryotic translation initiation factor 5B	3	35	0.11	0.95	0.1942
Cdc42ep4	IPI00124906	4	53.57	Cdc42ep4 Cdc42 effector protein 4	1	12	0.11	1.17	0.5543
Tra2a	IPI00377298	8	104.24	Tra2a transformer-2 alpha	9	118	0.10	1.76	0.1809
Afap1	IPI00467327	4	73.5	Gene_Symbol=Afap1 actin filament associated protein 1	2	27	0.10	0.96	0.8558
Mcm4	IPI00117016	1	15.58	Mcm4 DNA replication licensing factor MCM4	1	14	0.09	0.89	0.1880
Sfrs2	IPI00621131	2	36.22	Sfrs2 30 kDa protein	5	71	0.09	1.19	0.0393
Mical1	IPI00116371	1	15.87	Mical1 NEDD9-interacting protein with calponin homology and LIM domains	2	36	0.07	0.83	0.1982
Sfrs7	IPI00222763	3	41.94	Sfrs7 Isoform 1 of Splicing factor, arginine/serine-rich 7	2	56	0.05	0.87	0.1461
Saps3	IPI00122858	2	29.57	Saps3 Isoform 1 of SAPS domain family member 3	1	30	0.04	0.99	0.8132
Hsp90aa1	IPI00330804	5	79.96	Hsp90aa1 Heat shock protein HSP 90-alpha	2	93	0.03	1.71	0.0011
Hdgf	IPI00313817	5	80.07	Hdgf Hepatoma-derived growth factor	2	147	0.02	0.92	0.9189
3300001P08Rik	IPI00649422	1	17.09	3300001P08Rik Isoform 1 of Cisplatin resistance-associated overexpressed protein	0	5	0.00	0.90	0.4338
Nudc	IPI00132942	1	18.66	Nudc Nuclear migration protein nudC	0	5	0.00	1.01	0.5497
Paf1	IPI00331654	1	16.99	Paf1 Paf1, RNA polymerase II associated factor, homolog	0	5	0.00	1.01	0.5072
Nasp	IPI00830976	4	54.33	Nasp nuclear autoantigenic sperm protein isoform 2	0	5	0.00	1.16	0.0096
1810007M14Rik	IPI00457896	2	28.06	1810007M14Rik similar to GC-rich sequence DNA-binding factor homolog isoform 4	0	5	0.00	1.17	0.8663
Ppp1r12c	IPI00669561	3	41.71	Ppp1r12c Lung RCB-0558 LLC cDNA, RIKEN full-length enriched library, clone:G730023J22 product:protein phosphatase 1, regulatory (inhibitor) subunit 12C, full insert sequence	0	5	0.00	2.06	0.0121
Cebpz	IPI00752710	2	38.72	Cebpz Isoform CBF1 of CCAAT/enhancer-binding protein zeta	0	6	0.00	1.57	0.2569
Prpf40a	IPI00284213	2	38.37	Prpf40a Isoform 1 of Pre-mRNA-processing factor 40 homolog A	0	7	0.00	1.09	0.4500
D11Wsu99e	IPI00762255	2	26.65	D11Wsu99e hypothetical protein LOC28081	0	8	0.00	1.14	0.8602
Nisch	IPI00110435	3	46.17	Nisch nischarin	0	9	0.00	3.59	0.4612
Syne2	IPI00753411	7	118.16	Gene_Symbol=Syne2 similar to spectrin repeat containing, nuclear envelope 2 isoform a	0	12	0.00	0.85	0.3988
Myh9	IPI00788324	4	53.06	Myh9 15 days pregnant adult female amnion cDNA, RIKEN full-length enriched library, clone:M421002E03 product:myosin heavy chain IX, full insert sequence	0	12	0.00	0.92	0.9023
Utp18	IPI00353579	2	30.46	Utp18 U3 small nucleolar RNA-associated protein 18 homolog	0	12	0.00	1.05	0.5147
Ccdc16	IPI00320645	3	48.02	Ccdc16 Coiled-coil domain-containing protein 16	0	13	0.00	1.08	0.1974
Ccnk	IPI00654283	1	13.14	Ccnk NOD-derived CD11c +ve dendritic cells cDNA, RIKEN full-length enriched library, clone:F630103D11 product:cyclin K, full insert sequence	0	15	0.00	1.07	0.3764
Gm98	IPI00755018	3	54.69	Gm98 similar to CG3328-PA	0	15	0.00	6.29	0.0831
Slc9a3r1	IPI00109311	2	34.01	Slc9a3r1 Ezrin-radixin-moesin-binding phosphoprotein 50	0	17	0.00	1.01	0.3747
Thoc5	IPI00222687	1	19.92	Thoc5 9 days embryo whole body cDNA, RIKEN full-length enriched library, clone:D030012I15 product:ANONYMOUS (GENE FROM NF2/MENINGIOMA REGION OF 22Q12) homolog	0	18	0.00	0.82	0.1933
Leo1	IPI00474486	12	159.9	Leo1 Isoform 1 of RNA polymerase-associated protein LEO1	0	20	0.00	0.97	0.3068
Eif4g3	IPI00673296	3	36.67	Eif4g3 (eIF-4-gamma 3) (eIF-4G 3) (eIF4G 3) (eIF-4-gamma II) (eIF4GII) isoform 17	0	21	0.00	0.83	0.2188
Rnps1	IPI00462424	1	14.86	Rnps1 Isoform 1 of RNA-binding protein with serine-rich domain 1	0	23	0.00	1.08	0.6062
Zranb2	IPI00756485	4	59.91	Zranb2 37 kDa protein	0	24	0.00	1.00	0.1345
Bud13	IPI00153284	8	104.82	Bud13 Isoform 1 of BUD13 homolog	0	37	0.00	1.24	0.6556
Sf3a1	IPI00408796	1	18.22	Sf3a1 Splicing factor 3 subunit 1	0	104	0.00	1.02	0.3952

4.3.2 Phosphorylation of several spliceosome components that regulate exon-skipping is enriched in differentiated ES cells

To better understand the biological processes that these regulated proteins shared in common, I used the DAVID annotation database to assign GO biological process annotations to the proteins. I used standard settings with the exception of increasing the threshold count of proteins per term from 2 to 5. The top 5 GO_BP_3 categories of differentiated ES cells are presented in figure 4.1. The results show a striking increase in phosphorylation of mRNA splicing factors during differentiation, there was no similar striking enrichment of GO terms among the phosphoproteins enriched before differentiation of ES cells (Figure 4.2).

Further analysis revealed that these phosphorylated splicing factors were mostly of the Serine-Arginine rich (SR) family of proteins. SR proteins are regulators of splicing in metazoans and are known to influence alternative splicing by inhibiting exon skipping[220].

Figure 4.3 demonstrates iTRAQ ratios of underlying proteins from ES cells before and after differentiation, when proteins were identified in all three runs, error bars are included, in all cases these proteins' levels are not enriched after differentiation. In contrast, spectral counts of phosphorylated peptides are >5 fold higher after differentiation, and in the case of Sf3a1, the ratio is 0 to104. For this reason, I conclude that these SR proteins are being differentially phosphorylated during ES cell differentiation.

Phosphorylation of SR proteins by the kinase CLK1 leads to an inhibition SR protein activity and thus, and increase in alternative splicing[221-230]. Studies on Clk1 and SR proteins' roles in splicing have occurred in cancerous cell types, however recently published work has identified SR proteins responsible for alternative splicing of several transcripts during neural development in P19 cells in vitro[231] and ES/neural progenitor cells in vivo and in vitro[231, 232].

I hypothesized that the observed increase of phosphorylation of SR proteins during differentiation is a necessary aspect of development, limiting alternative splicing of target transcripts before differentiation and increasing alternative splicing as cells differentiate into the neural lineage.

**Top 5 GO_BP_3 Terms enriched in proteins with increased
phosphorylation after ES and EC cell differentiation**
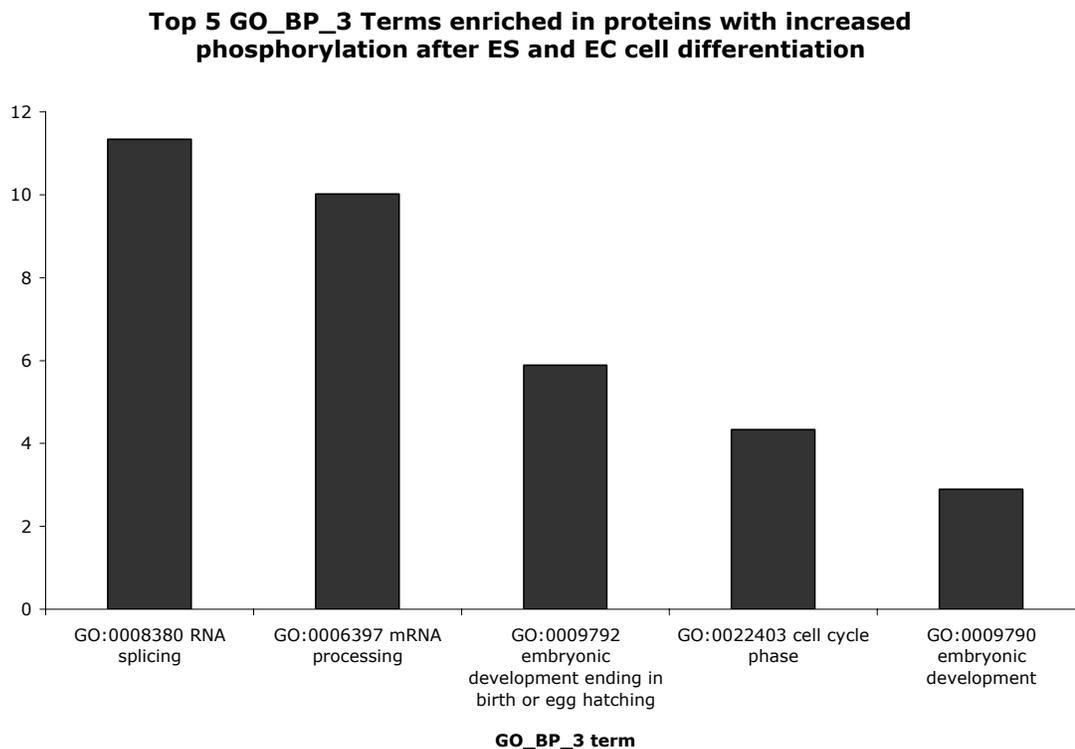
Figure 4.1: GO terms associated with phosphoproteins enriched after
differentiation. Top 5 Gene ontologies of phosphoproteins enriched after
differentiation whose underlying protein level are unchanged. The Y-axis refers to
the enrichment of gene symbols falling into a particular ontology over the amount
expected from a random list. Phosphoproteins involved in splicing and
processing (the categories are largely redundant) are >10 fold enriched over
what would be expected from a random list.

Figure 4.2: GO terms associated with phosphoproteins enriched before differentiation. Top 5 Gene ontologies of phosphoproteins enriched before differentiation whose underlying protein level are unchanged. The Y-axis refers to the enrichment of gene symbols falling into a particular ontology over the amount expected from a random list. No terms are more than 2 fold enriched.

iTRAQ ratio of Splicing factors measured in Undifferentiated / Differentiated 129 SVEV ES cells
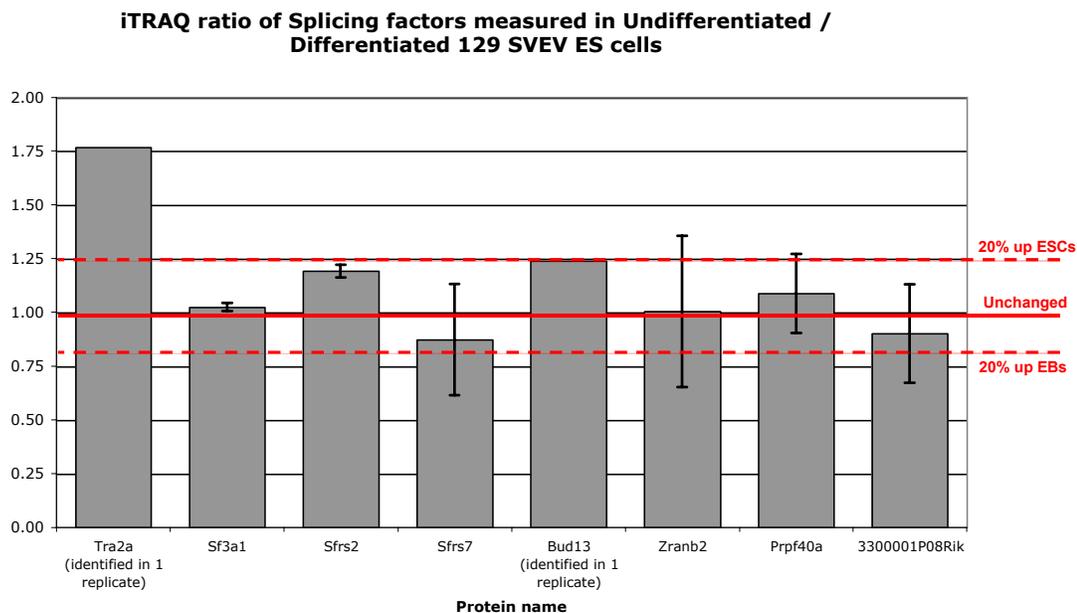
Figure 4.3: iTRAQ ratios of SR proteins before and after differentiation Histogram representing iTRAQ ratios of genes that fall into the GO_BP_3 category "RNA splicing" in figure 4.1. All eight proteins' levels are unchanged or decreased during differentiation of ES cells. Error bars represent standard deviation of three iTRAQ ratios measured in 3 whole proteome runs. Proteins with no error bars were only identified/quantified in one run.

**Spectral counts of spliceosome component phosphopeptides
identified before and after differentiation of ES cells**
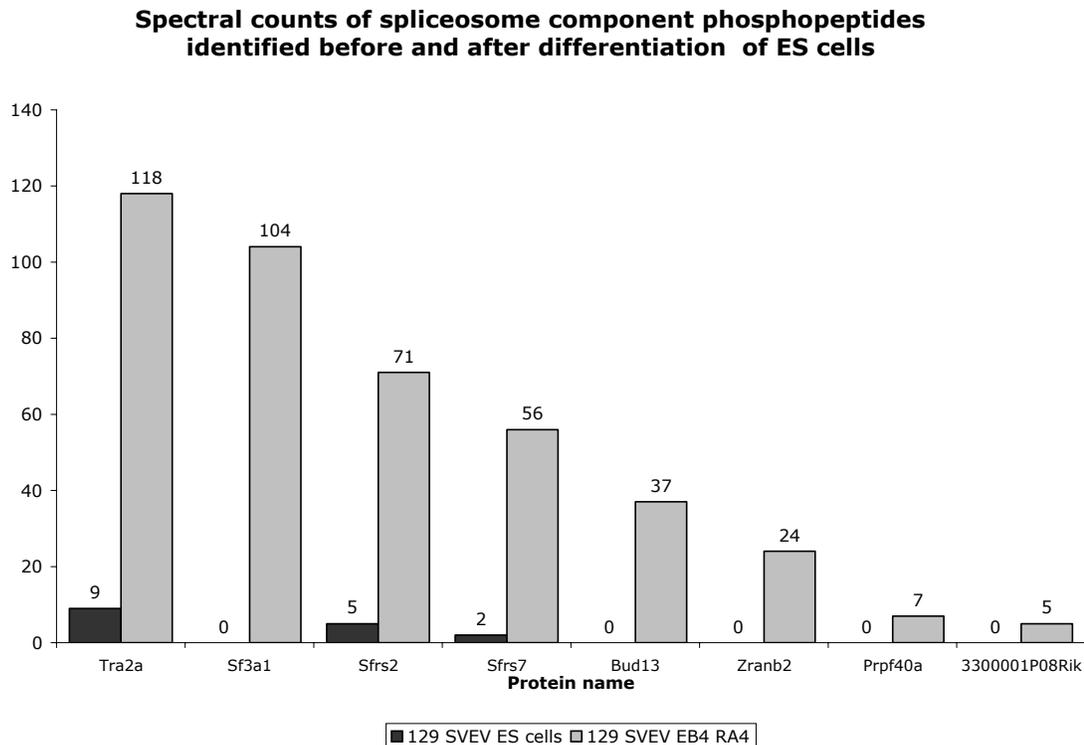
Figure 4.4: Spectral counts of SR peptides before and after differentiation.
Phosphorylation of SR splicing factors is highly enriched after differentiation of
ES cells. Histogram represents spectral counts of phosphopeptides from SR
proteins identified before (dark bars) and after (light bars) ATRA mediated
differentiation of ES cells. All eight proteins, though unchanged at the whole
protein level (figure 4.3), are more than 5 fold enriched in differentiated ES cells.
This strongly suggests that these eight SR proteins are being differentially
phosphorylated during differentiation of ESCs.

4.3.3 CLK1 is a putative kinase of SR-proteins

Much is known about regulation of SR-proteins and the role that phosphorylation of these factors plays in splicing. SR proteins are implicated in both constitutive and alternative splicing in insect cells, Cos and 293T cells. The kinase Clk1 phosphorylates SR proteins and leads to inactivation of SR-proteins and exon exclusion. Clk1 has not been implicated in neural differentiation, and knockout studies have failed to reveal any phenotype associated with neural differentiation. SR proteins have, on the other hand been implicated in neural differentiation, including the SR protein nSR100 (Srrm4), a protein induced during neurogenesis and necessary for alternative splicing of >100 transcripts and for neural development in vitro (ES cells) and in vivo (zebrafish). Based on these data I asked if Clk1 was the kinase responsible for phosphorylating the 8 SR proteins enriched after differentiation.

4.3.4 Inhibition of Clk1 changes the morphology of embryoid bodies only after addition of retinoic acid

I used the Clk family-specific kinase inhibitor Tg003, a chemical with an IC_{50} of 20 nM for Clk1, 200 nM for Clk2 and 15 nM for Clk4. Tg003 has negligible effects on Pkc, Pka and Clk3, with an IC_{50} of >10μM for all three and no effect on cell survival or growth rate. I pretreated ES cells with either 1μM Tg003 (stock solution 1mM in DMSO) or DMSO alone for 24 hours, and then differentiated the ES cells according to the same protocol I used previously. Cells were aggregated by trypsinization and replating on bacteriological plates in ES media lacking LIF. Cells remained in the presence of Tg003 (1μM, 0.1% DMSO) or DMSO (0.1%)

throughout the experiment. Media was refreshed at 2, 4 and 6 days, with 1 μM ATRA added to the media at 4 days post aggregation.

Undifferentiated cells were morphologically identical between drug treated and control cells as were EBs until the addition of retinoic acid on day 4. 24 hours after addition of ATRA, EBs treated with Tg003 attached to the surface of the bacteriological plates and began to form outgrowths (Figure 4.5), while control EBs remained free-floating and non-adherent. This remained the case until days 7 and 8, when control EBs also began to adhere to the bacteriological plates and form outgrowths (Figure 4.5).

I collected cells on days 0 and 8 to perform whole and phosphoproteome analyses. The first analysis was comparison of undifferentiated ES cells and 8 day control EBs. As expected, markers of pluripotency declined after differentiation, while markers of the RA- differentiated state increased (Table 4.4). The next analysis was a comparison between 8 day EBs differentiated in the presence or absence of Tg003 in order to identify differences in the whole and phosphoproteomes of pluripotent cells grown in the presence or absence of Tg003.

EBs differentiated in the presence of Tg003 adopt a flattened, adherent morphology 24 hours after addition of ATRA

Mock (DMSO)

Day 0 | Day 2 | Day 4 | Day 5 | Day 8

1 µM Tg003

Day 0 | Day 2 | Day 4 | Day 5 | Day 8
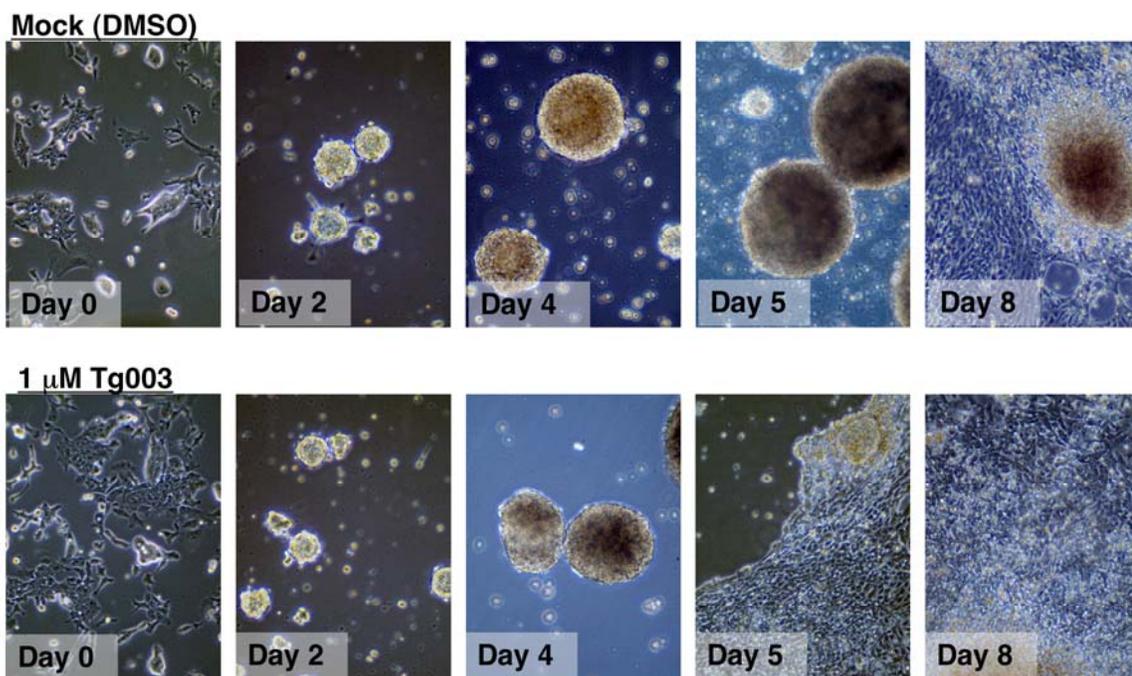
Figure 4.5: Differentiation of ES cells in the presence of Tg003. Differentiation of ES cells in the presence or absence of Tg003. ES cells were aggregated and grown in bacteriological plates in the absence of LIF. Media was refreshed at 2, 4 and 6 days with ATRA added at 4 days. EBs grown in the presence of Tg003 were identical in morphology to control cells until 24 hours after addition of ATRA, when they began to adhere to the plate and produce outgrowths.

Table 4.4: ES cells in the Tg003 experiment lose pluripotency markers as they differentiate. Proteins markers of pluripotency are enriched in ES cells differentiation, while markers of neural differentiation and ATRA response are enriched in differentiated cells.

total_114	Normalized_115	Normalized_116	Normalized_117	Group	Score	Unique_Peptide	Spectrum_num	Gene Symbol	Accession	Protein	ES/EB	p-value	Sum of iTRAQ intensities
9009.19	8544.09	1202.81	1154.15	1226	80.86	5	7	Pou5f1	IPI00117218	class 5, transcription factor 1	7.447	0.0009	19910.2354
631.92	675.35	90.50	66.76	2049	39.33	2	3	Cth	IPI00122344	lyase	8.312	0.0018	1464.5376
1359.40	1119.45	449.99	33.46	2110	37.58	2	5	Cdh1	IPI00318626	Cdh1 Epithelial-cadherin precursor	5.127	0.0534	2962.2957
110.90	95.41	23.21	23.26	574	169.18	10	17	Dnmt3b	IPI00268106	Dnmt3b Isoform 2 of DNA (cytosine-5)-methyltransferase 3B	4.440	0.0093	252.7768
825.82	668.61	178.00	177.90	2361	32.54	2	2	Dppa4	IPI00754115	Dppa4 ES cells cDNA, RIKEN full-length enriched library, clone:2410091M23 product:hypothetical protein, full insert	4.199	0.0185	1850.3334
68971.52	64570.27	18577.55	16252.46	263	263.85	14	73	LOC100043906	IPI00124973	OC100043906 Importin subunit alpha-2	3.834	0.0025	168371.7981
10250.30	9658.14	2791.51	2911.45	188	312.51	17	32	Rif1	IPI00626173	Telomere-associated protein RIF1	3.491	0.0018	25611.4114
25165.22	23149.24	5563.20	4423.96	816	126.45	7	14	Utf1	IPI00454162	Undifferentiated embryonic cell transcription factor 1	4.838	0.0036	58301.6203
4.04	17.98	205.66	202.06	2783	19.87	1	2	Gata4	IPI00119514	Gata4 Transcription factor GATA-4	0.054	0.0014	429.7365
391.47	306.44	7718.72	7740.82	1898	47.45	3	3	Hoxa5	IPI00132371	Hoxa5 Homeobox protein Hox-A5	0.045	0.0000	16157.4361
826.30	792.60	3072.65	3876.21	2199	35.75	2	2	Ncam1	IPI00230665	140 of Neural cell adhesion molecule 1, 180 kDa isoform precursor	0.233	0.0220	8567.7559
458.34	590.77	1763.05	1993.41	1727	53.59	3	4	Akt1	IPI00323969	serine/threonine-protein kinase	0.279	0.0095	4805.5833
5077.43	6975.72	20088.49	20514.15	1231	80.12	4	6	Crabp1	IPI00230721	Crabp1 Cellular retinoic acid-binding protein 1	0.297	0.0046	52655.7881
23.91	58.68	75.33	68.40	1972	43.16	2	5	Cdh2	IPI00323134	Cdh2 Cadherin-2 precursor	0.575	0.2267	226.3117
47.82	76.33	153.23	125.21	3301	15.63	1	1	Ppp1r9a	IPI00336313	specific F-actin binding protein	0.446	0.0610	402.5801

4.3.5 Phosphorylation of SR proteins during differentiation is not mediated by Clk1

To test the role of Clk1 mediated phosphorylation of SR-proteins during differentiation of ES cells, I used the Clk1/4 specific inhibitor Tg003[233] to block activation of Clk1 and closely related kinases Clk4 and Clk2 during differentiation of ES cells. I used mass spectrometry and iTRAQ mass tag labeling to analyze the whole proteome of 129/SvEv embryoid bodies differentiated in the presence of Tg003 or carrier (DMSO) to monitor SR protein phosphorylation and changes in the whole and phosphoproteomes of differentiated ES cells (EB day 8, RA day 4) in the presence and absence of the Clk1/4 inhibitor Tg003[233].

I hypothesized that inhibition of Clk1 would lead to a reduction of the phosphorylation of the phosphorylated SR-proteins enriched in ES cells after differentiation. This was not the case, phosphoproteome analysis identified 6 of the 8 phosphorylated SR proteins and I confirmed observations made in previous datasets, that the candidate SR proteins are indeed enriched after differentiation of ES cells (Table 4.7), but the levels of phosphorylation were not changed between the control and Tg003 treated samples (Figure 4.8). Interestingly, levels of phosphopeptides from two other SR proteins, Srrm1 and Srrm2 were reduced by >50% in Tg003 treated cells, suggesting that these sites (Srrm1 S751/S753, Srrm2 T727/S733 and Srrm2 T877) are targets of Clk1, Cl2 or Clk4 (Figure 4.9).

4.3.6 Inhibition of Clk family kinases during differentiation of ES cells leads to very few differences in whole and phosphoproteomes of Tg003-treated and control cells

Comparison of the whole and phosphoproteomes of DMSO and Tg003 treated cells revealed extensive similarities between the two samples, especially at the whole proteome level (Table 4.7). More than 80% of the proteins measured in the whole proteome analysis differ by <20% between EBs treated with Tg003 or control cells treated with DMSO, while only 0.77% of proteins identified differ by >50% and have iTRAQ reporter intensities of >100. Therefore, there seems to be little difference between EBs with inhibited Clk1 and EBs with active Clk1 (Table 4.6). Of the 26 proteins that are different between Tg003 treated and control EBs, the most striking change was an enrichment of three hemoglobin subunits (β, ζ, ϵ) in the Tg003 treated cells p<0.05, iTRAQ intensities all >50,000. The DMSO treated cells had the same number of proteins enriched in them compared with the Tg003 treated (13 proteins each), but there was no similar, striking class of proteins enriched.

The phosphoproteome of EBs treated with Tg003 differs from the phosphoproteome of EBs treated with DMSO slightly more than the whole proteome. In the case of the phosphoproteomes, 50% of the proteins identified change by <20%, while 5% are changed by more than 50% (Tables 4.9 and 4.10). For this reason, I conclude that changes to the cells when grown in the presence of a Clk1 inhibitor occur mainly at the level of the phosphoproteome.

I hypothesize that the phosphoproteins with >50% differences between the two datasets represent the "shadow cast" by inhibition of Clk family kinases by Tg003.

Table 4.5: Summary of data from Clk1 inhibition experiments. Summary of Tg003 treatment whole and phosphoproteome datasets.

	Total spectra	Total peptides	Total proteins	FDR	Proteins <20% changed	Proteins >50% changed ITRAQ >100
Phospho		1679		0.48% (8/1680)	854 (50.86%)	85 (5.06%)
Whole	71629	17120	3343	0.48% (16/3343)	2727 (81.57%)	26 (0.77%)

Table 4.6: iTRAQ quantitation shows no effect on phosphorylation of candidate SR proteins by Tg003 treatment.

total 114	Normalized 115	Normalized 116	Normalized 117	Peptide	Gene Symbol	Accession	Protein	DMSO / Tg003	p value	sum of iTRAQ intensitied
					Tra2a					
152	95	73	97	FGESEEVEMEVEsDEEDQEK	Sf3a1	IPI00408796	Sf3a1 Splicing factor 3 subunit 1	1.448	0.3424	268
2447	1541	1881	2148	sPPPVSK	Sfrs2	IPI00474430	Sfrs2 0 day neonate head cDNA, RIKEN full-length enriched library, clone:4833401I21 product:splicing factor, arginine/serine-rich 2 (SC-35), full insert sequence	0.990	0.9695	5571
2271	1705	2072	1775	sRsPPPVSK	Sfrs2	IPI00474430	Sfrs2 0 day neonate head cDNA, RIKEN full-length enriched library, clone:4833401I21 product:splicing factor, arginine/serine-rich 2 (SC-35), full insert sequence	1.033	0.8598	5555
733	629	691	354	SRsPPPVSK	Sfrs2	IPI00474430	Sfrs2 0 day neonate head cDNA, RIKEN full-length enriched library, clone:4833401I21 product:splicing factor, arginine/serine-rich 2 (SC-35), full insert sequence	1.303	0.4638	1677
1945	1953	1431	2114	sRsASLR	Sfrs7	IPI00474169	Sfrs7 Isoform 4 of Splicing factor, arginine/serine-rich 7	1.100	0.6566	5500
460	251	306	352	YEHDSDLsPPR	Bud13	IPI00153284	Bud13 Isoform 1 of BUD13 homolog	1.080	0.8284	911
1084	1428	864	966	YNLDAsEEEDSNK	Zranb2	IPI00845647	Zranb2 Isoform 2 of Zinc finger Ran-binding domain-containing protein 2	1.373	0.1974	3260
6389	7057	9330	7074	EEsDGEYDEFGR	Zranb2	IPI00845647	Zranb2 Isoform 2 of Zinc finger Ran-binding domain-containing protein 2	0.820	0.3356	23463
1701	1455	1676	2048	QETVSDFtPK	Prpf40a	IPI00284213	Prpf40a Isoform 1 of Pre-mRNA-processing factor 40 homolog A	0.848	0.3312	5180
					3300001P08Rik					

Table 4.7: iTRAQ quantitation of candidate SR protein phosphopeptides confirms that phosphorylation of some is increased during differentiation

total_114	Normalized_115	Normalized_116	Normalized_117	Peptide	Gene Symbol	Accession	Protein	Undifferentiated / Differentiated	p-value	Sum of iTRAQ
					Tra2a					
15	11	38	10	FGESEEVEMEVEsDEEDQEK	Sf3a1	IPI00408796	Sf3a1 Splicing factor 3 subunit 1	0.56	0.53202	74
730	702	742	750	sPPPVSK	Sfrs2	IPI00121135	Sfrs2 Splicing factor, arginine/serine-rich 2	0.96	0.18064	2925
14972	12717	18178	17418	sRsPPPVSK	Sfrs2	IPI00121135	Sfrs2 Splicing factor, arginine/serine-rich 2	0.78	0.07988	63284
377	292	489	348	sRSPPPVSK	Sfrs2	IPI00121135	Sfrs2 Splicing factor, arginine/serine-rich 2	0.80	0.41456	1506
15	14	34	44	sPPKSPEEEGAVSS	Sfrs2	IPI00121135	Sfrs2 Splicing factor, arginine/serine-rich 2	0.37	0.04234	107
11	7	95	88	SPSGsPHR	Sfrs7	IPI00474169	Sfrs7 Isoform 4 of Splicing factor, arginine/serine-rich 7	0.10	0.00186	201
26	39	134	113	SPsGSPHR	Sfrs7	IPI00474169	Sfrs7 Isoform 4 of Splicing factor, arginine/serine-rich 7	0.26	0.01780	313
204	199	158	182	YEHDSDLsPPR	Bud13	IPI00153284	Bud13 Isoform 1 of BUD13 homolog	1.18	0.12272	743
1310	1542	1865	2896	EEsDGEYDEFGR	Zranb2	IPI00756485	Zranb2 37 kDa protein	0.60	0.21240	7613
882	777	1006	998	YNLDAsEEEDSNK	Zranb2	IPI00756485	Zranb2 37 kDa protein	0.83	0.08277	3663
24	27	16	41	EsEGEEEDEDEDLSK	Zranb2	IPI00756485	Zranb2 37 kDa protein	0.88	0.81130	108
4	4	5	3	EVEDKEsEGEEEDEDEDLSK	Zranb2	IPI00756485	Zranb2 37 kDa protein	0.96	0.87104	16
34	37	45	24	sPESQVIGENIK	Zranb2	IPI00756485	Zranb2 37 kDa protein	1.02	0.95952	140
36	24	41	19	QETVSDFtPK	Prpf40a	IPI00284213	Prpf40a Isoform 1 of Pre-mRNA-processing factor 40 homolog A	1.00	0.99893	121
					3300001P08Rik					

Table 4.8: Proteins whose levels are enriched in DMSO or Tg003 treated embryoid bodies. Proteins whose levels are different between DMSO and Tg003 treated 8 day EBs. As with previous iTRAQ analyses P-values are reported but fold change (50%) and total iTRAQ intensities (>100) are were used as cutoff parameters.

total_114	Normalized_115	Normalized_116	Normalized_117	Group	Score	Unique_Peptide	Spectrum_num	Gene Symbol	Accession	Protein	DMSO / Tg003	p-value	Sum of iTRAQ intensities
70.66	118.23	28.83	11.84	3342	14.48	1	1	Gucy2e	IPI00130729	Gucy2e Guanylyl cyclase GC-E precursor	4.644	0.0992	229.5609
110.58	113.69	55.92	61.80	2857	18.6	1	1	Wipi2	IPI00131321	Wipi2 WD repeat domain phosphoinositide-interacting protein 2	1.905	0.0039	341.9883
244.74	184.16	127.41	109.78	1679	52.82	3	6	Smtnl2	IPI00229702	Smtnl2 RIKEN cDNA D130058I21 gene	1.808	0.0934	666.0941
492.68	565.57	322.07	265.69	2272	33.06	2	2	Pex14	IPI00720176	Pex14 similar to Peroxisomal membrane protein PEX14	1.800	0.0363	1646.0163
202.13	211.12	115.54	114.52	1918	40.96	2	3	Ddx10	IPI00336329	Ddx10 DEAD (Asp-Glu-Ala-Asp) box polypeptide 10	1.796	0.0024	643.3160
564.27	490.24	364.67	239.19	3306	14.87	1	1	Adfp	IPI00125653	Adfp Adipophilin	1.746	0.0905	1658.3725
43.25	40.92	24.24	24.41	2545	22.16	1	2	Gyg	IPI00264062	Gyg Glycogenin-1	1.730	0.0043	132.8217
186.75	175.86	95.32	116.65	2499	23.5	1	3	Mttp	IPI00309073	Mttp Microsomal triglyceride transfer protein large subunit precursor	1.711	0.0244	574.5816
67.69	68.97	48.93	32.10	1741	50.26	3	3	LOC100046049	IPI00316343	Coil,LOC100046049,ENSMUSG00000072564,2210409E12Rik coilin	1.687	0.0810	217.6777
689.05	594.64	443.75	346.68	3355	14.4	1	1	Clybl	IPI00153903	Clybl Citrate lyase beta subunit-like protein, mitochondrial precursor	1.624	0.0678	2074.1268
62.16	62.92	44.79	33.42	3067	16.8	1	1	Adar	IPI00395223	Adar Isoform 2 of Double-stranded RNA-specific adenosine deaminase	1.599	0.0543	203.2849
273.49	246.66	184.57	143.45	293	239.1	13	38	Rpl5	IPI00762217	Rpl5 Blastocyst blastocyst cDNA, RIKEN full-length enriched library, clone:I1C0019E04 product:ribosomal protein L5, full insert sequence	1.586	0.0595	848.1748
35.66	43.48	27.45	24.25	2210	34.23	2	2	Map2k4	IPI00649311	Map2k4 Mitogen activated protein kinase kinase 4, variant 1	1.531	0.0831	130.8455
12463.16	12502.80	19153.31	18347.65	1360	70.24	4	9	LOC100044141	IPI00555131	LOC100044141,Hbb-y Hemoglobin subunit epsilon-Y2	0.666	0.0041	62466.9130
507.13	394.25	688.74	692.76	2432	28.89	2	2	Cpox	IPI00400301	Cpox Coproporphyrinogen III oxidase, mitochondrial precursor	0.652	0.0511	2282.8794
273.24	358.30	486.35	499.79	1885	43.02	2	2	Polr1e	IPI00120981	Polr1e Isoform 2 of DNA-directed RNA polymerase I subunit RPA49	0.640	0.0542	1617.6817
98.58	121.16	184.16	162.55	2357	31.29	2	2	Tbl2	IPI00127026	Tbl2 Transducin beta-like 2 protein	0.634	0.0556	566.4448
300.21	255.63	491.19	406.05	1996	38.36	2	2	LOC100045283	IPI00469227	Gpc6,LOC100045283 glypican 6 isoform 1	0.619	0.0709	1453.0830
22.98	32.79	47.25	47.82	1862	44.4	2	2	Arfgap1	IPI00403723	Arfgap1 Isoform 2 of ADP-ribosylation factor GTPase-activating protein 1	0.587	0.0572	150.8393
98.69	116.08	171.57	202.18	2543	22.25	1	1	Dhcr24	IPI00453867	Dhcr24 24-dehydrocholesterol reductase precursor	0.575	0.0457	588.5137
32.71	30.48	66.08	54.71	1866	44.09	2	2	Dnajc10	IPI00655098	Dnajc10 Adult male testis cDNA, RIKEN full-length enriched library, clone:4930571C17 product:DKFZp434J1813 protein, full insert sequence (Fragment)	0.523	0.0382	183.9884
4985.00	5954.72	10023.75	11262.12	1255	75.49	4	10	Hbb	IPI00407504	Hbb-bh1,Hbb Hemoglobin subunit beta-H1	0.514	0.0223	32225.6021
27912.05	32027.98	62605.13	61643.18	992	98.51	5	16	Hba-x	IPI00230261	Hba-x Hemoglobin subunit zeta	0.482	0.0043	184188.3473
61.32	51.80	120.91	124.46	3247	15.48	1	1	BC030440	IPI00480293	BC030440 Uncharacterized membrane protein KIAA0286	0.461	0.0058	358.4974
65.47	108.86	228.26	170.46	3066	16.8	1	2	Vta1	IPI00133024	Vta1 Vacuolar protein sorting-associated protein VTA1 homolog	0.437	0.0900	573.0504
17.91	18.14	74.49	59.89	3274	15.23	1	1	Klc1	IPI00845696	Klc1 kinesin light chain 1 isoform 1H	0.268	0.0213	170.4268

Table 4.9: Phosphopeptides enriched in DMSO treated samples, reduced in Tg003 samples. Phosphopeptides enriched in DMSO treated EBs, reduced in Tg003 treated 8 day EBs. Lowercase letters in peptide sequence refer to phosphorylated residue. As with previous iTRAQ analyses P-values are reported but fold change (50%) and total iTRAQ intensities (>100) are were used as cutoff parameters.

total_114	Normalized_1 15	Normalized_1 16	Normalized_1 17	Peptide	Group	Gene Symbol	Accession	Protein	DMSO / Tg003	p-value	Total iTRAQ
106.04	122.96	0.00	34.64	GSEGSQsPGSSVDAEDDPSR	105.00	Zfpm1	IPI00132648	Zfpm1 Zinc finger protein ZFPM1	6.61	0.04	263.64
114.53	178.07	28.44	19.53	EMLLEDVGsEEEPEEDDEAPFQEK	55.00	Nucks1	IPI00341869	Nucks1 Nuclear ubiquitous casein and cyclin-dependent kinases substrate	6.10	0.06	340.56
52.73	50.05	3.80	15.33	QSSGEQsPDGGLSDSSDGQGERPLNLR	725.00	Nab1	IPI00319856	Nab1 Ngfi-A binding protein 1	5.37	0.02	121.90
44.02	54.71	9.17	11.72	STDsPIAIEPLSESDYR	247.00	Suhw2	IPI00228721	Suhw2 Activated spleen cDNA, RIKEN full-length enriched library, clone:F830217J08 product:hypothetical Zn-finger, C2H2 type containing protein, full insert sequence	4.73	0.02	119.62
55.74	60.13	17.08	23.56	QAQSStEIPLQAESGQGTEEEAAK	7.00	Akap12	IPI00123709	Akap12 SSeCKS	2.85	0.01	156.50
362.63	252.93	106.72	120.12	DWEDDsDEDMSNFDR	756.00	IPI00762817	IPI00762817	19 kDa protein	2.71	0.07	842.40
377.75	349.29	142.38	146.06	GtPDTGAASEEK	12.00	Eif5b	IPI00756424	Eif5b eukaryotic translation initiation factor 5B	2.52	0.00	1015.47
67.58	60.35	19.86	34.54	KLPPPPPQAPPEEENEsEPEEPSGVEGAAFQSR	298.00	Aof2	IPI00453837	Aof2 Lysine-specific histone demethylase 1	2.35	0.05	182.33
162.38	204.92	62.86	96.76	SGtPPRPGsVTNMQADECTATPQR	1.00	Srrm2	IPI00785384	Srrm2 Isoform 2 of Serine/arginine repetitive matrix protein 2	2.30	0.06	526.92
824.48	891.89	499.41	247.93	GHPSAGAEEEGGsDGsAAEAEPR	8.00	Eif3s9	IPI00761592	Eif3s9 109 kDa protein	2.30	0.07	2463.70
99.32	85.61	37.53	44.76	EGDIIPPLTGAtPPLIGHLK	3.00	Trp53bp1	IPI00229801	Trp53bp1 Transformation related protein 53 binding protein 1	2.25	0.02	267.22
81.57	60.16	32.03	31.72	SEQQAEALDsPQK	539.00	Lrrfip1	IPI00654388	Lrrfip1 Isoform 1 of Leucine-rich repeat flightless-interacting protein 1	2.22	0.07	205.48
1038.52	850.31	465.79	467.62	LQPLTSVDsDNDFVTPKPR	678.00	Ncapd2	IPI00275153	Ncapd2 Isoform 1 of Condensin complex subunit 1	2.02	0.04	2822.24
608.27	638.03	327.38	288.96	QGLAETSsPVAIsLR	4.00	Pcm1	IPI00127764	Pcm1 Isoform 1 of Pericentriolar material 1 protein	2.02	0.01	1862.64
413.32	396.48	201.28	212.05	LTSVLsPR	788.00	Arhgef17	IPI00465761	Arhgef17 Isoform 1 of Rho guanine nucleotide exchange factor 17	1.96	0.00	1223.12
286.00	311.25	105.32	200.20	AVSsPPTsPRPGSAATISSSASNIVPPR	31.00	Srp72	IPI00659258	Srp72 8 days embryo whole body cDNA, RIKEN full-length enriched library, clone:5730576P14 product:weakly similar to SIGNAL RECOGNITION PARTICLE 72 kDa PROTEIN (Fragment)	1.95	0.10	902.77
305.90	287.59	137.77	167.48	ISHSLYSGIEGLDEsPTR	318.00	Pard3	IPI00309259	Pard3 Par-3 (Partitioning defective 3) homolog	1.94	0.01	898.74
62.29	56.15	27.83	34.15	LASGDGDEEQDEEtEDEETEDHLGK	683.00	Ddx10	IPI00336329	Ddx10 DEAD (Asp-Glu-Ala-Asp) box polypeptide 10	1.91	0.02	180.42
41.23	34.24	18.82	21.28	sIRTPEPVVQTGPEFHPSTSTEQPDTR	9.00	Mdc1	IPI00461995	Mdc1 Mediator of DNA damage checkpoint protein 1	1.88	0.04	115.56

Table 4.9 continued

223.44	218.56	118.06	117.37	AVsSPPTsPRPGSAATISSSASNIVPPR	31.00	Srp72	IPI00659258	Srp72 8 days embryo whole body cDNA, RIKEN full-length enriched library, clone:5730576P14 product:weakly similar to SIGNAL RECOGNITION PARTICLE 72 kDa PROTEIN (Fragment)	1.88	0.00	677.43
33.76	31.42	21.54	13.36	GEGVSQVGPGtPPAPEsPR	26.00	Tnks1bp1	IPI00459443	Tnks1bp1 182 kDa tankyrase 1-binding protein	1.87	0.07	100.08
135.85	138.13	62.11	84.64	TDVSNFDEEFTGfAPTLsPPR	454.00	Pkn1	IPI00474711	Pkn1 Bone marrow macrophage cDNA, RIKEN full-length enriched library, clone:G530007J07 product:Cardiolipin/protease-activated protein kinase-1 homolog	1.87	0.03	420.73
1403.16	1455.46	793.90	825.54	SMLQtPPDQNLSGSK	1.00	Srrm2	IPI00785384	Srrm2 Isoform 2 of Serine/arginine repetitive matrix protein 2	1.77	0.00	4478.06
56.55	72.33	35.44	38.21	ELTDsPATTK	115.00	Atbf1	IPI00119958	Atbf1 Alpha-fetoprotein enhancer-binding protein	1.75	0.07	202.52
438.36	432.90	285.63	212.29	AAELAETQsEEEK	274.00	Wdhd1	IPI00282522	Wdhd1 Isoform 1 of WD repeat and HMG-box DNA-binding protein 1	1.75	0.04	1369.18
406.50	413.17	227.46	242.99	RPAAAAAGSAsPR	491.00	Wrnip1	IPI00648000	Wrnip1 Isoform 2 of ATPase WRNIP1	1.74	0.00	1290.12
199.30	239.60	109.10	143.37	VsPEVGSADVASIAQK	63.00	D6Wsu116e	IPI00468516	D6Wsu116e DNA segment, Chr 6, Wayne State University 116, expressed	1.74	0.07	691.38
1632.50	1600.69	923.51	959.21	DSVPAsPGVPAADFPAETEQSKPSf	90.00	Top2a	IPI00122223	Top2a DNA topoisomerase 2-alpha	1.72	0.00	5115.92
2455.71	2649.63	1380.38	1613.78	LLKPGEEPSEYtDEEDTK	822.00	Pgrmc2	IPI00351206	Pgrmc2 Membrane-associated progesterone receptor component 2	1.71	0.02	8099.51
38.97	39.57	23.94	22.13	MQAESQSPtNVDLEDK	26.00	Tnks1bp1	IPI00459443	Tnks1bp1 182 kDa tankyrase 1-binding protein	1.70	0.00	124.60
683.72	682.91	342.23	462.79	LLQDSSsPVDLAK	125.00	Ncoa2	IPI00116968	Ncoa2 Nuclear receptor coactivator 2	1.70	0.04	2171.64
145.83	174.17	86.52	102.30	AGQPEAGDGTTEsPTGAAGPEK	297.00	Gltscr2	IPI00122471	Gltscr2 Adult male pituitary gland cDNA, RIKEN full-length enriched library, clone:5330430H08 product:SIMILAR TO GLIOMA TUMOR SUPPRESSOR CANDIDATE REGION GENE 2 homolog	1.69	0.06	508.82
9851.09	11870.67	7157.85	5725.65	AAsPsPQSVR	6.00	Srrm1	IPI00605037	Srrm1 Isoform 1 of Serine/arginine repetitive matrix protein 1	1.69	0.07	34605.26
509.26	637.15	313.32	367.51	DIDLFGsDEEEEDK	47.00	Eef1d	IPI00831082	Eef1d 23 kDa protein	1.68	0.08	1827.24
292.46	360.95	178.66	210.73	ATLLNVPDLsDSIHSANASER	11.00	Tjp1	IPI00135971	Tjp1 Tight junction protein ZO-1	1.68	0.07	1042.79
345.31	306.38	173.86	217.95	VIAsPQAPASK	88.00	Ccdc86	IPI00402914	Ccdc86 Coiled-coil domain-containing protein 86	1.66	0.05	1043.51
2447.33	2638.84	1466.65	1639.56	sPTNSSEIFTPAHEENVR	537.00	BC003940	IPI00114944	BC003940 hypothetical protein LOC192173	1.64	0.02	8192.39
274.13	313.02	174.80	190.09	VLSDsEEEEKDADVPGTSTR	284.00	Phip	IPI00665494	Phip pleckstrin homology domain interacting protein isoform 11	1.61	0.03	952.03
217.87	276.40	160.32	148.59	AAQSPQQHSSGDPTEEEsPV	815.00	Slc16a1	IPI00137194	Slc16a1 Monocarboxylate transporter 1	1.60	0.09	803.18
347.33	362.02	203.15	247.58	VQGTGVtPPPTPLGTR	388.00	Pdxdc1	IPI00336503	Pdxdc1 2 days pregnant adult female ovary cDNA, RIKEN full-length enriched library, clone:E330016A09 product:hypothetical Pyridoxal-dependent decarboxylase family containing protein, full insert sequence	1.57	0.03	1160.08
2899.65	2875.94	1509.24	2174.36	DIVEEsPR	194.00	Gnl3	IPI00222461	Gnl3 Isoform 1 of Guanine nucleotide-binding protein-like 3	1.57	0.09	9459.19
877.44	733.94	540.43	490.76	SQsGSPAAPVEQVVIHTDTSGDPTLPQR	9.00	Mdc1	IPI00461995	Mdc1 Mediator of DNA damage checkpoint protein 1	1.56	0.06	2642.57
2546.80	2668.26	1898.21	1507.01	HSSLPTEsDEDIAPAQR	96.00	Ap3d1	IPI00117811	Ap3d1 AP-3 complex subunit delta-1	1.53	0.05	8620.29
144.61	131.31	90.81	89.90	sGDALTTVVVK	351.00	Snip1	IPI00308559	Snip1 Smad nuclear-interacting protein 1	1.53	0.02	456.64
1731.02	1574.65	1030.06	1142.94	QIsEDVDGPDNR	395.00	Nedd4	IPI00462445	Nedd4 E3 ubiquitin-protein ligase NEDD4	1.52	0.03	5478.67
210.79	255.47	159.44	148.65	GEAEQsEEEGEEDK	5.00	Mtap1b	IPI00130920	Mtap1b Microtubule-associated protein 1B	1.51	0.08	774.36
845.89	749.02	531.44	526.77	VAAAAGSGPsPPCSPGHDR	291.00	LOC100047698	IPI00755898	Pi4k2a,LOC100047698 similar to Pi4k2a protein	1.51	0.03	2653.13

Table 4.10: Phosphopeptides reduced in DMSO treated samples, enriched in Tg003 samples. Phosphopeptides enriched in Tg003 treated EBs, reduced in DMSO treated 8 day EBs. Lowercase letters in peptide sequence refer to phosphorylated residue. As with previous iTRAQ analyses P-values are reported but fold change (50%) and total iTRAQ intensities (>100) were used as cutoff parameters.

total_114	Normalized_115	Normalized_116	Normalized_117	Peptide	Group	Gene Symbol	Accession	Protein	DMSO / Tg003	p-value	Total iTRAQ
298.99	258.56	452.83	386.63	tSGPPVSELITK	799.00	Hist1h1e	IPI00223714	Hist1h1e Histone H1.4	0.66	0.07	1397.00
325.49	277.37	480.34	433.07	LEGPAsPDVELGK EETEESK	131.00	Mll2	IPI00757871	Mll2 similar to Myeloid/lymphoid or mixed-lineage leukemia protein 2	0.66	0.04	1516.27
149.43	156.89	249.29	221.07	ELLSPLsEPDDRY PLIVK	193.00	Aff4	IPI00113246	Aff4 AF4/FMR2 family member 4	0.65	0.03	776.69
72.65	74.33	106.38	120.34	LPSTSDDCPPIGtf VR	607.00	Pparbp	IPI00313307	Pparbp Isoform 4 of Peroxisome proliferator-activated receptor-binding protein	0.65	0.03	373.70
312.51	379.35	491.77	582.53	ANATNsPEGNK	99.00	AA407452	IPI00357818	AA407452 hypothetical protein LOC627049	0.64	0.08	1766.15
113.77	113.79	196.59	158.10	sPQLTTPGQTHP GEEECR	301.00	Setd5	IPI00844617	Setd5 Isoform 3 of SET domain-containing protein 5	0.64	0.08	582.25
166.88	195.83	266.30	304.71	DDDLVEFsDLESE DDERPR	142.00	IPI00720129	IPI00720129	138 kDa protein	0.64	0.05	933.73
486.79	383.85	708.37	685.27	VTLQDYHLPDsD EDEETAIQR	518.00	Zfyve19	IPI00751484	Zfyve19 Zfyve19 protein	0.62	0.04	2264.28
41.71	61.99	85.75	83.56	AAAAGLGHPSSP GGSEDGPPIsGDE DTAR	30.00	Arid3a	IPI00125956	Arid3a AT-rich interactive domain-containing protein 3A	0.61	0.08	273.00
219.75	190.05	307.72	368.29	ENPPVEDsSDEDI KR	257.00	Stt3b	IPI00316469	Stt3b Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B	0.61	0.06	1085.81
333.38	458.79	712.15	598.25	sPSQESIGAR	753.00	Caskin2	IPI00380698	Caskin2 Caskin-2	0.60	0.09	2102.58
532.75	695.07	983.29	1062.83	FDsLEDSPEER	633.00	1110058L19Rik	IPI00315560	1110058L19Rik UPF0369 protein C6orf57 homolog precursor	0.60	0.05	3273.94
350.27	474.31	727.70	647.52	EGsPAPLEPEPGA SQPK	358.00	Foxk2	IPI00808277	Foxk2 forkhead box K2	0.60	0.07	2199.79
112.11	140.53	232.94	192.98	TARPNsEAPLSGS EDADDSNK	12.00	Eif5b	IPI00756424	Eif5b eukaryotic translation initiation factor 5B	0.59	0.07	678.56
122.50	158.02	232.65	250.80	LSTTPsPTNSLHE DGVDDFR	679.00	Tox4	IPI00121251	Tox4 TOX high mobility group box family member 4	0.58	0.04	763.97
83.35	77.45	155.41	123.47	GNLETHEDSQVF sPK	59.00	Zc3h13	IPI00515528	Zc3h13 RIKEN cDNA 3110050K21	0.58	0.07	439.68
726.98	525.19	1127.88	1049.21	DVPPDILLDsPER	35.00	Nipbl	IPI00357096	Nipbl Isoform 2 of Nipped-B-like protein	0.58	0.05	3429.26
1304.97	1021.75	2038.10	2080.28	sASSDTSEELNSC DSPK	41.00	Slc9a3r1	IPI00109311	Slc9a3r1 Ezrin-radixin-moesin-binding phosphoprotein 50	0.56	0.02	6445.09
106.02	74.37	172.20	170.47	ESAsPTIPNLDLL EAHTK	379.00	Phf2	IPI00123868	Phf2 PHD finger protein 2	0.53	0.04	523.06
27.94	51.41	75.78	76.28	GSDAVsETSSVS HIEDLEK	388.00	Pdxdc1	IPI00336503	Pdxdc1 2 days pregnant adult female ovary cDNA, RIKEN full-length enriched library, clone:E330016A09 product:hypothetical Pyridoxal-dependent decarboxylase family containing protein, full insert sequence	0.52	0.09	231.40

Table 4.10 continued

256.91	263.11	490.18	509.34	LQAALDNEAGGR PAMEPGNGsLDL GGDAAGR	508.00	Hnrpu	IPI00458583	Hnrpu Osteoclast-like cell cDNA, RIKEN full-length enriched library, clone:I420039N16 product:heterogeneous nuclear ribonucleoprotein U, full insert sequence	0.52	0.00	1519.55
162.26	111.69	298.46	239.82	sPPEGDTTLFLSR	335.00	Dido1	IPI00453851	Dido1 Isoform 1 of Death-inducer obliterator 1	0.51	0.08	812.24
164.57	203.02	414.03	355.24	sPTGPSNSFLAN MGGTVAHK	264.00	Rbm17	IPI00170394	Rbm17 Splicing factor 45	0.48	0.03	1136.86
50.64	46.97	101.87	112.81	LQDISELLAtGVG LSDSEVEPDGEH NTTELPQAVAGR	326.00	Ppan	IPI00130510	Ppan Suppressor of SWI4 1 homolog	0.45	0.01	312.29
15.41	33.59	55.97	52.81	IPsDNLEEPVPK	738.00	Adnp	IPI00672180	Adnp,Dpm1 activity-dependent neuroprotective protein	0.45	0.08	157.78
64.73	92.04	204.03	146.64	VEEASPPAVQQP TDPASPtVATTP EPVGGDAGDK	37.00	Dnmt3a	IPI00131694	Dnmt3a Isoform 1 of DNA (cytosine-5)-methyltransferase 3A	0.45	0.09	507.44
52.28	70.84	154.41	158.01	LSVGEQPDVAHD PYEFLQsPEPAAP AK	690.00	Brd9	IPI00762915	Brd9 bromodomain containing 9	0.39	0.01	435.54
61.35	197.65	384.56	338.91	RLsEEACPGVLSV APTVTQPPGR	368.00	Akap1	IPI00230592	Akap1 Isoform 5 of A kinase anchor protein 1, mitochondrial precursor	0.36	0.08	982.47
120.29	111.11	329.96	327.24	TEEDRENTQIDDT EPLSPVSNsK	3.00	Trp53bp1	IPI00229801	Trp53bp1 Transformation related protein 53 binding protein 1	0.35	0.00	888.61
62.65	39.22	186.67	141.99	FPQGPPSQMGsP MGNR	858.00	Pspc1	IPI00321266	Pspc1 Paraspeckle protein 1	0.31	0.05	430.53
9.18	23.43	51.07	54.17	SAPPAGAAPSTL SAATPGTSEDHS QsATPPEQQR	765.00	Hoxc4	IPI00122942	Hoxc4 Homeobox protein Hox-C4	0.31	0.04	137.85
58.39	37.27	120.02	189.77	AEAPAGPALGLP sPEVESGLER	467.00	1200014J11Rik	IPI00278062	1200014J11Rik 6 days neonate spleen cDNA, RIKEN full-length enriched library, clone:F430012F01 product:hypothetical protein, full insert sequence	0.31	0.10	405.44
70.46	76.23	249.86	226.69	ASMSEFLEsEDGE VEQQR	523.00	Suz12	IPI00396676	Suz12 Polycomb protein Suz12	0.31	0.01	623.24
26.36	74.78	217.28	146.00	NsPNNISGISNPPG TPR	471.00	Ssbp3	IPI00649271	Ssbp3 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone:A630064O18 product:single-stranded DNA binding protein 3, full insert sequence	0.28	0.09	464.43
25.44	18.28	113.32	67.79	SLVSPIPsPTGTIS VPNSCPAsPR	98.00	Foxk1	IPI00656285	Foxk1 Forkhead box protein K1	0.24	0.10	224.82
16.21	0.00	59.40	77.15	EGLLPASDQEAG GPSDIFSSSsPLDI	63.00	D6Wsu116e	IPI00468516	D6Wsu116e DNA segment, Chr 6, Wayne State University 116, expressed	0.12	0.04	152.75
25.96	0.00	119.88	128.66	ETVVSGPLGVEDI SPSMsPDDK	3.00	Trp53bp1	IPI00229801	Trp53bp1 Transformation related protein 53 binding protein 1	0.10	0.01	274.49
0.08	0.00	110.20	122.46	EAsSPPTHIQLQE EMEPVEGPAPPV LIQLSPPNTDAGA R	15.00	Acin1	IPI00230677	Acin1 Isoform 4 of Apoptotic chromatin condensation inducer in the nucleus	0.00	0.00	232.74

4.4 Discussion

4.4.1 Identification of changes in phosphorylation independent of changes in underlying protein level

Phosphoproteomics in combination with whole proteome analysis allowed us to identify changes in phosphorylation of proteins before and after differentiation of pluripotent cells. Semi-quantitative measurement of phosphopeptides alone is insufficient to determine changes in phosphorylation since underlying protein levels are also subject to changes. I identified >1000 phosphoproteins in both ES and EC cells and was able to identify phosphoproteins enriched before and after differentiation by filtering for those proteins whose levels were unchanged at the whole proteome level. These global changes were previously invisible to systems biologists studying pluripotent cells during neural differentiation[106] and as such were not reported in similar studies of the transcriptome by Aiba et al.

4.4.2 Phosphorylation regulates splicing factors during differentiation

To identify changes in phosphorylation that are independent of changes in underlying protein levels, I looked only at proteins whose levels remained unchanged before and after differentiation in the iTRAQ data. Interestingly, the resulting set of proteins specifically phosphorylated in differentiated cells was made up almost entirely of regulators of splicing. This suggests that in addition to the other observed changes that occur in differentiating EC and ES cells, global splicing is also altered. Many of the identified splicing factors are phosphorylated at serines flanked by acidic residues that are putative targets of the acidophilic

kinases CK2 and Clk1. CK2 and Clk1 regulate key splicing factors that control alternative splicing of a wide variety of genes[221-229]. These observations lead me to hypothesize that phosphorylation of the observed splicing factors is due to activation of Clk2 and Clk1 and leads to global changes in splicing.

To test this hypothesis I inhibited Clk1 and 2 using the chemical inhibitor Tg003, and found that inhibition of Tg003 had no effect on the phosphorylation of these splicing factors.

Chapter 4 includes contributions from Zhouxin Shen and Steven Briggs. The dissertation author was the primary investigator and author of this material.

Chapter 5: Conclusions and future directions

In order to better to better understand post-transcriptional events associated with pluripotency, I performed deep, quantitative proteome analysis of two pluripotent cell types, ES and EC cells to identify proteins and protein phosphorylation events associated with pluripotency and with the differentiation of pluripotent cells. These analyses have identified 1) new markers of the pluripotent state that are common to undifferentiated ES and EC cells 2) Cases of post-transcriptional regulation of protein levels in pluripotent cells 3) phosphorylation of proteins associated with pluripotency 4) phosphorylation events enriched before and after differentiation due to changes in kinase or phosphatase activity, rather than changes in the underlying protein levels.

This deep, quantitative proteome analysis of ES cells is an important step in understanding not just transcriptional underpinnings of pluripotency, but also the structural, metabolic and adhesive characteristics of pluripotent cells. Such a broad analysis, with resulting large datasets inevitably leads to far more observations and hypotheses than can be tested in one graduate career.

5.1 UTF1

I chose to focus on phosphorylation of one ES-specific transcription factor, this limited focus lead to minor discoveries about the role of phosphorylation of UTF1, but was perhaps, too zoomed in and focused to lead to major conclusions about how regulation of transcription factors in ES cells affects pluripotency.

A result that intrigues me is the observation that UTF1 does not seem to repress Oct4 reporter expression in ES cells. This dovetails nicely with data from Zhao et al. 2008[219] reporting increased efficiency of reprogramming when iPSCs are reprogrammed with UTF1 in addition to the Yamanaka cocktail. I hypothesize that UTF1 is necessary for repression of non-pluripotency proteins during early development, and will repress most viral and pro-differentiation genes, while failing to repress the markers of pluripotency that I identified in chapter 2. If I had more time, I would profile UTF1 knockdown and control P19CL6 EC cells (courtesy of Bart Eggan) to test this hypothesis. As an alternative, I've tried to convince two other scientists to work on this idea. One works on silencing of herpes viruses in ES cells. Since TK is a herpes viral gene repressed by UTF1, I hypothesize that other herpes virus genes are also repressed by UTF1, and thus UTF1 is a factor that causes repression of herpes viruses.

The other scientist, Gerald Pao works on reprogramming of somatic cells, and found that enforced non-specific expression of the entire genome leads to a stable, iPS-like state, Gerald hypothesizes that this non-specific expression of genes activate pluripotency associated factors and repress non-pluripotency associated factors leading to a stable iPS state. UTF1 is a good candidate for such a repressor since it is the classic target the Oct4/Sox2 dimer-mediated activation[56, 63, 217] and is a powerful repressor of transcription.

I eagerly await learning the results of both these sets of experiments, and would have liked to work on them myself, if I were starting my dissertation research at this stage.

5.2 The proteomics approach

A major advantage of a proteomics approach to systems biology is the ability to identify the subcellular localization and relative abundance of with a cell: this is an area that I feel that I did not adequately explore during my graduate career. A targeted analysis of the major organelles of ES and EC cells before and after differentiation is likely to provide insights into pluripotency-specific differences in some organelles. I specifically would like to look at the Ribosome and mitochondria of ES cells, since these are the organelles with many subunits enriched compared with differentiated cells. It has been noted that the ribosomes of ES cells [111]

Similarly, phosphoproteomics offers the ability to identify proximal signaling events that influence self-renewal and differentiation. I believe a powerful future experiment would be to look at the phosphoproteome of ES cells after treatment with pro-differentiation signaling factors such as ATRA or Tgfβ to identify the short-term changes to the phosphoproteome of ES cells, before any confounding changes to the whole proteome have occurred.

5.3 Final thoughts

An interesting thing about finishing a dissertation such as this one is that I've had a chance to look back over m research and can see what I would do differently. This is not to say that I am not happy with my proteome analyses: I believe that they will add to our understanding of the post-transcriptional changes that occur in pluripotent cells during differentiation. My work on UTF1 would have

greatly benefited from a knock-in/knockout approach, and generating a UTF1 knockout ES line remains an important goal for this area of research. Finally, my last chapter on mRNA splicing factors is an exciting area of my thesis that I never took the time to fully follow up on after my candidate kinase turned out not to be responsible for phosphorylating the 8 SR proteins of interest. That said, I am quite happy with this approach to phosphoproteomics. Without whole proteome data to compare with the phosphoproteome, I don't think that phosphoproteomics will ever be more than a way to make a laundry list of phosphorylation events in a certain cell type. The approach would be especially useful when looking at changes to the phosphoproteome of ES cells after addition of a growth factor or inhibitor that causes differentiation.

Appendix: Materials and methods

A.1 Materials and Methods Chapter 2

A.1.1 Cell Culture

P19 cells (ATCC no. CRL-1825) were cultured as described in Aiba et al.[106] Cells were grown in high glucose DMEM (Gibco) supplemented with 10% Heat inactivated-Fetal Bovine Serum (HI-FBS) and 100 units penicillin/streptomycin. Cells were subcultured 1:10 every 2-3 days on 150 mm tissue culture dishes treated with 0.1% Gelatin.

For differentiation, P19 cells were plated on bacteriological plates in tissue culture media supplemented with 1μM RA (10 mM stock of ATRA dissolved in 100% ethanol) as described in Aiba et al 2006[106]. Media was replaced on day 2, EBs were collected for processing on day 4.

129/SvEv mouse ES cells from ATCC were cultured as described in Aiba et al. Cells were grown in high-glucose DMEM (Gibco) supplemented with glutamax, LIF, 15% HI-FBS and 100 units pen/strep, 1mM Sodium Pyruvate, 1 mM NEAA and 55 μM β-mercaptoethanol. Cells were subcultured 1:5 every 2-3 days on 150 mm tissue culture dishes treated with 0.1% Gelatin.

For differentiation, 129/SvEv mouse ES cells were plated on bacteriological plates in ES culture media minus LIF as described in Aiba et al 2006[106]. Media was replaced on day 2, day 4 and day 6. Cells were treated with 1μM RA on days 4-8.

A.1.2 Cell lysis, reduction and trypsinization

100 μL cell pellets were lysed in 250 μL lysis buffer (2% (w/v) RapiGest (Waters, 186002122), 1mM EDTA, and 50mM Hepes buffer (pH 7.2)). Cysteines were reduced and alkylated using 1 mM Tris(2-carboxyethyl) phosphine (TCEP, Fisher, AC36383)) at 95°C for 5 minutes followed by 2.5 mM iodoacetamide (Fisher, AC12227) at 37°C in the dark for 15 minutes. Protein concentrations were measured using Bradford assay (Pierce). Proteins were digested with trypsin (Roche, 03 708 969 001) at the enzyme-to-substrate ratio (w:w) = 1:50 overnight.

A.1.3 iTRAQ mass tagging of peptides

For iTRAQ (Applied Biosystems, Foster City, CA) derivatization an aliquot of each digested sample (100 μg of total protein) was treated with one tube of one of the iTRAQ reagents in 70% isopropanol at pH 7.2 for 2 hours at room temperature. Labeled samples were dried down in a vacuum concentrator. 100 μL of water was added to each tube to dissolve the peptides. Samples tagged with 4 different iTRAQ reagents were pooled together. 1% trifluoroacetic acid (TFA), pH 1.4 was added to precipitate RapiGest. Samples were incubated at 4°C overnight and then centrifuged at 16,100 g for 15 minutes. Supernatant was collected and centrifuged through a 0.22 μM filter and was used for LC-MS/MS analysis. iTRAQ labeling efficiency was calculated by searching the MS/MS data specifying 4 possible iTRAQ modifications: 1) fully labeled; 2) n-terminus-labeled only; 3) lysine-labeled only; and 4) non-labeled. Using the above protocol I obtained higher than 90% iTRAQ labeling efficiency for all datasets.

A.1.4 On-line separation of peptides by HPLC

An Agilent 1100 HPLC system (Agilent Technologies, Santa Clara, CA) delivered a flow rate of 300 nL per minute to a 3-phase capillary chromatography column through a splitter. Using a custom pressure cell, 5 μm Zorbax SB-C18 (Agilent) was packed into fused silica capillary tubing (200 μm ID, 360 μm OD, 20 cm long) to form the first reverse phase column (RP1). A 5 cm long strong cation exchange (SCX) column packed with 5 μm PolySulfoethyl (PolyLC, Inc.,) was connected to RP1 using a zero dead volume 1 μm filter (Upchurch, M548) attached to the exit of the RP1 column. A fused silica capillary (100 μm ID, 360 μm OD, 20 cm long) packed with 5 μm Zorbax SB-C18 (Agilent) was connected to SCX as the analytical column (the second reverse phase column). The electro-spray tip of the fused silica tubing was pulled to a sharp tip with the inner diameter smaller than 1 μm using a laser puller (Sutter P-2000). The peptide mixtures were loaded onto the RP1 using the custom pressure cell. Columns were not re-used. Peptides were first eluted from the RP1 to the SCX column using a 0 to 80% acetonitrile gradient for 150 minutes. The peptides were fractionated by the SCX column using a series of salt gradients (from 10 mM to 1 M ammonium acetate for 20 minutes), followed by high resolution reverse phase separation using an acetonitrile gradient of 0 to 80% for 120 minutes. Typically it takes 4 days (38 salt fractions) for each full proteome analysis.

A.1.5 MS/MS analysis

Spectra were acquired using a LTQ linear ion trap tandem mass spectrometer (Thermo Electron Corporation, San Jose, CA) employing

automated, data-dependent acquisition. The mass spectrometer was operated in positive ion mode with a source temperature of $150^{\circ}C$.

The full MS scan range of 400-2000 m/z was divided into 3 smaller scan ranges (400-800, 800-1050, 1050-2000) to improve the dynamic range. Both CID (Collision Induced Dissociation) and PQD (Pulsed-Q Dissociation) scans of the same parent ion were collected for protein identification and quantitation. Each MS scan was followed by 4 pairs of CID-PQD MS/MS scans of the most intense ions from the parent MS scan. A dynamic exclusion of 1 minute was used to improve the duty cycle of MS/MS scans. About 20,000 MS/MS spectra were collected for each salt step fractionation.

A.1.6 Data analysis

The raw data were extracted and searched using Spectrum Mill v3.03 (Agilent). The CID and PQD scans from the same parent ion were merged together. MS/MS spectra with a sequence tag length of 1 or less were considered to be poor spectra and were discarded. The remaining MS/MS spectra were searched against the International Protein Index (IPI) mouse database (v3.31, 56,555 protein sequences). The enzyme parameter was limited to fully tryptic peptides with a maximum miscleavage of 1. All other search parameters were set to default settings of Spectrum Mill (carbamidomethylation of cysteines, iTRAQ modification, +/- 2.5 Da for precursor ions, +/- 0.7 Da for fragment ions, and a minimum matched peak intensity (SPI%) of 50%). A concatenated forward-reverse database was constructed to calculate the in-situ false discovery rate (FDR). The total number of protein sequences in the

combined database is 113,110.   Cutoff scores were dynamically assigned to each dataset to maintain the false discovery rate less than 1% at the protein level. Only proteins with 2 or more unique peptides were validated and selected for following quantitative analysis. Proteins that share common peptides were grouped to address the database redundancy issue. The proteins within the same group shared the same set or subset of unique peptides.

A.1.7 Relative Quantification by iTRAQ mass tagging reagent

Protein iTRAQ intensities were calculated by summing the peptide iTRAQ intensities from each protein group.   Peptides shared among different protein groups were removed before quantitation. A minimal total iTRAQ intensity of 100 was used to filter out low intensity spectra. Isotope impurities of iTRAQ reagents were corrected using correction factors provided by the manufacturer (Applied Biosystems).

Median normalization was performed to normalize the protein iTRAQ reporter intensities in which the log ratios between different iTRAQ tags (115/114, 116/114, 117/114) are adjusted globally such that the median log ratio is zero.

Quantitative analysis was performed on the normalized protein iTRAQ intensities. Protein ratios between undifferentiated and differentiated cells were calculated by taking the ratios of the total iTRAQ intensities from the corresponding iTRAQ reporters. T-test (two tailed, paired) was used to calculate the p-values. Proteins with more than 50% change and p-values less than 0.05 were considered significantly changed after differentiation.

A.1.8 Reanalysis of microarray data

Microarray data from Aiba et al. were obtained from pubmed GEO datasets. Data were processed by normalizing expression values to reference data and generating undifferentiated:differentiated ratios of mRNA measurements from each replicate of the experiment. Ratios were averaged in excel using the geomean function, and the student's T-test was performed to obtain p values. Protein and mRNA data were combined in Microsoft access by matching Gene symbols and accession NCBI numbers between the two tables.

A.1.9 RT-PCR

RNA was isolated from undifferentiated and differentiated cells using Trizol (invitrogen cat #15596-026) according to manufacturer's instructions. RNA quality was checked by agarose gel electrophoresis, and 2μg total RNA was treated with DNAase Turbo (ambion catalog #AM2238) according to manufacture's instructions. After DNAse inactivation, cDNA was synthesized by reverse transcription with/without Superscript II (invitrogen) and 10 μM random hexamer primers. RT PCR reactions were performed using the following primer sequences:

Beta actin: F-GATCTGGCACCACACCTTCTACAATG R-CGTACATGGCTGGGGTGTTGAAG, Oct4: F-CTCCCGAGGAGTCCCAGGACAT, R-GATGGTGGTCTGGCTGAACACCT, Pdlim7: F-GCACTCAGGAGCAGGCACGATGG, R-CCTCCGGGCGTGAGCCG, Dpysl2: F-GTGACGCCCAAGACGGTGAC, R-ATGTTGTCGTCAATCTGAGCACCAG, H2afy F-

GGAGAAGAAGGGCGGCAAGG, R-GGCCTGCACTAATAGCAGCTC, Utf1 F-GGTTCGCCGCCGCTCTACTG, R-GCAGGGGCAGGTTCGTCATTTTC, Racgap1 F-TCCTTATGATCCACfFCTACAGAGAGTG, R-GCGCTCCACCACCTTG, Sall4 F-GGAGAGAAGCCTTTCGTGTG R-CTCTATGGCCAGCTTCCTTC. All RT-PCR reactions were performed at 22, 25 and 27 cycles to avoid saturation of the PCR products.

A.1.10 Western blotting

Cells were collected in RIPA buffer and protein levels were quantified by Bradford. Protein samples were prepared in 1X SDS loading buffer containing β-ME, sonicated and run on 10% tris-gylcine protein gels, transferred to PVDF membranes. Membranes were blocked with 5% W/V non-fat dry milk in PBS-0.05% Tween and probed with the appropriate antibodies at concentrations from 1:100 to 1:1000. Specific antibodies: αUTF1 (RB chemicon, 1:1000), αSall4 (Santa Cruz 1:1000), αPDLIM7 (RB 1:1000 Chemicon), αH2afy (1:100 Santa Cruz), αDpysl2 (1:1000 abnova), αActin (MS SCBT 1:1000) or αTubulin were used as loading controls.

A.1.11 Cellular Immunofluorescence

Undifferentiated cells and differentiated cells were plated on coverslips coated with poly-d-lysine and cultured overnight. Cells were fixed in 4% paraformaldehyde, permeablized with PBS-0.2% Triton, blocked with 10% BSA in PBS-0.1% triton for 10 hour and stained with the appropriate pair of primary antibodies diluted 1:1000 (Oct4 and NeuN or UTF1 and bIII-Tubulin) in blocking buffer or left in blocking buffer, so secondary alone controls. Cells were washed

3X for 15 minutes in PBS 0.1% Triton and incubated with secondary antibodies from molecular probes diluted 1:1000 (antiMouse-534 and antirabbit-488). Slides were washed 3X and mounted in vectashield (vector labs) plus dapi.

A.2 Materials and Methods Chapter 3

A.2.1 Cell Culture

P19 cells from ATCC were cultured as described in Aiba et al. [106] Cells were grown in DMEM supplemented with 10% HI-FBS and 100 units pen/strep. Cells were subcultured 1:10 every 2-3 days on 150 mm tissue culture dishes treated with 0.1% Gelatin.

For differentiation, P19 cells were plated on bacterialogical plates in tissue culture media supplemented with 1μM RA (1 mM RA dissolved in 100% ethanol) as described in Aiba et al 2006[106]. Media was replaced on day 2, and EBs were collected on day 4.

129/SvEv mouse ES cells from ATCC were cultured as described in Aiba et al. Cells were grown in DMEM supplemented with 15% HI-FBS and 100 units pen/strep, 1000 units of LIF, 1mM Sodium Pyruvate, 1 mM NEAA and 55 μM β-mercaptoethanol. Cells were subcultured 1:5 every 2-3 days on 150 mm tissue culture dishes treated with 0.1% Gelatin.

For differentiation, 129/SvEv mouse ES cells were plated on bacteriological plates in tissue ES culture media minus LIF as described in Aiba et al 2006[106]. media was replaced on day 2, day 4 and day 6. Cells were treated

with 1μM RA on days 4-8 (1:10,000 dilution of10 mM RA dissolved in 100%
ethanol).

A.2.2 Sample preparation

100 μL cell pellets were lysed in 250 μL lysis buffer (2% (w/v) RapiGest
(Waters, 186002122), 1mM EDTA, and 50mM Hepes buffer (pH 7.2)) plus
phosphatase inhibitors (NaF, Vanadate). Cysteines were reduced and alkylated
using 1 mM Tris(2-carboxyethyl) phosphine (TCEP, Fisher, AC36383)) at 95°C
for 5 minutes followed by 2.5 mM iodoacetamide (Fisher, AC12227) at 37°C in
the dark for 15 minutes. Protein concentrations were measured using Bradford
assay (Pierce). Proteins were digested with trypsin (Roche, 03 708 969 001) at
the enzyme-to-substrate ratio (w:w) = 1:50 overnight.

A.2.3 TiO$_2$ columns to enrich phosphopeptides

Phosphopeptide enrichment was performed using home-made TiO2
columns. 0.2g TiO2 powder (Sigma# 224227) was weighted and added to an
empty 0.22 uM filter (Fisher# 07-200-386). TiO2 was washed by 500 μL water
twice by centrifuging at 5,000 f for 5 minutes. Tryptic digested peptides (pH 1.4)
from above was added to the TiO$_2$ column. Samples were centrifuged at 1,000g
for 30 minutes until the column was dry. Flow through was discarded. Metal
oxide column was washed by 500μL 1% TFA twice by centrifuging at 5,000 f for
5 minutes. Enriched phosphopeptides were eluted by 200μL 100mM
(NH$_4$)$_2$HPO$_4$ by centrifuging at 1,000g for 30 minutes until the column was dry.
Eluted peptides were acidified by adding formic acid to a final concentration of
2%.

A.3 Materials and Methods Chapter 4

A.3.1 Clk 1 inhibition

Undifferentiated mouse ES cells of line 129/SvEv were plated at 40% confluence in feeder-free plates pretreated with 0.1% gelatin. Cells were grown in the presence of 0.1% DMSO or 1 μM Tg003 dissolved in DMSO for 24 hours. Cells were then aggregated in bacteriological plates without LIF in the presence of DMSO or Tg003. At day 4, 1 μM ATRA was added to EBs. After 8 days of differentiation in the presence of Tg003 or DMSO, EBs were collected and processed for whole and phosphoproteome analysis as described in methods for chapter 2 and 3.

REFERENCES

1. Gilbert, S.F. in Developmental Biology, Edn. 6th. (ed. S.F. Gilbert) Online edition, NCBI books (Sinauer Associates, Sunderland, MA; 2000).

2. Evans, M.J. & Kaufman, M.H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-156 (1981).

3. Kaufman, M.H., Robertson, E.J., Handyside, A.H. & Evans, M.J. Establishment of pluripotential cell lines from haploid mouse embryos. *J Embryol Exp Morphol* **73**, 249-261 (1983).

4. Thomson, J.A. et al. Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145-1147 (1998).

5. Alberts, B. in Molecular Biology of the Cell, Edn. 4th. (ed. D.B. Bruce Alberts, Julian Lewis, Martin Raff, Keith Roberts, and James D. Watson) NCBI online edition (Garland Science, New York, New York; 2002).

6. in Stem Cell Information, Vol. 2009 (National Institutes of Health, U.S. Department of Health and Human Services, 2009, Bethesda, MD; 2009).

7. Lajtha, L.G. Stem cell concepts. *Nouv Rev Fr Hematol* **21**, 59-65 (1979).

8. Thomas, E.D. et al. Allogeneic marrow grafting using HL-A matched donor-recipient sibling pairs. *Trans Assoc Am Physicians* **84**, 248-261 (1971).

9. Thomas, E.D. et al. Allogeneic marrow grafting for hematologic malignancy using HL-A matched donor-recipient sibling pairs. *Blood* **38**, 267-287 (1971).

10. Huntly, B.J. & Gilliland, D.G. Leukaemia stem cells and the evolution of cancer-stem-cell research. *Nat Rev Cancer* **5**, 311-321 (2005).

11. Chury, Z. & Tobiska, J. [Clinical findings & results of culture in a case of stem-cell leukemia with pluripotential properties of the stem cells.]. *Neoplasma* **5**, 220-231 (1958).

12. Sinkovics, J.G., Sykes, J.A., Shullenberger, C.C. & Howe, C.D. Patterns of growth in cultures deriving from human leukemic sources. *Tex Rep Biol Med* **25**, 446-467 (1967).

13. Goodell, M.A., Brose, K., Paradis, G., Conner, A.S. & Mulligan, R.C. Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo. *J Exp Med* **183**, 1797-1806 (1996).

14. Gage, F.H. Mammalian neural stem cells. *Science* **287**, 1433-1438 (2000).

15. Gage, F.H., Ray, J. & Fisher, L.J. Isolation, characterization, and use of stem cells from the CNS. *Annu Rev Neurosci* **18**, 159-192 (1995).

16. Kuhn, H.G., Dickinson-Anson, H. & Gage, F.H. Neurogenesis in the dentate gyrus of the adult rat: age-related decrease of neuronal progenitor proliferation. *J Neurosci* **16**, 2027-2033 (1996).

17. Palmer, T.D., Takahashi, J. & Gage, F.H. The adult rat hippocampus contains primordial neural stem cells. *Mol Cell Neurosci* **8**, 389-404 (1997).

18. Shihabuddin, L.S., Palmer, T.D. & Gage, F.H. The search for neural progenitor cells: prospects for the therapy of neurodegenerative disease. *Mol Med Today* **5**, 474-480 (1999).

19. Becker, A.J., Mc, C.E. & Till, J.E. Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. *Nature* **197**, 452-454 (1963).

20. Uccelli, A., Moretta, L. & Pistoia, V. Mesenchymal stem cells in health and disease. *Nat Rev Immunol* (2008).

21. Kiger, A.A., Jones, D.L., Schulz, C., Rogers, M.B. & Fuller, M.T. Stem cell self-renewal specified by JAK-STAT activation in response to a support cell cue. *Science* **294**, 2542-2545 (2001).

22. Blank, U., Karlsson, G. & Karlsson, S. Signaling pathways governing stem-cell fate. *Blood* **111**, 492-503 (2008).

23. Nemeth, M.J. & Bodine, D.M. Regulation of hematopoiesis and the hematopoietic stem cell niche by Wnt signaling pathways. *Cell Res* **17**, 746-758 (2007).

24. Walker, M.R., Patel, K.K. & Stappenbeck, T.S. The stem cell niche. *J Pathol* **217**, 169-180 (2009).

25. Gupta, P.B., Chaffer, C.L. & Weinberg, R.A. Cancer stem cells: mirage or reality? *Nat Med* **15**, 1010-1012 (2009).

26.   Ogden, A.T. et al. Identification of A2B5+CD133- tumor-initiating cells in adult human gliomas. *Neurosurgery* **62**, 505-514; discussion 514-505 (2008).

27.   Panagiotakos, G. & Tabar, V. Brain tumor stem cells. *Curr Neurol Neurosci Rep* **7**, 215-220 (2007).

28.   Zhu, Y., Ghosh, P., Charnay, P., Burns, D.K. & Parada, L.F. Neurofibromas in NF1: Schwann cell origin and role of tumor environment. *Science* **296**, 920-922 (2002).

29.   Grandics, P. The cancer stem cell: evidence for its origin as an injured autoreactive T cell. *Mol Cancer* **5**, 6 (2006).

30.   Reya, T., Morrison, S.J., Clarke, M.F. & Weissman, I.L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105-111 (2001).

31.   Clark, A.T. The stem cell identity of testicular cancer. *Stem Cell Rev* **3**, 49-59 (2007).

32.   Bridges, B. & Hussain, A. Testicular germ cell tumors. *Curr Opin Oncol* **19**, 222-228 (2007).

33.   Tummala, M.K. & Hussain, A. Recent developments in germ cell tumors of the testes. *Curr Opin Oncol* **20**, 287-293 (2008).

34.   van de Geijn, G.J., Hersmus, R. & Looijenga, L.H. Recent developments in testicular germ cell tumor research. *Birth Defects Res C Embryo Today* **87**, 96-113 (2009).

35.   Kleinsmith, L.J. & Pierce, G.B., Jr. Multipotentiality of Single Embryonal Carcinoma Cells. *Cancer Res* **24**, 1544-1551 (1964).

36.   Damjanov, I. Teratocarcinoma: neoplastic lessons about normal embryogenesis. *Int J Dev Biol* **37**, 39-46 (1993).

37.   Przyborski, S.A., Christie, V.B., Hayman, M.W., Stewart, R. & Horrocks, G.M. Human embryonal carcinoma stem cells: models of embryonic development in humans. *Stem Cells Dev* **13**, 400-408 (2004).

38.   Cervantes, R.B., Stringer, J.R., Shao, C., Tischfield, J.A. & Stambrook, P.J. Embryonic stem cells and somatic cells differ in mutation frequency and type. *Proc Natl Acad Sci U S A* **99**, 3586-3590 (2002).

39. Hentze, H. et al. Teratoma formation by human embryonic stem cells: Evaluation of essential parameters for future safety studies. *Stem Cell Res* (2009).

40. Sukoyan, M.A. et al. Embryonic stem cells derived from morulae, inner cell mass, and blastocysts of mink: comparisons of their pluripotencies. *Mol Reprod Dev* **36**, 148-158 (1993).

41. Lensch, M.W., Schlaeger, T.M., Zon, L.I. & Daley, G.Q. Teratoma formation assays with human embryonic stem cells: a rationale for one type of human-animal chimera. *Cell Stem Cell* **1**, 253-258 (2007).

42. Bain, G., Kitchens, D., Yao, M., Huettner, J.E. & Gottlieb, D.I. Embryonic stem cells express neuronal properties in vitro. *Dev Biol* **168**, 342-357 (1995).

43. Dushnik-Levinson, M. & Benvenisty, N. Embryogenesis in vitro: study of differentiation of embryonic stem cells. *Biol Neonate* **67**, 77-83 (1995).

44. Fraichard, A. et al. In vitro differentiation of embryonic stem cells into glial cells and functional neurons. *J Cell Sci* **108 (Pt 10)**, 3181-3188 (1995).

45. Keller, G.M. In vitro differentiation of embryonic stem cells. *Curr Opin Cell Biol* **7**, 862-869 (1995).

46. Habib, M., Caspi, O. & Gepstein, L. Human embryonic stem cells for cardiomyogenesis. *J Mol Cell Cardiol* **45**, 462-474 (2008).

47. Lev, S., Kehat, I. & Gepstein, L. Differentiation pathways in human embryonic stem cell-derived cardiomyocytes. *Ann N Y Acad Sci* **1047**, 50-65 (2005).

48. Niebruegge, S. et al. Generation of human embryonic stem cell-derived mesoderm and cardiac cells using size-specified aggregates in an oxygen-controlled bioreactor. *Biotechnol Bioeng* **102**, 493-507 (2009).

49. Dinsmore, J. et al. Embryonic stem cells differentiated in vitro as a novel source of cells for transplantation. *Cell Transplant* **5**, 131-143 (1996).

50. Doetschman, T., Shull, M., Kier, A. & Coffin, J.D. Embryonic stem cell model systems for vascular morphogenesis and cardiac disorders. *Hypertension* **22**, 618-629 (1993).

51.   Scholer, H.R., Balling, R., Hatzopoulos, A.K., Suzuki, N. & Gruss, P. Octamer binding proteins confer transcriptional activity in early mouse embryogenesis. *EMBO J* **8**, 2551-2557 (1989).

52.   Scholer, H.R., Hatzopoulos, A.K., Balling, R., Suzuki, N. & Gruss, P. A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. *EMBO J* **8**, 2543-2550 (1989).

53.   Takeda, J., Seino, S. & Bell, G.I. Human Oct3 gene family: cDNA sequences, alternative splicing, gene organization, chromosomal location, and expression at low levels in adult tissues. *Nucleic Acids Res* **20**, 4613-4620 (1992).

54.   Nichols, J. et al. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379-391 (1998).

55.   Avilion, A.A. et al. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* **17**, 126-140 (2003).

56.   Remenyi, A. et al. Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev* **17**, 2048-2059 (2003).

57.   Cavaleri, F. & Scholer, H.R. Nanog: a new recruit to the embryonic stem cell orchestra. *Cell* **113**, 551-552 (2003).

58.   Chambers, I. et al. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643-655 (2003).

59.   Mitsui, K. et al. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631-642 (2003).

60.   Catena, R. et al. Conserved POU binding DNA sites in the Sox2 upstream enhancer regulate gene expression in embryonic and neural stem cells. *J Biol Chem* **279**, 41846-41857 (2004).

61.   Kuroda, T. et al. Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression. *Mol Cell Biol* **25**, 2475-2485 (2005).

62.   Rodda, D.J. et al. Transcriptional regulation of nanog by OCT4 and SOX2. *J Biol Chem* **280**, 24731-24737 (2005).

63. Nishimoto, M. et al. Structural analyses of the UTF1 gene encoding a transcriptional coactivator expressed in pluripotent embryonic stem cells. *Biochem Biophys Res Commun* **285**, 945-953 (2001).

64. Chamberlain, S.J., Yee, D. & Magnuson, T. Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. *Stem Cells* **26**, 1496-1505 (2008).

65. Herranz, N. et al. Polycomb complex 2 is required for E-cadherin repression by the Snail1 transcription factor. *Mol Cell Biol* **28**, 4772-4781 (2008).

66. Lee, T.I. et al. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301-313 (2006).

67. Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M. & Helin, K. The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol Cell Biol* **27**, 3769-3779 (2007).

68. Chen, T., Ueda, Y., Dodge, J.E., Wang, Z. & Li, E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol Cell Biol* **23**, 5594-5605 (2003).

69. Datta, J., Ghoshal, K., Sharma, S.M., Tajima, S. & Jacob, S.T. Biochemical fractionation reveals association of DNA methyltransferase (Dnmt) 3b with Dnmt1 and that of Dnmt 3a with a histone H3 methyltransferase and Hdac1. *J Cell Biochem* **88**, 855-864 (2003).

70. Majumder, S. et al. Role of de novo DNA methyltransferases and methyl CpG-binding proteins in gene silencing in a rat hepatoma. *J Biol Chem* **277**, 16048-16058 (2002).

71. Okano, M., Bell, D.W., Haber, D.A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257 (1999).

72. Okano, M. & Li, E. Genetic analyses of DNA methyltransferase genes in mouse model system. *J Nutr* **132**, 2462S-2465S (2002).

73. Dejosez, M. et al. Ronin is essential for embryogenesis and the pluripotency of mouse embryonic stem cells. *Cell* **133**, 1162-1174 (2008).

74. van den Boom, V. et al. UTF1 is a chromatin-associated protein involved in ES cell differentiation. *J Cell Biol* **178**, 913-924 (2007).

75.	Coppo, P. et al. Constitutive and specific activation of STAT3 by BCR-ABL in embryonic stem cells. *Oncogene* **22**, 4102-4110 (2003).

76.	Ernst, M., Novak, U., Nicholson, S.E., Layton, J.E. & Dunn, A.R. The carboxyl-terminal domains of gp130-related cytokine receptors are necessary for suppressing embryonic stem cell differentiation. Involvement of STAT3. *J Biol Chem* **274**, 9729-9737 (1999).

77.	Matsuda, T. et al. STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *EMBO J* **18**, 4261-4269 (1999).

78.	Raz, R., Lee, C.K., Cannizzaro, L.A., d'Eustachio, P. & Levy, D.E. Essential role of STAT3 for embryonic stem cell pluripotency. *Proc Natl Acad Sci U S A* **96**, 2846-2851 (1999).

79.	Schuringa, J.J., van der Schaaf, S., Vellenga, E., Eggen, B.J. & Kruijer, W. LIF-induced STAT3 signaling in murine versus human embryonal carcinoma (EC) cells. *Exp Cell Res* **274**, 119-129 (2002).

80.	Qi, X. et al. BMP4 supports self-renewal of embryonic stem cells by inhibiting mitogen-activated protein kinase pathways. *Proc Natl Acad Sci U S A* **101**, 6027-6032 (2004).

81.	Domogatskaya, A., Rodin, S., Boutaud, A. & Tryggvason, K. Laminin-511 but not -332, -111, or -411 enables mouse embryonic stem cell self-renewal in vitro. *Stem Cells* **26**, 2800-2809 (2008).

82.	Bendall, S.C. et al. IGF and FGF cooperatively establish the regulatory stem cell niche of pluripotent human cells in vitro. *Nature* **448**, 1015-1021 (2007).

83.	Orsulic, S., Huber, O., Aberle, H., Arnold, S. & Kemler, R. E-cadherin binding prevents beta-catenin nuclear localization and beta-catenin/LEF-1-mediated transactivation. *J Cell Sci* **112 (Pt 8)**, 1237-1245 (1999).

84.	Watanabe, S. et al. Activation of Akt signaling is sufficient to maintain pluripotency in mouse and primate embryonic stem cells. *Oncogene* **25**, 2697-2707 (2006).

85.	Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676 (2006).

86.   Takahashi, K. et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861-872 (2007).

87.   Yu, J. et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917-1920 (2007).

88.   Henion, P.D. & Weston, J.A. Retinoic acid selectively promotes the survival and proliferation of neurogenic precursors in cultured neural crest cell populations. *Dev Biol* **161**, 243-250 (1994).

89.   Kelley, M.W., Turner, J.K. & Reh, T.A. Retinoic acid promotes differentiation of photoreceptors in vitro. *Development* **120**, 2091-2102 (1994).

90.   Bain, G., Ray, W.J., Yao, M. & Gottlieb, D.I. Retinoic acid promotes neural and represses mesodermal gene expression in mouse embryonic stem cells in culture. *Biochem Biophys Res Commun* **223**, 691-694 (1996).

91.   Ray, W.J. & Gottlieb, D.I. Regulation of protein abundance in pluripotent cells undergoing commitment to the neural lineage. *J Cell Physiol* **168**, 264-275 (1996).

92.   Schlett, K. & Madarasz, E. Retinoic acid induced neural differentiation in a neuroectodermal cell line immortalized by p53 deficiency. *J Neurosci Res* **47**, 405-415 (1997).

93.   Jones-Villeneuve, E.M., Rudnicki, M.A., Harris, J.F. & McBurney, M.W. Retinoic acid-induced neural differentiation of embryonal carcinoma cells. *Mol Cell Biol* **3**, 2271-2279 (1983).

94.   Husmann, M., Gorgen, I., Weisgerber, C. & Bitter-Suermann, D. Up-regulation of embryonic NCAM in an EC cell line by retinoic acid. *Dev Biol* **136**, 194-200 (1989).

95.   Aizawa, T., Haga, S. & Yoshikawa, K. Neural differentiation-associated generation of microglia-like phagocytes in murine embryonal carcinoma cell line. *Brain Res Dev Brain Res* **59**, 89-97 (1991).

96.   Dinsmore, J., Ratliff, J., Jacoby, D., Wunderlich, M. & Lindberg, C. Embryonic stem cells as a model for studying regulation of cellular differentiation. *Theriogenology* **49**, 145-151 (1998).

97.   Przyborski, S.A., Morton, I.E., Wood, A. & Andrews, P.W. Developmental regulation of neurogenesis in the pluripotent human embryonal carcinoma cell line NTERA-2. *Eur J Neurosci* **12**, 3521-3528 (2000).

98. Guan, K., Chang, H., Rolletschek, A. & Wobus, A.M. Embryonic stem cell-derived neurogenesis. Retinoic acid induction and lineage selection of neuronal cells. *Cell Tissue Res* **305**, 171-176 (2001).

99. Guo, X. et al. Proteomic characterization of early-stage differentiation of mouse embryonic stem cells into neural cells induced by all-trans retinoic acid in vitro. *Electrophoresis* **22**, 3067-3075 (2001).

100. Duggan, D.J., Bittner, M., Chen, Y., Meltzer, P. & Trent, J.M. Expression profiling using cDNA microarrays. *Nat Genet* **21**, 10-14 (1999).

101. Takahashi, K., Okita, K., Nakagawa, M. & Yamanaka, S. Induction of pluripotent stem cells from fibroblast cultures. *Nat Protoc* **2**, 3081-3089 (2007).

102. Kiel, M.J., Yilmaz, O.H., Iwashita, T., Terhorst, C. & Morrison, S.J. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109-1121 (2005).

103. Lopez-Ferrer, D., Ramos-Fernandez, A., Martinez-Bartolome, S., Garcia-Ruiz, P. & Vazquez, J. Quantitative proteomics using 16O/18O labeling and linear ion trap mass spectrometry. *Proteomics* **6 Suppl 1**, S4-11 (2006).

104. Graumann, J. et al. Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins. *Mol Cell Proteomics* **7**, 672-683 (2008).

105. Liu, H., Lin, D. & Yates, J.R., 3rd Multidimensional separations for protein/peptide analysis in the post-genomic era. *Biotechniques* **32**, 898, 900, 902 passim (2002).

106. Aiba, K. et al. Defining a developmental path to neural fate by global expression profiling of mouse embryonic stem cells and adult neural stem/progenitor cells. *Stem Cells* **24**, 889-895 (2006).

107. Van de Putte, T. et al. Mice with a homozygous gene trap vector insertion in mgcRacGAP die during pre-implantation development. *Mech Dev* **102**, 33-44 (2001).

108. Yamada, T., Kurosaki, T. & Hikida, M. Essential roles of mgcRacGAP in multilineage differentiation and survival of murine hematopoietic cells. *Biochem Biophys Res Commun* **372**, 941-946 (2008).

109. Arar, C., Ott, M.O., Toure, A. & Gacon, G. Structure and expression of murine mgcRacGAP: its developmental regulation suggests a role for the Rac/MgcRacGAP signalling pathway in neurogenesis. *Biochem J* **343 Pt 1**, 225-230 (1999).

110. Sucov, H.M. & Evans, R.M. Retinoic acid and retinoic acid receptors in development. *Mol Neurobiol* **10**, 169-184 (1995).

111. Sampath, P. et al. A hierarchical network controls protein translation during murine embryonic stem cell self-renewal and differentiation. *Cell Stem Cell* **2**, 448-460 (2008).

112. Hutchins, A.P. & Robson, P. Unraveling the human embryonic stem cell phosphoproteome. *Cell Stem Cell* **5**, 126-128 (2009).

113. Swaney, D.L., Wenger, C.D., Thomson, J.A. & Coon, J.J. Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci U S A* **106**, 995-1000 (2009).

114. Washburn, M.P., Wolters, D. & Yates, J.R., 3rd Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**, 242-247 (2001).

115. Wolters, D.A., Washburn, M.P. & Yates, J.R., 3rd An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* **73**, 5683-5690 (2001).

116. McDonald, W.H. & Yates, J.R., 3rd Shotgun proteomics and biomarker discovery. *Dis Markers* **18**, 99-105 (2002).

117. Washburn, M.P., Ulaszek, R., Deciu, C., Schieltz, D.M. & Yates, J.R., 3rd Analysis of quantitative proteomic data generated via multidimensional protein identification technology. *Anal Chem* **74**, 1650-1657 (2002).

118. Wu, C.C., MacCoss, M.J., Howell, K.E. & Yates, J.R., 3rd A method for the comprehensive proteomic analysis of membrane proteins. *Nat Biotechnol* **21**, 532-538 (2003).

119. Graumann, J. et al. Applicability of tandem affinity purification MudPIT to pathway proteomics in yeast. *Mol Cell Proteomics* **3**, 226-237 (2004).

120. Abdel-Rahman, B., Fiddler, M., Rappolee, D. & Pergament, E. Expression of transcription regulating genes in human preimplantation embryos. *Hum Reprod* **10**, 2787-2792 (1995).

121. Rathjen, J. et al. Formation of a primitive ectoderm like cell population, EPL cells, from ES cells in response to biologically derived factors. *J Cell Sci* **112 (Pt 5)**, 601-612 (1999).

122. Fuhrmann, G. et al. Mouse germline restriction of Oct4 expression by germ cell nuclear factor. *Dev Cell* **1**, 377-387 (2001).

123. Tada, M., Takahama, Y., Abe, K., Nakatsuji, N. & Tada, T. Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. *Curr Biol* **11**, 1553-1558 (2001).

124. Tan, S.M., Wang, S.T., Hentze, H. & Droge, P. A UTF1-based selection system for stable homogeneously pluripotent human embryonic stem cell cultures. *Nucleic Acids Res* **35**, e118 (2007).

125. Nishimoto, M., Fukushima, A., Okuda, A. & Muramatsu, M. The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. *Mol Cell Biol* **19**, 5453-5465 (1999).

126. Fukushima, A., Nishimoto, M., Okuda, A. & Muramatsu, M. Carboxy-terminally truncated form of a coactivator UTF1 stimulates transcription from a variety of gene promoters through the TATA Box. *Biochem Biophys Res Commun* **258**, 519-523 (1999).

127. Fukushima, A. et al. Characterization of functional domains of an embryonic stem cell coactivator UTF1 which are conserved and essential for potentiation of ATF-2 activity. *J Biol Chem* **273**, 25840-25849 (1998).

128. Okuda, A. et al. UTF1, a novel transcriptional coactivator expressed in pluripotent embryonic stem cells and extra-embryonic cells. *EMBO J* **17**, 2019-2032 (1998).

129. Hattori, N. et al. Preference of DNA methyltransferases for CpG islands in mouse embryonic stem cells. *Genome Res* **14**, 1733-1740 (2004).

130. Richards, M., Tan, S.P., Tan, J.H., Chan, W.K. & Bongso, A. The transcriptome profile of human embryonic stem cells as defined by SAGE. *Stem Cells* **22**, 51-64 (2004).

131. Pannell, D. et al. Retrovirus vector silencing is de novo methylase independent and marked by a repressive histone code. *EMBO J* **19**, 5884-5894 (2000).

132. Feraud, O., Cao, Y. & Vittet, D. Embryonic stem cell-derived embryoid bodies development in collagen gels recapitulates sprouting angiogenesis. *Lab Invest* **81**, 1669-1681 (2001).

133. Iwai, N., Kitajima, K., Sakai, K., Kimura, T. & Nakano, T. Alteration of cell adhesion and cell cycle properties of ES cells by an inducible dominant interfering Myb mutant. *Oncogene* **20**, 1425-1434 (2001).

134. Prozialeck, W.C. & Lamar, P.C. Cadmium (Cd2+) disrupts E-cadherin-dependent cell-cell junctions in MDCK cells. *In Vitro Cell Dev Biol Anim* **33**, 516-526 (1997).

135. Larue, L. et al. A role for cadherins in tissue formation. *Development* **122**, 3185-3194 (1996).

136. Hermiston, M.L., Wong, M.H. & Gordon, J.I. Forced expression of E-cadherin in the mouse intestinal epithelium slows cell migration and provides evidence for nonautonomous regulation of cell fate in a self-renewing system. *Genes Dev* **10**, 985-996 (1996).

137. Rohwedel, J., Horak, V., Hebrok, M., Fuchtbauer, E.M. & Wobus, A.M. M-twist expression inhibits mouse embryonic stem cell-derived myogenic differentiation in vitro. *Exp Cell Res* **220**, 92-100 (1995).

138. Burdsal, C.A., Damsky, C.H. & Pedersen, R.A. The role of E-cadherin and integrins in mesoderm differentiation and migration at the mammalian primitive streak. *Development* **118**, 829-844 (1993).

139. Mummery, C.L., Feyen, A., Freund, E. & Shen, S. Characteristics of embryonic stem cell differentiation: a comparison with two embryonal carcinoma cell lines. *Cell Differ Dev* **30**, 195-206 (1990).

140. Cole, M.F., Johnstone, S.E., Newman, J.J., Kagey, M.H. & Young, R.A. Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. *Genes Dev* **22**, 746-755 (2008).

141. Flagiello, D. et al. Assignment of the genes for cellular retinoic acid binding protein 1 (CRABP1) and 2 (CRABP2) to human chromosome band 15q24 and 1q21.3, respectively, by in situ hybridization. *Cytogenet Cell Genet* **76**, 17-18 (1997).

142. Faraonio, R., Galdieri, M. & Colantuoni, V. Cellular retinoic-acid-binding-protein and retinol-binding-protein mRNA expression in the cells of the rat seminiferous tubules and their regulation by retinoids. *Eur J Biochem* **211**, 835-842 (1993).

143. Bucco, R.A. et al. Regulation and localization of cellular retinol-binding protein, retinol-binding protein, cellular retinoic acid-binding protein (CRABP), and CRABP II in the uterus of the pseudopregnant rat. *Endocrinology* **137**, 3111-3122 (1996).

144. Boerman, M.H. & Napoli, J.L. Cellular retinol-binding protein-supported retinoic acid synthesis. Relative roles of microsomes and cytosol. *J Biol Chem* **271**, 5610-5616 (1996).

145. Siddiqui, N.A., Thomas, E.J., Dunlop, W. & Redfern, C.P. Retinoic acid receptors and retinoid binding proteins in endometrial adenocarcinoma: differential expression of cellular retinoid binding proteins in endometrioid tumours. *Int J Cancer* **64**, 253-263 (1995).

146. Berkovitz, B.K. & Maden, M. The distribution of cellular retinoic acid-binding protein I (CRABPI) and cellular retinol-binding protein I (CRBPI) during molar tooth development and eruption in the rat. *Connect Tissue Res* **32**, 191-199 (1995).

147. Ylikoski, J., Pirvola, U. & Eriksson, U. Cellular retinol-binding protein type I is prominently and differentially expressed in the sensory epithelium of the rat cochlea and vestibular organs. *J Comp Neurol* **349**, 596-602 (1994).

148. Siddiqui, N.A., Loughney, A., Thomas, E.J., Dunlop, W. & Redfern, C.P. Cellular retinoid binding proteins and nuclear retinoic acid receptors in endometrial epithelial cells. *Hum Reprod* **9**, 1410-1416 (1994).

149. Detrick, R.J., Dickey, D. & Kintner, C.R. The effects of N-cadherin misexpression on morphogenesis in Xenopus embryos. *Neuron* **4**, 493-506 (1990).

150. Balsamo, J. & Lilien, J. N-cadherin is stably associated with and is an acceptor for a cell surface N-acetylgalactosaminylphosphotransferase. *J Biol Chem* **265**, 2923-2928 (1990).

151. Duband, J.L. Role of adhesion molecules in the genesis of the peripheral nervous system in avians. *J Physiol (Paris)* **84**, 88-94 (1990).

152. Miyatani, S. et al. Neural cadherin: role in selective cell-cell adhesion. *Science* **245**, 631-635 (1989).

153. Rutishauser, U. N-cadherin: a cell adhesion molecule in neural development. *Trends Neurosci* **12**, 275-276 (1989).

154. Brackenbury, R. Expression of neural cell adhesion molecules in normal and pathologic tissues. *Ann N Y Acad Sci* **540**, 39-46 (1988).

155. Vlachakis, N., Choe, S.K. & Sagerstrom, C.G. Meis3 synergizes with Pbx4 and Hoxb1b in promoting hindbrain fates in the zebrafish. *Development* **128**, 1299-1312 (2001).

156. Davenne, M. et al. Hoxa2 and Hoxb2 control dorsoventral patterns of neuronal development in the rostral hindbrain. *Neuron* **22**, 677-691 (1999).

157. Van Maele-Fabry, G., Clotman, F., Gofflot, F., Bosschaert, J. & Picard, J.J. Postimplantation mouse embryos cultured in vitro. Assessment with whole-mount immunostaining and in situ hybridization. *Int J Dev Biol* **41**, 365-374 (1997).

158. Faiella, A., Zappavigna, V., Mavilio, F. & Boncinelli, E. Inhibition of retinoic acid-induced activation of 3' human HOXB genes by antisense oligonucleotides affects sequential activation of genes located upstream in the four HOX clusters. *Proc Natl Acad Sci U S A* **91**, 5335-5339 (1994).

159. Sham, M.H. et al. The zinc finger gene Krox20 regulates HoxB2 (Hox2.8) during hindbrain segmentation. *Cell* **72**, 183-196 (1993).

160. Dennis, G., Jr. et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**, P3 (2003).

161. Huang da, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).

162. Kawashima, T. et al. A Rac GTPase-activating protein, MgcRacGAP, is a nuclear localizing signal-containing nuclear chaperone in the activation of STAT transcription factors. *Mol Cell Biol* **29**, 1796-1813 (2009).

163. Toure, A. et al. Phosphoregulation of MgcRacGAP in mitosis involves Aurora B and Cdk1 protein kinases and the PP2A phosphatase. *FEBS Lett* **582**, 1182-1188 (2008).

164. Zhao, W.M. & Fang, G. MgcRacGAP controls the assembly of the contractile ring and the initiation of cytokinesis. *Proc Natl Acad Sci U S A* **102**, 13158-13163 (2005).

165. Oceguera-Yanez, F. et al. Ect2 and MgcRacGAP regulate the activation and function of Cdc42 in mitosis. *J Cell Biol* **168**, 221-232 (2005).

166. Ban, R., Irino, Y., Fukami, K. & Tanaka, H. Human mitotic spindle-associated protein PRC1 inhibits MgcRacGAP activity toward Cdc42 during the metaphase. *J Biol Chem* **279**, 16394-16402 (2004).

167. Yamada, T., Hikida, M. & Kurosaki, T. Regulation of cytokinesis by mgcRacGAP in B lymphocytes is independent of GAP activity. *Exp Cell Res* **312**, 3517-3525 (2006).

168. Betel D, W.M., Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. *Nucleic Acids Res.* **36**, D149-153 (2008).

169. Kim, E.J., Kho, J.H., Kang, M.R. & Um, S.J. Active regulator of SIRT1 cooperates with SIRT1 and facilitates suppression of p53 activity. *Mol Cell* **28**, 277-290 (2007).

170. Dai, M.S., Sun, X.X. & Lu, H. Aberrant expression of nucleostemin activates p53 and induces cell cycle arrest via inhibition of MDM2. *Mol Cell Biol* **28**, 4365-4376 (2008).

171. Meng, L., Lin, T. & Tsai, R.Y. Nucleoplasmic mobilization of nucleostemin stabilizes MDM2 and promotes G2-M progression and cell survival. *J Cell Sci* **121**, 4037-4046 (2008).

172. Pestov, D.G., Strezoska, Z. & Lau, L.F. Evidence of p53-dependent cross-talk between ribosome biogenesis and the cell cycle: effects of nucleolar protein Bop1 on G(1)/S transition. *Mol Cell Biol* **21**, 4246-4255 (2001).

173. Takagi, M., Sueishi, M., Saiwaki, T., Kametaka, A. & Yoneda, Y. A novel nucleolar protein, NIFK, interacts with the forkhead associated domain of Ki-67 antigen in mitosis. *J Biol Chem* **276**, 25386-25391 (2001).

174. Motoi, N. et al. Identification and characterization of nucleoplasmin 3 as a histone-binding protein in embryonic stem cells. *Dev Growth Differ* **50**, 307-320 (2008).

175. Johmura, Y., Osada, S., Nishizuka, M. & Imagawa, M. FAD24, a regulator of adipogenesis, is required for the regulation of DNA replication in cell proliferation. *Biol Pharm Bull* **31**, 1092-1095 (2008).

176. Johmura, Y., Osada, S., Nishizuka, M. & Imagawa, M. FAD24 acts in concert with histone acetyltransferase HBO1 to promote adipogenesis by controlling DNA replication. *J Biol Chem* **283**, 2265-2274 (2008).

177. Johmura, Y., Suzuki, M., Osada, S., Nishizuka, M. & Imagawa, M. FAD24, a regulator of adipogenesis and DNA replication, inhibits H-RAS-mediated transformation by repressing NF-kappaB activity. *Biochem Biophys Res Commun* **369**, 464-470 (2008).

178. Kirwan, M. & Dokal, I. Dyskeratosis congenita, stem cells and telomeres. *Biochim Biophys Acta* **1792**, 371-379 (2009).

179. Kondoh, H. et al. A high glycolytic flux supports the proliferative potential of murine embryonic stem cells. *Antioxid Redox Signal* **9**, 293-299 (2007).

180. Lopez-Lazaro, M. Role of oxygen in cancer: looking beyond hypoxia. *Anticancer Agents Med Chem* **9**, 517-525 (2009).

181. Samper, E. et al. Increase in mitochondrial biogenesis, oxidative stress, and glycolysis in murine lymphomas. *Free Radic Biol Med* **46**, 387-396 (2009).

182. Waki, A. et al. Dynamic changes in glucose metabolism accompanying the expression of the neural phenotype after differentiation in PC12 cells. *Brain Res* **894**, 88-94 (2001).

183. Mehrabian, Z., Liu, L.I., Fiskum, G., Rapoport, S.I. & Chandrasekaran, K. Regulation of mitochondrial gene expression by energy demand in neural cells. *J Neurochem* **93**, 850-860 (2005).

184. Vander Heiden, M.G., Cantley, L.C. & Thompson, C.B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029-1033 (2009).

185. Bohnsack, B.L., Lai, L., Dolle, P. & Hirschi, K.K. Signaling hierarchy downstream of retinoic acid that independently regulates vascular remodeling and endothelial cell proliferation. *Genes Dev* **18**, 1345-1358 (2004).

186. Schubert, M., Holland, N.D., Laudet, V. & Holland, L.Z. A retinoic acid-Hox hierarchy controls both anterior/posterior patterning and neuronal specification in the developing central nervous system of the cephalochordate amphioxus. *Dev Biol* **296**, 190-202 (2006).

187. Moustakas, A., Souchelnytskyi, S. & Heldin, C.H. Smad regulation in TGF-beta signal transduction. *J Cell Sci* **114**, 4359-4369 (2001).

188. Bischoff, J.R. & Plowman, G.D. The Aurora/Ipl1p kinase family: regulators of chromosome segregation and cytokinesis. *Trends Cell Biol* **9**, 454-459 (1999).

189. Jang, C.Y., Coppinger, J.A., Seki, A., Yates, J.R., 3rd & Fang, G. Plk1 and Aurora A regulate the depolymerase activity and the cellular localization of Kif2a. *J Cell Sci* **122**, 1334-1341 (2009).

190. Yu, X. et al. Chfr is required for tumor suppression and Aurora A regulation. *Nat Genet* **37**, 401-406 (2005).

191. Tomita, M. et al. Overexpression of Aurora A by loss of CHFR gene expression increases the growth and survival of HTLV-1-infected T cells through enhanced NF-kappaB activity. *Int J Cancer* **124**, 2607-2615 (2009).

192. Wada, N. et al. Overexpression of the mitotic spindle assembly checkpoint genes hBUB1, hBUBR1 and hMAD2 in thyroid carcinomas with aggressive nature. *Anticancer Res* **28**, 139-144 (2008).

193. Gao, F. et al. hBub1 negatively regulates p53 mediated early cell death upon mitotic checkpoint activation. *Cancer Biol Ther* **8**, 548-556 (2009).

194. Gao, F. et al. hBub1 deficiency triggers a novel p53 mediated early apoptotic checkpoint pathway in mitotic spindle damaged cells. *Cancer Biol Ther* **8**, 627-635 (2009).

195. Zhao, C., Gong, L., Li, W. & Chen, L. Overexpression of Plk1 promotes malignant progress in human esophageal squamous cell carcinoma. *J Cancer Res Clin Oncol* (2009).

196. Petronczki, M., Lenart, P. & Peters, J.M. Polo on the Rise-from Mitotic Entry to Cytokinesis with Plk1. *Dev Cell* **14**, 646-659 (2008).

197. Morrison, D.K. The 14-3-3 proteins: integrators of diverse signaling cues that impact cell fate and cancer development. *Trends Cell Biol* **19**, 16-23 (2009).

198. Obsilova, V., Silhan, J., Boura, E., Teisinger, J. & Obsil, T. 14-3-3 proteins: a family of versatile molecular regulators. *Physiol Res* **57 Suppl 3**, S11-21 (2008).

199. Fuentealba, L.C. et al. Integrating patterning signals: Wnt/GSK3 regulates the duration of the BMP/Smad1 signal. *Cell* **131**, 980-993 (2007).

200. Bone, H.K. et al. Involvement of GSK-3 in regulation of murine embryonic stem cell self-renewal revealed by a series of bisindolylmaleimides. *Chem Biol* **16**, 15-27 (2009).

201. Ragab, A. & Travers, A. HMG-D and histone H1 alter the local accessibility of nucleosomal DNA. *Nucleic Acids Res* **31**, 7083-7089 (2003).

202. Zlatanova, J., Caiafa, P. & Van Holde, K. Linker histone binding and displacement: versatile mechanism for transcriptional regulation. *FASEB J* **14**, 1697-1704 (2000).

203. Jerzmanowski, A. SWI/SNF chromatin remodeling and linker histones in plants. *Biochim Biophys Acta* **1769**, 330-345 (2007).

204. Martens, J.A. & Winston, F. Recent advances in understanding chromatin remodeling by Swi/Snf complexes. *Curr Opin Genet Dev* **13**, 136-142 (2003).

205. Xu, R.H. et al. Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. *Nat Methods* **2**, 185-190 (2005).

206. Burdon, T., Smith, A. & Savatier, P. Signalling, cell cycle and pluripotency in embryonic stem cells. *Trends Cell Biol* **12**, 432-438 (2002).

207. Armstrong, L. et al. The role of PI3K/AKT, MAPK/ERK and NFkappabeta signalling in the maintenance of human embryonic stem cell pluripotency and viability highlighted by transcriptional profiling and functional analysis. *Hum Mol Genet* **15**, 1894-1913 (2006).

208. Storm, M.P. et al. Regulation of Nanog expression by phosphoinositide 3-kinase-dependent signaling in murine embryonic stem cells. *J Biol Chem* **282**, 6265-6273 (2007).

209. Nagaoka, M. et al. E-cadherin-coated plates maintain pluripotent ES cells without colony formation. *PLoS ONE* **1**, e15 (2006).

210. Kishi, M. et al. Requirement of Sox2-mediated signaling for differentiation of early Xenopus neuroectoderm. *Development* **127**, 791-800 (2000).

211. Elling, U., Klasen, C., Eisenberger, T., Anlag, K. & Treier, M. Murine inner cell mass-derived lineages depend on Sall4 function. *Proc Natl Acad Sci U S A* **103**, 16319-16324 (2006).

212. Wu, Q. et al. Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. *J Biol Chem* **281**, 24090-24094 (2006).

213. Zhang, J. et al. Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. *Nat Cell Biol* **8**, 1114-1123 (2006).

214. Anton, R., Kestler, H.A. & Kuhl, M. Beta-catenin signaling contributes to stemness and regulates early differentiation in murine embryonic stem cells. *FEBS Lett* **581**, 5247-5254 (2007).

215. Okumura-Nakanishi, S., Saito, M., Niwa, H. & Ishikawa, F. Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. *J Biol Chem* **280**, 5307-5317 (2005).

216. Tomioka, M. et al. Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. *Nucleic Acids Res* **30**, 3202-3213 (2002).

217. Boer, B. et al. Differential activity of the FGF-4 enhancer in F9 and P19 embryonal carcinoma cells. *J Cell Physiol* **208**, 97-108 (2006).

218. Nishimoto, M. et al. Oct-3/4 maintains the proliferative embryonic stem cell state via specific binding to a variant octamer sequence in the regulatory region of the UTF1 locus. *Mol Cell Biol* **25**, 5084-5094 (2005).

219. Zhao, Y. et al. Two supporting factors greatly improve the efficiency of human iPSC generation. *Cell Stem Cell* **3**, 475-479 (2008).

220. Ibrahim, E.C., Schaal, T.D., Hertel, K.J., Reed, R. & Maniatis, T. Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers. *Proc Natl Acad Sci U S A* **102**, 5002-5007 (2005).

221. Trembley, J.H., Hu, D., Slaughter, C.A., Lahti, J.M. & Kidd, V.J. Casein kinase 2 interacts with cyclin-dependent kinase 11 (CDK11) in vivo and phosphorylates both the RNA polymerase II carboxyl-terminal domain and CDK11 in vitro. *J Biol Chem* **278**, 2265-2270 (2003).

222. Trembley, J.H. et al. Activation of pre-mRNA splicing by human RNPS1 is regulated by CK2 phosphorylation. *Mol Cell Biol* **25**, 1446-1457 (2005).

223. Dorr, J., Kartarius, S., Gotz, C. & Montenarh, M. Contribution of the individual subunits of protein kinase CK2 and of hPrp3p to the splicing process. *Mol Cell Biochem* **316**, 187-193 (2008).

224. Lehnert, S., Gotz, C., Kartarius, S., Schafer, B. & Montenarh, M. Protein kinase CK2 interacts with the splicing factor hPrp3p. *Oncogene* **27**, 2390-2400 (2008).

225. Duncan, P.I., Stojdl, D.F., Marius, R.M. & Bell, J.C. In vivo regulation of alternative pre-mRNA splicing by the Clk1 protein kinase. *Mol Cell Biol* **17**, 5996-6001 (1997).

226. Kojima, T., Zama, T., Wada, K., Onogi, H. & Hagiwara, M. Cloning of human PRP4 reveals interaction with Clk1. *J Biol Chem* **276**, 32247-32256 (2001).

227. Umehara, H. et al. Effect of cisplatin treatment on speckled distribution of a serine/arginine-rich nuclear protein CROP/Luc7A. *Biochem Biophys Res Commun* **301**, 324-329 (2003).

228. Schwertz, H. et al. Signal-dependent splicing of tissue factor pre-mRNA modulates the thrombogenicity of human platelets. *J Exp Med* **203**, 2433-2440 (2006).

229. Nowak, D.G. et al. Expression of pro- and anti-angiogenic isoforms of VEGF is differentially regulated by splicing and growth factors. *J Cell Sci* **121**, 3487-3495 (2008).

230. Prasad, J., Colwill, K., Pawson, T. & Manley, J.L. The protein kinase Clk/Sty directly modulates SR protein activity: both hyper- and hypophosphorylation inhibit splicing. *Mol Cell Biol* **19**, 6991-7000 (1999).

231. Chen, X. et al. Tra2betal regulates P19 neuronal differentiation and the splicing of FGF-2R and GluR-B minigenes. *Cell Biol Int* **28**, 791-799 (2004).

232. Calarco, J.A. et al. Regulation of vertebrate nervous system alternative splicing and development by an SR-related protein. *Cell* **138**, 898-910 (2009).

233. Muraki, M. et al. Manipulation of alternative splicing by a newly developed inhibitor of Clks. *J Biol Chem* **279**, 24246-24254 (2004).