QUANTITATIVE PROTEOMICS: METHODOLOGIES AND COMPARATIVE ANALYSIS OF DIFFERENTIATING MURINE EMBRYONIC STEM CELLS

by

PEGGI M.ANGEL

(Under the direction of Ron Orlando)

ABSTRACT

Embryonic stem cells are derived from the inner cell mass of the preimplantation embryo and are defined by properties of pluripotency and the ability to remain undifferentiated while proliferating in cell culture. When these cells are provoked towards spontaneous or directed differentiation, they provide investigative tools for studying the dynamic cellular mechanisms that occur during early development. While there are many reports of the transcriptional profiling of differentiating embryonic stem cells, there is little information on the translational events that occur during differentiation. In this work, we explore and develop methodologies for producing a quantitative description of translational events. We conclude by providing a comparative proteomic description of differentiating embryonic stem cells, applying the observed protein expression changes as an *in vitro* model of early embryogenesis.

First, we investigate quantitation of a proteomic population using stable isotopic labeling (SIL) methods. Trypsin commonly is used to facilitate incorporation of two ^{18}O atoms into the C-termini of the peptide population. We examine methods for preventing loss of the stable isotopes, a phenomenon that will hinder quantitative efforts. Removal of trypsin from the solution was found to be the best way to prevent loss of the label.

In a related observation, we found that trypsin facilitated ^{18}O incorporation has the potential to greatly increase the false discovery rate of sites of N-linked glycosylation. We demonstrate that this process occurs on an N-glycosylated standard protein and explore ways to prevent the problem. This

problem was best solved by removal of trypsin from solution prior to removal of the N-linked glycosylate from the protein.

We develop an alternative method for SIL using urea, a chaotrope used to denature and dissolve protein pellets. This method was comparable to other stable isotopic labeling strategies; however, the addition of the tag to C-terminal lysine residues decreased the intensity of the precursor ion during MALDI-TOF analysis. This method may prove useful for samples that require a strong chaotrope prior to proteolysis.

Last, we use spectral counts, a simple label-free method, to comparatively describe the protein population of differentiating embryonic stem cells. A total of 1498 proteins were identified, 22% of which were predicted and hypothetical proteins. Regulated protein comprised 15% of the population. We discuss the identified protein expression comparative to known embryogenetic events.

INDEX WORDS: $^{16}O/^{18}O$ labeling, N-linked glycosylation site mapping, stable isotopic labeling, proteomics, embryonic stem cell

QUANTITATIVE PROTEOMICS: METHODOLOGIES AND COMPARATIVE ANALYSIS OF DIFFERENTIATING MURINE EMBRYONIC STEM CELLS

by

Peggi M. Angel

B.S. Georgia Southern University, 1999

A Dissertation Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2007

QUANTITATIVE PROTEOMICS: METHODOLOGIES AND COMPARATIVE ANALYSIS OF DIFFERENTIATING MURINE EMBRYONIC STEM CELLS

by

PEGGI M. ANGEL

Major Professor:	Ron Orlando
Committee:	Jonathan Amster
	Lance Wells

Electronic version approved:
Maureen Grasso
Dean of the Graduate School
University of Georgia
May 2007

DEDICATION

To my beloved husband, my best friend and true love, William Hugh Angel: You have listened to every talk and numerous research ideas all with lively interest and feedback. I can never hope to express how grateful I am for your support in this venture.

To my sister, Pamela Kay Bass: My best friend and commuting companion.

To my mother and father: Thank you for instilling in me the merits of hard work, dedication, and perseverance.

ACKNOWLEDGEMENTS

I would like to express my gratitude and appreciation for the opportunity to work with the many great scientists at the Complex Carbohydrate Research Center at the University of Georgia. Each one of you has been in some way a role model for me during my graduate career. I profoundly appreciate the time given to answering questions and the freedom to use the equipment in your laboratories. I would like to especially express my gratitude to Ron Orlando, for the patient discussions and guidance you have provided. I deeply appreciate the invaluable exposure to scientists and science that you have provided for me. I would also like to thank Carl Bergman, who has been a friend and a mentor to me. I am grateful for your open door policy and open-ended access to your laboratory. Last, I acknowledge specific people: Punit Shah, James Atwood, Fernanda Ludolf, Gerardo Gutierrez-Sanchez, Lin Lin, Lei Cheng, Art Nuccio, Nicholas Seyfried, Karen Howard, Kumar Kolli, Todd Mize, Tasneem Bahrainwala, and Tracy Andacht. Each of you has touched my life in some way. I give thanks to all of you for the fun, the laughter and the help that was given during this wonderful experience.

TABLE OF CONTENTS

PAGE

CHAPTER 1

INTRODUCTION

Embryonic stem cells are pluripotent, self-renewing cells derived from the inner cell mass of the preimplantation blastocyst.[1] These cells have the ability to propagate indefinitely in culture, providing investigators with an unlimited supply of cells with which to model disease and developmental progression. A primary use for these cells is the investigation of early development or embryogenesis. Prior to the ability to culture embryonic stem cells, studies involving mammalian embryogenesis were performed using murine knockouts and *in situ hybridization* to explore the importance of a only few genes at a time. The ability to culture of embryonic stem cells has allowed researchers to use these cells as an *in vitro* model of embryogenesis, permitting global investigation of transcriptional events. Genetic expression of differentiating ES cells has been compared to *in vivo* embryogenetic activity, thus correlating the *in vitro* and *in vivo* models of embryogenesis [2-5]. A further layer of cellular dynamics that has not been thoroughly explored is the translational events that occur during embryogenesis. Previous studies have focused on the expression of a few proteins important to genetic studies, or have attempted to characterize the protein complement of embryonic stem cells as a singular population.[6-8] In this work, our goal is to quantitatively and qualitatively describe the protein expression of differentiating embryonic stem cells as an *in vitro* model of embryogenesis. A report that details quantitative and comprehensive protein expression of embryogenesis, even in an *in vitro* model, will provide an informative landscape for future developmental studies. However, describing both qualitative and quantitative aspects of a complex protein population is a complicated and challenging task, as will be discussed in the literature review.

In chapter 3, we begin our investigation into quantitative proteomics by exploring one of the more conventional methods of this field, stable isotopic labeling by incorporation of two ^{18}O atoms into the C-termini of peptides.[9] In this method, the application of the two ^{18}O onto the C-termini is enzyme facilitated, primarily by trypsin, a serine protease specific to lysine and arginine residues. One problem associated with this field is the post-label exchange of one or both of the ^{18}O atoms with ^{16}O, a phenomenon that obstructs quantitative efforts. We explore the reasons behind loss of the stable isotopes and discuss several methods for prevention of isotopic back exchange.

In chapter 4, we discuss the implications of trypsin use in an ^{18}O based N-linked glycosylation site mapping workflow. Here, our studies led us to the realization that trypsin has the potential to greatly increase the false discovery rate of N-linked site identification. We present a short manuscript detailing the problem and discussing several methods for avoiding this potential pitfall.

Chapter 5 describes the development of a new stable isotopic labeling method. This method uses urea to attach a carbamyl group containing either a light or a heavy stable isotope to primary amines after digestion of the protein population. Urea is a strong chaotrope commonly used in proteomics, especially for solubilizing difficult to dissolve samples. The development and application of the technique is illustrated on standard protein populations. A discussion on the advantages and disadvantages of this labeling method concludes this chapter.

In the final chapter, chapter 6, the simple and label-free method of spectral counting is used to quantify the protein expression of embryonic stem cell, early primitive ectoderm-like cells and embryoid bodies. Discussions include a survey of identified proteins, a description of the control set used to evaluate the normalization procedure, and a comparison of the protein population to reported protein expression occurring during embryogenesis.

REFERENCES

1. Evans, M. J.; Kaufman, M. H. *Nature* 1981, 292, 154-156.
2. Rathjen, J. *Journal of Cell Science* 1999, 112, 601-612.
3. Gardner, R. L. *Journal of Cell Science, Supplement* 1988, 10, 16.
4. Leahy, A.; Xiong, J. W.; Kuhnert, F.; Stuhlmann, H. *Journal of Experimental Zoology* 1999, 284, (1), 67-81.
5. Pelton, T. A.; Sharma, S.; Schulz, T. C.; Rathjen, J.; Rathjen, P. D., T. *Journal of Cell Science* 2002, 115, (2), 329-339.
6. Elliott, S. T.; Crider, D. G.; Graham, C. P.; Boheler, K. R.; Van Eyk, J. E. *Proteomics* 2004, 4, 3813-3832.
7. Nagano, K.; Taoka, M.; Yamauchi, Y.; Itagaki, C.; Shinkawa, T.; Nunomura, K.; Okamura, N.; Takahashi, N.; Izumi, T.; Isobe, T. *Proteomics* 2005, 5, 1346-1361.
8. Van Hoof, D.; Passier, R.; Ward-Van Oostwaard, D.; Pinkse, M. W. H.; Heck, A. J. R.; Mummery, C. L.; Krijgsveld, J. *Molecular & Cellular Proteomics* 2006, 5, (7), 1261.
9. Schnolzer, M.; Jedrzejewski, P.; Lehmann, W. D. *Electrophoresis* 1996, 17, (5), 945-53.

CHAPTER 2

LITERATURE REVIEW

Mouse embryonic stem cells: Tools for exploring our developmental origins.

Fertilization of the mammalian egg combines the genetic code from two individuals and initiates cellular processes that, if successful, will lead to the development of a mature and functional organism. The critical time period from fertilization of the egg until representation of all major organs is referred to as embryogenesis. In terms of embryogenesis, the mouse is the most studied mammalian species. *Mus musculus* is used to study embryogenesis comparative to humans for both ethical and technical reasons. The mouse is a very good, although not exact model for human development. This species is genetically comparative to humans; a rough draft of the mouse genome completed in 2002 revealed that the mouse and human genome each seem to contain about 30,000 protein coding genes.[2] Approximately 80% of all mouse genes were identified with a single orthologue in the human genome, while the proportion of mouse genes without any homologue detectable in the human genome was less than 1%.[2]

Early development of the mouse is well characterized.[3-6] Fertilization leads to the formation of a blastocyst at 4.75 days post coitum (d.p.c.) consisting of a collection of ~25 cells called the inner cell mass surrounded by a trophoblast shell. The appearance of the blastocystic structure is nearly concomitant with implantation into the uterine wall. The collection of cells composing the inner cell mass continues to proliferate and differentiate, forming the hypoblast and the epiblast. The trophoblast shell and the hypoblast differentiate, forming the extra embryonic tissues of the developing embryo. The pluripotent epiblast continues to proliferate and by 6.0 d.p.c is a pool of approximately 600 cells called primitive ectoderm. At 6.5 d.p.c., the primitive ectoderm population undergoes a tremendous burst of activity. A columnar extension of differentiating cells called the primitive streak extends from the epiblast. Precursors of the primordial germ layers compose specific portions of the primitive streak. Ectodermal precursors are located on the tip and center of the primitive streak, whilst mesodermal cells form the sides of the primitive streak. Embryonic endoderm arises posterior to the extending primitive streak. The precursor cells of the three primary germ layers continue the course of massive migration and differentiation initiated during extension of the primitive streak. These cells will follow genetically fated paths, ultimately resulting in the concise formation of tissues and then organs of the developing organism.

Investigation of early embryogenetic events that transform the spherical cluster of the inner cell mass to a complex and multifunctional organism may be studied *in vitro* using embryonic stem cells as a model for early embryogenesis. Advances in cell culture have produced three cell populations originating from the mouse species that are representative of early embryogenesis. The embryonic stem cell is derived from the inner cell mass of the preimplantation blastocyst and thus represents that population of undefined cells.[1, 7, 8] Embryonic stem cells are defined by the ability to remain undifferentiated while propagating indefinitely in culture.[7] This allows a virtually unlimited supply of cells as compared to the small collection of cells composing the inner cell mass. Embryonic stem cells are also pluripotent, able to differentiate into almost any tissue type, characteristic of this property of the inner cell mass.[9] Mouse embryonic stem cells may be cultured to yield a second pluripotent population, termed early primitive ectoderm-like cells (EPL).[10] The genetic and morphological profiles of these cells have been correlated to the primitive ectoderm of the embryo.[11, 12] Therefore, these cells may be used as models representative of the primitive ectoderm, equivalent of a time point preceding differentiation into the three primary germ layers. Currently, there is not a human model equivalent of the primitive ectoderm, although attempts are being made to develop this cell type.[13] Spontaneous differentiation of either ES or EPL cells produces embryoid bodies.[12, 14] Embryoid bodies are heterogeneous clusters of all three tissue types and may be used to represent the differentiated cellular state. Altogether, the mouse embryonic stem cell, the early primitive ectoderm-like cell and the embryoid body are useful tools towards the modeling of the cellular dynamics occurring during early embryogenesis.

The genetic mechanisms of differentiating ES cells continue to be well investigated, especially in regards to the pluripotent nature of the embryonic stem cells.[14-20] Genetic events surrounding *in vivo* embryogenesis have been compared to embryonic stem cells, thus correlating the *in vivo* and *in vitro* models at a genetic level.[12, 20-22] While the transcriptional activities surrounding early embryogenesis continue to be reported on, studies describing the comprehensive expression of the translational events taking place during this time period are limited at best. A specific reason for an investigation into the deeper layer of cellular mechanisms driving translational events is that protein abundance does not always

correlate with RNA abundance, the product of transcriptional events. In the past, positive connections between RNA and protein levels have been a controversial topic, with investigators trying to establish a correlation between comprehensive genetic expression and protein abundance.[23-26] However, a recent study suggests that correspondence between protein and RNA is probably function specific rather than there being a global connection between the two regulatory events.[27] Discordant RNA and protein expression has also been suggested to be an artifact of post-translational modifications.[26] In either case, it is clear that investigations between functional gene expression and protein abundance levels must be performed for a full understanding of cellular regulation.

Currently, there are only three studies that depict the global protein expression of embryonic stem cells.[28-30] These studies are limited to characterizing protein expression of the murine embryonic stem cell[28, 29], or comparing the embryonic stem cell to the embryoid body to obtain information on embryonic stem cell proteins.[30] Thus far, there has not been a study that specifically attempts to detail or quantitate large-scale protein expression occurring during early embryogenesis. Therefore, a qualitative and quantitative description of global protein expression occurring during embryogenesis will facilitate a further layer of understanding to the cellular mechanisms that regulate mammalian embryogenesis.

Comprehensive proteomic profiling of embryonic stem cells.

To date, there are two methods that have been used to detail comprehensive protein expression of embryonic stem cells. These methods are two dimensional gel electrophoresis[28] and liquid chromatography coupled to mass spectrometry (LC-MS/MS).[29, 30] These techniques represent the standard approaches utilized when describing large scale protein populations.[31] Each of these techniques has specific advantages and disadvantages. Herein, we describe the general approach to each of these methods. Each approach is followed by a description of the protein expression of embryonic stem cells reported using a particular method, allowing for methodological comparisons specific to the topic as well as detailing the fundamental portions of a descriptive proteomic experiment.

Two dimensional gel electrophoresis has proven a useful tool for comprehensive protein profiling since the technique was first described in 1975.[32, 33] A typical two dimensional gel electrophoresis begins

by denaturing the protein pellet, reducing the cysteine bridges and alkylating the cysteine residues. Proteins are then separated by isoelectric point using an immobilized pH gradient gel (IPG) and combined application of a current. Proteins migrate according to charge, stopping only when at the isoelectric point where net charge = zero. In the second dimension, the proteins are separated according to size by SDS PAGE (Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis). Proteins may be visualized using specialized software by staining the gel after both separations are complete, or by fluorescently labeling the denatured and reduced proteins prior to separation.[33] After visualization, further specialized software is used to select spots for excision. The proteins in the spots are then trypsin digested. Mass spectrometry, typically MALDI-TOF, is used to identify the protein digest by exact masses of the peptide fingerprint. Additional information on peptide and protein identification may be gleaned using MS/MS techniques such as MALDI-TOF/TOF, or electrospray ionization coupled to MS/MS.[33] Peptide mass fingerprints or fragments from MS/MS are compared to a theoretical database using publicly or commercially available search algorithms.[34, 35] These automated search algorithms compare experimentally observed peptide fingerprints or fragments to theoretical fragments produced by tryptic digest of a theoretical protein database pertaining to the organism of interest. The primary advantage to 2D gel electrophoresis technology is that the technique adds to the confidence of identification through the additionally acquired information of experimentally determined molecular weight and pI of protein. Another advantage is the ability to easily visualize protein expression and protein modification. Disadvantages lie in the limited effectiveness of the method at separating proteins at a low molecular weight (Mw<15kDa) or a high molecular weight (Mw> 200 kDa), as well as a limited dynamic range and sensitivity (4 orders of magnitude, ~50,000-100,000 copies of protein per cell).[33, 36] The chief cause for the decrease in effectiveness during separation of high molecular weight species is the limited mobility of these hydrophobic proteins during the separation processes.[37]

For embryonic stem cells, a two dimensional gel electrophoresis strategy was used to construct a database of proteins expressed by R1 embryonic stem cells, with the purpose to build a database of proteins expressed by embryonic stem cells.[28] Protein loads in the range of 500 to 2000 micrograms were

loaded onto IPG strips varying from pH 3-10, pH 4-7, and pH 6-11, and afterwards separated by SDS-PAGE as described previously. A total of 700 protein spots were digested and identified by peptide mass fingerprint using a MALDI-TOF calibrated to <150 ppm. Spots that could not be identified by peptide mass fingerprinting were loaded onto a nanoflow HPLC coupled to an API QSTAR pulsar mass spectrometer. Mass spectrometry analysis led to the identification of 331 proteins (123 from pH 3-10 IPG , 147 from pH 4-7 IPG , and 61 from the pH 6-11 IPG), that were further grouped to yield 218 proteins. Identified proteins were categorized into functional groups including 1) DNA maintenance, translation, transcription, protein processing (~50%), 2) Cellular energetics (~15%), 3) structural (~10%), 4) Cell signaling (~15%), and 5) Other (~10%), which were proteins that did not fit into the other four categories. The proteins identified in this study represent the first attempt to catalog embryonic stem cell proteins. The study provides a demonstration of the more abundant proteins associated with murine embryonic stem cells, with the major limitation on identification appearing to be a result of limited ability to select defined spots due to streaking. Streaking of proteins indicates a change in the pI of the protein and could be a sign of either experimental or biological modifications.

Two strategies utilizing a liquid chromatography coupled to mass spectrometry approach were employed to investigate protein expression that may be related to embryonic stem cell pluripotency and proliferation.[29, 30] Like 2D gel electrophoresis, a classic LC-MS/MS technique for identification of large scale protein expression begins with denaturation of the protein, reduction of cysteine bridges, and alkylation of the exposed cysteine residue. Unlike 2DGE, the proteins are enzymatically digested before the separation techniques. In most cases, a trypsin digest is used; however, other enzymes as well as chemical hydrolysis may be employed for especially hydrophobic samples.[38] Trypsin offers an advantage over other enzymes in that it places the basic residue at the C-terminus, influencing fragmentation to produce more C-terminal y-ions. This leads to a higher mass series of ions that is set apart from any internal fragmentation ion, and thus facilitates sequencing of the MS/MS spectra.[39] After digestion, peptides may be separated by conventional liquid chromatography techniques such as reverse phase, size exclusion, or ion exchange chromatography.[40] Since digestion of proteins converts the already complex

population into a more complex population, liquid chromatography techniques are coupled together to increase peak capacity. Selection of the coupled separation methods depends upon what population is being examined. In many proteomic studies, an ion exchange method (either strong cation or anion) is coupled to a reverse phase method that elutes directly into the mass spectrometer for sequencing of peptides.[41] Reverse phase chromatography is typically selected as the last separation method because the mobile phase modifiers for this separation are compatible with the mass spectrometry analysis. After fragmentation, peptides are sequenced using a search algorithm as described previously. One additional complexity to an LC-MS/MS approach is grouping peptides to proteins. Advances in bioinformatics have produced software that allows the user to statistically group peptides to proteins, providing a measure of confidence in the protein identification.[42, 43] Proteins may be further grouped into homology groups based upon the sequence of the identified peptides. The advantage to the use of LC-MS/MS over 2D-GE is that virtually any protein can be analyzed by enzymatic or chemical digestion into peptides, regardless of molecular weight or pI,. Second, separation techniques may be combined to enrich for a particular population. A drawback to the technique is that the complexity of the sample dictates that the peak capacity of the system is often over exceeded. Since the acquisition of MS/MS is data dependent, with precursors selected for MS/MS dependent upon when they elute from the column, and what other peaks elute at the same time, variations in data collection occur. Each technical replicate that is performed acquires more peptide identifications than previous replicates, with reportedly up to at least seven replicates before the baseline for increasing number of identification stabilizes.[44]

A two dimensional LC-MS/MS experiment was performed to characterize the proteomic expression of the E14-1 mouse embryonic stem cell line.[29] In this approach, an anion exchange column was used to separate digested soluble proteins in the first dimension of separation followed by online fractionation to a reverse phase column. The second dimension reverse phase culminated in elution into quadrupole time of flight mass spectrometer (QToF-2). Six technical replicates of the 2D LC-MS/MS strategy were performed. Mascot facilitated database searching was performed against human, rat and mouse databases (separately) in order to identify proteins that may have previously characterized protein

homologues in these species that are similar to mouse. In-house developed software was used to select peptides with a p-value > 0.05 (based upon MASCOT calculated probability scoring), remove redundant peptides, group peptides from all three databases to proteins, and assign sequence coverage to each protein. This analysis resulted in 6032 unique peptide mapping to 1985 proteins. After removal of proteins with interspecies redundancy, 1790 proteins remained. Each technical replicate was reported to result in approximately 20% more new identifications than the previous technical replicate. Proteins with isoelectric points ranging from 3.67 to 12.02 were identified. Molecular weights of identified proteins ranged from 5 kDa to 802 kDa. Proteins were identified that were reported to have abundancies greater than 10^4 copies per cell. Lowest levels of abundancy were not explored. Functional annotation of the 62% of the proteome (1117 proteins) was performed, with two of the primary functions being protein metabolism and cellular signaling. Interestingly, the cellular signaling category contained several proteins previously associated with a differentiated cellular state, implying that ES cells have ready mechanisms with which to respond to differentiation stimuli. This study also covered post translational modifications of proteins, independently searching the database for acetylation, phosphorylation, and pyroglutamination. The study identified 151 acetylated peptides and 67 phosphoproteins. This study shows some advantages over the previously discussed 2D gel electrophoresis study: 1) a wider range of protein molecular weight and pI were identified, 2) more proteins were identified, probably due to a more automated approach to data acquisition, and 3) peptides with modifications could be analyzed “on the fly”, at the same time as unmodified peptides. Drawbacks to this study appeared to be the complexity of assigning peptides to proteins particularly at a very high confidence level. Peptides were selected based upon the MASCOT calculated peptide probability of 95% for a particular search and were then grouped to proteins. In addition, it was not clear whether spectrum that had previously been identified were removed prior to searching for post translational modifications, presenting a chance that some spectra could have been associated with more than one peptide.

In the last approach that reports on murine embryonic stem cells, investigators used a one dimensional gel separation coupled with high resolution LC-MS/MS to compare HES-2 human and the

D3 murine embryonic stem cells to their differentiated counterparts in an attempt to isolate a stem cell specific population of proteins.[30] A total of 1.5 mg (each) of protein from four cells (human undifferentiated and differentiated and mouse differentiated and undifferentiated) was separated on an SDS-PAGE gel using previously described strategies. After separation by molecular weight, gel lanes were cut into 26 fractions, followed by dicing into smaller fractions. Proteins were digested following an in gel digestion protocol, using trypsin for enzymatic proteolysis. Peptides were extracted from the gel pieces and trapped on a 100 micron I.D. capillary C18 column packed in-house connected to a 50 micron x 15 centimeter C18 column. Peptides were eluted off the column into an LTQ-FT mass spectrometer. Only one technical replicate of each fraction was performed. The Mascot search algorithm was used to search against human and then mouse databases. The human and mouse database were reversed, producing a random database. Peptides identified in both databases were manually removed from the datasets. Only fully tryptic peptides as well as peptides with a score >25 were allowed to be grouped to proteins. The minimum protein score allowed was 60. Redundant proteins were removed from the list of acceptably identified proteins. This proteomic analysis marks the first attempt to quantitatively describe the differentiation process, although the manuscript focused on proteins that were unique or upregulated in the embryonic stem cell population. Peptide ratios were used for quantitation. Basically, the technique takes a ratio of the number of peptides from each differently expressed state to produce a fold change value. This value expression was further correlated by western blots performed on the differential cell states. Overall, the study yielded a total of 2295 murine protein identifications, with approximately 50% of the identifications shared between the embryonic cell state and embryoid bodies. Embryonic stem cell specific proteins accounted for 32% of the identifications, whilst embryoid body unique proteins accounted for 18% of the identifications. Since the primary purpose of the paper was to reveal embryonic stem cell specific proteins, only proteins that were unique or upregulated in embryonic stem cells were discussed in the manuscript. This pool of 888 mouse embryonic stem cell proteins contained 743 unique proteins and 145 proteins that were upregulated more than three fold over embryoid bodies. Western blotting techniques were used for the identification and regulation of 26 proteins that were also present in

the human embryonic stem cell pool. Of these proteins, 18 were recognized by antibodies during western blotting. Approximately half of the proteins investigated by western blotting had corresponding regulatory patterns. Aberrant expression values were reported as due to either denaturing of the protein during western blotting procedure or due to the presence of mixed colonies of cells. This study shows an ability to provide high quality, high throughput comprehensive proteomic coverage to embryonic stem cells, with rough correlations between different quantitative methods. A disadvantage to this technique is that it is still limited by the front end gel approach, possibly discriminating against very low MW, high MW and very basic species. The plethora of fractions produced from the gel run probably aided in providing additional peak capacity even if the proteins had precipitated or were not separated well.

In conclusion, three methods have been used to provide qualitative information on comprehensive protein expression for embryonic stem cells. Comparison of all three methods highlights the essential procedures for a modern proteomic experiment: 1) Sample preparation, 2) enzymatic digestion, 3) the use of robust and unbiased separation procedures, 4) mass spectrometry to identify peptides and proteins, 5) search algorithms for automated searching of acquired spectra, and 6) the use of bioinformatics software to provide an automated statistical approach for the identification of proteins and for grouping peptides to proteins. In the case of two dimensional gel electrophoresis, robotics and software are also required to select spots for analysis. Given the above requirements, the burden of the modern proteomics experiment now lies on the analysis of the identified proteins in attempting to put together functional pictures that are instructive of the biological topic.

Qualitatively, of the three methods selected for protein profiling of embryonic stem cells, the in gel digestion followed by LC-MS/MS provided the most identifications, probably due to a combination of the gel-fractionation technique and the high mass accuracy of the Fourier transform mass spectrometer. However, the laboratory used a specialized setup that allowed automated analysis of the large number of samples (104) produced in a single pass experiment. Therefore, it seems that a simple LC- MS/MS approach would be better suited for the comprehensive profiling of differentiating embryonic stem cells as this would allow automated analysis of a large number of samples as well as replicate analyses.

Clearly, from the survey of these embryonic stem cell investigations, a study entailing comprehensive proteomic profiling of differentiating embryonic stem cells has the capabilities to yield an informative depiction of the early events of embryogenesis.

Quantitative proteomics

Performing quantitative analysis on the multi-component and chemically complex proteomic population is a difficult and challenging task. Currently, there are two classes of strategies for obtaining quantitative information on a large-scale proteomic population. These strategies take either a stable isotopic labeling approach[45, 46] or a label-free approach in attempts to extract quantitive expression values from comparative proteomic populations. Each of these approaches utilizes a specific qualitative analyses strategy with a quantitative method to compare proteomes.

Stable isotopic labeling strategies use either enzymatic or chemical reactions to apply a light or a heavy stable isotopic label to biologically differential proteomes and combine the two proteomes in a one-to-one ratio for simultaneous analysis. This produces pairs of co-eluting peaks that differ only in a mass shift difference between the light and heavy applied stable isotopes. Relative quantitation is performed by taking a ratio of the peak area or peak heights of an eluting peptide pair from the MS or the MS/MS spectrum. Light or heavy stable isotopic pairs consist of ^{12}C-^{13}C, ^{14}N-^{15}N, ^{1}H-^{2}H, or ^{16}O-^{18}O.[45] These isotopes may be incorporated into a chemical compound in a variety of combinations, or may be applied singly to a chemical moiety on the protein or peptide. Ideally, the mass shift between the two species should be sufficient so that the light mass and the heavy mass do not have overlapping isotopic envelopes, as this can complicate quantification efforts. Additionally, the different labels should not cause a chromatographic shift, as differential elution may cause a difference in the ionization efficiency for the sample and thus affect quantitation of the sample.

Stable isotopic labeling schemes may be classified into two categories, labeling at the protein level[47-50] and labeling at the peptide level[51-55] [Figure 2.1]. Labeling at the protein level can be accomplished by metabolic labeling in cell culture[48, 49], or by labeling the complete protein before separations.[50] Metabolic labeling entails growing the cells in media that has been enriched for a particular

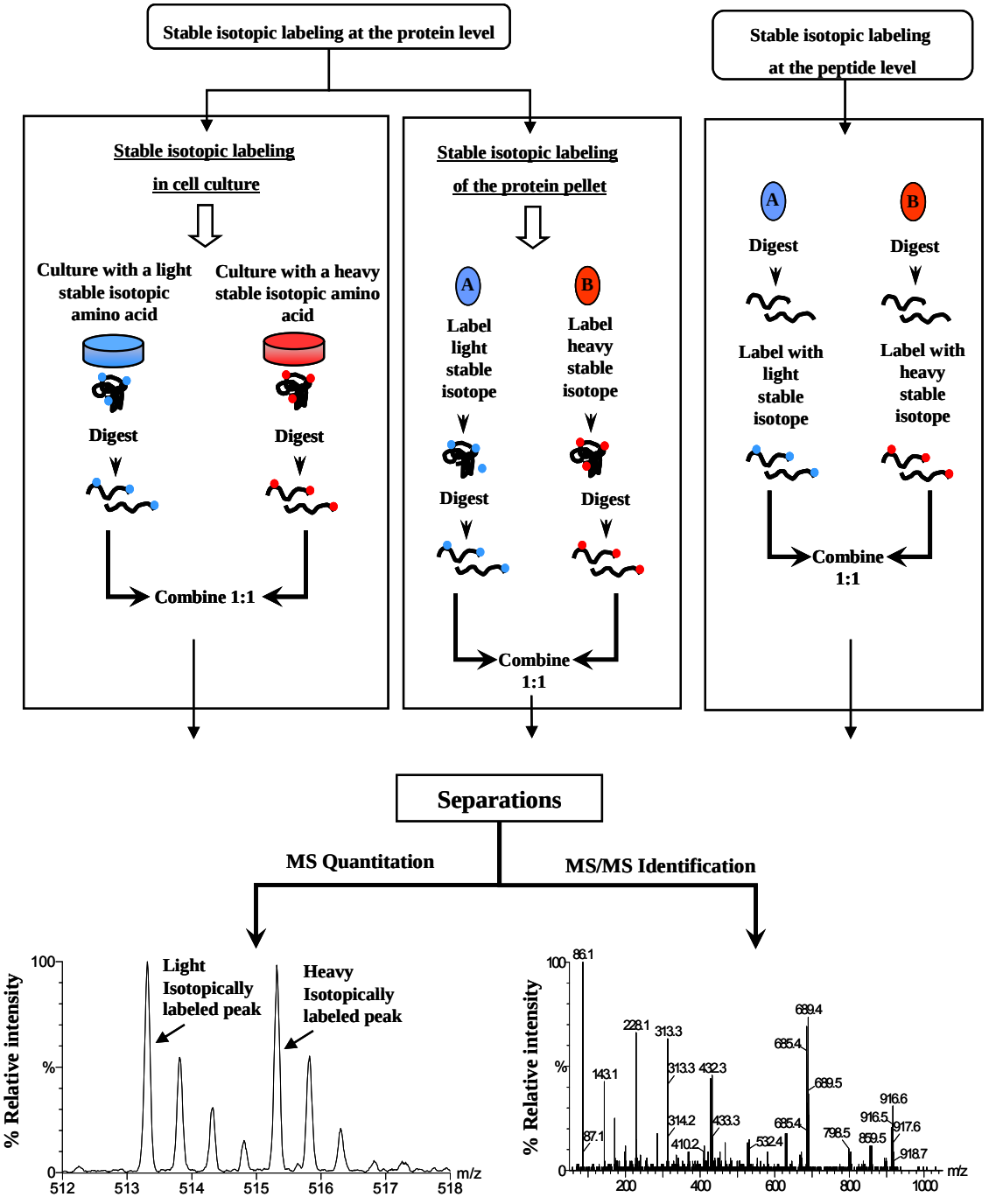


Figure 2.1: General schema for stable isotopic labeling of proteomic samples. All schema conclude with MS and MS/MS analysis where relative quantitation is performed by comparison of the peak areas or intensities.

stable isotope such as 13C, 15N or 2H[47, 48] or that contains isotopically labeled amino acids[49]. This produces consistently and nearly completely labeled proteins, unbiased towards type of protein, abundancy of protein, or post translational modifications. A major disadvantage to this technique is that is can only be applied to cells that are capable of being grown in culture. Therefore, methods to label at the protein and peptide level have been developed. An example of a strategy for labeling at the protein level involves the chemical application of N-nicotinoyloxy-succinimide (Nic-NHS) to primary amines of the protein prior to separations or digestion.[50] Here, the light isotopic Nic-NHS contains the naturally abundant isotopes of the chemical compound, while the heavy Nic-NHS contains four deuteriums to produce a 4 Da mass shift for every α –amine (N-terminus) or ε-amine (lysine residue). After labeling, proteins may be separated by gel-based methods and then digested, or may be directly digested and then separated using LC-MS/MS techniques. Disadvantages to labeling at the protein pellet level include a decrease in efficiency of tryptic digestion (all lysine residues are now blocked from proteolysis) and a potential change in the pI of the amine site. If gel-based separation methods are used, a change in the pI of protein will influence the migration of the protein and may lead to a decreased separation capability. Changing the pI of the amine site also affects the efficiency of ionization for mass spectrometry identification. For instance, a decrease in basicity of the amine site has been shown to limit the number of identifications that may be produced.[56]

The last general scheme for isotopic labeling involves applying the stable isotope to the peptide. Labeling at the peptide level generally targets chemical moieties in common with the entire peptide population, such as the C-terminus or the N-terminus. One of the earlier approaches for isotopic labeling of peptides from a global proteomic population used trypsin to enzymatically add two ^{18}O isotopes to the C-terminus of tryptic fragments during digestion.[51] Investigators performed this technique by simply digesting the proteins in ^{18}O enriched water. The tryptic mechanism utilizes the native water of the solution for hydrolysis, with one ^{18}O isotope being incorporated into the C-terminus of the peptide during cleavage from the polypeptide chain. Since the cleaved peptide may bind again to the tryptic active site, in highly enriched water a second ^{18}O atom is incorporated into the C-terminus.[57] This produces a 4 Da

shift for every tryptic peptide in the sample. Factors that influence amount of incorporation of ^{18}O include pH, temperature, hydrophobicity of the peptide, post translational modifications, and difference in association rates between lysines and arginines.[57-59] Isotopic labeling with ^{18}O has an advantage in the simplicity of the technique and the universal application to any tryptic population. However, due to the continued interaction of trypsin with the tryptic peptide, trypsin induced back exchange may occur and must be prevented for accurate quantitation. Labeling the C-terminus of peptides is an example of targeting chemical moieties in common with the majority of the population. Other techniques take this approach by targeting the N-terminus of the peptide, another universally present moiety in the peptide population. Labeling the N-terminus usually results in labeling of the ε-amine of the lysine residue. Labeling moieties that are in common with the majority of the peptide population has the advantage in that more peptides may be used for quantification of the whole protein.

In a label-free quantitative approach, a parallel work flow is employed for comparative analysis of each biological state. Like stable isotopic labeling techniques, a plethora of label-free approaches exist. Features of the eluting peptide peak such as normalized ion intensities,[60] spectral count,[44] mass, scan number and signal intensity,[61] or accurate mass plus retention time[62] have all been evaluated as a way to gain quantitative information from the peptide and protein populations. Other proteomic studies have employed aspects of the identified population such as percent peptide coverage[63] or peptide ratios[30] obtain semi-quantitative information between biological states. Of these techniques, spectral counts have become the most analytically well developed method for label-free comparative proteomics, standing up to rigorous statistical evaluation on both soluble and membrane sample preparations. [64-66]

Simply put, a spectrum count is the acquisition of one fragmentation event for a peptide. All MS/MS events for a sequenced peptide are summed together to produce a spectral count for that peptide. After clustering peptides to proteins, a spectral count at the protein level may be obtained. Initial studies on the quantitative aspects of accumulating spectral counts were performed by creating six calibration curves by spiking six standard proteins digests into a yeast proteome digest at an abundance range from 0.0417% to 4.17%.[44] The six standard proteins selected were albumin, ovalbumin, phosphorylase B,

trypsin inhibitor, carbonic anhydrase, and lysozyme, providing a good variation in size and characteristic features, such as post translational modification of the proteins. Spectral counts, number of peptides per protein and sequence coverage were all investigated in relation to protein abundancy and linearity. Correlation between sequence coverage and protein abundancy was poor, with a correlation coefficient ranging from 0.205 for lysozyme and 0.797 for carbonic anhydrase. Peptide counts, or number of peptides per protein, had a better correlation, with R^2 ranging from 0.819 for lysozyme to 0.9909 for carbonic anhydrase. However, spectral counts correlated linearly with protein abundance, having R^2 0.996 for phosphorylase B to 0.999 for trypsin inhibitor. Reproducibility was also examined. Reproducibility among three technical replicates for proteins that were spiked in at abundancies of 0.25% ranged from 62.4% (lysozyme) to 20.0% (trypsin inhibitor). At a concentration of 25%, reproducibility varied less, from 7.2% (lysozyme) to 14.4% for carbonic anhydrase. Thus the study showed that spectral counts correlate linearly with abundance over at least two orders of magnitude, and are very reproducible for higher concentration, while reproducibility varies more at lower concentrations. In another report, spectral counts versus signal intensity was investigated in regards to quantitation at both the protein level or at the peptide level using the *Porphyromonas gingivalis* proteome.[66] Protein level quantitation was performed using the total number of spectral counts or intensities from all peptides assigned to the protein in each run. Peptide level quantitation was performed by averaging the peptide spectral counts/signal intensities from two or three replicates or if only one replicate resulted in a peptide hit, reporting the value of the peptide. Spectral count methods were found to have better correlations than the use of signal intensities, hypothesized to be due to the fact the spectral counts were less noisy than signal intensities. Both methods performed poorly at the peptide quantitation level. This study concluded that spectral count at the protein level was the best approach to quantitative analysis due to a higher reproducibility at the protein level versus the peptide level. In regards to the application of spectral counts to samples, spectral counts have been used for quantification of a wide range of sample types including the a soluble preparation of aforementioned organisms, *Saccharomyces cerevisiae*[44] and *Porphyromonas gingivalis*,[66] human blood serum,[64] the membrane proteome of *S. cerevisiae*[67], and mouse tissue.[27]

Spectral counts have been compared to stable isotopic labeling.[68] In one study, investigators grew *S. cerevisiae* in rich and minimal media using ^{14}N and ^{15}N ammonium sulfate as the sole nitrogen source, thereby incorporating the stable isotopes into the organism.[68] After digestion, samples were combined in a 1:1 ratio and analyzed in quadruplicate by strong cation-reverse phase terminating in a linear ion trap mass spectrometer. Spectral counts and stable isotopic protein ratios were found to have a strong correlation, that was enhanced when only proteins with a ratio>1.5 were considered. Deviations from correlation between the two methods were also attributed to the wider dynamic range of spectral counts over stable isotopic labeling. While both methods correlated well with each other, it was concluded that spectral counts were more useful due in part to their larger dynamic range. The stable isotopic labeling method was concluded to be more effective at detecting changes at the peptide level, as well at appearing to be more sensitive in calculating less than two fold changes in protein expression.

In summary, the acquisition of a spectral count is an intrinsic part of a proteomic experiment, yielding not only the sequence of a peptide, but also a facile way to obtain quantitative information on protein regulation. Spectral counts have a linear dynamic range of at least two orders of magnitude, are reproducible, can accurately quantitate at fold changes of greater than approximately 2.0[68, 69] to 2.5[64], and have been demonstrated on a wide range of samples. Spectral counts correlate well with stable isotopic labeling, and are considered superior due to a wider dynamic range. In limited samples, stable isotopic labeling may be a better option since this strategy appears to be more accurate for single peptides. However, as a label free method, spectral counts have an additional advantage over a stable isotopic labeling strategy in that no extraneous sample handling is required, thus decreasing the potential to bias the comparative proteomes through experimental error. Due to all of the above mentioned merits, a spectral count method appears to be the one of the best approaches for obtaining both qualitative and quantitative information on the translational regulation of the embryonic stem cell, early primitive ectoderm-like cell and embryoid bodies.

Summary of objectives

As mentioned previously, there is a lack of studies examining the translational regulation of early embryogenetic events. Current technology in murine embryonic stem cell research has produced differentiated cells that appear to be genetically representative of pivotal time points occurring during early mammalian development. In addition, recent advancements in quantitative proteomic research have produced methods for global analysis of protein regulation. Therefore, it is now possible to investigate at a translational level the cellular dynamics that produce a multi-celled and complex functioning organism. In this work presented here, three manuscripts deal with investigation and development of quantitative proteomic methodologies. In the fourth and final manuscript, the proteomes of embryonic stem cells, early primitive ectoderm-like cells, and embryoid bodies are qualitatively and quantitatively compared. These studies will advance the field of quantitative proteomics as well as provide new information on the translation events of early embryogenesis.

REFERENCES

1. Waterston, R. H.; Lindblad-Toh, K.; Birney, E.; Rogers, J.; Abril, J. F.; Agarwal, P.; Agarwala, R.; Ainscough, R.; Alexandersson, M.; An, P. *Nature* 2002, 420, (6915), 520-562.

2. Snow, M. H. L. *Journal of Embryology and Experimental Morphology*, 1977; Vol. 42, pp 293-303.

3. Gardner, R. L. *Journal of Cell Science, Supplement* 1988, 10, 16.

4. Quinlan, G. A. *Mammalian Development*, Lonai, P., Ed. Oversease Publishers Association: Amsterdam, 1996; pp 1-27.

5. Zernicka-Goetz, M. *Nature Reviews Molecular Cell Biology* 2005, 6, (12), 919.
6. Evans, M. J.; Kaufman, M. H. *Nature* 1981, 292, 154-156.

7. Martin, G. R. *Proceedings of the National Academy of Sciences* 1981, 78, (12), 7634-7638.

8. Nagy, A.; Rossant, J.; Nagy, R.; Abramow-Newerly, W.; Roder, J. C. *Proceedings of the National Academy of Sciences* 1993, 90, (18), 8424-8428.

9. O'Shea, K. S. *Biology of Reproduction* 71, (6), 1755-1765.

10. Rathjen, J. *Journal of Cell Science* 1999, 112, 601-612.

11. Stead, E.; White, J.; Faast, R.; Conn, S.; Goldstone, S.; Rathjen, J.; Dhingra, U.; Rathjen, P.; Walker, D.; Dalton, S. *Oncogene* 2002, 21, (54), 8320-8333.

12. Pelton, T. A.; Sharma, S.; Schulz, T. C.; Rathjen, J.; Rathjen, P. D. *Journal of Cell Science* 2002, 115, (2), 329-339.

13. Darr, H.; Mayshar, Y.; Benvenisty, N. *Development* 2006, 133, (6), 1193-1201

14. Leahy, A.; Xiong, J. W.; Kuhnert, F.; Stuhlmann, H. *Journal of Experimental Zoology* 1999, 284, (1), 67-81.

15. Lars Palmqvist Clive H. Glover, L. H., Min Lu, Bolette Bossen, James M. Piret, R.Keith Humphiries, Cheryl D. Helgason. *Stem Cells* 2005, 23, 18.

16. Ginis, I.; Luo, Y.; Miura, T.; Thies, S.; Brandenberger, R.; Gerecht-Nir, S.; Amit, M.; Hoke, A.; Carpenter, M. K.; Itskovitz-Eldor, J. *Dev Biol* 2004, 269, (2), 360-80.

17. Ramalho-Santos, M.; Yoon, S.; Matsuzaki, Y.; Mulligan, R. C.; Melton, D. A. *Science* 2002, 298, (5593), 597-600.

18. Kelly, D. L.; Rizzino, A. *Molecular Reproduction and Development* 2000, 56, (2), 113-123.

19. Houbaviy, H. B.; Murray, M. F.; Sharp, P. A. *Dev. Cell* 2003, 5, 351-358.

20. Sharov, A. A.; Piao, Y.; Matoba, R.; Dudekula, D. B.; Qian, Y.; VanBuren, V.; Falco, G.; Martin, P. R.; Stagg, C. A.; Bassey, U. C. *PLoS Biol* 2003, 1, (3), E74.

21. Wang, Q. T.; Piotrowska, K.; Ciemerych, M. A.; Milenkovic, L.; Scott, M. P.; Davis, R. W.; Zernicka-Goetz, M. *Dev Cell* 2004, 6, (1), 133-44.

22. Hamatani, T.; Carter, M. G.; Sharov, A. A.; Ko, M. S. *Dev Cell* 2004, 6, (1), 117-31.

23. Gygi, S. P.; Rochon, Y.; Franza, B. R.; Aebersold, R. *Molecular and Cellular Biology* 1999, 19, (3), 1720-1730.

24. Chen, G.; Gharib, T. G.; Huang, C. C.; Taylor, J. M. G.; Misek, D. E.; Kardia, S. L. R.; Giordano, T. J.; Iannettoni, M. D.; Orringer, M. B.; Hanash, S. M. *Molecular & Cellular Proteomics* 2002, 1, (4), 304-313.

25. Corbin, R. W.; Paliy, O.; Yang, F.; Shabanowitz, J.; Platt, M.; Lyons, C. E.; Root, K.; McAuliffe, J.; Jordan, M. I.; Kustu, S. *Proceedings of the National Academy of Sciences* 2003, 100, (16), 9232-9237.

26. Mootha, V. K.; Bunkenborg, J.; Olsen, J. V.; Hjerrild, M.; Wisniewski, J. R.; Stahl, E.; Bolouri, M. S.; Ray, H. N.; Sihag, S.; Kamal, M. *Cell* 2003, 115, (5), 629-640.

27. Kislinger, T.; Cox, B.; Kannan, A.; Chung, C.; Hu, P.; Ignatchenko, A.; Scott, M. S.; Gramolini, A. O.; Morris, Q.; Hallett, M. T. *Cell* 2006, 125, (1), 173-186.

28. Elliott, S. T.; Crider, D. G.; Graham, C. P.; Boheler, K. R.; Van Eyk, J. E. *Proteomics* 2004, 4, 3813-3832.

29. Nagano, K.; Taoka, M.; Yamauchi, Y.; Itagaki, C.; Shinkawa, T.; Nunomura, K.; Okamura, N.; Takahashi, N.; Izumi, T.; Isobe, T. *Proteomics* 2005, 5, 1346-1361

.

30. Van Hoof, D.; Passier, R.; Ward-Van Oostwaard, D.; Pinkse, M. W. H.; Heck, A. J. R.; Mummery, C. L.; Krijgsveld, J., *Molecular & Cellular Proteomics* 2006, 5, (7), 1261.

31 Aebersold, R.; Mann, M. *Nature* 2003, 422, 198-207.

32. O'Farrell, P. H. *Journal of Biological Chemistry* 1975, 250, (10), 4007-4021.

33. Gorg, A.; Weiss, W.; Dunn, M. J. *Proteomics* 2004, 4, (12), 3665-85.

34. Eng, J. K.; McCormack, A. L.; Yates, J. R. *J. Am. Soc. Mass Spectrom* 1994, 5, (11), 976-989.

35. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* 1999, 20, (18), 3551-67.

36. Corthals, G. L.; Wasinger, V. C.; Hochstrasser, D. F.; Sanchez, J. C. *Electrophoresis* 2000, 21, (6), 1104-15.

37. Hochstrasser, D. F. *Electrophoresis* 1998, 19, 1501-1505.

38. Wu, C. C.; Yates, J. R. *Nature Biotechnology* 2003, 21, (3), 262-267.

39. Bieman, K. *Methods in Enzymology*, 1990; Vol. 193, pp 455-479.

40. Romijn, E. P.; Krijgsveld, J.; Heck, A. J. *J Chromatogr A* 2003, 1000, (1-2), 589-608.

41. Wang, H.; Hanash, S. *J Chromatogr B Analyt Technol Biomed Life Sci* 2003, 787, (1), 11-8.

42. Weatherly, D. B.; Atwood, J. A.; Minning, T. A.; Cavola, C.; Tarleton, R. L.; Orlando, R. *Molecular & Cellular Proteomics* 2005, 4, (6), 762-772.

43. Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. *Anal. Chem* 2003, 75, (17), 4646-4658.

44. Liu, H.; Sadygov, R. G.; Yates Iii, J. R. *Anal. Chem* 2004, 76, (14), 4193-4201.

45. Goshe, M. B.; Smith, R. D. *Curr Opin Biotechnol* 2003, 14, (1), 101-9.

46. Julka, S.; Regnier, F. *J. Proteome Res* 2004, 3, (3), 350-363.

47. Pasa-Tolic, L.; Jensen, P. K.; Anderson, G. A.; Lipton, M. S.; Peden, K. K.; Martinovic, S.; Tolic, N.; Bruce, J. E.; Smith, R. D. *J. Am. Chem. Soc* 1999, 121, 7949-7950.

48. Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T. *Proceedings of the National Academy of Sciences of the United States of America* 1999, 6591-6596.

49. Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. *Molecular & Cellular Proteomics* 2002, 1, (5), 376-386.

50. Schmidt, A.; Kellermann, J.; Lottspeich, F., *Proteomics* 2005, 5, (1), 4-15.

51. Schnolzer, M.; Jedrzejewski, P.; Lehmann, W. D. *Electrophoresis* 1996, 17, (5), 945-53.

52. Hsu, J. L.; Huang, S. Y.; Chow, N. H.; Chen, S. H. *Anal Chem* 2003, 75, (24), 6843-52.

53. Brancia, F. L.; Montgomery, H.; Tanaka, K.; Kumashiro, S. *Anal. Chem* 2004, 76, (10), 2748-2755.

54. Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nature Biotechnology* 1999, 17, 994-999.

55. Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S. *Molecular & Cellular Proteomics* 2004, 3, (12), 1154-1169.

56. Krause, E.; Wenschuh, H.; Jungblut, P. R. *Anal. Chem* 1999, 71, (19), 4160-4165.

57. Yao, X.; Afonso, C.; Fenselau, C. *J. Proteome Res* 2003, 2, 147-152.

58. Stewart, II; Thomson, T.; Figeys, D. *Rapid Communications in Mass Spectrometry* 2001, 15, (24), 2456-2465.

59. Staes, A.; Demol, H.; Van Damme, J.; Martens, L.; Vandekerckhove, J.; Gevaert, K., *J. Proteome Res* 2004, 3, 786-791.

60. Wang, W.; Zhou, H.; Lin, H.; Roy, S.; Shaler, T. A.; Hill, L. R.; Norton, S.; Kumar, P.; Anderle, M.; Becker, C. H. *Nat. Biotechnol* 1999, 17, 994-999.

61. Radulovic, D.; Jelveh, S.; Ryu, S.; Hamilton, T. G.; Foss, E.; Mao, Y.; Emili, A. S. *Molecular & Cellular Proteomics* 2004, 3, (10), 984-997.

62. Silva, J. C.; Denny, R.; Dorschel, C. A.; Gorenstein, M.; Kass, I. J.; Li, G. Z.; McKenna, T.; Nold, M. J.; Richardson, K.; Young, P. *Nat. Biotechnol* 1999, 17, 994-999.

63. Florens, L.; Washburn, M. P.; Raine, J. D.; Anthony, R. M.; Grainger, M.; Haynes, J. D.; Moch, J. K.; Muster, N.; Sacci, J. B.; Tabb, D. L. *Nature* 2002, 419, 520-526.

64. Old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G. *Molecular & Cellular Proteomics* 2005, 4, (10), 1487-1502.

65. Boris Zybailov, A. L. M., Mihaela E. Sardiu, Michael K. Coleman, Laurence Florens, Michael P. Washburn. *Journal of Proteome Research* 2006, 5, 9.

66. Xia, Q., Wang, Tiansong, Park, Yoonsuk, Lamont, Richard J., Hackett, Murray, *International Journal of Mass Spectrometry* 2007, 259, 105-116.

67. Zybailov, B.; Mosley, A. L.; Sardiu, M. E.; Coleman, M. K.; Florens, L.; Washburn, M. P. *Journal of Proteome Research* 2006, 5, (9), 2339-2347.

68. Zybailov, B.; Coleman, M. K.; Florens, L.; Washburn, M. P. *Anal. Chem* 2005, 77, (19), 6218-6224.

69. Hendrickson, E. L.; Xia, Q.; Wang, T.; Leigh, J. A.; Hackett, M. *The Analyst* 2006, 131, 7.

CHAPTER 3

TRYPSIN IS THE PRIMARY MECHANISM BY WHICH THE 18O LABEL IS LOST IN QUANTITATIVE PROTEOMIC STUDIES[1]

[1] Angel, Peggi M., Orlando, R. *Analytical Biochemistry*, 2006, 359: 26-34.
Reprinted here with permission from Elsevier.

ABSTRACT

Labeling with ^{18}O is currently one of the most commonly used methods for incorporating a stable isotopic label into samples for comparative proteomic studies. In this approach, isotopic labeling involves the enzymatic digestion, typically performed with trypsin, of a protein population in ^{18}O water, which incorporates the stable isotope into the C-termini of the newly formed peptides. Although trypsin is often used to facilitate isotopic incorporation after digestion, it is typically overlooked that this same mechanism can lead to isotopic loss even under conditions such as low pH where it is assumed that trypsin is inactive. To examine the role trypsin plays in isotopic loss, several experiments were performed on the rate of de-labeling under conditions relevant to multidimensional proteomic experiments. Results from these studies demonstrate that enzyme facilitated exchange of ^{18}O in the peptide with ^{16}O in the aqueous solvent was the major process by which the label is removed from the peptides, even under conditions of low pH and temperature where trypsin is thought to be inactive. This study brings the rapid, tryptic facilitated exchange to the attention of laboratories using this scheme in order to prevent inaccuracies in quantitative labeling due to loss of the isotopic label.

INTRODUCTION

Quantitative proteomics attempts to provide changes in protein expression between two biologically different states for purposes of disease research or as a complement to genomic research. The field of shotgun proteomics has developed numerous methods for correlating various features of the eluting peptides to the abundance levels of the proteins in the proteome. Label free methods using normalized ion intensities[1], spectral counts[2], mass, scan number and signal intensity[3], or accurate mass plus retention time[4] have all been used with success to link these aspects of eluting peptides to protein expression level for comparative quantitation. However, ion intensities and retention times are variable, limiting the accuracy of these approaches. The alternative strategy for relative quantitation involves the simultaneous analysis of isotopically labeled and unlabeled peptides. This approach has advantages over label free methods in that the heavy and light labeled peptide pairs elute together, allowing a direct comparison of relative abundances for that peptide and that two or more differential populations may be combined for a single analysis, greatly decreasing the time spent analyzing the samples.

Recent years have seen a broad array of strategies for introduction of the stable isotope to a particular population. The isotope may be introduced *in vivo* by metabolic incorporation of a stable isotopically labeled amino acid in cell culture (SILAC)[5], or by using media enriched for particular isotope(s) such as 13C, 15N or 2H.[6,7] These methods produce consistent incorporation of an isotope into a population. However, metabolic incorporation is not practical for differential comparison of whole animals. Methods to introduce a stable isotope *in vitro* have thus been developed. For example, isotope-coded affinity tags (ICAT)[8] chemically target specific amino acids, typically cysteine, in the peptide sequence for differential labeling. A potential limitation of strategies targeting specific amino acids, such as ICAT, is that only populations that contain the target residue can be selected for quantification purposes. Other *in vitro* approaches target functional groups common to all peptides, such as the primary

amines of the N-termini and/or the carboxyl groups of the C-termini.[9-14] This allows all observable peptides from each protein to be used for quantification purposes.

^{18}O labeling utilizing tryptic activity to label the C-termini of peptides with the stable isotope[9] is the most widely used universal labeling procedure for quantitative proteomics and has been applied to a wide variety of analyses, most notably including samples of limited amount[15] and membrane populations.[16] The resulting mass shift from ^{18}O incorporation does not alter the chromatographic separation or the ionization efficiency of the labeled peptides, thus providing a versatile, nondiscriminatory, global labeling system for relative quantitative proteomics. ^{18}O labeling for proteomics is different from other methods of isotopic incorporation that target universally present functional groups in that the mechanism is enzyme driven, allowing fast incorporation of the label without jeopardizing biologically derived modifications that may be of interest. The technique is performed by enzymatically digesting the proteins in ^{18}O enriched water using trypsin or another serine protease, thus producing a +4 Da shift with the incorporation of the two ^{18}O isotopes. Samples may also be labeled after digestion has occurred, which permits smaller volumes of ^{18}O water to be used while allowing conditions to be optimized separately for digestion and the labeling of a particular population.[17,18] Enzyme facilitated labeling after digestion is possible due to the cleaved peptide acting as a pseudo substrate to serine proteases (reactions III_F+IV_F/III_R+IV_R in Figure 3.1).[9,18,19] It is for this reason that trypsin,[9] chymotrypsin,[18] Lys-C,[9] and Glu-C[9,20] catalyze the addition of two ^{18}O atoms into the C terminus of the proteolytic peptides. It should be noted that other classes of proteases can lead to the incorporation of a single ^{18}O isotope, as has been reported using the metalloendopeptidase Lys-N.[10] However, trypsin is by far the most commonly used enzyme because it consistently produces uniform peptides that are amenable to interpretable fragmentation by tandem mass spectrometry.[21]

Comparative proteomics relies on minimal loss of the isotopic label for accurate relative quantitation. For ^{18}O labeling, the loss of the isotopic label has been reported via chemical reaction at low

pHs[9,22] as well as through the same enzyme facilitated mechanisms that were used to apply the isotopic label,[9, 18,19,22] which appears to be the faster process. The use of immobilized trypsin eliminates this process as it allows the trypsin to be removed from the sample after the labeling step.[23] Other approaches have minimized the loss of the isotopic label via the enzymatic route by employing conditions that deactivate trypsin, such as reducing and alkylating after labeling,[22] boiling the enzyme at digestive pH followed by lowering the pH[16] or the more common practice of inhibiting the tryptic mechanism by lowered pH of the sample.[9, 15,16, 19] These disabling approaches are based on the prediction that conditions which minimize the proteolytic rate of trypsin (reactions I - IV in Figure 3.1) will lead to a similar decrease in the rate of isotopic loss from enzymatic back exchange (reactions III_F+IV_F/III_R+IV_R in Figure 3.1).

Here we present a study demonstrating that trypsin is the main facilitator of loss of the ^{18}O label even at low pH and temperature, conditions that are often used in a proteomic study. Since ^{18}O isotopic labeling is currently one of the most utilized isotopic labeling techniques available for a wide range of *in vitro* samples types, it is our aim to bring the rapid, tryptic facilitated exchange to the attention of laboratories using this scheme in order to prevent inaccuracies in quantitative labeling due to loss of the isotopic label.

EXPERIMENTAL PROCEDURES

Materials

Alpha-casein, ammonium bicarbonate, trifluoroacetic acid, α-cyanohydroxy cinnamic acid (CHCA), L-1-Chloro-3-[4-tosyl-amido]-7-amino-2-heptanone-HCI (TLCK) were purchased from Sigma (St. Louis, MO). Isotopic water (95% ^{18}O) was supplied by Isotec (Miamisburg, OH). Urea, formic acid and acetonitrile were supplied by J.T. Baker (Phillipsburg NJ). Sequencing grade trypsin was purchased from Promega (Madison, WI). Immobilized trypsin beads were purchased from Pierce (Rockford, IL).

Digestion and labeling. A solution of approximately 50 pmole/μL α-casein in 50 mM ammonium bicarbonate was digested overnight at 37°C with trypsin using a 1:50 ratio of protease to substrate (w/w). For all cases, ^{18}O labeling was performed after this initial digest using the following protocol: A 60 μL portion of the digested protein was dried under vacuum for 30 min at <37°C. A 60 μL volume of 50 mM ammonium bicarbonate in ^{18}O enriched water, prepared by adding the appropriate amount of ammonium bicarbonate salt to the isotopically enriched water, was added to the samples, and labeling allowed to proceed overnight. No additional trypsin was needed as previous studies had shown that trypsin remained viable after the drying process had occurred.

Loss of label in the absence of trypsin, molecular weight cutoff filtration. A 400 μL portion of the 50 picomole/μL tryptically digested α-casein peptides was filtered with a molecular weight cutoff filter of 10,000 (Amicron; Beverly, MA) to remove trypsin. After filtration, the ^{18}O labeled peptides were diluted 1:15 with 5% formic acid to reach a pH of 2. A sample at a pH of 7 was prepared in the same fashion with dilution of the labeled peptides in natural abundance water. Back exchange was measured once every 10 min for the first hour, once an hour for 5 h after that, and then once every 24 h thereafter.

Loss of label in the absence of trypsin, immobilized beads. Exactly 20 μL of immobilized trypsin beads was added to a 50 μL solution of 50 picomole/μL solution of α-casein protein in natural abundance water. The sample was dried by vacuum centrifugation and reconstituted in 40 μL of 50 mM ammonium bicarbonate prepared in 95% $H_2{}^{18}O$ with 10 μL acetonitrile and incubated at room temperature overnight with shaking. After incubation, the sample was spun at 10,000 g for 3 min. A portion of the supernatant was removed and diluted 1:10 in either formic acid or water to attain the desired pH of either 2 or 7. Back exchange was tracked by 1:5 dilution in the MALDI matrix at time zero, 6 h after time zero, and every 24 h thereafter.

Loss of label, temperature dependent. For pH studies, 20 μL of the 50 picomole/μL labeled solution was diluted 1:10 with varying concentrations of formic acid to reach a pH of either 2 or 7. The

pH of each solution was checked using litmus paper (Colorphast, EMD Chemicals, Gibbstown, NJ). Back exchange was measured at room temperature (approximately 23˚C) and at 4 C every 10 min for the first 50 min, every 25 min for the next 75 min, and once 24 h after time zero. At the specified time, a portion of the sample was withdrawn and diluted 1:5 with MALDI matrix, and analyzed by MALDI-MS, as described below. For the experiment performed at 4° C, samples and matrix were kept on ice during the course of the investigation, with overnight storage on ice inside a refrigerator maintained at 2-8 °C.

Loss of label in the presence of high organic and low pH. The effects of lyophilization and reconstitution in strong cation exchange buffer were examined. A fully labeled solution of α-casein sample was diluted 1:1 in natural abundance water. The sample was instantly frozen within seconds after addition of the natural abundance water by suspending the sample in liquid nitrogen for 15 sec. The sample was immediately lyophilized. To prevent melting before sublimination of the sample, the sample was maintained on an alcohol/dry ice bath. Visual inspection confirmed that no melting of the samples occurred before the drying process had begun. After lyophilization, samples were reconstituted in 20% methanol, 0.1% formic acid, imitating a strong cation exchange separation. Back exchange was measured at room temperature (approximately 23˚C) every 10 min for the first 50 min, every 25 min for the next 75 min, and once 24 h after time zero. At the specified time, a portion of the sample was withdrawn and diluted 1:5 in MALDI matrix, and analyzed by MALDI-MS, as described below.

Loss of label, lowered pH with heat denaturation. The effects of simultaneous heat denaturation and lowering of pH versus isotopic loss were measured. The fully labeled solution of α-casein was adjusted to a pH of 2 using concentrated formic acid. The small amount of natural abundance water in the formic acid did not impact the study, since the relative change of label was being recorded. The solution was then placed in boiling water for 10 min. After cooling to room temperature, a fraction of this solution was diluted in natural abundance water to a final pH of 7. Back exchange was measured at room temperature (approximately 23˚C) every 10 min for the first 50 min, every 25 min for the next 75 min,

and once 24 h after time zero. At the specified time, a portion of the sample was withdrawn and diluted 1:5 in MALDI matrix, and analyzed by MALDI-MS, as described below.

Loss of label, lowered pH with inhibitor. L-1-Chloro-3-[4-tosyl-amido]-7-amino-2-heptanone-HCI (TLCK) was used to inhibit tryptic activity. This molecule is not stable at pHs greater that 7.5, therefore a stock solution of 1 mg/mL TLCK was prepared in 5% formic acid/natural abundance water. Exactly 1 µL of the 1 mg/mL TLCK stock solution was diluted 1:50 in 95% ^{18}O labeled water to a final concentration of 20 µg/mL TLCK. The pH of a labeled sample solution was lowered to below pH 6.5 by addition of 1 µL of 10% formic acid. TLCK was added in a ratio of 1:30 enzyme: inhibitor to a 50 µL portion of labeled α-casein sample. The labeled sample with TLCK was incubated overnight at room temperature. A 10 µL aliquot of the sample was diluted to 100 µL in 0.1% formic acid/natural abundance water to a measured pH of 2. Back exchange was measured at time zero, every 10 min for the first 50 min, followed by time points at 75 min and 100 min. A final measurement was taken at 24 h from time zero.

MALDI-MS. All samples were analyzed by MALDI-ToF on an ABI 4700 Proteomics Analyzer (Applied Biosystems, Applied Biosystems, Foster City, CA) equipped with an Nd:YAG laser operating at 355 nm with a 200 Hz laser rate. All samples were acquired in the positive ion reflector mod**e** with an acquisition mass range from 900 - 2000 m/z and a focus at 1500 m/z. Each spectrum was an accumulation of 1000 shots obtained with a laser setting of 3600, accelerating voltage of 20 kV, source chamber pressure of 6.0 x 10^{-8} torr, and a reflector chamber pressure of 2.0 x 10^{-8} torr. External calibration was performed using four standards des-arg^{1}-bradykinin (904.468), angiotensin I (1296.685), Glu^{1}-fibrinopeptide B (1570.678), and neurotensin (1672.92) (singly charged monoisotopic species denoted).

Samples for all measurements were mixed with α-cyanohydroxy cinnamic acid (CHCA) to a 1 picomole/µL concentration and spotted onto a MALDI target plate. Mixing with the MALDI matrix and spotting were performed in less than 1 min. Back exchange of a sample dried onto the MALDI target was

measured at time zero and after 24 h and found to be negligible. For all experiments, where possible, the same stock solution of labeled digest was used to minimize variances due to enzyme, peptide, and ^{18}O water concentration. Peptide mass fingerprinting was performed using Protein prospector (http://prospector.ucsf.edu) against the Swissprot databank with the settings enzyme- trypsin, species bos taurus, one missed cleavage, peptide tolerance 200 ppm, instrument setting MALDI-TOF. Calculation of isotopic contributions was performed using MS-Isotope (http://prospector.ucsf.edu), using the peptide sequence function for known peptides, and the averagine function to calculate isotopic contribution for unknown peptides.

RESULTS

Proteolysis with trypsin begins with association of the polypeptide chain with the enzyme by formation of a hydrogen bond between the peptide backbone and the enzyme[24] and subsequent locking of the charged residue onto the catalytic serine residue[25] (Fig. 3.1, reaction I). This step is followed by cleavage of the amide bond after arginine (Arg) and lysine (Lys) residues (Fig. 1, reaction II). During this process, the newly created peptide terminating with Lys/Arg becomes covalently attached to the enzyme. Nucleophilic attack by water on the peptide-enzyme complex liberates the peptide from the enzyme and results in the incorporation of a single oxygen atom from the solvent into the C-terminus of the newly formed peptide (Fig. 3.1, reactions III_F and IV_F). The Lys/Arg terminated peptide may again bind to the active site of the enzyme releasing one of the two equivalent C-terminal oxygen atoms (Fig. 3.1, reactions III_R and IV_R). Once again this peptide-enzyme complex dissociates after hydrolysis, incorporating another oxygen atom from the solvent into the C-terminus. Since Lys/Arg terminated peptides are in constant equilibrium with the peptide-enzyme complex, these two reactions (Fig. 3.1, reactions $III_F + IV_F$ and $III_R + IV_R$) occur repeatedly and explain why two ^{18}O labels are incorporated into the C-terminus of trypsin generated peptides when this process is performed in water highly enriched with the isotope. This equilibrium implies that the extent of isotopic incorporation on the peptides will proceed towards that of the solvent in which they are dissolved. In other words, ^{18}O containing peptides will lose their isotopic

labels, i.e., be de-labeled, when the isotopic content of the solvent is decreased. A similarly problematic, although rarely discussed, consequence of this equilibrium is that non-labeled peptides will incorporate ^{18}O (i.e., become labeled) when they are introduced to an isotopically enriched solvent. This situation can occur when a ^{16}O labeled sample dissolved in ^{16}O water is combined with an ^{18}O labeled sample in ^{18}O water. These concerns prompted us to evaluate the extent of isotopic scrambling resulting from tryptic activity under conditions applicable to a proteomic scheme.

Initial studies on the loss of isotopic label were performed with α-casein that was post digestively labeled and thus contained the maximum amount of double isotopic incorporation.[16-18] The extent of ^{18}O delabeling was determined by MALDI-TOF analysis. MALDI was selected for these studies because it allows a shorter time period from sample preparation to analysis than either ESI-MS or LC-MS, and thus offered the smallest possible window for loss of the isotope and a more accurate representation of the isotopic abundances of the peptides in solution. Prior to these studies, the extent of back exchange in the MALDI matrix was experimentally determined to be below the detection limits (data not shown), which is in agreement with a previous report.[19]

Early studies identified that the primary process limiting isotopic incorporation was the initial proteolytic cleavage (reactions I-IV_f).[18] The use of a post-digestive labeling strategy circumvents this issue.[17,18] In this latter approach, it is expected that the rate at which ^{18}O is incorporated will be similar to the rate of enzyme facilitated back exchange since both of these processes proceed via reactions III_f+IV_f/III_r+IV_r, i.e., peptides that are slow to label are also anticipated to experience slow back exchange. To account for variability introduced by differing rates of back exchange due to peptide length or composition,[18,19] and thus establish a general trend, five peptides with mass to charge ratios (*m/z*) of 971, 1134, 1271, 1384, and 1759 were followed under the various conditions described in subsequent paragraphs. An average of these five peptides was used to calculate the relative percent double label remaining at the selected time point. All isotopic incorporation percentages were calculated for the double

labeled peptides via dividing the peak area from the double labeled ion by the sum of peak areas from the un-labeled, single labeled and double labeled ions after adjusting these ion areas for the naturally occurring isotopes of the peptides analyzed. Differences in labeling efficiencies between experiments were accounted for by calculating the loss of the double labeled ion relative to the value obtained at time zero for each set of conditions. This allowed for direct comparisons of the extent of isotopic loss from each set of experiments. Loss of the double ^{18}O label was followed as the resulting mass shift (+4 Da) minimizes isotopic overlap and thus produces more reliable quantitation.

The conditions for this investigation were selected to imitate those that would be typically encountered in a proteomic study. In particular, a pH of 2 was intended to mimic a sample being re-suspended in an acidic solution prior to injection onto a reversed phase column, while a pH of 7 mimics combining samples immediately after digesting/labeling without adjustment of pH. Temperature was also taken into account, with the de-labeling process being investigated at both room temperature and at 4°C to account for cases where the sample is placed in an auto sampler with a cooling system. While these conditions do not represent every conceivable possibility, they do cover a wide spectrum of those typically used in proteomic studies and demonstrate general trends for trypsin catalyzed isotopic loss.

To set a benchmark for chemical exchange versus enzymatic exchange, the loss of the double ^{18}O label was first observed at acidic and neutral pH in the absence of trypsin while at room temperature. Two different samples were prepared for this study. One of these samples consisted of α-casein digested/labeled in ^{18}O water, followed by removal of the trypsin with a 10,000 molecular weight cutoff filter. Once trypsin was removed, samples were diluted 1:10 with natural abundance water or a formic acid solution to reach a pH of 7 or 2. In this sample, the relative percent loss of double label after 9 d storage at room temperature was 3.6 ± 6.3 at pH 7, and 16.2 ± 3.1% at pH 2 (Table 3.1 and Fig. 3.2). A second sample was prepared by digesting/labeling α-casein in ^{18}O water using immobilized trypsin. This solution was centrifuged and the supernatant removed for examination of back exchange, as previously reported.[23] The loss of label from the sample at pH 7 was 4.2 ± 3.3% after 9 d (Table 3.1 and Fig. 3.2),

while that of the same sample stored at pH 2 was 11.2 ± 6.4% for this time period (Table 3.1 and Fig. 3.2). The extent of back exchange for these two samples are not experimentally distinguishable, and agree well with previous reports on chemical methods for incorporating/exchanging oxygen atoms into the carboxylic groups of peptides under acidic conditions.[19,26] These values are also in agreement with a study that utilized acid catalyzed labeling of the peptide.[27]

Investigation of the delabeling process in the presence of trypsin at a neutral pH was performed. At room temperature, loss of the ^{18}O isotope labels was found to be very rapid, with a 78.0 ± 11.6 % loss of the double isotopic label in 10 min, when a fully labeled α-casein digest was mixed in a 1:10 ratio with a solution of $H_2{}^{16}O$ (Table 3.1, Fig. 3.3-3.5). Although de-labeling at this pH can be expected, the rapidity of loss of the isotopic label is surprising. In particular, the time for de-labeling is much shorter than the timeframe normally used for tryptic digestion in solutions with a pH near the optimum for this protease. An additional experiment with the same conditions but at a temperature of 4°C showed a 22.1 ±14.3% loss of double label after 2 h and a 74.8 ± 19.4% loss of label after 24 h (Table 3.1). The decrease in temperature at neutral pH thus slows down the tryptically facilitated de-labeling process, but significant isotopic loss still occurs. The rate of isotopic loss in both of these experiments is much greater than that observed in the solutions which did not contain trypsin, leading us to conclude that the enzyme is catalyzing this loss. These observations show that significant isotopic losses occur quickly after adding an ^{18}O labeled sample to a solution of $H_2{}^{16}O$, as would happen when an ^{18}O labeled sample is mixed with a ^{16}O labeled sample in $H_2{}^{16}O$ at pH 7 in the presence of trypsin. Furthermore, if this solution is allowed to stand at a neutral pH, the $^{16}O/^{18}O$ ratio in the tryptic peptides will quickly approach that of the bulk solution. The rapidity of isotopic exchange observed under these conditions can clearly lead to erroneous results and is therefore unsatisfactory for quantitative proteomic studies.

Loss of the isotopic label at a pH of 2 in the presence of trypsin was also investigated. Acidic conditions are often used to limit tryptic activity,[9,15,16,19] yet past work[28] has reported that trypsin retains

activity even at low pH. The tryptically facilitated delabeling process at acidic pH was explored by diluting the fully labeled α-casein digest 1:10 in a 5% formic acid solution. At room temperature, a 14.9 ± 7.9% loss of label was observed after 24 h (Table 3.1 and Fig. 3.6). A second sample cooled to 4°C, acidified to pH 2, and stored at this temperature showed the loss of double label to be 6.1 ± 6.4% after 24 h and 17.2 ± 6.7 % after 4 d. Obviously the isotopic loss is more rapid than can be accounted for by acidic back exchange as shown in the previous experiment where loss of label was examined in the absence of trypsin. The overall sum of these results leads to the conclusion that trypsin is the main facilitator of the delabeling process, even at low pH and temperature. These results imply that some loss of ^{18}O to the solvent can occur at pH 2, and that this loss can become significant when the samples are stored for longer than several hours even at a lowered temperature. This situation can arise when samples are deposited into a queue on an auto injector or subjected to further treatment in $H_2{}^{16}O$ based solvents.

Since strong cation exchange and reverse phase are currently the two primary forms of separation for proteomics, another experiment examined the delabeling process using conditions characteristic of a peptide separation by strong cation exchange chromatography. In this case, a fully labeled α-casein digest containing trypsin was combined with an equal amount of natural abundance water, imitating a 1:1 combination of two populations. The sample was immediately snap frozen for 15 sec in liquid nitrogen and lyophilized for 8 h. The sample was then reconstituted in 20% methanol, 0.1% formic acid. After 9 d at room temperature, a 6.8 ± 12.3 relative percent loss of label was calculated (Table 3.1). The extent of back exchange in this case is comparable to the investigation where trypsin was removed from solution and maintained at a neutral pH, indicating the presence of 20% methanol at low pH minimizes the tryptically catalyzed nucleophilic attack of water on the C-terminus of the cleaved peptide (Fig. 3.1, Reactions III and IV). Based on these observations, lowering the pH and adding methanol appears to be a good method to minimize isotopic loss. However, these conditions preclude subsequent procedures, such as enzymatic de-glycosylation, from being performed after isotopic labeling.

A series of experiments were performed to examine other methods for minimizing tryptic activity, thus reducing isotopic loss. The first of these evaluated was heat denaturation, as this was the simplest conceivable procedure to inactivate the enzyme. In these experiments, the fully labeled α-casein digest in $H_2{}^{18}O$ containing the trypsin used to digest/label the peptides was boiled for 10 min. Because heating was performed in 95% $H_2{}^{18}O$, exchange of oxygen atoms on the C-terminus with water via a chemical process would not lead to loss of the isotopic label. This solution was allowed to cool, then diluted 1:10 with natural abundance water. Experiments performed at a pH near the enzymes optimal activity (pH ~7.8) showed a rapid loss of isotope (data not shown for brevity). This rate matched that seen previously for samples that were not boiled, suggesting that trypsin quickly re-natures after being boiled. Thus, simply boiling the solution is not a satisfactory method to minimize loss of the isotopic label. Heat denaturing under low pH was also examined. This was accomplished by acidifying the solution of labeled peptides in $H_2{}^{18}O$ with formic acid prior to boiling for 10 min. This solution was allowed to cool, then diluted 1:10 with natural abundance water. Once again, heating was performed in 95% $H_2{}^{18}O$ water to negate any chemical exchange which may result from the elevated temperature and reduced pH. Results from this investigation revealed that this procedure slows the rate of isotopic exchange with the solvent compared to simply lowering the pH (Table 3.1 and Fig. 3.6). The isotopic loss observed under these conditions could indicate that trypsin renatures after heating, albeit, more slowly than at a pH of 7. The similarity in isotopic loss with and without boiling suggests that this is not an effective method to reduce isotopic loss. In addition, boiling peptides under acidic conditions can lead to numerous unwanted reactions, such as the hydrolysis of aspartic acid-proline linkages, the loss of sialic acid, and deamidation of aspartic acid residues. For these reasons heat denaturation was deemed unacceptable.

Chemical inhibition of trypsin with TLCK was also examined as a means of slowing/stopping the enzymatic facilitation of isotopic exchange with the solvent. In these experiments, TLCK was added (30:1 inhibitor to enzyme ratio) to a solution of fully labeled α-casein digest in $H_2{}^{18}O$ containing the

trypsin used to digest/label the peptides. This solution was allowed to stand at room temperature for 24 h, then diluted 1:10 with natural abundance water. TLCK is not stable at pHs over 7.5, therefore, the pH of this solution was lowered to a pH of 2 with concentrated formic acid prior to addition of the TLCK solution. Under these conditions, it is expected that trypsin would be significantly inhibited due to the high ratio of inhibitor: enzyme. MALDI-MS revealed that after 3 d, the relative loss of double label was 22.1 ± 14.1% (Table 3.1 and Fig. 3.6). This level of isotope loss is experimentally indistinguishable from that observed after heat denaturation at this pH (discussed above), although it is several orders of magnitude slower than that observed without trypsin denaturation. The need to maintain a low pH to reduce isotopic loss, both with TLCK and heat denaturation, prohibits additional steps, such as enzymatic de-glycosylation, from being performed after isotopic labeling. This limitation, combined with the substantial isotopic loss, lead us to conclude that TLCK treatment to reduce isotopic exchange is not a satisfactory method.

DISCUSSION

The use of enzyme facilitated ^{18}O labeling as a method of stable isotope incorporation for quantitative proteomics requires control of isotopic loss for accurate quantitation. The continued interaction of the cleaved peptide with trypsin, even under conditions previously thought to severely limit proteolytic activity, adds further evidence[22] that the mechanism for proteolysis (reactions I-IV in Fig. 3.1) and delabeling/labeling (reactions III_F+ IV_F/III_R + IV_R in Fig. 3.1) can occur under different conditions. In particular, the labeling/delabeling reactions appear to readily occur at a lower pH than proteolysis. Using a serine protease such as trypsin to digest and label peptides for quantitative proteomics is useful in that it can de-couple the digestion step from the labeling procedure. Unfortunately, this continued interaction with the cleaved peptide allows enzyme facilitated exchange of the isotope on the peptides with the solvent anytime the sample and enzyme are present in the same sample. Therefore, the experimental design for a quantitative proteomic scheme utilizing a $^{16}O/^{18}O$ enzyme facilitated labeling scheme to differentiate populations must take into account the capabilities for continued interaction

between the cleaved peptide and the serine protease. As demonstrated here, loss of the label becomes negligible after removal of the trypsin from solution and maintaining the solution at a neutral pH prior to analysis. This study serves to alert investigators to the fact that the enzymatic mechanism of incorporating the ^{18}O stable isotope also serves as the main process by which the stable isotope tag is removed. In conclusion, trypsin is the primary mechanism by which the ^{18}O isotopic label is lost in quantitative proteomic studies, even under conditions of low pH and low temperature.

REFERENCES

1. W. Wang, H. Zhou, H. Lin, S. Roy, T.A. Shaler, L.R. Hill, S. Notron, P. Kumar, M. Anderle, C.H. Becker. *Anal. Chem*. 2003; 75: 4818-4826.

2. H. Liu, R.G. Sadygov, J.R. Yates. Anal. Chem. 2004; 76: 4193-4201.

3. D. Radulovic, J. Salomeh, S. Ryu, G. Hamilton, E. Foss, Y. Mao, A. Emili. *Mol. Cell. Proteomics* 2004; 3: 984-997.

4. J. Silva, R. Denny, C.A. Dorschel, M. Gorenstein, I.J. Kass, G. Li, T. McKenna, M.J. Nold, K. Richardson, P. Young, S. Geromanos. *Anal. Chem.* 2005; 77: 2187-2200.

5. S.E. Ong, B. Blagoev, I. Kratchmarova, D.B. Kristensen, H. Steen, A. Pandey, M. Mann. *Mol. Cell. Proteomics* 2002; 1:376-386.

6. Y. Oda, K. Huang, F.R. Cross, D. Cowburn, B.T. Chait. Proc. Natl. Acad. Sci. USA 1999; 96: 6591-6596.

7. L. Paša-Tolić, P.K. Jensen, G.A. Anderson, M.S. Lipton, K.K. Peden, S. Martinović, N. Tolić, J.E. Bruce, R.D. Smith. *J. Am. Chem. Soc.* 1999; 121: 7949-7950.

8. S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold. *Nature Biotechnol*. 1999; 17: 994-999.

9. M. Schnozler, P. Jedrzejewski, W. Lehmann. *Electrophoresis* 1996; 17: 945-953.

10 K.C.S. Rao, R.T. Carruth, M. Myagi. *J. Proteome Res*., 2005; 4: 507-514.

11. P. Liu, F. Regnier. *J. Proteome Res*. 2002; 1: 443-450.

12. F. Che, L.D. Fricker. *Anal. Chem.* 2002; 74: 3190-3198.

13. J.-L. Hsu, S.-Y. Huang, N.-H. Chow, S.-H. Chen. *Anal. Chem*. 2003; 75: 6843-6852.

14. M. Jin, H. Bateup, J. Padovan, P. Greengard, A.C. Nairn, B.T. Chait. *Anal. Chem*. 2005; 77: 7845-7851.

15. L. Zang, D. Palmer Toy, W.S. Hancock, D.C. Sgroi, B.L. Karger. *J. Proteome Res*. 2004; 3: 604-612.

16. J. Blonder, M.L. Hale, K.C. Chan, L.R. Yu, D.A. Lucas, T.P. Conrads, M. Zhou, M.R. Popoff, H.J. Issaq, B.G. Stiles, T.D. Veenstra. *J. Proteome Res*. 2005; 4: 523-531.

17. Y.K. Wang, Z. Ma, D.F. Quinn, E.W. Fu. *Anal. Chem*. 2001; 73: 3742-3750.

18. X. Yao, C. Afonso, C. Fenselau.*J. Proteome Res*. 2003; 2: 147-152.

19. I.I. Stewart, T. Thomson, D. Figeys. *Rapid Commun. Mass Spec.* 2001; 15: 2456-2465.

20. K.J. Reynolds, X. Yao, C. Fenselau. *J. Proteome Res.* 2002; 1:27-33.

21. S.A. Martin, R.S. Johnson, C.E. Costello, K. Bieman, in: C.J. McNeal (Ed.) The Analysis of Peptides and Proteins by Mass Spectrometry, P*roceedings of the Fourth Texas Symposium on Mass Spectrometry*, John Wiley & Sons, New York, 1988, pp. 135-150.

22. A. Staes, H. Demol, J. Van Damme, L. Martens, J. Vanderckhove, K. Gevaert,. *J. Proteome Res*. 2004; 3: 786-791.

23. K.J. Brown, C. Fenselau. *J. Proteome Res*. 2004; 3:455-462.

24. K. Fodor, V. Harmat, R. Neutze, L. Szilágyi, L. Gráf, G. Katona. *Biochemistry* 2006; 45:2114-2121.

25. Fink, A.LIn *Enzyme Mechanisms*; Page, M. I., Williams, A., Eds.; The Royal Society of Chemistry, London, UK, 1987; pp 159-177.

26. R.C. Murphy, K.L. Clay. *Methods Enzymol.* 1990; 193: 338-348.

27. D.M. Desiderio, M.Kai. *Biomed. Mass Spec*. 1983; 10:471-479.

28. M. Kunitz, J.H. Northrop. *J. Gen. Phys*. 1934; 17: 591-615.

Figure 3.1: A mechanism depicting the process by which peptides are cleaved from the protein and isotopically labeled, where E is the enzyme, Pro is the intact protein, Pep-K/R is the tryptic peptide terminated with either lysine or arginine, and Pep is the other cleaved peptide. It is likely that the optimal conditions for Reaction III where the enzyme is interacting with the cleaved peptide, are different than the conditions for Reaction I and Reaction II where the enzyme is interacting with the intact protein to induce cleavage of the peptide backbone.

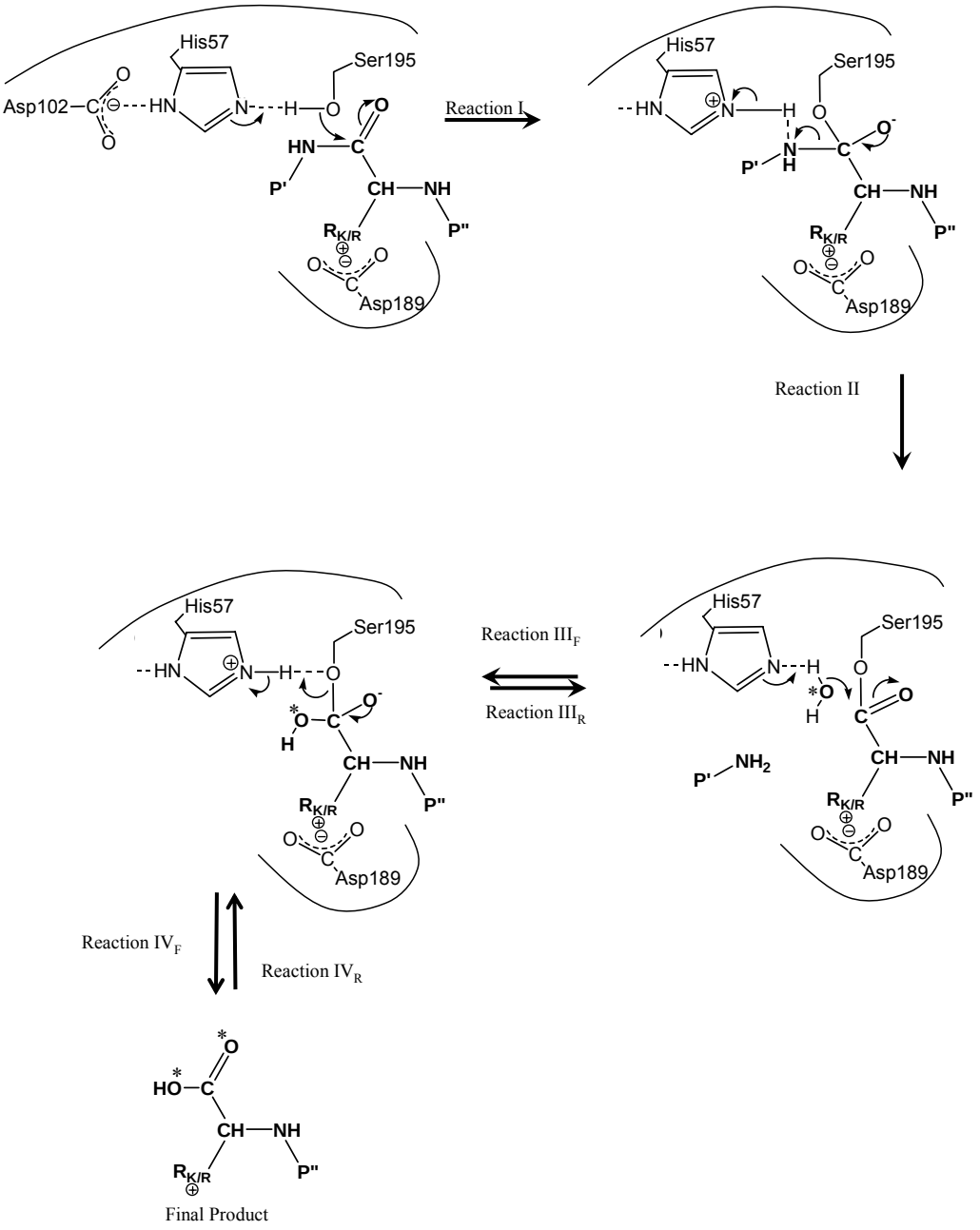


Figure 3.2: Relative percent loss of the double ^{18}O label from the fully labeled casein digest as a function of time stored in solutions free of trypsin at room temperature after being mixed in a 1:10 ratio with a solution of $H_2{}^{16}O$ at either a pH of 2 or 7.

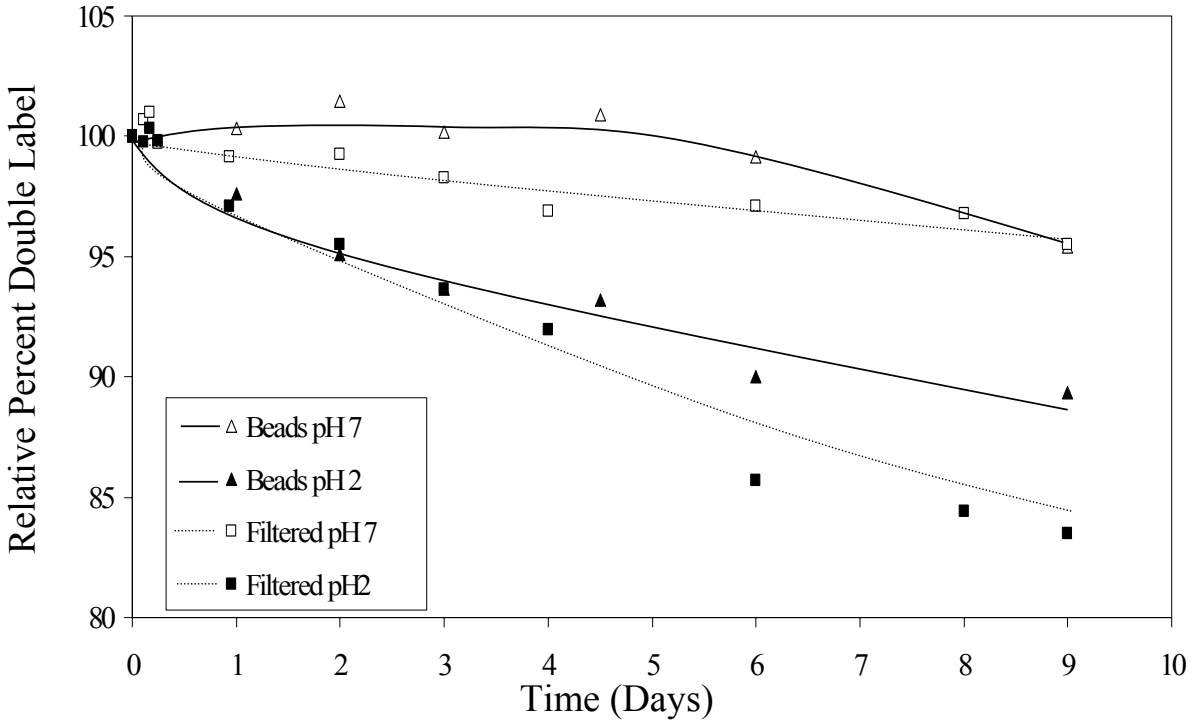


Figure 3.3: Region of the MALDI mass spectrum showing the peptide with a molecular mass of 1759 (HQGLPQEVLNENLLR) from the analysis of α casein digested in 95% $H_2{}^{18}O$, time zero.

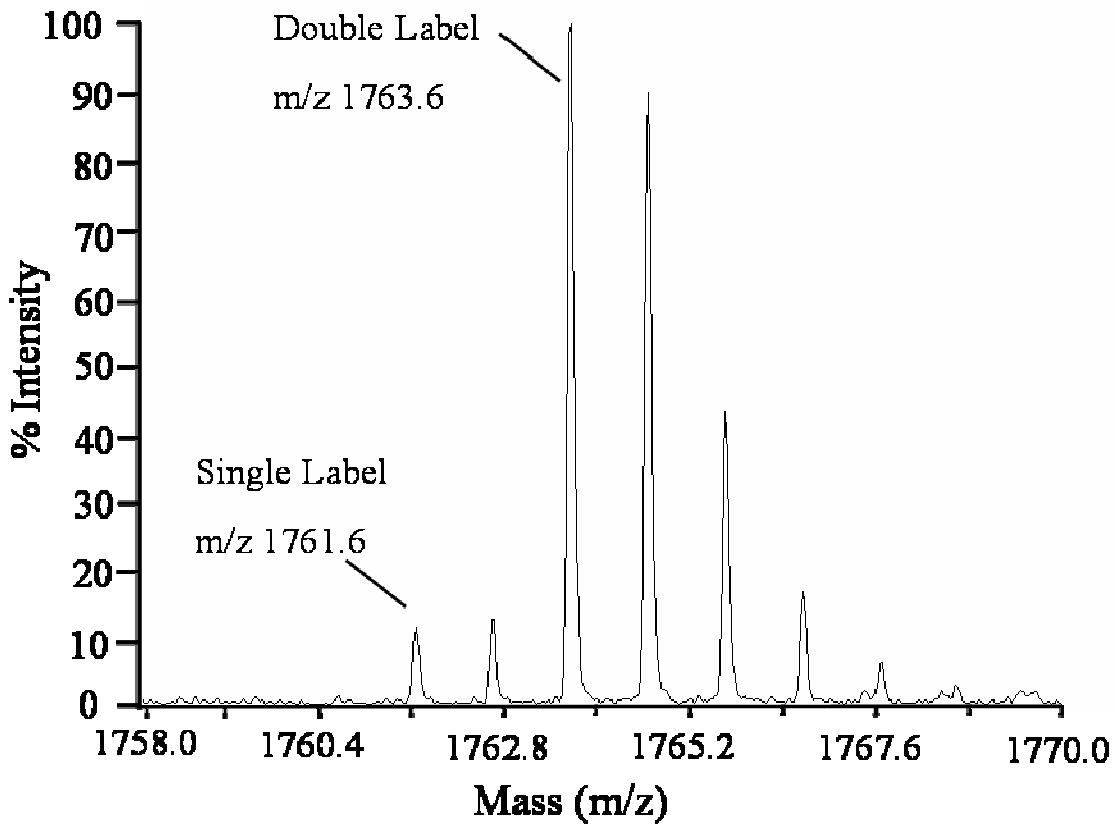


Figure 3.4**:** Region of the MALDI mass spectrum showing the peptide with a molecular mass of 1759 (HQGLPQEVLNENLLR) from the analysis of α casein digested in 95% $H_2{}^{18}O$and then mixed in a 1:10 ratio with a solution of $H_2{}^{16}O$ at a pH of 7 for 10 minutes.

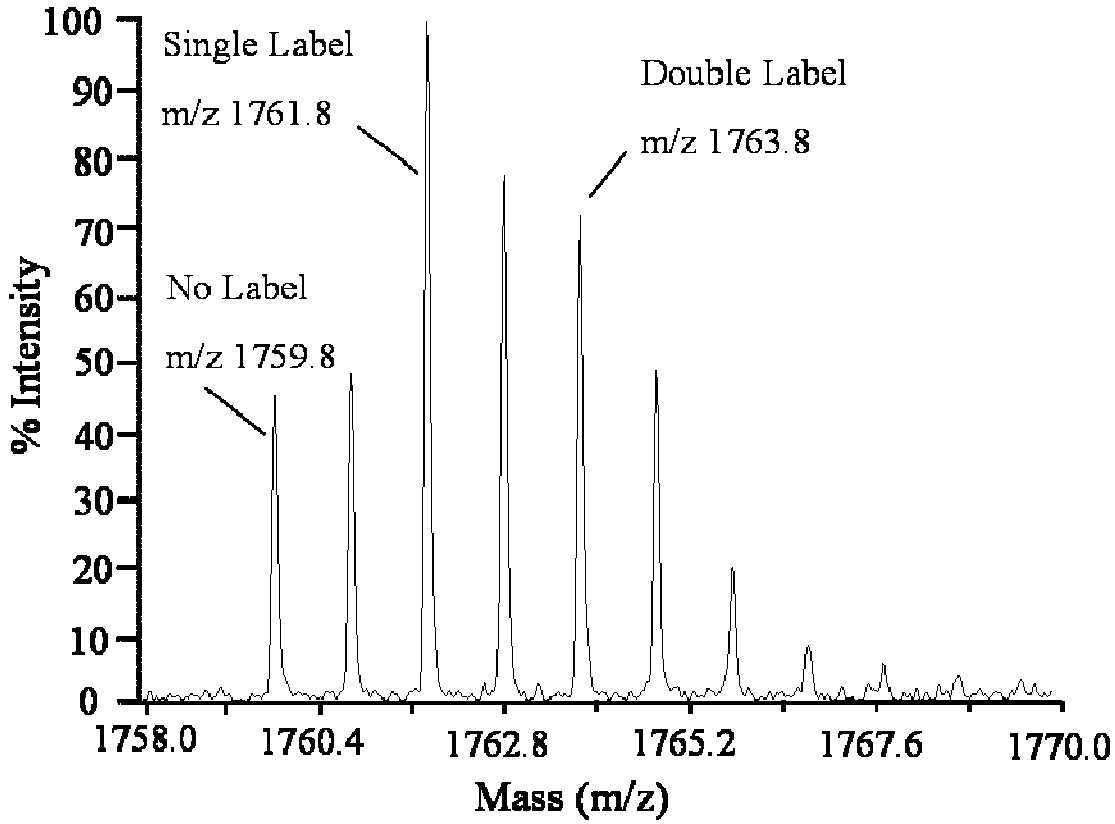


Figure 3.5**:** Region of the MALDI mass spectrum showing the peptide with a molecular mass of 1759 (HQGLPQEVLNENLLR) from the analysis of α casein digested in 95% $H_2{}^{18}O$and then mixed in a 1:10 ratio with a solution of $H_2{}^{16}O$ at a pH of 7 for 2 hours.

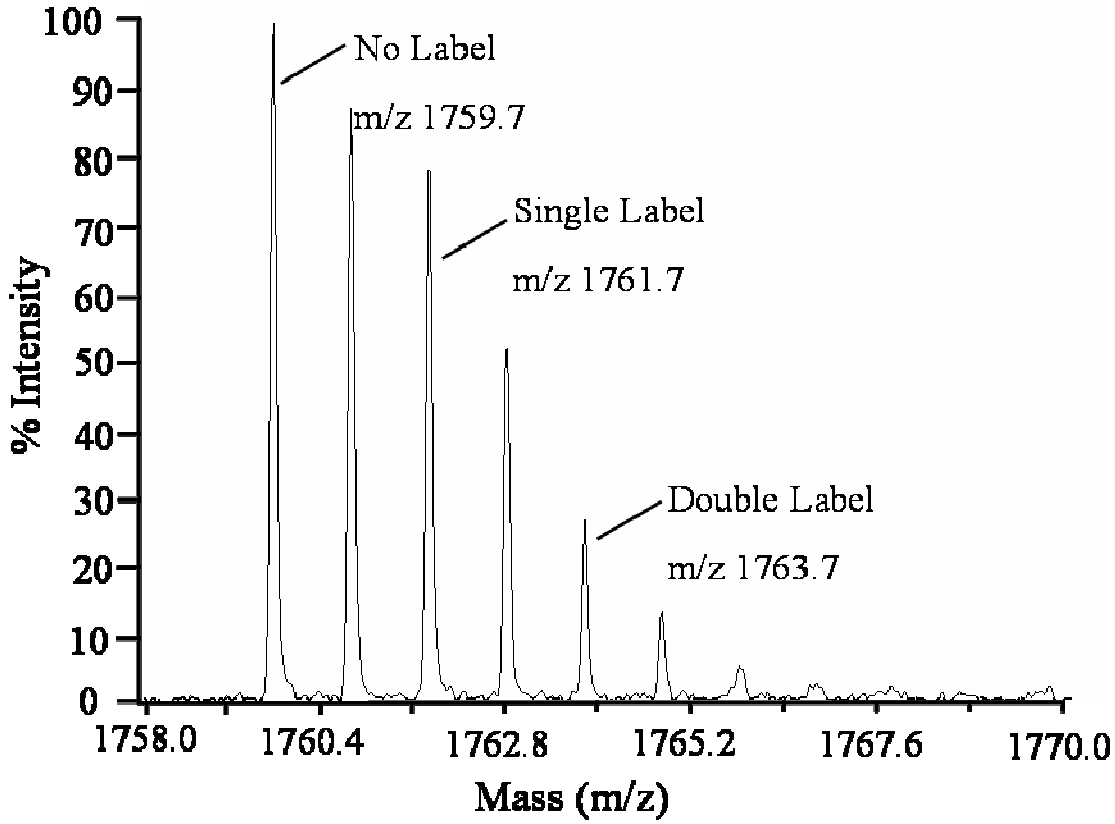


Figure 3.6: Relative percent loss of the double ^{18}O label from the fully labeled casein digest under conditions as a function of time stored at room temperature after being mixed in a 1:10 ratio with a solution of $H_2{}^{16}O$ at a pH of 2. In this figure, conditions were used to inactivate trypsin by lowering the pH, heat denaturation, and the addition of a trypsin inhibitor (TLCK).

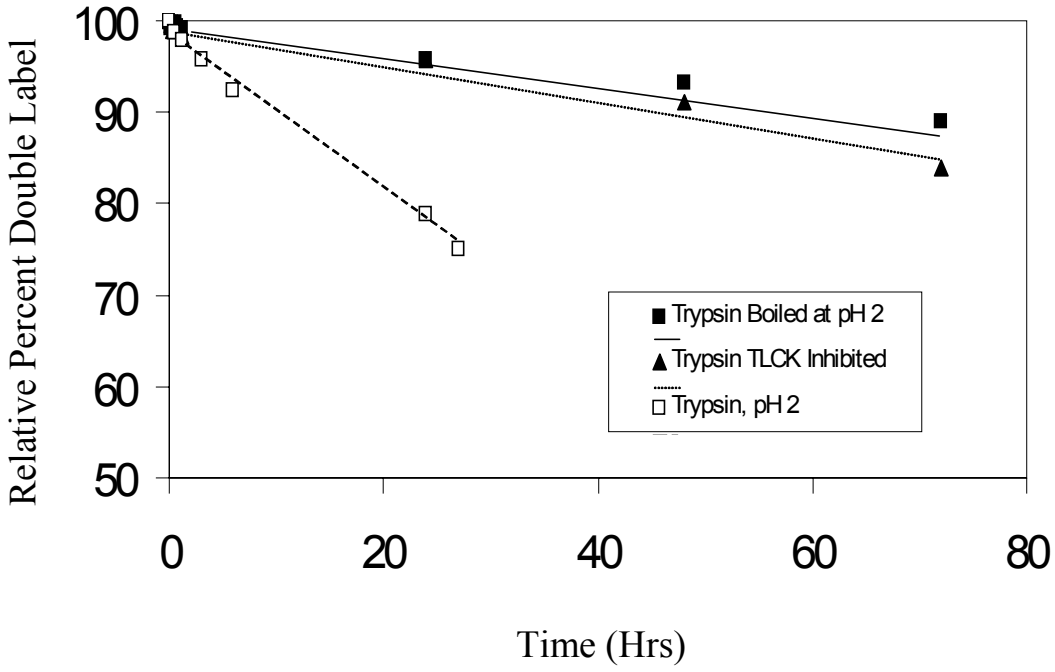


Table 3.1: Summary of the results for the relative percent loss of double ^{18}O label. Percent double label at each time point equals ($^{18}O2$ peak/[^{16}O peak + $^{18}O2$ peak]) - isotopic contributions x 100. The percent double label for each peptide was calculated and an average over all peptides was taken for each time point. Relative percent change is calculated relative to the initial labeling value.

Condition	Final Timepoint	Relative % Change
Trypsin, pH 7, RT*	2 hr	95.2 ± 6.3
Trypsin, pH 2, RT*	24 hrs	14.9 ± 7.9
Trypsin, pH 7, 4°C*	2 hr	22.1 ± 14.3
Trypsin, pH 2, 4°C*	4 Days	17.9 ± 6.7
Beads, pH 7*	9 Days	4.2 ± 3.3
Beads pH 2*	9 Days	11.2 ± 6.4
Trypsin, Filtered pH 7†	9 Days	3.6 ± 6.3
Trypsin, Filtered pH 2†	9 Days	16.2 ± 3.1
Trypsin, TLCK†	3 Days	22.1 ± 14.1
Trypsin, heat pH 2†	3 Days	16.7 ± 12.2
Trypsin, 20% methanol, pH 2	9 Days	6.8 ± 12.3

CHAPTER 4

A POTENTIAL PITFALL IN 18O BASED N-LINKED GLYCOSYLATION SITE MAPPING[2]

[2] Angel, Peggi M., Lim, J., Wells, L., Bergmann, C., Orlando, R. 2007
Rapid Commun Mass Spectrom. 21:674-682.

ABSTRACT

A common procedure for identifying N-linked glycosylation sites involves tryptic digestion of the glycoprotein, followed by the conversion of glycosylated asparagine residues into ^{18}O labeled aspartic acids by PNGase F digestion in ^{18}O water. The 3 Da mass tag created by this process is readily observable by LC-MS/MS analysis, and is often used to identify the sites of N-linked glycosylation. While using this procedure, we noticed that 60% of the asparagines identified as being glycosylated were not part of the consensus sequence required for N-linked glycosylation, and thus were not biologically possible. Investigation into the source of this unacceptably high false positive rate demonstrated that even after reverse phase cleanup and heat denaturation, the trypsin used for proteolysis was still active and led to the incorporation of ^{18}O into the C-termini of the peptides during the de-glycosylation step. The resulting mass shift accounted for most of the false positive sites, as the database search algorithm confused it with an ^{18}O labeled Asp residue near the C-terminus of a peptide. This problem can be overcome by eliminating trypsin from the solution prior to performing the deglycosylation process, by resuspending the peptides in natural abundance water following deglycosylation, or by allowing ^{18}O incorporation into the C-terminus as a variable modification during the database search. These methods have been demonstrated on a model protein, and are applicable to the analyses of glycoproteins that are digested with trypsin or another serine protease prior to enzymatic release of the carbohydrate side chains. This study should alert investigators in the field to this potential and unexpected pitfall and provide strategies to overcome this phenomenon.

INTRODUCTION

Glycosylation of an asparagine (Asn) residue via the amido group forms a β-aspartylglycosylamine linkage to the polypeptide chain and is the basis of eukaryotic N-linked glycosylation[1]. Asparagine linked glycosylation in eukaryotic cells occurs specifically in the consensus sequence Asn-Xxx-Ser/Thr where Xxx can be any amino acid excluding proline[1,2]. N-linked glycosylation has been demonstrated to play a defined role in intracellular processes such as protein folding, cell transport and signaling, and cell proliferation[3-5]. The study of N-linked glycosylation has shown that disruptions to this process may have significant biological implications, ranging from congenital defects to pathogenesis.[6,7] N-linked glycosylation has also been implicated in cellular invasion by bacterial, viral, and eukaryotic parasites by facilitating adherence and infiltration of the host cells[7-10]. Thus the identification of sites of N-linked glycosylation has become an integral part of proteomic examinations of congenital and infectious disease pathways.

Identifying sites of N-linked glycosylation is challenging for a number of reasons. The first challenge is the low abundance of the glycan-containing peptide population relative to the total peptide population. This may be overcome by enriching for the N-linked moiety using techniques such as lectin affinity, hydrophilic, or size exclusion chromatography[11-14]. The heterogeneity of the glycans also complicates many of the strategies used for peptide separations by causing such phenomena as peak broadening or the presence of multiple peaks during LC separations. Additionally, the MS/MS spectra of glycopeptides are typically dominated by neutral losses from the glycan, which often interferes with the database search strategies used in high throughput proteomic experiments. For these reasons, N-linked glycans are typically released prior to the use of LC-MS/MS analysis for the identification of sites of glycosylation.

Release of N-linked glycans is readily performed using glycosidases such as PNGase F and PNGase A[15,16]. These peptide-N-glycosidases convert glycosylated asparagine (Asn) residues into aspartic acids (Asp), thus producing a mass shift of 0.9840 Da at the site of modification. However, identification of

Asp within the consensus sequence Asn-Xxx-Ser/Thr is not definitive for assigning potential sites of N-linked glycosylation, as both naturally occurring and experimentally induced deamidation of asparagine may convert Asn into Asp, resulting in the same 0.9840 Da mass shift. Various mass spectrometry techniques have been developed to differentiate between sites of N-linked glycosylation versus deamidation. These include performing the deglycosylation in heavy isotopic water[12, 17] or isolating the glycopeptides by chemically binding the glycopeptides to a solid support before PNGase F release and subsequent LC-MS/MS analysis [18]. In the former technique, glycosidase digests are performed in ^{18}O water, resulting in a +2.9840 Da mass shift due to incorporation of the ^{18}O isotope into the newly formed Asp [17]. This mass shift is more easily recognizable than the 0.9840 Da mass shift observed when deglycosylation is performed in ^{16}O water, thus easing the mass accuracy required for the experiment. Isotopic labeling also allows site(s) of N-linked glycosylation to be differentiated from sites of deamidation.

Recently, we began an investigation into the sites of N-linked glycosylation on a polygalacturonase-inhibiting protein (PGIP) isolated from *Pyrus communis* using PNGase F deglycosylation in ^{18}O water followed by LC-MS/MS. The data resulted in the ambiguous identification and incorrect assignments of N-linked glycosylation sites. Specifically, the raw data showed mass shifts of 2-4 Da beyond what was expected for incorporation of one ^{18}O into those peptides containing formerly glycosylated sites. These mass shifts were also seen in peptides lacking the N-linked sequon. A review of the experimental protocol led us to postulate that the additional mass shift was due to the incorporation of one or two ^{18}O into the C-terminus of the tryptically digested peptides. This incorporation could occur since the tryptic peptides continue to interact with the trypsin that was used for the initial proteolytic digestion [19-23].

Here, we investigate the possibility of tryptic ^{18}O incorporation during PNGase F deglycosylation using fetuin from fetal calf serum, a well known protein containing three sites of N-linked glycosylation. We demonstrate that tryptic incorporation of ^{18}O into the C-terminus of peptides during N-linked glycosylation site mapping can occur and that this unexpected modification can lead to incorrect and

ambiguous identification of sites of N-linked glycosylation. The intent of this study is to alert investigators in the field to the potential and unexpected errors resulting from this phenomenon and to suggest strategies to overcome this pitfall.

EXPERIMENTAL PROCEDURES

Materials

Fetuin from fetal calf serum, glycopeptidase F (PNGase F), ammonium bicarbonate, trifluoroacetic acid, dithiothreitol (DTT), iodoacetamide (IDA), and α-cyanohydroxy cinnamic acid (CHCA) were purchased from Sigma. Isotopic water (95% ^{18}O) was supplied by Isotec through Sigma. Formic acid and acetonitrile were supplied by J.T.Baker. Sequence grade trypsin was purchased from Promega.

Preparation of PGIP digest -- PGIP from pears (*Pyrus communis*, cv Bartlett) was isolated and purified as previously described[24]. Approximately 2 μg aliquots of the PGIP were reduced with DTT followed by carbamidomethylation with iodoacetamide. Trypsin was added in a 1:1 enzyme:substrate ratio and the sample incubated overnight at 37°C. The high enzyme:substrate ratio was used because previous experience with PGIPs has shown that these proteins are extremely resistant to digestion[25]. After digestion the samples were desalted using a C_{18} Microspin Column (Nestgroup) following the manufacturer's protocol, and dried under vacuum at 37°C.

Deglycosylation of PGIPs -- A 7.5 U/mL stock solution of PNGase F was prepared in 100 mM sodium phosphate, pH 7.5 in natural abundance water. The dried PGIP digests were brought up in 9 μL of 95% ^{18}O enriched water. Exactly 1 μL of the PNGaseF stock solution was added to each PGIP sample. Samples were incubated for 5 h at 37°C. Exactly 10 μL of 1% TFA was added to the solution to terminate the reaction. Samples were then dried under vacuum at 37°C before being analyzed on a Finnigan LTQ (Thermo Electron, San Jose, CA).

PGIP analysis by RP-ESI-LTQ -- PGIP samples were analyzed on a Finnigan LTQ interfaced to an Agilent 1100 quaternary pump (Agilent Technologies, Palo Alto, CA). The PGIP digest was pressure loaded onto a PicoTip Emitter column (New Objective, Woburn, MA) packed with 8.5 cm of 5 µm, 120 Å C18 (Waters, Manchester, MA). The loaded column was then washed with 0.1% formic acid in water prior to being interfaced with the LTQ. Mobile phase A was 0.1% formic acid in water, and B was 80% acetonitrile, 0.1% formic acid. Separation of peptides proceeded with a gradient from 5% to 15% B over 20 min., followed by 15-40% B over 15 min. Each full MS scan from 350 to 2000 *m/z* was collected in centroid mode, followed by MS/MS on the top nine precursors.

PGIP data analysis -- MS/MS spectra were searched against a PGIP only database using TurboSequest [26]. The parameters employed were strict trypsin specificity with up to three missed cleavages, and allowed for variable modifications of cysteine (+57.02), oxidation of methionine (+15.98), PNGase facilitated ^{18}O incorporation into the asparagine residue (+2.98), and deamidation of asparagine (+0.98). The resulting peptide summary was filtered using Xcorr values of 2, 2.5, and 3.0, for +1, +2, and +3 charged peptides, respectively.

Preparation of fetuin digest -- A stock solution of 200 pmol/µL fetuin was prepared and reduced with 10 mM DTT followed by carbamidomethylation with iodoacetamide. Trypsin was then added in a 1:50 enzyme to substrate ratio and the sample incubated overnight at 37°C. After digestion, the sample was desalted using a C_{18} Microspin Column. A portion of this desalted stock solution was set aside for PNGase F digestion following the removal of trypsin by ultrafiltration (10,000 MWCO, Amicon).

Peptide –N-glycosidase F deglycosylation of fetuin peptides -- Ten microliter aliquots of the fetuin digest were dried under vacuum at 37°C. The dried fetuin digest was brought up in 10 µL of 95% ^{18}O enriched water. PNGase F was prepared in natural abundance water to 0.1 units of activity per microliter. Exactly 2.5 µL of the fetuin digest was added to 10.5 µL of 50 mM ammonium bicarbonate in ^{18}O enriched water, followed by the addition of 2.5 µL of the PNGase F preparation. Samples were incubated

overnight at 37°C before being analyzed by MALDI-ToF (4700 Proteomics Analyzer, Applied Biosystems) in the positive reflector mode.

Treatment of fetuin peptides to remove C-terminal label -- Following MALDI-ToF analysis of one aliquot of the ^{18}O tryptically labeled fetuin peptides, a portion of the sample was dried under vacuum at 37°C. The sample was then resuspended in natural abundance water, incubated overnight at 37°C, and reanalyzed by MALDI-ToF.

Fetuin analysis by RP-ESI-QToF -- Two aliquots of the fetuin digest that were deglycosylated in the presence of trypsin were analyzed by μRP-ESI-QToF using a 150 μm x 150 mm, 5 μm, 300 Å C18 column (GraceVydac) on a Waters CapLC (Milford, MA) interfaced to a nanoelectrospray source on a QToF-2 (Micromass, Manchester, UK). Mobile phase A was 0.1% formic acid in water; mobile phase B was 0.1% formic acid in acetonitrile. Samples were analyzed using a 1% B/min gradient from 5% to 60% B at 1 μL/min. Resulting spectra were processed by Masslynx v3.5 (Micromass, UK).

Data analysis of fetuin peptides -- Peptides analyzed by MALDI-ToF were identified by comparison with a theoretical digest of fetuin using MS-Digest [http://prospector.ucsf.edu/mshome4.0.htm]. For ESI-ToF data, peptides were identified using the Mascot search algorithm [matrixscience.com] with strict trypsin specificity, one missed cleavage, carbamidomethylation as a fixed modification, and variable modifications of one or two ^{18}O at the C-terminus of lysine or arginine. Peptide-N-glycosidase F incorporation of ^{18}O was accounted for using a variable asparagine modification of 117.03563 as the monoisotopic mass. A database consisting solely of bovine fetuin was used to search obtained spectra.

RESULTS AND DISCUSSION

Recent attempts in our laboratory to map sites of N-linked glycosylation on a plant polygalacturonase inhibiting protein (PGIP) by employing PNGase F deglycosylation in ^{18}O water

followed by LC-MS/MS produced puzzling and ambiguous results. In particular, three of the five N-linked sites of glycosylation determined with high confidence were not found within the Asn-Xxx-Ser/Thr consensus sequence required for N-linked glycosylation (Table 4.1). The apparent 60% false positive rate of this experiment brings into question the legitimacy of the two "identified" sites, as it leaves open the possibility that these were simply randomly identified sites that fortuitously occurred within a consensus sequence. Thus, an approach which simply limits the "identified" N-linked glycosylation sites to Asn residues found within a consensus sequence may yield erroneous results [27].

These ambiguities led us to examine the PGIP data in greater detail. One characteristic shared by the majority of falsely identified glycosylation sites was their close proximity to the C-terminus of large tryptic peptides. Another observation was that the precursor masses for most of the peptides appeared between 2 and 4 mass units greater than that predicted by a theoretical digest of the protein. Careful inspection of the protocol used to map these putative sites of N-linked glycosylation led us to hypothesize that the presence of active trypsin during the de-glycosylation step would lead to the incorporation of ^{18}O into the C-terminus of tryptic peptides. Such an unexpected incorporation of the ^{18}O isotope into the C-terminus would account for the observed mass shifts.

To test this hypothesis, we performed multiple experiments on fetuin from fetal calf serum, a well-characterized glycoprotein with three sites of N-glycosylation. The tryptic digest of fetuin was first performed in natural abundance water, followed by PNGase F catalyzed hydrolysis of the N-linked glycans in natural abundance water. The masses observed for peptides in the MALDI-ToF analysis of this sample appear at those calculated from the amino acid sequence [see the representative peptide in Fig. 4.1A], while peptides with occupied N-linked glycosylation sites appear 0.98 Da above their calculated values as expected from the Asn to Asp conversion associated with deglycosylation [see the representative glycopeptide in Fig. 4.1B]. This experiment was followed by performing the deglycosylation event in the absence of trypsin to demonstrate that this procedure only incorporates ^{18}O at sites occupied by N-linked glycans. For these experiments, an aliquot of the tryptic digest was subjected

to 10 kDa MWCO ultrafiltration to remove trypsin from solution. The sample was then dried, resuspended in 95% abundance ^{18}O water, and digested with PNGase F overnight. MALDI-ToF analysis of this sample clearly shows a mass shift consistent with the incorporation of a single ^{18}O into sites of N-linked glycosylation [Figs. 4.2A and 4.2B], demonstrating that PNGase F only adds the isotope at sites of N-linked glycosylation, as expected.

The deglycosylation process was then examined under normal experimental conditions by taking a portion of the fetuin digest that had been desalted on a C_{18} microspin column drying it and resuspending in ^{18}O enriched water for deglycosylation. Under this protocol, trypsin co-elutes from the microspin column with the peptides/glycopeptides and thus is present during the de-glycosylation step. Inspection of the MALDI-ToF spectrum of the representative peptides (Figs. 4.3A and 4.3B) clearly shows that the dominant ion for each peptide, including the deglycosylated peptides, appears 4 Da above that predicted from the amino acid sequence allowing for the Asn to Asp conversion in ^{18}O water. These spectra also contain peaks from the incorporation of a single ^{18}O into their C-termini (1476.75 and 1745.74 in Figs. 4.3A and 4.3B, respectively). Although unexpected, this incorporation is not surprising as it has previously been shown that trypsin retains its activity after undergoing desalting on a C_{18} column followed by elution with 0.1% triflouroacetic acid in 80% acetonitrile [28]. Similar mass shifts from isotopic incorporation were observed after adding trypsin to a sample that had been de-glycosylated in ^{18}O water in the absence of trypsin (data not shown for brevity). These experiments clearly demonstrate that tryptic activity during the deglycosylation step leads to an unexpected incorporation of ^{18}O into these peptides.

An experiment was also performed to evaluate the effects of heat denaturation on the ability of trypsin to facilitate ^{18}O incorporation since this procedure is commonly used prior to deglycosylation with PNGase F digestion. In this experiment, a dried aliquot of the desalted fetuin digest used in the previous experiments was resuspended in natural abundance water, boiled for 10 min, dried, resuspended in 95% abundance ^{18}O water, and digested with PNGase F overnight. The MALDI spectrum obtained from this

sample, as shown by the representative peptide in Fig. 4A, demonstrates that approximately one-half of this peptide's population has been labeled with the isotope at the C-termini. The level of ^{18}O incorporation is not the same for all peptides/glycopeptides, as shown by the MALDI-ToF spectrum of the representative glycopeptide (Fig 4.4B) where the most intense ion corresponds to the incorporation of a single ^{18}O into the site of glycosylation as confirmed by MS/MS analysis (data not shown for brevity). It is known that the rates of ^{18}O incorporation into individual peptides vary widely, and are dependent on features such as the peptide's length, hydrophobicity, and the amino acid at its C-terminus.[20, 22, 23] In addition to these factors, the N-linked glycans appear to further reduce the level of ^{18}O incorporation into the C-terminus of the glycopeptide (Fig. 4.4B), perhaps via steric hinderance. Clearly boiling has reduced, but not eliminated, isotopic incorporation via the tryptic pathway, and thus can lead to falsely identified sites of N-glycosylation. In addition, boiling the sample between digestions can lead to numerous unwanted reactions, such as the loss of sialic acid from the glycans and deamidation of aspartic acid, the latter of which can clearly lead to misidentifications. The conclusion from this experiment is that ^{18}O incorporation into the C-terminus can occur even after elution from a C_{18} column in 80% acetonitrile with 0.1% triflouroacetic acid followed by heat denaturation.

To confirm that the unexpected incorporation of ^{18}O is at the C-terminus of these peptides, as would result from tryptic activity, a portion of the trypsin-containing deglycosylated digests was subjected to LC-MS/MS analysis. This experiment was performed on a Q-ToF instrument to provide the fragment ion mass accuracy, resolution, and increase the ion fragmentation range as imposed by the ion trap's "1/3 rule". All of these factors were needed to unmistakably identify the type and location of the modification. The resulting spectrum (Fig. 4.5) clearly shows that the experimentally determined masses of the y-ion series from these peptides are increased by 4 Da when compared to the values calculated theoretically. This spectrum localized the 4 Da shift to the C-terminus, which confirms our hypothesis of tryptic incorporation of the isotope during the deglycosylation event.

Incorporation of ^{18}O into the C-terminus offers several explanations for the difficulty in identifying the sites of N-glycosylation on PGIP. First, the ion trap MS system and the standard database search parameters used to search the MS/MS spectra in these experiments do not have sufficient mass accuracy to differentiate between the molecular ions of a peptide with one or two ^{18}O incorporated into the C-termini (a 2 or 4 Da shift) and a peptide with an ^{18}O modified Asp (a 3 Da shift). With large multiply charged ions, which represent the majority of peptides which were misidentified in Table 1, the resolution of the ion trap combined with the presence of ^{13}C peaks and various ^{18}O incorporated species does not allow the unambiguous assignment of the additional mass. The fragmentation patterns of trypsin labeled peptides are also very similar to those with a labeled Asn residue near the C-terminus. As observed above with tryptic incorporation of ^{18}O (Fig.4.5), the isotopic addition occurs on the C-terminus, and thus the mass to charge values of the b-ions are unchanged while those of the y-ions are increased by the mass of the isotope(s). Another example of this phenomenon is shown in the MS/MS spectra from the peptide LQSFDEYSYFHN*R (Fig. 4.6), one of the PGIP peptides which database searching identified as possessing an N-linked glycosylation site outside of a consensus sequon. Careful manual inspection of this data (Table 4.2) identifies that the mass to charge values of the y-ions are consistent with this peptide as having two ^{18}O at the C-terminus. The database search algorithm, however, identified the spectrum in Figure 6 as having one ^{18}O incorporated into the asparagine residue. This misidentification presumably occurred because the 1 Da difference between the calculated y-ion values from these two species is within the allowable mass tolerance of the search. In addition, ion traps are not capable of observing the low *m/z* ions, y_1 in this example, which would allow these two species to be readily differentiated. This reasoning predicts that the problem of misidentifications will decrease as the number of amino acids between the Asn residue and the C-terminus increase, as is observed in the experiment with PGIP where three of the five incorrectly assigned glycosylation sites are found adjacent to the C-terminus. The combination of instrumental factors, database search parameters, and an unexpected incorporation of ^{18}O into the C-terminus of the peptides can thus lead to misidentification of N-linked glycosylation sites by MS/MS analysis.

A primary factor that led to these ambiguous and incorrect assignments appears to be that the addition of ^{18}O to the C-termini was not anticipated, and therefore, the search program was not permitted this assignment as an option. In other words the search algorithm matched the experimental spectrum to the best allowable theoretical spectrum, which in these examples were incorrect. If this reasoning is sound, a solution to this dilemma is to perform database searches that allow for variable modifications of one and two tryptic ^{18}O incorporation(s) into the C-terminus, as well as a variable modification at Asn residues. In this manner the search program can determine which of these species provides the best match with the experimental data. When searches were performed in this manner on the PGIP data, all of the peptides listed in Table 4.1 were designated as having two ^{18}O incorporated into their C-terminus. In addition, all of the N-linked glycosylation sites outside of the required sequon were eliminated, while the two glycosyation sites in consensus sequons were identified as having a modified Asn, indicating that these sites are occupied, which is presumably accurate. These results agree with manual interpretation of these spectra (as was done above in Table 4.2). Thus, this approach has eliminated all of the apparent false positive N-linked glycosylation site identifications without affecting the discovery of actual glycosylation sites. One potential detrimental effect from using this computational solution is that increasing the number of possible peptides searched has been shown to increase the probability of incorrect peptide identifications; [27, 29] however this topic is outside of the scope of this manuscript. Furthermore, allowing for additional variable modifications significantly increases search times. Despite these concerns, simply permitting ^{18}O incorporation into the C-terminus during the database search appears to be a satisfactory solution to the incorrect assignment of N-linked glycosylation sites when using the ^{18}O labeling approach.

An alternative approach to this problem is to either eliminate ^{18}O incorporation into the C-terminus during de-glycosylation, or to remove this modification prior to LC-MS/MS analysis. One solution is to remove trypsin from the solution during the enzymatic deglycosylation. This can be accomplished by the use of immobilized trypsin. Similarly, molecular weight cutoff filtering can be used to remove trypsin, as

was done in the experiment shown above in Figure 4.2. This procedure suffers from the potential loss of peptides resulting from binding to the filter. Alternatively, the C-terminal ^{18}O label can be removed by exploiting the continued tryptic activity on the cleaved peptides. In this method, the sample is simply dried, re-suspended in natural abundance water, and allowed to stand overnight prior to analysis, thus permitting trypsin to re-incorporate ^{16}O into the peptides C-termini. This procedure was performed on the fully ^{18}O labeled sample shown previously in Figure 3, and the resulting MALDI-ToF spectra (Fig. 4.7A and 4.7B) demonstrate the complete removal of the isotopes from the C-terminus of the peptides. This procedure has no effect on the ^{18}O incorporated into the newly formed Asp residue since trypsin only catalyzes the exchange of oxygen atoms on the C-terminal carboxylic acid of peptides terminating in Lys or Arg -- not from the side-chains of the amino acids. The deglycosylated peptide is no longer a substrate for the glycosidase and thus the ^{18}O incorporated into the newly formed Asp is unaffected by the presence of this enzyme. This approach effortlessly eliminates the problem without incurring significant sample losses. When this modified approach was used on the PGIP digest, all of the N-linked glycosylation sites outside of the required sequon were not identified as being glycosylated, while the two sites in consensus sequons were recognized as having a modified Asn, indicating that these sites are occupied. These results are in agreement with those obtained by searching the database with a variable C-terminus modification of +2 and+4 Da, discussed above. This modified approach was ultimately used to fully characterize the N-linked glycosylation sites present on this PGIP (manuscript in preparation). Thus, removal of the ^{18}O from the C-terminus seems to have eliminated all of the misidentified N-linked glycosylation sites.

CONCLUSIONS

Tryptic addition of one and two ^{18}O into the peptide C-terminus during N-linked deglycosylation can result in the erroneous identification of N-linked sites of glycosylation. It is expected that the presence of other serine proteases, such as Lys-C and Glu-C, will also lead to the addition of ^{18}O during de-

glycosylation in ^{18}O water, since these enzymes, like trypsin, catalyze the addition of two ^{18}O atoms into the C terminus of the proteolytic peptides[20, 22, 30]. Because trypsin continues to interact with the cleaved peptides even at lower pHs [21], this potential pitfall can be extended to other glycosidases such as PNGase A, which has an optimal pH of 5.0. This problem can be overcome by eliminating trypsin from the solution prior to performing the deglycosylation process, by resuspending the peptides in natural abundance water following deglycosylation, or by allowing ^{18}O incorporation into the C-terminus as a variable modification during the database search. These methods have been demonstrated on a model and an unknown protein, and are applicable to the analyses of glycoproteins that are digested with trypsin or another serine protease prior to enzymatic release of the carbohydrate side chains. This study should alert investigators in the field of this potential and unexpected pitfall.

REFERENCES

1. Kornfield, R., and Kornfield, S. *Annu. Rev. Biochem,*1985, *54*, 661-664.

2. Bause, E., and Hettkamp, H. *FEBS Lett.* 1979, *108*, 341-344.

3. Helenius, A., and Aebi, M. *Science* 2001, *291*, 2364-2369.

4. Trombetta, E.S. *Glycobiology* 2003, *13*, 77-91.

5. Davis, B. *Chem. Rev.* 2002, *102*, 579-601.

6. Schachter, H. *Cell Mol. Life Sci.* 2001, *58*, 1085-1104.

7. Freeze, H., and Westphal, V. *Biochimie* 2001, **83**, 791-799.

8. Karlyshev, A. V., Everest, P., Linton, D., Cawthraw, S., Newell, D. G., and Wren, B. W. *Microbiology* 2004, *150*, 1957-1964.

9. Poignard, P., Saphire, E., Parren, P., and Burton, D. R. *Annu. Rev. Immunol.* 2001, *19*, 253–274.

10. Petri, W.A., Haque, R., and Mann, B. *Entamoeba histolytica*. *Annu. Rev. Microbiol.* 2002, *56*, 39-64.

11. Xiong, L., and Regnier, F. *J. Chrom. B* 2002, *782*, 405-418.

12. Kaji, H., Saito, H., Yamauchi, Y., Shinkawa, T., Toka, M., Hirabayashi, J., Kasai, K., Takahashi, N., Isobe, T. *Nature Biotech* 2003, **21**, 667-672.

13. Tajiri, M., Yoshida, S., and Wada, Y. *Glycobiology* 2005, *15*, 1332-1340.

14. Alvarez-Manilla, G., Atwood, J., Guo, Y., Warren, N., Orlando, R., and Pierce, M. *J. Proteome Res.* 2006, *5*, 701-708.

15. Tarentino, A., Gómez, C., and Plummer, T. *Biochemistry* 1985, *24*, 4665-4671.

16. Altmann, F., Paschinger, K., Dalik, T., and Vorauer, K. *Euro. J. Biochem.* 1998, *252*, 118-123.

17. Gonzalez, J., Takao, T., Hor, H., Besada, V., Rodríguez, R., Padron, G., and Shimonishi, Y. *Anal. Biochem.* 1992, *205*, 151-158.

18. Zhang, H., Xiao-jun, L., Martin, D., and Aebersold, R. *Nature Biotech.* 2003, *21*, 660-666.

19. Wang, Y. K., Ma, Z., Quinn, D. F., and Fu, E.W. *Anal. Chem*. 2001, *73*, 3742-3750.

20. Yao, X., Afonso, C., and Fenselau, C. *J. Proteome Res*. 2003, *2*, 147-152.

21. Staes, A., Demol, H., Van Damme, J., Martens, L., Vanderckhove, J., and Gevaert, K. *J. Proteome Res.* 2004, *4*, 786-791.

22. Schnözler, M., Jedrzejewski, P., and Lehmann, W. *Electrophoresis* 1996, *17*, 945-953.

23. Stewart, I.I., Thomson, T., and Figeys, D. *Rapid Commun. Mass Spectrom*. 2001, *15*, 2456-2465.

24. Stotz, H., Powell, A., Damon, S., Greve, L., Bennett, A., and Labavitch, J. *Plant Physiol.* 1993, *102*, 133-138.

25. King, D., Bergmann, C., Orlando, R., Benen, J.A.E., Kester, H.C.M., and Visser, J. *Biochemistry* 2002, *41*, 10225-10233.

26. Eng, J., McCormack, A., and Yates, J. *J. Am. Mass Spectrom.* 1994, 5, 976-989.

27. Atwood, J., Sahoo, S., Manilla, G., Weatherly, D., Koli, K., Orlando, R., and York, W. *Rapid Commun. Mass Spectrom.* 2005, *19*, 3002-3006.

28. Titani, K., Sasagawa, T., Resing, K., and Walsh, K. *Anal. Biochem.* 1982, *123*, 408-412.

29. Resing, K., Meyer-Andt, K., Mendoza, A., Aveline-Wolf, L., Jonscher, K., Pierce, K., Old, W., Cheung, H., Russell, S., Wattawa, J., Goehle, G., Knight, R., and Ahn, N. *Anal. Chem.* 2004, *76*, 3556-3568

30. Reynolds, K.J., Yao, X., Fenselau, C. *J. Proteome Res.* 2002, *1*, 27-33.

Table 4.1. Ambiguous identification of N-linked peptides during PNGase F release in ^{18}O water and in the presence of trypsin. Peptides identified as containing N-linked glycosylation sites in the analysis of pear polygalacturonase inhibiting protein. Sites "identified" as being glycosylated are denoted with as N*, and the consensus sequences for N-linked glycosylation (N-X-S/T) are underlined. In this experiment, care was not taken to prevent tryptic activity during PNGase F deglycosylation in ^{18}O enriched water.

Denoted N-linked glycosylation sites	Charge	XCorr	ΔC_n
IYGSIPVEFTQLN*FQFLN*VSYN*R(L)	3	5.46	0.71
IYGSIPVEFTQLNFQFLN*VSYN*R(L)	2	4.58	0.10
LQSFDEYSYFHN*R(C)	2	3.48	0.18
NKLEGDASVIFGLN*K(T)	3	4.22	0.12

Table 4.2. Examination of MSMS from a peptide containing a falsely identified site of glycosylation. Theoretical and experimental m/z values for the b- and y-ion fragments of the peptide identified as LQSFDEYSYFHNR. The values were calculated with no modification, ^{18}O incorporation into the Asp residue produced by deglycosylation, and with the addition of two ^{18}O into the C-terminus from tyrptic activity. The experimental values were obtained from the LC-MS/MS analysis of trypsin digested and PNGase F deglycosylated polygalacturonase inhibiting protein (PGIP).

	b-ion	Theoretical Ions			Observed
Sequence	**No.**	**Unmodified**	**Deglycosylation**	**Two $^{18}O_2$**	**b-ion series**
L					
Q	2	242.2	242.2	242.2	**242.2**
S	3	329.2	329.2	329.2	**329.2**
F	4	476.3	476.3	476.3	
D	5	591.3	591.3	591.3	**591.3**
E	6	720.3	720.3	720.3	
Y	7	883.4	883.4	883.4	**883.4**
S	8	970.4	970.4	970.4	**970.4**
Y	9	1133.5	1133.5	1133.5	**1133.5**
F	10	1280.6	1280.6	1280.6	**1280.5**
H	11	1417.6	1417.6	1417.6	**1417.6**
N	12	1531.7	1534.6	1531.7	**1531.6**
R					

	y-ion	Theoretical Ions			Observed
Sequence	**No.**	**Unmodified**	**Deglycosylation**	**Two $^{18}O_2$**	**y-ion series**
L					
Q	12	1592.7	1595.7	1596.7	
S	11	1464.6	1467.6	1468.6	**1468.6**
F	10	1377.6	1380.6	1381.6	**1381.6**
D	9	1230.5	1233.5	1234.5	**1234.5**
E	8	1115.5	1118.5	1119.5	**1119.5**
Y	7	986.5	989.4	990.5	**990.5**
S	6	823.4	826.4	827.4	**827.5**
Y	5	736.4	739.3	740.4	**740.4**
F	4	573.3	576.3	577.3	**577.4**
H	3	426.2	429.2	430.2	**430.3**
N	2	289.2	292.1	293.2	**293.2**
R	1	175.1	175.1	179.1	

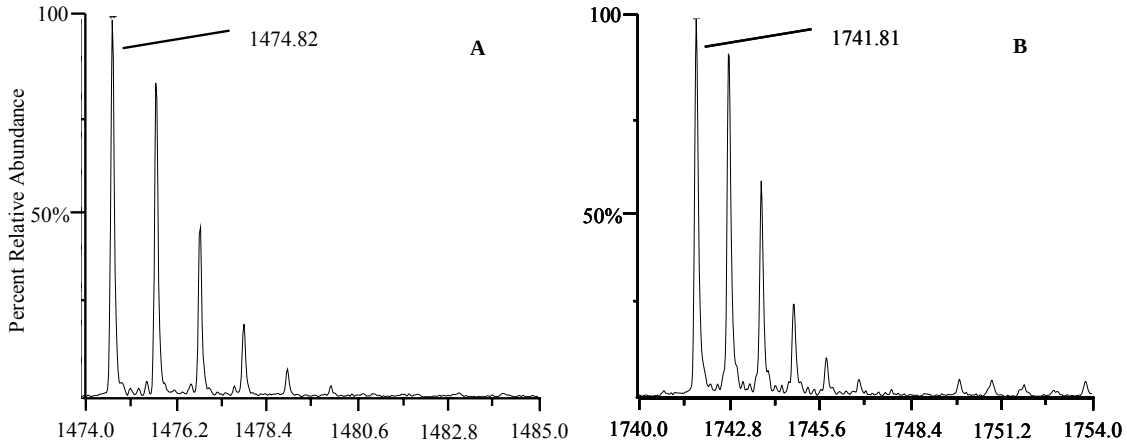


Figure 4.1. Transferrin digested and deglycosylated in natural abundance water. MALDI-ToF spectrum of trypsin digested bovine fetuin showing (A) a representative tryptic peptide **(**TPIVGQPSIPGGPVR) and (B) a representative N-linked glycopeptide (LCPDCPLLAPLN*****DSR) after PNGase F cleavage in natural abundance water. The molecular ion for the peptide appears at 1474.8 and that of the glycopeptide after deglycosylation appears at 1741.8.

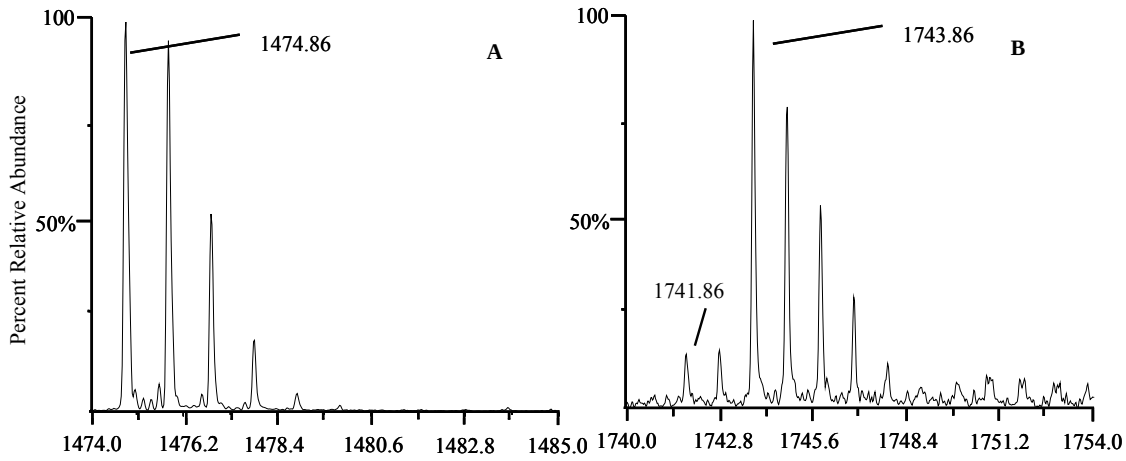


Figure 4.2. Deglycosylation of transferrin in the absence of trypsin and in the presence of ^{18}O enriched water. MALDI-ToF spectrum of trypsin digested bovine fetuin showing (A) a representative tryptic peptide **(**TPIVGQPSIPGGPVR) and (B) a representative N-linked glycopeptide (LCPDCPLLAPLN*****DSR) after PNGase F cleavage in 95% ^{18}O enriched water in the absence of trypsin. The molecular ion for the peptide appears at 1474.8, showing no incorporation of ^{18}O. The glycopeptide appears at 1741.7 and 1743.7 indicating the incorporation of either a ^{16}O or an ^{18}O into the Asp formed during deglycosylation. The ratio of these two ions reflects the $^{16}O/^{18}O$ ratio of the solution.

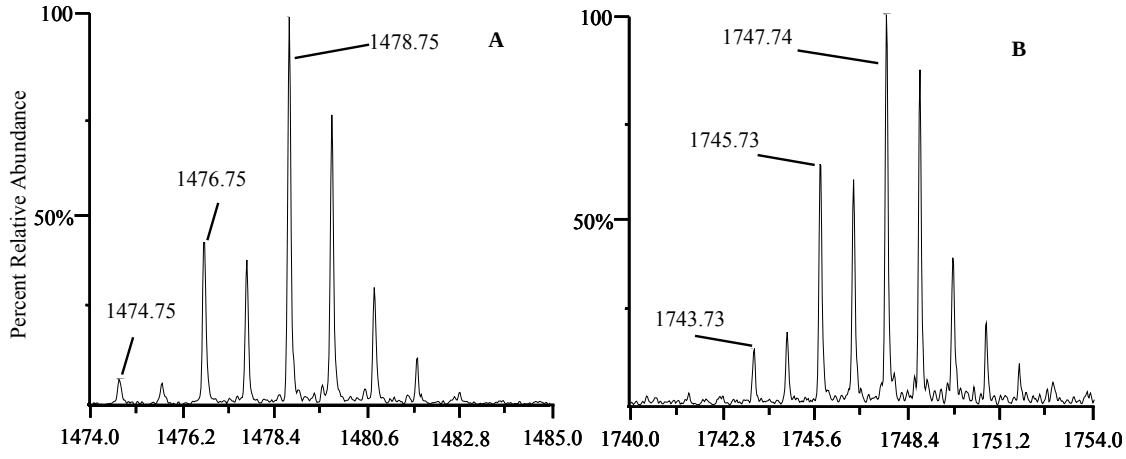


Figure 4.3**.** Tryptically induced ^{18}O incorporation during PNGase F deglycosylation. MALDI-ToF spectrum of trypsin digested bovine fetuin showing (A) a representative tryptic peptide **(**TPIVGQPSIPGGPVR) and (B) a representative N-linked glycopeptide (LCPDCPLLAPLN*****DSR). In this experiment, an aliquot of the fetuin digest was desalting on a C_{18} column, eluted with 0.1% triflouroacetic acid in 80% acetonitrile, dried, resuspended in 95% abundance ^{18}O water, and digested with PNGase F overnight. No effort was made to remove the trypsin from this sample prior to deglycosylation. The molecular ion for the peptide appears at 1474.8 (no ^{18}O), 1476.8 (one ^{18}O) and 1478.8 (two ^{18}O). The glycopeptide appears at 1743.7 (one ^{18}O), 1745.7 (two ^{18}O), and 1747.7 (three ^{18}O). The ratio of these ions reflects the $^{16}O/^{18}O$ ratio of the solution.

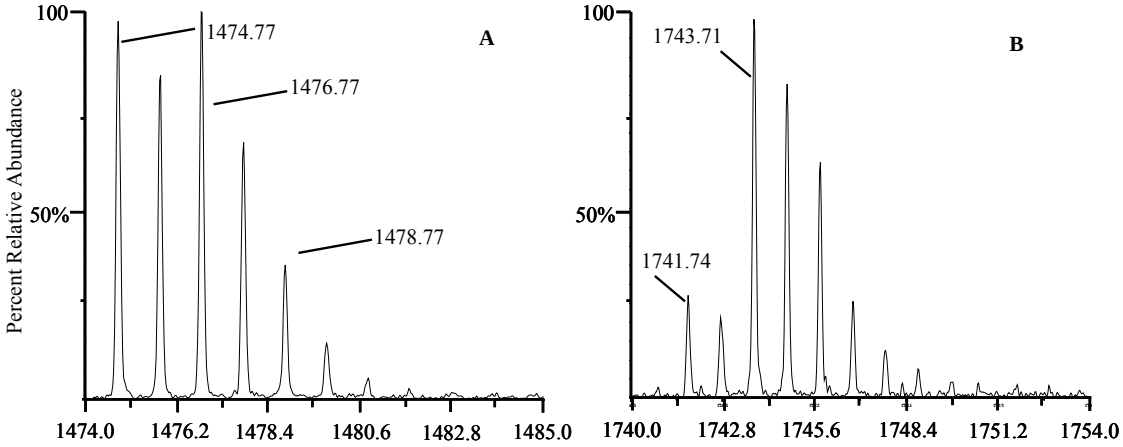


Figure 4.4. Tryptically induced ^{18}O incorporation after boiling prior to PNGase F deglycosylation. MALDI-ToF spectrum of trypsin digested bovine fetuin showing (A) a representative tryptic peptide **(**TPIVGQPSIPGGPVR) and (B) a representative N-linked glycopeptide (LCPDCPLLAPLN*****DSR). In this experiment, an aliquot of the fetuin digest was desalting on a C_{18} column, eluted with 0.1% triflouroacetic acid in 80% acetonitrile, dried, resuspended in natural abundance water, boiled for 10 min, dried, resuspended in 95% abundance ^{18}O water, and digested with PNGase F overnight. No effort was made to remove the trypsin from this sample prior to deglycosylation. The molecular ion for the peptide appears at 1474.8 (no ^{18}O), 1476.8 (one ^{18}O) and 1478.8 (two ^{18}O). The glycopeptide appears at 1741.7 and 1743.7 indicating the incorporation of either a ^{16}O or an ^{18}O into the Asp formed during deglycosylation.

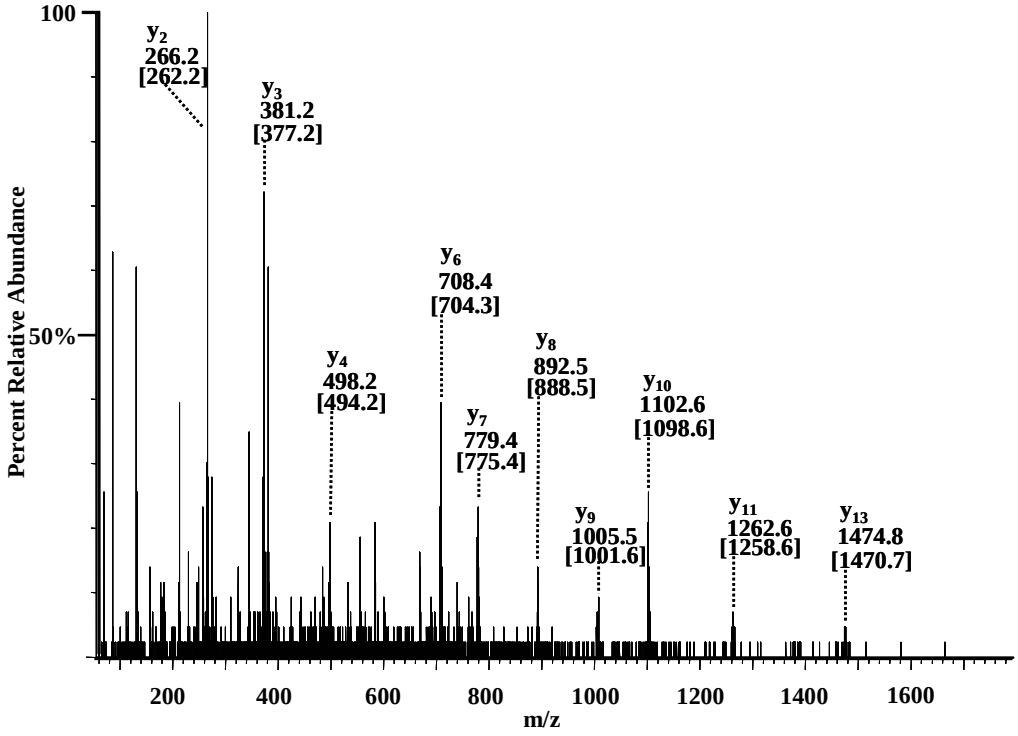


Figure 4.5. Identification of sites of ^{18}O tryptic incorporation on a formerly N-linked glycosylated peptide from transferrin. LC-MS/MS spectrum of the doubly charged (*m/z* = 872.9) N-linked glycopeptide LCPDCPLLAPLN*DSR after PNGase F cleavage in 95% ^{18}O enriched water in the presence of trypsin (i.e., the sample analyzed in Figure 4.3). Bracketed numbers represent theoretical values for fragmentation of the peptide with ^{18}O inclusion only into the Asp produced by deglycosylation.

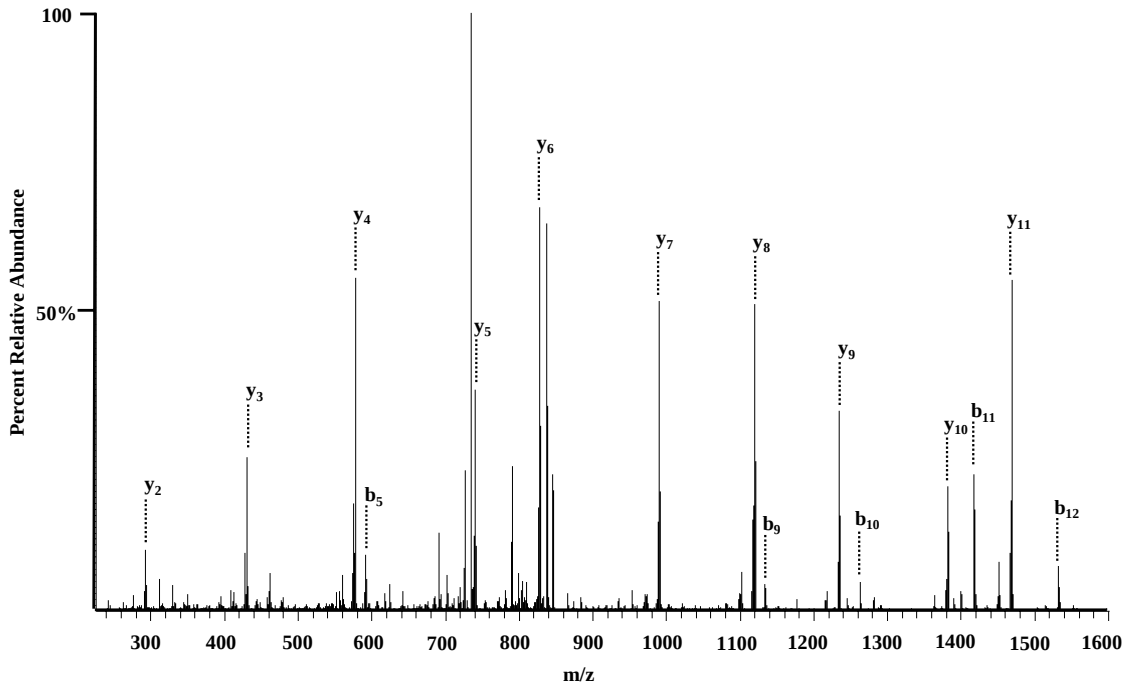


Figure 4.6. MSMS of peptide LQSFDEYSYFHNR, incorrectly identified as having a site of N-linked glycosylation. LC-MS/MS spectrum of a peptide from polygalacturonase inhibiting protein (PGIP) with a doubly charged precursor ion at 855.7 m/z units. Sequest identified this peptide as the formerly glycosylated peptide LQSFDEYSYFHN*R, where * denotes the site of glycosylation. The b- and y-ion assignments resulted from the data base search. In this experiment, the PGIP was trypsin digested, desalting on a C_{18} column, eluted with 0.1% triflouroacetic acid in 80% acetonitrile, dried, resuspended in 95% abundance ^{18}O water, and digested with PNGase F overnight.

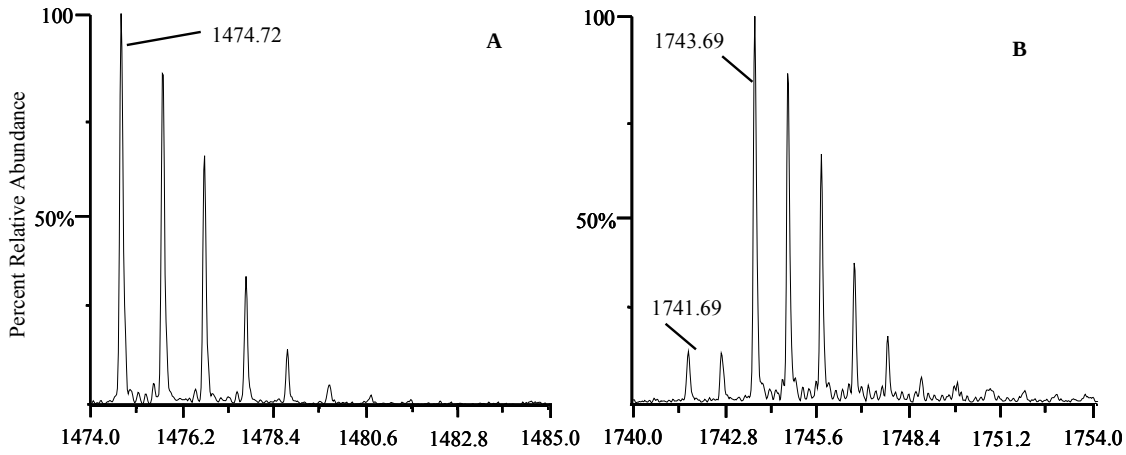


Figure 4.7. Removal of the tryptically applied ^{18}O tag after re-suspending in 16O water. MALDI-ToF spectrum of trypsin digested bovine fetuin showing (A) a representative tryptic peptide **(**TPIVGQPSIPGGPVR) and (B) a representative N-linked glycopeptide (LCPDCPLLAPLN*****DSR) after PNGase F cleavage in 95% ^{18}O enriched water in the presence of trypsin (i.e., the sample analyzed in Fig. 4.3) followed by resuspension in natural abundance water. The molecular ion for the peptide appears at 1474.8 (zero ^{18}O), indicating complete exchange of ^{18}O on the C-terminus with ^{16}O after this treatment. The glycopeptide appears at 1741.7 and 1743.7 resulting from incorporation of either a ^{16}O or an ^{18}O into the Asp formed during deglycosylation.

CHAPTER 5

CARBAMYLATION AS A STABLE ISOTOPIC LABELING METHOD FOR QUANTITATIVE PROTEOMICS[3]

[3] Angel, P., Orlando, R.
In press, *Rapid Communications in Mass Spectrometry.*

ABSTRACT

A method was developed that uses urea to both solubilize and isotopically label biological samples for comparative proteomics. This approach uses either light or heavy urea ($^{12}CH_4{}^{14}N_2O$ or $^{13}CH_4{}^{15}N_2O$, respectively) at a concentration of 8 M and a pH of 7 to dissolve the samples prior to digestion. After the sample is digested using standard proteomic protocols and dried, isotopic labeling is completed by resuspending the sample in a solution of 8 M urea at a pH of 8.5, using the same isotopic species of urea as used for digestion and incubating for 4 hours at 80°C. Under these conditions, carbamylation occurs only on the primary amines of the peptides. The effects of complete carbamylation on MALDI-TOF MS and ESI-MS/MS (CID) were examined. Peptides that had a C-terminal carbamylated lysine residue were found to have a reduced intensity when viewed by MALDI-TOF MS. Collision induced dissociation of a tryptic peptide that was carbamylated on both the N-terminus and the C-terminus was found to have a more uniform distribution of b- and y-ions, as well as prominent ions from loss of water. Reverse-phase chromatography coupled to ESI-MS/MS was used to identify and quantify the isotopically labeled standard proteins, bovine serum albumin (BSA), bovine transferring, and bovine alpha casein. Quantitative error between theoretical and observed data ranged from 1.7 – 10.0%. Relative standard deviations for protein quantitation ranged from 5.2% – 21.5% over a dynamic range from 0.1 – 10 (L/H). The development of a method utilizing urea-assisted carbamylation of lysines and N-termini to globally labeled samples for comparative proteomics may prove useful for samples that require a strong chaotrope prior to proteolysis.

INTRODUCTION

The field of comparative proteomics strives to provide a means to study the up and down regulation in protein expression caused by environmentally or biologically driven changes to an organism. Shotgun proteomic approaches for comparative analyses between differential biological states are performed by measuring the relative abundances of proteins either by label free[1-5] or stable isotopic labeling methods.[6,7] Label free techniques employ a parallel work flow between the two biological states, and use aspects of the eluting peak such as normalized ion intensities,[1] spectral counts,[2,3] mass, scan number and signal intensity,[4] or accurate mass plus retention time[5] to gain semi-quantitative information from the peptide or protein populations. Stable isotopic labeling schemes involve labeling each population with either a light or a heavy stable isotope and combining the two proteomes for simultaneous analysis.[6,7] Relative quantification between the isotopically labeled populations is performed by taking a ratio of the area or the intensity of the light and heavy monoisotopic peaks from the MS or MS/MS. Co-elution of the stable isotopically labeled species has a particular advantage for mass spectrometric analysis in that it ensures that the two isotopic species have similar ionization efficiencies, allowing for more accurate quantification. Both of these strategies begin with the solubilization and denaturation of the protein pellet.

Shotgun proteomic approaches to dissolve and disrupt protein structure have employed organic solvents, acid-labile detergents, chaotropes, or a combination of detergents and chaotropes.[8-13] While detergents have been shown to interfere with mass spectrometry based methods[13-16], they may be used successfully in proteomic experiments if they are removed prior to LC-MS.[17-19] Chaotropes such as urea or guanidine hydrochloride are excellent solubilizing agents that may be easily removed by reverse phase chromatography, thus fitting well with the commonly used system of reverse phase coupled with mass spectrometry. Urea in particular has proven to be especially useful for dissolving and producing efficient digestions of even the most difficult of samples, for instance membrane[13] and lipid raft protein preparations.[20] However, a drawback to the use of urea for the denaturation and solubilization of a

protein pellet is that cyanic acid, a decomposition product of urea, reacts readily with several functional groups on proteins.[21-25] While the formation of cyanic acid may be temporarily overcome through the use of ion exchanger resins to remove the species from solution, urea readily produces cyanic acid when heated, as discussed below. Thus, during the course of a proteomic experiment, when the sample is incubated at elevated temperatures during disulfide bond reduction for instance, cyanic acid will form and may react with protein functional groups.

The presence of cyanic acid in a solution of urea was first reported by Walker and Hambley as an equilibrium between urea and ammonium cyanate in aqueous solutions [Figure 5.1A].[26] When the solution is heated even mildly (e.g. 37°C) or allowed to age, the reaction is driven towards the breakdown of urea and the formation of cyanate.[25,26] Once formed, cyanic acid is susceptible to nucleophilic attack by a variety of functional groups depending upon the pH of the solution.[21] At a pH of 5 or below, cyanate is capable of reacting with the carboxyl,[21,24] sulfhydryl,[21,23,29] or phosphate[30] moieties. At neutral to slightly basic pHs (i.e., pH 7-9), cyanic acid reacts exclusively with primary amines, resulting in the formation of a stable carbamylamino group [Figure 5.1B].[21,24,28] For proteins, the reactivity of a primary amine with cyanic acid depends not only upon the pH of the solution but also on the location of the primary amine. Carbamylation of the α-amines of the N-termini has been shown to be approximately 100 times faster at a neutral pH than at the aliphatic ε-amines of the lysine residues.[21, 24] As the basicity of the solution is increased to a pH of 9, the ε-amines are deprotonated and become accessible to carbamylation.[21] Hence, the reaction of residual isocyanate often results in partial carbamylation of the primary amines in the peptide or protein population in protocols where a solution of urea is used to dissolve and denature the sample. The field of proteomics has been aware of this reaction, therefore when urea is used as a denaturant, the variable modification of carbamylation (+CHNO, Δm = 43.006 Da) is allowed for when performing database searches.[31,32] However, for comparative proteomics, a consequence of this modification is that partial (or complete) carbamylation blocks N-terminal amines as well as the ε-amines of C-terminal lysines, sites that could potentially be used to apply a stable isotopic

tag. Another downside to this reaction is that the peptide population is split among the non-carbamylated, partially carbamylated, and fully carbamylated species. These peptides will elute at different times during a chromatographic separation, potentially causing changes in ionization efficiencies. The partial modification also divides the peptide concentration into multiple peaks, diluting the overall signal of the eluting peptide thus decreasing its detectability. Carbamylation occurring during sample processing is thus detrimental to both label free and stable isotopic labeling comparative proteomic strategies.

The high concentration of urea used for dissolution of protein samples combined with the ready reaction of the byproducts of urea degradation with primary amines poses an interesting prospect for a new isotopic labeling strategy. The use of urea as a stable isotopic labeling method would cause the modifications that naturally occur during digestion in this medium to become a part of the isotopically labeled population. Isotopically labeling with urea would thus produce a homogeneously labeled population and provide an alternative, facile, and high-throughput method for comparative proteomics. In this work, we establish a procedure to provide full carbamylation of peptides, determine the effect of this modification on peptides observed by MALDI-TOF MS and ESI-MS/MS, and investigate quantitation of carbamylated peptides by ESI-MS.

EXPERIMENTAL

Materials and reagents

Ammonium bicarbonate, α-cyano-4-hydroxycinnamic acid (CHCA), dithiothreitol (DTT), iodoacetamide (IDA), tris-HCl, trifluoroacetic acid (TFA), the protein standards bovine serum albumin (BSA), bovine (apo)-transferrin and bovine alpha casein, along with the peptide standard melittin were purchased from Sigma (St Louis, MO, USA). Sequencing grade trypsin was purchased from Promega (Madison, WI, USA). Light urea ($^{12}CH_4{}^{14}N_2O$), formic acid, and acetonitrile were purchased from J.T. Baker Corporation (Phillipsburg, NJ, USA). Heavy urea ($^{13}CH_4{}^{15}N_2O$) was purchased from Cambridge Isotope Laboratories, Inc. (Andover, MA, USA).

Digestion and labeling procedure

A typical protocol for digestion and labeling of samples is provided using (apo)- transferrin as a model protein. Urea for digestion and urea for labeling are prepared separately as control of pH is the critical factor for minimizing carbamylation of lysine residues before proteolysis. In all cases, the urea solution was freshly prepared prior to use.

Approximately 10 nmol of standard protein was dissolved in 100 µL light isotopic (^{12}C, $^{14}N_2$) 8 M urea/200 mM Tris-HCl, pH 7.4. A separate solution of approximately 10 nanomoles of the same protein was dissolved in 100 µL heavy isotopic (^{13}C, $^{15}N_2$) 8 M urea/200 mM Tris-HCl, pH 7.4. Each protein solution was reduced with 20 mM DTT for 2 h at 50°C, followed by carbamidomethylation with 45 mM IDA at room temperature for 1 h. The denatured, reduced, and alkylated protein solutions were diluted by 7 parts with 50 mM ammonium bicarbonate to reduce the concentration of urea to 1 M. Digestion proceeded overnight at 37°C. After digestion, the solutions were dried under vacuum at 37°C for 45 min.

The dried light isotopically labeled solution was brought up in 300 µL light isotopic (^{12}C, $^{14}N_2$) 8 M urea/ 200 mM, Tris-HCl pH 8.5. The separate solution that was originally dissolved in heavy isotopic urea was brought up in heavy isotopic (^{13}C, $^{15}N_2$) 8 M urea/ 200 mM, Tris-HCl pH 8.5. The addition of more solvent was sometimes required to fully solubilize all of the urea. This was accomplished by the adding 200 mM Tris, pH 8.5, in a drop wise fashion. After vortexing for 45 seconds, the samples were inspected for solid particulates. If solids were observed, more solvent was added until all solids were dissolved. Drying down the sample that was dissolved in the initial volume of urea and then adding more urea produced more complete labeling results than by simply drying down the initial sample and bringing it back up in buffer. After ensuring complete solubilization, both samples were incubated simultaneously for 4 hours at 80°C, with periodic vortexing of the samples. After labeling, samples were desalted for mass spectrometry analysis.

Stability of the carbamylated amine

A stock solution consisting of 100 pmole/µL BSA digest was carbamylated with heavy urea ($^{13}CH_4{}^{15}N_2O$) using the labeling method described above. To examine the long term pH-dependent stability of the carbamylated amine, the stock solution of labeled BSA peptides was diluted 1:10 with varying concentrations of formic acid to attain a pH of 2 or 7 and stored at 4°C. Stability of the labeled peptides was investigated at T_0, 2 weeks, 1 month, and approximately every 3 months, up to an endpoint of 15 months.

To examine the long term stability of the label in 8 M urea/200 mM Tris, pH 8.5, a portion of the 100 pmol/µL stock solution of heavy isotopically carbamylated BSA peptides was diluted 1:20 in 8 M light urea ($^{12}CH_4{}^{14}N_2O$)/200 mM Tris, pH 8.5 and stored at room temperature. This sample was examined was examined at T_0, 1 day, 5 days, 2 weeks, 5 months and 7 months. At the specified time points, a portion of each sample was withdrawn, desalted and analyzed by MALDI-TOF MS as described below. Spectra recorded at each time point were examined for mass shifts and the appearance of new ion masses relative to T_0.

Preparation of standards for the calibration curves

The standard proteins bovine serum albumin and bovine transferrin were digested and labeled as described above. This resulted in four solutions, a light and heavy isotopically labeled solution of BSA digest, and a light and heavy isotopically labeled solution of bovine transferrin. Each solution was desalted using C_{18} spin columns [Nestgroup, Southborough, MA]. After desalting, samples were dried under vacuum and brought up in 200 µL of 0.1% formic acid. Absorption measurements at 280 nm for each solution were used as an aid to estimate concentration prior to combining ratios for each curve. Total protein concentration for each ratio was approximately 15 pmol/µL.

Preparation of a simple mixture of standard proteins

Separate solutions of three standard proteins, alpha casein, BSA, and bovine transferrin were prepared. Each protein was dissolved in ammonium bicarbonate to a concentration of 4 nmol/μL, 250 pmol/μL, and 25 pmol/μL, respectively. For the light isotopically labeled solution, 5 μL of alpha casein, 16 μL of BSA, and 32 μL of transferrin was placed in a tube and dried under vacuum. For the heavy isotopically labeled solution, 5 μL of alpha casein, 48 μL of BSA and 4 μL of transferrin was placed in a tube and dried under vacuum. Each solution was treated according to the digestion and labeling procedure described previously. After labeling, the separate solutions were combined and desalted on an analytical reverse phase column (Jupiter C18, Phenomenex, Torrance, CA) and lyophilized to dryness. The combined light and heavy populations were then brought up in 100 μL of 0.1% formic acid. A portion of this was diluted to 1:6 prior to reverse phase analysis. Total protein concentration for the analyzed solution was approximately 33 pmol/μL.

MALDI-TOF MS analysis

A portion from each of the samples that were to be analyzed by MALDI-TOF MS was diluted to 1 pmol/μL. Samples were desalted by ZipTip ^{TM}, following the manufacturers protocol with the exception that desalted peptides were eluted from the ZipTip ^{TM} with the CHCA matrix prepared in 50% (v/v) acetonitrile/0.1% (v/v) TFA. All MALDI-TOF MS was performed on an ABI 4700 Proteomics Analyzer (Applied Biosystems, Foster City, CA) equipped with an Nd:YAG laser operating at 355 nm with a 200 Hz laser rate. All samples were acquired in the positive ion reflector mode with an acquisition range from 900 - 3000 *m/z* and a focus at 1800 *m/z*. Each spectrum was an accumulation of 1000 shots obtained with a laser setting of 3600, accelerating voltage of 20 kV, source chamber pressure of 6.0×10^{-8} torr, and a reflector chamber pressure of 2.0×10^{-8} torr. External calibration was performed using four standards: des-arg^1-bradykinin (904.468), angiotensin I (1296.685), Glu^1-fibrinopeptide B (1570.678), and

neurotensin (1672.92) (singly charged monoisotopic species denoted). Calculation of light to heavy isotopic ratios was performed manually for each isotopically labeled peptide pair. After accounting for the overlapping isotopic envelope from the light isotopic ion by inputting the peptide sequence into MS-Isotope [http://prospector.ucsf.edu] and utilizing the peptide sequence option with the appropriate modifications, total ion counts for the area under each isotopic peak were used to calculate the light/heavy isotopic ratios.

CID of the tryptic peptide VLTTGLPALISWIK

Two solutions of the tryptic peptide VLTTGLPALISWIK were prepared to investigate the influence of carbamylation on the collision induced dissociation spectra of a peptide. The unmodified peptide was dissolved in 50/50 methanol, 1% acetic acid at a concentration of 3 pmol/µL. The doubly carbamylated species of the same peptide was produced by incubating the peptide in a solution containing 8 M urea ($^{12}CH_4{}^{14}N_2O$), pH 8..5 for 2.5 h at 80°C. After carbamylation, the peptide was desalted on a spin column and dried under vacuum. The dried peptide was resuspended in 50/50 methanol, 1% acetic acid at a concentration of 3 pmol/µL and directly infused into a Q-TOF II [Micromass, Wythenshawe, UK]. Collisionally induced fragmentation was performed on the precursor ion 756.46^{+2} for the unmodified peptide and 799.46^{+2} for the doubly carbamylated peptide. The nanospray capillary voltage was set at 3.2 kV and the cone voltage was set at 40 V. Collision gas was maintained at 10 psi. Collision energy for the unmodified peptide was found to be optimal at 33 eV, while the collision energy for the fully carbamylated peptide was found to be optimal at 24 eV. Source temperature was maintained at 80°C.

ESI MS/MS analysis

Samples for construction of the calibration curve and the simple mixture of proteins were analyzed by reverse phase LC coupled to a Q-TOF II [Micromass, Wythenshawe, UK]. This mass spectrometer is equipped with a nanoelectrospray source and was operated in positive-ion mode, with a MS acquisition range of 400-2000 *m/z*. The nanospray capillary voltage was set at 3.2 kV and the cone voltage was set at 40 V. Collision gas was maintained at 10 psi, with collision energy at 10 eV. Source temperature was

maintained at 80°C. Mobile phase A was water containing 0.1% formic acid and mobile phase B was acetonitrile containing 0.1% formic acid. Peptides were eluted from a 150 μm x 150 mm C18 column [Grace Vydac, Hesperia, CA] using a linear gradient of 5-60% B with a flow rate of 1 μL/min. Spectra were collected and processed using Masslynx v3.5. Identification of peptides was performed using the Mascot algorithm [www.matrixscience.com], allowing for the variable modifications as stated previously. Total ion chromatograms for peptides identified at a ≥ 95% confidence level were extracted and combined to obtain the spectra for quantification of each ion pair. Calculation of light to heavy isotopic ratios was performed manually for each isotopically labeled peptide pair. The overlapping isotopic envelope from the light isotopic ion was accounted for by inputting the peptide sequence into MS-Isotope [http://prospector.ucsf.edu] and utilizing the peptide sequence option with the appropriate modifications. Total ion count for the area under each isotopic peak was used to calculate the light/heavy isotopic ratios. Isotopic contributions from unidentified peptides were calculated using an averagine mass.

RESULTS AND DISCUSSION

The widespread use of urea to solubilize proteins combined with the apparent ease of carbamylation led us to investigate the use of isotopically enriched urea as a derivatization reagent for quantitative proteomics. Prior to attempting to quantify actual protein populations, a stable isotopic labeling strategy was developed for use in the context of a proteomic work flow. This strategy utilizes urea to denature and solubilize the protein pellet as well as to incorporate a stable isotopic label into the peptide population.

The labeling strategy begins by dissolving the protein pellet in 8 M of either light or heavy isotopic urea at a neutral pH, followed by reduction and alkylation. At a neutral pH, unprotonated ε-amines are limited in concentration, and the chief reaction that occurs is the modification of the α-amines of the N-termini.[21, 24] Since carbamylation of the ε-amines of lysine residues will block subsequent proteolysis at those sites, pH is maintained at a neutral pH throughout reduction, alkylation and digestion. When reduction and alkylation are complete, the sample is diluted to decrease the concentration of urea to

< 1 M for tryptic digestion.[11, 12] After digestion, the sample is dried to reduce the volume and resuspended in 8 M urea prepared at a pH of 8.5, utilizing the same isotopic form of urea that was initially employed to dissolve the protein pellet. The increase in pH deprotonates the primary amines on the lysine residues, thus ensuring reaction with both the α-amines of the N-termini and the ε-amines of the lysine residues. The peptides are then isotopically labeled by incubating for 4 h at 80°C. The high temperature drives the breakdown of urea [Figure 5.1A],[26, 27] thus increasing concentration of the isocyanate available for production of carbamylated peptides. Light or heavy isotopically carbamylated peptides are produced via nucleophilic attack by available primary amines on the carbon atom of cyanate [Figure 5.1B]. This adds a carbamyl group (chemical composition CHNO, Δm = 43.006 Da for the light label) to each primary amine. After incubation, the light and heavy isotopic populations are combined in a one-to-one ratio, desalted, and analyzed by MS and MS/MS for identification and relative quantification between the light and heavy isotopic pairs.

Carbamylation of peptides

The ability to fully carbamylate a peptide was first evaluated by incubating a standard peptide, melittin (GIGAVLKVLTTGLPALISWIKRKRQQ, monoisotopic mass 2845.74), in 8 M urea at pH 8.5 for 2.5 h at 80°C. Melittin has four primary amines, including three internal lysine residues, thus providing a "worst case" scenario and an example of internal as well as N-terminal formation of carbamylamino species. A comparison of the spectra obtained before and after derivatization reveals a total mass difference of 172 Da, consistent with the predicted mass shift of four carbamylated residues. Furthermore this comparison does not reveal peaks from any under carbamylated species (Figure 5.2A and 5.2B). Since noncarbamylated peptides have a greater intensity than their carbamylated counterparts when analyzed by MALDI-TOF MS, as discussed below, the lack of ions from under carbamylated species in Figure 5.2B indicates that this reaction has gone to completion. The lower abundance ions at *m/z* values below the MH^+ of melittin (*m/z* = 2846.74) are impurities associated with the standard solution and are observed in both of these spectra. The ability to fully carbamylate melittin displays the utility of the

labeling procedure for situations such as tryptic peptides that contain a missed cleavage, or where a protease other than trypsin or Lys-C is used to cleave the polypeptide chain.

To investigate complete carbamylation on a protein digest, a control solution of tryptically digested unmodified BSA peptides was prepared and analyzed by MALDI-TOF MS. Aliquots of this digest were also incubated at 80°C in 8 M light ($^{12}CH_4{}^{14}N_2O$) or heavy urea ($^{13}CH_4{}^{15}N_2O$), pH 8.5, for 4 h, which was necessary to produce consistent carbamylation of this mixture. The MALDI-TOF MS spectra of each these solutions are shown in Figures 5.3A, B, and C. Peptides from each of these samples were assigned by comparing the mass to charge values obtained by MALDI-TOF MS with an *in silico* digest of this protein. A list of peptides that could be assigned in each spectrum allowing for variable modifications at the N-terminus and the lysine residues is presented in Table 1. Twelve peptides in the mass range 927-1749 *m/z* were identified across all three spectra on the basis of mass accuracy (<200 ppm). Further confidence in identification of peptide masses could be gained by calculating the number of heavy isotopes incorporated and comparing that with the theoretical number of labels that should be applied for each peptide according to sequence. Evaluation of the list of labeled peptides revealed that seven of these peptides contain either a C-terminal or internal lysine residue. Five other peptides terminate in arginine and, therefore, were only N-terminally labeled. One of these arginine containing peptides (Peptide 19) has an N-terminal lysine residue (Table 5.1). Both the lysine and the N-terminal amine of this peptide are carbamylated. The apparent full carbamylation of these peptides, as indicated by the number of heavy isotopes incorporated, demonstrates the utility of this method towards universal incorporation of a stable isotope into a mixed peptide population.

Effect of carbamylation on peptides analyzed by MALDI-TOF MS

Comparison of peptides from a control (non-carbamylated) BSA digest to the carbamylated digest by MALDI-TOF MS indicated a change in the relative peptide intensity between these samples. An initial observation of the non-carbamylated digest showed more peptides (21 peptides) compared to the

carbamylated digest (12 peptides). An examination of Table 1 revealed that the missing peptides, which were seen in the unmodified digest, have C-terminal lysine residues. In addition, the peptides with C-termini lysine residues that are observed in the MALDI spectrum of the carbamylated digest (peptides 8, 9, 10, 11, 13) appear to have decreased ion intensity after carbamylation relative to the control (Figure 5.3). For instance**,** peptide 8 has a 25% decrease in intensity after carbamylation, relative to the peptide 16, the base peak. Likewise, other peptides with a C-terminal lysine demonstrate a decrease in intensity after carbamylation: peptide 9 (50% decrease), peptide 10 (35% decrease), peptide 11 (40% decrease), and peptide 13 (10% decrease). These observations are consistent with the decrease in basicity of the lysine residues after carbamylation. A similar decrease in ion intensity and missing peptides has been reported when comparing lysine terminating peptides to arginine terminating peptides by MALDI analysis.[33] In this report, the preferential observation of arginine terminating peptides versus lysine terminating peptides was attributed to the higher basicity of arginine. We believe this same explanation can be used when comparing unmodified lysines versus carbamylated lysines. Arginine terminating peptides (peptides 1,5,7,16,19,21) are modified only at the N-terminus and appear to have similar ion intensities when compared to the unmodified digest (Figure 3). This labeling method may thus be most useful, at least in relation to MALDI ionized samples, when Arg-C digests are performed.

Fragmentation of doubly carbamylated peptides

The influence of carbamylation on collision induced dissociation (CID) of a peptide was examined. Only a few reports exist on the effect of carbamylation on the CID of a peptide and these are limited to species that are carbamylated on either the C-terminus or the N-terminus.[31,32] Fragmentation of doubly carbamylated tryptic peptides was initially investigated using the tryptic peptide VLTTGLPALISWIK, monoisotopic mass 1510.92, a peptide with potential sites for carbamylation on both the N-terminus and the C-terminus. Comparison of this peptide before and after carbamylation is shown in Figure 4. A few things may be noted from the CID spectra. In the MS/MS spectrum of the unmodified peptide, the y-ion series represents 83% of the identified fragment ions, while in that obtained

from the doubly carbamylated peptide, the y-ion series composes 60% of the identified fragment ions. This change appears to result from an increase in the number of detectable b-ions (Fig. 5.4). This tendency toward a more pronounced b-ion series after carbamylation was also seen in the LC-MS/MS spectra obtained from analysis of the digested and labeled protein standards used for the quantification studies. (Fig. 5.4). These observations may result from a reduction in the charge localization, causing a more uniform distribution of b- and y-ions.[36] This rationale is consistent with the observation that peptides terminating with a Lys residue have reduced relative intensities in MALDI-TOF MS analysis, discussed above. It is apparent from this example and others[34,35] that carbamylated peptides that are fragmented by ESI-CID yield sequence rich information.

Stability of the carbamylated amine

Previous reports have investigated the stability of the carbamylamino groups under strenuous acidic and alkaline conditions. [21,22,28, 38] The ε-carbamylamino group has been reported as being resistant to acidic hydrolysis, with approximately 24% of the carbamyl group cleaved after incubation in 6 N HCl at 110°C for 22 h.[22,28] Both the α- and ε-carbamylated amines have been reported as being stable when stored for several months at -20°C in 50% acetic acid.[38] In dilute alkaline situations, carbamylaminos appear to be susceptible to hydrolysis.[21,28,38] Long-term stability of the carbamyl modification was re-investigated here using solvents that are typically seen in current proteomic separation protocols. A solution of heavy carbamylated BSA peptides was stored at 4°C at a pH of 2 and 7 for 15 months. At specified time points, a portion of the sample was withdrawn, desalted, and analyzed by MALDI-TOF MS for mass shifts relative to T_0, data not shown for brevity. Specifically, the MALDI-TOF MS spectra were examined for the loss of one or more of the heavy carbamyl groups ($^{13}CH^{15}NO$, $\Delta m = 45$) and the emergence of new ion masses. These solutions showed no observable change in the MALDI-TOF MS spectra, indicating that there was no loss of or reaction with the stable isotopic label for well over 1 year (15 months) at a pH of 2 or a pH of 7.

A similar experiment was performed with the heavy carbamylated BSA peptides incubated in 8 M light urea, pH 8.5 at room temperature. This mimicked a situation where samples are stored in 8 M urea after carbamylation, or where sample populations have been combined prior to labeling. At the time points indicated in the experimental section, MALDI-TOF MS spectra were collected and inspected for the loss of the stable isotopic label as well as examined for new mass shifts that would indicate further modification of the peptides. There was no observable change in the heavy carbamylated BSA peptide mixture after storing the sample for 15 months under these conditions. Both of these stability studies indicate that there does not appear to be any loss of or reaction with the stable isotopic tag when mixing the two populations prior to separation, when queued on an autosampler, or during the separation processes.

Quantification of carbamylated proteins by ESI- MS and MS/MS

The ability of this stable isotopic labeling method to perform quantitative ESI-MS/MS analysis was investigated using two standard proteins, BSA and bovine transferrin. Each protein was isotopically labeled with either light or heavy isotopic forms of urea as indicated in the experimental section. Two calibration curves were prepared, one from BSA peptides and one from bovine transferrin peptides. BSA was combined in light to heavy stable isotopic ratios of 0.1, 0.2, 0.5, 0.8, 2.0, 3.3, and 10.0. Bovine transferrin was combined in light to heavy stable isotopic ratios of 0.3, 0.5, 1.0, 3.0, 5.0, and 8.0. Coverage for bovine serum albumin after carbamylation ranged from 26 – 37% at each calibration point, while for transferrin coverage ranged from 32 – 41% for each calibration point.

Quantitative labeling was first examined at the peptide level by quantifying the relative abundances of each peptide after identification. Figure 5 demonstrates quantitative labeling at the peptide level using the doubly carbamylated transferrin peptide, ENFEVLCK, and the singly carbamylated peptide ELPDPQESIQR. Both of these peptides show excellent correlation values across the explored

ratios and are in very good agreement with each other. This indicates that the method is capable of providing accurate quantitative information for proteins that are identified by only a single peptide.

Quantitation at the protein level was performed by taking an average of all peptides at a specified light and heavy isotopic ratio for calculation of the relative abundance of the protein. Peptides used for protein quantification typically spanned a *m/z* range of approximately 550-2500, and were comprised of singly, doubly and triply charged species. Figure 5.6 shows the calibration curves and correlation values associated with each standard protein. For BSA, there were 78 observations of peptides over the whole curve yielding an overall error of 8.8% and a relative standard deviation of 21.5%. For transferrin, there were 101 observations of peptides over the whole curve, yielding an overall error 1.7% with a relative standard deviation of 13.5%. For both proteins, the main cause of the high percent relative standard deviation was the difficulty in quantitation of the heavy isotopic abundance when the light isotopic abundance was much higher (≥ 8, L/H). For these situations, the monoisotopic peak from the heavy isotopic species is only a small portion of the overlapping peak from the natural occurrence of ^{13}C in the light isotopic species. For example, the BSA calibration point prepared in a ratio of 10:1 light to heavy isotopic abundance had a calculated relative standard deviation of 27.8%, due to variation associated with calculating the small percent abundance of the heavy isotopic peak. This problem has been indicated in other stable isotopic labeling methods.[38] Although not investigated, this cause of inaccuracy is expected to decrease when a higher mass isotopic variant, such as the $^{13}CH_4{}^{15}N_2{}^{18}O$ form of urea which would result in a 4 Da shift for peptides that do not have a lysine residue, is used, since this would shift the monoisotopic peak from the heavy isotopic labeled population further away from the dominant peaks of the light isotopic envelope. The overall explored dynamic range was from 0.1 to 10 (L/H), which showed a linear range of two orders of magnitude, which is comparable to other stable isotopic labeling methods.[38,39] The quantitation of these two standard proteins may be used as examples of protein quantitation by quantitative carbamylation of primary amines.

Isotopic effect on chromatographic retention

The use of a hydrogen-deuterium pair to isotopically label samples has been reported to produce an "isotopic effect", resulting in partial resolution between isotopic pairs.[7] This phenomenon has the potential to influence the ionization efficiency of the differentially eluting pairs, thus reducing the accuracy of the quantitative procedure. While the $^{12}C/^{13}C$ isotopic pair has not been reported to produce an isotopic shift in LC retention, the $^{14}N/^{15}N$ pair has been associated with a small shift in LC retention times.[7] We investigated the $^{12}C^{14}N$ and $^{13}C^{15}N$ pairs for the presence of an isotopic effect by examining the extracted ion chromatogram for several of the identified peptides and evaluating differences in retention times between each of the isotopically labeled pairs of peptides. Figure 5.7 demonstrates the co-elution of the isotopic pairs for the peptide KNYELLCGDNTR. This peptide has two primary amines, therefore any minor isotopic effects should be more noticeable than in a singly labeled peptide. The light labeled (two carbamyl groups, each $^{12}CH^{14}NO$) peptide and the heavy labeled (two carbamyl groups, each $^{13}CH^{15}NO$) peptides co-elute, each having a retention time of 38.88 min. In all cases investigated, the retention times were identical for isotopically labeled pairs of ions. From this analysis, it appears that the potential isotopic effect due to the use of the $^{14}N/^{15}N$ pair is insignificant and does not affect quantitation of the peptides.

Quantification of a simple mixture of carbamylated proteins by ESI- MS and MS/MS

Carbamylation as a stable isotopic labeling method for simple protein mixtures that vary greatly in abundancies between proteins as well as relative abundancies was examined by preparing two solutions of proteins. Both solutions contained alpha casein, bovine serum albumin, and transferrin in molar abundancy ratios of 5, 1, and 0.2, respectively; thus, bovine serum was five times less abundant than alpha casein and transferrin was 25 times less abundant than alpha casein. For the two mixtures, relative abundance between the light and heavy populations were varied by combining alpha casein in a 1:1 light–to-heavy ratio, bovine serum albumin in a 1:2.85 light to heavy ratio, and transferrin in a 7.5:1 light to heavy ratio. Table 5.2 shows representative peptides from this run as well as the results from this experiment. Multiply tagged peptides were encountered for each protein, including the transferrin peptide

KNYELLCGDNTR that contains a missed cleavage and terminates in a lysine residue. The relative standard deviation of transferrin is higher (15.2%) than the other two more abundant proteins. This may be explained by the lower concentration of this protein in the solution, combined with the previously discussed difficulty in quantitation of isotopic ratios where the light isotope is a large contributor of the monoisotopic heavy labeled peak. As stated above, labeling with a version of urea that contains an additional mass shift, such as the $^{13}CH_4{}^{15}N_2{}^{18}O$ form of urea, would increase the mass shift between light and heavy isotopes, minimizing the variation in peak due to overlapping isotopes. Overall, quantitative efforts of this protein mixture showed good agreement between the theoretical and experimental data.

CONCLUSIONS

The use of urea as a reagent for quantitative proteomics has both advantages and disadvantages. The foremost advantage to the technique is the ability to use a strong chaotrope to solubilize and isotopically label the sample. This aids in efficient digestion and produces comprehensively labeled populations of peptides. A second advantage is that no extraneous sample handling is required to generate the labeled samples. Third, since the labeling is comprehensive over the whole peptide population, more peptides are useful in quantitation of protein levels. Peptides labeled in this manner provide LC-MS/MS spectra rich in complementary b- and y-ion series for sequence identification. Last of all, the reagents required for the labeling process are inexpensive and are readily available in a variety of isotopic combinations.

The downside to carbamylation as a quantitative stable isotopic labeling technique is that carbamylation of the lysines reduces the basicity of these sites, exacerbating a problem inherent in the analysis of peptides terminating in this amino acid.[33] This results in an apparent decrease in signal intensity, demonstrated here on lysine terminating peptides viewed by MALDI-TOF MS. On the other hand, peptides that are carbamylated on the N-terminus and that have an arginine C-terminus do not have a noticeable loss in ion intensity.

In conclusion, the use of urea as a stable isotopic labeling method is a facile process, fitting in well with current proteomic workflows. Quantitative carbamylation of all primary amines has been demonstrated here for the labeling of peptides. This method may prove especially useful for situations that require the use of a strong chaotrope such as the analysis of membrane proteins. Future work will investigate the usefulness of the method towards more complex mixtures of proteins and comparative membrane proteomics.

REFERENCES

1. Wang W, Zhou H, Lin H, Roy S, Shaler TA, Hill LR, Notron S, Kumar P, Anderle M, Becker CH. *Anal. Chem.* 2003; 75: 4818.

2. Liu H, Sadygov RG, Yates JR. *Anal. Chem*. 2004; 76: 4193.

3. Old WM., Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mednoza A, Sevinsky JR, Resing KA, Ahn NG. *Mol. Cell Prot.* 2005; 4.10: 1487.

4. Radulovic C, Jelveh S, Ryu S, Hamilton TG, Foss E, Mao Y, Emili A. *Mol. Cell. Proteom.* 2004; 3: 984.

5. Silva JC, Denny R, Dorschel CA, Gorenstein M, Kass IJ, Li GZ, McKenna T, Nold MJ, Richardson K, Young P, Geromanos S. *Anal. Chem*. 2005; 77: 2187.

6. Tao, WA., Aebersold, RA. *Curr. Opin. Biotechnol.* 2003; 14: 110.

7. Goshe, MG, Smith, RD. *Curr. Opin. Biotechnol.* 2003, 14: 101.

8. Washburn, MP, Wolters D, Yates III JR. *Nat. Biotechnol*. 2001; 19: 242.

9. Blonder J, Conrads TP, Yu LR, Terunuma A, Janini GM, Issaq HW, Vogel JC, Veenstra TD. *Proteomics.* 2004; 4: 31.

10. Ying-Qing Y, Gilar M, Lee PJ, Bouvier ESP, Gebler JC. *Anal. Chem.* 2003; 75: 6023.

11. Pace NC, Grimsley GR, Scholtz M. Protein Folding Handbook 2005; 1: 45.

12. Lee TD, Shively JE. *Methods Enzymol.* 1990; 361-374.

13. Ruth MC, Old WM, Emrick MA, Meyer-Arendt K, Aveline-Wolf LD, Pierce KG, Mendoza AM, Sevinsky JR, Hamady M, Knight RD, Resing KA, and Ahn NA. *J. Prot. Res.* 2006; 5: 709.

14. Börnsen KO, Gass MAS, Buin GJM, von Adrichem JHM, Biro MC, Kresbach GM, Erhat M. *Rapid Comm. Mass Specrom.* 1997; 11: 603-609.

15. Loo RRO, Dales N, Andrews PC. *Protein Sci.* 1994; 3: 1975.

16. Funk J, Li X., Franz T. *Rapid Comm. Mass Spectrom.* 2005; 19: 2986.

17. Lovet, LM, Ouyan,WJ, Eaton, DL, Stults, JT. *J Prot Res*, 2005; 4(2), 400-409.

18. Hansen, KC, Schmitt-Ulms, G, Chalkley, RJ, Hirsch, J, Baldwin, MA, Burlingame, AL. *Mol Cell Prot*, 2003; 2.5: 299-314.

19. Swiderek, KM, Klein, ML, Davis, MT, Hefta, SA, Shively, JE. 1994. Techniques in Protein Chemistry VI, J. W. Crabb, ed., (Academic Press, New York), 267-275.

20. Foster L, Hoog C, Mann M. *Proc. Natl. Acad. USA* 2003; 100: 5813.

21. Stark G. *Methods Enzymol.* 1967; 11: 590.

22. Stark G, Stein W, Moore, S. *J. Biol. Chem.* 1960; 235: 3177.

23. Stark G. *J. Biol. Chem.* 1964; 239: 1411.

24. Stark G. *Biochem.* 1965; 4: 1030.

25. Stark G. *Biochem.* 1965; 4: 2363.

26. Walker J, Hambly FJ. *J. Chem. Soc.* 1895: 67: 746.

27. Dirnhuber P, Schtz F. *Biochem. J.* 1948; 42: 628.

28. Stark G. *Methods Enzymol.* 1967; 11: 594.

29. Lippincott J, Apostol I. *Anal. Biochem.* 1999; 267: 57.

30. Spector L, Jones ME, Lipmann F. *Methods Enzymol.* 1957; 3: 653.

31. Searle BC, Dasari S, Turner M, Reddy AP, Choi D, Wilmarth PA, McCormack AL, David LL, Nagalla SR. *Anal. Chem.* 2004; 76: 2220.

32. Wu CC, MacCoss MJ, Howell KE, Yates JR. *Nat. Biotech.* 2003; 21: 532.

33. Krause E, Wenschuh H, Jungblut PR. *Anal. Chem.* 1999; 71: 4160.

34. Van Driessche G, Vandenberghe I, Jaquemotte F, Devreese B, Van Beeumen JJ. *J. Mass Spectrom.* 2002; 37: 858.

35. Lapko VN, Smith DL, Smith JB. *Protein Sci.* 2001; 10: 1130.

36. Biemann K. *Methods Enzymol.* 1990; VOL: 193.

37. Stark GR, Smyth DG. *J. Biol. Chem.* 1963; 238: 214.

38. Hsu J, Huang S, Chow N, Chen S. *Anal. Chem.* 2003; 75: 6843.

39. Brancia FL, Montgomer H, Tanaka K, Kumashiro S. *Anal. Chem*. 2004; 76: 2748.

Table 5.1: A summary of the peptides observed by MALDI-TOF MS analysis of tryptic digested bovine serum albumin (BSA).

Peptide No.	Peptide Sequence	Unmodified Theor. Mass (MH+)	Observed Mass (MH+)	Theoretical No. Labels	Light Isotopic Label Theoretical Mass (MH+)	Light Isotopic Label Observed Mass (MH+)	Heavy Isotopic Label Observed Mass (MH+)	Calculated No. Labels
1	(K)YLYEIAR(R)	927.49	927.34	1	970.50	970.37	972.41	1
2	(K)DLGEEHFK(G)	974.46	974.32	2	1060.47	1060.33	1064.38	2
3	(K)QTALVELLK(H)	1014.62	1014.62	2	1100.63	-	-	
4	(K)QNCDQFEK(L)	1068.44	1068.29	2	1154.45	-	-	
5	(K)CCTESLVNR(R)	1138.50	1138.34	1	1181.50	1181.36	1183.41	1
6	(K)LVNELTEFAK(T)	1163.63	1163.48	2	1249.64	-	-	
7	(R)CCTKPESER(M)	1166.49	1166.34	2	1252.51	1252.36	-	
8	(R)FKDLGEEHFK (G)	1249.62	1249.46	3	1378.64	1378.48	1384.54	3
9	(K)HLVDEPQNLIK (Q)	1305.72	1305.55	2	1391.73	1391.57	1395.63	2
10	(K)TVMENFVAFVDK(C)	1399.69	1399.51	2	1485.71	1485.51	-	
11	(K)SLHTLFGDELCK (V)	1419.69	1419.52	2	1505.71	1505.54	1509.61	2
12	(K)YICDNQDTISSK(L)	1443.64	1443.47	2	1529.65	-	-	
13	(K)TCVADESHAGCEK (S)	1463.59	1463.41	2	1549.60	1549.45	1553.51	2
14	(R)ETYGDMADCCEK(Q)	1478.52	1478.35	2	1564.54	-	1568.77	2
15	(K)EYEATLEECCAK (D)	1502.61	1502.43	2	1588.63	-	-	
16	(K)VPQVSTPTLVEVSR (S)	1511.84	1511.67	1	1554.85	1554.69	1556.75	1
17	(K)LKECCDKPLLEK(S)	1532.78	1532.60	4	1704.81	-	1712.46	4
18	(K)LKPDPNTLCDEFK(A)	1576.77	1576.59	3	1705.79	-	-	
19	(R)KVPQVSTPTLVEVSR(S)	1639.94	1639.75	2	1725.95	1725.79	1729.86	2
20	(K)YNGVFQECCQAEDK(G)	1747.71	1747.51	2	1833.72	-	-	
21	(K)ECCHGDLLECADDR(A)	1749.66	1749.48	1	1792.67	1792.51	1794.58	1

Table 5.2**:** Quantitative results from the LC-MS analysis of a simple protein mixture composed of BSA, alpha casein, and bovine transferrin.

Protein	**Abundancy (Rel. A-cas)**	**Peptide Sequence**	**Observed Ratio (L/H)**	**Mean ± SD**	**Expected Ratio (L/H)**	**% Error**
Serum albumin (bovine)	0.2	VLTSSAR	0.358	0.355 ± 0.0186	0.331	7.3
		QTALVELLK	0.348			
		SLHTLFGDELCK	0.379			
		VPQVSTPTLVEVSR	0.335			
alpha casein (bovine)	1.0	YLGYLEQLLR	1.06	1.10 ± 0.0613	1.00	10.0
		NAVPITPTLNR	1.17			
		ALNEINQFYQK	1.07			
transferrin (bovine)	0.04	KNYELLCGDNTR	6.44	7.30 ± 1.11	7.54	3.27
		HSTVFDNLPNPEDR	6.90			
		FDEFFSAGCAPGSPR	8.55			

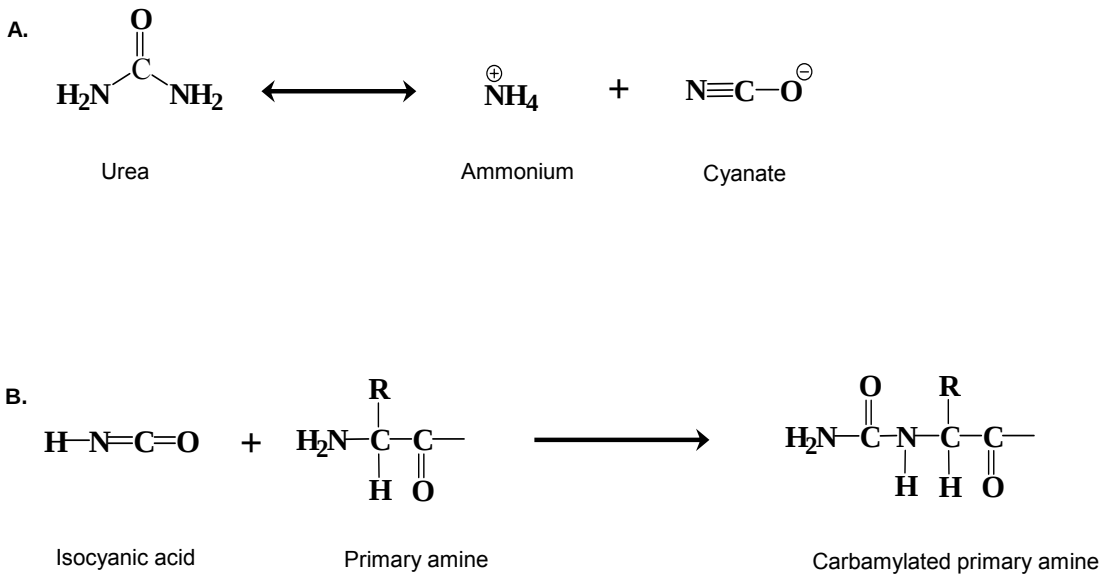
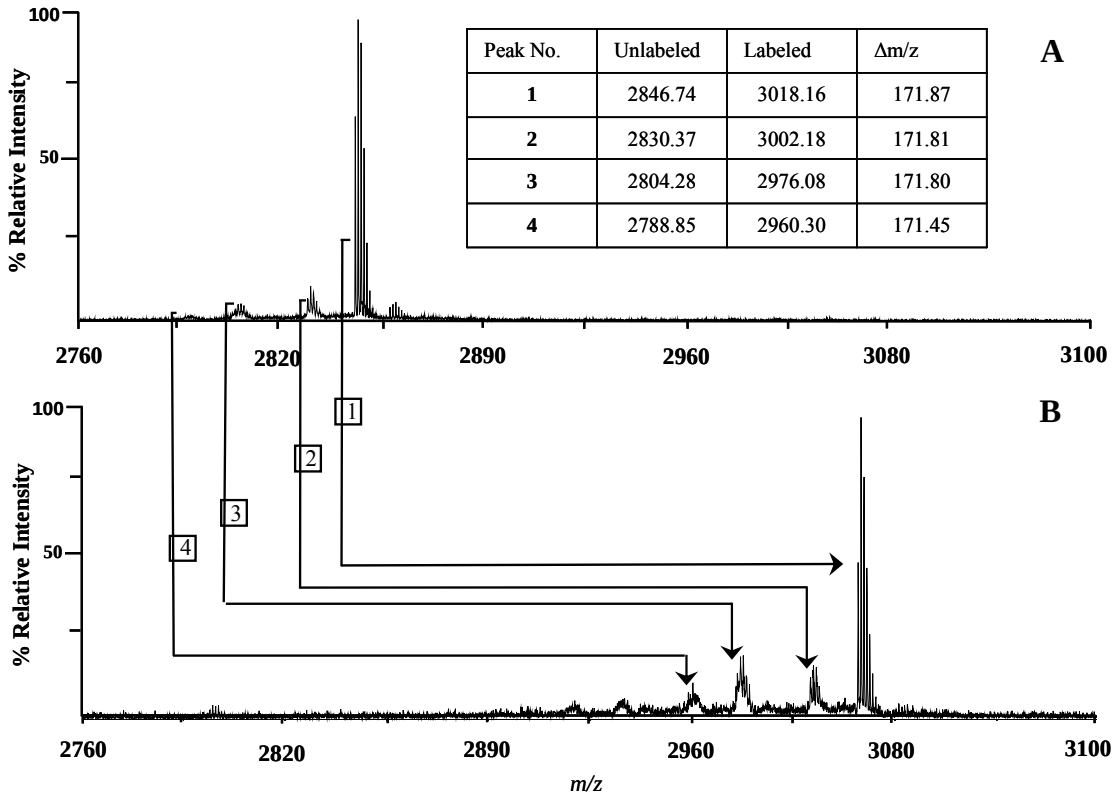


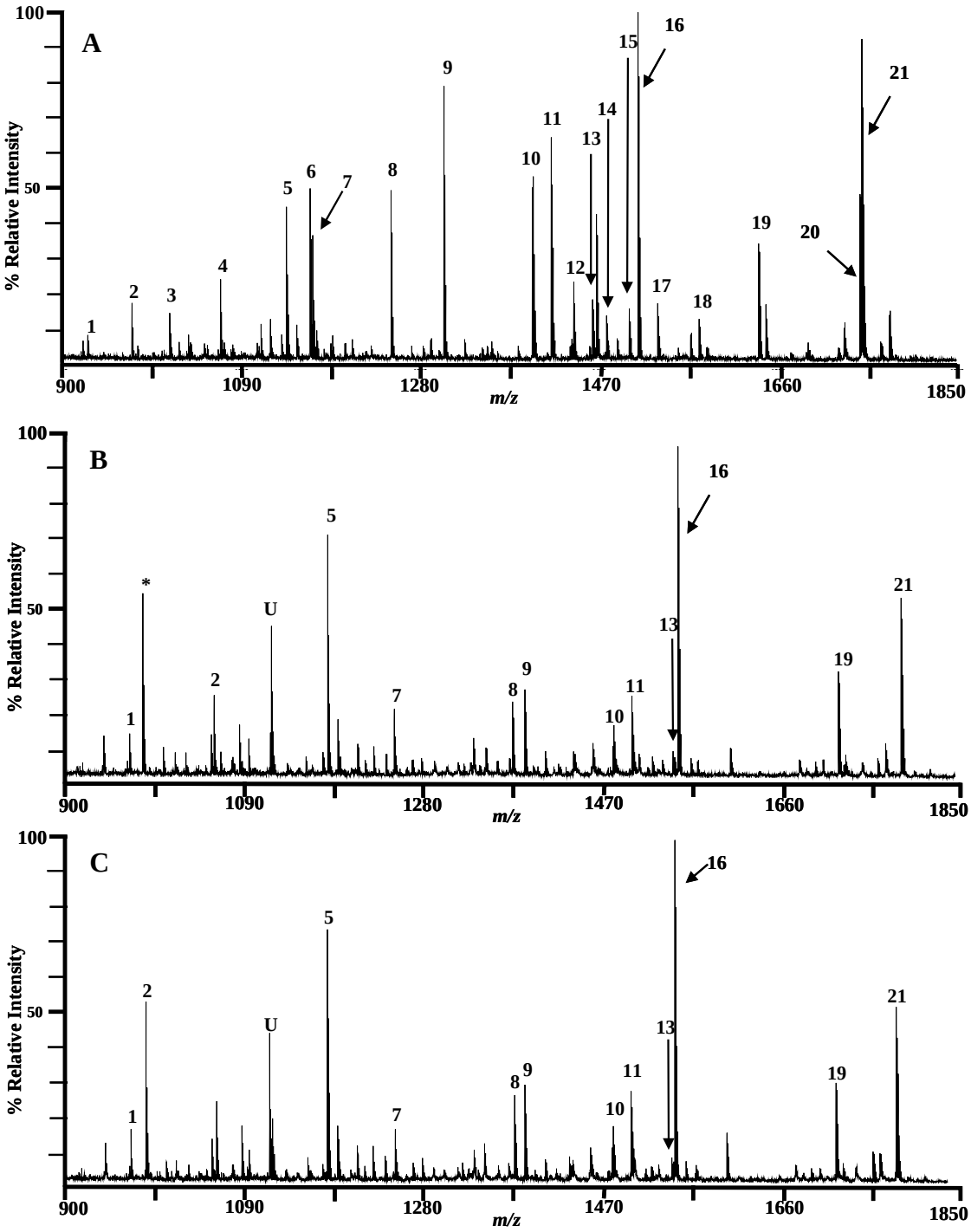
Figure 5.1: Reaction mechanism for the carbamylation of primary amines. (A) Urea is in equilibrium with ammonium and cyanate. Heat and/or time drive the reaction towards the breakdown of urea, causing a buildup of cyanic acid. (B) At neutral to basic pH, the cyanic acid undergoes nucleophilic attack by primary amines forming the carbamylated amine.



Peak No.	Unlabeled	Labeled	Δm/z
1	2846.74	3018.16	171.87
2	2830.37	3002.18	171.81
3	2804.28	2976.08	171.80
4	2788.85	2960.30	171.45

Figure 5.2**:** MALDI-TOF MS analysis of the standard peptide melittin, GIGAVLKVLTTGLPALISWIKRKRQQ, monoisotopic protonated mass 2846.74. Spectrum were obtained from (A) the unmodified peptide, (B) the peptide after undergoing carbamylation with light urea ($^{12}CH_4{}^{14}N_2O$), and (C) the peptide after undergoing carbamylation with heavy urea ($^{13}CH_4{}^{15}N_2O$). Mass shifts: peak 1, Δm = 171.87; peak 2, Δm = 171.81; peak 3, Δm = 171.80; peak 4, Δm = 171.45.

Figure 5.3**:** MALDI-TOF MS analysis of trypsin digested bovine serum albumin (BSA). Shown spectra are from (A) the unmodified sample, (B) the sample carbamylated in light urea ($^{12}CH_4{}^{14}N_2O$), and (C) the sample carbamylated in heavy urea ($^{13}CH_4{}^{15}N_2O$).


A
B
C
100
50
% Relative Intensity
900
1090
1280
1470
1660
1850
m/z
*
U
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21

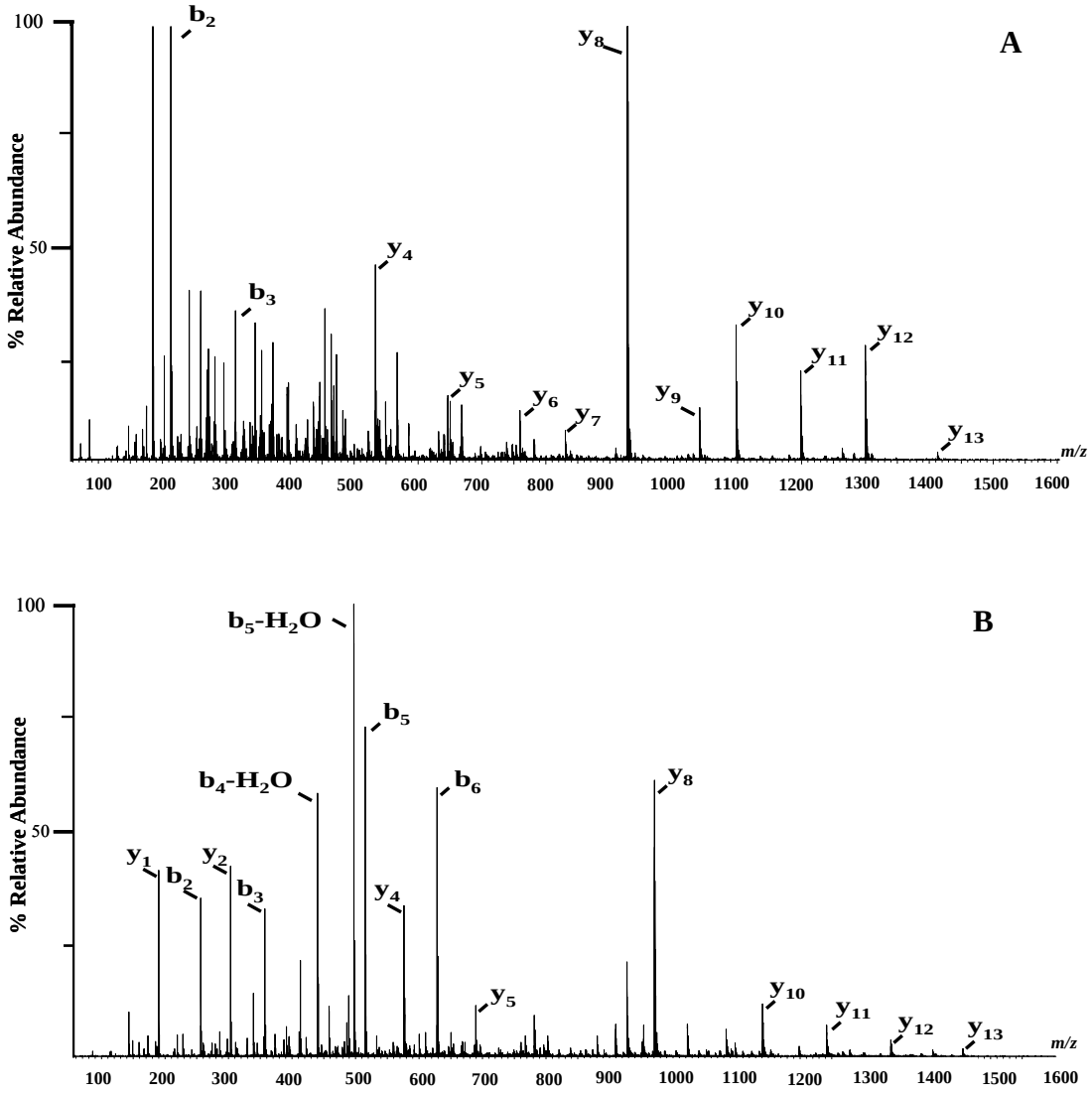
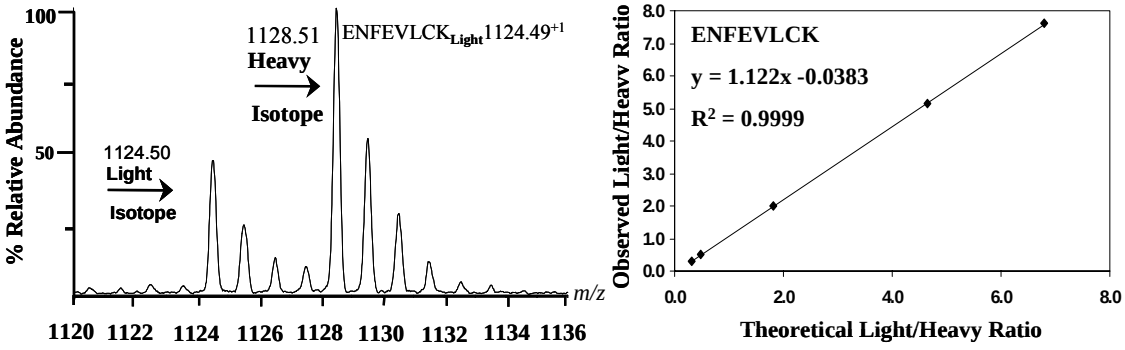
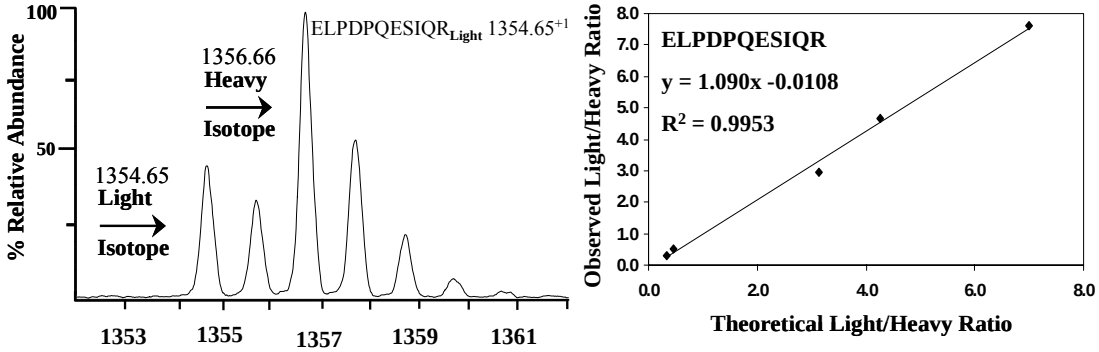


Figure 5.4: ESI-MS/MS spectra from the tryptic peptide VLTTGLPALISWIK. Spectrum (A) was obtained from the unmodified peptide, the precursor was the ($M+2H^{+2}$) ion at 756.46 *m/z*, while spectrum (B) was obtained from the peptide that was carbamylated at both the N- and C-termini, the precursor was the ($M+2H^{+2}$) ion at 799.46 *m/z*.

Figure 5.5. Results from quantitative isotopic labeling at the peptide level. (A) LC-MS spectrum of the fully carbamylated transferrin peptide, ENFEVLCK, the singly protonated monoisotopic *m/z* of the peptide with the light isotopic labeled appears at 1124.49. Since this peptide contains two primary amines, the corresponding ion with the heavy label appears 4 *m/z* higher. The theoretical light to heavy ratio for this peptide was 0.510 (L/H) while the calculated ratio was 0.482 (L/H), a -5.5% deviation from the true value. (B) Calibration curve obtained from ENFEVLCK. (C) LC-MS spectrum of the fully carbamylated transferrin peptide, ELPDPQESIQR, the singly protonated monoisotopic *m/z* of the peptide with the light isotopic labeled appears at 1354.65. This peptide contains a single primary amine, thus the corresponding ion with the heavy label appears 2 *m/z* higher. The theoretical light to heavy ratio for this peptide was 0.510 (L/H) while the calculated ratio was 0.537 (L/H), a + 5.3% deviation from the true value. (D) Calibration curve obtained from ELPDPQESIQR.


100
50
% Relative Abundance
1128.51
Heavy
Isotope
ENFEVLCK$_{Light}$1124.49^{+1}
1124.50
Light
Isotope
m/z
1120 1122 1124 1126 1128 1130 1132 1134 1136
Observed Light/Heavy Ratio
ENFEVLCK
y = 1.122x -0.0383
R^2 = 0.9999
Theoretical Light/Heavy Ratio


100
50
% Relative Abundance
1356.66
Heavy
Isotope
ELPDPQESIQR$_{Light}$ 1354.65^{+1}
1354.65
Light
Isotope
1353 1355 1357 1359 1361
Observed Light/Heavy Ratio
ELPDPQESIQR
y = 1.090x -0.0108
R^2 = 0.9953
Theoretical Light/Heavy Ratio

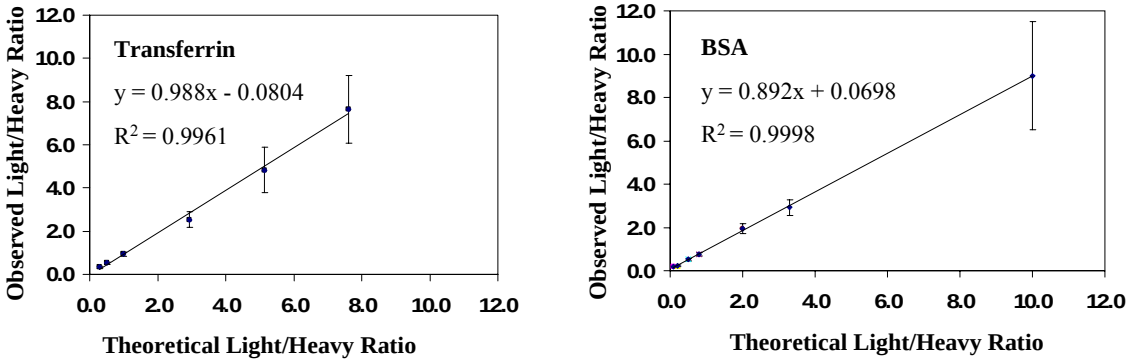


Figure 5.6: Relative quantitation of the standard proteins transferrin and bovine serum albumin (BSA). Each point represents an average of the relative abundancies of 5-14 peptides that have been isotopically labeled by carbamylation of the primary amines.

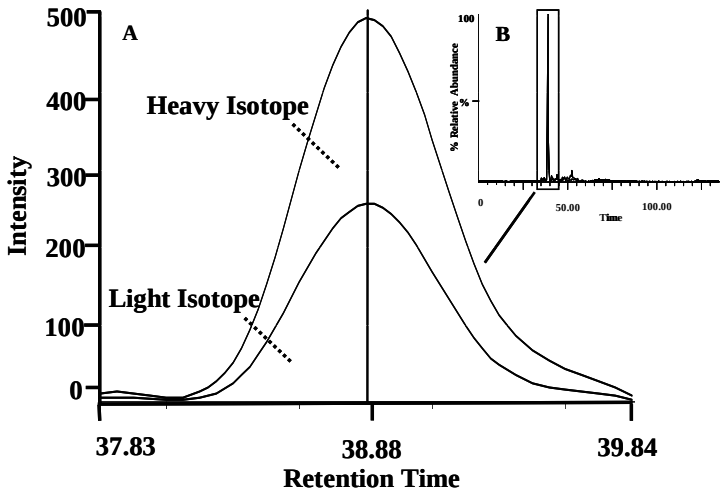


Figure 5.7. (A) Extracted ion chromatograms form the doubly carbamylated transferrin peptide KNYELLCGDNTR labeled with either the light (monoisotopic protonated ion = 1568.71 *m/z*) or heavy urea (monoisotopic protonated ion = 1572.71). The retention time was measured to be exactly 38.88 for both forms of this peptide, demonstrating the co-elution of the isotopic pair. (B) The overlaid extracted ion chromatograms for the two isotopically labeled peaks showing the total run time of the analysis.

CHAPTER 6

COMPARATIVE PROTEOMICS OF DIFFERENTIATING MURINE EMBRYONIC STEM CELLS: AN *IN VITRO* MODEL OF EMBRYOGENESIS[4]

[4] Angel, P., Nash, R., L. Lin, Weatherly, D.B., Dalton, S., Orlando, R.
To be submitted *Journal of Proteome Research*

ABSTRACT

Genetic profiling of early embryogenetic events continues to be well studied using differentiating embryonic stem cells as models to rapidly explore global genetic expression. Protein expression occurring during embryogenesis has been limited to genetic studies that investigate genetic deletion or expression of a particular gene during the course of development. Here, we describe the global protein expression of differentiating murine embryonic stem cells in terms of the events of embryogenesis. Embryonic stem cells (ES), early primitive ectoderm-like cells (EPL) and embryoid bodies (EB) were used as *in vitro* models of the inner cell mass, the primitive ectoderm, and the differentiated cell state, respectively. A two dimensional reverse phase strategy culminating in MS/MS was used to sequence the peptides. Nonredundant peptides were grouped to proteins. Proteins were further grouped according to homology of identified peptides. A total of 3328 mouse proteins (1033 protein homology groups) for ES, 1852 mouse proteins (558 protein homology groups) for EPL, and 2911 mouse proteins (1004 protein homology groups) for embryoid bodies. Uniquely identified proteins were selected as the proteins that had the most complete peptide coverage in the homology group. Protein identifications were sorted according to comparative and single populations, forming a pool of 1498 proteins across the three cell states. Approximately 22% of these proteins were predicted and hypothetical proteins. Average spectral count from the nonredundant peptides grouped to proteins was used to provide quantitive information on protein expression between the differentiated cell states. Literature searches as well as functional analysis using Ingenuity software were able to validate discovered protein expression and regulation with previously explored embryogenetic events. The use of embryonic stem cells, early primitive ectoderm-like cells and embryoid bodies as an *in vitro* model of embryogenesis provides a tool to explore the global protein expression that occurs during early developmental events.

INTRODUCTION

Successful mammalian development depends upon correct execution of the early events of embryogenesis. These events ultimately lead to the formation of both the embryo and the extra embryonic tissues that create the embryonic environment during gestation. In mouse, fertilization leads to the formation of a blastocyst at 4.75 d.p.c.[1, 2], consisting of a trophoblast shell surrounding an inner cell mass. Continued development of the 4.75 d.p.c. blastocyst results in the inner cell mass proliferating and differentiating to form the hypoblast and the epiblast. The trophoblast and the hypoblast will go on to form the extra embryonic tissues of the developing embryo, ensuring embryonic survival in the maternal host. The pluripotent[3] epiblast continues to propagate and forms the primitive ectoderm of the embryonic epiblast. At approximately 6.0 d.p.c., the primitive ectoderm population undergoes a tremendous burst of activity. Cells begin to migrate and differentiate, forming the three primary germ layers of the embryo. These primary germ lineages continue to multiply and follow a directed differentiation path. By 7.5 d.p.c.[2,4], developmental events have resulted in the establishment of the body plan for the developing organism.

The cellular mechanisms that transform the 25-30 cells of the inner cell mass into a complex, multi-cellular functioning organism may be studied *in vitro* using differentiating embryonic stem cells as a model for embryogenesis [Figure 6.1]. Mouse embryonic stem (ES) cells are derived from inner cell mass of the 4.5 d.p.c pre-implantation blastocyst[5-7]. Therefore, as an *in vitro* model of embryogenesis, these ES cells represent the time point prior to the division of the inner cell mass into the trophoblast and the epiblast. Embryonic stem cell retain the pluripotent properties of the inner cell mass and are thus able to develop into almost any tissue type. They are defined by their ability to self-renew in an undifferentiated state, thereby presenting researchers a virtually limitless supply of cells for investigation. A second pluripotent population, termed early primitive ectoderm-like (EPL) cells has been cultured from ES cells[8]. These cells are have a genetic and morphological profile similar to the early primitive ectoderm of the embryonic epiblast[8,9], characteristic of the time point just prior to differentiation into the three

primary germ layers. If ES or EPLs are allowed to spontaneously differentiate, they form embryoid bodies[9,10], heterogeneous clusters of all three of the primary tissue types, representative of differentiated tissue. Taken together, the embryonic stem cells, early primitive ectoderm-like cells and embryoid bodies present an *in vitro* model useful for profiling the cellular mechanisms of early embryogenesis.

Genetic expression of differentiating embryonic stem cells continues to be well studied, especially in relation to the pluripotent properties of these cells.[10-16] Genetic expression of differentiating ES cells has been compared to *in vivo* embryogenetic activity, thus correlating the *in vitro* and *in vivo* models of embryogenesis.[9, 16, 17] One area lacking in exploration in regards to embryogenesis is the investigation of comprehensive protein expression. Proteins are produced by translation of mRNA, and are an additional layer of cellular dynamics. Examination of protein expression thus allows a more detailed glimpse into functionality of the cell state, providing information that is complementary to genomic information. Since protein abundancies are not always related to mRNA abundancies,[18-23] quantification of regulated protein expression is useful for identification of dynamically controlled pathways and functions corresponding to each biological condition.

Currently, comprehensive protein profiling of embryonic stem cells has been limited to a few recent studies.[24-26] These proteomic studies have used either a 2D gel electrophoresis[24] or a liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS)[25,26] approach to describe the protein expression of embryonic stem cells. However, these studies do not focus specifically on embryogenetic events, and instead use embryonic stem cells to investigate the protein expression of the pluripotent cell state[24,25] or compare the pluripotent cell state to the differentiated state[26]. To date, there has not been a report specifically attempting to detail or quantitate the global protein expression occurring during embryogenesis.

In this study, we use LC -MS/MS techniques and spectral counts[27] to characterize the protein expression of embryonic stem cells, early primitive ectoderm-like cells, and embryoid bodies as an *in vitro* model of early embryogenesis. We qualitatively describe the functionality of these cell states as well as provide quantitative information on the translational regulation of shared populations. In addition, we

compare functional aspects of our proteomic findings with previously investigated events associated with embryogenesis. The analysis of these three cell states by LC-MS/MS techniques represents the first time that comprehensive protein expression occurring during early embryogenesis has been investigated. The study of functionality and regulation at a translational level complements genomic studies and yields new topics for investigative studies relating to early embryogenesis.

EXPERIMENTAL

Materials

Ammonium bicarbonate, α-cyano-4-hydroxycinnamic acid (CHCA), dithiothreitol (DTT), guanidine hydrochloride, iodoacetamide (IDA), tris-HCl, trifluoroacetic acid (TFA), were purchased from Sigma (St Louis, MO). Ethanol, USP grade, 200 proof was purchased from AAPER Alcohol and Chemical company (Shelbyville, KY). Acetonitrile (ACN), isopropyl alcohol (IPA), formic acid (FA), was purchased from J.T. Baker (Phillipsburg, NJ). Sequencing grade trypsin was purchased from Promega (Madison, WI, USA).

Cell Culture

R1 mES cells[6] were cultured in the absence of feeders on tissue culture grade plastic-ware(VWR) pre-coated with 0.2% gelatin-phosphate buffered saline (PBS), as described previously.[8] ES cell culture complete medium consisted of Dulbecco's Modified Eagle Medium (DMEM, Gibco BRL) supplemented with 10% fetal calf serum (Gibco), 1% L-glutamine, 1% sodium pyruvate (Gibco), 0.1 mM 2-mercaptoethanol, 1% penicillin/streptomycin and 1000 U/ml recombinant human LIF (ESGRO,Chemicon) at 37°C under 10% CO_2. Differentiation was induced by removing the ES cells from the plates through 0.25% trypsin (Gibco) washed in PBS and re-plated on 0.2% gelatin-coated plates in non-conditioned complete media i the abscence of LIF. After 2 days cells were trypsinized, washed 3 times with ice cold PBS and stored at -80°C.

R1 mEPL cells were formed and maintained by culturing as described previously[8] in ES-media containing 50% conditioned media (MedII) supplemented with 1 mM L-glutamine, 0.1 mM 2-

mercaptoethanol, 100 U/ml penicillin, 100 U/ml streptomycin and 1,000 U/ml recombinant LIF at a density of 2 x 104 cells/cm^2. mEPL cells were passaged every 48 hours. Med II conditioned media was prepared from HepG2 cells (ATCC HB-8065) grown for 5 days in DMEM supplemented with 10% FCS, 40 mg/ml gentamycin and 1 mM L-glutamine, after being seeded at 5 x 104 cells/ml. Cell supernatants were collected, filter sterilized then stored at 4°C for up to 2 weeks. NIH 3T3 mouse fibroblast and early passage (passage 2-3) mouse embryonic fibroblasts (MEFs), prepared from outcrossed Swiss mice[28] were cultured in DMEM supplemented with 10% foetal calf serum, 1 mM L-glutamine, 100 U/ml streptomycin and 100 U/ml penicillin.

Flow cytometry

Fluorescein labeled Dolichous Biflorus Agglutinin (DBA) lectin was purchased from Vector Laboratories. Cells were harvested to a single cell suspension by trypsinization then washed and resuspended in 0.1% BSA-PBS. Aliquots of 100 µl containing 105 cells were mixed with FITC-DBA-lectin (1:100) and incubated for 60 min on ice. Cells were then washed once in PBS and resuspended in 500 µl of PBS and analyzed on a Beckman-Coulter flow cytometer with WinMDi software.

Preparation of cell extracts

Proteins were extracted using the TriZol® Reagent kit (Invitrogen, Life Sciences) and following the manufacturer's protocol. Briefly, 6 cell pellets from each population (~20 million cell each pellet) were lysed in 1-mL TriZol reagent by sonication followed by incubation at room temperature to permit dissociation of nucleoprotein complexes. Exactly 0.2 mL of chloroform was added to each pellet and the samples centrifuged at 12,000g for 5 minutes at 4°C. The aqueous portion containing the RNA was removed from the samples. This allowed extraction of the RNA from the samples for use in the subsequent transcriptome analysis. DNA was removed by adding 0.3 mL of ethanol and centrifuging at 2,000 g for 5 minutes at 4°C. The phenol/ethanol supernatant containing the proteins was removed to clean microcentrifuge tubes. Proteins were precipitated from this using 1.5 mL of isopropanol, incubating for 20 minutes at room temperature before pelleting at 12,000 g for 10 minutes at 4°C. Protein pellets were washed by incubating twenty minutes at room temperature in a 0.3 M Guanidine HCl/95% ethanol

solution, followed by centrifugation at 8,000 g for 5 minutes at 4°. Supernatant was removed and the wash repeated a minimum of three times for each pellet.

Protein Digestion

Each pellet was dissolved in 200 µL of 6 M Guanidine HCl to the pellet and sonicating for 30 minutes at room temperature. Proteins were reduced with 50 mM dithiothreitol, incubating for 2 hours at 55°C, followed by alkylation with 125 mM iodoacetamide, incubating in the dark at room temperature for 1 hour with periodic vortexing. The solution was diluted to reduce the concentration of 6 M guanidine HCl to < 1 M using 50 mM ammonium bicarbonate. Proteins were digested in solution by incubating with 1% w/w sequencing grade trypsin at 37°C for 72 hours. After digestion, samples were filtered with a 0.2 µM Nanosep filter to remove particulates, followed by a 10,000 MWCO filtration to remove undigested proteins. Peptides were desalted on a 4.6 x 150 mm Jupiter C18 column (Phenomenex, Torrance, CA), eluting in 80% acetonitrile. After desalting, peptides were dried by lyophilization.

First Dimension Reverse Phase Separation

A BCA kit (Pierce, Rockford, Ill) was used to according to the manufacture's protocol to quantitate the desalted peptide population. Exactly 2.5 mg of each peptide population was aliquotted for first dimension reverse phase analysis. Peptides were separated by reverse phase using a using a 4.6 x 150 mm Jupiter C18 column. Mobile phase A was 0.1% trifluoroacetic acid in water, mobile phase B was 0.1% trifluoroacetic acid in acetonitrile. Peptides were eluted using a linear gradient from 5% B to 45% B, collecting 3 minute fractions offline. Each collected fraction was dried by lyophilization and diluted to approximately 1µg/µL, based upon the A 280 trace.

Second Dimension Reverse Phase Separation

Fractions were analyzed on a Finnegan LTQ coupled to an Agilent 1100 series HPLC (Agilent Technologies, Palo Alto, CA) equipped with a quaternary pump and an autosampler. Exactly 5 µL of each 3 minute fraction was injected onto a 150 µM x 150 mm , 5 µm C18 column (GraceVydac, Hesperia, CA). Mobile phase A was 0.1% formic acid in water, mobile phase B was 0.1% formic acid in acetonitrile. Flow rate was set at 1.0 µL/min. Gradient elution of the peptide was achieved using a

concentration of mobile phase B that bracketed each collected fraction by 10% and eluted over a period of 100 minutes. Each separation was followed by a column wash at 60%B for 15 minutes and re-equilibration for 45 minutes. Peptides were eluted directly into a Finnegan LTQ equipped with a nanoelectrospray source. Source voltage was set at 2kV. Dynamic exclusion parameters were set for an exclusion time of 160 seconds, maximum list of 50, and a repeat count of 3. Each full MS scan from 450-2000 m/z was collected in centroid mode, followed by MS/MS on the top nine precursors. Each 3 minute fraction collected from the first dimension of reverse phase was analyzed three times by second dimension reverse phase, resulting in a total of 72 runs over all three proteomes.

Database Searching

Raw files obtained from the LTQ analysis were converted to mzXML files and then to PKL files using in-house software to perform the conversions. Peak Lists (PKL files) were searched using Mascot (www.matrixscience.com) against a combined NCBI non-redundant mouse database downloaded on 7/27/06 (46,905 entries) and bovine database (35,907 entries). Search parameters were set to trypsin as the enzyme specification, 2 missed cleavages, 450 ppm for the precursor and 0.6 Da for the fragment masses. Variable modifications used were carbamidomethylation of cysteines, guanidination of lysines, and oxidation of methionine (experimental modifications). The combined NCBI mouse and cow database was reversed and all data searched again using the same search parameters, producing a number of randomly identified peptides.

Protein Identification

Results from the real and random mascot searches were input into ProValt[29] using a 1% false discovery rate to group non-redundant peptides to proteins. ProValt utilizes three user defined input parameters, protein false discovery rate, peptide coverage, and minimum peptide score, to calculate the mascot score threshold for single and multiple peptides grouping to proteins. In this case, false discovery rate was set at 1.0, peptide coverage was set to 5, and minimum peptide score was set to 18. From this, single peptides identifying a protein were calculated to have a required mascot score of $\geq$51 to meet the 1.0% FDR criteria. Likewise, for proteins identified by two or more peptides, the minimum peptide

mascot score threshold associated with a 1.0% FDR of the protein was calculated to be two peptides with a score ≥ 36, three peptides with a score ≥ 30, four peptides with a score ≥ 27, and five peptide with a score ≥ 24. ProValt clusters nonredundant peptides falling within the user defined criteria to protein homology groups based upon sequence homology of identified peptides. The protein with the most sequence coverage is reported as the top scoring protein. Homology groups mapping to bovine proteins were eliminated from the ProValt results, producing a list of identified mouse homology groups. Top scoring proteins are reported with total mascot score, average spectral counts, number of non-redundant peptides associated to the protein, and percent coverage of the protein.

Data Analysis.

Only spectral events that were linked to protein identification at a 1% false discovery rate were used to quantitatively compare protein populations. Choosing only spectral observations that resulted in a statistically linked match eliminated spectral events due to background noise and randomly matched spectra. It may be noted that these spectral counts signify the sampling of the database by the system, thus spectra resulting from biological modifications and/or unknown proteins that have not been identified are not associated with the statistical analysis of the proteomes of each cell population. For quantitation, average spectral counts were calculated by averaging the two or three replicates where the protein was reported. If the protein appeared in only one replicate, that value is used for the spectral count. Normalization of spectral counts proceeded using a variation of a method recently described by Zybailov et al.[30] Individual spectral counts were first divided by the number of amino acids. This ratio is then divided by the sum of the ratio of all spectral counts over total number of amino acids for each proteome (Equ A). We found it was helpful to introduce a correction factor that accounted for variances in sampling of the database by the system. This correction factor consisted of the total spectral counts of a given population (recorded from sampling of the database) divided by the total spectral counts of the control population and was multiplied to the normalization factor (Equ B).

Equation A) $$\frac{SC_{individual}/\#AA_{individual}}{\Sigma\,(SC_{Total}/AA_{Total})}$$

Equation B) $$\frac{CF = \Sigma\, SC_{Control}}{\Sigma\, SC_{Measured}}$$

After normalization, proteins were sorted by comparative proteomes. This produced four comparative populations: proteins in common with all three cell types, proteins in common with ES and EPL, proteins in common with ES and EB, and proteins in common with EPL and EB. Proteins that only mapped to one of the three cell types were labeled as single proteomes. Normalized spectra were used to calculate a comparative ratio signifying protein expression for comparative populations. Each comparative ratio is taken relative to a control population, either embryonic stem cell or embryoid body.

Prediction of Protein Subcellular location.

Prediction of protein subcellular location was made using WoLF PSORT.[31]

Ingenuity Functional Analysis.

The Ingenuity Pathways Analysis software (www.ingenuity.com) was used to analyze functional aspects of the comparative proteomes. Ingenuity is a software application that employs an expert curated knowledge base derived from literature findings to calculate cellular function for groups of proteins. Ingenuity reports a significance value for each reported functional or pathway analysis. The significance values associated with functions analysis is a measure for how likely it is that genes from the user defined dataset file participate in that function. The significance is expressed as a p-value, which is calculated using the right-tailed Fisher's Exact Test. Ingenuity calculates the p-value by comparison of the number of proteins from the user's dataset that participate in a given function or pathway, relative to the total number of occurrences of these genes in all functional/pathway annotations stored in the Ingenuity Pathways Knowledge Base.

RESULTS AND DISCUSSION

Proteomic analysis of embryonic stem cells, early primitive ectoderm-like cells and embryoid bodies. Embryonic stem cells (ES), early primitive ectoderm-like cells (EPL), and embryoid bodies [Figure 6.2] from the R1 mouse stem cell line[6] were analyzed for protein expression as an in *vitro model* of embryogenesis. A two dimensional reverse phase strategy was used to qualitatively and quantitatively describe the three cell populations that are representative of critical time points in early embryogenesis. After protein extraction and digestion, peptide populations were separated by reverse phase, collecting 3

minute fractions offline. Offline fractionation allowed the same fraction to be stored and repeatedly analyzed. Each collected fraction was analyzed three times by a second reverse phase analysis culminating in collection of MS/MS spectra by a Finnegan LTQ for peptide identification. To increase the confidence in peptide identification, fragmentation spectra were searched against a forward and reversed combined mouse and bovine database using the Mascot[32] search algorithm. Results from the Mascot search were processed using Provalt[29] to group nonredundant peptides from all runs to protein identification at a ≤1% false discovery rate (FDR). Proteins were further grouped according to homology of identified peptides. For each homology group, the top protein was the protein with the most unique peptides mapping to it. Although other proteins from the homology group may be present in the sample, only the top proteins were used for comparative purposes. Bovine proteins were removed from the list of uniquely identified proteins, leaving only the proteins identified as unique to mouse. Overall, this type of analyses resulted in a total of 3328 mouse proteins (1033 protein groups) for ES, 1852 mouse proteins (558 protein homology groups) for EPL, and 2911 mouse proteins (1004 protein homology groups) for embryoid bodies. Uniquely identified proteins were sorted according to comparative and single proteomes, resulting in a distribution of 1498 proteins across the three populations [Figure 6.3]. Grouping according to single and comparative proteomes avoided comparing proteins that were not observed across between comparative populations. A smaller number of proteins were identified in the EPL population at a ≤1% FDR than in either embryonic stem cells or embryoid bodies. The under representation of this population is probably due to bovine serum contamination from the media in spite of washing the cells in excess of three times. This was indicated by a higher amount of bovine albumin in EPL versus either ES or EB (3.90 EPL/ES, 5.21 EPL/EB).

Proteins were evaluated according to cellular location to determine the composition of the sample preparation [Figure 6.3]. Since this was a TRIzol® extraction, the population should consist largely of soluble proteins, mostly from the cellular cytoplasm. In fact, 46% of the identified proteins could be identified as having cytoplasmic origins. Membrane proteins composed 3.4% of the uniquely identified

population. Therefore, prediction of probable location of protein correlates with the performed sample preparation.

Evaluation of a control set of proteins. A set of proteins was selected from the comparative ES, EPL and EB protein population, producing a collection of control proteins. These proteins consisted of alpha tubulins (TUBA), beta tubulins (TUBBs), gamma actin (ACTG), and glyceraldehyde 3-phosphate dehydrogenase (GAPDH). Alpha tubulin was selected since it has previously been used as a loading control during immunoblotting analyses of embryonic stem cells, early primitive ectoderm-like cells and embryoid bodies.[33] Beta tubulin (TUBB), gamma actin and GAPDH were selected as part of the control set primarily since these cytoplasmic proteins are often used as loading controls during immunoblotting assays.

Translating control proteins from immunoblotting techniques to LC-MS techniques required us to consider differences between the techniques, mainly the inability of the immunoblotting technique to discern between isoforms. For instance, the alpha tubulin antibody used as an immunoblotting loading control for ES, EPL, and EB analysis[33] is specific to a highly homologous region for alpha tubulins, the sequence EGMEEGEFSEA, occurring between amino acids 414-421. In an immunoblotting assay, this antibody is unable to discern between isoforms of alpha tubulin. However, in this proteomic study, we were able to confidently identify individual alpha tubulins in our data set. For analysis of alpha tubulin as a control protein, we include all individual alpha tubulins that contained this epitope in our control set. A similar issue resulted from the use of beta tubulin and GAPDH as a control protein. Concerning the beta tubulins, beta tubulin 1, 2, 3 and 5 were all identified in comparative datasets. Beta tubulin 3 was not used in our control dataset as there have been reports of variable expression of this isoform between different cell types.[34] On the other hand, several antibodies used for loading controls during immunoblotting recognize epitopes in common with beta tubulin 1, 2, and 5 so these tubulins were evaluated as part of our control set of proteins. Glyceraldehyde 3-phosphate dehydrogenase is part of the glycolysis cycle and is a highly conserved protein. We found a total of ten variants of GAPDH in our proteome samples, each with

only small differences (≤10%) in amino acid composition. We treated each of these unique identifications as individual proteins in our control dataset.

Evaluation of our control protein dataset after normalization is demonstrated in Figure 6.4. Based upon accuracy of the ratios for each comparative set and standard deviation associated with these measurements, we set a ≥3 fold change limit for reporting up or down regulation. This is only slightly more conservative than previous reports where spectral counts have been used to report protein expression at 2-fold [30,36,] or 2.5-fold[37] expression level for biological samples. Overall ratios for the control proteins for comparative proteome were 0.94 ± 0.41 for EPL/ES, 1.19 ± 0.22 for EB/ES, and 0.88 ± 0.34 for EPL/EB. Taken as a whole, our control protein dataset had 18 individual proteins, comprising approximately 1% of the total protein population and allowing 35 individual comparisons between proteomes. We believe that the selected proteins might also be useful as part of a constant feature set for other mammalian proteomic studies. Table 6.1 demonstrates the distribution of regulated, predicted and hypothetical proteins after the normalization procedure.

Single Proteomes: Comparison to Previous Studies. We inspected the single proteomes of our *in vitro* model of embryogenesis for genes and proteins that have been previously correlated to embryogenetic events and continued development of *mus musculus*. By investigating genes seen exclusively in embryonic stem cells, early primitive ectoderm-like cells, or embryoid bodies, we sought to substantiate the embryogenetic time points for our *in vitro* model. After sorting to comparative and single proteomes, literature searches were performed on the single proteome proteins and their associated genes, which are listed in entirety in Tables 6.2, 6.3, 6.4 (ES, EPL and EB, respectively). To increase our confidence that proteins were connected to a specified cell state, we only investigated protein identifications that were seen in at least two technical replicates. We also include protein and gene expression that has been reported to be associated with embryonic stem cells. We did not consider predicted proteins in our literature search.

The identified protein expression that can be correlated to previous embryogenetic or developmental research is shown in Table 6.5. Significantly, the proteins unique to each differentiated state have been linked to events corresponding to the appropriate *in vivo* embryogenetic processes.

The embryonic stem cell unique proteome produced thirteen genes that had been explored previously in terms of embryogenesis and were reliably seen in that population. These proteins have properties corresponding to the equivalent in *vivo* time point, the inner cell mass, or have been seen in prior studies characterizing embryonic stem cells. For instance, we identified the developmental pluripotency associated 5 (DPPA5) protein only in our embryonic stem cell population. This gene has previously been reported to be expressed by the human inner cell mass, human and mouse embryonic stem cells, and human primordial germ cells.[38, 39] DPPA5 is not expressed by differentiated cells or by carcinomal cells and thus has been proposed to be a robust identifier of pluripotency.[39] Our study further substantiates the proposal that this protein could be used as an indicator of a pluripotent cell population. Two other proteins, basic transcription factor 3 (BTF3) and the transcription factor ELYS (AHCTF1), were identified that have been linked to the inner cell mass during embryogenesis. Basic transcription factor 3 has been reported to be required for post implantation survival.[40] BTF3 was also seen in the 2D gel electrophoresis of R1 mouse stem cell,[24] and the 2D LC MS/MS of the E14-1 mouse stem cell line,[25] but was not reported by Hoof et al. in the D3 mouse stem cell line when comparing embryonic stem cell protein expression to embryoid body expression.[26] Gene disruption of another protein, the transcription factor ELYS, also known as AT hook containing transcription factor 1 (AHCTF1), showed that this gene is required for survival and development of the inner cell mass. Homozygous embryos died between 3.5 and 5.5 d.p.c., whereas the wild type embryo reportedly expressed AHCTF1 strongly throughout 6.5 d.p.c.[41] Later stages of development have been explored, reporting down regulation of this protein as the embryo develops and finally segregation to haematopoietic tissue by postnatal day four.[42] AHCTF1 has not been reported by other large scale proteomic studies. On the whole, the embryonic stem cell proteins in this study appear to correspond with the *in vivo* time point of embryogenesis, prior to differentiation of

the inner cell mass to hypoblast and epiblast or can be correlated to previous studies of embryonic stem cells.

Early primitive ectoderm-like cells were explored in terms of embryogenesis [Table 6.5]. Proteins identified exclusively in this population were related to tissue modeling and early systems development, corresponding with *in vivo* activities of embryogenesis [Fig 6.1]. An example of tissue remodeling occurs with the observation of the well studied protein SPARC, secreted acidic cysteine rich glycoprotein (Reviewed in 43).[44-46] In terms of embryogenesis, the SPARC-null mice have been shown to progress normally, but postnatally develop cataracts.[45,46] Immunochemistry has associated the appearance of the SPARC protein with cells involved in cell proliferation, migration and remodeling.[46] SPARC is considered to participate in these processes by counteradhesive activity, growth factor modulation, cell shape alteration, as well as regulation of tissue differentiation.[43] The appearance of SPARC only in the EPL population suggests that these cells are initiative to or are undergoing the morphogenetic activities of gastrulation, previously proposed by genetic profile comparison to the primitive ectoderm.[8,9] Proteins corresponding to early systems development were also seen unique to this population. Eukaryotic translation initiation factor 4A2 (EIFA2)[47] and utrophin (UTRN)[48] have been shown to be upregulated during neural development, one of the early emerging systems during embryogenesis. Morgan and Behringer observed that EIFA2 was sharply upregulated *in vitro* during neuralization of embryonic ectoderm and *in vivo* during neural induction.[47] The appearance of EIFA2 only in the EPL population may signify that this cell population is progressed towards a differentiated state. The gene utrophin was systematically investigated during early mouse embryogenesis. The first significant accumulation of utrophin was reported at E8.5 in the region of the developing neural groove.[48] Other proteins that have been linked to systems development were annexin-2A (neoangiogenesis)[49], and Luc7 homolog (S. cerevisiae)-like isoform 2 (regulation of myogenesis).[50] It is noted that many of the species investigated have been studied in terms of genetic expression rather than protein expression. Based upon literature research, it appears that for many of these proteins, expression is upregulated initiative of the differentiation process, rather that being indicative of a definitive differentiation. However, the

observation of these specific proteins indicates that the EPL cell population is further progressed towards differentiation than the embryonic stem cell population. Overall, the profile of the proteins uniquely associated with this population is reflective of morphogenesis and differentiation, associating this time point with a primitive ectoderm commencing towards differentiation.

Evaluation of the embryoid body protein expression resulted in twenty four species that have previously been investigated during early development [Table 6.5]. Many of these species are linked to upregulation in genetic expression as differentiation proceeds. One identified species that has been relatively well studied in terms of gene expression is the Eph receptor A2 (EPHA2, also known as ECK). EPHA2 mRNA first appears in the ectodermal regions adjacent to the distal end of the forming primitive streak during gastrulation.[51] As development progresses, EPHA2 localizes to the tip of the primitive streak, and displays a spatially restricted pattern in the developing hindbrain region, suggestive of a role in guiding cell-cell interactions as this development proceeds.[51] In murine adults, EPHA2 demonstrates widespread expression, with highest levels in cells of epithelial origin.[52] The varied presence of EPHA2 in the different portions of the lifecycle of the animal led to a further study by performed by Brantley-Sieders et al.[53] In this report, adult mice and mouse endothelial cells were used to investigate the regulatory activities of this receptor. It was found that EPHA2 was critical for postnatal angiogenesis, probably due to EPHA2 participating in epithelial cell migration. The expression of the EPHA2 protein in connection with the embryoid body population thus correlates with nervous system development and later processes that are essential for survival of the adult animal. Another species seen only in the embryoid body population, four and a half LIM domains 1 (FHL1) has been reported as observed first in the 8.5 d.p.c heart outflow tract and with advancing development is displayed in the heart, limb buds, and neural tube (10.5 d.p.c).[54] Other research has implicated the expression of FHL1 protein as essential to the process of skeletal muscle elongation, studied in developing myoblasts.[55] In this study, FHL1 protein expression decreased during the first 40 hours of differentiation, followed by a sharp increase as definitive differentiation progressed. This decreasing and then increasing expression appears to be a common motif during embryogenesis, and is a way for the organism to maximize efficient use of coded genetic

information. Other proteins uniquely associated with the embryoid body population, including EIF4G2, HNRPU, and USP9X, were related to an advanced differentiation state. For example, EIF4G2 is essential to gastrulation processes, with EIF4G2 (-/-) mice dying at onset of gastrulation and provoked EIF4G2 (-/-) ES cells demonstrating impaired differentiation.[56] In situ hybridization of HNRPU mRNA demonstrated a restricted expression pattern during embryogenesis with weak expression in the primitive ectoderm followed by vigorous expression at 8.5 -9.5 d.p.c in the tail bud and mesoderm as well as in the forebrain and limb buds.[57] USP9X mRNA and protein was detected using *in situ* hybridization and immunochemistry starting at 9.5 d.p.c and was thereafter expressed in primordial germ cells with weaker expression in the surrounding somatic cells.[58] USP9X has an apparent gender dependent expression pattern, with expression in the male primordial gonads deteriorating by 15.5 d.p.c, while female expression remains stable.[58] On the whole, the proteins identified in the embryoid body pool are associated with later stage development systems and processes such as skeletal/muscle differentiation, nephrogenesis, and limb budding, or have been linked to continued viability of the embryo.

Generally speaking, the proteins that are exclusively linked to each of the three cell populations appear to correlate our *in vitro* model of embryogenesis to the *in vivo* processes shown to be essential to early murine development. However, the majority of the proteins that could be connected to unique populations have not yet been investigated in terms of embryogenesis, either at a genetic level or a protein level. In addition, many predicted proteins were also seen in each unique population that could be of further interest to the field of embryogenesis. Thus, while we have shown that the protein expression of our *in vitro* model corresponds to previously investigated *in vivo* expression, the study has identified additional genes that merit the attention of future embryogenetic studies.

Functional Analysis of the in vitro model of embryogenesis. The main purpose of this work was to explore functionality and regulation of the global protein expression occurring during embryogenesis. With this in mind, we examined the *in vitro* model of embryogenesis for functionality between shared proteomes. Ingenuity software (www.ingenuity.com) was used to compute biological functions of co-expressed proteins from comparative proteomes. Ingenuity compares the set of identified proteins to a

database consisting of genes and their corresponding functions, supported by literature observations. The program calculates a probability (p-value) based upon the number of proteins from the user dataset that participate in a function relative to the total number of occurrences of these genes in all functional annotations stored in Ingenuity's literature curated database. The smaller the p-value, the more likely that activities related to that function are occurring. We input proteins observed in comparative proteomes into the Ingenuity software. After functional searching using Ingenuity, we further filtered the identified functions, keeping only those functions associated with a p-value less than 5.0E-3. Some proteins were implicated in multiple functions. Tables 6.6 to 6.9 list the statistically significantly identified functions for each comparative proteome, as well as the proteins associated with each function, whilst Tables 6.10-6.13 list the shared proteomes in entirety.

Functional aspects of embryonic stem cells and early primitive ectoderm-like cells. We examined the functionality and regulation of the proteins shared by ES cells and EPL cells. Table 6.10 demonstrates the complete list of proteins identified in this comparative set, including calculated regulation. This comparative proteome contained 60 proteins, including 13 predicted proteins, and one hypothetical protein. The top cellular function computed by Ingenuity software was cellular movement (p-value 2.63E-4) [Table 6.6]. One protein corresponding to cellular movement that appeared to be significantly upregulated in the EPL population versus the embryonic stem cell population was basigin (BSG), with an EPL/ES ratio of 10.7. Basigin, a transmembrane glycoprotein, has been shown to be essential for intercellular recognition between the uterine decidua and the embryo, allowing successful implantation of the blastocyst.[59,60] Basigin protein expression in the embryo begins at approximately 4 d.p.c., around the time of implantation.[59,60] Null mutants have been shown to develop normally until the time of implantation, at which time a majority of the mutants die.[59] Survival of the atypically small null mutants is thought to be enabled by expression of embigin, a protein with homologies to basigin.[59] Basigin has also been identified in cancer cells, where upregulation in expression has been positively correlated to changes in tissue morphology and invasiveness of cancer.[61] Thus, the up regulation of BSG protein in the

EPL population, as seen in this study, seems to indicate that BSG plays a further functional role in embryogenesis by participating in intercellular recognition during primitive ectoderm differentiation.

The majority (78%) of the proteins in this comparative proteome do not have a significant change in regulation between the two cell types. However, many of the highly upregulated proteins (EPL/ES), such as BSG (10.7), HYOU1 (5.9), and GMFB (13.7) can be connected to processes of tissue /cell morphology and cellular movement. Therefore, the calculated functions, as well as regulation of the associated proteins, appear to be a realistic depiction of the progression from a pluripotent state towards processes of gastrulation.

Functional aspects of early primitive ectoderm-like cells and embryoid bodies. The comparative population consisting of proteins co-expressed by early primitive ectoderm-like cells and embryoid bodies was investigated. This protein set consisted of 50 proteins, 11 of which were predicted proteins as well as one hypothetical protein. A complete listing of the protein and quantitative ratios is provided in Table 6.11. Regulated proteins represent 40% of the identified population with sixteen upregulated proteins and four down regulated proteins (EPL/EB ratios). All observed proteins were investigated using Ingenuity to compute probable functions of the comparative population, demonstrated in Table 6.7. Primary functions of these further differentiated cells include cellular organization and assembly, molecular transport and cell growth/proliferation. Two proteins, FABP3 and LAMC1 are seen with a significantly regulated expression in the EPL population versus the EB population. Fatty acid binding protein 3 (FABP3) has is appears heavily upregulated in the EPL population versus the embryoid body population (EPL/EB 101.1). FABP3, a transporter molecule, binds lipids for transport throughout the cellular cytoplasm.[62] FABP3 null mice develop normally with apparent compensation for loss of this gene by other fatty acid binding proteins, but as adults develop cardiac hypertrophy.[62, 63] Interestingly, in these studies, high levels of FABP3 were also observed in brain tissue; however, implications for the adult animal in the absence of FABP3 have not yet been investigated. While apparently nonessential at early stages of development, the expression of FABP3 appears to be necessary for maintenance of a healthy adult organism.[64] Regulatory expression of FABP3 has not been investigated during embryogenesis. The current study implies that

FABP3 regulation plays a role in early stage development, indicating a need for further studies on this particular fatty acid binding protein.

Another protein, laminin gamma 1 (LAMC1), also appeared to be highly regulated. This protein was not associated with a particular cellular function by Ingenuity software at the time of our investigation. LAMC1 is an extracellular, complex glycoprotein associated with the basement membrane. Laminins play a complex role in development, implicated in the attachment, migration and organization of cells into tissue types.[65, 66] The essential role of LAMC1 during embryogenesis is demonstrated by LAMC1 null mice which display embryonic lethality at an estimated 6.0 d.p.c.[67] Investigation of these LAMC1 (-/-) embryos revealed the absence of any basement membranes and failure to differentiate endoderm, therefore this protein is essential to the formation of at least one of the primary germ layers. To our knowledge, and like FABP3, LAMC1 has not been quantitatively investigated during the processes of early embryogenesis. This study shows the substantial upregulation in the EPL population versus the EB population, an observation that is can be expected as by the previously shown requirements for the successful development of the embryo.

In vivo, the 6.0 d.p.c time period is marked by the differentiation of the primitive ectoderm into the three primary germ layers. Rapidly accelerated growth of the epiblast leads to the extension of the primitive streak followed by dispersal of cells as the process of gastrulation initiates massive migration and morphogenesis of cells.[3] The comparative protein population of early primitive ectoderm-like cells and embryoid bodies appears to recapitulate this scenario, demonstrating the functional and regulatory activities of this process including the regulation of both predicted and hypothetical proteins.

Functional aspects of embryonic stem cells and embryoid bodies. A total of 349 proteins were found in common between embryonic stem cells and embryoid bodies [Table 6.12]. The larger number of common proteins in this population compared to other populations could be partly due to under sampling of the EPL population, which occurred due to bovine serum contamination in spite of copious washing of the cells. In this population, 50 proteins (~14%) of the proteins were predicted proteins and seven proteins (2%) were hypothetical proteins. Thirty four proteins were down regulated and 43 proteins were

upregulated in the embryonic stem cell population versus the embryoid body population, accounting for 22% of the identified population. The top primary function shared between embryonic stem cells and embryoid bodies was RNA post-transcriptional modification (p-value 6.47E-5), 40% of which were upregulated in ES versus EB [See Table 6.8]. Among the regulated proteins associated with RNA post transcriptional modification was the DEAD box protein 52 (DDX52). DDX52 appeared to be more highly expressed in embryonic stem cells over embryoid bodies (ES/EB 19.6). DDX52 is a nuclear enzyme possessing ATP hydrolysis activity[68] that, to the best of our knowledge, has not been studied from a developmental aspect. The DEAD box proteins have been reported to have a range of RNA associated activities, including pre-mRNA splicing and export, RNA winding and unwinding, translation initiation, organelle gene expression, and RNA decay.[69, 70] It is not immediately clear what specific role DDX52 may have in embryonic stem cells versus embryoid bodies, and due to the magnitude of the upregulation, this protein may be of further interest in future studies. Also associated with RNA post transcriptional activity was the eukaryotic translation factor 4A1 (EIAFA1). EIAF41 was down regulated in the observed in the ES population versus the EB population (0.29 ES/EB). Again, we did not find a study depicting this protein during the course of embryogenesis. However, over expression of EIF4A1 mRNA has been associated with the metastasis of cancer,[71,72] implicating that the EIF4A1 protein may be upregulated during differentiation processes. This protein has previously been identified in a 2D gel analysis of R1 embryonic stem cells, but was not reported in the LC-MS/MS analysis of the D3 embryonic stem cell line by Hoof et al.[24,26]

One protein that did not correlate to a defined cellular function and that was found in common between embryonic stem cells and embryoid bodies was the undifferentiated transcription factor 1 (UTF1). Undifferentiated transcription factor 1 (UTF1) has been shown to be differentially regulated during development by *in situ* hybridization of mRNA.[73] Rapid induction of UTF1 mRNA expression begins at 3.5 d.p.c. and is expressed by the inner cell mass of the preimplantation blastocyst and not the trophoectoderm. UTF1 mRNA continues to be expressed during formation of the primitive ectoderm and is down regulated as differentiation proceeds, becoming undetectable by 14.5 d.p.c.[73] Although there are

no reports on the expression of the UTF1 *protein* during *in vivo* embryogenesis, the UTF1 gene and protein have been explored in singular to embryonic stem cells. Investigation of UTF1 genetic expression in embryonic stem cells has indicated that the UTF1 gene is activated by the OCT3/4 protein,[74] a transcription factor known to be essential for the formation and maintenance of pluripotent cells.[75] The interplay between this gene and the OCT3/4 appears to play a critical role in the proliferation of the pluripotent population.[74] The UTF1 protein was reported to be uniquely identified in the D3 and E14-1 but not the R1 mouse embryonic stem cell line during comprehensive proteomic studies of embryonic stem cells.[24-26] In our study, UTF1 was expressed in both embryonic stem cells and embryoid bodies, although with a lower level observed in embryoid bodies (normalized EB/ES ratio 0.46).

Both UTF1 and the previously discussed EIF4A1 provide examples of differential protein expression across cell lines. While some of the variation may be due to differences in the capabilities of the analytical techniques, this type of deviation in expression is not surprising as genetic analysis of other embryonic stem cell lines have revealed unique signatures for each line, attributed to differing growth conditions and inherited genetic variation.[11,76] Further characterization of protein regulation across cell lines would provide an additional understanding of the dynamically regulated cellular processes crucial to maintaining the pluripotent embryonic state and directing differentiation processes.

Functional and qualitative aspects of embryonic stem cells, early primitive ectoderm-like cells and embryoid bodies.. There were 319 proteins identified in common with embryonic stem cells, early primitive ectoderm-like cells and embryoid bodies [Table 6.13]. Predicted proteins (45) composed 14% of this population. Five hypothetical proteins identified. Regulated protein composed approximately one third of the population, with differential regulation amongst the three cell populations.

The histones were one group of proteins that was readily apparent in this population versus other comparative populations. Histone proteins are the core component of the chromatin complex, acting as a protein spool for winding and condensing strands of DNA. Post translational modifications of histones activate or silence genomic transcription by initiating or repressing winding of DNA strands.[77] Here, we observed several histones that appeared to be differentially regulated across the three proteomes.

Theoretically, since the histone proteins are a part of the fundamental genomic material, the core histone protein should not be differentially expressed. This apparent regulation can be explained as follows. During our database searches of MS/MS fragments, we only allowed for experimental modifications. Biological modifications were excluded from our database searches since allowing for these multitude of modifications would cause a significant increase in the number of false positive identifications and as well as the time spent searching the data.[78] Since biological searches were not included in our database searches, apparent regulation may be a sign of differential post translational modification, as the biologically modified peptide would not be a part of the quantitative pool for the identified protein. We believe that the apparent regulation of histones in our study is in fact differential post translational modification. This is corroborated by several studies that have shown that modifications of histone proteins play a fundamental role in the events of embryogenesis.[79]

Functional aspects of the protein population shared between ES, EPL, and EB were calculated by Ingenuity software [Table 6.9]. The top function in common with all three proteomes was protein folding, represented by thirteen proteins with a calculated p-value of 3.89E-9. Most of the proteins associated with this function have similar expression values, as is anticipated for basic cellular functions. One of the interesting functions in regards to embryogenesis that appeared to be highly regulated was cellular growth and proliferation (p-value 2.84E-3, 46 proteins). Cellular growth and proliferation is obviously one of the vital processes of early embryogenesis, as the 25 pluripotent cells of the inner cell mass transform into the multi-celled tissues and organs of a complex organism. Several of the proteins exclusive to this function were uniquely regulated across the three cell populations. For example, the protein S-phase kinase-associated protein 1A (SKP1A) was upregulated in the ES cells versus the EPL population, and directionally upregulated versus the EB population (0.11 EPL/ES, 0.56 EB/ES). SKP1A is a multi-associative protein, that forms part of the ubiquitin-protein ligase complex SKP1A-cullin-Fbox, required in the early G1/S phase of the cell cycle.[81] The rapid proliferation of embryonic stem cells has been shown to be correlated with CDK2 driven cell cycle rate that is slows as differentiation proceeds.[34] Inactivation of the SKP1 protein during embryogenesis promotes a genetic instability, producing tumors

in mice that are unable to form the SKP1A-cullin-FBox complex.[81] The upregulation of SKP1A in our study seems to correspond to an embryonic cell cycle, with the proceeding down regulation in the early primitive ectoderm-like cells and two day embryoid bodies correlating with events occurring during cellular proliferation and differentiation. Another protein, serine (or cysteine) proteinase inhibitor, clade H, member 1 (SERPINH1), appeared upregulated only in the EPL population (11.3 EPL/ES, 0.93 EB/ES). SERPINH1 is a collagen binding glycoprotein that is expressed in spatially distinct patterns from 7.5 d.p.c on, and is also associated with proliferative cell regions.[82] SERPINH1 expression levels correlate positively with expression of the protein's substrate, collagen.[82] SERPINH1 null mice are reported to have a disordered neuroepithelial cell distribution and obvious growth retardation by 9.5 d.p.c.[83] SERPINH1 is thus essential for both somatic cell growth and is an essential component for cell morphogenesis during embryonic development. Therefore it appears that the upregulation of SERPINH1 associated with the EPL population is coincident to the controlled dispersal and morphogenetic activities of this population during cellular growth and proliferation.

The analysis of proteins expressed in common between embryonic stem cells, early primitive ectoderm-like cells and embryoid bodies revealed proteins and regulation that have not been previously investigated in terms of early embryogenesis. Most of the functions associated with this subset of proteins were related to basic cellular activities. However, regulation of the proteins associated with these functions varied considerably among the three populations, demonstrating a dynamic control of the differentiating cell.

CONCLUSIONS

In conclusion, we have demonstrated the largest scale study of protein expression occurring during embryogenesis (to date). We have used LC-MS/MS techniques along with spectral counts to qualitatively and quantitatively describe protein expression of differentiating embryonic stem cells in relation to the events occurring during embryogenesis. Approximately 22% of the total protein population was composed of predicted and hypothetical proteins. Many of the predicted proteins share regulatory aspects with their sequence aligned counterparts. The observations on the regulation and functionality of unique

and shared populations have indicated several candidates and topics useful for future embryogenetic studies. In addition, by describing the qualitative aspects of protein expression for the R1 embryonic stem cell line, we have provided another study that may be used for comparison of global protein expression among murine embryonic stem cell lines.

REFERENCES

1. Snow, M. H. L. (1977) Gastrulation in the mouse: growth and regionalization of the epiblast. *J Embryol Exper Morphol*, 42, 293-303.
2. Quinlan, G. A. (1996) Organisation of the Body Plan: Cell Fate and Gene Activity During Gastrulation of the Mouse Embryo. In *Mammalian Development*, Lonai, P., Ed. Overseas Publishers Association: Amsterdam, 1-27.
3. Gardner, R. L. (1988) Multi-lineage 'stem' cells in the mammalian embryo. *J Cell Sci Suppl***,** 10, 16.
4. Poelmann, R. E.(1980) Differential mitosis and degeneration patterns in relation to the alterations in the shape of the embryonic ectoderm of early post-implantation mouse embryos. *J Embryol Exp Morphol***,** 55, 33-51.
5. Martin, G. R. (1981) Isolation of a Pluripotent Cell Line from Early Mouse Embryos Cultured in Medium Conditioned by Teratocarcinoma Stem Cells. *Proc Natl Acad Sci USA***,** 78, (12), 7634-7638.
6. Nagy, A.; Rossant, J.; Nagy, R.(1993) Abramow-Newerly, W.; Roder, J. C., Derivation of Completely Cell Culture-Derived Mice from Early-Passage Embryonic Stem Cells. *Proc Natl Acad Sci USA,* 90, 8424-8428.
7. Evans, M. J.; Kaufman, M. H. (1981) Establishment in culture of pluripotential cells from mouse embryos. *Nature***,** 292, 154-156.
8. Rathjen, J. (1999) Formation of a primitive ectoderm like cell population, EPL cells, from ES cells in response to biologically derived factors. *J Cell Sci***,** 112, 601-612.
9. Pelton, T. A.; Sharma, S.; Schulz, T. C.; Rathjen, J.; Rathjen, P. D. (2002) Transient pluripotent cell populations during primitive ectoderm formation: correlation of in vivo and in vitro pluripotent cell development. *J Cell Sci***,** 115, (2), 329-339.

10. Leahy, A.; Xiong, J. W.; Kuhnert, F.; Stuhlmann, H. (1999) Use of developmental marker genes to define temporal and spatial patterns of differentiation during embryoid body formation. *J Exp Zool*, 284, (1), 67-81.

11. Lars Palmqvist Clive H. Glover, L. H., Min Lu, Bolette Bossen, James M. Piret, R.Keith Humphiries, Cheryl D. Helgason. (2005) Correlation of murine embryonic stem cell gene expression profiles with functional measures of pluripotency. *Stem Cells*, 23, 663-680.

12. Ginis, I.; Luo, Y.; Miura, T.; Thies, S.; Brandenberger, R.; Gerecht-Nir, S.; Amit, M.; Hoke, A.; Carpenter, M. K.; Itskovitz-Eldor, J. (2004) Differences between human and mouse embryonic stem cells. *Dev Biol*, 269, (2), 360-80.

13. Ramalho-Santos, M.; Yoon, S.; Matsuzaki, Y.; Mulligan, R. C.; Melton, D. A. (2002) " Stemness": Transcriptional Profiling of Embryonic and Adult Stem Cells. *Science*, 298, (5593), 597-600.

14. Kelly, D. L.; Rizzino, A. (2000) DNA microarray analyses of genes regulated during the differentiation of embryonic stem cells. *Mol Reprod Dev*, 56, (2), 113-123.

15. Houbaviy, H. B.; Murray, M. F.; Sharp, P. A. (2003) Embryonic stem cell-specific MicroRNAs. *Dev. Cell*, 5, 351-358.

16. Sharov, A. A.; Piao, Y.; Matoba, R.; Dudekula, D. B.; Qian, Y.; VanBuren, V.; Falco, G.; Martin, P. R.; Stagg, C. A.; Bassey, U. C. (2003) Transcriptome analysis of mouse stem cells and early embryos. *PLoS Biol*, 1, (3), E74.

17. Shen, M. M.; Leder, P.(1992) Leukemia Inhibitory Factor is Expressed by the Preimplantation Uterus and Selectively Blocks Primitive Ectoderm Formation in vitro. *Proc Natl Acad Sci USA*, 89, 8240-8244.

18. Gygi, S. P.; Rochon, Y.; Franza, B. R.; Aebersold, R.(1999) Correlation between Protein and mRNA Abundance in Yeast. *Mol Cell Biol* 19, 1720-1730.

19. Chen, G.; Gharib, T. G.; Huang, C. C.; Taylor, J. M. G.; Misek, D. E.; Kardia, S. L. R.; Giordano, T. J.; Iannettoni, M. D.; Orringer, M. B.; Hanash, S. M. (2002) Discordant Protein and mRNA Expression in Lung Adenocarcinomas. *Mol Cell Proteomics,* 1, 304-313.

20. Mootha, V. K.; Bunkenborg, J.; Olsen, J. V.; Hjerrild, M.; Wisniewski, J. R.; Stahl, E.; Bolouri, M. S.; Ray, H. N.; Sihag, S.; Kamal, M. (2003) Integrated Analysis of Protein Composition, Tissue Diversity, and Gene Regulation in Mouse Mitochondria. *Cell,* 115, 629-640.

21. Corbin, R. W.; Paliy, O.; Yang, F.; Shabanowitz, J.; Platt, M.; Lyons, C. E.; Root, K.; McAuliffe, J.; Jordan, M. I.; Kustu, S.(2003) Toward a protein profile of Escherichia coli: Comparison to its transcription profile. *Proc Natl Acad Sci USA,* 100, 9232-9237.

22. Kislinger, T.; Cox, B.; Kannan, A.; Chung, C.; Hu, P.; Ignatchenko, A.; Scott, M. S.; Gramolini, A. O.; Morris, Q.; Hallett, M. T. (2006) Global Survey of Organ and Organelle Protein Expression in Mouse: Combined Proteomic and Transcriptomic Profiling. *Cell,* 125, 173-186.

23. Schmidt, M. W., Hosueman, Andres, Ivanov, Alexander, R., Wolf, Dieter, A.(2007) Comparative proteomic and transcriptomic profiling of the fission yeast *Schizosaccharomyces pombe*. *Mol Syst Biol,* 3, 1-12.

24. Elliott, S. T.; Crider, D. G.; Graham, C. P.; Boheler, K. R.; Van Eyk, J. E.(2004) Two-dimensional gel electrophoresis database of cultured murine R1 embryonic stem cells. *Proteomics,* 4, 3813-3832.

25. Nagano, K.; Taoka, M.; Yamauchi, Y.; Itagaki, C.; Shinkawa, T.; Nunomura, K.; Okamura, N.; Takahashi, N.; Izumi, T.; Isobe, T.(2005) Large-scale identification of proteins expressed in mouse embryonic stem cells. *Proteomics,* 5, 1346-1361.

26. Van Hoof, D.; Passier, R.; Ward-Van Oostwaard, D.; Pinkse, M. W. H.; Heck, A. J. R.; Mummery, C. L.; Krijgsveld, J. (2006) A Quest for Human and Mouse Embryonic Stem Cell-specific Proteins. *Mol Cell Proteomics,* 5, 1261-1273.

27. Liu, H.; Sadygov, R. G.; Yates Iii, J. R. (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem,* 76, 4193-4201.

28. Robertson, E. J.(1987) Teratocarcinomas and embryonic stem cells: A practical approach. In Robertson, E. J., Ed. IRL Press, pp 71-112.

29. Weatherly, D. B.; Atwood, J. A.; Minning, T. A.; Cavola, C.; Tarleton, R. L.; Orlando, R. (2005) A Heuristic Method for Assigning a False-discovery Rate for Protein Identifications from Mascot Database Search Results. *Mol Cell Proteomics,* 4, 762-772.

30. Zybailov, B.; Mosley, A. L.; Sardiu, M. E.; Coleman, M. K.; Florens, L.; Washburn, M. P. (2006) Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. *J Proteome Res,* 5, 2339-2347.

31. Horton, P.; Park, K. J.; Obayashi, T.; Nakai, K., Protein Subcellular Localization Prediction with WoLF PSORT. (2006) *Proc 4th annual Asia PaciWcBioinformatics Conference APBC06, Taipei, Taiwan* 39-48.

32. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S.(1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551-67.

33. Stead, E.; White, J.; Faast, R.; Conn, S.; Goldstone, S.; Rathjen, J.; Dhingra, U.; Rathjen, P.; Walker, D.; Dalton, S.(2002) Pluripotent cell division cycles are driven by ectopic Cdk2, cyclin A/E and E2F activities. *Oncogene,* 21, 8320-8333.

34. White, J.; Stead, E.; Faast, R.; Conn, S.; Cartwright, P.; Dalton, S.(2005) Developmental Activation of the Rb-E2F Pathway and Establishment of Cell Cycle Regulated Cdk Activity During Embryonic Stem Cell Differentiation. *Mol Biol Cell* 16, 2018–2027.

35. Katsetos, C. D.; Herman, M. M.; Mork, S. J. (2003) Class III beta-tubulin in human development and cancer. *Cell Motil Cytoskeleton,* 55, 77-96.

36. Hendrickson, E. L.; Xia, Q.; Wang, T.; Leigh, J. A.; Hackett, M. (2006) Comparison of spectral counting and metabolic stable isotope labeling for use with quantitative microbial proteomics. *Analyst,* 131, 1335-1345.

37. Old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G. (2005) Comparison of Label-free Methods for Quantifying Human Proteins by Shotgun Proteomics. *Mol Cell Proteomics***,** 4, 1487-1502.

38. Adjaye, J.; Huntriss, J.; Herwig, R.; BenKahla, A.; Brink, T. C.; Wierling, C.; Hultschig, C.; Groth, D.; Yaspo, M. L.; Picton, H. M.(2005) Primary Differentiation in the Human Blastocyst: Comparative Molecular Portraits of Inner Cell Mass and Trophectoderm Cells. *Stem Cells***,** 23, 1514-1525.

39. Kim, S. K.; Suh, M. R.; Yoon, H. S.; Lee, J. B.; Oh, S. K.; Moon, S. Y.; Moon, S. H.; Lee, J. Y.; Hwang, J. H.; Cho, W. J. (2005) Identification of Developmental Pluripotency Associated 5 Expression in Human Pluripotent Stem Cells. *Stem Cells, 23, 458-462.*

40. Deng, J. M.; Behringer, R. R.(1995) An insertional mutation in theBTF3 transcription factor gene leads to an early postimplantation lethality in mice. *Transgenic Research* 4, 264-269.

41. Okita, K.; Kiyonari, H.; Nobuhisa, I.; Kimura, N.; Aizawa, S.; Taga, T. (2004) Targeted disruption of the mouse ELYS gene results in embryonic death at peri-implantation development. *Genes Cells,* 9, 1083-1091.

42. Kimura, N.; Takizawa, M.; Okita, K.; Natori, O.; Igarashi, K.; Ueno, M.; Nakashima, K.; Nobuhisa, I.; Taga, T. (2002) Identification of a novel transcription factor, ELYS, expressed predominantly in mouse foetal haematopoietic tissues. *Genes Cells,* 7, 435-446.

43. Bradshaw, A. D.; Sage, E. H. (2001) SPARC, a matricellular protein that functions in cellular differentiation and tissue response to injury. *J Clin Invest***,** 107, 1049-1054.

44. Sage, H.; Vernon, R. B.; Decker, J.; Funk, S.; Iruela-Arispe, M. L.(1989) Distribution of the calcium-binding protein SPARC in tissues of embryonic and adult mice. *J Histochem Cytochem,* 37, 819-829.

45. Gilmour, D. T.; Lyon, G. J.; Carlton, M. B. L.; Sanes, J. R.; Cunningham, J. M.; Anderson, J. R.; Hogan, B. L. M.; Evans, M. J.; Colledge, W. H.(1998) Mice deficient for the secreted

glycoprotein SPARC/osteonectin/BM40 develop normally but show severe age-onset cataract formation and disruption of the lens. *The EMBO Journal,* 17, 1860-1870.

46. Brekken, R. A.; Puolakkainen, P.; Graves, D. C.; Workman, G.; Lubkin, S. R.; Sage, E. H.(2003) Enhanced growth of tumors in SPARC null mice is associated with changes in the ECM. *J Clin Invest,* 111, 487-495.

47. Morgan, R., Sargent, Michael G. (1997) The role in neural patterning of translation initiation factor eIF4AII; induction of neural fold genes. *Development,* 124, 2751-2760.

48. Schofield, J. H., D.; Davies, K.; Buckingham, M.; Edwards, Y.H.(1993) Expression of the dystrophin-related protein (utrophin) gene during mouse development. *Develop Dyn,* 198, 254-264.

49. Ling, Q.; Jacovina, A. T.; Deora, A.; Febbraio, M.; Simantov, R.; Silverstein, R. L.; Hempstead, B.; Mark, W. H.; Hajjar, K. A.(2004) Annexin II regulates fibrin homeostasis and neoangiogenesis in vivo. *J Clin Invest,* 113, 38-48.

50. Kimura, E.; Hidaka, K.; Kida, Y.; Morisaki, H.; Shirai, M.; Araki, K.; Suzuki, M.; Yamamura, K. I.; Morisaki, T. (2004) Serine-arginine-rich nuclear protein Luc7l regulates myogenesis in mice. *Gene,* 341, 41-47.

51. Ruiz, J. C.; Robertson, E. J. (1994) The expression of the receptor-protein tyrosine kinase gene, eck, is highly restricted during early mouse development. *Mech Dev,* 46, 87-100.

52. Lindberg, R. A.; Hunter, T. (1990) cDNA cloning and characterization of eck, an epithelial cell receptor protein-tyrosine kinase in the eph/elk family of protein kinases. *Mol Cell Biol,* 10, 6316-6324.

53. Brantley-Sieders, D. M.; Caughron, J.; Hicks, D.; Pozzi, A.; Ruiz, J. C.; Chen, J. (2004) EphA2 receptor tyrosine kinase regulates endothelial cell migration and vascular assembly through phosphoinositide 3-kinase-mediated Rac1 GTPase activation. *J Cell Sci,* 117, 2037-2049.

54. Chu, P. H.; Ruiz-Lozano, P.; Zhou, Q.; Cai, C.; Chen, J. (2000) Expression patterns of FHL/SLIM family members suggest important functional roles in skeletal muscle and cardiovascular system. *Mech Dev,* 95, 259-265.

55. McGrath, M. J.; Mitchell, C. A.; Coghill, I. D.; Robinson, P. A.; Brown, S. (2003) Skeletal muscle LIM protein 1 (SLIM1/FHL1) induces a5ß1-integrin-dependent myocyte elongation. *Am J Physiol Cell Physiol,* 285, C1513-C1526.

56. Yamanaka, S.; Zhang, X. Y.; Maeda, M.; Miura, K.; Wang, S.; Farese Jr, R. V.; Iwao, H.; Innerarity, T. L. (2000) Essential role of NAT1/p97/DAP5 in embryonic differentiation and the retinoic acid pathway. *The EMBO Journal,* 19, 5533-5541.

57. Sousa-Nunes, R.; Rana, A. A.; Kettleborough, R.; Brickman, J. M.; Clements, M.; Forrest, A.; Grimmond, S.; Avner, P.; Smith, J. C.; Dunwoodie, S. L. (2003) Characterizing Embryonic Gene Expression Patterns in the Mouse Using Nonredundant Sequence-Based Selection. *Genome Res,* 13, 2609-2620.

58. Noma, T.; Kanai, Y.; Kanai-Azuma, M.; Ishii, M.; Fujisawa, M.; Kurohmaru, M.; Kawakami, H.; Wood, S. A.; Hayashi, Y. (2002) Stage-and sex-dependent expressions of Usp9x, an X-linked mouse ortholog of Drosophila Fat facets, during gonadal development and oogenesis in mice. *Gene Expr Patterns,* 2, 87-91.

59. Igakura, T.; Kadomatsu, K.; Kaname, T.; Muramatsu, H.; Fan, Q. W.; Miyauchi, T.; Toyama, Y.; Kuno, N.; Yuasa, S.; Takahashi, M. (1998) A Null Mutation in Basigin, an Immunoglobulin Superfamily Member, Indicates Its Important Roles in Peri-implantation Development and Spermatogenesis. *Dev Biol,* 194, 152-165.

60. Xiao, L. J.; Chang, H.; Ding, N. Z.; Ni, H.; Kadomatsu, K.; Yang, Z. M.(2002) Basigin expression and hormonal regulation in mouse uterus during the peri-implantation period. *MolReprod Devel,* 63, 47-54.

61. I Yan, Li, Zucker, Stanley, Toole, Bryan P. (2005) Roles of the multifunctional glycoprotein, emmprin (basigin; CD147), in tumour progression. *Thromb Haemost,* 93, 199-204.

62. Glatz, J. F.; Storch, J.(2001) Unravelling the significance of cellular fatty acid-binding proteins. *Curr Opin Lipidol***,** 12, 267-274.

63. Binas, B.; Danneberg, H.; McWhir, J. I. M.; Mullins, L.; Clark, A. J.(1999) Requirement for the heart-type fatty acid binding protein in cardiac fatty acid utilization. *FASEB J,* 13, 805-812.

64. Haunerland, N. H.; Spener, F. (2004) Progress in Lipid Research. *Progress in Lipid Res***,** 43, 328-349.

65. Miner, J. H.; Yurchenco, P. D. (2004) Laminin functions in tissue morphogenesis. *Annu Rev Cell Dev Biol***,** 20, 255-284.

66. Hallmann, R.; Horn, N.; Selg, M.; Wendler, O.; Pausch, F.; Sorokin, L. M. (2005) Expression and Function of Laminins in the Embryonic and Mature Vasculature. *Physiol Rev***,** 85, 979-1000.

67. Smyth, N.; Vatansever, H. S.; Murray, P.; Meyer, M.; Frie, C.; Paulsson, M.; Edgar, D.(1999) Absence of Basement Membranes after Targeting the LAMC1 Gene Results in Embryonic Lethality Due to Failure of Endoderm Differentiation. *J Cell Biol* 144, 151-160.

68. Oh, J. Y.; Kim, J.(1999) ATP hydrolysis activity of the DEAD box protein Rok1p is required for in vivo ROK1 function. *Nucleic Acids Res* 27, 2753-2759.

69. Rocak, S.; Linder, P.(2004) DEAD-box proteins: the driving forces behind RNA metabolism. *Nat Rev Mol Cell Biol* 5, (3), 232-241.

70. Cordin, O.; Banroques, J.; Tanner, N. K.; Linder, P. (2006) The DEAD-box protein family of RNA helicases. *Gene***,** 367, 17-37.

71. Eberle, J.; Krasagakis, K.; Orfanos, C. E.(1997) Translation initiaton factor EIF-4A1 mRNA is consistently overexpressed in human melanoma cells in vitro *Int J Cancer***,** 71, 396-401.

72. Ji, P.; Diederichs, S.; Wang, W.; Böing, S.; Metzger, R.; Schneider, P. M.; Tidow, N.; Brandt, B.; Buerger, H.; Bulk, E. (2003) MALAT-1, a novel noncoding RNA, and thymosin bold italic beta 4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene,* 22, 8031-8041.

73. Okuda, A.; Fukushima, A.; Nishimoto, M.; Orimo, A.; Yamagishi, T.; Nabeshima, Y.; Kuro-o, M.; Nabeshima, Y.; Boon, K.; Keaveney, M. (1998) UTF1, a novel transcriptional coactivator expressed in pluripotent embryonic stem cells and extra-embryonic cells. *The EMBO Journal,* 17, 2019-2032.

74. Nishimoto, M.; Miyagi, S.; Yamagishi, T.; Sakaguchi, T.; Niwa, H.; Muramatsu, M.; Okuda, A. (2005) 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.

75. Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., Smith, A. (1998) Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct-4. *Cell,* 95, 379-391.

76. Abeyta, M. J.; Clark, A. T.; Rodriguez, R. T.; Bodnar, M. S.; Pera, R. A. R.; Firpo, M. T. (2004) Unique gene expression signatures of independently-derived human embryonic stem cell lines. *Hum Mol Genet,* 13, 601-608.

77. Fischle, W.; Wang, Y.; Allis, C. D.(2003) Histone and chromatin cross-talk. *Curr Opin Cell Biol,* 15, 172-183.

78. Resing, K. A.; Meyer-Arendt, K.; Mendoza, A. M.; Aveline-Wolf, L. D.; Jonscher, K. R.; Pierce, K. G.; Old, W. M.; Cheung, H. T.; Russell, S.; Wattawa, J. L.(2004) Improving reproducibility and sensitivity in identifying human proteins by shotgun proteomics. *Anal Chem,* 76, 3556-3568.

79. Li, E. (2002) Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet,* 3, 662-673.

80. DeSalle, L. M.; Pagano, M. (2001) Regulation of the G1 to S transition by the ubiquitin pathway. *FEBS Lett,* 490, 179-189.

81. Piva, R.; Liu, J.; Chiarle, R.; Podda, A.; Pagano, M.; Inghirami, G. (2002) In Vivo Interference with Skp1 Function Leads to Genetic Instability and Neoplastic Transformation. *Mol Cell Biol,* 22, 8375-8387.

82. Masuda, H.; Hosokawa, N.; Nagata, K. (1998) Expression and Localization of Collagen-Binding Stress Protein Hsp47 in Mouse Embryo Development: Comparison with Types I and II Collagen. *Cell Stress Chaperones,* 3, 256-264.

83. Nagai, N.; Hosokawa, M.; Itohara, S.; Adachi, E.; Matsushita, T.; Hosokawa, N.; Nagata, K. (2000) Embryonic Lethality of Molecular Chaperone Hsp47 Knockout Mice Is Associated with Defects in Collagen Biosynthesis. *J Cell Biol,* 150, 1499-1506.

84. Yang, Y.; Mahaffey, C. L.; Berube, N.; Frankel, W. N. (2006) Interaction between Fidgetin and Protein Kinase A-anchoring Protein AKAP95 Is Critical for Palatogenesis in the Mouse. *J Biol Chem* 281, 22352-22359.

85. Zeng, X.; Miura, T.; Luo, Y.; Bhattacharya, B.; Condie, B.; Chen, J.; Ginis, I.; Lyons, I.; Mejido, J.; Puri, R. K.(2004) Properties of Pluripotent Human Embryonic Stem Cells BG01 and BG02. *Stem Cells* 22, 292-312.

86. Lee, J.; Prohaska, J. R.; Thiele, D. J. (2001) Essential role for mammalian copper transporter Ctr1 in copper homeostasis and embryonic development. *Proc Natl Acad Sci USA,* 98, 6842-6847.

87. Shimizu, K.; Chen, W.; Ashique, A. M.; Moroi, R.; Li, Y. P. (2003) Molecular cloning, developmental expression, promoter analysis and functional characterization of the mouse CNBP gene. *Gene* 307, 51-62.

88. Takahashi, Y.; Kako, K.; Kashiwabara, S.; Takehara, A.; Inada, Y.; Arai, H.; Nakada, K.; Kodama, H.; Hayashi, J.; Baba, T. (2002) Mammalian Copper Chaperone Cox17p Has an Essential Role in Activation of Cytochrome c Oxidase and Embryonic Development. *Mol Cell Biol* 22, 7614-7621.

89. Kim, S.-K.; Suh, M. R.; Yoon, H. S.; Lee, J. B.; Oh, S. K.; Moon, S. Y.; Moon, S.-H.; Lee, J. Y.; Hwang, J. H.; Cho, W. J.; Kim, K.-S. (2005) Identification of Developmental Pluripotency Associated 5 Expression in Human Pluripotent Stem Cells. *Stem Cells,* 23, 458-462.

90. Adjaye, J.; Huntriss, J.; Herwig, R.; BenKahla, A.; Brink, T. C.; Wierling, C.; Hultschig, C.; Groth, D.; Yaspo, M.-L.; Picton, H. M.; Gosden, R. G.; Lehrach, H. (2005) Primary

Differentiation in the Human Blastocyst: Comparative Molecular Portraits of Inner Cell Mass and Trophectoderm Cells. *Stem Cells,* 23, 1514-1525.

91. Luo, J.; Sladek, R.; Bader, J. A.; Matthyssen, A.; Rossant, J.; Giguere, V. (1997) Placental abnormalities in mouse embryos lacking the orphan nuclear receptor ERR-beta. *Nature,* 388, 778-782.

92. Mahe, D.; Fischer, N.; Decimo, D.; Fuchs, J. P. (2000) Spatiotemporal regulation of hnRNP M and 2H9 gene expression during mouse embryonic development. *Biochim Biophys Acta,* 1492, 414-424.

93. Sibug, R. M.; Datson, N.; Tijssen, A. M. I.; Morsink, M.; de Koning, J.; de Kloet, E. R.; Helmerhorst, F. M. (2007) Effects of urinary and recombinant gonadotrophins on gene expression profiles during the murine peri-implantation period. *Hum Reprod,* 22, 75-82.

94. Yu, T. S.; Moctezuma-Anaya, M.; Kubo, A.; Keller, G.; Robertson, S. (2002) The heart LIM protein gene (Hlp), expressed in the developing and adult heart, defines a new tissue-specific LIM-only protein family. *Mech Dev,* 116, 187-192.

95. Tzouanacou, E.; Tweedie, S.; Wilson, V. (2003) Identification of Jade1, a Gene Encoding a PHD Zinc Finger Protein, in a Gene Trap Mutagenesis Screen for Genes Involved in Anteroposterior Axis Development. *Mol Cell Biol,* 23, 8553-8562.

96. Okita, K.; Kiyonari, H.; Nobuhisa, I.; Kimura, N.; Aizawa, S.; Taga, T. (2004) Targeted disruption of the mouse ELYS gene results in embryonic death at peri-implantation development. *Genes Cells,* 9, 1083-1091.

97. Cui, X. S.; Li, X. Y.; Jeong, Y. J.; Jun, J. H.; Kim, N. H. (2006) Gene Expression of Cox5a, 5b, or 6b1 and Their Roles in Preimplantation Mouse Embryos. *Biol Reprod,* 74, 601-610.

98. Ibdah, J. A.; Paul, H.; Zhao, Y.; Binford, S.; Salleng, K.; Cline, M.; Matern, D.; Bennett, M. J.; Rinaldo, P.; Strauss, A. W. (2001) Lack of mitochondrial trifunctional protein in mice causes neonatal hypoglycemia and sudden death. *J Clin Invest,* 107, 1403-1409.

99. Aherne, A.; Kennan, A.; Kenna, P. F.; McNally, N.; Lloyd, D. G.; Alberts, I. L.; Kiang, A. S.; Humphries, M. M.; Ayuso, C.; Engel, P. C. (2004) On the molecular pathology of neurodegeneration in IMPDH1-based retinitis pigmentosa. *Hum Mol Genet* 13, 641-650.

100. Ciaudo, C.; Bourdet, A.; Cohen-Tannoudji, M.; Dietz, H. C.; Rougeulle, C.; Avner, P. (2006) Nuclear mRNA Degradation Pathway (s) Are Implicated in Xist Regulation and X Chromosome Inactivation. *PLoS Genet,* 2, e94: 0874-0882.

101. Harbers, K.; Muller, U.; Grams, A.; Li, E.; Jaenisch, R.; Franz, T. (1996) Provirus integration into a gene encoding a ubiquitin-conjugating enzyme results in a placental defect and embryonic lethality. *Proc Natl Acad Sci U S A,* 93, 12412-1217.

102. Schwartz, D. R.; Homanics, G. E.; Hoyt, D. G.; Klein, E.; Abernethy, J.; Lazo, J. S. (1999) The neutral cysteine protease bleomycin hydrolase is essential for epidermal integrity and bleomycin resistance. *Proc Natl Acad Sci USA,* 96, 4680-4685.

103. Riethmacher, D.; Brinkmann, V.; Birchmeier, C. (1995) A Targeted Mutation in the Mouse E-Cadherin Gene Results in Defective Preimplantation Development. *Proc Natl Acad Sci USA,* 92, 855-859.

104. Vestweber, D.; Gossler, A.; Boller, K.; Kemler, R. (1987) Expression and distribution of cell adhesion molecule uvomorulin in mouse preimplantation embryos. *Dev Biol,* 124, 451-456.

105. Gentile, L.; Monti, M.; Sebastiano, V.; Merico, V.; Nicolai, R.; Calvani, M.; Garagna, S.; Redi, C. A.; Zuccotti, M. (2004) Single-cell quantitative RT-PCR analysis of Cpt1b and Cpt2 gene expression in mouse antral oocytes and in preimplantation embryos. *Cytogenet Genome Res,* 105, 215-221.

106. Lu, Z. H.; Books, J. T.; Ley, T. J. (2006) Cold Shock Domain Family Members YB-1 and MSY4 Share Essential Functions during Murine Embryogenesis. *Mol Cell Biol,* 26, 8410-8417.

107. Guo, H.; Miao, H.; Gerber, L.; Singh, J.; Denning, M. F.; Gilliam, A. C.; Wang, B. (2006) Disruption of EphA2 Receptor Tyrosine Kinase Leads to Increased Susceptibility to Carcinogenesis in Mouse Skin. *Cancer Res,* 66, 7050-7058.

108. Williams, T. M.; Williams, M. E.; Kuick, R.; Misek, D.; McDonagh, K.; Hanash, S.; Innis, J. W. (2005) Candidate downstream regulated genes of HOX group 13 transcription factors with and without monomeric DNA binding capability. *Dev Biol* 279, 462-480.

109. Ristevski, S.; O'Leary, D. A.; Thornell, A. P.; Owen, M. J.; Kolas, I.; Hertzog, P. J. (2003) The ETS Transcription Factor GABPa Is Essential for Early Embryogenesis. *Mol Cell Biol* 24, 5844-5849.

110. Tsai, R. Y. L.; McKay, R. D. G. (2002) A nucleolar mechanism controlling cell proliferation in stem cells and cancer cells. *Genes De,* 16, 2991-3003.

111. Pawlak, M. R.; Scherer, C. A.; Chen, J.; Roshon, M. J.; Ruley, H. E. (2000) Arginine N-Methyltransferase 1 Is Required for Early Postimplantation Mouse Development, but Cells Deficient in the Enzyme Are Viable. *Mol Cell Biol,* 20, 4859-4869.

112. Ogawa, H.; Wu, Q.; Komiyama, J.; Obata, Y.; Kono, T. (2006) Disruption of parental-specific expression of imprinted genes in uniparental fetuses. *FEBS Lett,* 580, 5377-5384.

113. Pereira, C. M.; Sattlegger, E.; Jiang, H. Y.; Longo, B. M.; Jaqueta, C. B.; Hinnebusch, A. G.; Wek, R. C.; Mello, L.; Castilho, B. A. (2005) IMPACT, a Protein Preferentially Expressed in the Mouse Brain, Binds GCN1 and Inhibits GCN2 Activation. *J Biol Chem,* 280, 28316-28323.

114. Shirahama-Noda, K.; Yamamoto, A.; Sugihara, K.; Hashimoto, N.; Asano, M.; Nishimura, M.; Hara-Nishimura, I. (2003) Biosynthetic processing of cathepsins and lysosomal degradation are abolished in asparaginyl endopeptidase-deficient mice. *J Biol Chem,* 278, 33194-33199.

115. Rossi, D. J.; Londesborough, A.; Korsisaari, N.; Pihlak, A.; Lehtonen, E.; Henkemeyer, M.; Mäkelä, T. P. (2001) Inability to enter S phase and defective RNA polymerase II CTD phosphorylation in mice lacking Mat1. *The EMBO Journal,* 20, 2844-2856.

116. Bonner, A. E.; Lemon, W. J.; You, M. (2003) Gene expression signatures identify novel regulatory pathways during murine lung development: implications for lung tumorigenesis. *J Med Genet,* 40, 408-417.

117. Shafi, R.; Iyer, S. P. N.; Ellies, L. G.; O'Donnell, N.; Marek, K. W.; Chui, D.; Hart, G. W.; Marth, J. D. (2000) The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. *Proc Natl Acad Sci USA,* 97, 5735-5739.

118. O'Donnell, N.; Zachara, N. E.; Hart, G. W.; Marth, J. D. (2004) Ogt-Dependent X-chromosome-linked protein glycosylation is a requisite modification in somatic cell function and embryo viability. *Mol Cell Biol,* 24, 1680-1690.

119. Yamamoto, H.; Tsukahara, K.; Kanaoka, Y.; Jinno, S.; Okayama, H. (1999) Isolation of a mammalian homologue of a fission yeast differentiation regulator. *Mol Cell Biol* 19, 3829-3841.

120. Boussadia, O.; Amiot, F.; Cases, S.; Triqueneaux, G.; Jacquemin-Sablon, H.; Dautry, F. (1997) Transcription of unr (upstream of N-ras) down-modulates N-ras expression i*n vivo*. *FEBS Lett,* 420, 20-24.

121. Sinasac, D. S.; Crackower, M. A.; Lee, J. R.; Kobayashi, K.; Saheki, T.; Scherer, S. W.; Tsui, L. C. (1999) Genomic structure of the adult-onset type II Citrullinemia Gene, SLC25A13, and cloning and expression of its mouse homologue. *Genomics,* 95, 289-292.

122. Sinasac, D. S.; Moriyama, M.; Jalil, M. A.; Begum, L.; Li, M. X.; Iijima, M.; Horiuchi, M.; Robinson, B. H.; Kobayashi, K.; Saheki, T. (2004) Slc25a13-knockout mice harbor metabolic deficits but fail to display hallmarks of adult-onset type II Citrullinemia. *Mol Cell Biol,* 24, 527-536.

123. Wakamiya, M.; Matsuura, T.; Liu, Y.; Schuster, G. C.; Gao, R.; Xu, W.; Sarkar, P. S.; Lin, X.; Ashizawa, T.(2006) The role of ataxin 10 in the pathogenesis of spinocerebellar ataxia type 10. *Neurology,* 67, 607-613.

124. Echtermeyer, F.; Streit, M.; Wilcox-Adelman, S.; Saoncella, S.; Denhez, F.; Detmar, M.; Goetinck, P. F. (2001) Delayed wound repair and impaired angiogenesis in mice lacking syndecan-4. *The J Clin Invest,* 107, R9-R14.

125. Pantaleon, M.; Kanai-Azuma, M.; Mattick, J. S.; Kaibuchi, K.; Kaye, P. L.; Wood, S. A.(2001) FAM deubiquitylating enzyme is essential for preimplantation mouse embryo development. *Mech Dev*, 109, 151-160.

126. Jouret, F.; Auzanneau, C.; Debaix, H.; Wada, G. H. S.; Pretto, C.; Marbaix, E.; Karet, F. E.; Courtoy, P. J.; Devuyst, O. (2005) Ubiquitous and kidney-specific subunits of vacuolar H+-ATPase are differentially expressed during nephrogenesis. *J Am Soc Nephrol*, 16, 3235-3246.

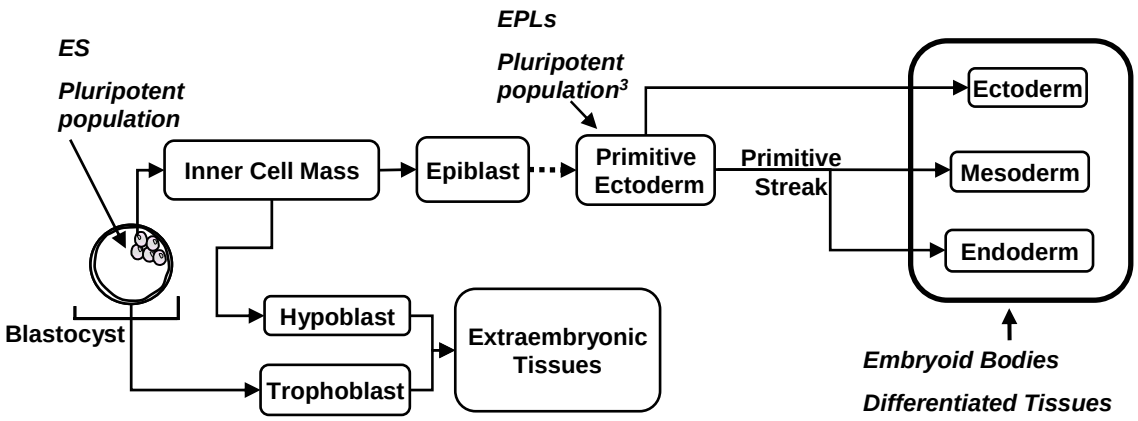
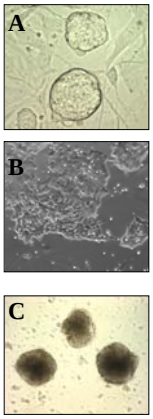


Figure 6.1. Adapted from References 1 and 3. Embryonic stem cells (ES), early primitive ectoderm-like cells (EPL) and embryoid bodies as an *in vitro* model of early embryogenesis. Embryonic stem cells are equivalent to the pluripotent inner cell mass of the 4.5 d.p.c. blastocyst. Early primitive ectoderm-like cells have a genetic profile similar to the primitive ectoderm, a second pluripotent population appearing at approximately 5.5 d.p.c. Embryoid bodies are heterogeneous clusters of all three tissue types, representing tissue differentiation occuring at 7.5 - 8 d.p.c.

Figure 6.2. Contrast light microscopy showing differentiating R1 mouse embryonic stem cells. A) Embryonic stem cells cultured on MEF plates B) embryonic stem cells cultured on 50% Med II, partially differentiating to form early primitive ectoderm-like cells , and C) embryoid bodies.



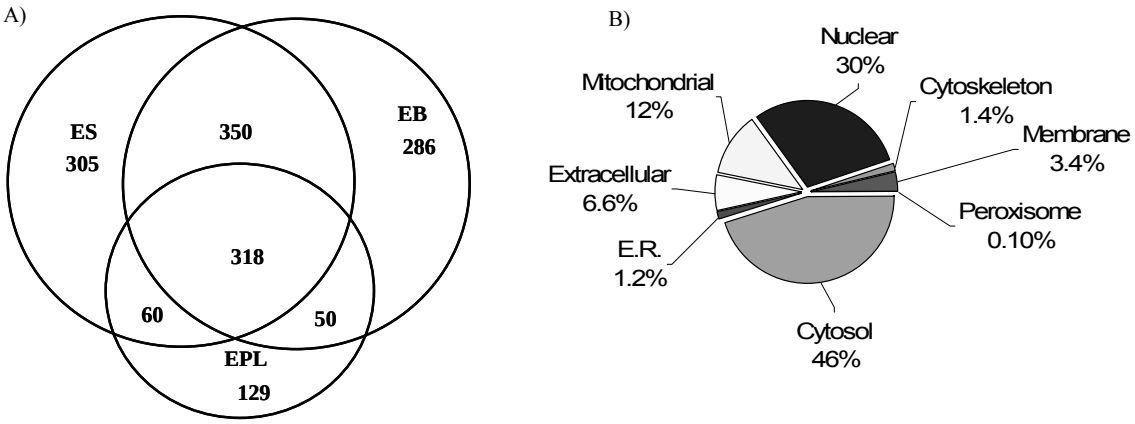


Figure 6.3. Characterization of the proteomes. A) Uniquely identified proteins were sorted according to single and shared proteomes. Approximately 52% of all proteins were shared between at least two proteomes, with 20% of the total population in common with all three proteomes. B) Predicted subcellular locations of the total population of uniquely identified proteins (1498 proteins) using the WoLF PSORT algorithm.

Table 6.1. Distribution of predicted, hypothetical and regulated proteins across the comparative and single proteomes.

Population	EPL/ES	EPL/EB	EB/ES	ES/EB/EPL			ES	EPL	EB
				EPL/ES	EB/ES	EPL/EB			
Total proteins	**60**	**50**	**350**	**318**					
Hypothetical	1	1	7	5			14	8	13
Predicted	13	11	51	44			67	30	63
Upregulated	12	16	34	47	33	46	-	-	-
Downregulated	1	6	44	60	28	61	-	-	-
No Change	47	28	272	211	257	211	-	-	-

Figure 6.4. Comparison of control set of protein between embryonic stem cells, early primitive ectoderm-like cells and embryoid bodies after normalization. From this evaluation, regulation was defined as having a greater than three fold change.

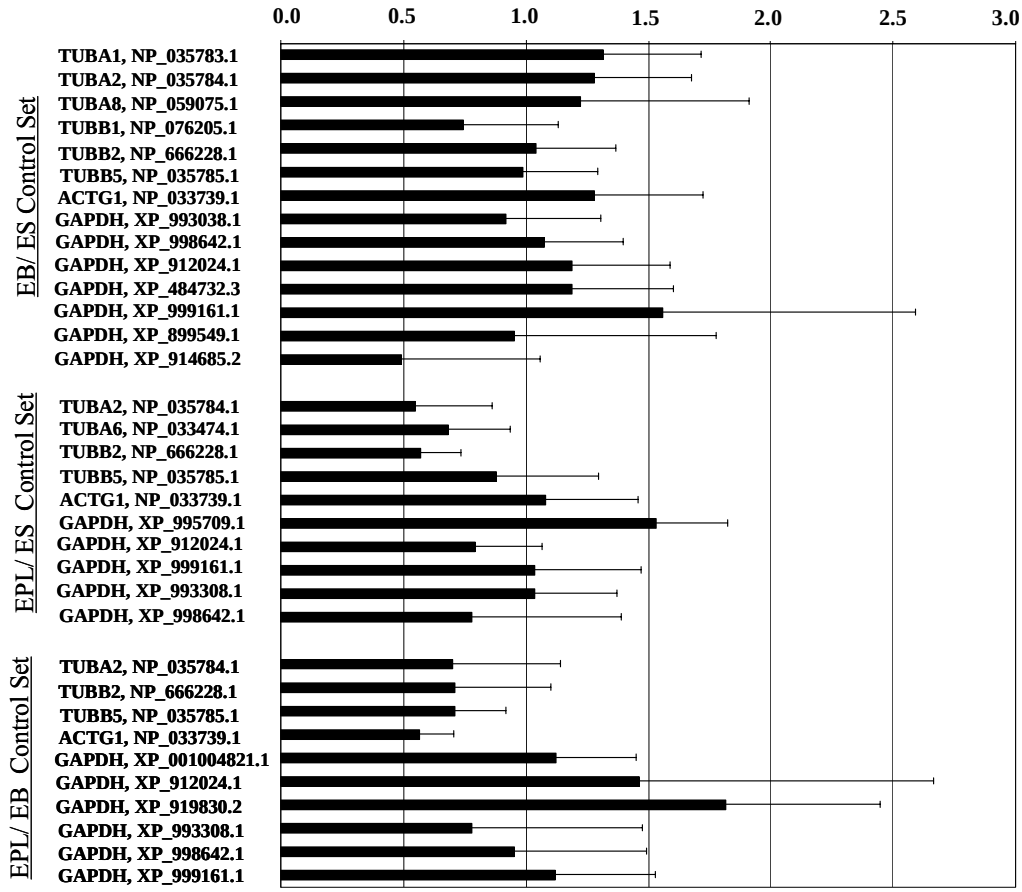


Table 6.2. Proteins identified only in the embryonic stem cell proteome. Relative percent abundance is calculated by dividing the spectral counts for an individual protein by the total spectral counts from all proteins in the embryonic stem cell proteome (unique and shared). Mascot score, number of peptides, and coverage are totals from nonredundant peptides seen over all three replicate analyses.

			Embryonic stem cells				Percent Relative Abundance in Identified Proteome
Gene	NCBI Accession	Protein Description	Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Embryonic stem cells AVG N_SC ± STDEV	
HIBADH	NP_663542.1	3-hydroxyisobutyrate dehydrogenase, mitochondrial precursor	99	1	4	4.94E-05 ± 0.00E+00	0.005
NIPSNAP1	NP_032724.1	4-nitrophenylphosphatase domain and non-neuronal SNAP25-like protein homolog 1	82	1	8	1.46E-04 ± 4.12E-05	0.015
AKAP8	NP_062748.2	A kinase anchor protein 8	105	1	2	9.64E-05 ± 1.02E-04	0.010
ACAA2	NP_803421.1	acetyl-Coenzyme A acyltransferase 2 (mitochondrial 3-oxoacyl-Coenzyme A thiolase)	77	1	4	1.67E-04 ± -	0.017
ACP6	NP_062774.2	acid phosphatase 6, lysophosphatidic	118	3	11	7.92E-05 ± -	0.008
ANP32A	NP_033802.2	acidic (leucine-rich) nuclear phosphoprotein 32 family, member A	273	6	23	5.14E-04 ± 2.71E-04	0.052
ARPC5	NP_080645.1	actin related protein 2/3 complex, subunit 5	101	1	16	2.19E-04 ± -	0.022
ACOT9	NP_062710.2	acyl-Coenzyme A thioesterase 2, mitochondrial	64	1	3	1.13E-04 ± -	0.011
AP3B1	NP_033810.1	adaptor-related protein complex 3, beta 1 subunit	53	1	1	1.50E-05 ± -	0.002
ADAT1	NP_038953.1	adenosine deaminase, tRNA-specific 1	53	1	3	6.63E-05 ± -	0.007
AK2	NP_001029138.1	adenylate kinase 2 isoform a	123	2	11	2.08E-04 ± -	0.021
AKAP1	NP_033778.1	A-kinase anchor protein 1	77	1	3	1.93E-05 ± -	0.002
ALPL	NP_031457.1	alkaline phosphatase 2, liver	57	1	3	1.90E-04 ± -	0.019
ANLN	NP_082666.1	anillin	67	1	2	1.48E-05 ± -	0.001
ATOX1	NP_033850.1	antioxidant protein 1	132	2	49	1.22E-03 ± 3.44E-04	0.122
AATF	NP_062790.1	apoptosis antagonizing transcription factor	82	1	3	3.15E-05 ± -	0.003
AIP	NP_057875.1	aryl-hydrocarbon receptor-interacting protein	59	1	4	1.50E-04 ± -	0.015
ATP6V1H	NP_598587.2	ATPase, H+ transporting, lysosomal, V1 subunit H	53	1	5	6.85E-05 ± -	0.007
ATP6V1B1	NP_598918.1	ATPase, H+ transporting, V1 subunit B, isoform 1	118	1	4	6.45E-05 ± -	0.006
ATP1A2	NP_848492.1	ATPase, Na+/K+ transporting, alpha 2 polypeptide	133	2	4	4.06E-05 ± 1.15E-05	0.004
ABCF2	NP_038881.1	ATP-binding cassette, sub-family F (GCN20), member 2	125	2	5	6.15E-05 ± 1.52E-05	0.006
BTF3	NP_663430.2	basic transcription factor 3	271	3	24	2.35E-03 ± 3.54E-04	0.236
BTF3L4	NP_081729.1	basic transcription factor 3-like 4	56	1	19	2.10E-04 ± 1.48E-04	0.021
BAX	NP_031553.1	Bcl2-associated X protein	62	1	7	8.62E-04 ± -	0.087
BCAS3	NP_619622.2	breast carcinoma amplified sequence 3-like	82	1	2	1.78E-05 ± -	0.002
SSSCA1	NP_065237.3	C184L-22	89	1	14	3.33E-04 ± -	0.033
CNN3	NP_082320.1	calponin 3, acidic	89	2	7	3.01E-04 ± -	0.030
CTNNB1	NP_031640.1	catenin (cadherin associated protein), beta 1, 88kDa	89	1	2	1.06E-04 ± -	0.011
CTNND1	NP_031641.2	catenin (cadherin associated protein), delta 1	99	2	3	1.76E-05 ± -	0.002
EXOSC8	NP_081424.2	CBP-interacting protein 3	62	1	4	6.00E-05 ± -	0.006
CNOT8	NP_081225.1	CCR4-NOT transcription complex, subunit 8	70	1	5	1.13E-04 ± -	0.011
CDCA3	NP_038566.1	cell division cycle associated 3	106	1	7	6.22E-05 ± -	0.006
C20ORF77	NP_081710.1	cell-cycle related and expression-elevated protein in tumor	184	2	10	2.38E-04 ± 1.28E-04	0.024
ZNF9	NP_038521.1	cellular nucleic acid binding protein 1	216	3	22	7.75E-04 ± 1.02E-03	0.078
CHMP5	NP_084090.1	chromatin modifying protein 5	74	1	7	3.02E-04 ± -	0.030
RCC1 (includes EG:1104)	NP_598639.1	chromosome condensation 1	134	4	13	2.36E-04 ± -	0.024
CLASP2	NP_083909.1	CLIP-associating protein CLASP2	62	1	1	2.57E-05 ± -	0.003
COPB2	NP_056642.1	coatomer protein complex, subunit beta 2 (beta prime)	100	1	2	3.66E-05 ± 1.83E-05	0.004
COIL	NP_057915.1	coilin	55	1	2	5.82E-05 ± -	0.006
5830416A07RIK	NP_001025165.1	conserved nuclear protein Nhn1 isoform b	53	1	2	1.75E-05 ± -	0.002
COBL	NP_766084.2	cordon-bleu protein	57	1	1	2.40E-05 ± -	0.002
CTTN	NP_031829.2	cortactin	130	3	7	2.43E-04 ± 3.00E-04	0.024
CFDP1	NP_035931.1	craniofacial development protein 1	102	2	10	2.24E-04 ± -	0.023
COX17	NP_001017429.1	cytochrome c oxidase subunit XVII assembly protein homolog	83	1	25	1.18E-03 ± 1.30E-03	0.119
CYP51A1	NP_064394.2	cytochrome P450, family 51	55	1	3	3.29E-05 ± -	0.003
CIP29	NP_079640.1	cytokine induced protein 29 kDa	118	2	14	7.88E-05 ± -	0.008
DDX19A	NP_031942.1	Ddx19-like protein	127	2	8	8.66E-05 ± 7.35E-05	0.009
DDX42	NP_082350.2	DEAD (Asp-Glu-Ala-Asp) box polypeptide 42	128	3	7	7.15E-05 ± 4.33E-05	0.007
DIDO1	NP_780760.2	death inducer-obliterator 1 isoform 3	64	1	1	1.47E-05 ± -	0.001
DEK	NP_080176.1	DEK oncogene (DNA binding)	55	1	3	4.36E-05 ± -	0.004
DSTN	NP_062745.1	destrin	77	1	10	1.00E-04 ± -	0.010
DPPA5	NP_079550.2	developmental pluripotency associated 5	103	2	24	9.12E-04 ± 1.09E-03	0.092
DDI2	NP_001017966.1	DNA-damage inducible protein 2	141	3	13	6.22E-05 ± 2.93E-05	0.006
DNAJC11	NP_766292.1	DnaJ (Hsp40) homolog, subfamily C, member 11	79	1	6	5.06E-05 ± -	0.005
ZRF1	NP_033610.1	DnaJ (Hsp40) homolog, subfamily C, member 2	86	2	5	1.87E-04 ± -	0.019
DNAJC9	NP_598842.1	DnaJ homolog, subfamily C, member 9	74	2	5	3.20E-04 ± -	0.032
DYNC1I2	NP_034194.1	dynein, cytoplasmic, intermediate chain 2	62	1	4	2.70E-05 ± -	0.003
DYNC1LI1	NP_666341.1	dynein, cytoplasmic, light intermediate chain 1	84	2	7	3.16E-05 ± -	0.003
ELAC2	NP_075968.1	elaC homolog 2	63	1	2	1.99E-05 ± -	0.002
CDCA8	NP_080836.2	embryonic stem cell-related	58	1	5	5.73E-05 ± -	0.006
ERH	NP_031977.1	enhancer of rudimentary homolog	147	3	22	3.18E-04 ± -	0.032
EZH2	NP_031997.1	enhancer of zeste homolog 2	80	1	2	2.22E-05 ± -	0.002
ECHDC2	NP_081004.1	enoyl Coenzyme A hydratase domain containing 2	67	1	7	6.73E-05 ± -	0.007
LEREPO4	NP_081210.2	erythropoietin 4 immediate early response	102	1	4	3.89E-05 ± -	0.004
ESRRB	NP_036064.2	estrogen related receptor, beta	147	3	9	9.56E-05 ± 2.70E-05	0.010
EIF1AX	NP_034250.2	eukaryotic translation initiation factor 1A	55	1	11	1.72E-03 ± -	0.173
EIF3S3	NP_542366.1	eukaryotic translation initiation factor 3, subunit 3 (gamma)	202	3	14	1.88E-04 ± -	0.019
EIF3S7	NP_061219.1	eukaryotic translation initiation factor 3, subunit 7 (zeta)	64	1	4	3.03E-05 ± -	0.003
EXOSC10	NP_057908.1	exosome component 10	52	1	2	1.87E-05 ± -	0.002
EXOSC3	NP_079789.1	exosome component 3	51	1	8	1.21E-04 ± -	0.012
FSCN1	NP_032010.1	fascin homolog 1, actin bundling protein (Strongylocentrotus) purpuratus)	60	1	3	5.04E-05 ± 2.37E-05	0.005
FTO	NP_036066.1	fatso	106	2	6	9.89E-05 ± -	0.010
FIP1L1	NP_077145.2	FIP1 like 1	56	1	3	3.04E-05 ± -	0.003
FRAP1	NP_064393.1	FK506 binding protein 12-rapamycin associated protein 1	66	1	1	6.49E-06 ± -	0.001
FKBP2	NP_032046.1	FK506 binding protein 2	57	1	9	1.18E-04 ± -	0.012
FKBP3	NP_038930.1	FK506 binding protein 3	114	2	17	5.17E-04 ± -	0.052
FEN1	NP_032025.2	flap structure specific endonuclease 1	202	4	14	2.18E-04 ± -	0.022
FNBP4	NP_061298.1	formin binding protein 4	59	1	3	1.54E-05 ± -	0.002
SIAHBP1	NP_598452.2	fuse-binding protein-interacting repressor isoform b	221	4	13	4.24E-04 ± 2.64E-04	0.043
GTF2F1	NP_598562.1	general transcription factor IIF, polypeptide 1	103	1	4	3.26E-05 ± -	0.003
GRWD1	NP_700468.1	glutamate-rich WD repeat containing 1	155	2	9	6.45E-05 ± 2.79E-05	0.006

Proteins in found in embryonic stem cells only, continued.

Gene	NCBI Accession	Protein Description	Embryonic stem cells: Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Embryonic stem cells: AVG N_SC ± STDEV	Percent Relative Abundance in Identified Proteome
GPX1	NP_032186.2	glutathione peroxidase 1	88	1	9	1.65E-04 ± -	0.017
GSTO1	NP_034492.1	glutathione S-transferase omega 1	65	1	4	6.90E-05 ± -	0.007
GRHPR	NP_525028.1	glyoxylate reductase/hydroxypyruvate reductase	117	2	10	1.01E-04 ± 0.00E+00	0.010
GOLPH3	NP_079949.1	golgi phosphoprotein 3	59	1	5	5.55E-05 ± -	0.006
TNRC15	NP_666224.2	Grb10 interacting GYF protein 2	83	1	2	2.69E-04 ± -	0.027
GTPBP1	NP_038846.2	GTP binding protein 1	80	1	2	9.08E-05 ± 1.43E-05	0.009
GTPBP4	NP_081276.2	GTP binding protein 4	98	1	2	5.22E-05 ± -	0.005
GAPVD1	NP_079985.2	GTPase activating protein and VPS9 domains 1	127	1	1	4.61E-05 ± -	0.005
HSPA14	NP_056580.2	heat shock 70kDa protein 14 isoform 1	96	2	5	2.28E-04 ± -	0.023
HNRPM	NP_084080.1	heterogeneous nuclear ribonucleoprotein M	278	6	10	1.21E-04 ± 3.47E-05	0.012
HMGB1	NP_034569.1	high mobility group box 1	233	5	29	2.82E-04 ± 4.44E-05	0.028
BAT3	NP_476512.1	HLA-B-associated transcript 3	73	1	1	1.29E-04 ± -	0.013
HSPBAP1	NP_780320.1	Hspb associated protein 1	80	1	4	6.85E-05 ± 0.00E+00	0.007
HSD17B4	NP_032318.2	hydroxysteroid (17-beta) dehydrogenase 4	101	2	4	2.25E-05 ± -	0.002
-	NP_862902.1	hypothetical protein LOC101994	72	1	2	4.13E-05 ± -	0.004
-	NP_001028430.1	hypothetical protein LOC215821	77	2	1	4.58E-05 ± -	0.005
-	NP_663437.1	hypothetical protein LOC219072	76	1	7	4.56E-05 ± -	0.005
-	NP_665835.1	hypothetical protein LOC238330	105	2	4	3.20E-05 ± 1.51E-05	0.003
-	NP_766531.2	hypothetical protein LOC268420	103	1	8	4.19E-05 ± -	0.004
-	NP_766535.1	hypothetical protein LOC268490	67	2	9	3.40E-04 ± -	0.034
-	NP_082319.1	hypothetical protein LOC52846	66	1	5	2.29E-04 ± -	0.023
-	NP_848727.1	hypothetical protein LOC66455	87	1	9	2.70E-04 ± -	0.027
-	NP_080277.1	hypothetical protein LOC67153	92	1	7	5.37E-05 ± -	0.005
-	NP_080305.1	hypothetical protein LOC67201	67	1	6	5.55E-05 ± -	0.006
-	NP_083630.1	hypothetical protein LOC72083	67	1	13	1.04E-04 ± -	0.010
-	NP_776105.1	hypothetical protein LOC72148	57	1	8	1.36E-04 ± 6.43E-05	0.014
-	NP_083996.1	hypothetical protein LOC76737	75	1	5	2.84E-04 ± -	0.028
-	NP_001028363.1	hypothetical protein LOC97243	54	1	10	7.59E-05 ± -	0.008
IMP3	NP_598737.1	IMP3, U3 small nucleolar ribonucleoprotein, homolog	69	1	8	9.00E-05 ± -	0.009
IDH3B	NP_570954.1	isocitrate dehydrogenase 3, beta subunit	69	1	3	4.31E-05 ± -	0.004
TNPO2	NP_663365.2	karyopherin (importin) beta 2b	72	2	2	3.69E-05 ± 2.61E-05	0.004
KLHDC4	NP_663580.1	kelch domain containing 4	100	1	2	7.56E-05 ± 4.33E-05	0.008
KIF3C	NP_032471.2	kinesin family member 3C	100	3	2	6.45E-04 ± -	0.065
KLF4	NP_034767.1	Kruppel-like factor 4 (gut)	71	1	4	3.49E-05 ± -	0.004
KLF5	NP_033899.2	Kruppel-like factor 5	66	1	7	3.71E-05 ± -	0.004
LALBA	NP_034809.1	lactalbumin, alpha	76	2	4	6.37E-04 ± 8.18E-05	0.064
LMNB2	NP_034852.1	lamin B2	101	2	4	6.94E-05 ± 1.96E-05	0.007
PSIP1	NP_598709.1	lens epithelium-derived growth factor	189	4	10	3.13E-04 ± -	0.031
CRIP2	NP_077185.1	LIM only protein HLP	63	1	15	2.39E-04 ± 2.25E-04	0.024
LSM4	NP_056631.1	LSM4 homolog, U6 small nuclear RNA associated	54	1	7	1.21E-04 ± -	0.012
M6PR	NP_034879.2	mannose-6-phosphate receptor, cation dependent	87	1	6	5.95E-05 ± -	0.006
MSTO1	NP_659147.1	misato	54	1	2	2.98E-05 ± -	0.003
MRPS7	NP_079581.1	mitchondrial ribosomal protein S7	82	2	9	2.33E-03 ± -	0.233
MRPL10	NP_080430.1	mitochondrial ribosomal protein L10	61	1	7	1.26E-04 ± -	0.013
MRPS27	NP_776118.2	mitochondrial ribosomal protein S27	96	2	6	7.98E-05 ± -	0.008
MORF4L2	NP_062742.3	mortality factor 4 like 2	60	1	6	5.75E-05 ± -	0.006
SAP130	NP_766553.1	mSin3A-associated protein 130	60	1	2	1.57E-05 ± -	0.002
MSI2	NP_473384.1	Musashi homolog 2	60	1	4	9.57E-05 ± -	0.010
MINA	NP_080186.2	myc induced nuclear antigen	60	1	3	3.56E-05 ± -	0.004
MARCKS	NP_032564.1	myristoylated alanine rich protein kinase C substrate	132	2	20	5.36E-05 ± 0.00E+00	0.005
NANS	NP_444409.1	N-acetylneuraminic acid synthase (sialic acid synthase)	97	2	9	1.38E-04 ± -	0.014
NDUFB10	NP_080960.1	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 10	80	1	11	4.08E-04 ± 5.43E-05	0.041
NAPA	NP_080174.1	N-ethylmaleimide sensitive fusion protein attachment protein alpha	89	1	5	1.12E-04 ± -	0.011
NISCH	NP_073147.1	nischarin	52	1	1	1.22E-05 ± -	0.001
NOSIP	NP_079809.1	nitric oxide synthase interacting protein	124	2	11	1.92E-04 ± 1.17E-04	0.019
NFATC2IP	NP_035030.1	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2 interacting protein	60	1	4	4.02E-05 ± -	0.004
NPLOC4	NP_955763.1	nuclear protein localization 4	72	1	3	2.87E-05 ± -	0.003
NUCB1	NP_032775.1	nucleobindin 1	141	2	6	7.28E-05 ± -	0.007
NOL1	NP_620086.1	nucleolar protein 1	165	3	7	9.39E-05 ± 4.43E-05	0.009
NOL5A	NP_077155.1	nucleolar protein 5A	219	4	12	1.62E-04 ± 1.19E-04	0.016
NOL6	NP_631983.1	nucleolar RNA-associated protein short isoform	220	3	6	6.28E-05 ± -	0.006
NUP35	NP_081367.1	nucleoporin 35	133	2	10	2.21E-04 ± 2.94E-05	0.022
NUP54	NP_899248.1	nucleoporin 54	116	2	6	1.30E-04 ± -	0.013
NUP43	NP_663752.1	nucleoporin Nup43	51	1	8	3.85E-04 ± -	0.039
NUDT5	NP_058614.1	nudix (nucleoside diphosphate linked moiety X)-type motif 5	62	1	7	7.59E-05 ± -	0.008
PAF1	NP_062331.2	Paf1, RNA polymerase II associated factor, homolog	63	1	2	6.19E-05 ± -	0.006
PPIC	NP_032934.1	peptidylprolyl isomerase C	92	2	8	7.42E-04 ± 5.52E-05	0.074
PRDX5	NP_036151.1	peroxiredoxin 5 precursor	52	1	11	1.58E-04 ± 0.00E+00	0.016
PECR	NP_076012.2	peroxisomal trans 2-enoyl CoA reductase	53	1	6	3.28E-04 ± -	0.033
PHF17	NP_758507.3	PHD zinc finger protein Jade1	66	1	2	5.29E-05 ± 5.73E-05	0.005
PNN	NP_032917.1	pinin	54	1	2	4.56E-05 ± -	0.005
PABPN1	NP_062275.1	poly(A) binding protein, nuclear 1	101	2	19	5.48E-05 ± -	0.006
POLR2C	NP_033116.2	polymerase (RNA) II (DNA directed) polypeptide C	85	1	7	1.20E-04 ± 0.00E+00	0.012
POLR2D	NP_081278.1	polymerase (RNA) II (DNA directed) polypeptide D isoform 1	122	2	22	1.17E-04 ± -	0.012
GCN1L1	XP_358358.4	PREDICTED: GCN1 general control of amino-acid synthesis 1-like 1 isoform 1	296	6	4	2.89E-05 ± 1.29E-05	0.003
-	XP_001052174.1	PREDICTED: high mobility group box 2	185	4	17	2.70E-04 ± 1.20E-04	0.027
-	XP_918687.2	PREDICTED: hypothetical protein	118	3	8	3.35E-04 ± -	0.034
-	XP_920296.1	PREDICTED: hypothetical protein	51	1	13	9.85E-05 ± -	0.010
-	XP_283973.4	PREDICTED: hypothetical protein LOC230393 isoform 1	92	3	1	6.44E-05 ± -	0.006
-	XP_283218.5	PREDICTED: kinesin family member 13B isoform 1	129	2	2	8.98E-06 ± -	0.001
-	XP_127535.3	PREDICTED: serologically defined colon cancer antigen 10 isoform 1	116	2	7	7.06E-05 ± -	0.007
-	XP_132143.4	PREDICTED: signal recognition particle 72 isoform 1	152	3	7	4.93E-05 ± -	0.005
-	XP_001002026.1	PREDICTED: similar to 40S ribosomal protein S2	222	6	24	8.28E-04 ± 2.90E-04	0.083
-	XP_984175.1	PREDICTED: similar to 40S ribosomal protein S25	190	4	26	2.25E-03 ± -	0.226
-	XP_978957.1	PREDICTED: similar to 40S ribosomal protein S26	101	2	27	1.94E-03 ± 8.48E-04	0.195
-	XP_899567.1	PREDICTED: similar to 40S ribosomal protein S6 isoform 2	459	7	26	2.95E-03 ± 1.94E-03	0.296
-	XP_921694.1	PREDICTED: similar to 40S ribosomal protein SA (p40) (34/67 kDa laminin receptor) isoform 1	505	8	37	2.09E-03 ± 2.12E-04	0.210
-	XP_930844.2	PREDICTED: similar to 60S ribosomal protein L26 (Silica-induced gene 20 protein) (SIG-20)	108	2	8	5.52E-04 ± -	0.055
-	XP_485430.1	PREDICTED: similar to 60S ribosomal protein L3 (L4)	443	9	30	5.20E-04 ± 2.73E-04	0.052
-	XP_001004898.1	PREDICTED: similar to 60S ribosomal protein L5	184	5	12	1.15E-03 ± 9.87E-04	0.116
-	XP_914375.1	PREDICTED: similar to 60S ribosomal protein L7a (Surfeit locus protein 3)	199	5	19	7.88E-04 ± 4.37E-04	0.079
-	XP_917582.2	PREDICTED: similar to abnormal cell LINeage family member (lin-41)	54	1	2	1.94E-05 ± -	0.002
-	XP_996193.1	PREDICTED: similar to ADP/ATP translocase 1 (mANC1) isoform 2	101	2	13	3.89E-04 ± -	0.039
-	XP_916103.2	PREDICTED: similar to Afadin (Af-6 protein) isoform 4	112	3	3	4.57E-05 ± -	0.005
-	XP_979694.1	PREDICTED: similar to aldo-keto reductase family 1, member A4 (aldehyde reductase)	161	3	16	2.24E-04 ± 7.91E-05	0.022
-	XP_001002965.1	PREDICTED: similar to AP-2 complex subunit alpha-1	198	4	6	8.47E-05 ± -	0.009

Proteins in found in embryonic stem cells only, continued.

Gene	NCBI Accession	Protein Description	Embryonic stem cells: Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Embryonic stem cells: AVG N_SC ± STDEV	Percent Relative Abundance in Identified Proteome
-	XP_992936.1	PREDICTED: similar to carnitine deficiency-associated gene expressed in ventricle 3 isoform CDV3B	84	1	8	5.11E-04 ± 6.39E-05	0.051
-	XP_928308.1	PREDICTED: similar to CCR4-NOT transcription complex, subunit 1 isoform a isoform 30	103	3	2	3.47E-05 ± -	0.003
-	XP_982995.1	PREDICTED: similar to CG10889-PA	54	1	1	3.61E-05 ± -	0.004
-	XP_910951.1	PREDICTED: similar to CG11412-PA, isoform A isoform 3	60	1	4	1.36E-04 ± 0.00E+00	0.014
-	XP_977719.1	PREDICTED: similar to coiled-coil domain containing 16	86	1	6	4.56E-05 ± -	0.005
-	XP_990671.1	PREDICTED: similar to Cullin-associated NEDD8-dissociated protein 1	247	4	5	4.65E-05 ± -	0.005
-	XP_987982.1	PREDICTED: similar to Cytochrome c, somatic	79	1	13	6.31E-04 ± -	0.063
-	XP_996932.1	PREDICTED: similar to cytokine induced apoptosis inhibitor 1	78	1	5	5.36E-05 ± -	0.005
-	XP_620101.2	PREDICTED: similar to Ferritin light chain 1 (Ferritin L subunit 1)	240	3	63	1.52E-03 ± 5.71E-04	0.153
-	XP_986001.1	PREDICTED: similar to Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	113	4	15	3.62E-04 ± 1.71E-04	0.036
-	XP_983635.1	PREDICTED: similar to guanine nucleotide binding protein-like 3 (nucleolar)-like isoform 4	144	2	7	7.65E-05 ± 5.97E-05	0.008
-	XP_001000476.1	PREDICTED: similar to H2A histone family, member Z isoform 2	269	5	35	3.15E-03 ± 4.90E-04	0.316
-	XP_923168.1	PREDICTED: similar to H3 histone, family 3B	205	5	23	4.91E-03 ± 1.13E-03	0.493
-	XP_983732.1	PREDICTED: similar to heat shock protein 1 (chaperonin)	409	5	11	1.09E-03 ± 3.06E-04	0.110
-	XP_908932.1	PREDICTED: similar to heterogeneous nuclear ribonucleoprotein A3 isoform 5	850	15	36	2.38E-03 ± 1.16E-03	0.239
-	XP_997579.1	PREDICTED: similar to Heterogeneous nuclear ribonucleoprotein G (RNA-binding motif protein, X chromosom	144	3	23	2.72E-04 ± 7.70E-05	0.027
-	XP_891398.1	PREDICTED: similar to histone H2A	180	3	32	1.25E-02 ± 1.95E-03	1.257
-	XP_001003792.1	PREDICTED: similar to La-related protein 4 (La ribonucleoprotein domain family member 4)	60	1	2	4.07E-05 ± -	0.004
-	XP_997235.1	PREDICTED: similar to myosin, light polypeptide 6, alkali, smooth muscle and non-muscle	128	2	19	3.27E-04 ± 0.00E+00	0.033
-	XP_894449.1	PREDICTED: similar to NADH dehydrogenase (ubiquinone) Fe-S protein 3	109	2	10	1.26E-04 ± -	0.013
-	XP_989473.1	PREDICTED: similar to neuronal differentiation-related gene	114	2	6	9.85E-05 ± -	0.010
-	XP_996697.1	PREDICTED: similar to NSFL1 (p97) cofactor (p47)	108	2	5	1.68E-04 ± 4.52E-05	0.017
-	XP_910294.1	PREDICTED: similar to Nuclear protein SkiP (Ski-interacting protein)	54	1	3	3.09E-05 ± 0.00E+00	0.003
-	XP_001000561.1	PREDICTED: similar to optic atrophy 1 isoform 4	70	1	2	4.01E-05 ± -	0.004
-	XP_904983.1	PREDICTED: similar to peptidylprolyl isomerase G isoform 6	65	1	2	1.98E-04 ± -	0.020
-	XP_891322.1	PREDICTED: similar to Proteasome activator complex subunit 2	88	1	10	6.93E-05 ± -	0.007
-	XP_911955.2	PREDICTED: similar to protein phosphatase 1, regulatory (inhibitor) subunit 14B	129	2	49	1.84E-03 ± -	0.185
-	XP_921971.1	PREDICTED: similar to putative aminopeptidase Fxna isoform 6	52	1	1	1.84E-05 ± -	0.002
-	XP_999944.1	PREDICTED: similar to RAN, member RAS oncogene family	243	5	35	5.08E-04 ± 3.08E-04	0.051
-	XP_990136.1	PREDICTED: similar to RAS related protein 1b	77	1	3	3.96E-05 ± -	0.004
-	XP_983037.1	PREDICTED: similar to ribophorin II	98	1	3	7.87E-05 ± -	0.008
-	XP_899268.2	PREDICTED: similar to ribosomal protein L10	86	2	17	7.46E-04 ± -	0.075
-	XP_999014.1	PREDICTED: similar to ribosomal protein L21	70	1	9	5.17E-04 ± -	0.052
-	XP_910742.1	PREDICTED: similar to ribosomal protein L27	110	3	29	9.40E-04 ± 6.21E-04	0.094
-	XP_985262.1	PREDICTED: similar to ribosomal protein L36	183	3	14	6.22E-04 ± 1.35E-04	0.062
-	XP_994281.1	PREDICTED: similar to ribosomal protein S18	89	2	11	2.41E-03 ± 6.20E-04	0.242
-	XP_001000635.1	PREDICTED: similar to ribosomal protein S24	59	1	8	3.08E-03 ± 2.20E-03	0.309
-	XP_898612.1	PREDICTED: similar to RING1 and YY1 binding protein	102	1	7	1.46E-04 ± 3.15E-05	0.015
-	XP_990616.1	PREDICTED: similar to short coiled-coil protein (predicted)	71	1	9	2.02E-04 ± -	0.020
-	XP_981485.1	PREDICTED: similar to signal recognition particle 54 isoform 3	77	1	2	8.65E-05 ± -	0.009
-	XP_001005155.1	PREDICTED: similar to Signal transducer and activator of transcription 3	89	1	6	6.46E-05 ± -	0.006
-	XP_001006366.1	PREDICTED: similar to small nuclear ribonucleoprotein polypeptide A	52	1	9	5.20E-05 ± -	0.005
-	XP_001006276.1	PREDICTED: similar to Splicing factor, arginine/serine-rich 3	149	3	25	3.72E-03 ± 4.78E-04	0.374
-	XP_001003884.1	PREDICTED: similar to Tetratricopeptide repeat protein 5 (TPR repeat protein 5)	52	1	3	4.35E-04 ± 8.02E-05	0.044
-	XP_356994.2	PREDICTED: similar to ubiquitin A-52 residue ribosomal protein fusion product 1	153	4	27	4.97E-04 ± 4.14E-04	0.050
-	XP_910906.1	PREDICTED: similar to UPF0315 protein	69	1	10	4.63E-04 ± 2.81E-04	0.047
SBNO1	XP_355637.3	PREDICTED: sno, strawberry notch homolog 1 isoform 1	121	1	1	1.19E-05 ± -	0.001
SPTAN1	XP_993811.1	PREDICTED: spectrin alpha 2 isoform 14	315	6	4	6.16E-05 ± 3.87E-05	0.006
SF3A3	XP_620099.1	PREDICTED: splicing factor 3a, subunit 3 isoform 1	71	1	4	4.96E-05 ± 2.34E-05	0.005
POP1	NP_080616.1	processing of precursors 1 homolog isoform 2	146	3	3	2.45E-05 ± 1.15E-05	0.002
PDCD8	NP_036149.1	programmed cell death 8	51	1	2	5.41E-05 ± -	0.005
PSAP	NP_035309.2	prosaposin	73	2	4	3.86E-04 ± -	0.039
PSME1	NP_035319.1	proteasome (prosome, macropain) 28 subunit, alpha	51	1	5	1.13E-03 ± -	0.113
PSMA5	NP_036097.1	proteasome (prosome, macropain) subunit, alpha type 5	200	4	20	6.52E-04 ± 5.34E-04	0.065
PSMB7	NP_035317.1	proteasome (prosome, macropain) subunit, beta type 7	66	1	7	3.59E-04 ± -	0.036
PSMD2	NP_598862.1	proteasome 26S non-ATPase subunit 2	80	2	4	1.09E-04 ± -	0.011
PIN1	NP_075860.1	protein (peptidyl-prolyl cis/trans isomerase) NIMA-interacting 1	145	2	24	5.02E-04 ± -	0.050
PPFIA2	NP_796347.2	PTPRF interacting protein alpha 2	104	2	2	1.32E-05 ± -	0.001
ALDH18A1	NP_705782.1	pyrroline-5-carboxylate synthetase isoform 2	59	1	2	2.09E-05 ± -	0.002
PC	NP_032823.1	pyruvate carboxylase	56	1	2	1.41E-05 ± -	0.001
RAB40B	NP_631886.1	Rab40b, member RAS oncogene family	69	2	5	1.79E-04 ± -	0.018
RFC1	NP_035388.1	replication factor C 1	80	2	2	8.78E-05 ± -	0.009
RCN1	NP_033063.1	reticulocalbin 1	105	2	7	1.53E-04 ± -	0.015
RTN4	NP_918943.1	reticulon 4 isoform A	80	1	1	2.85E-05 ± -	0.003
ARHGEF7	NP_059098.2	Rho guanine nucleotide exchange factor (GEF7)	103	2	5	1.28E-04 ± -	0.013
ARHGDIB	NP_031512.1	Rho, GDP dissociation inhibitor (GDI) beta	67	1	8	2.48E-04 ± -	0.025
RPL24	NP_077180.1	ribosomal protein L24	209	4	27	9.84E-04 ± 8.84E-04	0.099
RPL29	NP_033108.1	ribosomal protein L29	79	2	13	1.34E-03 ± -	0.135
RPS20	NP_080423.1	ribosomal protein S20	108	2	11	3.48E-04 ± 9.83E-05	0.035
KIAA1008	NP_082591.1	RIKEN cDNA 2810028N01	112	2	5	2.92E-05 ± -	0.003
RBM19	NP_083038.1	RNA binding motif protein 19	182	3	4	3.48E-05 ± -	0.003
RBPMS	NP_062707.1	RNA binding protein gene with multiple splicing	251	4	30	3.78E-04 ± 5.94E-05	0.038
POLR1C	NP_033111.1	RNA polymerase 1-1	53	1	3	2.33E-04 ± -	0.023
POLR3B	NP_081699.1	RNA polymerase III subunit RPC2	57	1	1	1.46E-05 ± -	0.001
BXDC5	NP_081608.1	RNA processing factor 1 isoform 1	67	1	12	7.81E-05 ± -	0.008
RTCD1	NP_079793.1	RNA terminal phosphate cyclase domain 1	62	1	4	3.62E-04 ± -	0.036
SALL4	NP_958797.1	sal-like 4 isoform b	97	1	4	5.31E-05 ± -	0.005
"SEPT1"	NP_059489.1	septin 1	59	1	3	2.49E-04 ± 1.60E-04	0.025
SRRM2	NP_780438.1	serine/arginine repetitive matrix 2	308	6	3	7.20E-05 ± 3.83E-05	0.007
SETD3	NP_082538.1	SET domain containing 3	58	1	2	2.79E-05 ± -	0.003
SAP18	NP_033145.1	Sin3-associated polypeptide 18	54	1	10	7.57E-04 ± -	0.076
SNRPE	NP_033253.1	small nuclear ribonucleoprotein E	55	1	13	1.53E-03 ± 3.82E-04	0.153
SNRPA	NP_056597.2	small nuclear ribonucleoprotein polypeptide A	98	2	10	1.73E-04 ± -	0.017
SPIN	NP_666155.1	spindlin isoform 2	54	1	6	1.26E-04 ± -	0.013
SF3A1	NP_080451.3	splicing factor 3a, subunit 1	167	5	9	1.67E-04 ± -	0.017
SFRS7	NP_666195.1	splicing factor, arginine/serine-rich 7	116	2	9	1.92E-03 ± 2.24E-04	0.193
STMN1	NP_062615.1	stathmin 1	147	3	15	7.78E-04 ± -	0.078
SDHB	NP_075863.2	succinate dehydrogenase Ip subunit	80	1	5	5.87E-05 ± -	0.006
SUPT16H	NP_291096.1	suppressor of Ty 16 homolog	53	1	2	1.58E-05 ± -	0.002
TAF10	NP_064408.2	TAF10 RNA polymerase II, TATA box binding protein (TBP)-associated factor	104	1	11	1.52E-04 ± 0.00E+00	0.015
TXNL2	NP_075629.2	thioredoxin-like 2	157	3	13	2.62E-04 ± 2.84E-05	0.026
TMPO	NP_035735.1	thymopoietin	174	3	7	9.55E-05 ± -	0.010
TIPRL	NP_663488.1	TIP41, TOR signalling pathway regulator-like	61	1	5	3.05E-04 ± -	0.031
TPX2	NP_082385.2	TPX2, microtubule-associated protein homolog	180	2	4	5.55E-05 ± 1.57E-05	0.006
TCEA3	NP_035672.1	transcription elongation factor A (SII), 3	147	3	9	4.29E-04 ± -	0.043
AHCTF1	NP_080651.1	transcription factor ELYS	214	6	4	6.15E-05 ± 2.25E-05	0.006

Proteins in found in embryonic stem cells only, continued.

Gene	NCBI Accession	Protein Description	Embryonic stem cells: Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Embryonic stem cells: AVG N_SC ± STDEV	Percent Relative Abundance in Identified Proteome
GATAD2B	NP_647465.1	transcription repressor p66 component of the MeCP1 complex	82	1	3	5.57E-05 ± -	0.006
TIMM8A	NP_038926.1	translocase of inner mitochondrial membrane 8 homolog a	64	1	11	1.71E-04 ± -	0.017
TMOD3	NP_058659.1	tropomodulin 3	79	1	5	7.05E-05 ± 3.32E-05	0.007
WARS	NP_035840.2	tryptophanyl-tRNA synthetase	70	1	3	6.88E-05 ± -	0.007
TUBB4	NP_033477.2	tubulin, beta 4	1263	22	68	5.09E-03 ± 5.14E-04	0.511
TBCD	NP_084154.1	tubulin-specific chaperone d	114	1	2	2.77E-05 ± 1.96E-05	0.003
TPD52	NP_001020433.1	tumor protein D52 isoform 2	99	1	8	7.39E-05 ± -	0.007
SNRPC	NP_035562.1	U1 small nuclear ribonucleoprotein 1C	55	1	8	3.12E-04 ± -	0.031
SNRP70	NP_033250.3	U1 small nuclear ribonucleoprotein 70 kDa	74	2	4	3.69E-04 ± -	0.037
SNRPA1	NP_067311.3	U2 small nuclear ribonucleoprotein polypeptide A'	92	2	10	2.60E-04 ± -	0.026
IMP4	NP_848716.1	U3 snoRNP protein 4 homolog	58	1	10	5.69E-05 ± -	0.006
UQCRC2	NP_080175.1	ubiquinol cytochrome c reductase core protein 2	68	1	4	3.65E-05 ± -	0.004
USP14	NP_001033678.1	ubiquitin specific protease 14 isoform 2	93	2	7	3.61E-05 ± -	0.004
USP8	NP_062703.2	ubiquitin specific protease 8	67	1	2	1.53E-05 ± -	0.002
UBE2V2	NP_076074.2	ubiquitin-conjugating enzyme E2 variant 2	131	3	21	2.28E-03 ± -	0.229
UBE2D2	NP_064296.1	ubiquitin-conjugating enzyme E2D 2	83	1	16	8.44E-04 ± 1.03E-03	0.085
UBE2N	NP_542127.1	ubiquitin-conjugating enzyme E2N	162	4	37	4.36E-04 ± -	0.044
UBTF	NP_035681.1	upstream binding transcription factor, RNA polymerase I	137	2	3	8.67E-05 ± -	0.009
UTP14A	NP_082552.1	UTP14, U3 small nucleolar ribonucleoprotein, homolog A	142	2	5	4.32E-05 ± 3.05E-05	0.004
VPS25	NP_081052.2	vacuolar protein sorting 25	73	1	9	9.40E-05 ± -	0.009
WDHD1	NP_766186.2	WD repeat and HMG-box DNA binding protein 1	93	1	2	1.53E-05 ± 0.00E+00	0.002
WDR3	NP_780761.1	WD repeat domain 3	135	2	3	5.27E-05 ± -	0.005
XAB1	NP_598517.1	XPA binding protein 1	52	1	4	2.67E-04 ± -	0.027
ZC3H8	NP_065619.2	zinc finger CCCH type containing 8	94	1	5	5.43E-05 ± -	0.005
ZNF289	NP_076343.1	zinc finger protein 289	69	1	3	9.55E-05 ± -	0.010
ZNF593	NP_077177.1	zinc finger protein 593	62	1	16	2.85E-04 ± 1.43E-04	0.029
ZNF706	NP_080797.1	zinc finger protein 706	63	1	16	1.74E-03 ± -	0.175
IPO9	NP_722469.1	importin 9	91	1	2	1.59E-05 ± -	0.002
NUP214	NP_758472.2	nucleoporin 214	66	1	1	2.38E-05 ± -	0.002
OXSR1	NP_598746.1	oxidative-stress responsive 1	76	1	3	7.03E-05 ± -	0.007
PMPCB	NP_082707.1	peptidase (mitochondrial processing) beta	95	2	6	1.02E-04 ± -	0.010
RBM25	NP_081625.2	RNA binding motif protein 25	119	2	6	6.96E-05 ± 1.51E-05	0.007

Table 6 3. Proteins identified only in the early primitive ectoderm-like cell proteome. Relative percent abundance is calculated by dividing the spectral counts for an individual protein by the total spectral counts from all proteins in the early primitive ectoderm-like cell proteome (unique and shared). Mascot score, number of peptides, and coverage are totals from nonredundant peptides seen over all three replicate analyses.

			Early Primitive Ectoderm-like				Percent Relative
Gene	NCBI Accession	Protein Description	Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Early primitive ectoderm-like cells AVG N_SC ± STDEV	Abundance in Identified Proteome
AKAP2	NP_033779.2	A kinase (PRKA) anchor protein 2 isoform 3	54	1	2	1.22E-04 ± -	0.012
ATAD3A	NP_849534.1	AAA-ATPase TOB3	72	1	4	5.47E-05 ± -	0.006
ACAT2	NP_033364.1	acetyl-Coenzyme A acetyltransferase 2	52	1	3	1.63E-04 ± -	0.017
ACBD5	NP_083069.1	acyl-Coenzyme A binding domain containing 5	76	1	4	6.83E-05 ± -	0.007
ALDH6A1	NP_598803.1	aldehyde dehydrogenase family 6, subfamily A1	155	2	6	8.45E-04 ± -	0.086
OS-9	NP_808282.1	amplified in osteosarcoma	54	1	3	5.24E-05 ± -	0.005
ANXA2	NP_031611.1	annexin A2	271	5	17	2.57E-03 ± 1.49E-03	0.261
ANXA5	NP_033803.1	annexin A5	139	3	8	9.11E-04 ± -	0.092
API5	NP_031492.1	apoptosis inhibitor 5	96	1	4	6.41E-05 ± 0.00E+00	0.007
RARSL	NP_852071.1	arginyl-tRNA synthetase-like	88	1	3	1.40E-04 ± 1.19E-04	0.014
ATBF1	NP_031522.1	AT motif binding factor 1	107	4	1	3.47E-05 ± -	0.004
CARHSP1	NP_080097.1	calcium-regulated heat-stable protein (24kD)	78	1	11	2.18E-04 ± -	0.022
CALU	NP_031620.1	calumenin isoform 1	166	2	4	1.64E-03 ± 1.89E-03	0.166
CAPZA2	NP_031630.1	capping protein (actin filament) muscle Z-line, alpha 2	72	1	3	2.26E-04 ± -	0.023
KIT	NP_066922.1	c-kit	89	1	2	6.63E-05 ± -	0.007
CNDP2	NP_075638.2	CNDP dipeptidase 2 (metallopeptidase M20 family)	416	8	28	2.95E-04 ± 1.57E-04	0.030
COTL1	NP_082347.1	coactosin-like 1	180	3	21	4.55E-04 ± -	0.046
CFL2	NP_031714.1	cofilin 2, muscle	104	1	12	2.34E-03 ± -	0.237
COX5A	NP_031773.1	cytochrome c oxidase, subunit Va	197	5	47	1.78E-03 ± -	0.181
COX6B1	NP_079904.1	cytochrome c oxidase, subunit VIb polypeptide 1	78	1	16	1.13E-03 ± 1.06E-03	0.114
N-PAC	NP_082996.1	cytokine-like nuclear factor n-pac	92	2	5	5.92E-05 ± -	0.006
DDX50	NP_444413.1	DEAD (Asp-Glu-Ala-Asp) box polypeptide 50	66	1	3	8.80E-05 ± -	0.009
DBT	NP_034152.1	dihydrolipoamide branched chain transacylase E2	72	1	2	6.70E-05 ± -	0.007
DAB2	NP_075607.2	disabled homolog 2 isoform a	79	2	5	5.06E-04 ± -	0.051
DSCR3	NP_031860.1	Down syndrome critical region protein A	57	1	6	2.18E-04 ± -	0.022
TXNDC12	NP_079610.1	endoplasmic reticulum protein ERp19	52	1	8	1.90E-04 ± -	0.019
EIF4A2	NP_038534.1	eukaryotic translation initiation factor 4A2	85	2	6	8.33E-04 ± 2.81E-04	0.085
FAU	NP_032016.1	Finkel-Biskis-Reilly murine sarcoma virus (FBR-MuSV) ubiquitously expressed (fox derived)	86	2	8	9.71E-04 ± -	0.099
FKBP11	NP_077131.2	FK506 binding protein 11	64	1	9	2.68E-04 ± 9.28E-05	0.027
GTF2E2	NP_080860.2	general transcription factor II E, polypeptide 2 (beta subunit)	82	1	7	2.21E-04 ± 1.56E-04	0.022
GOT2	NP_034455.1	glutamate oxaloacetate transaminase 2, mitochondrial	344	6	23	5.76E-04 ± 3.04E-04	0.058
EPRS	NP_084011.1	glutamyl-prolyl-tRNA synthetase	324	7	7	1.85E-04 ± 2.11E-04	0.019
GLT25D1	NP_666323.1	glycosyltransferase 25 domain containing 1	114	2	6	7.85E-05 ± 3.70E-05	0.008
PTD004	NP_080218.1	GTP-binding protein PTD004 isoform a	80	1	4	1.63E-04 ± 8.16E-05	0.017
HIST1H1D	NP_663759.2	histone 1, H1d	405	7	24	7.75E-03 ± 3.46E-03	0.786
HIST1H2AE	NP_835492.1	histone 1, H2ao	618	10	41	4.70E-02 ± 3.48E-03	4.772
HIST1H3G	NP_659539.1	histone 1, H3g	422	10	60	2.80E-02 ± 1.83E-02	2.843
HADHA	NP_849209.1	hydroxyacyl-Coenzyme A dehydrogenase alpha subunit	198	4	6	2.26E-04 ± 6.47E-05	0.023
HMBS	NP_038579.1	hydroxymethylbilane synthase	90	1	6	2.68E-04 ± -	0.027
-	NP_001013046.1	hypothetical protein LOC103266	96	1	23	7.02E-04 ± 1.99E-04	0.071
-	NP_001034289.1	hypothetical protein LOC230866 isoform 2	67	1	2	6.50E-05 ± -	0.007
-	NP_001032835.1	hypothetical protein LOC328099	58	1	7	2.57E-04 ± -	0.026
-	NP_080618.1	hypothetical protein LOC67726	108	1	5	9.89E-05 ± 4.66E-05	0.010
-	NP_653108.2	hypothetical protein LOC68796	67	1	2	4.70E-05 ± -	0.005
-	NP_081997.1	hypothetical protein LOC71206	62	2	4	3.95E-04 ± 4.47E-04	0.040
-	NP_082279.1	hypothetical protein LOC71919	81	1	3	7.34E-05 ± 3.46E-05	0.007
IMPDH1	NP_035959.2	inosine 5'-phosphate dehydrogenase 2	132	3	6	2.93E-04 ± 3.46E-04	0.030
KRT19	NP_032497.1	keratin complex 1, acidic, gene 19	152	2	7	2.32E-03 ± 1.13E-04	0.236
HADHSC	NP_032238.1	L-3-hydroxyacyl-Coenzyme A dehydrogenase, short chain	57	1	4	4.26E-04 ± -	0.043
LMNA	NP_001002011.1	lamin A isoform A	88	2	6	2.91E-04 ± -	0.030
LMAN1	NP_081676.1	lectin, mannose-binding, 1	84	2	4	6.25E-05 ± -	0.006
LUC7L	NP_082466.1	Luc7 homolog (S. cerevisiae)-like isoform 2	69	1	4	2.18E-04 ± 1.85E-04	0.022
MVP	NP_542369.1	major vault protein	51	1	4	3.75E-05 ± -	0.004
MTA2	NP_035972.3	metastasis-associated protein 2	214	3	6	1.45E-04 ± 1.28E-04	0.015
MAT2B	NP_598778.1	methionine adenosyltransferase II, beta	83	2	8	2.90E-04 ± -	0.029
MBD3	NP_038623.1	methyl-CpG binding domain protein 3	123	2	8	2.27E-04 ± -	0.023
MRPS31	NP_065585.1	mitochondrial ribosomal protein S31	59	1	7	8.41E-05 ± -	0.009
MRCL3	NP_080340.2	myosin light chain, regulatory B-like	97	1	12	1.13E-03 ± 8.19E-04	0.114
MRLC2	NP_075891.1	myosin regulatory light chain-like	57	1	12	1.56E-03 ± 1.25E-03	0.159
MTPN	NP_032124.1	myotrophin	147	3	42	6.84E-04 ± 1.94E-04	0.069
NDUFA8	NP_080979.1	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 8	58	1	15	1.88E-04 ± -	0.019
NCKAP1	NP_058661.1	NCK-associated protein 1	77	1	2	2.86E-05 ± -	0.003
SSFA1	NP_035612.2	NHP2 non-histone chromosome protein 2-like 1	95	1	14	2.52E-04 ± -	0.026
NRK	NP_038752.1	Nik related kinase	56	1	1	2.22E-05 ± -	0.002
SYNCRIP	NP_062640.1	NS1-associated protein 1 isoform 1	287	5	12	3.10E-04 ± 2.25E-04	0.031
NRBP1	NP_671734.1	nuclear receptor binding protein	64	1	3	3.62E-04 ± -	0.037

Proteins in found in early primitive ectoderm-like cells only, continued.			Early Primitive Ectoderm-like				Percent Relative
Gene	NCBI Accession	Protein Description	Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Early primitive ectoderm-like cells AVG N_SC ± STDEV	Abundance in Identified Proteome
PAK3	NP_032804.1	p21 (CDKN1A)-activated kinase 3	54	1	3	8.91E-05 ± 4.20E-05	0.009
PGM2	NP_079976.1	phosphoglucomutase 1	75	1	3	5.21E-05 ± -	0.005
PLEC1	NP_958787.1	plectin 1 isoform 2	106	2	1	3.56E-05 ± 1.01E-05	0.004
ARFIP1	XP_130985.3	PREDICTED: ADP-ribosylation factor interacting protein 1 isoform 1	76	1	4	5.20E-04 ± 2.45E-04	0.053
-	NP_080679.1	PREDICTED: hypothetical protein LOC67842	74	1	3	5.08E-05 ± -	0.005
-	XP_996816.1	PREDICTED: similar to 40S ribosomal protein S19	222	5	34	1.04E-03 ± 7.16E-04	0.105
-	XP_988349.1	PREDICTED: similar to 40S ribosomal protein S2	88	2	7	6.17E-04 ± -	0.063
-	XP_921699.1	PREDICTED: similar to 40S ribosomal protein S4, X isoform isoform 1	225	5	14	2.09E-03 ± 1.93E-03	0.212
-	XP_999884.1	PREDICTED: similar to 60S ribosomal protein L13	97	2	8	6.24E-04 ± 2.21E-04	0.063
-	XP_001003798.1	PREDICTED: similar to 60S ribosomal protein L21	75	1	10	6.25E-04 ± -	0.063
-	XP_141727.3	PREDICTED: similar to 60S ribosomal protein L32	126	2	22	7.18E-04 ± -	0.073
-	XP_984865.1	PREDICTED: similar to 60S ribosomal protein L7a	292	7	26	1.12E-03 ± 2.56E-04	0.113
-	XP_903693.1	PREDICTED: similar to 60S ribosomal protein L7a (Surfeit locus protein 3) isoform 2	402	9	35	9.31E-04 ± 3.71E-04	0.094
-	XP_981516.1	PREDICTED: similar to AP-3 complex subunit sigma-1	69	1	9	1.67E-04 ± -	0.017
-	XP_987162.1	PREDICTED: similar to ATP synthase gamma chain, mitochondrial precursor	55	1	3	7.90E-05 ± -	0.008
-	NP_082320.1	PREDICTED: similar to calponin 3, acidic	93	1	5	4.89E-04 ± -	0.050
-	XP_985395.1	PREDICTED: similar to cell division cycle 42	91	1	17	8.08E-04 ± 7.12E-04	0.082
-	XP_996538.1	PREDICTED: similar to Cytoplasmic dynein light chain isoform 2	82	1	16	4.29E-04 ± 2.02E-04	0.043
-	XP_979560.1	PREDICTED: similar to DnaJ-like protein 2 isoform 2	65	2	6	1.65E-04 ± -	0.017
-	XP_989052.1	PREDICTED: similar to enolase 1, alpha non-neuron	245	5	20	1.23E-03 ± 4.89E-04	0.125
-	XP_979011.1	PREDICTED: similar to GLI pathogenesis-related 2	59	1	6	2.90E-04 ± -	0.029
-	XP_986710.1	PREDICTED: similar to H2A histone family, member O	639	10	40	3.86E-02 ± 7.61E-03	3.913
-	XP_921752.1	PREDICTED: similar to High mobility group protein 1	53	1	7	3.00E-04 ± -	0.030
-	XP_982238.1	PREDICTED: similar to IgE-binding protein	89	1	4	1.84E-04 ± 8.65E-05	0.019
-	XP_991268.1	PREDICTED: similar to Inner nuclear membrane protein Man1 (LEM domain containing protein 3)	109	1	3	3.51E-05 ± -	0.004
-	XP_001004685.1	PREDICTED: similar to NADP-dependent malic enzyme (NADP-ME) (Malic enzyme 1) isoform 4	133	2	10	1.76E-04 ± -	0.018
-	XP_990943.1	PREDICTED: similar to nascent polypeptide-associated complex alpha polypeptide	101	1	8	2.47E-04 ± 1.17E-04	0.025
-	XP_909486.2	PREDICTED: similar to Nidogen-1 precursor (Entactin)	317	5	25	6.09E-04 ± 6.93E-04	0.062
-	XP_982306.1	PREDICTED: similar to nuclease sensitive element binding protein 1	198	4	22	1.67E-03 ± 1.58E-03	0.170
-	XP_979585.1	PREDICTED: similar to Nucleophosmin	241	6	24	4.49E-03 ± 1.69E-03	0.456
-	XP_001002745.1	PREDICTED: similar to Proteasome subunit alpha type 5 isoform 1	180	2	13	1.88E-03 ± -	0.190
-	XP_622874.1	PREDICTED: similar to protein phosphatase-1 regulatory subunit 7 isoform 3	85	1	5	3.58E-04 ± -	0.036
-	XP_988366.1	PREDICTED: similar to ribosomal protein L31	62	1	8	3.49E-03 ± 5.48E-04	0.354
-	XP_895313.1	PREDICTED: similar to ribosomal protein L38	138	3	39	4.61E-03 ± -	0.468
-	XP_919999.1	PREDICTED: similar to ribosomal protein S23	78	1	8	9.04E-04 ± -	0.092
PLOD3	NP_036092.1	procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3	64	1	3	8.72E-05 ± -	0.009
P4HA2	NP_035161.1	procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), alpha II polypeptide	111	2	8	9.02E-05 ± 4.25E-05	0.009
PROSC	NP_473398.1	proline synthetase co-transcribed isoform a	55	1	6	1.18E-04 ± -	0.012
PTPN9	NP_062625.2	protein tyrosine phosphatase, non-receptor type 9	63	1	3	6.35E-04 ± 3.00E-04	0.064
PRUNE	NP_775482.1	PRUNEM1	83	1	4	1.42E-04 ± -	0.014
PURB	NP_035351.1	purine rich element binding protein B	140	1	13	9.97E-05 ± -	0.010
PKLR	NP_038659.1	pyruvate kinase liver and red blood cell	158	4	5	1.97E-04 ± 3.98E-05	0.020
RAB10	NP_057885.1	RAB10, member RAS oncogene family	73	1	6	4.85E-04 ± -	0.049
RABEP1	NP_062273.1	rabaptin, RAB GTPase binding effector protein 1	109	2	3	2.25E-04 ± -	0.023
RCN2	NP_036122.1	reticulocalbin 2	126	1	6	2.15E-03 ± 1.41E-03	0.218
RCN3	NP_080831.1	reticulocalbin 3	70	1	4	1.77E-03 ± -	0.180
SPARC	NP_033268.1	secreted acidic cysteine rich glycoprotein	151	2	10	2.67E-04 ± 7.56E-05	0.027
SPTLC1	NP_033295.2	serine palmitoyltransferase subunit 1	56	1	3	2.73E-04 ± -	0.028
SMC4L1	NP_598547.1	SMC4 structural maintenance of chromosomes 4-like 1	94	2	2	7.54E-05 ± -	0.008
EIF1	NP_035638.1	suppressor of initiator codon mutations, related sequence 1	112	1	16	1.24E-03 ± 1.19E-03	0.126
STXBP2	NP_035633.1	syntaxin binding protein 2	52	1	3	5.45E-05 ± -	0.006
TXNDC4	NP_083848.1	thioredoxin domain containing 4	158	2	10	1.19E-04 ± 5.63E-05	0.012
TXNDC5	NP_663342.3	thioredoxin domain containing 5	203	4	13	5.42E-04 ± 3.29E-04	0.055
TRIP10	NP_598886.1	thyroid hormone receptor interactor 10	66	1	4	1.18E-04 ± -	0.012
TBL2	NP_038791.2	transducin (beta)-like 2	81	1	3	7.31E-05 ± -	0.007
TRNT1	NP_081572.1	tRNA nucleotidyl transferase, CCA-adding, 1	64	1	4	2.98E-04 ± -	0.030
TRUB1	NP_082391.1	TruB pseudouridine (psi) synthase homolog 1 isoform 1	70	1	6	9.56E-05 ± -	0.010
TINAGL1	NP_075965.1	tubulointerstitial nephritis antigen-like	227	4	11	3.63E-04 ± 3.59E-04	0.037
NFRSF10E	NP_064671.2	tumor necrosis factor receptor superfamily, member 10b	94	1	4	2.54E-04 ± -	0.026
UBC	NP_062613.2	ubiquitin C	352	6	5	6.93E-04 ± 7.32E-04	0.070
UBE3C	NP_598668.1	ubiquitin protein ligase E3C	61	1	2	2.98E-05 ± -	0.003
UAP1L1	NP_001028465.1	UDP-N-acteylglucosamine pyrophosphorylase 1-like 1	81	1	4	1.27E-04 ± 0.00E+00	0.013
UTRN	NP_035812.2	utrophin	151	5	2	1.22E-04 ± 0.00E+00	0.012

Table 6 4. Proteins identified only in the embryoid body proteome. Relative percent abundance is calculated by dividing the spectral counts for an individual protein by the total spectral counts from all proteins in the embryoid body proteome (unique and shared). Mascot score, number of peptides, and coverage are totals from nonredundant peptides seen over all three replicate analyses.

			Embryoid Bodies				Percent Relative Abundance in Identified Proteome
Gene	NCBI Accession	Protein Description	Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Embryoid Bodies AVG N_SC ± STDEV	
NT5C	NP_056622.1	5',3'-nucleotidase, cytosolic	55	1	7	6.98E-05 ± -	0.007
ARBP	NP_031501.1	acidic ribosomal phosphoprotein P0	377	6	25	1.47E-03 ± 5.12E-03	0.148
ACADS	NP_031409.2	acyl-Coenzyme A dehydrogenase, short chain	80	1	3	3.39E-05 ± 0.00E+00	0.003
APRT	NP_033828.1	adenine phosphoribosyl transferase	54	1	9	1.55E-04 ± -	0.016
ARF3	NP_031504.1	ADP-ribosylation factor 3	162	4	18	3.86E-04 ± 1.54E-03	0.039
ARF5	NP_031506.1	ADP-ribosylation factor 5	125	3	14	1.55E-04 ± 7.75E-04	0.016
ARL2	NP_062696.2	ADP-ribosylation factor-like 2	57	1	12	7.59E-05 ± -	0.008
ARL6	NP_062639.2	ADP-ribosylation factor-like 6	80	1	10	2.25E-04 ± 1.06E-03	0.023
ASPSCR1	NP_081153.1	alveolar soft part sarcoma chromosome region, candidate 1 long isoform	89	1	3	5.08E-05 ± -	0.005
ALAD	NP_032551.3	aminolevulinate, delta-, dehydratase	97	1	6	2.33E-04 ± 1.50E-03	0.023
APPBP1	NP_659180.1	amyloid beta precursor protein binding protein 1	81	1	3	5.23E-05 ± 3.70E-04	0.005
APRIN	NP_780519.3	androgen-induced prostate proliferative shutoff associated protein AS3	182	3	3	4.83E-05 ± 1.67E-04	0.005
ASL	NP_598529.1	argininosuccinate lyase	155	2	7	7.02E-05 ± 1.73E-04	0.007
ACTR1A	NP_058556.1	ARP1 actin-related protein 1 homolog A	53	1	5	3.71E-05 ± -	0.004
ASPH	NP_075553.1	aspartyl beta-hydroxylase isoform 1	65	1	3	1.89E-05 ± -	0.002
ATP6V1F	NP_079657.1	ATPase, H+ transporting, V1 subunit F	66	1	10	3.52E-04 ± 1.66E-03	0.035
ABCD3	NP_033017.2	ATP-binding cassette, sub-family D, member 3	127	2	6	6.35E-05 ± -	0.006
BRRN1	NP_659067.1	barren homolog	81	1	2	1.93E-05 ± -	0.002
BAG3	NP_038891.3	Bcl2-associated athanogene 3	68	1	4	4.84E-05 ± 0.00E+00	0.005
BLMH	NP_848760.1	bleomycin hydrolase	267	6	17	4.60E-04 ± 2.96E-03	0.046
BCAT2	NP_033867.1	branched chain aminotransferase 2, mitochondrial	90	1	4	9.47E-05 ± 4.10E-04	0.010
ACTL6A	NP_062647.1	BRG1/brm-associated factor 53A	148	2	10	6.51E-04 ± -	0.065
BUB1B	NP_033903.1	budding uninhibited by benzimidazoles 1 homolog, beta	73	2	2	3.98E-05 ± -	0.004
CDH1	NP_033994.1	cadherin 1	138	2	5	6.84E-05 ± 3.28E-04	0.007
CPT2	NP_034079.1	carnitine palmitoyltransferase 2	117	1	4	1.06E-04 ± 3.67E-04	0.011
CSNK2A1	NP_031814.2	casein kinase II, alpha 1 polypeptide	327	5	16	2.62E-04 ± 1.35E-03	0.026
CEBPZ	NP_001019977.1	CCAAT/enhancer binding protein zeta	62	1	1	2.65E-05 ± -	0.003
CDC23	NP_848124.1	CDC23 (cell division cycle 23, yeast, homolog)	111	2	5	2.34E-05 ± -	0.002
CRABP1	NP_038524.1	cellular retinoic acid binding protein I	129	1	9	1.53E-04 ± 7.21E-04	0.015
CENPE	NP_776123.2	centromere protein E	75	2	1	1.13E-05 ± 7.99E-05	0.001
CLNS1A	NP_076160.1	chloride channel, nucleotide-sensitive, 1A	79	1	5	5.79E-05 ± -	0.006
CHAF1B	NP_082359.1	chromatin assembly factor 1 subunit B	62	1	4	2.44E-05 ± -	0.002
CNAP1	NP_666283.1	chromosome condensation-related SMC-associated protein 1	232	4	5	4.68E-05 ± 2.52E-04	0.005
CML66	NP_080425.2	chronic myelogenous leukemia tumor antigen 66	63	1	3	2.40E-05 ± -	0.002
CSTF3	NP_663504.1	cleavage stimulation factor subunit 3 isoform 1	62	1	2	1.95E-05 ± -	0.002
COPB	NP_203534.1	coatomer protein complex, subunit beta 1	59	1	1	2.93E-05 ± -	0.003
COPG2	NP_059506.1	coatomer protein complex, subunit gamma 2	54	1	2	3.21E-05 ± -	0.003
CSDA	NP_620817.2	cold shock domain protein A long isoform	199	2	10	3.39E-03 ± 2.44E-03	0.340
COPS4	NP_036131.1	COP9 signalosome subunit 4	61	1	3	2.75E-04 ± -	0.028
CORO1B	NP_035908.1	coronin, actin binding protein 1B	105	1	5	5.77E-05 ± 0.00E+00	0.006
CUL1	NP_036172.1	cullin 1	88	1	2	3.60E-05 ± -	0.004
CDK2	NP_904326.1	cyclin-dependent kinase 2 isoform 1	68	1	4	8.07E-05 ± -	0.008
COX4I1	NP_034071.1	cytochrome c oxidase subunit IV isoform 1	53	1	7	1.65E-04 ± -	0.017
CDC123	NP_598598.1	D123 gene product	86	1	5	4.15E-05 ± -	0.004
DDB1	NP_056550.1	damage specific DNA binding protein 1	104	1	2	4.90E-05 ± 0.00E+00	0.005
DAZAP1	NP_573451	DAZ associated protein 1	59	1	4	1.72E-04 ± -	0.017
DDT	NP_034157.1	D-dopachrome tautomerase	52	1	10	1.18E-04 ± -	0.012
NIPBL	NP_957684.1	delangin isoform B	59	1	1	7.78E-06 ± 3.66E-05	0.001
PPP2R5D	NP_033384.2	delta isoform of regulatory subunit B56, protein phosphatase 2A	100	2	5	7.05E-05 ± -	0.007
PCID1	NP_663355.1	dendritic cell protein GA17	382	6	23	1.87E-04 ± 1.35E-03	0.019
DSTN	NP_062745	destrin	60	1	5	1.97E-04 ± 9.75E-04	0.020
DRG2	NP_067329.1	developmentally regulated GTP binding protein 2	57	1	5	3.83E-05 ± -	0.004
CYB5R3	NP_084063.1	diaphorase 1	51	1	4	5.10E-04 ± -	0.051
DAB2	NP_001008702.1	disabled homolog 2 isoform b	111	2	5	4.25E-05 ± 1.47E-04	0.004
CRLZ1	NP_075541.1	disrupter of silencing SAS10	82	1	3	1.19E-04 ± -	0.012
DNMT3L	NP_062321.1	DNA (cytosine-5-)-methyltransferase 3-like	192	4	13	2.98E-04 ± -	0.030
PRIM1	NP_032947.1	DNA primase small subunit, 49kDa	77	1	4	1.12E-04 ± 3.88E-04	0.011
DNAJA2	NP_062768.1	DnaJ (Hsp40) homolog, subfamily A, member 2	175	2	14	1.58E-04 ± 1.28E-03	0.016
DNAJC3	NP_032955.2	DnaJ (Hsp40) homolog, subfamily C, member 3	71	1	6	5.54E-05 ± -	0.006
DPM1	NP_034202.1	dolichol-phosphate (beta-D) mannosyltransferase 1	78	1	5	5.37E-05 ± -	0.005
DSCR2	NP_062410.1	Down syndrome critical region protein 2	58	1	5	9.66E-05 ± -	0.010
DNM2	NP_001034609.1	dynamin 2	63	1	2	4.82E-05 ± -	0.005
EPHA2	NP_034269.2	Eph receptor A2	110	1	2	7.14E-05 ± 3.78E-04	0.007
EPS15L1	NP_031970.1	epidermal growth factor receptor pathway substrate 15, related	80	1	2	1.54E-05 ± -	0.002
EPB41L3	NP_038841.1	erythrocyte protein band 4.1-like 3	60	1	3	3.51E-05 ± 1.74E-04	0.004
EIF3S1	NP_653128.1	eukaryotic translation initiation factor 3, subunit 1 alpha	120	2	12	7.96E-03 ± 7.01E-02	0.799
EIF4G3	NP_766291.2	eukaryotic translation initiation factor 4 gamma, 3	67	2	2	2.21E-04 ± -	0.022
EIF4G2	NP_038535.2	eukaryotic translation initiation factor 4, gamma 2	128	4	3	9.24E-05 ± 7.06E-04	0.009
EIF4B	NP_663600.1	eukaryotic translation initiation factor 4B	110	1	5	1.41E-04 ± 8.01E-04	0.014
SITPEC	NP_036159.1	evolutionarily conserved signaling intermediate in Toll pathway	82	1	7	9.63E-05 ± -	0.010
XPO5	NP_082474.1	exportin 5	173	4	5	4.06E-05 ± 8.20E-05	0.004
XPO7	NP_075532.1	exportin 7	70	1	2	3.85E-05 ± -	0.004
FNTA	NP_032059.1	farnesyltransferase, CAAX box, alpha	95	1	4	3.70E-05 ± -	0.004
FHL1	NP_034341.1	four and a half LIM domains 1	81	1	5	2.49E-04 ± 2.11E-03	0.025
GABPA	NP_032091.2	GA repeat binding protein, alpha	54	1	3	9.07E-04 ± 6.53E-04	0.091
GCN1L1	NP_766307	GCN1 general control of amino-acid synthesis 1-like 1	300	7	4	5.40E-05 ± 2.11E-04	0.005
GEMIN4	NP_796341.1	gem (nuclear organelle) associated protein 4	52	1	1	5.28E-05 ± -	0.005
GEMIN5	NP_766146.1	gem (nuclear organelle) associated protein 5	55	1	1	4.18E-05 ± 3.29E-04	0.004
GTF2E1	NP_083088.1	general transcription factor II E, polypeptide 1 (alpha subunit)	69	1	3	3.17E-05 ± -	0.003
GTF2I	NP_034495.1	general transcription factor II I	104	3	7	2.85E-05 ± -	0.003
GLE1L	NP_083199.1	GLE1 RNA export mediator-like (yeast	96	1	3	5.99E-05 ± -	0.006

Proteins in found in embryoid bodies only, continued.

Gene	NCBI Accession	Protein Description	Embryoid Bodies: Total Mascot Score	Embryoid Bodies: Total No. Peptides	Embryoid Bodies: Total % Protein Coverage	Embryoid Bodies: AVG N_SC ± STDEV	Percent Relative Abundance in Identified Proteome
GNPDA1	NP_036067.1	glucosamine-6-phosphate deaminase 1	63	1	6	3.62E-04 ± 2.39E-03	0.036
GNPNAT1	NP_062298.1	glucosamine-phosphate N-acetyltransferase 1	52	1	8	1.52E-04 ± -	0.015
GCLC	NP_034425.1	glutamate-cysteine ligase, catalytic subunit	56	1	2	2.19E-05 ± -	0.002
GSR	NP_034474.3	glutathione reductase 1	109	1	5	2.79E-05 ± 0.00E+00	0.003
MGC68323	NP_955766.1	glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)-like	353	7	24	5.08E-03 ± 8.79E-03	0.510
GPD2	NP_034404.2	glycerol phosphate dehydrogenase 2, mitochondrial	65	1	2	1.92E-05 ± -	0.002
GLDC	NP_613061.1	glycine decarboxylase	83	1	2	1.36E-05 ± 0.00E+00	0.001
GSK3A	NP_001026837.1	glycogen synthase kinase 3 alpha	60	1	3	8.55E-05 ± -	0.009
GAS8	NP_061343.1	growth arrest specific 8	110	4	4	5.99E-04 ± -	0.060
GNL3	NP_849174.1	guanine nucleotide binding protein-like 3 (nucleolar) short isoform	148	3	11	6.91E-05 ± 5.87E-04	0.007
GNB2	NP_034442.1	guanine nucleotide-binding protein, beta-2 subunit	97	1	6	8.21E-05 ± -	0.008
HNRPR	NP_083147.1	heterogeneous nuclear ribonucleoprotein R	103	2	5	1.10E-04 ± -	0.011
PRMT1	NP_062804.1	heterogeneous nuclear ribonucleoproteins methyltransferase-like 2	281	5	18	3.26E-04 ± 3.06E-03	0.033
HNRPU	NP_058085.1	heterogenous nuclear ribonucleoprotein U	260	6	8	2.45E-04 ± 4.63E-04	0.025
HRB2	NP_848725.1	HIV-1 Rev binding protein 2	70	1	3	3.67E-05 ± -	0.004
HIP2	NP_058066.2	huntingtin interacting protein 2	79	2	11	2.79E-04 ± -	0.028
-	NP_663515.2	hypothetical protein LOC229543	214	3	6	4.02E-05 ± -	0.004
-	NP_001004147.1	hypothetical protein LOC237730	70	2	7	4.45E-05 ± -	0.004
-	NP_766625.1	hypothetical protein LOC272538	53	1	2	6.47E-05 ± -	0.006
-	NP_001025047.1	hypothetical protein LOC382038	102	2	2	1.83E-05 ± -	0.002
-	NP_001008551.1	hypothetical protein LOC434632	397	6	19	5.80E-04 ± 1.28E-03	0.058
-	NP_080894.1	hypothetical protein LOC52469	57	1	9	2.58E-04 ± -	0.026
-	NP_080470.1	hypothetical protein LOC67490	69	2	3	5.28E-05 ± -	0.005
-	NP_080545.2	hypothetical protein LOC67604	84	1	5	4.27E-05 ± -	0.004
-	NP_081024.1	hypothetical protein LOC68510	71	1	1	6.57E-06 ± -	0.001
-	NP_081029.1	hypothetical protein LOC68523	60	1	13	1.71E-04 ± 1.21E-03	0.017
-	NP_997594.1	hypothetical protein LOC68734	67	1	2	1.70E-05 ± -	0.002
-	NP_081165.2	hypothetical protein LOC68964 isoform a	65	1	2	2.30E-05 ± -	0.002
-	NP_077222.3	hypothetical protein LOC73288	65	1	1	2.90E-05 ± -	0.003
IPO11	NP_083941.2	importin 11	57	1	2	2.86E-05 ± -	0.003
IMPACT	NP_032404.1	imprinted and ancient	200	2	9	4.39E-04 ± 4.35E-03	0.044
IMMT	NP_083949.2	inner membrane protein, mitochondrial	157	2	4	2.77E-05 ± 1.30E-04	0.003
IDE	NP_112419	insulin degrading enzyme	220	4	6	9.09E-05 ± 2.19E-04	0.009
INTS7	NP_848747.3	integrator complex subunit 7	74	1	2	7.22E-05 ± -	0.007
IFI30	NP_075552.1	interferon gamma inducible protein 30	54	1	10	5.63E-05 ± -	0.006
IDH3A	NP_083849.1	isocitrate dehydrogenase 3 (NAD+) alpha	136	2	10	5.72E-05 ± 2.69E-04	0.006
IARS2	NP_941055.1	isoleucine-tRNA synthetase 2, mitochondrial	61	1	2	1.38E-05 ± -	0.001
KPNA1	NP_032491.2	karyopherin (importin) alpha 1	65	1	3	2.59E-05 ± 0.00E+00	0.003
LAS1L	NP_690035.2	LAS1-like isoform 1	80	1	2	1.84E-05 ± -	0.002
LGALS2	NP_079898.1	lectin, galactose-binding, soluble 2	93	1	12	6.44E-04 ± 4.55E-03	0.065
LEFTY2	NP_796073.1	Left-right determination factor 2	74	1	5	5.06E-05 ± 2.19E-04	0.005
LGMN	NP_035305.1	legumain	61	1	4	8.02E-05 ± 6.81E-04	0.008
LUC7L	NP_080157.1	Luc7 homolog (S. cerevisiae)-like isoform 1	101	1	5	7.87E-04 ± 4.30E-03	0.079
MAGO	NP_079840.1	mago-nashi homolog	192	3	38	2.80E-03 ± 1.64E-02	0.281
MLSTD2	NP_080419.2	male sterility domain containing 2	102	2	5	3.79E-04 ± -	0.038
MNAT1	NP_032638.1	menage a trois 1	87	1	4	1.51E-04 ± 9.42E-04	0.015
MTX2	NP_058084.2	metaxin 2	112	1	7	7.96E-05 ± 3.75E-04	0.008
MAT2A	NP_663544.1	methionine adenosyltransferase II, alpha	204	4	16	3.53E-05 ± -	0.004
RHOT1	NP_067511.4	mitochondrial Rho 1	56	1	3	4.42E-05 ± -	0.004
MRLP55	NP_080311.1	mitochondrial ribosomal protein L55	86	1	11	1.83E-04 ± 1.27E-03	0.018
MSRP9	NP_076003.2	mitochondrial ribosomal protein S9	61	1	4	7.16E-05 ± -	0.007
MAP3K4	NP_036078.1	mitogen activated protein kinase kinase kinase 4	75	1	1	1.36E-05 ± 6.41E-05	0.001
MKI67IP	NP_080748.2	Mki67 (FHA domain) interacting nucleolar phosphoprotein	51	1	9	8.81E-05 ± -	0.009
MMS19L (includes	NP_082428.1	MMS19 (MET18 S. cerevisiae)-like	56	1	2	1.35E-05 ± -	0.001
MOV10	NP_032645.1	Moloney leukemia virus 10	79	1	1	1.39E-05 ± -	0.001
MYO6	NP_001034635.1	myosin VI	64	1	1	1.11E-05 ± -	0.001
MYL6	NP_034990.1	myosin, light polypeptide 6, alkali, smooth muscle and non-muscle	108	2	16	1.85E-04 ± -	0.019
ATP1B3	NP_031528.1	Na+/K+ -ATPase beta 3 subunit	58	1	6	1.00E-04 ± -	0.010
ARD1A	NP_063923.1	N-acetyltransferase ARD1	69	1	6	5.94E-05 ± -	0.006
NYREN18	NP_058016.1	Nedd8 ultimate buster-1	53	1	2	2.06E-05 ± -	0.002
NMT2	NP_032734.1	N-myristoyltransferase 2	152	3	7	2.64E-05 ± 0.00E+00	0.003
NCBP1	NP_001028373.1	nuclear cap binding protein subunit 1, 80kDa	141	2	3	1.47E-04 ± 5.09E-04	0.015
NFX1	NP_076228.1	nuclear transcription factor, X-box binding 1	55	1	2	1.25E-05 ± -	0.001
NUCKS1	NP_780503.1	nuclear, casein kinase and cyclin-dependent kinase substrate	110	1	8	1.80E-04 ± 1.70E-03	0.018
NOL6	NP_631982.1	nucleolar RNA-associated protein long isoform	65	1	2	1.22E-05 ± -	0.001
NUP160	NP_067487.1	nucleoporin 160	57	1	1	9.96E-06 ± -	0.001
NUDT21	NP_080899.1	nudix (nucleoside diphosphate linked moiety X)-type motif 21	84	1	7	4.00E-04 ± 2.18E-03	0.040
OGT	NP_631883.2	O-linked N-acetylglucosamine transferase	145	2	3	9.34E-05 ± 7.55E-04	0.009
OSBPL9	NP_775485.1	oxysterol binding protein-like 9 isoform b	71	1	2	2.44E-05 ± -	0.002
GATAD2A	NP_663571.2	p66 alpha homolog isoform 1	75	1	2	2.22E-05 ± -	0.002
-	NP_081569.1	pad-1-like isoform 1	122	3	2	3.04E-05 ± -	0.003
KIAA1274	NP_038781.1	paladin	64	1	2	1.62E-05 ± -	0.002
PCM1	NP_076151.1	pericentriolar material 1	58	1	1	1.38E-05 ± -	0.001
PES1	NP_075027.1	pescadillo homolog 1, containing BRCT domain	162	2	10	4.78E-05 ± -	0.005
PDLC3	NP_081126.2	phosducin-like 3	58	1	5	3.78E-04 ± 2.06E-03	0.038
PIP5K2C	NP_473438.1	phosphatidylinositol-4-phosphate 5-kinase, type II, gamma	51	1	3	3.32E-05 ± -	0.003
PSPH	NP_598661.1	phosphoserine phosphatase	55	1	5	1.24E-04 ± -	0.012
JARID1B	NP_690855.1	PLU1	69	1	1	9.04E-06 ± -	0.001
POLD1	NP_035261.2	polymerase (DNA directed), delta 1, catalytic subunit	56	1	1	1.26E-05 ± -	0.001
POLD2	NP_032920.1	polymerase (DNA directed), delta 2, regulatory subunit	86	1	3	2.98E-05 ± -	0.003
PNKP	NP_067524.1	polynucleotide kinase 3'- phosphatase	58	1	3	2.83E-05 ± -	0.003
POU2F1	NP_945151.1	POU domain, class 2, transcription factor 1 isoform L	62	1	3	1.84E-05 ± -	0.002
CUBN	XP_130038.5	PREDICTED: cubilin (intrinsic factor-cobalamin receptor) isoform 1	89	2	1	7.71E-06 ± -	0.001
-	XP_125867.4	PREDICTED: down-regulated in metastasis isoform 1	78	1	1	3.00E-05 ± -	0.003
EDD1	XP_196130.5	PREDICTED: extraembryonic development protein isoform 1	157	4	1	6.50E-05 ± -	0.007
MLLT4	XP_895540.2	PREDICTED: myeloid/lymphoid or mixed lineage-leukemia translocation to 4 homolog isoform 1	83	1	1	1.15E-05 ± 5.42E-05	0.001
POLR2I	XP_133304.3	PREDICTED: polymerase (RNA) II (DNA directed) polypeptide I isoform 1	97	1	18	1.12E-04 ± -	0.011
PGRMC2	XP_130859.7	PREDICTED: progesterone membrane binding protein isoform 1	109	1	9	6.43E-05 ± 0.00E+00	0.006
-	XP_916456.1	PREDICTED: similar to 10 kDa heat shock protein, mitochondrial (Hsp10) (10 kDa chaperonin) (CPN10)	118	1	14	7.12E-03 ± 1.19E-02	0.715
-	XP_984363.1	PREDICTED: similar to 3-phosphoglycerate dehydrogenase	133	1	9	7.66E-04 ± 9.69E-03	0.077
-	XP_620258.1	PREDICTED: similar to 40S ribosomal protein S2 isoform 1	252	6	22	4.13E-04 ± 1.46E-03	0.041

Proteins in found in embryoid bodies only, continued.

Gene	NCBI Accession	Protein Description	Embryoid Bodies: Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Embryoid Bodies: AVG N_SC ± STDEV	Percent Relative Abundance in Identified Proteome
-	XP_125109.1	PREDICTED: similar to 40S ribosomal protein S6	110	2	6	3.92E-04 ± 1.58E-03	0.039
-	XP_484667.1	PREDICTED: similar to 40S ribosomal protein SA (p40) (34/67 kDa laminin receptor)	572	8	45	3.27E-03 ± 5.00E-03	0.328
-	XP_001000750.1	PREDICTED: similar to 60S acidic ribosomal protein P0 (L10E)	270	5	17	1.13E-03 ± 4.37E-03	0.113
-	XP_286185.2	PREDICTED: similar to 60S ribosomal protein L11	133	2	18	9.16E-04 ± 3.98E-03	0.092
-	XP_001002512.1	PREDICTED: similar to 60S ribosomal protein L18	216	3	11	9.26E-03 ± 8.55E-02	0.929
-	XP_981783.1	PREDICTED: similar to 60S ribosomal protein L5	302	5	21	2.73E-03 ± 3.69E-03	0.274
-	XP_001002625.1	PREDICTED: similar to Adenosylhomocysteinase	210	3	19	3.08E-04 ± 1.66E-03	0.031
-	XP_996891.1	PREDICTED: similar to Casein kinase II subunit alpha (CK II)	81	2	5	8.43E-05 ± 5.96E-04	0.008
CNOT1	XP_919363.2	PREDICTED: similar to CCR4-NOT transcription complex, subunit 1 isoform a isoform 16	228	5	3	8.23E-05 ± 8.31E-05	0.008
-	XP_917492.1	PREDICTED: similar to CG7593-PA isoform 9	53	1	5	1.24E-03 ± -	0.124
-	XP_980525.1	PREDICTED: similar to Clusterin precursor (Sulfated glycoprotein 2)	64	1	4	3.12E-05 ± 0.00E+00	0.003
CAND1	XP_909201.2	PREDICTED: similar to Cullin-associated NEDD8-dissociated protein 1 isoform 4	429	9	8	1.18E-04 ± 1.02E-03	0.012
POLR2E	XP_999085.1	PREDICTED: similar to DNA-directed RNA polymerase II 23 kDa polypeptide isoform 1	93	1	8	9.97E-05 ± 4.70E-04	0.010
-	XP_995436.1	PREDICTED: similar to ERO1-like	54	1	3	2.41E-04 ± -	0.024
-	XP_921237.1	PREDICTED: similar to eukaryotic translation initiation factor 4A, isoform 1 isoform 7	344	5	13	1.29E-03 ± 8.29E-03	0.129
EXOSC7	XP_922330.1	PREDICTED: similar to Exosome complex exonuclease RRP42 isoform 4	108	2	10	2.16E-04 ± 1.02E-03	0.022
-	XP_001006050.1	PREDICTED: similar to Ferritin light chain 1 (Ferritin L subunit 1)	162	2	17	4.32E-04 ± 8.81E-04	0.043
-	XP_979290.1	PREDICTED: similar to glyceraldehyde-3-phosphate dehydrogenase	111	2	12	3.03E-03 ± 1.10E-02	0.304
-	XP_979208.1	PREDICTED: similar to heterogeneous nuclear ribonucleoprotein A3 isoform 4	715	12	33	2.40E-03 ± 5.50E-03	0.241
-	XP_619939.3	PREDICTED: similar to High mobility group protein 1 (Heparin-binding protein p30) isoform 1	59	1	9	2.48E-04 ± -	0.025
-	XP_001003011.1	PREDICTED: similar to High mobility group protein 2 (HMG-2)	74	2	20	1.93E-04 ± -	0.019
IPO8	XP_914156.1	PREDICTED: similar to Importin-8 (Imp8) (Ran-binding protein 8) (RanBP8) isoform 8	198	2	3	6.91E-05 ± 2.39E-04	0.007
-	XP_998315.1	PREDICTED: similar to keratin Kb40	88	2	5	3.60E-04 ± -	0.036
-	XP_994889.1	PREDICTED: similar to Laminin alpha-5 chain precursor	165	2	1	2.31E-05 ± -	0.002
-	XP_988418.1	PREDICTED: similar to melanoma antigen family D, 2	110	1	10	8.43E-05 ± -	0.008
-	XP_988053.1	PREDICTED: similar to Midasin (MIDAS-containing protein)	471	10	2	2.00E-05 ± 1.00E-04	0.002
-	XP_989537.1	PREDICTED: similar to nucleosome assembly protein 1-like 1	55	1	17	1.16E-03 ± -	0.116
-	XP_988293.1	PREDICTED: similar to oxoglutarate dehydrogenase-like	168	3	4	9.04E-05 ± 6.26E-04	0.009
-	XP_924476.1	PREDICTED: similar to polyglutamine binding protein 1 isoform 3	63	1	9	4.75E-05 ± -	0.005
-	XP_911613.1	PREDICTED: similar to proteasome 26S non-ATPase subunit 2 isoform 3	246	7	8	3.30E-04 ± 4.88E-04	0.033
-	XP_922555.1	PREDICTED: similar to Protein C10orf86 isoform 5	54	1	6	3.66E-05 ± -	0.004
-	XP_990021.1	PREDICTED: similar to Protein phosphatase 2C isoform beta (PP2C-beta) isoform 2	85	2	5	2.05E-04 ± -	0.021
-	XP_912927.2	PREDICTED: similar to Ribonuclease HI large subunit (RNase HI large subunit) (Ribonuclease H2) (RNase H2	67	1	20	5.27E-04 ± -	0.053
-	XP_996449.1	PREDICTED: similar to ribosomal protein L21	87	1	9	1.74E-04 ± 1.23E-03	0.017
-	XP_991783.1	PREDICTED: similar to ribosomal protein L24	128	4	28	1.24E-03 ± -	0.124
-	XP_992604.1	PREDICTED: similar to ribosomal protein S10	159	3	24	2.34E-03 ± 1.54E-02	0.235
-	XP_978973.1	PREDICTED: similar to ribosomal protein S24	57	1	9	2.22E-03 ± 2.39E-02	0.223
-	XP_979136.1	PREDICTED: similar to ribosomal protein S8 isoform 1	344	6	35	1.61E-03 ± 1.77E-03	0.162
RBM25	XP_997451.1	PREDICTED: similar to RNA binding motif protein 25 isoform 6	86	2	4	6.51E-05 ± -	0.007
SEC31L1	XP_987087.1	PREDICTED: similar to SEC31-like 1	100	1	2	2.10E-05 ± -	0.002
SPTAN1	XP_992183.1	PREDICTED: similar to Spectrin alpha chain, brain isoform 11	591	10	6	8.43E-05 ± 5.65E-05	0.008
-	XP_001001258.1	PREDICTED: similar to splicing factor 3b, subunit 4 (predicted) isoform 2	76	1	3	4.94E-05 ± 2.33E-04	0.005
-	XP_990615.1	PREDICTED: similar to sterile alpha motif domain containing 1	64	1	5	5.38E-05 ± -	0.005
-	XP_990178.1	PREDICTED: similar to Succinyl-CoA ligase [GDP-forming] beta-chain, mitochondrial precursor	90	1	5	2.91E-05 ± 0.00E+00	0.003
-	XP_982539.1	PREDICTED: similar to threonyl-tRNA synthetase isoform 2	74	1	3	9.67E-05 ± 1.09E-03	0.010
-	XP_993084.1	PREDICTED: similar to threonyl-tRNA synthetase-like 1	63	1	1	8.23E-06 ± -	0.001
-	XP_990665.1	PREDICTED: similar to Transitional endoplasmic reticulum ATPase isoform 1	1069	16	32	1.95E-03 ± 1.07E-02	0.196
-	XP_998765.1	PREDICTED: similar to tubulin, alpha 2 isoform 2 isoform 2	741	11	38	7.39E-03 ± 2.89E-02	0.742
-	XP_622904.2	PREDICTED: similar to Tyrosyl-tRNA synthetase, cytoplasmic (Tyrosyl--tRNA ligase) (TyrRS)	69	1	2	2.60E-05 ± 0.00E+00	0.003
-	XP_990345.1	PREDICTED: similar to vacuolar H+ ATPase G1	67	1	9	7.10E-04 ± 1.68E-03	0.071
-	XP_001000597.1	PREDICTED: similar to Zinc finger A20 domain-containing protein 2 (Zinc finger protein 216) isoform 3	57	1	12	1.31E-04 ± -	0.013
SYNE2	XP_922176.2	PREDICTED: synaptic nuclear envelope 2	126	4	1	1.42E-05 ± -	0.001
-	XP_109868.6	PREDICTED: tensin 3	73	1	2	2.20E-05 ± 6.23E-05	0.002
-	XP_128374.5	PREDICTED: zinc finger protein 294 isoform 1	68	1	1	2.37E-05 ± 7.89E-05	0.002
	NP_001034166.1	proline synthetase co-transcribed isoform b	61	1	8	2.74E-04 ± -	0.028
PREP	NP_035286.1	prolyl endopeptidase	166	3	7	5.90E-05 ± 3.40E-04	0.006
PSMA4	NP_036096.1	proteasome (prosome, macropain) subunit, alpha type 4	54	1	8	1.07E-04 ± -	0.011
PSMD13	NP_036005.1	proteasome 26S non-ATPase subunit 13	175	2	12	6.19E-05 ± 4.29E-04	0.006
PSMD8	NP_080821.2	proteasome 26S non-ATPase subunit 8	204	5	16	6.04E-04 ± 2.39E-03	0.061
PPP1CB	NP_766295.1	protein phosphatase 1, catalytic subunit, beta isoform	241	5	20	3.13E-04 ± 2.61E-03	0.031
PPP2R4	NP_620087.2	protein phosphatase 2A, regulatory subunit B (PR 53)	51	1	3	1.30E-04 ± -	0.013
PDK1	NP_766253.1	pyruvate dehydrogenase kinase, isoenzyme 1	119	2	8	4.82E-05 ± 2.27E-04	0.005
RAB2	NP_067493.1	RAB2, member RAS oncogene family	57	1	7	6.58E-05 ± -	0.007
RAD21	NP_033035.2	RAD21 homolog	54	1	2	6.59E-05 ± -	0.007
RSU1	NP_033131.2	Ras suppressor protein 1	58	1	6	1.76E-04 ± 1.07E-03	0.018
	NP_033054.1	RAS-like, family 2, locus 9	288	5	25	2.86E-03 ± 2.43E-02	0.287
RFC4	NP_663455.1	replication factor C (activator 1) 4	95	2	9	3.83E-05 ± -	0.004
RPA2	NP_035414.2	replication protein A2	54	1	6	2.58E-04 ± -	0.026
RNPS1	NP_033096.1	ribonucleic acid binding protein S1	53	1	4	9.37E-04 ± 2.44E-03	0.094
RPN2	NP_062616.2	ribophorin II	208	3	11	1.55E-04 ± -	0.016
RPL22L1	NP_080793.1	ribosomal protein L22 like 1	84	1	20	4.58E-04 ± -	0.046
RPL9	NP_035422.1	ribosomal protein L9	379	7	58	2.28E-03 ± 1.09E-02	0.229
TRA2A	NP_932770.1	RIKEN cDNA G430041M01	51	1	5	4.97E-05 ± -	0.005
ROD1	NP_835458.1	ROD1 regulator of differentiation 1 isoform 2	112	2	5	2.80E-04 ± 1.70E-03	0.028
SEC22L1	NP_035472.1	SEC22 vesicle trafficking protein-like 1	60	1	7	1.30E-04 ± -	0.013
SEC23A	NP_033173.2	SEC23A	116	2	4	1.82E-05 ± 0.00E+00	0.002
EXOC4	NP_033174.2	SEC8	77	1	2	1.43E-05 ± -	0.001
SRPK1	NP_058075.2	serine/arginine-rich protein specific kinase 1	104	2	2	4.09E-04 ± -	0.041
STRAP	NP_035629.1	serine/threonine kinase receptor associated protein	182	3	11	6.36E-04 ± 3.98E-03	0.064
SLC25A13	NP_056644.1	solute carrier family 25 (mitochondrial carrier, adenine nucleotide translocator), member 13	239	5	11	3.82E-04 ± 1.45E-04	0.038
SLC25A12	NP_766024.1	solute carrier family 25 (mitochondrial carrier, Aralar), member 12	86	2	4	2.06E-04 ± -	0.021
SLC9A3R1	NP_036160.1	solute carrier family 9 (sodium/hydrogen exchanger), isoform 3 regulator 1	87	1	4	7.86E-05 ± 5.56E-04	0.008
SNX5	NP_077187.1	sorting nexin 5	81	2	7	1.16E-04 ± -	0.012
ATXN10	NP_058539.2	spinocerebellar ataxia 10 homolog	186	4	11	1.03E-04 ± 1.04E-03	0.010
SF3B2	NP_084385.1	splicing factor 3b, subunit 2	118	1	2	1.38E-04 ± 1.01E-03	0.014
SART1	NP_058578.2	squamous cell carcinoma antigen recognized by T-cells 1	58	1	2	3.46E-05 ± 2.45E-04	0.003
STAG2	NP_067440.2	stromal antigen 2	93	1	2	4.20E-05 ± 4.24E-04	0.004
SVH	NP_080310.1	SVH protein	94	1	8	4.56E-05 ± -	0.005
SDC4	NP_035651.1	syndecan 4	56	1	7	2.11E-04 ± 1.99E-03	0.021
RIC8	NP_444424.1	synembryn	64	1	4	2.63E-05 ± -	0.003
TBC1D15	NP_079982.2	TBC1 domain family, member 15	66	1	2	2.08E-05 ± -	0.002

Proteins in found in embryoid bodies only, continued.

Gene	NCBI Accession	Protein Description	Embryoid Bodies: Total Mascot Score	Embryoid Bodies: Total No. Peptides	Embryoid Bodies: Total % Protein Coverage	Embryoid Bodies: AVG N_SC ± STDEV	Percent Relative Abundance in Identified Proteome
TST	NP_033463.1	thiosulfate sulfurtransferase, mitochondrial	70	1	4	1.88E-04 ± -	0.019
TSNAX	NP_058605.1	translin-associated factor X	58	1	6	9.63E-05 ± -	0.010
TIMM50	NP_079892.1	translocase of inner mitochondrial membrane 50 homolog	187	2	9	2.77E-04 ± 1.81E-03	0.028
TMEM11	NP_775655.1	transmembrane protein 11	66	1	8	7.35E-05 ± -	0.007
TNRC5	NP_082341.1	trinucleotide repeat containing 5	97	1	7	7.59E-05 ± 3.57E-04	0.008
TPP2	NP_033444.1	tripeptidyl peptidase II	70	1	1	4.98E-05 ± 3.91E-04	0.005
TSFM	NP_079813.1	Ts translation elongation factor, mitochondrial	134	2	12	7.18E-05 ± 2.48E-04	0.007
TSC1	NP_075025.2	tuberous sclerosis 1	86	1	3	1.20E-05 ± -	0.001
TBCC	NP_848472.2	tubulin-specific chaperone c	71	1	7	4.09E-05 ± -	0.004
TPD52	NP_001020432.1	tumor protein D52 isoform 1	60	1	6	1.70E-04 ± -	0.017
KB37	NP_001003669.1	type II keratin Kb37	82	1	2	5.64E-05 ± -	0.006
PRPF6	NP_598462.1	U5 snRNP-associated 102 kDa protein	55	1	1	2.97E-05 ± 0.00E+00	0.003
UBR2	NP_666190.2	ubiquitin protein ligase E3 component n-recognin 2	105	1	1	1.59E-05 ± 1.12E-04	0.002
USP10	NP_033488.1	ubiquitin specific protease 10	160	3	5	7.92E-05 ± 1.24E-04	0.008
USP15	NP_081880.2	ubiquitin specific protease 15	71	1	2	4.27E-05 ± -	0.004
USP9X	NP_033507.1	ubiquitin specific protease 9, X chromosome	571	12	5	1.25E-04 ± 4.74E-04	0.013
UBAP2	NP_081148.1	ubiquitin-associated protein 2	67	1	1	6.58E-05 ± 7.11E-05	0.007
UBE2D3	NP_079632.1	ubiquitin-conjugating enzyme E2D 3 (UBC4/5 homolog, yeast)	72	1	16	2.18E-03 ± -	0.219
UGCGL1	NP_942602.1	UDP-glucose ceramide glucosyltransferase-like 1	61	1	1	9.00E-06 ± -	0.001
ATP6V1B2	NP_031535.2	vacuolar H+ATPase B2	151	2	6	9.56E-05 ± 1.93E-04	0.010
VDP	NP_062363.1	vesicle docking protein	132	4	4	2.91E-05 ± 0.00E+00	0.003
ZFR	NP_035897.1	zinc finger RNA binding protein	68	1	2	2.65E-05 ± 1.87E-04	0.003
ZMAT	NP_079870.1	zinc finger, matrin type 2	93	1	10	7.01E-05 ± -	0.007

Table 6.5. Proteins unique to embryonic stem cells, early primitive ectoderm-like cells or embryoid bodies that were seen in at least two replicate analyses of each proteome and have been previously investigated in terms of embryogenesis. Abbreviations: E = Embryonic day, ES Cell = Embryonic stem cell.

Gene	Refseq ID	Description	Essential Process during Embryogenesis
Embryonic stem cell unique population			
AKAP8	NP_062748.2	A kinase anchor protein 8	Palatogenesis (84)
ANP32A	NP_033802.2	Acidic (leucine-rich) nuclear phosphoprotein 32 family, member A	Embryonic stem cell characterization (85)
ATOX1	NP_033850.1	Antioxidant protein 1	Partial embryonic survival/ defined postnatal survival (86)
BTF3	NP_663430.2	Basic transcription factor 3	Postimplantation survival (40)
			Embryonic stem cell characterization (24,25)
CNBP1	NP_038521.1	Cellular nucleic acid binding protein 1	Cell proliferation/Face and limb morphogenesis (87)
COX17	NP_001017429.1	Cytochrome c oxidase subunit XVII assembly protein homolog	Gastrulation (88)
DPPA5	NP_079550.2	Developmental pluripotency associated 5	Pluripotency (89)
			Pluripotency (90)
ESRRB	NP_036064.2	Estrogen related receptor, beta	Placental development (91)
			Embryonic stem cell characterization (25)
HNRNP M	NP_084080.1	Heterogeneous nuclear ribonucleoprotein M	Regulation of splicing during differentiation (92)
KLHDC4	NP_663580.1	Kelch domain containing 4	Implicated peri-implantation (93)
HLP	NP_077185.1	LIM only protein HLP	Cardiogenesis (94)
PHF17	NP_758507.3	PHD zinc finger protein Jade1	Specification of embryonic tissue lineages (95)
AHCTF1	NP_080651.1	Transcription factor ELYS	Survival and development of the inner cell mass (96)
Early primitive ectoderm-like cell unique population			
ANXA2	NP_031611.1	Annexin A2	Neoangiogenesis (49)
COX6B1	NP_079904.1	Cytochrome c oxidase, subunit VIb polypeptide 1	Blastocyst survival (97)
EIF4A2	NP_038534.1	Eukaryotic translation initiation factor 4A2	Neural plate induction (47)

HADHA	NP_849209.1	Enoyl-Coenzyme A hydratase (trifunctional protein), alpha subunit	Intrauterine growth (98)
			Postnatal survival (98)
IMPDH1	NP_035959.2	Inosine 5'-phosphate dehydrogenase 2	Retinal viability (99)
LUC7L	NP_082466.1	Luc7 homolog (S. cerevisiae)-like isoform 2	Regulation of muscle differentiation (50)
MTPN	NP_032124.1	Myotrophin	Embryonic stem cell characterization (25)
SPARC	NP_033268.1	Secreted acidic cysteine rich glycoprotein	Cell proliferation/Matrix remodeling (44)
			Initiates/facilitates cell migration (43)
			Lens morphology (45)
EIF1	NP_035638.1	Suppressor of initiator codon mutations, related sequence 1	Sex-linked inactivation (100)
UBC	NP_062613.2	Ubiquitin C	Placental development (101)
UTRN	NP_035812.2	Utrophin	Neural development (48)
BLMH	NP_848760.1	Bleomycin hydrolase	Neonatal survival/ Epidermal integrity (102)
CDH1	NP_033994.1	Cadherin 1	Preimplantation survival (103)
			Epithelial cell formation (104)
CPT2	NP_034079.1	Carnitine palmitoyltransferase 2	Implicated requirement for embryonic development (105)
MSY4	NP_620817.2	Cold shock domain protein A long isoform (MSY4)	Essential for normal growth (106)
			Viability of embryo (106)
			Expression declines as the development progresses (106)
EPHA2	NP_034269.2	Eph receptor A2	Early hindbrain development (51)
			Downregulation linked to decreased cell motility (107)
			Promotion of postnatal angiogenesis (53)
			Endothelial cell migration (53)
EIF4G2	NP_038535.2	Eukaryotic translation initiation factor 4, gamma 2	ES Cell Differentiation/Embryonic gastrulation (56)
			Primary germlayer formation (56)
			Limb budding (108)
FHL1	NP_034341.1	Four and a half LIM domains 1	Skeletal/muscle differentiation (54)
GABPA	NP_032091.2	GA repeat binding protein, alpha	Mitochondrial transcription/preimplantation survival (109)
GNL3	NP_849174.1	Guanine nucleotide binding protein-like 3	Regulation of cell proliferation (110)

		(nucleolar) short isoform	
HRMT1L2	NP_062804.1	Heterogeneous nuclear ribonucleoproteins methyltransferase-like 2	Postimplantation development (111)
HNRPU	NP_058085.1	Heterogeneous nuclear ribonucleoprotein U	Upregulation forebrain, mesoderm, limb buds (57)
IMPACT	NP_032404.1	Imprinted and ancient	Paternally controlled genetic expression (112)
			Expressed in the brain only (113)
LGMN	NP_035305.1	Legumain	Kidney tubule cell function (114)
LUC7L	NP_080157.1	Luc7 homolog (S. cerevisiae)-like isoform 1	Regulation of myogenesis (50)
MAT1	NP_032638.1	Menage a trois 1	Postimplantation survival (115)
CPSF5	NP_080899.1	Nudix (nucleoside diphosphate linked moiety X)-type motif 21	Upregulated during lung development (116)
OGT	NP_631883.2	O-linked N-acetylglucosamine transferase	ES viability/Embryo viability (117)
			Somatic cell function/Embryo viability (118)
ROD1	NP_835458.1	ROD1 regulator of differentiation 1 isoform 2	Proposed hematopoiesis (119)
UNRIP	NP_035629.1	Serine/threonine kinase receptor associated protein	Embryonic survival (120)
SLC25A13	NP_056644.1	Citrin	Implicated in limb budding (121)
			May play a role in genetic response to environmental or aging stresses (122)
ATXN10	NP_058539.2	Ataxin 10	Postimplantation survival (123)
SDC4	NP_035651.1	Syndecan 4	Neoangiogenesis/Tissue repair (124)
USP9X	NP_033507.1	Ubiquitin specific protease 9, X chromosome	Sex-linked differentiation (58)
			Preimplantation development (125)
VAT2	NP_031535.2	Vacuolar H+ATPase B2	Plays a role in nephrogenesis (126)

Table 6.6. Top cellular functions of proteins in common with early primitive ectoderm-like cells and embryonic stem cells. Functions were assigned using Ingenuity Functions/Pathways Analysis and selecting computed functions represented by at least one p-value of ≤5.00E-3. Proteins are abbreviated by gene synonym and may map to more than one function. Direction of EPL/ES regulation is indicated: ↑ (up), ↓ (down). Proteins not marked by regulatory indication have expression ratios falling below the threshold for determination of fold change regulation (3 fold).

Function description	**Significance (p-value)**	**Gene ID**
Cellular Movement	2.63E-4 to 4.94E-2	BSG ↑, CRK↑, RHOA, RPSA
Tissue Morphology	3.61E-03 to 2.50E-2	BSG↑, HYOU1↑, RHOA
Cellular Compromise	3.61E-03 to 2.85E-2	GMFB↑, CRK↑, RHOA, CKAP5
Cell Morphology	3.61E-03 to 3.21E-2	CRK↑, CKAP5, RHOA, VDAC3, VCP
Molecular Transport	3.61E-03 to 4.60E-2	TXNRD1↑, IDH1, RHOA, VDAC3
Small Molecule Biochemistry	3.61E-03 to 4.60E-2	BSG↑, TXNRD1↑, ACAA1, HSPE1, IDH1, RHOA, RPSA, VDAC3

Table 6.7. Top cellular functions of proteins in common with early primitive ectoderm-like cells and embryoid bodies. Functions were assigned using Ingenuity Functions/Pathways Analysis and selecting computed functions represented by at least one p-value of ≤5.00E-3. Proteins are abbreviated by gene synonym and may map to more than one function. Direction of EPL/EB regulation is indicated: ↑ (up), ↓ (down). Proteins not marked by regulatory indication have expression ratios falling below the threshold for determination of fold change regulation (3 fold).

Function description	**Significance (p-value)**	**Gene ID**
Cellular Assembly and Organization	2.76E-3	HSPG2↑,GDI1
Molecular Transport	2.76E-03 to 3.00E-2	COPG↑, FABP3↑, M6PRBP1, ATP1A1, AP1B1, DDX19B, ↑, SNX2, VPS35
Cellular Growth and Proliferation	5.52E-3 to 3.60E-2	HSPG2↑, PRDX4, SCYE1,
Carbohydrate Metabolism	2.76E-03	FH↑
Gene Expression	2.76E-03	SIAHBP1↓
RNA Post-Transcriptional Modification	5.52E-03 to 4.07E-2	CPSF6, SCYE1
Cell Death	5.52E-03 to 3.27E-2	PPAT↓, SCYE1

Table 6.8. Primary cellular functions of proteins in common with embryonic stem cells and embryoid bodies. Functions were assigned using Ingenuity Functions/Pathways Analysis and selecting computed functions represented by at least one p-value of ≤5.00E-3. Proteins are abbreviated by gene synonym and may map to more than one function. Direction of ES/EB regulation is indicated: ↑ (up), ↓ (down). Proteins not marked by regulatory indication have expression ratios falling below the threshold for determination of fold change regulation (3-fold).

Function description	**Significance (p-value)**	**Gene ID**
RNA Post-Transcriptional Modification	6.47E-05 to 3.18E-2	CDK7↓, CSTF2↑, DDX52↑, DNAJB11, EIF4A1↓, HCFC1↑, HNRPUL1↑, NOP5/NOP58↑, SFRS2↑, EBNA1BP2, PABPC4, POLR2A, SFRS1, SUPT5H, SYNCRIP
Gene Expression	2.06E-03 to 4.42E-2	DNMT1↑, EIF3S10↓, EIF4A↓, SSRP1↓, XPO1↓, MAPK1, MYH11, PML, POLR2A, SUPT6H, YBX1
Cellular Development	3.39E-03 to 3.73E-2	CBX1↑, PML, MAPK1,
Cell Cycle	5.02E-03 to 4.66E-2	CBX1↑, HCFC1↑, LIG1↓, EBNA1BP2, ESPL1, MSH2, BAZ1B, PCNA, PML, RBBP4, SMARCC1

Table 6.9. Top cellular functions of proteins in common with embryonic stem cells, early primitive ectoderm-like cells, and embryoid bodies. Functions were assigned using Ingenuity Functions/Pathways Analysis and selecting computed functions represented by at least one p-value of ≤5.00E-3. Proteins are abbreviated by gene synonym and may map to more than one function. Direction of regulation is indicated as: ↑ (up), ↓ (down), or no change (●) and is annotated as [EPL/ES]/[EB/ES]/[EPL/EB]. Proteins not marked by regulatory indication have expression ratios falling below the threshold for determination of fold change regulation (3-fold).

Function description	**Significance (p-value)**	**Gene ID**
Protein Folding	3.89E-09	CALR [●/●/↑],ERP29 [●/●/↑],LRPAP1 [●/●/↑], TXN [↓/●/↓], FKBP4 [↓/●/↓], CCT3, CCT4, CCT6A, IQGAP1,PDIA6, PPIA, PPID, TCP1
Cellular Assembly and Organization	3.04E-05 to 3.81E-2	PDIA3 [↑/↓/↑], TLN1 [↑/↓/↑], BASP1 [●/↓/↑], LAMA1[↑/●/↑],TMEM4 [↑/●/↑], SLC25A5 [↑/●/↑],MYH9 [↑/●/●],TXN [↓/●/↓], MYH10, PM3, PPID
Cell Cycle	8.49E-05 to 4.10E-2	DNM1L [↓/↓/●],SMARCA4 [●/↑/↓], MIF [↑/●/↑],DCTN2 [↓/●/●], FASN [↓/●/●], CFL1, SMARCA5
DNA Replication, Recombination, and Repair	8.49E-05 to 3.93E-2	RANBP2 [↓/●/↓],SMARCA4 [●/↑/↓],KPNB1 [●/↑/↓], DCTN2[↓/●/●],PPIB [↓/●/●], GSTP1 [●/●/↓],NME1, NME2, PPIA, SMARCA5
Protein Synthesis	1.67E-04 to 2.27E-2	HSPB1 [↓/↓/●],PABPC1 [↓/↓/●],RPL18A [↓/↓/●],RPL17 [↓/↓/●], FARSLB [↓/↓/●], RPLP1 [●/↓/↑], RPL35 [●/↓/↑], RPL14 [●/↓/↑], TUFM [↓/●/↓],EIF2S1 [↓/●/↓],THOP1 [↑/●/↑], DPP3 [●/↓/●], RPL7 [●/↓/●], EEF2 [●/●/↑], EIF4G1 [●/●/↓], PTBP1, RPL4, RPL22 , RPS5, SSB, UBE2L3, XPNPEP1
Small Molecule Biochemistry	2.55E-04 to 3.93E-2	STIP1[↓/↓/↓], P4HB [↑/●/↑], P4HA1[↑/●/↑], FKBP4 [↓/●/↓], HSP90AB1[↓/●/↓], PPIB [↓/●/●], ADSS, IMPDH2, PPIA
Protein Trafficking	1.04E-03 to 1.75E-2	PDIA3 [↑/↓/↑], RANBP2 [↓/●/↓],CALR [●/●/↑],ERP29 [●/●/↑], ARFGAP1 [●/●/●], CFL1 [●/●/●], HSPA9B (includes EG:3313), NME2, NPM1, TPR
Cellular Growth and Proliferation	2.84E-03 to 4.67E-2	PDIA3 [↑/↓/↑],PSMC3 [↓/↓/●], AKAP12 [↓/●/↓], FKBP4 [↓/●/↓],HNRPAB [↓/●/↓], FASN [↓/●/●],CCT5 [●/↑/●],PPP2R1A [↑/●/●], CALR [●/●/↑], EIF4G1 [●/●/↓], EEF1D [↑/●/↑],GSTP1[●/●/↓], HNRPA1 [↑/●/●],HNRPC [↓/●/↓],HNRPF [↑/↑/●], LAMB1 [↑/↑/↑],MIF [↑/●/↑],PA2G4 [↓/●/↓],SERPINH1 [↑/●/↑], SKP1A [↓/●/↓], SLC3A2 [●/●/↑], SMARCA4 [●/↑/↓],TRAP1 [↓/●/↓], TRIM28 [↓/●/↓],TXN [↓/●/↓]ACTG1, ACTN1, ACTN4, ATP5A1, ATP5B, CCT2, CCT3, DYNC1H1, EIF5A, GGA2, HSPD1, MTAP, NME1, NME2, NPM1, PFN1, PKM2, PSMC5, SMARCA5, TCP1, TPR
Cell Morphology	3.55E-03 to 3.93E-2	PDIA3 [↑/↓/↑], BASP1 [●/↓/↑], FKBP4 [↓/●/↓], LAMA1 [↑/●/↑],TMEM4 [↑/●/↑], MYH9 [↑/●/●],TXN [↓/●/↓], MYH10 , IQGAP1, TPM3
Cell Death	4.32E-03 to 4.10E-2	HSP90B1 [↑/↑/●],HSPA5 [↑/●/↑],HSPB1 [↓/↓/●],TRIM28 [↓/●/↓], MYH9 [↑/●/●], GSTP1 [●/●/↓], CFL1, GLO1, HSPA9B (includes EG:3313), NPM1, PPID

Table 6.10. Identified proteins in common with embryonic stem cells and early primitive ectoderm-like cells. Mascot score, number of peptides, and coverage are totals from nonredundant peptides seen over all three replicate analyses.

			Embryonic stem			Early Primitive Ectoderm-like						
Gene	NCBI Accession	Protein Description	Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Embryonic stem cells Avg N_SC ± STDEV	Early primitive ectoderm-like cells Avg N_SC ± STDEV	EPL/ES Ratio	ES/EPL Ratio
-	NP_079694.2	1110059P08Rik protein	129	1	6	66	1	6	8.93E-05 ± 6.19E-05	1.05E-04 ± -	1.17	0.85
ACAA1	NP_570934.1	acetyl-Coenzyme A acyltransferase 1	112	2	8	267	5	16	9.76E-05 ± 8.29E-05	2.29E-04 ± 2.15E-04	2.34	0.43
AP1M1	NP_001030931.1	adaptor-related protein complex 2, beta 1 subunit isoform a	101	1	2	58	1	2	1.74E-05 ± -	3.40E-05 ± -	1.95	0.51
AP2B1	NP_031482.1	adaptor-related protein complex AP-1, mu subunit 1	52	1	4	61	1	4	3.91E-05 ± -	7.64E-05 ± -	1.95	0.51
AHSA1	NP_666148.1	AHA1, activator of heat shock 90kDa protein ATPase homolog 1	93	1	7	73	1	7	1.47E-04 ± -	1.91E-04 ± -	1.30	0.77
ACIN1	NP_062513.1	apoptotic chromatin condensation inducer 1 isoform 1	120	3	9	54	1	4	8.55E-05 ± -	5.56E-05 ± -	0.65	1.54
ARMET	NP_083379.1	arginine-rich, mutated in early stage tumors	74	1	9	118	2	13	1.85E-04 ± -	1.08E-03 ± 1.28E-03	5.85	0.17
BSG	NP_033898.1	basigin	111	2	6	142	4	11	1.28E-04 ± 6.03E-05	1.37E-03 ± 5.28E-04	10.7	0.09
-	NP_853625.1	cDNA sequence BC022641	141	2	3	174	4	6	2.78E-05 ± 1.82E-05	1.63E-04 ± -	5.85	0.17
CSTF1	NP_077161.1	cleavage stimulation factor, 3' pre-RNA, subunit 1	83	1	6	59	1	3	3.84E-05 ± -	2.25E-04 ± -	5.85	0.17
CKAP5	NP_083713.1	cytoskeleton associated protein 5	92	2	1	75	1	1	1.65E-05 ± -	1.61E-05 ± 0	0.98	1.02
DLD	NP_031887.2	dihydrolipoamide dehydrogenase	254	4	12	98	2	6	2.17E-04 ± 1.61E-04	1.90E-04 ± -	0.88	1.14
DBNL	NP_038838.1	drebrin-like	167	3	12	73	1	5	1.91E-04 ± -	7.46E-05 ± -	0.39	2.56
FUBP1	NP_476513.2	far upstream element (FUSE) binding protein 1	155	3	6	117	2	4	1.80E-04 ± -	1.26E-04 ± 1.07E-04	0.70	1.43
GMFB	NP_071306.1	glia maturation factor, beta	71	1	16	66	1	18	1.17E-04 ± -	1.59E-03 ± -	13.7	0.07
GPI	NP_032181.1	glucose phosphate isomerase 1	98	2	3	79	1	3	1.48E-04 ± -	5.79E-05 ± -	0.39	2.56
HSPE1	NP_032329.1	heat shock protein 1 (chaperonin 10)	351	6	59	419	8	57	1.41E-02 ± 5.29E-03	2.71E-02 ± 3.49E-03	1.93	0.52
HSPA1L	NP_038586.1	heat shock protein 1-like	291	4	9	223	4	9	9.90E-04 ± 4.73E-04	5.38E-04 ± 2.78E-04	0.54	1.84
-	NP_079881.1	hypothetical protein LOC66508	88	1	8	58	1	8	1.03E-04 ± -	2.01E-04 ± -	1.95	0.51
HYOU1	NP_067370.2	hypoxia up-regulated 1	119	2	3	230	4	7	1.08E-04 ± 1.29E-04	6.36E-04 ± 1.95E-04	5.90	0.17
IDH1	NP_034627.2	isocitrate dehydrogenase 1 (NADP+), soluble	167	3	8	145	4	12	9.99E-05 ± 2.83E-05	7.80E-05 ± -	0.78	1.28
RPSA	NP_035159.2	laminin receptor 1 (ribosomal protein SA)	620	9	43	261	4	25	3.05E-03 ± 1.16E-04	2.59E-03 ± 2.46E-03	0.85	1.18
LRRC47	NP_957678.1	leucine rich repeat containing 47	143	2	7	60	1	4	2.85E-05 ± 0.00E+00	5.56E-05 ± 0	1.95	0.51
MAP4	NP_032659.1	microtubule-associated protein 4	177	5	6	63	1	1	2.13E-04 ± 1.04E-05	1.15E-04 ± -	0.54	1.86
NOLC1	NP_001034440.1	nucleolar and coiled-body phosphoprotein 1 isoform B	91	2	2	163	2	5	1.26E-04 ± 3.60E-05	9.23E-05 ± -	0.73	1.37
NUDCD2	NP_080299.3	nudC domain containing 2	85	1	11	66	1	11	1.05E-04 ± -	3.09E-04 ± 1.45E-04	2.93	0.34
FARSLA	NP_079924.1	phenylalanine-tRNA synthetase-like, alpha subunit	63	1	7	217	4	26	5.87E-05 ± -	2.86E-04 ± 2.43E-04	4.88	0.20
PICALM	NP_666306.2	phosphatidylinositol-binding clathrin assembly protein	203	4	10	64	1	3	1.00E-04 ± -	4.89E-05 ± -	0.49	2.05
PITRM1	NP_660113.1	pitrilysin metalloprotease 1	62	1	2	66	1	2	1.60E-05 ± -	2.49E-04 ± -	15.6	0.06
-	XP_001003532.1	PREDICTED: similar to 3-phosphoglycerate dehydrogenase	305	5	12	265	3	10	3.42E-04 ± 2.64E-04	5.15E-04 ± 6.43E-04	1.51	0.66
-	XP_921641.1	PREDICTED: similar to 60S ribosomal protein L13a	133	3	15	62	1	5	4.70E-03 ± 2.64E-03	9.02E-04 ± 9.05E-04	0.19	5.21
-	XP_990185.1	PREDICTED: similar to 60S ribosomal protein L7a (Surfeit locus protein 3)	423	9	28	380	9	30	1.42E-03 ± 4.23E-04	1.16E-03 ± 1.43E-04	0.82	1.23
-	XP_916290.1	PREDICTED: similar to Adenylate kinase isoenzyme 4, mitochondrial	175	2	24	71	1	16	1.98E-04 ± 4.30E-05	1.45E-04 ± -	0.73	1.37
-	XP_995709.1	PREDICTED: similar to Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	69	1	10	58	1	10	6.02E-04 ± 1.42E-04	8.81E-04 ± 6.92E-04	1.46	0.68
-	XP_001002974.1	PREDICTED: similar to Heat shock 70 kDa protein 1B (HSP70.1) isoform 6	294	4	10	136	3	7	8.37E-04 ± 3.75E-04	3.96E-04 ± 2.84E-04	0.47	2.12
-	XP_990050.1	PREDICTED: similar to heterogeneous nuclear ribonucleoprotein A3	918	15	28	483	7	22	2.40E-03 ± 1.11E-03	1.05E-03 ± 9.35E-04	0.44	2.29
-	XP_980342.1	PREDICTED: similar to heterogenous nuclear ribonucleoprotein U	525	11	15	153	2	3	3.11E-04 ± 9.52E-05	1.82E-04 ± 1.43E-04	0.59	1.71
-	XP_994132.1	PREDICTED: similar to peptidylprolyl isomerase D	75	2	19	82	2	13	8.58E-04 ± -	1.91E-03 ± -	2.23	0.45
-	XP_980250.1	PREDICTED: similar to ribosomal protein L27a	204	4	30	258	3	23	1.79E-03 ± 8.74E-04	1.67E-03 ± 5.04E-04	0.94	1.07
-	XP_484272.1	PREDICTED: similar to ribosomal protein L9	246	5	32	176	3	27	2.01E-03 ± 1.19E-03	3.03E-03 ± 1.27E-03	1.51	0.66
RHOA	NP_058082.2	ras homolog gene family, member A	81	1	7	73	1	7	8.58E-05 ± -	1.67E-04 ± -	1.95	0.51
RPL18	NP_033103.2	ribosomal protein L18	263	4	22	266	3	15	3.93E-03 ± 1.25E-03	3.84E-03 ± 1.69E-03	0.98	1.02
RPL19	NP_033104.1	ribosomal protein L19	228	3	14	103	1	9	5.35E-04 ± 3.51E-04	2.97E-03 ± -	5.55	0.18
RPL5	NP_058676.1	ribosomal protein L5	276	6	21	102	1	6	2.02E-03 ± 1.24E-03	2.21E-03 ± 5.02E-04	1.09	0.92
RPS10	NP_080239.1	ribosomal protein S10	375	5	33	167	2	15	1.40E-03 ± 4.58E-04	7.18E-04 ± 7.41E-04	0.51	1.96
RPS18	NP_035426.1	ribosomal protein S18	152	3	17	58	1	6	1.42E-03 ± 1.08E-03	6.38E-04 ± -	0.45	2.22

Proteins in common with embryonic stem cells and early primitive ectoderm-like cells, continued.

Gene	NCBI Accession	Protein Description	Embryonic stem: Total Mascot Score	Embryonic stem: Total No. Peptides	Embryonic stem: Total % Protein Coverage	Early Primitive Ectoderm-like: Total Mascot Score	Early Primitive Ectoderm-like: Total No. Peptides	Early Primitive Ectoderm-like: Total % Protein Coverage	Embryonic stem cells: Avg N_SC ± STDEV	Early primitive ectoderm-like cells: Avg N_SC ± STDEV	EPL/ES Ratio	ES/EPL Ratio
SET	NP_076360.1	SET translocation	358	6	23	164	2	15	6.30E-04 ± 2.29E-04	1.12E-03 ± 4.03E-04	1.77	0.56
SDHA	NP_075770.1	succinate dehydrogenase Fp subunit	58	1	3	153	2	5	7.48E-05 ± -	9.73E-05 ± 6.88E-05	1.30	0.77
SUCLG1	NP_063932.1	succinate-CoA ligase, GDP-forming, alpha subunit	54	1	5	60	1	5	9.94E-05 ± -	9.70E-05 ± -	0.98	1.02
TXNRD1	NP_056577.2	thioredoxin reductase 1	95	1	3	64	1	3	3.32E-05 ± -	1.29E-04 ± -	3.90	0.26
TCEA1	NP_035671.1	transcription elongation factor A (SII) 1	103	1	4	51	1	4	3.85E-04 ± 2.20E-04	1.61E-04 ± 7.59E-05	0.42	2.39
TIMM13	NP_038923.1	translocase of inner mitochondrial membrane 9 homolog	136	3	37	79	1	15	1.28E-03 ± 2.01E-04	1.36E-03 ± 1.44E-03	1.06	0.94
TIMM44	NP_035722.1	translocator of inner mitochondrial membrane 44	120	3	6	75	1	4	1.46E-04 ± -	7.15E-05 ± -	0.49	2.05
TUBA6	NP_033474.1	tubulin, alpha 6	1007	14	43	722	12	40	7.72E-03 ± 9.80E-04	5.44E-03 ± 2.90E-03	0.71	1.42
YARS	NP_598912.2	tyrosyl-tRNA synthetase	272	5	10	80	1	3	2.51E-04 ± 6.28E-05	1.53E-04 ± 1.30E-04	0.61	1.64
VCP	NP_033529.2	valosin containing protein	1223	19	35	1096	16	29	1.88E-03 ± 5.08E-05	1.99E-03 ± 1.37E-03	1.06	0.95
CRK	NP_598417.2	v-crk sarcoma virus CT10 oncogene homolog	66	1	6	158	4	17	5.44E-05 ± -	1.59E-04 ± 7.51E-05	2.93	0.34
VDAC3	NP_035826.1	voltage-dependent anion channel 3	75	1	4	80	2	10	4.68E-04 ± -	1.14E-03 ± -	2.44	0.41
VBP1	NP_035822.1	von Hippel-Lindau binding protein 1	88	1	8	62	1	6	2.07E-04 ± -	4.04E-04 ± -	1.95	0.51
WDR74	NP_598900.1	WD repeat domain 74	66	1	4	82	1	5	1.08E-04 ± 9.17E-05	8.41E-05 ± -	0.78	1.28

Table 6.11. Identified proteins in common with early primitive ectoderm-like cells and embryoid bodies. Mascot score, number of peptides, and coverage are totals from nonredundant peptides seen over all three replicate analyses.

			Early Primitive Ectoderm-like			Embryoid Bodies							
Gene	NCBI Accession	Protein Description	Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Early primitive ectoderm-like cells AVG N_SC ± STDEV	Embryoid Bodies AVG N_SC ± STDEV	EPL/EB Ratio	EB/EPL Ratio
AP1B1	NP_031480.2	adaptor protein complex AP-1, beta 1 subunit	137	1	3	217	3	5	1.14E-04 ± 7.91E-05	5.92E-05 ± -	1.93	0.52
ARCN1	NP_666097.2	archain 1	146	4	9	84	1	2	7.03E-05 ± -	5.07E-06 ± 0	13.89	0.07
DNPEP	NP_058574.2	aspartyl aminopeptidase	66	1	3	96	1	3	4.78E-04 ± -	9.44E-04 ± -	0.51	1.98
CDC42	NP_033991.1	cell division cycle 42 homolog	193	4	27	95	2	16	7.89E-04 ± 5.17E-04	2.92E-04 ± -	2.70	0.37
CPSF6	NP_001013409.1	cleavage and polyadenylation specific factor 6	75	1	3	76	1	3	1.17E-04 ± -	5.91E-05 ± 3.87E-05	1.98	0.50
COPG	NP_059505.1	coatomer protein complex, subunit gamma isoform 1	296	4	7	98	1	2	2.71E-04 ± 1.40E-04	5.86E-05 ± 4.88E-05	4.63	0.22
CRYZ	NP_034098.1	crystallin, zeta	103	1	8	96	1	7	9.76E-05 ± -	2.53E-04 ± -	0.39	2.59
COX2	NP_904331.1	cytochrome c oxidase subunit II	89	1	9	156	3	20	1.42E-04 ± -	3.07E-04 ± -	0.46	2.16
DDX19B	NP_758488.1	DDX19 homolog	53	1	3	68	1	2	6.74E-05 ± -	8.74E-05 ± -	0.77	1.30
DDX48	NP_619610.1	DEAD (Asp-Glu-Ala-Asp) box polypeptide 48	114	2	9	250	5	13	1.57E-04 ± 7.86E-05	1.92E-04 ± 1.04E-04	0.82	1.22
DLST	NP_084501.1	dihydrolipoamide S-succinyltransferase (E2 component of 2-oxo-glutarate complex)	210	2	7	253	5	15	6.05E-04 ± 5.03E-05	3.28E-04 ± 2.33E-04	1.84	0.54
FABP3	NP_034304.1	fatty acid binding protein 3	407	7	44	64	1	11	1.06E-02 ± 3.40E-03	1.05E-04 ± 0	101.1	0.01
FH	NP_034339.1	fumarate hydratase 1	81	1	4	89	1	4	1.27E-04 ± -	2.75E-05 ± -	4.63	0.22
SIAHBP1	NP_082640.1	fuse-binding protein-interacting repressor isoform a	92	1	3	64	1	3	5.73E-05 ± -	1.98E-04 ± -	0.29	3.46
GDI1	NP_034403.1	guanosine diphosphate (GDP) dissociation inhibitor 1	195	2	9	221	4	12	1.08E-04 ± 5.11E-05	1.04E-04 ± 6.50E-05	1.04	0.96
HADHB	NP_663533.1	hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A h	134	2	9	77	1	3	1.81E-04 ± 7.85E-05	2.06E-04 ±	0.88	1.13
HSPG2	XP_983973.1	heparan sulfate proteoglycan 2 (perlecan)	325	5	2	149	2	1	1.03E-04 ± 1.25E-04	1.75E-05 ± 1.12E-05	5.89	0.17
KIAA0251	NP_444411.2	hypothetical protein LOC94184 isoform 1	99	1	3	58	1	2	1.23E-04 ± 7.11E-05	1.77E-05 ±	6.94	0.14
ITGB1	NP_034708.1	integrin beta 1 (fibronectin receptor beta)	72	1	2	69	1	2	1.01E-04 ± 2.86E-05	6.41E-05 ± 5.05E-05	1.58	0.63
LAMC1	NP_034813.1	laminin, gamma 1	301	5	21	81	2	7	2.06E-03 ± 5.59E-04	1.74E-05 ± 8.69E-06	118.8	0.01
M6PRBP1	NP_080112.1	mannose-6-phosphate receptor binding protein 1	109	2	9	51	1	3	7.39E-04 ± -	3.19E-05 ± -	23.14	0.04
ATP1A1	NP_659149.1	Na+/K+ -ATPase alpha 1 subunit	300	4	7	64	1	1	6.84E-04 ± 1.82E-04	1.36E-05 ± -	50.1	0.02
NID2	NP_032721.2	nidogen 2	236	6	9	157	2	3	1.15E-04 ± 9.77E-05	6.96E-05 ± -	1.65	0.60
OAT	NP_058674.1	ornithine aminotransferase	161	2	7	186	5	11	1.47E-04 ± 0	9.54E-05 ± 0	1.54	0.65
PAK2	NP_796300.1	p21-activated kinase 2	54	1	3	61	1	3	9.25E-05 ± 4.36E-05	2.66E-05 ± -	3.47	0.29
PPAT	XP_979032.1	phosphoribosyl pyrophosphate amidotransferase	76	1	3	108	1	3	6.25E-05 ± -	1.75E-04 ± 5.73E-05	0.36	2.81
PRDX4	NP_058044.1	peroxiredoxin 4	95	2	7	95	2	7	2.36E-04 ± 0	3.06E-04 ± 5.09E-05	0.77	1.30
PBEF1	NP_067499.1	pre-B-cell colony-enhancing factor 1	61	1	6	152	2	8	1.32E-04 ± 0	1.42E-04 ± 1.24E-04	0.93	1.08
-	XP_991849.1	PREDICTED: similar to 40S ribosomal protein S25	115	2	18	180	4	26	2.58E-03 ± -	7.03E-03 ± 8.69E-03	0.37	2.72
-	XP_898748.1	PREDICTED: similar to 40S ribosomal protein S26	106	2	23	152	3	31	1.12E-03 ± 1.19E-03	4.61E-03 ± 7.38E-04	0.24	4.11
-	XP_919631.1	PREDICTED: similar to 40S ribosomal protein S6 isoform 1	130	3	12	285	5	17	7.13E-04 ± 2.75E-04	3.27E-03 ± 2.38E-03	0.22	4.58
-	XP_991128.1	PREDICTED: similar to basic transcription factor 3	51	1	12	100	1	12	1.39E-03 ± -	3.43E-04 ± 2.42E-04	4.05	0.25
-	XP_001004821.1	PREDICTED: similar to Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	258	5	13	288	6	16	3.60E-03 ± 1.24E-03	3.34E-03 ± 5.35E-04	1.08	0.93
-	XP_919830.2	PREDICTED: similar to glyceraldehyde-3-phosphate dehydrogenase	133	2	16	151	2	14	3.11E-03 ± 6.73E-04	3.94E-03 ± 6.75E-04	0.79	1.27
-	XP_994513.1	PREDICTED: similar to H2A histone family, member V isoform 1	363	6	28	244	4	20	1.69E-02 ± 5.80E-03	1.54E-03 ± 1.98E-04	10.98	0.09
-	XP_981174.1	PREDICTED: similar to pyruvate dehydrogenase (lipoamide) beta	98	2	4	51	1	2	2.17E-04 ± -	9.37E-05 ± -	2.31	0.43
-	XP_996982.1	PREDICTED: similar to Ribosome-binding protein 1 (Ribosome receptor protein) (mRRp)	305	4	3	121	3	2	4.58E-05 ± 2.27E-05	6.79E-05 ± -	0.67	1.48
-	XP_984595.1	PREDICTED: similar to Signal recognition particle receptor beta subunit (SR-beta) isoform 2	61	1	5	161	2	11	1.20E-04 ± -	2.42E-04 ± 1.59E-04	0.50	2.02
-	XP_994029.1	PREDICTED: spectrin alpha 2 isoform 20	591	9	7	594	10	6	1.13E-04 ± 7.51E-06	8.43E-05 ± 1.12E-05	1.34	0.75
PLOD2	NP_036091.1	procollagen lysine, 2-oxoglutarate 5-dioxygenase 2	227	5	10	114	2	4	3.73E-04 ± 9.30E-05	5.68E-05 ± 2.68E-05	6.56	0.15
PSMA6	NP_036098.1	proteasome (prosome, macropain) subunit, alpha type 6	67	1	7	105	2	12	6.13E-04 ± 2.01E-04	2.55E-04 ± 4.01E-05	2.40	0.42
PRPF3	NP_081817.2	PRP3 pre-mRNA processing factor 3 homolog	56	1	3	144	2	4	9.46E-05 ± -	8.17E-05 ± -	1.16	0.86
RPN1	NP_598694.2	ribophorin I	198	4	5	150	3	5	2.13E-04 ± 7.51E-05	4.59E-05 ± -	4.63	0.22
RPE	NP_079959.2	ribulose-5-phosphate-3-epimerase	97	1	8	86	1	8	2.13E-04 ± 1.00E-04	6.12E-05 ± -	3.47	0.29
S100G	NP_033919.1	S100 calcium binding protein G	59	1	15	58	1	15	1.23E-03 ± 5.78E-04	6.01E-03 ± -	0.20	4.90
SCOC	NP_062682.1	short coiled-coil protein isoform b	159	2	24	60	1	22	5.25E-04 ± 2.27E-04	1.70E-04 ± -	3.09	0.32
SCYE1	NP_031952.1	small inducible cytokine subfamily E, member 1	147	2	6	78	1	4	1.04E-04 ± -	4.50E-05 ± -	2.31	0.43
SNX2	NP_080662.1	sorting nexin 2	56	1	5	164	2	7	6.22E-05 ± -	4.48E-05 ± 1.55E-05	1.39	0.72
VPS35	NP_075373.1	vacuolar protein sorting 35	51	1	3	162	2	4	4.06E-05 ± 0	7.01E-05 ± -	0.58	1.73
VDAC2	NP_035825.1	voltage-dependent anion channel 2	288	7	39	100	2	10	7.30E-04 ± 5.51E-04	3.31E-04 ± -	2.20	0.45

Table 6.12. Proteins in common with embryonic stem cells and embryoid bodies. Mascot score, number of peptides, and coverage are totals from nonredundant peptides seen over all three replicate analyses.

			Embryonic stem cells			Embryoid Bodies						
Gene	NCBI Accession	Protein Description	Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Embryonic stem cells AVG N_SC ± STDEV	Embryoid Bodies AVG N_SC ± STDEV	ES/EB Ratio	EB/ES Ratio
-	NP_666205.1	2B28	124	2	16	115	2	16	2.04E-04 ± 8.51E-05	1.64E-04 ± 9.97E-05	1.24	0.80
YWHAG	NP_061359.2	3-monooxgenase/tryptophan 5-monooxygenase activation protein, gamma polypeptide	415	6	24	476	7	34	2.21E-03 ± 1.08E-03	3.05E-03 ± 1.00E-03	0.72	1.38
ACP1	NP_067305.2	acid phosphatase 1, soluble	90	1	11	88	1	11	1.47E-03 ± -	2.36E-04 ± 1.84E-04	6.23	0.16
ANP32E	NP_075699.2	acidic (leucine-rich) nuclear phosphoprotein 32 family, member E	57	1	5	317	4	18	1.27E-04 ± -	2.33E-04 ± 1.24E-04	0.55	1.83
ACO2	NP_542364.1	aconitase 2, mitochondrial	170	5	10	121	1	2	1.27E-04 ± -	7.16E-05 ± 5.06E-05	1.78	0.56
ADSL	NP_033764.2	adenylosuccinate lyase	75	1	4	77	1	4	1.37E-04 ± -	2.40E-04 ± 1.30E-04	0.57	1.76
AARS	NP_666329.2	alanyl-tRNA synthetase	339	7	10	104	2	4	1.08E-04 ± 8.61E-05	1.59E-04 ± -	0.68	1.46
ALDH2	NP_033786.1	aldehyde dehydrogenase 2, mitochondrial	161	4	9	271	5	16	1.91E-04 ± -	1.88E-04 ± 1.14E-04	1.02	0.98
AGPS	NP_766254.1	alkyldihydroxyacetone phosphate synthase	112	2	6	131	2	7	7.70E-05 ± -	2.49E-04 ± 4.59E-05	0.31	3.23
AOF2	NP_598633.1	amine oxidase (flavin containing) domain 2	165	3	9	60	1	2	4.12E-05 ± -	1.04E-04 ± -	0.40	2.53
NPEPPS	NP_032968.2	aminopeptidase puromycin sensitive	119	3	5	125	2	3	1.26E-04 ± -	5.31E-05 ± 1.07E-05	2.37	0.42
ANXA6	NP_038500.2	annexin A6	122	2	4	94	1	4	1.23E-04 ± -	2.07E-05 ± -	5.93	0.17
APOE	NP_033826.1	apolipoprotein E	128	2	9	202	4	16	2.84E-04 ± 1.23E-04	1.24E-03 ± 1.73E-03	0.23	4.37
APEX1	NP_033817.1	apurinic/apyrimidinic endonuclease 1	248	4	26	256	4	25	2.78E-04 ± 1.31E-04	3.08E-04 ± 1.16E-04	0.90	1.11
ASS	NP_031520.1	argininosuccinate synthetase	139	3	11	289	5	17	1.87E-04 ± 1.16E-04	3.73E-04 ± 3.91E-04	0.50	1.99
RARS	NP_080212.1	arginyl-tRNA synthetase	358	6	12	268	6	13	2.51E-04 ± 1.15E-04	9.52E-05 ± 7.48E-05	2.64	0.38
ARS2	NP_113582.1	arsenate resistance protein 2	64	1	3	74	1	3	5.67E-05 ± 0.00E+00	6.38E-05 ± 2.76E-05	0.89	1.12
ASNS	NP_036185.1	asparagine synthetase	182	3	7	221	5	12	3.84E-04 ± -	8.71E-05 ± 5.28E-05	4.40	0.23
NARS	NP_081626.1	asparaginyl-tRNA synthetase	248	4	8	245	4	8	4.14E-04 ± 1.75E-04	2.04E-04 ± 9.20E-05	2.03	0.49
DARS	NP_663482.1	aspartyl-tRNA synthetase	188	3	10	136	2	6	3.63E-04 ± 1.87E-04	3.76E-04 ± 4.14E-04	0.97	1.03
ACLY	NP_598798.1	ATP citrate lyase	334	7	8	278	7	6	1.57E-04 ± 1.08E-04	1.47E-04 ± 9.05E-06	1.07	0.94
ATP5H	NP_082138.1	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit d	118	2	18	97	2	18	7.88E-04 ± 3.14E-04	8.67E-04 ± -	0.91	1.10
ATP5C1	NP_065640.1	ATP synthase, H+ transporting, mitochondrial F1 complex, gamma subunit	81	1	4	138	3	12	2.22E-04 ± -	2.34E-04 ± 2.65E-04	0.95	1.05
ATP5O	NP_613063.1	ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit	155	2	15	105	1	10	3.89E-04 ± 1.55E-04	2.84E-04 ± 1.36E-04	1.37	0.73
ABCE1	NP_056566.1	ATP-binding cassette, subfamily E, member 1	125	2	4	284	5	11	9.67E-05 ± 1.95E-05	2.56E-04 ± 1.30E-04	0.38	2.65
BIRC6	NP_031592.1	baculoviral IAP repeat-containing 6	54	1	0	59	1	0	3.42E-06 ± -	8.64E-06 ± -	0.40	2.53
HEATR1	NP_659084.3	BAP28 protein	158	4	2	159	3	2	2.32E-05 ± 0.00E+00	1.63E-05 ± 1.38E-05	1.42	0.70
BZW2	NP_080116.2	basic leucine zipper and W2 domains 2	119	1	5	112	1	5	5.93E-05 ± 2.79E-05	7.77E-05 ± 5.09E-05	0.76	1.31
BID	NP_031570.2	BH3 interacting domain death agonist	182	2	15	160	2	15	8.91E-04 ± 5.40E-04	1.00E-03 ± -	0.89	1.12
BPY2IP1	NP_766601.1	BPY2 interacting protein 1	61	1	2	75	1	2	1.70E-05 ± -	4.30E-05 ± 2.48E-05	0.40	2.53
BXDC2	NP_080672.2	brix domain containing 2	66	1	5	248	5	33	1.20E-04 ± 8.48E-05	1.85E-04 ± 1.27E-04	0.65	1.55
BAZ1B	NP_035844.1	bromodomain adjacent to zinc finger domain, 1B	139	3	2	147	2	3	4.48E-05 ± 1.12E-05	3.46E-05 ± 5.45E-06	1.29	0.77
CACYBP	NP_033916.1	calcyclin binding protein	52	1	10	83	1	10	5.78E-04 ± -	1.93E-03 ± 2.46E-04	0.30	3.34
CANX	NP_031623.1	calnexin	317	5	12	344	5	8	9.43E-04 ± 6.18E-04	4.80E-04 ± 5.94E-05	1.96	0.51
A1 (includes E	NP_033927.1	capping protein (actin filament) muscle Z-line, alpha 1	87	1	5	52	1	6	1.74E-04 ± 8.18E-05	1.22E-03 ± -	0.14	7.03
CTNNA1	NP_033948.1	catenin (cadherin associated protein), alpha 1	181	3	5	198	3	7	6.39E-05 ± 1.29E-05	4.62E-05 ± 0.00E+00	1.38	0.72
CDC2	NP_031685.2	cell division cycle 2 homolog A	150	2	10	299	4	23	2.23E-04 ± -	1.72E-04 ± 7.18E-05	1.29	0.77
ZW10	NP_036169.1	centromere/kinetochore protein zw10	54	1	3	88	1	3	2.12E-05 ± -	5.38E-05 ± -	0.40	2.53
CSPG6	NP_031816.2	chondroitin sulfate proteoglycan 6	69	1	1	185	3	4	1.36E-05 ± -	1.91E-05 ± 1.32E-05	0.71	1.41
CHMP4B	NP_083638.1	chromatin modifying protein 4B	54	1	12	54	1	12	2.22E-04 ± -	2.49E-04 ± -	0.89	1.12
CBX1	NP_031648.1	chromobox homolog 1	107	1	9	170	2	25	1.01E-03 ± 3.14E-04	2.51E-04 ± 1.15E-04	4.03	0.25
CHD4	NP_666091.1	chromodomain helicase DNA binding protein 4	254	5	5	166	3	4	2.02E-05 ± 4.99E-06	2.92E-05 ± 3.09E-05	0.69	1.45
CHD4	NP_076054.1	chromosome segregation 1-like	81	1	2	415	7	11	1.02E-04 ± -	2.20E-04 ± 8.66E-05	0.46	2.16
CROP	NP_080589.1	cisplatin resistance-associated overexpressed protein	57	1	2	67	1	2	1.01E-04 ± -	1.71E-04 ± 2.84E-05	0.59	1.69
CSTF2	NP_573459.1	cleavage stimulation factor, 3' pre-RNA subunit 2	138	2	8	64	1	2	1.05E-04 ± 4.36E-05	2.41E-05 ± -	4.35	0.23
COPZ1	NP_062791.1	coatomer protein complex, subunit zeta 1	85	1	13	150	2	13	9.35E-05 ± -	6.31E-04 ± 3.61E-04	0.15	6.75
C1QBP	NP_031599.1	complement component 1, q subcomponent binding protein	155	2	15	302	4	29	6.35E-04 ± 1.37E-04	2.51E-03 ± 1.13E-03	0.25	3.95
COPS3	NP_036121.1	COP9 (constitutive photomorphogenic), subunit 3	94	2	9	159	2	11	7.83E-05 ± -	1.21E-04 ± 8.30E-05	0.65	1.55
CKB	NP_067248.1	creatine kinase, brain	300	6	23	207	3	15	9.41E-04 ± 4.35E-04	7.33E-05 ± 6.35E-05	12.85	0.08
CUL3	NP_057925.1	cullin 3	85	1	2	102	1	2	5.39E-05 ± 1.52E-05	4.54E-05 ± 1.29E-05	1.19	0.84
CUL5	NP_082083.1	cullin 5	68	1	2	62	1	2	2.12E-05 ± -	1.79E-05 ± -	1.19	0.84
CSE1L	NP_598838.2	cutaneous T-cell lymphoma tumor antigen se70-2	63	1	3	62	1	3	4.25E-05 ± 1.84E-05	4.03E-05 ± -	1.05	0.95
CDK7	NP_034004.1	cyclin-dependent kinase 7 (homolog of Xenopus MO15 cdk-activating kinase)	60	1	4	94	1	4	4.78E-05 ± -	1.75E-04 ± 9.32E-05	0.27	3.65
CARS	NP_038770.2	cysteinyl-tRNA synthetase	332	6	11	117	1	3	1.10E-04 ± 7.04E-05	3.36E-05 ± -	3.26	0.31
CYC1	NP_058028.1	cytidine 5'-triphosphate synthase	170	3	6	220	4	8	1.40E-04 ± -	8.27E-05 ± 8.35E-05	1.69	0.59
N-PAC	NP_079843.1	cytochrome c-1	104	2	11	145	2	11	1.53E-04 ± -	2.58E-04 ± -	0.59	1.69

Proteins in common with embryonic stem cells and embryoid bodies, continued.

Gene	NCBI Accession	Protein Description	Embryonic stem cells: Total Mascot Score	Embryonic stem cells: Total No. Peptides	Embryonic stem cells: Total % Protein Coverage	Embryoid Bodies: Total Mascot Score	Embryoid Bodies: Total No. Peptides	Embryoid Bodies: Total % Protein Coverage	Embryonic stem cells: AVG N_SC ± STDEV	Embryoid Bodies: AVG N_SC ± STDEV	ES/EB Ratio	EB/ES Ratio
-	NP_803425.1	D19Bwg1357e protein	185	4	10	92	2	4	1.28E-04 ± 1.08E-04	1.08E-04 ± -	1.19	0.84
DDX18	NP_080136.2	DEAD (Asp-Glu-Ala-Asp) box polypeptide 18	218	4	10	263	5	15	7.52E-05 ± 7.09E-05	9.16E-05 ± 2.44E-05	0.82	1.22
DDX39	NP_932099.2	DEAD (Asp-Glu-Ala-Asp) box polypeptide 39	301	6	15	386	9	24	4.52E-04 ± 4.49E-04	5.77E-04 ± 3.77E-05	0.78	1.28
DDX51	NP_081432.2	DEAD (Asp-Glu-Ala-Asp) box polypeptide 51	98	2	6	75	1	3	7.77E-05 ± -	4.37E-05 ± -	1.78	0.56
DDX52	NP_084372.1	DEAD (Asp-Glu-Ala-Asp) box polypeptide 52	158	4	7	52	1	3	4.57E-04 ± 4.89E-04	2.33E-05 ± -	19.57	0.05
DDX6	NP_031867.1	DEAD (Asp-Glu-Ala-Asp) box polypeptide 6	151	3	9	180	3	10	2.74E-04 ± 1.94E-04	5.97E-04 ± 4.42E-04	0.46	2.18
DHX15	NP_031865	DEAH (Asp-Glu-Ala-His) box polypeptide 15	53	1	2	89	1	2	4.16E-05 ± -	2.98E-04 ± 1.24E-04	0.14	7.17
DHX29	NP_766182.2	DEAH (Asp-Glu-Ala-His) box polypeptide 29	80	2	2	56	1	1	4.85E-05 ± -	1.02E-05 ± -	4.74	0.21
DENR	NP_080879.1	density-regulated protein	104	2	17	60	1	8	1.67E-04 ± -	4.23E-04 ± -	0.40	2.53
DHFR	NP_064360.1	differential display and activated by p53	149	2	12	63	1	6	1.01E-04 ± 5.03E-05	8.48E-05 ± 6.00E-05	1.19	0.84
DAK	NP_034179.1	dihydrofolate reductase	59	1	9	55	1	9	2.83E-03 ± 5.01E-04	4.48E-04 ± -	6.32	0.16
DNMT1	NP_034196.2	DNA methyltransferase (cytosine-5) 1	294	8	6	87	1	1	4.09E-05 ± 4.34E-05	8.62E-06 ± -	4.74	0.21
PRIM2A	NP_032948.1	DNA primase, p58 subunit	133	3	6	81	2	4	1.31E-04 ± 9.83E-05	6.91E-05 ± 5.86E-05	1.90	0.53
DNAJB11	NP_080676.2	DnaJ (Hsp40) homolog, subfamily B, member 11	87	1	4	61	1	4	1.23E-04 ± 5.34E-05	9.75E-05 ± 2.76E-05	1.26	0.79
DNAJC7	NP_062769.2	DnaJ (Hsp40) homolog, subfamily C, member 7	109	2	6	56	1	3	8.93E-05 ± 1.93E-05	1.13E-04 ± 0.00E+00	0.79	1.26
DUSP9	NP_083628.3	dual specificity phosphatase 9	151	3	13	60	1	3	9.15E-05 ± 7.77E-05	9.26E-05 ± -	0.99	1.01
EHD1	NP_034249.1	EH-domain containing 1	182	5	16	57	1	2	2.48E-04 ± 0.00E+00	2.61E-05 ± 0.00E+00	9.49	0.11
HNRPUL1	NP_659171.1	EIB-55kDa associated protein 5 isoform 1	94	1	4	100	1	4	2.70E-04 ± -	3.79E-05 ± 2.48E-05	7.11	0.14
ELAVL1	NP_034615.2	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 1 (Hu antigen R)	426	8	28	98	3	7	6.60E-04 ± 2.03E-04	2.57E-04 ± 1.21E-04	2.57	0.39
EMD	NP_031953.1	emerin	127	2	15	51	1	6	1.05E-03 ± 4.97E-04	7.01E-04 ± -	1.51	0.66
EPS8	NP_031971.2	epidermal growth factor receptor pathway substrate 8	56	1	2	78	1	2	4.03E-05 ± -	3.40E-05 ± 1.70E-05	1.19	0.84
EIF2S2	NP_080306.1	eukaryotic translation initiation factor 2, subunit 2 (beta)	323	5	19	311	5	18	8.00E-04 ± 2.65E-04	5.76E-04 ± 3.51E-04	1.39	0.72
EIF2S3	NP_036140.1	eukaryotic translation initiation factor 2, subunit 3, structural gene X-linked	406	7	21	251	4	16	1.75E-04 ± 7.01E-05	1.08E-04 ± 1.71E-05	1.62	0.62
EIF2B4	NP_034252.1	eukaryotic translation initiation factor 2B, subunit 4 delta	89	1	3	78	1	3	3.16E-05 ± -	1.33E-04 ± -	0.24	4.22
EIF3S6IP	NP_660121.2	eukaryotic translation initiation factor 3 subunit 6 interacting protein	89	2	4	544	8	18	8.80E-05 ± -	8.33E-04 ± 6.14E-04	0.11	9.46
EIF3S10	NP_034253.2	eukaryotic translation initiation factor 3, subunit 10 (theta)	308	8	8	417	7	7	9.85E-05 ± 8.71E-05	4.22E-04 ± 2.94E-04	0.23	4.29
EIF3S2	NP_061269.1	eukaryotic translation initiation factor 3, subunit 2 (beta)	151	4	26	52	1	10	2.04E-04 ± -	3.01E-04 ± -	0.68	1.48
EIF3S4	NP_058572.2	eukaryotic translation initiation factor 3, subunit 4 (delta)	223	4	24	178	4	24	3.97E-04 ± 1.82E-04	4.36E-04 ± 1.23E-04	0.91	1.10
EIF3S5	NP_079620.1	eukaryotic translation initiation factor 3, subunit 5 (epsilon)	122	2	7	105	1	5	3.21E-04 ± 1.95E-04	3.87E-05 ± -	8.30	0.12
EIF4A1	NP_659207.1	eukaryotic translation initiation factor 4A1	715	9	27	754	9	27	1.64E-03 ± 1.65E-04	5.65E-03 ± 1.87E-03	0.29	3.44
EIF5	NP_775539.1	eukaryotic translation initiation factor 5	133	3	10	113	2	8	7.72E-05 ± -	2.60E-04 ± -	0.30	3.37
XPO1	NP_001030303.1	exportin 1, CRM1 homolog	73	1	2	256	5	7	2.32E-05 ± 1.09E-05	2.35E-04 ± 0.00E+00	0.10	10.12
ESPL1	NP_001014976.1	extra spindle poles-like 1	120	2	2	95	1	1	7.81E-06 ± -	1.32E-05 ± -	0.59	1.69
FTL	NP_034370.1	ferritin light chain 1	295	5	40	179	3	28	9.35E-04 ± 4.46E-04	1.81E-03 ± 4.66E-04	0.52	1.93
FYTTD1	NP_081502.2	forty-two-three domain containing 1	87	1	5	54	1	5	8.70E-05 ± 6.03E-05	3.96E-04 ± -	0.22	4.55
FTSJ3	NP_079586.1	FtsJ homolog 3	335	4	10	188	3	8	9.88E-05 ± 1.12E-04	2.33E-04 ± 1.65E-04	0.42	2.36
GMNN	NP_065592.1	geminin	75	1	7	124	1	7	8.03E-05 ± -	2.03E-04 ± 1.92E-04	0.40	2.53
GOT1	NP_034454.1	glutamate oxaloacetate transaminase 1, soluble	187	3	12	51	1	4	2.40E-04 ± -	3.38E-05 ± -	7.11	0.14
GORASP2	NP_081628.2	golgi reassembly stacking protein 2	153	2	8	131	2	8	2.02E-04 ± 2.59E-05	1.44E-04 ± 6.44E-05	1.40	0.72
GRPEL1	NP_077798.1	GrpE-like 1, mitochondrial	66	1	5	63	1	5	4.32E-04 ± 1.17E-04	1.50E-04 ± 9.83E-05	2.88	0.35
GMPS	NP_001028472.2	guanine monphosphate synthetase	289	5	11	296	5	11	1.19E-04 ± 6.76E-05	8.06E-05 ± 2.85E-05	1.48	0.67
GNB2L1	NP_032169.1	guanine nucleotide binding protein (G protein), beta polypeptide 2 like 1	391	7	40	444	7	36	5.48E-04 ± 4.06E-04	6.46E-04 ± 2.54E-05	0.85	1.18
GNA11	NP_034431.1	guanine nucleotide binding protein, alpha 11	52	1	4	56	1	4	9.22E-05 ± -	7.78E-05 ± -	1.19	0.84
H2AFY	NP_036145.1	H2A histone family, member Y	93	1	5	150	2	8	1.78E-04 ± 7.71E-05	1.25E-04 ± 7.81E-05	1.42	0.70
HELLS	NP_032260.2	helicase, lymphoid specific	203	3	5	163	3	4	6.72E-05 ± 3.08E-05	8.50E-05 ± 4.50E-05	0.79	1.26
HDGF	NP_032257.2	hepatoma-derived growth factor	214	3	33	110	2	18	1.75E-03 ± 1.27E-03	7.66E-04 ± -	2.28	0.44
HNRPDL	NP_057899.1	heterogeneous nuclear ribonucleoprotein D-like	138	3	8	93	2	6	1.18E-03 ± 2.72E-04	2.18E-03 ± 3.02E-03	0.54	1.84
HARS	NP_032240.2	histidyl-tRNA synthetase	57	1	3	55	1	4	3.25E-05 ± -	2.74E-05 ± -	1.19	0.84
HAT1	NP_080391.2	histone aminotransferase 1	152	3	10	375	5	21	1.19E-04 ± -	3.13E-04 ± 1.08E-04	0.38	2.62
HCFC1	NP_032250.2	host cell factor C1	242	6	6	73	1	1	4.59E-05 ± 1.68E-05	1.02E-05 ± 4.83E-06	4.48	0.22
HSPBP1	NP_077134.1	hsp70-interacting protein	63	1	3	103	2	8	2.32E-04 ± -	1.17E-04 ± -	1.98	0.51
HMMR	NP_038580.1	hyaluronan mediated motility receptor (RHAMM)	140	2	3	129	2	3	1.15E-04 ± 1.47E-05	1.23E-04 ± 1.23E-04	0.93	1.07
-	NP_082497.2	hypothetical protein LOC102122	93	1	6	68	1	6	1.30E-04 ± 0.00E+00	1.65E-04 ± -	0.79	1.26
-	NP_955518.1	hypothetical protein LOC107094	135	2	4	132	2	2	7.03E-05 ± 4.52E-05	2.16E-05 ± 0.00E+00	3.26	0.31
-	NP_659095.1	hypothetical protein LOC223601	79	1	4	65	1	4	5.11E-05 ± 0.00E+00	4.31E-05 ± 0.00E+00	1.19	0.84
-	NP_663471.1	hypothetical protein LOC225913	85	1	4	102	1	4	8.59E-05 ± -	8.45E-05 ± 8.54E-05	1.02	0.98
-	NP_780706.1	hypothetical protein LOC238880	349	7	25	382	8	28	2.11E-03 ± 9.03E-04	3.44E-03 ± 4.99E-04	0.61	1.63
-	NP_898911.1	hypothetical protein LOC71970	105	1	4	173	2	4	1.70E-04 ± 1.20E-04	8.38E-05 ± 1.69E-05	2.03	0.49
HPRT1	NP_038584.1	hypoxanthine guanine phosphoribosyl transferase 1	158	3	15	137	2	11	1.52E-04 ± 1.07E-04	2.56E-04 ± -	0.59	1.69
IPO7	NP_852658.2	importin 7	179	4	5	383	6	9	3.99E-05 ± 1.13E-05	1.08E-04 ± -	0.37	2.70
2 (includes EG	NP_898850.1	insulin-like growth factor 2 mRNA binding protein 2	105	3	8	76	1	3	7.59E-05 ± 6.44E-05	2.05E-04 ± 2.56E-05	0.37	2.70
IGF2BP3	NP_076159.2	insulin-like growth factor 2, binding protein 3	144	3	9	80	1	3	1.05E-04 ± 1.08E-04	8.04E-05 ± 2.78E-05	1.30	0.77
ILF3	NP_034691.1	interleukin enhancer binding factor 3	88	1	2	138	2	5	1.03E-04 ± 6.88E-05	3.83E-05 ± 3.25E-05	2.69	0.37
IDI1	NP_808875.1	isopentenyl-diphosphate delta isomerase isoform 2	82	1	5	112	1	9	3.65E-04 ± -	6.15E-05 ± -	5.93	0.17
KPNA3	NP_032492.1	karyopherin (importin) alpha 3	159	2	9	165	3	10	9.53E-05 ± -	5.36E-05 ± 3.79E-05	1.78	0.56

Proteins in common with embryonic stem cells and embryoid bodies, continued.

Gene	NCBI Accession	Protein Description	Embryonic stem cells: Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Embryoid Bodies: Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Embryonic stem cells: AVG N_SC ± STDEV	Embryoid Bodies: AVG N_SC ± STDEV	ES/EB Ratio	EB/ES Ratio
KTN1	NP_032503.1	kinectin 1	134	3	3	126	2	2	3.74E-05 ± 1.76E-05	4.73E-05 ± 3.72E-05	0.79	1.26
KIF20A	NP_033030.1	kinesin family member 20A	61	1	2	75	1	2	1.87E-05 ± -	2.36E-05 ± 1.11E-05	0.79	1.26
KIF5C	NP_032475.2	kinesin family member 5C	66	2	3	124	2	3	2.84E-03 ± -	2.63E-04 ± -	10.80	0.09
KIF15	NP_034750.1	kinesin family member 15	161	2	2	57	1	1	2.39E-05 ± 1.69E-05	1.01E-05 ± -	2.37	0.42
LACTB2	NP_663356.1	lactamase, beta 2	62	1	6	77	1	6	5.75E-05 ± -	1.13E-04 ± 7.40E-05	0.51	1.97
LDHB	NP_032518.1	lactate dehydrogenase 2, B chain	195	3	13	395	6	18	6.44E-04 ± 6.31E-04	7.10E-04 ± 1.11E-04	0.91	1.10
LBR	NP_598576.1	lamin B receptor	61	1	3	110	2	4	2.64E-05 ± -	4.46E-05 ± 3.15E-05	0.59	1.69
LRRC40	NP_077156.2	leucine rich repeat containing 40	80	1	4	76	1	2	8.25E-05 ± -	2.32E-05 ± 0.00E+00	3.56	0.28
LZTFL1	NP_201579.1	leucine zipper transcription factor-like 1	104	1	5	97	1	5	1.11E-04 ± -	1.87E-04 ± -	0.59	1.69
LIG1	NP_034845.1	ligase I, DNA, ATP-dependent	59	1	1	150	2	3	1.81E-05 ± -	6.86E-05 ± 7.54E-05	0.26	3.79
TDH	NP_067455.4	L-threonine dehydrogenase	171	2	12	278	3	17	2.07E-04 ± 2.05E-04	5.36E-04 ± 7.79E-05	0.39	2.59
LYAR	NP_079557.1	Ly1 antibody reactive clone	122	2	5	83	2	5	4.55E-04 ± 2.71E-04	2.52E-04 ± 3.05E-04	1.81	0.55
MDH1	NP_032644.2	malate dehydrogenase 1, NAD (soluble)	218	4	14	56	1	4	4.79E-04 ± 1.03E-04	8.36E-05 ± -	5.73	0.17
MARCKSL1	NP_034937.1	MARCKS-like protein	181	3	39	88	1	17	1.90E-03 ± 5.85E-04	1.63E-04 ± 1.07E-04	11.69	0.09
(includes EG:2	NP_001008705.1	BUD31 homolog (yeast)	76	1	18	79	1	18	4.82E-04 ± -	3.39E-04 ± 2.87E-04	1.42	0.70
MARS	NP_001003913.1	methionine-tRNA synthetase	287	4	9	347	6	13	3.79E-04 ± 2.12E-05	9.80E-05 ± 4.97E-05	3.87	0.26
MTHFD1L	NP_758512.2	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1-like	160	4	5	141	3	4	1.44E-04 ± 3.59E-05	8.57E-05 ± 1.43E-05	1.68	0.60
MAP1B	NP_032660.1	microtubule-associated protein 1 B	79	1	1	118	2	1	6.72E-06 ± -	2.27E-05 ± -	0.30	3.37
MCM5	NP_032592.1	minichromosome maintenance deficient 5, cell division cycle 46	303	6	11	221	5	11	1.65E-04 ± 1.92E-04	1.90E-04 ± 8.07E-05	0.87	1.15
MCM4	NP_032591.2	minichromosome maintenance protein 4	364	8	12	338	6	8	1.66E-04 ± 6.17E-05	1.46E-04 ± 7.42E-05	1.14	0.88
MCM6	NP_032593.1	minichromosome maintenance protein 6	283	5	9	300	5	11	4.13E-04 ± 4.99E-04	1.76E-04 ± 1.57E-04	2.35	0.43
MCM7	NP_032594.1	minichromosome maintenance protein 7	327	5	8	461	7	13	1.69E-04 ± 1.16E-04	4.92E-04 ± 2.32E-04	0.34	2.91
MRPS2	NP_536700.2	mitochondrial ribosomal protein S2	71	1	7	84	1	7	8.53E-05 ± 4.02E-05	4.80E-05 ± -	1.78	0.56
MAPK1	NP_036079.1	mitogen activated protein kinase 1	209	4	17	198	4	18	2.93E-04 ± 1.62E-04	1.82E-04 ± 9.00E-05	1.61	0.62
MAPK3	NP_036082.1	mitogen activated protein kinase 3	84	1	4	128	2	12	2.03E-04 ± 2.06E-04	1.10E-04 ± 1.04E-04	1.84	0.54
MSN	NP_034963.2	moesin	247	5	8	170	3	7	1.05E-04 ± 1.09E-04	2.30E-04 ± 1.71E-05	0.46	2.18
MSH2	NP_032654.1	mutS homolog 2	182	2	4	216	2	4	1.95E-04 ± -	1.34E-04 ± -	1.45	0.69
MYBBP1A	NP_058056.2	MYB binding protein (P160) 1a	1094	15	16	1294	22	21	4.43E-04 ± 3.63E-04	8.17E-04 ± 1.59E-04	0.54	1.84
MYH11	NP_038635.1	myosin, heavy polypeptide 11, smooth muscle	178	5	3	86	2	1	1.54E-04 ± -	6.48E-05 ± -	2.37	0.42
MYST2	NP_808287.1	MYST histone acetyltransferase 2	89	1	3	125	2	7	6.34E-05 ± -	5.35E-05 ± 3.78E-05	1.19	0.84
NDUFA4	NP_035016.1	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4	76	1	15	73	1	15	1.21E-03 ± 5.34E-04	5.11E-04 ± 2.41E-04	2.37	0.42
NSF	NP_032766.2	N-ethylmaleimide sensitive fusion protein	52	1	2	52	1	2	4.45E-05 ± -	3.75E-05 ± -	1.19	0.84
NEDD4	NP_035020.2	neural precursor cell expressed, developmentally down-regulated gene 4	374	6	10	252	4	7	1.21E-04 ± 9.24E-05	9.97E-05 ± 4.54E-05	1.22	0.82
UBAP2L	NP_705693.1	Nice-4 protein homolog isoform 2	205	3	5	158	2	4	1.09E-04 ± -	3.92E-05 ± -	2.77	0.36
NONO	NP_075633.1	non-POU-domain-containing, octamer binding protein	200	4	15	187	3	7	4.20E-04 ± 3.05E-04	5.90E-04 ± 2.30E-04	0.71	1.41
SYNCRIP	NP_062770.1	NS1-associated protein 1 isoform 2	470	8	23	163	2	7	3.73E-04 ± 3.51E-04	1.49E-04 ± 8.60E-05	2.50	0.40
NUDC	NP_035078.1	nuclear distribution gene C homolog	175	5	15	226	4	13	4.74E-04 ± 3.53E-05	6.45E-04 ± 1.28E-04	0.73	1.36
NVL	NP_080447.1	nuclear VCP-like	296	5	9	211	4	5	1.03E-04 ± 9.55E-05	2.18E-04 ± 6.18E-05	0.47	2.11
YBX1	NP_035862.1	nuclease sensitive element binding protein 1	213	2	11	211	2	11	7.37E-03 ± 2.06E-03	3.77E-03 ± 2.05E-04	1.96	0.51
NOL1	NP_620086.1	nucleolar protein 1	199	4	5	73	2	3	9.39E-05 ± 4.43E-05	1.06E-04 ± -	0.89	1.12
NOP5/NOP58	NP_061356.1	nucleolar protein 5	182	3	12	231	4	12	2.62E-04 ± 2.47E-05	6.89E-05 ± 6.81E-05	3.81	0.26
NOLA2	NP_080907.1	nucleolar protein family A, member 2	73	1	10	70	1	10	4.33E-04 ± 2.16E-04	2.74E-04 ± -	1.58	0.63
NUP155	NP_573490.2	nucleoporin 155	87	2	2	163	4	4	4.76E-05 ± -	1.27E-04 ± 4.18E-05	0.37	2.67
NUP62	NP_444304.1	nucleoporin 62	142	2	5	138	2	5	3.78E-04 ± 3.29E-04	1.77E-04 ± 8.53E-05	2.13	0.47
NUP93	NP_765998.1	nucleoporin 93	159	3	4	195	3	6	1.41E-04 ± 0.00E+00	6.82E-05 ± -	2.08	0.48
NAP1L1	NP_056596.1	nucleosome assembly protein 1-like 1	512	6	20	218	3	12	8.47E-04 ± 2.36E-04	4.52E-04 ± 2.68E-04	1.87	0.53
NAP1L4	NP_032698.1	nucleosome assembly protein 1-like 4	107	2	10	83	1	5	1.32E-04 ± -	1.12E-04 ± -	1.19	0.84
PDAP1	NP_001028485.1	PDGFA associated protein 1	143	2	16	72	1	7	9.75E-04 ± 7.99E-04	3.86E-04 ± 2.04E-04	2.53	0.40
PPWD1	NP_766395.1	peptidylprolyl isomerase domain and WD repeat containing 1	85	1	3	63	1	3	6.41E-05 ± 1.81E-05	4.32E-05 ± -	1.48	0.67
PRDX2	NP_035693.2	peroxiredoxin 2	88	1	9	112	2	17	1.67E-04 ± -	2.11E-04 ± 9.97E-05	0.79	1.26
PFKL	NP_032852.2	phosphofructokinase, liver, B-type	124	2	3	291	5	10	7.43E-05 ± 1.50E-05	7.75E-05 ± 7.45E-05	0.96	1.04
PFKP	NP_062677.1	phosphofructokinase, platelet	130	2	4	60	1	2	1.76E-04 ± 1.22E-05	1.60E-04 ± -	1.10	0.91
PGK2	NP_112467.1	phosphoglycerate kinase 2	124	2	6	157	3	7	2.58E-04 ± 2.81E-05	2.34E-04 ± -	1.10	0.91
PLAA	NP_766283.1	phospholipase A2, activating protein	74	1	3	57	1	2	2.08E-05 ± -	1.76E-05 ± -	1.19	0.84
PSAT1	NP_803155.1	phosphoserine aminotransferase 1	176	2	7	77	1	3	4.03E-04 ± -	1.51E-04 ± 0.00E+00	2.67	0.37
FUS	NP_631888.1	pigpen	119	2	3	149	2	3	2.40E-04 ± 6.78E-05	4.67E-04 ± 5.27E-04	0.51	1.95
PARP1	NP_031441.2	poly (ADP-ribose) polymerase family, member 1	797	14	18	615	13	19	6.69E-04 ± 1.02E-04	4.82E-04 ± 3.35E-04	1.39	0.72
PABPC4	NP_570951.2	poly(A) binding protein, cytoplasmic 4 isoform 1	431	7	16	232	3	7	7.27E-04 ± 1.57E-04	2.64E-04 ± 2.54E-04	2.75	0.36
POLR2A	NP_033115.1	polymerase (RNA) II (DNA directed) polypeptide A	324	5	5	193	3	3	6.00E-05 ± 2.42E-05	2.89E-05 ± 7.22E-06	2.08	0.48
-	XP_001002647.1	PREDICTED: hypothetical protein	85	3	7	83	2	8	7.94E-04 ± -	1.03E-04 ± -	7.71	0.13
-	NP_001038994	PREDICTED: microrchidia 3 isoform 1	144	2	4	175	3	5	4.10E-05 ± 4.06E-05	4.45E-05 ± 0.00E+00	0.92	1.08
-	XP_133073.3	PREDICTED: nucleoporin 205 isoform 1	133	3	2	246	6	4	2.01E-05 ± 5.68E-06	2.03E-05 ± 0.00E+00	0.99	1.01
-	XP_622094.1	PREDICTED: phosphogluconate dehydrogenase isoform 1	99	1	4	82	1	4	3.43E-05 ± -	2.89E-05 ± -	1.19	0.84
-	XP_001002609.1	PREDICTED: scaffold attachment factor B	95	1	2	135	2	2	3.34E-05 ± -	2.11E-05 ± 9.96E-06	1.58	0.63
-	XP_907394.1	PREDICTED: similar to 40S ribosomal protein S2 isoform 1	102	3	10	182	4	14	2.49E-04 ± 1.45E-04	1.13E-04 ± 2.80E-05	2.20	0.45

Proteins in common with embryonic stem cells and embryoid bodies, continued.

Gene	NCBI Accession	Protein Description	Embryonic stem cells: Total Mascot Score	Embryonic stem cells: Total No. Peptides	Embryonic stem cells: Total % Protein Coverage	Embryoid Bodies: Total Mascot Score	Embryoid Bodies: Total No. Peptides	Embryoid Bodies: Total % Protein Coverage	Embryonic stem cells: AVG N_SC ± STDEV	Embryoid Bodies: AVG N_SC ± STDEV	ES/EB Ratio	EB/ES Ratio
-	XP_124146.3	PREDICTED: similar to 40S ribosomal protein S4, X isoform	89	3	13	109	3	17	3.72E-04 ± 4.69E-04	2.14E-04 ± 6.06E-05	1.74	0.57
-	XP_001005850.1	PREDICTED: similar to 60S ribosomal protein L12	222	3	51	221	3	51	1.20E-03 ± 7.19E-04	2.13E-03 ± 1.56E-03	0.56	1.78
-	XP_982298.1	PREDICTED: similar to 60S ribosomal protein L3 (J1 protein)	469	9	28	153	3	11	5.73E-04 ± 3.17E-04	1.27E-04 ± 1.28E-04	4.52	0.22
-	XP_994673.1	PREDICTED: similar to Activator 1 38 kDa subunit	190	3	9	110	1	5	1.30E-04 ± 1.23E-04	2.20E-04 ± -	0.59	1.69
-	XP_986918.1	PREDICTED: similar to Annexin A11 (Annexin XI) (Calcyclin-associated annexin 50) (CAP-50)	100	1	3	125	2	5	3.29E-05 ± 0.00E+00	2.77E-05 ± -	1.19	0.84
-	XP_001000692.1	PREDICTED: similar to Antigen KI-67	202	5	2	73	1	1	1.69E-05 ± -	7.13E-06 ± 3.36E-06	2.37	0.42
-	XP_923115.2	PREDICTED: similar to Aspartate aminotransferase, mitochondrial precursor	413	5	18	219	3	12	4.75E-04 ± 1.94E-04	1.95E-04 ± 5.62E-05	2.44	0.41
-	NP_080100	PREDICTED: similar to basic leucine zipper and W2 domains 1	59	1	3	86	2	5	3.95E-05 ± -	4.00E-04 ± -	0.10	10.12
-	XP_913237.2	PREDICTED: similar to Bifunctional aminoacyl-tRNA synthetase	669	12	11	303	4	4	2.10E-04 ± 3.95E-05	1.66E-04 ± 7.39E-05	1.26	0.79
-	XP_980555.1	PREDICTED: similar to chromosome condensation protein G isoform 2	89	1	1	359	6	7	1.65E-05 ± -	9.73E-05 ± 7.22E-05	0.17	5.90
-	NP_082996	PREDICTED: similar to cytokine-like nuclear factor n-pac	60	1	3	81	1	3	3.03E-05 ± -	6.82E-05 ± 7.38E-05	0.44	2.25
-	NP_081208	PREDICTED: similar to EBNA1 binding protein 2	79	1	10	71	1	10	5.41E-05 ± 0.00E+00	4.56E-05 ± -	1.19	0.84
-	XP_996339.1	PREDICTED: similar to General transcription factor II-I	205	3	7	195	4	10	8.89E-05 ± 3.56E-05	1.80E-04 ± -	0.49	2.02
-	XP_001003154.1	PREDICTED: similar to Glucose-6-phosphate isomerase isoform 2	81	2	7	215	3	12	1.94E-04 ± -	5.12E-04 ± 3.18E-04	0.38	2.64
-	XP_484732.3	PREDICTED: similar to Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	406	6	24	491	8	28	5.07E-03 ± 1.37E-03	6.06E-03 ± 1.29E-03	0.84	1.19
-	XP_914685.2	PREDICTED: similar to Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	216	4	18	111	2	15	3.11E-04 ± 8.90E-05	6.38E-04 ± 7.30E-04	0.49	2.05
-	XP_990390.1	PREDICTED: similar to Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	423	6	24	321	6	26	5.08E-03 ± 1.37E-03	5.54E-03 ± 7.47E-04	0.92	1.09
-	XP_899549.1	PREDICTED: similar to Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	270	5	23	151	4	20	1.19E-03 ± 8.57E-04	1.12E-03 ± 5.22E-02	1.38	0.95
-	XP_908892.1	PREDICTED: similar to H/ACA ribonucleoprotein complex subunit 4 (Dyskerin) 7	171	3	10	119	2	8	2.17E-04 ± -	9.15E-05 ± 1.85E-05	2.37	0.42
-	XP_918522.1	PREDICTED: similar to Leucyl-tRNA synthetase, cytoplasmic (Leucine--tRNA ligase) (LeuRS) isoform 8	403	7	11	324	6	8	2.11E-04 ± 6.44E-05	1.07E-04 ± 1.18E-05	1.98	0.51
-	XP_909898.1	PREDICTED: similar to lysyl-tRNA synthetase isoform 1	295	5	12	100	2	6	3.48E-04 ± 1.97E-05	4.69E-05 ± -	7.41	0.13
-	XP_996938.1	PREDICTED: similar to microfilament and actin filament cross-linker protein isoform a isoform 7	488	12	3	565	8	2	4.38E-05 ± 9.81E-06	4.98E-05 ± 1.16E-05	0.88	1.14
-	XP_980444.1	PREDICTED: similar to NAD-dependent deacetylase sirtuin-1	151	2	6	148	2	6	8.82E-05 ± 1.78E-05	7.08E-05 ± 4.42E-05	1.25	0.80
-	XP_979797.1	PREDICTED: similar to nascent polypeptide-associated complex alpha polypeptide	263	3	19	337	5	25	5.66E-04 ± 3.08E-04	1.95E-03 ± 6.36E-04	0.29	3.45
-	XP_922162.1	PREDICTED: similar to Nucleophosmin	119	3	25	167	4	26	1.55E-03 ± 8.61E-04	2.10E-03 ± 8.36E-04	0.74	1.35
-	XP_921493.1	PREDICTED: similar to peptidase (prosome, macropain) 26S subunit, ATPase 1	200	3	9	207	3	9	1.69E-04 ± 2.25E-05	1.42E-04 ± 9.48E-05	1.19	0.84
-	XP_979315.1	PREDICTED: similar to PHD finger protein 3 isoform 9	100	3	2	135	3	2	4.09E-05 ± -	4.14E-05 ± -	0.99	1.01
-	XP_001005759.1	PREDICTED: similar to prefoldin subunit 4	121	2	8	148	2	8	2.43E-04 ± 6.88E-05	3.28E-04 ± 2.17E-04	0.74	1.35
-	XP_992759.1	PREDICTED: similar to Protein diaphanous homolog 1	95	1	1	57	1	1	2.17E-05 ± 7.50E-06	3.29E-05 ± 1.55E-05	0.66	1.52
-	XP_923088.1	PREDICTED: similar to Putative eukaryotic translation initiation factor 3 subunit (eIF-3) isoform 7	110	1	2	59	1	2	3.66E-05 ± -	1.03E-05 ± -	3.56	0.28
-	XP_979289.1	PREDICTED: similar to Putative RNA-binding protein 3 (RNA-binding motif protein 3) isoform 2	114	2	30	117	2	30	2.49E-03 ± 4.59E-04	6.99E-04 ± 3.80E-04	3.56	0.28
-	NP_038744.1	PREDICTED: similar to Ras-GTPase-activating protein binding protein 1	138	2	5	123	2	5	3.20E-04 ± 4.03E-04	6.00E-05 ± 3.00E-05	5.34	0.19
-	XP_996529.1	PREDICTED: similar to RCC1-like isoform 3	167	5	13	198	4	12	5.58E-04 ± 7.63E-05	2.63E-04 ± 1.53E-04	2.12	0.47
-	NP_067438	PREDICTED: similar to Ribose-phosphate pyrophosphokinase I	62	1	5	85	1	5	1.04E-04 ± -	3.22E-04 ± 1.27E-04	0.32	3.09
-	XP_996461.1	PREDICTED: similar to ribosomal protein L31	211	3	21	145	2	15	2.90E-03 ± 1.18E-03	3.07E-03 ± 1.45E-03	0.94	1.06
-	XP_978248.1	PREDICTED: similar to ribosomal protein L38	301	4	36	145	2	33	4.18E-03 ± 3.62E-03	2.53E-03 ± 2.09E-03	1.65	0.60
-	XP_001004633.1	PREDICTED: similar to ribosomal protein S27a	352	8	53	334	6	38	7.57E-03 ± 6.22E-04	6.77E-03 ± 1.18E-03	1.12	0.89
-	XP_915304.2	PREDICTED: similar to ribosomal protein S27a	105	2	14	95	2	14	4.18E-04 ± 1.45E-04	2.82E-04 ± 6.10E-05	1.48	0.67
-	XP_930062.1	PREDICTED: similar to ribosomal protein S8 isoform 2	326	7	37	293	6	30	2.94E-03 ± 6.21E-04	1.95E-03 ± 2.01E-04	1.51	0.66
-	XP_919775.1	PREDICTED: similar to SCC-112 protein isoform 9	100	1	2	100	1	2	3.73E-05 ± -	2.79E-05 ± 6.05E-06	1.33	0.75
-	XP_891995.1	PREDICTED: similar to Splicing factor, arginine/serine-rich 5	77	1	6	89	1	4	2.46E-04 ± 2.61E-04	2.25E-04 ± 2.16E-04	1.09	0.91
-	XP_998318.1	PREDICTED: similar to Stress-70 protein, mitochondrial precursor (Mortalin)	263	4	13	258	4	13	4.92E-04 ± 3.25E-04	9.33E-04 ± 3.40E-04	0.53	1.90
-	XP_912105.2	PREDICTED: similar to Stretchin-Mlck CG18255-PA, isoform A	77	2	1	208	7	2	7.15E-05 ± -	6.03E-05 ± -	1.19	0.84
-	XP_981884.1	PREDICTED: similar to Testis expressed sequence 10 protein	184	4	6	68	1	2	7.55E-05 ± -	1.59E-05 ± -	4.74	0.21
-	XP_999265.1	PREDICTED: similar to thyroid hormone receptor associated protein 3 isoform 3	70	2	3	52	1	2	5.22E-05 ± -	2.94E-05 ± -	1.78	0.56
-	XP_992738.1	PREDICTED: similar to Transcriptional activator protein Pur-alpha isoform 1	331	4	27	152	2	16	3.18E-04 ± 1.41E-04	3.31E-04 ± 6.69E-05	0.96	1.04
PFDN2	NP_035200.2	prefoldin 2	72	1	8	51	1	8	2.15E-04 ± -	9.06E-05 ± -	2.37	0.42
PRPF8	NP_619600.1	pre-mRNA processing factor 8	211	5	3	155	3	2	6.73E-05 ± 2.51E-05	5.38E-05 ± -	1.25	0.80
PGRMC1	NP_058063.2	progesterone receptor membrane component	69	1	7	86	1	10	2.55E-04 ± -	3.58E-04 ± 2.02E-04	0.71	1.41
PCNA	NP_035175.1	proliferating cell nuclear antigen	142	2	14	135	2	14	4.44E-04 ± 3.53E-04	2.67E-04 ± 1.07E-04	1.66	0.60
PELP1	NP_083507.2	proline-, glutamic acid-, leucine-rich protein 1	142	2	4	195	2	4	1.03E-04 ± -	1.04E-04 ± 8.73E-05	1.00	1.00
PML	NP_032910.2	promyelocytic leukemia isoform 1	145	2	4	205	2	4	4.10E-05 ± -	9.50E-05 ± 1.22E-05	0.43	2.32
PTGES3	NP_062740.1	prostaglandin E synthase 3 (cytosolic)	316	5	41	473	7	55	1.31E-03 ± 3.16E-04	3.05E-03 ± 1.32E-03	0.43	2.33
PSMA2	NP_032970.1	proteasome (prosome, macropain) subunit, alpha type 2	187	4	20	129	2	9	3.18E-04 ± 3.50E-04	5.07E-04 ± 4.22E-05	0.63	1.59
PSMC4	NP_036004.1	proteasome 26S ATPase subunit 4	205	3	13	270	4	21	5.02E-04 ± 3.55E-04	2.67E-04 ± 1.20E-04	1.88	0.53
PSMD12	NP_080170.1	proteasome 26S non-ATPase subunit 12	195	4	11	205	4	9	9.68E-04 ± 2.58E-04	8.37E-04 ± 3.39E-04	1.16	0.86
PSMD3	NP_033465.1	proteasome 26S non-ATPase subunit 3	116	1	3	242	5	12	1.87E-04 ± -	1.71E-04 ± 1.68E-04	1.09	0.91
PSMD4	NP_032977.1	proteasome 26S non-ATPase subunit 4	86	1	7	89	1	5	3.52E-04 ± 7.62E-05	1.86E-04 ± 3.71E-05	1.90	0.53
PSME3	NP_035322.1	proteasome activator subunit 3	107	1	6	170	3	13	1.95E-04 ± 1.30E-04	3.30E-04 ± 1.55E-04	0.59	1.69
PSMB3	NP_036101.1	proteasome beta 3 subunit	58	1	8	108	2	17	1.61E-04 ± 8.07E-05	1.45E-03 ± 9.85E-04	0.11	9.00
PSMD6	NP_079826.2	proteasome, 26S, non-ATPase regulatory subunit 6	196	3	10	233	4	14	8.51E-05 ± 0.00E+00	1.44E-04 ± 1.01E-04	0.59	1.69
PRMT5	NP_038796.1	protein arginine N-methyltransferase 5	68	1	3	204	5	10	1.04E-04 ± -	2.12E-04 ± 1.10E-04	0.49	2.04
PDIA4	NP_033917.2	protein disulfide isomerase associated 4	227	5	10	212	4	10	1.81E-04 ± 1.81E-04	2.72E-04 ± 1.54E-05	0.66	1.51
PRKCSH	NP_032951.1	protein kinase C substrate 80K-H	65	1	2	68	1	5	4.77E-05 ± 2.25E-05	2.68E-05 ± -	1.78	0.56
PPP1CA	NP_114074.1	protein phosphatase 1, catalytic subunit, alpha	234	5	13	317	6	25	3.18E-04 ± 2.03E-04	4.23E-04 ± 2.24E-04	0.75	1.33

Proteins in common with embryonic stem cells and embryoid bodies, continued.

Gene	NCBI Accession	Protein Description	Embryonic stem cells: Total Mascot Score	Embryonic stem cells: Total No. Peptides	Embryonic stem cells: Total % Protein Coverage	Embryoid Bodies: Total Mascot Score	Embryoid Bodies: Total No. Peptides	Embryoid Bodies: Total % Protein Coverage	Embryonic stem cells: AVG N_SC ± STDEV	Embryoid Bodies: AVG N_SC ± STDEV	ES/EB Ratio	EB/ES Ratio
PPP1CC	NP_038664.2	protein phosphatase 1, catalytic subunit, gamma isoform	189	3	13	348	6	27	3.07E-04 ± 2.66E-04	4.75E-04 ± 1.88E-04	0.65	1.55
B (includes EG	NP_082890.2	protein phosphatase 2 (formerly 2A), regulatory subunit A (PR 65), beta isoform isoform b	150	3	7	190	2	6	1.10E-04 ± -	1.32E-04 ± 1.05E-04	0.84	1.19
PPP2CA	NP_062284.1	protein phosphatase 2a, catalytic subunit, alpha isoform	74	1	4	69	1	4	7.14E-05 ± 3.09E-05	1.81E-04 ± 9.03E-05	0.40	2.53
PTPN6	NP_038573.1	protein tyrosine phosphatase, non-receptor type 6	60	1	3	52	1	3	6.95E-05 ± 5.90E-05	2.35E-05 ± -	2.96	0.34
PPA1	NP_080714.2	pyrophosphatase	224	3	18	253	4	20	5.54E-04 ± 3.31E-05	3.54E-04 ± 2.75E-04	1.56	0.64
RAB1A	NP_033022.1	RAB1, member RAS oncogene family	74	1	8	67	1	8	1.61E-04 ± -	2.04E-04 ± -	0.79	1.26
RAB18	NP_035355.1	RAB18, member RAS oncogene family	55	1	7	51	1	7	1.61E-04 ± 1.14E-04	1.36E-04 ± -	1.19	0.84
RAD23A	NP_033036.2	RAD23a homolog	69	2	10	115	1	7	1.83E-04 ± -	1.41E-04 ± 5.89E-05	1.29	0.77
RAD23B	NP_033037.1	RAD23b homolog	349	5	24	142	1	11	4.11E-04 ± 6.08E-05	1.68E-04 ± 6.71E-05	2.45	0.41
RANBP5	NP_076068.1	RAN binding protein 5	712	11	14	466	9	10	3.27E-04 ± 1.28E-04	2.46E-04 ± 6.01E-05	1.33	0.75
RANGAP1	NP_035371.3	RAN GTPase activating protein 1	255	6	15	284	5	12	3.00E-04 ± 3.16E-04	3.08E-04 ± 1.85E-04	0.97	1.03
RAN	NP_033417.1	RAN, member RAS oncogene family	374	7	40	536	8	43	1.38E-03 ± 6.55E-04	3.51E-03 ± 2.74E-03	0.39	2.55
IPO4	NP_077229.4	Importin 4	130	3	3	54	1	1	1.07E-04 ± -	2.58E-05 ± -	4.15	0.24
G3BP2	NP_035946.2	Ras-GTPase-activating protein (GAP120) SH3-domain binding protein 2	100	1	4	78	1	4	2.06E-04 ± -	5.79E-05 ± -	3.56	0.28
RAVER1	NP_082187.1	Raver1	106	2	4	98	2	4	1.55E-04 ± -	3.73E-05 ± -	4.15	0.24
UPF1	NP_109605.1	regulator of nonsense transcripts 1	82	2	2	106	3	4	1.19E-04 ± -	6.27E-05 ± -	1.90	0.53
RBBP4	NP_033056.2	retinoblastoma binding protein 4	168	3	6	108	2	5	5.58E-04 ± 6.21E-04	1.97E-04 ± 3.28E-05	2.83	0.35
RRM1	NP_033129.2	ribonucleotide reductase M1	84	1	2	280	5	10	6.27E-05 ± -	9.40E-05 ± 3.67E-05	0.67	1.50
RRM2	NP_033130.1	ribonucleotide reductase M2	57	1	4	128	2	6	8.49E-05 ± -	5.61E-04 ± 2.87E-04	0.15	6.61
RSL1D1	NP_079822.1	ribosomal L1 domain containing 1	212	3	11	170	2	7	8.54E-05 ± 4.23E-05	1.44E-04 ± 1.78E-05	0.59	1.69
RPL13	NP_058018.2	ribosomal protein L13	286	7	27	191	5	21	1.88E-03 ± 5.66E-04	5.62E-04 ± 4.68E-05	3.35	0.30
RPL13A	NP_033464.2	ribosomal protein L13a	194	5	20	229	6	26	5.00E-03 ± 2.53E-03	4.35E-03 ± 2.24E-03	1.15	0.87
RPL15	NP_079862.1	ribosomal protein L15	202	5	28	266	6	35	1.83E-03 ± 9.75E-04	2.26E-03 ± 2.61E-03	0.81	1.24
RPL32	NP_742083.1	ribosomal protein L32	208	6	35	169	4	27	5.31E-04 ± 1.42E-04	1.03E-04 ± 0.00E+00	5.14	0.19
RPS14	NP_065625.2	ribosomal protein S14	404	5	36	248	4	29	2.23E-03 ± 7.46E-04	9.24E-04 ± 5.15E-04	2.41	0.41
RPS19	NP_075622.1	ribosomal protein S19	274	6	34	197	5	27	2.51E-03 ± 6.04E-04	1.01E-03 ± 8.85E-04	2.48	0.40
RPS23	NP_077137.1	ribosomal protein S23	346	8	60	213	4	38	3.13E-03 ± 1.96E-03	1.40E-03 ± 5.38E-04	2.23	0.45
RPS4X	NP_033120.1	ribosomal protein S4, X-linked	640	12	39	586	11	41	1.89E-03 ± 4.54E-04	1.73E-03 ± 3.24E-04	1.09	0.92
RPS9	NP_084043.1	ribosomal protein S9-like	403	9	34	332	9	33	1.05E-03 ± 7.93E-04	7.19E-04 ± 3.74E-04	1.46	0.68
RBM14	NP_063922.2	RNA binding motif protein 14	152	3	5	96	1	2	2.47E-05 ± -	4.17E-05 ± -	0.59	1.69
RPAP1	NP_796268.2	RNA polymerase II associated protein 1	155	2	3	84	1	1	1.17E-05 ± 0.00E+00	8.92E-05 ± 0.00E+00	0.13	7.59
RUVBL1	NP_062659.1	RuvB-like protein 1	233	4	12	395	5	17	5.08E-04 ± -	3.26E-04 ± 1.69E-04	1.56	0.64
AHCY	NP_057870.2	S-adenosylhomocysteine hydrolase	275	4	12	192	3	12	2.87E-04 ± 2.71E-05	2.26E-04 ± 9.14E-05	1.27	0.79
SEC23B	NP_062761.2	SEC23B	90	1	2	171	2	4	6.47E-05 ± -	1.46E-04 ± 3.64E-05	0.44	2.25
SHMT1	NP_033197.2	serine hydroxymethyl transferase 1 (soluble)	69	1	3	316	5	15	1.04E-04 ± 0.00E+00	3.70E-04 ± 2.15E-04	0.28	3.56
SHMT2	NP_082506.1	serine hydroxymethyl transferase 2 (mitochondrial)	312	4	12	372	4	13	2.74E-04 ± 1.33E-04	4.62E-04 ± 2.91E-04	0.59	1.69
SERBP1	NP_080090.1	SERPINE1 mRNA binding protein 1	254	3	12	179	2	8	1.76E-04 ± 1.43E-04	1.03E-04 ± 9.07E-05	1.71	0.58
SARS	NP_035449.1	seryl-aminoacyl-tRNA synthetase 1	121	2	3	93	2	5	1.03E-04 ± 1.78E-05	1.04E-04 ± -	0.99	1.01
SRP68	NP_666144.3	signal recognition particle 68	80	1	2	151	3	6	7.94E-05 ± 2.65E-05	8.93E-05 ± -	0.89	1.12
SNRPD2	NP_081219.1	small nuclear ribonucleoprotein D2	291	5	47	145	3	24	2.31E-03 ± 2.68E-03	2.60E-03 ± -	0.89	1.12
SNRPN	NP_038698.1	small nuclear ribonucleoprotein N	70	2	7	52	1	4	2.07E-04 ± -	5.82E-05 ± -	3.56	0.28
SMU1	NP_067510.2	smu-1 suppressor of mec-8 and unc-52 homolog	118	2	5	61	1	2	3.23E-05 ± 0.00E+00	4.53E-05 ± 1.57E-05	0.71	1.41
LSM2	NP_085100.1	snRNP core protein SMX5	105	1	15	116	1	15	2.53E-04 ± 1.26E-04	4.62E-04 ± 6.15E-05	0.55	1.83
SLC2A1	NP_035530.1	solute carrier family 2 (facilitated glucose transporter), member 1	90	1	3	133	2	5	5.61E-05 ± 1.94E-05	9.93E-05 ± 1.00E-04	0.56	1.77
SLC25A3	NP_598429.1	solute carrier family 25 (mitochondrial carrier, phosphate carrier), member 3	58	1	8	191	4	17	3.71E-04 ± -	1.37E-04 ± 1.38E-04	2.71	0.37
SNX3	NP_059500.2	sorting nexin 3	52	1	9	70	1	9	2.04E-04 ± -	1.72E-04 ± 0.00E+00	1.19	0.84
SF3B1	NP_112456.1	splicing factor 3b, subunit 1	176	6	6	179	4	4	6.34E-05 ± -	5.35E-05 ± -	1.19	0.84
SF3B3	NP_598714.1	splicing factor 3b, subunit 3	91	1	1	234	3	4	7.25E-05 ± 2.08E-05	5.73E-05 ± 0.00E+00	1.26	0.79
SFRS1	NP_775550.2	splicing factor, arginine/serine-rich 1 (ASF/SF2)	429	8	36	259	5	24	2.02E-03 ± 9.54E-04	5.07E-04 ± 3.90E-04	4.00	0.25
SFRS2	NP_035488.1	splicing factor, arginine/serine-rich 2	125	2	11	52	1	4	1.87E-04 ± 5.30E-05	6.32E-05 ± -	2.96	0.34
SFN	NP_061224.1	stratifin	303	5	17	189	4	10	1.94E-03 ± 9.69E-04	2.35E-03 ± 1.63E-03	0.83	1.21
SMC2L1	NP_032043.3	structural maintenance of chromosomes 2-like 1	72	2	2	119	4	4	6.95E-05 ± -	1.52E-04 ± -	0.46	2.19
SSRP1	NP_892035.1	structure specific recognition protein 1	93	1	2	165	4	12	4.68E-05 ± -	5.13E-04 ± 2.79E-05	0.09	10.96
SKIV2L2	NP_082427.1	superkiller viralicidic activity 2-like 2	86	2	3	170	3	4	4.77E-05 ± -	2.68E-05 ± 2.32E-05	1.78	0.56
SUPT5H	NP_038704.1	suppressor of Ty 5 homolog	101	2	2	71	1	3	3.06E-05 ± -	1.29E-05 ± -	2.37	0.42
SUPT6H	NP_033323.1	suppressor of Ty 6 homolog	83	1	1	96	1	1	9.59E-06 ± -	1.62E-05 ± 8.09E-06	0.59	1.69
SMARCC1	NP_033237.1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily c, member 1	53	1	1	194	3	4	4.50E-05 ± -	3.16E-05 ± 2.68E-05	1.42	0.70
RIF1	NP_780447.3	telomere-associated protein RIF1	376	5	3	406	6	5	3.88E-05 ± 2.77E-05	4.43E-05 ± 3.76E-05	0.88	1.14
TXNL1	NP_058072.2	thioredoxin-like 1	98	1	5	171	2	13	4.01E-04 ± -	1.45E-04 ± 9.66E-05	2.77	0.36
TRIP12	NP_598736.3	thyroid hormone receptor interactor 12	124	3	2	323	6	5	1.31E-04 ± -	5.86E-05 ± 2.44E-05	2.23	0.45
TOP2A	NP_035753.1	topoisomerase (DNA) II alpha	411	8	7	413	8	7	1.59E-04 ± 8.41E-05	2.31E-04 ± 4.32E-05	0.69	1.46
SIN3A	NP_035508.1	transcriptional regulator, SIN3A	61	1	2	192	3	3	5.20E-05 ± -	1.64E-05 ± 7.75E-06	3.16	0.32
TFRC	NP_035768.1	transferrin receptor	158	2	5	125	1	2	7.59E-05 ± 7.67E-05	9.15E-05 ± -	0.83	1.20
TRPC1	NP_035773.1	transient receptor potential cation channel, subfamily C, member 1	76	2	1	106	3	3	1.13E-04 ± 1.45E-05	8.63E-05 ± 2.44E-05	1.30	0.77
TRPC7	NP_036165.1	transient receptor potential cation channel, subfamily C, member 7	76	2	1	76	2	1	1.06E-04 ± 1.36E-05	6.48E-05 ± -	1.63	0.61

Proteins in common with embryonic stem cells and embryoid bodies, continued.

Gene	NCBI Accession	Protein Description	Embryonic stem cells: Total Mascot Score	Embryonic stem cells: Total No. Peptides	Embryonic stem cells: Total % Protein Coverage	Embryoid Bodies: Total Mascot Score	Embryoid Bodies: Total No. Peptides	Embryoid Bodies: Total % Protein Coverage	Embryonic stem cells AVG N_SC ± STDEV	Embryoid Bodies AVG N_SC ± STDEV	ES/EB Ratio	EB/ES Ratio
TSN	NP_035780.1	translin	100	1	9	162	2	15	1.21E-04 ± 4.19E-05	3.67E-04 ± 2.21E-04	0.33	3.04
TOMM34	NP_080272.1	translocase of outer mitochondrial membrane 34	75	1	5	111	2	10	1.61E-04 ± -	2.71E-04 ± -	0.59	1.69
TNPO1	NP_848831.1	transportin 1	360	7	12	461	8	15	2.03E-04 ± -	1.92E-04 ± 1.06E-04	1.06	0.95
TCOF1	NP_035682.1	treacle	224	3	3	74	1	1	5.85E-05 ± 3.83E-05	2.11E-05 ± -	2.77	0.36
TRIM25	NP_033572.1	tripartite motif protein 25	60	1	3	69	1	2	2.61E-05 ± -	6.60E-05 ± -	0.40	2.53
TBCA	NP_033347.1	tubulin cofactor a	319	3	24	365	4	24	8.68E-04 ± 3.86E-04	6.46E-04 ± 0.00E+00	1.34	0.74
TUBA1	NP_035783.1	tubulin, alpha 1	937	14	47	1023	14	49	7.57E-03 ± 7.54E-04	1.00E-02 ± 2.92E-03	0.76	1.32
TUBA3	NP_033472.1	tubulin, alpha 3	784	12	40	907	14	49	5.26E-03 ± 5.75E-04	6.51E-03 ± 3.76E-03	0.81	1.24
TUBB	NP_076205.1	tubulin, beta	1398	21	69	1219	18	51	6.17E-03 ± 1.99E-03	4.59E-03 ± 1.89E-03	1.35	0.74
TUB6	NP_080749.2	tubulin, beta 6	627	10	35	485	8	30	1.80E-03 ± 8.13E-04	2.44E-03 ± 1.40E-03	0.74	1.35
TUBA8	NP_059075.1	tubulin, alpha 8	505	8	24	532	8	29	3.62E-03 ± 5.67E-04	0.004434996 ± 2.37E-01	1.22	0.82
TPD52L2	NP_079758.2	tumor protein D52-like 2	67	1	10	52	1	10	2.01E-04 ± 4.34E-05	0.000126889 ± -	1.58	0.63
TWF1	NP_032997.3	twinfilin	89	1	5	82	1	5	4.73E-05 ± -	5.32E-05 ± 2.30E-05	0.89	1.12
YWHAB	NP_061223.2	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide	321	5	18	289	5	18	2.15E-03 ± 8.59E-04	2.69E-03 ± 1.30E-03	0.80	1.25
ASCC3L1	NP_796188.2	activating signal cointegrator 1 complex subunit 3-like 1	323	6	4	590	11	8	4.65E-05 ± 2.32E-05	9.15E-05 ± 5.81E-05	0.51	1.97
UBQLN1	NP_689420.1	ubiquilin 1 isoform 2	66	1	3	52	1	3	2.99E-05 ± -	6.30E-05 ± 1.78E-05	0.47	2.11
UBQLN4	NP_277068.1	ubiquilin 4	54	1	4	68	1	2	5.55E-05 ± -	2.11E-04 ± -	0.26	3.79
UQCRC1	NP_079683.2	ubiquinol-cytochrome c reductase core protein 1	203	3	13	251	4	13	1.26E-04 ± 8.68E-05	6.69E-04 ± 1.45E-04	0.19	5.29
UCHL3	NP_057932.1	ubiquitin carboxyl-terminal esterase L3 (ubiquitin thiolesterase)	57	1	4	62	1	7	7.20E-05 ± -	6.07E-05 ± -	1.19	0.84
UCHL5	NP_062508.1	ubiquitin C-terminal hydrolase 37	72	1	7	187	3	13	5.03E-05 ± -	5.09E-04 ± 4.20E-04	0.10	10.12
USP5	NP_038728.1	ubiquitin specific protease 5 (isopeptidase T)	459	8	16	390	6	11	7.01E-04 ± 5.63E-04	6.07E-04 ± 2.87E-04	1.15	0.87
USP7	NP_001003918.1	ubiquitin specific protease 7	55	1	2	90	1	2	1.50E-05 ± -	3.37E-05 ± 2.63E-05	0.44	2.25
UBE2V1	NP_075719.1	ubiquitin-conjugating enzyme E2 variant 1	144	3	21	63	1	8	1.35E-03 ± 1.27E-03	1.57E-03 ± 3.36E-04	0.86	1.16
UBE2C	NP_081061.1	ubiquitin-conjugating enzyme E2C	67	1	9	68	1	9	3.24E-04 ± 6.54E-05	5.85E-04 ± 1.65E-04	0.55	1.81
UTF1	NP_033508.1	undifferentiated embryonic cell transcription factor 1	141	3	18	62	1	10	5.37E-04 ± -	2.47E-04 ± 0.00E+00	2.17	0.46
CSDE1	NP_659150.1	upstream of NRAS	81	2	3	143	4	5	4.15E-05 ± -	1.57E-04 ± 2.47E-05	0.26	3.79
UPP1	NP_033503.1	uridine phosphorylase 1	87	2	7	104	3	11	1.60E-04 ± -	1.26E-03 ± -	0.13	7.87
ATP6V1E1	NP_031536.2	vacuolar H+ ATPase E1	90	1	6	139	2	10	1.46E-04 ± -	2.47E-04 ± -	0.59	1.69
VARS	NP_035820.2	valyl-tRNA synthetase 2	179	3	3	188	4	4	1.01E-04 ± 1.54E-04	4.10E-05 ± -	2.47	0.40
CRKL	NP_031790.2	v-crk sarcoma virus CT10 oncogene homolog (avian)-like	51	1	6	117	2	10	1.09E-04 ± -	6.68E-04 ± 2.93E-04	0.16	6.11
VIL2	NP_033536.2	villin 2	347	6	15	178	4	11	2.54E-04 ± 1.76E-04	2.14E-04 ± 1.68E-04	1.19	0.84
VCL	NP_033528.3	vinculin	72	1	1	76	1	1	1.24E-04 ± -	3.93E-05 ± 3.70E-05	3.16	0.32
WDR12	NP_067287.1	WD repeat domain 12	89	1	4	66	1	4	7.83E-05 ± -	7.70E-05 ± 1.91E-05	1.02	0.98
WIBG	NP_084376.1	within bgcn homolog	209	4	46	80	1	17	4.89E-04 ± 4.08E-04	1.38E-04 ± -	3.56	0.28

Table 6.13. Proteins in common with embryonic stem cells, early primitive ectoderm-like cells and embryoid bodies. Mascot score, number of peptides, and coverage are totals from nonredundant peptides seen over all three replicate analyses.

			Embryonic stem cells			Early Primitive Ectoderm-like			Embryoid Bodies								
Gene	NCBI Accession	Protein Description	Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Embryonic stem cells AVG N_SC ± STDEV	Early primitive ectoderm-like cells AVG N_SC ± STDEV	Embryoid Bodies AVG N_SC ± STDEV	EPL/ES Ratio	EB/ES Ratio	EPL/EB Ratio
ATIC	NP_080471.1	5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase	474	7	17	188	4	11	312	4	10	4.66E-04 ± 5.28E-04	2.18E-04 ± 2.31E-04	1.73E-04 ± 1.53E-02	0.47	0.37	1.26
AKAP12	NP_112462.1	A kinase (PRKA) anchor protein (gravin) 12	834	12	14	125	1	1	641	9	10	1.87E-04 ± 4.50E-05	2.88E-05 ± 1.36E-05	2.24E-04 ± 1.88E-02	0.15	1.20	0.13
ACTC	NP_033738.1	actin, alpha, cardiac	588	11	37	627	14	38	573	11	34	7.42E-03 ± 1.65E-03	9.11E-03 ± 1.97E-03	6.60E-03 ± 2.37E-01	1.23	0.89	1.38
ACTG1	NP_033739.1	actin, gamma, cytoplasmic 1	1041	17	51	931	18	53	1116	18	51	9.14E-03 ± 1.23E-03	1.03E-02 ± 2.69E-03	1.17E-02 ± 3.82E-01	1.12	1.28	0.88
ACTN4	NP_068695.1	actinin alpha 4	810	13	21	421	7	10	680	11	19	5.87E-04 ± 5.52E-04	5.90E-04 ± 4.27E-04	6.89E-04 ± 9.31E-03	1.01	1.17	0.86
ACTN1	NP_598917.1	actinin, alpha 1	466	7	10	537	8	12	555	11	16	3.40E-04 ± 4.07E-04	5.79E-04 ± 4.45E-04	5.27E-04 ± 6.51E-03	1.70	1.55	1.10
ADSS	NP_031448.2	adenylosuccinate synthetase, non muscle	153	2	6	92	1	5	282	4	15	1.81E-04 ± 2.05E-04	2.48E-04 ± 1.50E-04	1.22E-04 ± 5.30E-03	1.37	0.67	2.02
ARFGAP1	NP_665703.2	ADP-ribosylation factor GTPase activating protein 1	66	1	4	59	1	4	53	1	4	8.00E-05 ± 0.00E+00	1.56E-04 ± 1.10E-04	6.74E-05 ± -	1.95	0.84	2.31
AHNAK	NP_033773.1	AHNAK nucleoprotein isoform 1	319	10	3	715	17	5	140	4	1	7.71E-05 ± 2.27E-05	1.41E-04 ± 4.05E-05	9.13E-05 ± -	1.83	1.18	1.54
ADH5	NP_031436.2	alcohol dehydrogenase 5 (class III), chi polypeptide	140	2	10	58	1	8	227	3	12	2.21E-04 ± 1.60E-04	8.64E-05 ± -	1.44E-03 ± 5.30E-02	0.39	6.52	0.06
AKR1B1	NP_033788.2	aldo-keto reductase family 1, member B3 (aldose reductase)	319	8	32	51	1	4	242	5	22	1.71E-03 ± 6.03E-04	1.02E-04 ± -	6.92E-04 ± 4.51E-02	0.06	0.40	0.15
ALDOA	NP_031464.1	aldolase 1, A isoform	822	14	63	352	7	24	527	10	48	5.62E-03 ± 4.60E-04	2.78E-03 ± 7.12E-04	4.52E-03 ± 8.36E-02	0.49	0.80	0.61
GANAB	NP_032086.1	alpha glucosidase 2 alpha neutral subunit	310	5	8	230	4	7	271	5	8	1.09E-04 ± 3.57E-05	8.92E-05 ± 6.96E-05	1.59E-04 ± 4.09E-03	0.82	1.46	0.56
PPP2R1A	NP_058587.1	alpha isoform of regulatory subunit A, protein phosphatase 2	212	4	10	196	4	10	262	3	8	1.59E-04 ± 1.60E-04	4.94E-04 ± 1.45E-04	2.84E-04 ± 9.48E-03	3.10	1.79	1.74
ACTR3	NP_076224.1	ARP3 actin-related protein 3 homolog	275	5	19	63	1	5	251	4	18	2.11E-04 ± 1.27E-04	1.55E-04 ± 0.00E+00	3.12E-04 ± 1.35E-02	0.73	1.48	0.50
ATP5B	NP_058054.2	ATP synthase, H+ transporting mitochondrial F1 complex, beta subunit	813	12	26	617	9	24	742	10	28	1.19E-03 ± 7.68E-04	2.26E-03 ± 1.06E-03	1.81E-03 ± 1.00E-01	1.90	1.52	1.25
ATP5A1	NP_031531.1	ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit, isoform 1	895	15	32	770	12	28	894	14	30	3.49E-03 ± 1.05E-03	3.15E-03 ± 8.60E-04	3.87E-03 ± 1.47E-01	0.90	1.11	0.81
ATP2A2	NP_033852.1	ATPase, Ca++ transporting, cardiac muscle, slow twitch 2	241	3	5	188	3	7	116	2	3	9.95E-05 ± 4.98E-05	1.13E-04 ± 6.87E-05	1.40E-05 ± -	1.14	0.14	8.10
ATP6V1A	NP_031534.2	ATPase, H+ transporting, V1 subunit A, isoform 1	65	1	3	177	2	7	231	3	10	2.68E-05 ± -	2.27E-04 ± 3.02E-05	1.89E-04 ± 2.10E-02	8.46	7.03	1.20
SSB	NP_033304.1	autoantigen La	402	7	18	200	5	13	314	5	20	8.11E-04 ± 3.71E-04	1.25E-03 ± 1.10E-03	6.61E-04 ± 3.37E-02	1.54	0.82	1.88
BANF1	NP_035923.1	barrier to autointegration factor 1	185	3	42	56	1	18	158	2	40	2.79E-03 ± 1.97E-03	1.09E-03 ± 5.13E-04	1.20E-03 ± 9.96E-02	0.39	0.43	0.91
PHB2	NP_031557.1	B-cell receptor-associated protein 37	230	4	16	181	5	23	176	4	14	6.29E-04 ± 4.72E-04	5.96E-04 ± 5.37E-04	2.97E-04 ± 7.15E-03	0.95	0.47	2.01
BASP1	NP_081671.1	brain abundant, membrane attached signal protein 1	330	5	40	54	1	14	157	2	28	7.81E-04 ± 8.23E-04	5.72E-04 ± -	1.24E-04 ± 8.73E-03	0.73	0.16	4.63
CALR	NP_031617.1	calreticulin	250	6	28	345	8	24	226	6	20	9.55E-04 ± 3.47E-04	2.25E-03 ± 7.41E-04	7.21E-04 ± 2.37E-03	2.36	0.76	3.12
CAP1	NP_031624.1	CAP, adenylate cyclase-associated protein 1	106	1	8	110	1	8	229	2	12	1.51E-04 ± 8.79E-05	1.36E-04 ± -	1.57E-04 ± 9.47E-03	0.90	1.04	0.87
CAD	NP_076014.1	carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase	553	8	4	174	3	2	512	8	6	7.44E-05 ± 1.97E-05	5.08E-05 ± 5.13E-05	8.16E-05 ± 6.05E-03	0.68	1.10	0.62
CDC37	NP_058022.1	cell division cycle 37 homolog	230	3	14	60	1	6	224	3	14	1.34E-03 ± 5.21E-04	2.98E-04 ± 1.81E-04	2.82E-04 ± 2.22E-02	0.22	0.21	1.06
CDC5L	NP_690023.1	cell division cycle 5-like	269	5	10	119	2	5	141	2	4	1.99E-04 ± 1.31E-04	4.03E-05 ± -	1.16E-04 ± 3.62E-03	0.20	0.58	0.35
CCT2	NP_031662.1	chaperonin subunit 2 (beta)	745	11	30	392	5	14	1102	14	34	2.47E-03 ± 1.28E-03	1.29E-03 ± 9.22E-05	1.72E-03 ± 2.75E-02	0.52	0.70	0.75
CCT3	NP_033966.1	chaperonin subunit 3 (gamma)	643	11	22	193	5	14	689	12	25	7.39E-04 ± 3.96E-04	6.82E-04 ± 4.61E-04	7.77E-04 ± 1.46E-02	0.92	1.05	0.88
CCT4	NP_033967.1	chaperonin subunit 4 (delta)	517	10	24	288	6	19	831	16	39	5.42E-04 ± 7.04E-04	5.79E-04 ± 4.50E-04	8.98E-04 ± 2.85E-02	1.07	1.65	0.65
CCT5	NP_031663.1	chaperonin subunit 5 (epsilon)	155	4	10	182	3	8	299	6	13	1.73E-04 ± 1.97E-04	4.38E-04 ± 2.69E-04	5.50E-04 ± 3.18E-02	2.53	3.17	0.80
CCT6A	NP_033968.1	chaperonin subunit 6a (zeta)	268	6	14	263	5	14	347	7	23	9.97E-04 ± 5.54E-04	2.92E-03 ± 6.35E-04	1.02E-03 ± 3.29E-02	2.93	1.02	2.87
CCT7	NP_031664.2	chaperonin subunit 7 (eta)	459	9	23	177	5	18	437	8	18	4.16E-04 ± 1.07E-04	2.67E-04 ± 4.20E-05	8.89E-04 ± 5.93E-03	0.64	2.14	0.30
CCT8	NP_033970.2	chaperonin subunit 8 (theta)	407	7	18	158	2	5	631	9	24	8.96E-04 ± 2.53E-04	3.73E-04 ± 3.02E-04	1.27E-03 ± 4.57E-02	0.42	1.41	0.30
CLIC1	NP_254279.1	chloride intracellular channel 1	186	2	17	58	1	7	286	4	24	4.58E-04 ± 1.73E-04	1.34E-04 ± -	3.47E-04 ± 2.09E-02	0.29	0.76	0.39
CLIC6	NP_766057.1	chloride intracellular channel 6	68	1	3	89	2	3	186	3	8	2.78E-05 ± -	3.79E-04 ± -	1.76E-04 ± 4.97E-03	13.66	6.32	2.16
CS	NP_080720.1	citrate synthase	155	3	8	64	1	3	132	4	9	9.51E-05 ± 4.12E-05	7.66E-04 ± -	6.92E-04 ± 8.08E-02	8.05	7.27	1.11
CLTC	NP_001003908.1	clathrin, heavy polypeptide (Hc)	621	11	9	377	7	5	1123	18	14	1.28E-04 ± 6.01E-05	1.22E-04 ± 6.20E-05	8.14E-04 ± 4.43E-02	0.95	6.34	0.15
COPA	NP_034068.2	coatomer protein complex subunit alpha	92	2	2	231	3	4	165	4	5	1.08E-04 ± -	9.68E-05 ± 6.09E-05	4.56E-05 ± -	0.89	0.42	2.12
CFL1	NP_031713.1	cofilin 1, non-muscle	765	12	67	476	6	43	616	9	56	1.51E-02 ± 5.98E-03	1.58E-02 ± 2.17E-02	9.05E-03 ± 6.16E-01	1.05	0.60	1.75
CUL4B	NP_082564.2	cullin 4B	261	4	7	119	1	3	103	2	2	6.83E-05 ± 0.00E+00	1.33E-04 ± -	2.01E-04 ± -	1.95	2.95	0.66
CKAP4	NP_780660.1	cytoskeleton-associated protein 4	147	4	9	663	11	23	285	6	10	1.15E-04 ± -	1.69E-03 ± 2.97E-04	1.70E-04 ± 3.43E-03	14.64	1.48	9.92
DDX1	NP_598801.1	DEAD (Asp-Glu-Ala-Asp) box polypeptide 1	295	6	14	111	1	3	206	3	8	5.59E-05 ± 4.74E-05	1.89E-04 ± 5.04E-05	1.13E-04 ± 3.27E-03	3.38	2.02	1.67
DDX21	NP_062426.2	DEAD (Asp-Glu-Ala-Asp) box polypeptide 21	316	6	10	145	2	4	416	9	15	9.72E-05 ± 1.94E-05	5.69E-05 ± 2.68E-05	2.13E-04 ± 1.86E-02	0.59	2.19	0.27
DDX5	NP_031866.2	DEAD (Asp-Glu-Ala-Asp) box polypeptide 5	494	10	20	404	7	13	580	11	21	5.83E-04 ± 1.02E-04	7.88E-04 ± 7.43E-05	9.53E-04 ± 5.50E-02	1.35	1.63	0.83
DDX17	NP_951062.1	DEAD box polypeptide 17 isoform 1	384	5	10	232	4	8	367	5	10	8.80E-04 ± 6.03E-04	4.13E-04 ± 4.60E-04	4.42E-04 ± 2.68E-02	0.47	0.50	0.93
DDX3X	NP_034158.1	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 3, X-linked	622	10	19	264	4	10	417	7	14	3.83E-04 ± 3.19E-04	1.95E-04 ± 1.29E-04	5.13E-04 ± 2.35E-02	0.51	1.34	0.38
DHX9	NP_031868.1	DEAH (Asp-Glu-Ala-His) box polypeptide 9	209	4	3	96	2	2	238	5	6	3.20E-05 ± 3.46E-05	7.02E-05 ± -	8.09E-05 ± 5.63E-03	2.20	2.53	0.87
DUT	NP_076084.2	deoxyuridine triphosphatase	244	4	33	93	2	17	239	4	33	9.20E-04 ± 2.70E-04	5.98E-04 ± -	1.58E-03 ± 5.19E-02	0.65	1.72	0.38
DLAT	NP_663589.2	dihydrolipoamide S-acetyltransferase (E2 component of pyruvate dehydrogenase complex)	93	1	3	142	2	5	154	3	8	6.87E-05 ± 3.94E-05	5.03E-05 ± 0.00E+00	1.59E-04 ± 5.02E-03	0.73	2.32	0.32
DPP3	NP_598564.1	dipeptidyl peptidase III	158	2	5	63	1	2	227	4	9	1.42E-04 ± 9.34E-05	8.75E-05 ± -	4.73E-05 ± 1.34E-03	0.62	0.33	1.85
PARK7	NP_065594.2	DJ-1 protein	392	6	47	98	1	13	109	2	18	1.66E-03 ± 1.14E-03	3.42E-04 ± -	2.95E-04 ± 1.04E-02	0.21	0.18	1.16
DCTN2	NP_081427.1	dynactin 2	232	3	12	66	1	5	173	2	8	2.06E-04 ± 7.13E-05	8.04E-05 ± -	1.16E-04 ± 2.00E-03	0.39	0.56	0.69
DNM1L	NP_001021118.1	dynamin 1-like isoform b	164	5	7	64	1	3	186	4	7	1.89E-04 ± -	4.62E-05 ± -	5.32E-05 ± 5.76E-03	0.24	0.28	0.87
DYNC1H1	NP_084514.1	dynein, cytoplasmic, heavy chain 1	472	10	3	406	6	2	1070	19	5	4.28E-05 ± 2.17E-05	4.87E-05 ± 6.06E-05	8.82E-05 ± 1.66E-03	1.14	2.06	0.55
ETFA	NP_663590.2	electron transferring flavoprotein, alpha polypeptide	275	5	23	284	4	19	69	1	5	2.32E-04 ± 7.59E-05	1.07E-03 ± 6.86E-04	1.68E-04 ± 0.00E+00	4.60	0.72	6.36
EFTUD2	NP_035561.1	elongation factor Tu GTP binding domain containing 2	249	4	7	82	1	2	409	8	11	8.52E-05 ± 2.95E-05	6.65E-05 ± -	1.25E-04 ± 5.05E-03	0.78	1.46	0.53
ENAH	NP_034265.1	enabled homolog	83	1	3	84	1	3	96	1	3	2.06E-05 ± 0.00E+00	4.03E-05 ± -	5.80E-05 ± 5.59E-03	1.95	2.81	0.69
ERP29	NP_080405.1	endoplasmic reticulum protein ERp29 precursor	56	1	3	253	5	21	110	3	13	4.74E-04 ± 1.34E-04	1.05E-03 ± 7.85E-04	1.07E-04 ± -	2.21	0.22	9.84
ENO3	NP_031959.1	enolase 3, beta muscle	378	3	11	484	6	21	456	6	20	4.59E-03 ± 6.21E-04	6.70E-03 ± 1.64E-03	6.41E-03 ± 3.05E-01	1.46	1.40	1.04
ECHS1	NP_444349.1	enoyl Coenzyme A hydratase, short chain, 1, mitochondrial	73	1	7	128	2	11	62	1	4	1.14E-04 ± 0.00E+00	2.78E-04 ± 2.36E-04	4.81E-05 ± -	2.44	0.42	5.79
EEF1A1	NP_034236.1	eukaryotic translation elongation factor 1 alpha 1	1029	18	40	735	13	34	1078	18	42	1.36E-02 ± 3.03E-03	6.50E-03 ± 2.37E-03	1.56E-02 ± 5.61E-01	0.48	1.15	0.42
EEF1B2	NP_061266.2	eukaryotic translation elongation factor 1 beta 2	361	6	30	277	6	40	284	4	23	1.10E-03 ± 1.47E-04	5.02E-04 ± 1.02E-04	1.41E-03 ± 6.86E-02	0.46	1.27	0.36
EEF1D	NP_083939.1	eukaryotic translation elongation factor 1 delta isoform a	145	2	4	331	5	10	204	3	7	2.93E-04 ± 3.83E-05	1.32E-03 ± 1.33E-03	1.41E-04 ± 1.04E-02	4.52	0.48	9.37
EEF1G	NP_080283.2	eukaryotic translation elongation factor 1 gamma	363	7	17	117	2	6	444	8	22	7.45E-04 ± 2.79E-04	1.72E-04 ± 1.71E-04	1.15E-03 ± 4.23E-02	0.23	1.54	0.15
EEF2	NP_031933.1	eukaryotic translation elongation factor 2	1412	24	34	1003	17	29	967	17	26	2.62E-03 ± 9.25E-04	4.09E-03 ± 2.51E-04	1.06E-03 ± 4.80E-02	1.56	0.41	3.85
EIF2S1	NP_080390.1	eukaryotic translation initiation factor 2, subunit 1 alpha	269	4	20	185	2	10	596	8	35	2.15E-03 ± 4.58E-04	1.71E-04 ± 5.92E-05	1.91E-03 ± 6.53E-02	0.08	0.88	0.09
EIF3S8	NP_666312.1	eukaryotic translation initiation factor 3, subunit 8	201	5	6	160	2	3	159	5	7	9.08E-05 ± 1.03E-04	3.55E-05 ± 0.00E+00	4.60E-05 ± 4.33E-03	0.39	0.51	0.77
EIF4G1	NP_001005331.1	eukaryotic translation initiation factor 4, gamma 1 isoform b	301	6	4	129	2	1	388	7	7	1.18E-04 ± 6.26E-05	5.07E-05 ± 1.43E-05	2.48E-04 ± 2.04E-02	0.43	2.11	0.20
EIF5A	NP_853613.1	eukaryotic translation initiation factor 5A	357	7	60	171	2	23	245	4	31	1.04E-03 ± 7.77E-04	1.54E-03 ± 4.37E-04	1.42E-03 ± 2.91E-02	1.48	1.37	1.08
FUBP3	NP_001028561.1	far upstream element (FUSE) binding protein 3	231	4	15	68	1	4	146	2	9	3.36E-04 ± 1.59E-04	5.62E-05 ± -	7.28E-05 ± -	0.17	0.22	0.77
FASN	NP_032014.2	fatty acid synthase	359	5	3	116	2	1	337	5	3	1.41E-04 ± 3.82E-05	2.58E-05 ± 1.29E-05	5.95E-05 ± 3.07E-03	0.18	0.42	0.43
FBL	NP_032017.2	fibrillarin	275	5	27	90	2	12	410	8	36	1.35E-04 ± 5.84E-05	9.88E-05 ± 0.00E+00	7.83E-04 ± 6.18E-02	0.73	5.80	0.13
FKBP4	NP_034349.1	FK506 binding protein 4	293	5	17	55	1	3	457	6	23	2.41E-04 ± 7.52E-05	7.05E-05 ± -	3.56E-04 ± 1.44E-02	0.29	1.48	0.20
GLUD1	NP_032159.1	glutamate dehydrogenase 1	147	2	5	108	2	4	234	4	11	4.94E-05 ± 1.71E-05	1.74E-04 ± 0.00E+00	1.50E-04 ± 5.00E-03	3.51	3.04	1.16

Proteins in common with embryonic stem cells, early primitive ectoderm-like cells and embryoid bodies, continued.

			Embryonic stem cells			Early Primitive Ectoderm-like			Embryoid Bodies								
Gene	NCBI Accession	Protein Description	Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Embryonic stem cells AVG N_SC ± STDEV	Early primitive ectoderm-like cells AVG N_SC ± STDEV	Embryoid Bodies AVG N_SC ± STDEV	EPL/ES Ratio	EB/ES Ratio	EPL/EB Ratio
RANBP1	NP_035369.1	glutamate dehydrogenase 1	74	1	5	78	1	5	86	1	5	1.36E-04 ± 4.71E-05	4.77E-04 ± 2.25E-04	5.27E-04 ± 1.73E-02	3.51	3.88	0.91
QARS	NP_598555.1	glutaminyl-tRNA synthetase	111	2	3	52	1	4	64	1	2	1.39E-04 ± 1.51E-05	4.17E-05 ± -	1.80E-05 ± 0.00E+00	0.30	0.13	2.31
GSTP1	NP_038569.1	glutathione S-transferase, pi 1	137	2	17	108	1	8	142	2	17	1.13E-03 ± 4.34E-04	4.61E-04 ± -	1.40E-03 ± 5.02E-02	0.41	1.24	0.33
GARS	NP_851009.1	glycyl-tRNA synthetase	145	3	6	130	2	5	300	5	10	7.57E-05 ± 9.18E-05	1.11E-04 ± 3.13E-05	1.85E-04 ± 1.17E-02	1.46	2.45	0.60
GLO1	NP_079650.2	glyoxalase 1	202	4	30	158	2	16	86	1	11	3.30E-04 ± 3.41E-04	8.78E-04 ± 0.00E+00	3.79E-04 ± -	2.66	1.15	2.31
GLG1	NP_033175.1	golgi apparatus protein 1	297	5	7	100	1	1	212	3	4	7.04E-05 ± 3.98E-05	2.20E-04 ± -	9.50E-05 ± 3.14E-03	3.12	1.35	2.31
GGA2	NP_083034.1	golgi associated, gamma adaptin ear containing, ARF binding protein 2	84	2	7	173	2	11	52	1	4	5.49E-05 ± -	1.07E-04 ± 7.58E-05	4.63E-05 ± -	1.95	0.84	2.31
GPIAP1	NP_058019.2	GPI-anchored membrane protein 1	325	7	18	139	2	5	71	1	6	3.04E-04 ± 2.65E-04	1.37E-04 ± -	9.87E-05 ± 1.12E-02	0.45	0.32	1.39
H2AFX	NP_034566.1	H2A histone family, member X	185	3	27	525	7	33	184	3	27	1.22E-02 ± 1.50E-03	2.30E-02 ± 5.26E-03	8.39E-03 ± 3.39E-01	1.89	0.69	2.74
HSPA5	NP_071705.2	heat shock 70kD protein 5 (glucose-regulated protein)	1094	16	33	1861	28	48	781	11	24	1.23E-03 ± 3.45E-04	1.03E-02 ± 2.27E-03	2.42E-03 ± 1.23E-01	8.35	1.97	4.24
HSPB1	NP_038588.1	heat shock protein 1	371	6	36	169	4	35	361	6	40	4.91E-03 ± 9.72E-04	9.32E-04 ± 6.59E-04	1.52E-03 ± 2.16E-02	0.19	0.31	0.61
HSPD1	NP_034607.2	heat shock protein 1 (chaperonin)	1823	24	51	1330	22	43	1452	21	44	3.60E-03 ± 1.68E-03	8.44E-03 ± 2.02E-03	4.30E-03 ± 7.00E-02	2.34	1.19	1.96
HSP90AA1	NP_034610.1	heat shock protein 1, alpha	1911	32	50	1019	22	36	2186	37	48	7.59E-03 ± 8.71E-04	1.85E-03 ± 9.92E-04	8.09E-03 ± 1.68E-01	0.24	1.07	0.23
HSP90AB1	NP_032328.2	heat shock protein 1, beta	1984	36	55	1333	24	40	2378	41	55	1.17E-02 ± 2.65E-03	2.33E-03 ± 1.57E-03	1.30E-02 ± 4.76E-01	0.20	1.11	0.18
HSPH1	NP_038587.1	heat shock protein 105	454	8	16	102	2	3	478	6	13	1.86E-04 ± 9.71E-05	1.51E-04 ± 1.06E-04	1.79E-04 ± 8.61E-03	0.81	0.96	0.84
HSPA2	NP_001002012.1	heat shock protein 2	495	7	9	411	8	14	534	8	12	1.59E-03 ± 6.92E-04	7.65E-04 ± 8.84E-05	2.84E-03 ± 2.07E-01	0.48	1.78	0.27
HSPA4	NP_032326.2	heat shock protein 4	1231	18	36	169	4	9	1006	14	29	1.10E-03 ± 2.27E-04	2.69E-04 ± 2.69E-04	7.85E-04 ± 2.74E-02	0.24	0.71	0.34
HSPA8	NP_112442.2	heat shock protein 8	1687	27	51	1144	20	45	1402	20	42	5.47E-03 ± 8.44E-04	3.73E-03 ± 2.55E-03	6.86E-03 ± 3.37E-01	0.68	1.26	0.54
A9B (includes EG:3	NP_034611.1	heat shock protein 9A	823	13	28	249	4	13	719	12	27	1.27E-03 ± 2.00E-04	1.06E-03 ± 9.18E-04	1.01E-03 ± 2.06E-02	0.84	0.79	1.05
HUWE1	NP_067498.3	HECT, UBA and WWE domain containing 1	277	3	1	177	2	1	314	5	1	1.76E-05 ± 1.43E-05	4.43E-05 ± 4.49E-05	1.91E-05 ± 2.25E-03	2.51	1.08	2.31
HNRPAB	NP_034578.1	heterogeneous nuclear ribonucleoprotein A/B	495	10	32	173	4	16	263	4	17	2.86E-03 ± 6.10E-04	2.83E-04 ± 8.01E-05	4.83E-03 ± 5.20E-01	0.10	1.69	0.06
HNRPA1	NP_034577.1	heterogeneous nuclear ribonucleoprotein A1 isoform a	747	14	38	321	5	20	777	13	37	1.91E-03 ± 7.72E-04	6.70E-03 ± 1.70E-03	2.73E-03 ± 5.92E-02	3.50	1.43	2.45
HNRPA2B1	NP_058086.1	heterogeneous nuclear ribonucleoprotein A2/B1 isoform 1	681	12	24	420	6	23	666	11	29	2.15E-03 ± 1.37E-03	2.08E-03 ± 1.64E-04	8.35E-03 ± 1.50E-01	0.97	3.88	0.25
HNRPC	NP_058580.1	heterogeneous nuclear ribonucleoprotein C	257	7	25	253	7	27	118	3	10	1.13E-03 ± 6.31E-04	2.06E-04 ± 0.00E+00	1.25E-03 ± -	0.18	1.11	0.17
HNRPD	NP_031542.1	heterogeneous nuclear ribonucleoprotein D	220	5	18	160	3	12	193	4	13	1.19E-03 ± 1.53E-04	5.28E-04 ± -	1.66E-03 ± 2.71E-01	0.44	1.39	0.32
HNRPF	NP_598595.1	heterogeneous nuclear ribonucleoprotein F	227	4	14	281	4	18	350	5	16	6.38E-04 ± 3.93E-04	3.63E-03 ± 1.16E-03	4.34E-03 ± 4.74E-02	5.69	6.80	0.84
HNRPH1	NP_067485.1	heterogeneous nuclear ribonucleoprotein H1	338	5	17	124	2	6	348	6	20	4.18E-04 ± 2.45E-04	6.23E-04 ± 5.20E-04	4.56E-04 ± 1.29E-02	1.49	1.09	1.37
HNRPK	NP_079555.1	heterogeneous nuclear ribonucleoprotein K	591	10	28	683	10	32	806	12	38	1.66E-03 ± 6.90E-04	1.07E-03 ± 1.07E-04	2.62E-03 ± 7.24E-02	0.65	1.58	0.41
HNRPL	NP_796275.2	heterogeneous nuclear ribonucleoprotein L	217	4	13	71	1	2	125	2	5	7.46E-05 ± 2.11E-05	9.02E-04 ± 1.11E-03	1.76E-04 ± -	12.10	2.36	5.12
HDLBP	NP_598569.1	high density lipoprotein binding protein	293	6	7	278	7	7	167	3	4	1.44E-04 ± 5.22E-05	1.91E-04 ± 5.40E-05	9.54E-05 ± 7.96E-03	1.33	0.66	2.00
DCPS	NP_081306.1	histidine triad protein member 5	220	4	17	58	1	6	146	2	10	3.18E-04 ± 1.73E-04	1.91E-04 ± -	1.65E-04 ± -	0.60	0.52	1.16
HIST1H1A	NP_085112.1	histone 1, H1a	472	9	38	488	6	26	324	5	19	1.86E-03 ± 4.11E-04	9.10E-03 ± 6.15E-03	7.43E-04 ± 1.65E-02	4.88	0.40	12.25
HIST1H1B	NP_064418.1	histone 1, H1b	242	5	14	147	3	16	134	3	14	1.86E-03 ± 1.75E-03	1.26E-03 ± 9.31E-04	9.81E-04 ± 8.86E-02	0.68	0.53	1.28
HIST1H1C	NP_056601.1	histone 1, H1c	366	7	20	455	8	27	372	7	23	4.09E-03 ± 1.96E-03	1.69E-02 ± 8.45E-03	8.52E-03 ± 1.15E+00	4.13	2.08	1.98
HIST1H1E	NP_056602.1	histone 1, H1e	333	6	16	394	7	28	328	6	19	2.67E-03 ± 1.31E-03	1.30E-02 ± 6.31E-03	6.29E-03 ± 8.91E-01	4.88	2.35	2.07
HIST1H1T	NP_034507.1	histone 1, H1t	225	5	15	214	4	11	253	5	13	2.11E-03 ± 1.13E-03	2.32E-03 ± 1.34E-03	1.16E-03 ± 3.36E-02	1.10	0.55	2.00
HIST1H2BN	NP_075911.2	histone 1, H2bc	346	5	27	589	10	52	224	4	27	9.85E-03 ± 1.82E-03	3.31E-02 ± 7.85E-03	1.07E-02 ± 3.25E-01	3.36	1.09	3.09
HIST1H4K	NP_835583.1	histone 1, H4k	682	10	51	839	15	65	593	10	59	3.71E-02 ± 7.16E-03	6.41E-02 ± 5.36E-03	4.63E-02 ± 3.27E+00	1.73	1.25	1.39
BAT1	NP_062667.1	HLA-B-associated transcript 1A	387	9	23	102	2	6	313	7	13	6.45E-04 ± 4.40E-04	1.77E-03 ± 9.07E-04	4.67E-04 ± 9.96E-03	2.75	0.73	3.79
-	NP_001020559.1	hypothetical protein LOC433182	1291	18	61	1383	19	55	1274	18	62	1.39E-02 ± 1.18E-03	1.42E-02 ± 3.11E-03	1.72E-02 ± 3.60E-01	1.03	1.24	0.83
-	NP_079681.2	hypothetical protein LOC66184	515	10	32	170	4	14	380	7	26	1.73E-03 ± 5.37E-04	1.05E-03 ± 4.36E-04	9.59E-04 ± 1.07E-02	0.61	0.56	1.09
-	NP_075692.1	hypothetical protein LOC66422	108	1	11	79	1	14	136	2	14	3.89E-04 ± -	1.90E-04 ± -	2.46E-04 ± 0.00E+00	0.49	0.63	0.77
-	NP_080804.1	hypothetical protein LOC68045	151	2	15	185	2	12	174	2	15	1.36E-04 ± -	3.53E-04 ± 3.82E-04	2.86E-04 ± -	2.60	2.11	1.23
-	NP_598568.1	hypothetical protein LOC98238	511	8	36	338	6	27	339	6	26	1.20E-03 ± 3.29E-04	1.44E-03 ± 4.86E-04	1.18E-03 ± 3.18E-02	1.19	0.98	1.22
IMPDH2	NP_035960.2	inosine 5'-phosphate dehydrogenase 2	505	8	17	348	6	18	161	3	9	3.01E-04 ± 1.77E-04	2.93E-04 ± 3.46E-04	2.53E-04 ± 1.64E-02	0.98	0.84	1.16
ITPA	NP_080198.2	inosine triphosphatase	269	4	38	107	2	18	210	3	27	3.90E-04 ± 9.65E-05	9.79E-04 ± -	4.46E-04 ± 2.67E-02	2.51	1.14	2.19
IGF2BP1	NP_034081.1	insulin-like growth factor 2, binding protein 1	277	5	13	279	5	14	231	4	11	1.19E-03 ± 8.89E-04	4.48E-04 ± 2.91E-04	4.27E-04 ± 1.01E-02	0.38	0.36	1.05
IQGAP1	NP_057930.1	IQ motif containing GTPase activating protein 1	169	4	4	85	1	2	439	9	9	4.99E-05 ± -	3.25E-05 ± 2.25E-05	5.05E-05 ± 8.42E-04	0.65	1.01	0.64
IDH2	NP_766599.1	isocitrate dehydrogenase 2 (NADP+), mitochondrial	96	2	6	134	3	6	228	5	13	9.15E-05 ± 7.77E-05	2.86E-04 ± 3.03E-04	1.85E-04 ± 0.00E+00	3.12	2.02	1.54
IARS	NP_742012.1	isoleucine-tRNA synthetase	189	4	5	70	1	2	181	4	4	1.14E-04 ± 1.00E-04	2.56E-05 ± -	1.66E-05 ± 7.82E-04	0.23	0.15	1.54
KPNB1	NP_032405.3	karyopherin (importin) beta 1	478	8	13	164	2	3	940	14	24	2.90E-04 ± 2.25E-04	2.77E-04 ± 2.35E-04	1.06E-03 ± 2.58E-02	0.95	3.65	0.26
KHSRP	NP_034743.2	KH-type splicing regulatory protein	223	3	6	95	2	4	109	2	4	3.36E-04 ± 1.62E-04	4.77E-04 ± -	1.42E-03 ± 2.19E-02	1.42	4.22	0.34
KIF5B	NP_032474.2	kinesin family member 5B	150	4	5	205	4	8	295	5	7	5.16E-05 ± -	7.83E-05 ± 7.75E-05	2.90E-05 ± 0.00E+00	1.52	0.56	2.70
LDHA	NP_034829.1	lactate dehydrogenase 1, A chain	624	9	27	137	3	9	600	11	25	2.74E-03 ± 1.03E-03	4.86E-04 ± 2.92E-04	3.31E-03 ± 1.15E-01	0.18	1.21	0.15
LMNB1	NP_034851.1	lamin B1	719	11	22	195	3	7	369	6	19	9.38E-04 ± 2.78E-04	4.94E-04 ± 1.98E-04	1.58E-04 ± 8.99E-03	0.53	0.17	3.12
LAMB1	NP_032508.1	laminin B1 subunit 1	196	3	3	2193	33	25	553	8	8	2.11E-05 ± 1.38E-05	2.87E-03 ± 1.93E-03	2.08E-04 ± 9.64E-03	136.33	9.88	13.80
LAMA1	NP_032506.1	laminin, alpha 1	81	2	1	1738	30	14	157	2	1	2.15E-05 ± -	7.82E-04 ± 1.44E-04	3.17E-05 ± 0.00E+00	36.43	1.48	24.68
LAP3	NP_077754.2	leucine aminopeptidase 3	238	3	10	220	4	11	368	5	15	6.27E-04 ± 9.74E-05	1.18E-03 ± 7.92E-04	7.71E-04 ± 1.73E-02	1.89	1.23	1.53
LRPPRC	NP_082509.1	leucine-rich PPR motif-containing protein	274	7	6	59	1	1	299	6	7	4.82E-04 ± -	2.47E-05 ± -	7.05E-04 ± 9.80E-03	0.05	1.46	0.04
LRPAP1	NP_038615.2	low density lipoprotein receptor-related protein associated protein 1	91	1	4	166	3	7	94	1	4	2.53E-04 ± 1.63E-04	5.38E-04 ± 4.11E-04	1.74E-04 ± 2.74E-03	2.13	0.69	3.09
LUC7L2	NP_619621.1	LUC7-like 2	107	2	6	81	1	4	105	1	4	3.38E-04 ± -	1.57E-03 ± 1.28E-03	9.73E-04 ± 3.51E-02	4.64	2.88	1.61
MIF	NP_034928.1	macrophage migration inhibitory factor	146	2	17	64	1	8	72	1	8	8.16E-04 ± 3.00E-04	3.79E-03 ± 5.96E-04	3.64E-04 ± 1.72E-02	4.65	0.45	10.41
MDH2	NP_032643.2	malate dehydrogenase 2, NAD (mitochondrial)	531	8	34	510	7	32	407	6	26	2.79E-03 ± 8.24E-04	1.21E-03 ± 1.26E-03	2.74E-03 ± 2.93E-02	0.43	0.98	0.44
ME2	NP_663469.1	malic enzyme 2, NAD(+)-dependent, mitochondrial	54	1	4	51	1	4	73	1	4	2.81E-05 ± -	1.65E-04 ± 0.00E+00	4.74E-05 ± 3.35E-03	5.85	1.69	3.47
MATR3	NP_034901.2	matrin 3	103	2	3	235	5	11	153	3	6	1.57E-04 ± 1.94E-04	2.29E-04 ± 1.08E-04	2.31E-04 ± -	1.46	1.48	0.99
MTHFD1	NP_620084.1	methylenetetrahydrofolate dehydrogenase 1	892	13	18	132	2	3	780	13	18	1.17E-03 ± 2.43E-04	2.07E-04 ± -	5.47E-04 ± 1.85E-02	0.18	0.47	0.38
MTAP	NP_077753.1	methylthioadenosine phosphorylase	78	1	6	137	2	10	125	2	10	2.92E-04 ± 1.75E-04	4.57E-04 ± 1.61E-04	2.96E-04 ± 1.30E-02	1.56	1.01	1.54
MCM2	NP_032590.1	minichromosome maintenance deficient 2 mitotin	91	3	3	53	1	1	239	5	9	1.28E-04 ± -	1.43E-04 ± -	1.08E-04 ± 1.54E-03	1.12	0.84	1.32
MCM3	NP_032589.1	minichromosome maintenance deficient 3	340	7	11	162	2	5	253	5	10	2.34E-04 ± 2.45E-04	1.19E-04 ± -	8.02E-05 ± 8.12E-03	0.51	0.34	1.49
PCK2	NP_083270.1	mitochondrial phosphoenolpyruvate carboxykinase 2	87	1	3	336	4	11	56	1	2	6.62E-05 ± 5.17E-05	2.74E-04 ± 1.83E-04	1.46E-04 ± -	4.15	2.21	1.87
MSH6	NP_034960.1	mutS homolog 6	281	4	5	123	2	2	492	9	10	1.58E-04 ± 6.89E-05	9.51E-05 ± -	5.48E-05 ± 1.57E-03	0.60	0.35	1.74
MYH10	NP_780469.1	myosin heavy chain 10, non-muscle	995	15	12	584	9	7	1030	15	13	1.48E-04 ± 4.61E-05	2.29E-04 ± 1.63E-05	1.37E-04 ± 8.43E-03	1.55	0.92	1.68
MYH9	NP_071855.1	myosin, heavy polypeptide 9, non-muscle isoform 1	1533	19	15	2442	35	21	1592	18	14	1.86E-04 ± 6.38E-05	1.17E-03 ± 1.78E-04	4.53E-04 ± 1.76E-02	6.30	2.44	2.58
ASAH1	NP_062708.1	N-acylsphingosine amidohydrolase (acid ceramidase) 1	104	2	8	138	2	9	74	1	5	1.26E-04 ± -	4.10E-04 ± -	1.42E-04 ± -	3.25	1.12	2.89
NDUFA10	NP_077159.1	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 10	132	2	9	109	2	10	133	1	6	4.20E-04 ± -	2.73E-04 ± -	1.57E-04 ± -	0.65	0.37	1.74
NDUFS1	NP_663493.1	NADH dehydrogenase (ubiquinone) Fe-S protein 1	182	3	7	51	1	3	185	4	9	7.97E-05 ± 1.61E-05	8.89E-05 ± -	1.28E-04 ± 1.11E-03	1.12	1.61	0.69
NMT1	NP_032733.1	N-myristoyltransferase 1	107	2	4	92	2	5	160	4	11	6.67E-05 ± -	1.30E-04 ± -	8.44E-05 ± -	1.95	1.26	1.54
NSUN2	NP_663329.2	NOL1/NOP2/Sun domain family 2	117	3	6	83	2	4	137	3	7	1.44E-04 ± -	1.87E-04 ± -	1.62E-04 ± -	1.30	1.12	1.16
NASP	NP_058057.2	nuclear autoantigenic sperm protein	825	14	31	151	4	10	464	7	14	1.32E-03 ± 6.98E-04	2.93E-04 ± -	9.57E-04 ± 1.10E-02	0.22	0.72	0.31
NUMA1	NP_598708.2	nuclear mitotic apparatus protein 1	106	1	1	121	1	1	68	1	1	4.74E-05 ± 2.09E-05	1.54E-05 ± -	6.67E-06 ± -	0.33	0.14	2.31
NCL	NP_035010.3	nucleolin	1096	20	27	1088	18	25	1000	18	25	5.44E-03 ± 1.25E-03	2.89E-03 ± 5.05E-04	4.66E-03 ± 1.18E-01	0.53	0.86	0.62

Proteins in common with embryonic stem cells, early primitive ectoderm-like cells and embryoid bodies, continued.

			Embryonic stem cells			Early Primitive Ectoderm-like			Embryoid Bodies								
Gene	NCBI Accession	Protein Description	Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Embryonic stem cells AVG N_SC ± STDEV	Early primitive ectoderm-like cells AVG N_SC ± STDEV	Embryoid Bodies AVG N_SC ± STDEV	EPL/ES Ratio	EB/ES Ratio	EPL/EB Ratio
NPM1	NP_032748.1	nucleophosmin 1	763	12	36	315	6	23	783	11	34	7.39E-03 ± 7.55E-04	4.37E-03 ± 1.64E-03	4.32E-03 ± 5.16E-02	0.59	0.58	1.01
NUP133	NP_758492.1	nucleoporin 133	60	1	1	61	1	2	236	4	6	1.43E-05 ± 0.00E+00	2.80E-05 ± -	4.83E-05 ± 1.71E-03	1.95	3.37	0.58
NME1	NP_032730.1	nucleoside-diphosphate kinase 1	415	8	57	286	7	47	296	5	38	1.78E-03 ± 1.10E-03	1.70E-03 ± 1.18E-03	1.16E-03 ± 1.91E-02	0.96	0.65	1.46
NME2	NP_032731.1	nucleoside-diphosphate kinase 2	416	7	57	397	10	61	391	7	57	1.45E-03 ± 9.26E-04	2.13E-03 ± 1.12E-03	1.10E-03 ± 4.59E-02	1.46	0.76	1.93
NSFL1C	NP_938085.1	p47 protein	344	5	23	83	1	6	79	1	6	5.64E-04 ± 1.03E-04	8.68E-05 ± -	1.38E-04 ± 1.15E-02	0.15	0.24	0.63
PSPC1	NP_079958.2	paraspeckle protein 1	66	1	4	67	1	4	234	5	13	6.33E-05 ± -	1.24E-04 ± -	9.79E-05 ± 7.70E-03	1.95	1.55	1.26
PPIA	NP_032933.1	peptidylprolyl isomerase A	750	12	74	461	8	50	518	9	60	1.06E-02 ± 5.53E-03	1.62E-02 ± 5.68E-03	1.12E-02 ± 2.53E-01	1.53	1.06	1.44
PPIB	NP_035279.2	peptidylprolyl isomerase B	100	2	7	181	4	19	216	5	23	5.11E-04 ± 3.62E-04	1.99E-03 ± 9.61E-04	1.12E-03 ± 6.14E-02	3.90	2.19	1.78
PPID	NP_080628.1	peptidylprolyl isomerase D	184	3	10	177	2	9	172	3	10	4.47E-04 ± -	3.06E-04 ± 6.17E-05	1.89E-04 ± 5.33E-03	0.68	0.42	1.62
PRDX1	NP_035164.1	peroxiredoxin 1	404	8	43	280	5	25	382	9	54	1.39E-03 ± 4.19E-04	1.70E-03 ± 1.15E-04	1.33E-03 ± 1.40E-02	1.23	0.96	1.28
PRDX6	NP_031479.1	peroxiredoxin 6	389	5	35	87	1	9	320	5	34	8.62E-04 ± 3.72E-04	6.49E-04 ± 7.14E-04	7.27E-04 ± 2.59E-02	0.75	0.84	0.89
FARSLB	NP_035941.2	phenylalanine-tRNA synthetase-like, beta subunit	335	6	12	59	1	3	196	3	7	2.44E-04 ± 1.17E-04	5.48E-05 ± 0.00E+00	7.11E-05 ± 0.00E+00	0.23	0.29	0.77
PEBP1	NP_061346.2	phosphatidylethanolamine binding protein 1	433	7	68	141	2	11	343	5	57	2.01E-03 ± 2.56E-04	3.97E-03 ± 4.89E-03	1.72E-03 ± 3.42E-02	1.98	0.86	2.31
PGAM1	NP_075907.1	phosphoglycerate mutase 1	368	7	36	137	3	21	244	4	20	1.27E-02 ± 2.70E-03	2.16E-03 ± 2.34E-03	7.95E-03 ± 4.96E-02	0.17	0.63	0.27
PAICS	NP_080215.1	phosphoribosylaminoimidazole carboxylase	230	4	14	128	2	7	253	5	15	3.63E-04 ± 2.47E-04	1.90E-04 ± 1.61E-04	4.27E-04 ± 2.48E-02	0.52	1.17	0.45
GART	NP_034386.1	phosphoribosylglycinamide formyltransferase	622	9	17	243	3	7	436	7	13	1.09E-04 ± 1.33E-04	1.39E-04 ± 3.69E-05	2.81E-04 ± 2.84E-02	1.27	2.57	0.49
PABPC1	NP_032800.2	poly A binding protein, cytoplasmic 1	747	11	24	256	6	12	319	5	11	1.01E-03 ± 6.55E-05	1.52E-04 ± 5.08E-05	2.56E-04 ± 2.20E-02	0.15	0.25	0.60
PCBP1	NP_035995.1	poly(rC) binding protein 1	396	7	27	82	2	8	186	3	12	1.60E-03 ± 7.84E-04	4.54E-04 ± -	3.27E-04 ± 2.26E-03	0.28	0.20	1.39
PCBP2	NP_035172.1	poly(rC) binding protein 2	238	4	14	131	2	8	123	2	8	4.43E-04 ± 4.48E-04	4.32E-04 ± 2.14E-04	1.20E-04 ± 4.00E-03	0.98	0.27	3.60
PTBP1	NP_032982.1	polypyrimidine tract binding protein 1	437	5	20	321	4	13	426	5	17	3.56E-04 ± 2.52E-04	6.95E-04 ± 3.08E-04	3.27E-04 ± 1.92E-02	1.95	0.92	2.13
FLNB	XP_995248.1	PREDICTED: filamin B, beta isoform 6	708	14	9	1592	28	21	545	9	7	1.25E-04 ± 1.32E-05	3.56E-04 ± 2.16E-04	7.33E-05 ± 2.03E-03	2.84	0.59	4.85
-	XP_001052621.1	PREDICTED: guanosine diphosphate (GDP) dissociation inhibitor 2 isoform 1	558	7	22	390	5	16	766	15	56	6.20E-04 ± 5.52E-04	7.11E-04 ± 8.94E-04	1.11E-03 ± 4.97E-02	1.15	1.80	0.64
-	XP_919960.2	PREDICTED: hypothetical protein LOC381591	563	8	13	172	2	4	578	8	15	7.43E-04 ± 9.72E-05	8.28E-05 ± 0.00E+00	4.18E-04 ± 2.97E-02	0.11	0.56	0.20
-	XP_985520.1	PREDICTED: similar to 40S ribosomal protein S17	342	6	51	90	1	15	122	1	15	3.85E-03 ± 5.67E-04	2.21E-04 ± -	2.39E-03 ± 9.22E-02	0.06	0.62	0.09
-	XP_921519.1	PREDICTED: similar to 40S ribosomal protein S28	81	1	17	112	2	23	77	1	17	8.88E-03 ± 2.77E-03	6.09E-03 ± -	2.16E-03 ± 5.09E-02	0.69	0.24	2.82
-	XP_987138.1	PREDICTED: similar to 40S ribosomal protein S7 (S8)	273	6	36	499	8	38	228	5	30	3.07E-03 ± 1.43E-03	3.22E-03 ± 8.21E-04	2.01E-03 ± 4.38E-02	1.05	0.66	1.60
-	XP_996602.1	PREDICTED: similar to 60S acidic ribosomal protein P0 (L10E)	405	6	29	114	2	9	374	6	26	9.05E-04 ± 4.50E-04	1.07E-03 ± 6.48E-04	1.04E-03 ± 2.69E-02	1.18	1.15	1.03
-	XP_001003945.1	PREDICTED: similar to 60S ribosomal protein L12	411	6	42	399	8	47	322	5	38	4.45E-03 ± 1.40E-03	3.66E-03 ± 2.95E-03	3.68E-03 ± 5.53E-02	0.82	0.83	0.99
-	XP_989377.1	PREDICTED: similar to 60S ribosomal protein L23a	282	6	35	183	3	20	176	3	18	1.89E-03 ± 9.21E-04	8.34E-04 ± 2.08E-04	1.65E-03 ± 1.11E-01	0.44	0.88	0.50
-	XP_911243.1	PREDICTED: similar to TAX-responsive enhancer element binding protein 107	490	9	29	354	7	20	449	8	26	1.88E-03 ± 7.75E-04	1.75E-03 ± 6.17E-04	3.18E-03 ± 2.07E-01	0.93	1.69	0.55
-	XP_001003284.1	PREDICTED: similar to 60S ribosomal protein L7a (Surfeit locus protein 3)	624	12	35	419	9	33	474	8	30	2.05E-03 ± 5.53E-04	1.17E-03 ± 2.53E-04	4.36E-03 ± 5.47E-01	0.57	2.12	0.27
-	XP_980537.1	PREDICTED: similar to acidic nuclear phosphoprotein 32 family, member B	280	5	13	51	1	4	83	1	4	3.45E-04 ± 3.51E-05	1.19E-04 ± -	1.03E-04 ± -	0.34	0.30	1.16
-	XP_899571.1	PREDICTED: similar to chromobox homolog 3	172	4	33	156	4	32	222	4	30	8.14E-04 ± -	1.41E-03 ± -	6.86E-04 ± 4.04E-02	1.73	0.84	2.06
-	XP_001006329.1	PREDICTED: similar to cytoplasmic beta-actin	345	6	36	320	7	44	454	7	36	2.35E-03 ± 7.60E-04	3.62E-03 ± 1.72E-03	9.84E-03 ± 2.98E-01	1.54	4.19	0.37
-	XP_001003118.1	PREDICTED: similar to electron transferring flavoprotein, beta polypeptide	279	6	26	205	4	22	82	1	8	9.23E-04 ± 1.87E-04	4.50E-04 ± 4.55E-04	3.89E-04 ± 1.67E-02	0.49	0.42	1.16
-	XP_913060.1	PREDICTED: similar to Filamin-A isoform 19	538	10	7	482	8	4	521	9	6	5.63E-05 ± 1.25E-05	1.34E-04 ± 1.04E-04	6.85E-05 ± 3.69E-03	2.39	1.22	1.96
-	XP_999161.1	PREDICTED: similar to Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	286	6	21	53	1	7	151	5	21	3.98E-04 ± 2.26E-04	3.10E-04 ± 1.08E-04	2.55E-04 ± 8.87E-05	0.78	0.64	3.65
-	XP_998642.1	PREDICTED: similar to Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	552	9	31	345	6	30	329	6	25	4.66E-03 ± 1.10E-03	5.22E-03 ± 1.44E-03	5.04E-03 ± 9.44E-02	1.12	1.08	1.04
-	XP_993038.1	PREDICTED: similar to Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	856	13	46	636	10	46	797	13	46	8.48E-03 ± 3.33E-03	8.08E-03 ± 3.30E-03	7.81E-03 ± 1.21E-01	0.95	0.92	1.03
-	XP_912024.1	PREDICTED: similar to Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	540	7	30	422	6	35	511	8	32	5.30E-03 ± 1.45E-03	9.64E-03 ± 2.06E-03	6.29E-03 ± 1.24E-01	1.82	1.19	1.53
-	XP_990083.1	PREDICTED: similar to GTP-binding nuclear protein Ran (GTPase Ran) (Ras-like protein TC4)	214	4	22	56	1	5	229	5	25	9.71E-04 ± 4.22E-04	4.49E-04 ± 0.00E+00	7.97E-04 ± 3.19E-02	0.46	0.82	0.56
-	XP_992835.1	PREDICTED: similar to H2A histone family, member Z	109	2	16	62	1	12	76	1	12	2.83E-04 ± 2.00E-04	8.28E-04 ± -	1.19E-04 ± 0.00E+00	2.93	0.42	6.94
-	XP_902036.1	PREDICTED: similar to H3 histone, family 3B isoform 3	256	6	26	342	9	38	194	5	26	1.09E-02 ± 1.86E-03	2.79E-02 ± 1.85E-02	1.08E-02 ± 5.72E-01	2.56	0.99	2.59
-	XP_001001311.1	PREDICTED: similar to heterogeneous nuclear ribonucleoprotein A0	443	6	21	126	2	8	382	6	22	1.25E-03 ± 1.96E-04	1.77E-04 ± 6.11E-05	8.70E-04 ± 2.86E-02	0.14	0.70	0.20
-	XP_893918.1	PREDICTED: similar to heterogeneous nuclear ribonucleoprotein K	305	5	37	173	2	16	375	6	44	1.94E-03 ± 7.36E-04	2.69E-03 ± 2.95E-03	3.23E-03 ± 1.54E-01	1.39	1.66	0.83
-	XP_987047.1	PREDICTED: similar to Importin alpha-2 subunit	412	7	18	53	1	3	474	7	18	2.50E-04 ± 1.43E-04	1.22E-04 ± -	7.04E-04 ± 2.13E-02	0.49	2.81	0.17
-	XP_995699.1	PREDICTED: similar to Leukotriene A-4 hydrolase (LTA-4 hydrolase) (Leukotriene A(4) hydrolase)	104	1	3	115	1	3	234	3	8	6.04E-05 ± -	1.18E-04 ± -	6.96E-04 ± 4.88E-02	1.95	11.53	0.17
-	XP_987060.1	PREDICTED: similar to NADH-ubiquinone oxidoreductase 24 kDa subunit, mitochondrial precursor isoform 2	64	1	4	56	1	8	135	2	9	2.34E-04 ± 4.72E-05	1.30E-04 ± -	2.53E-04 ± 1.99E-02	0.56	1.08	0.51
-	XP_999910.1	PREDICTED: similar to NonO/p54nrb homolog isoform 3	391	8	14	85	1	3	210	4	5	3.65E-04 ± 3.05E-04	2.33E-04 ± 1.90E-04	1.48E-04 ± 6.73E-03	0.64	0.41	1.57
-	XP_991486.1	PREDICTED: similar to OTU domain, ubiquitin aldehyde binding 1 isoform 2	75	1	6	84	1	7	90	1	6	6.11E-05 ± 0.00E+00	1.19E-04 ± -	2.58E-04 ± -	1.95	4.22	0.46
-	XP_485239.1	PREDICTED: similar to phosphoglycerate kinase 1	820	14	43	800	11	35	798	13	41	2.16E-03 ± 6.33E-04	2.89E-03 ± 2.66E-03	2.29E-03 ± 5.88E-02	1.34	1.06	1.26
-	XP_999777.1	PREDICTED: similar to Phosphoglycerate mutase 1	559	8	42	181	3	21	452	6	26	1.39E-02 ± 3.23E-03	1.27E-03 ± 1.21E-03	8.37E-03 ± 5.82E-02	0.09	0.60	0.15
-	XP_913594.1	PREDICTED: similar to phosphoribosylformylglycinamidine synthase	199	4	6	60	1	1	134	3	4	4.54E-05 ± 3.78E-05	2.42E-05 ± -	8.35E-05 ± 2.95E-03	0.53	1.84	0.29
-	XP_979065.1	PREDICTED: similar to prohibitin	141	2	11	384	6	26	162	4	25	1.83E-04 ± 1.61E-04	8.31E-04 ± 1.68E-04	7.70E-04 ± -	4.55	4.22	1.08
-	XP_912973.1	PREDICTED: similar to protein disulfide isomerase-associated 6 isoform 2	532	6	24	588	9	29	489	6	24	2.55E-03 ± 4.34E-04	2.97E-03 ± 1.36E-03	1.44E-03 ± 4.82E-02	1.16	0.56	2.07
-	XP_999518.1	PREDICTED: similar to ribosomal protein L18a	83	1	8	55	1	8	95	1	8	1.48E-03 ± 5.46E-05	3.69E-04 ± -	6.38E-03 ± 1.80E-01	0.25	4.31	0.06
-	XP_001000091.1	PREDICTED: similar to ribosomal protein L30	189	3	35	154	2	24	288	4	40	4.27E-03 ± 2.05E-03	1.50E-03 ± 1.44E-03	7.73E-03 ± 5.47E-02	0.35	1.81	0.19
-	XP_988610.1	PREDICTED: similar to ribosomal protein S12	363	7	41	252	5	45	203	4	33	2.34E-03 ± 1.67E-03	2.94E-03 ± 1.73E-03	8.11E-04 ± 2.66E-02	1.25	0.35	3.62
-	XP_894950.1	PREDICTED: similar to ribosomal protein S13	322	6	31	183	4	21	185	4	29	2.03E-03 ± 5.43E-04	4.99E-04 ± 1.24E-04	1.23E-03 ± 3.50E-02	0.25	0.61	0.40
-	XP_915990.1	PREDICTED: similar to ribosomal protein S27a	73	1	38	153	2	43	81	1	38	2.04E-02 ± 3.35E-03	9.49E-03 ± 1.13E-02	1.19E-02 ± 1.26E-01	0.47	0.58	0.80
-	XP_992209.1	PREDICTED: similar to SGT1, suppressor of G2 allele of SKP1	91	1	4	112	1	4	103	1	4	2.08E-04 ± 5.88E-05	1.44E-04 ± 1.13E-04	2.34E-05 ± -	0.69	0.11	6.17
-	XP_979873.1	PREDICTED: similar to splicing factor, arginine/serine-rich 10 (transformer 2 homolog, Drosophila)	110	2	10	139	2	13	71	1	6	1.15E-04 ± -	1.87E-03 ± 1.77E-03	9.69E-05 ± -	16.26	0.84	19.28
-	XP_998748.1	PREDICTED: similar to transgelin 2	281	4	24	573	7	38	161	2	13	4.62E-04 ± 5.15E-04	1.33E-03 ± 5.80E-04	3.34E-04 ± 0.00E+00	2.88	0.72	3.99
-	XP_899768.1	PREDICTED: similar to ubiquitin A-52 residue ribosomal protein fusion product 1	240	4	33	265	5	33	244	4	33	7.90E-03 ± 9.33E-04	3.21E-03 ± 3.51E-03	5.73E-03 ± 3.76E-02	0.41	0.72	0.56
P4HA1	NP_035160.1	procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), alpha 1 polypeptide	67	1	3	537	8	14	158	2	6	3.10E-05 ± -	7.26E-04 ± 5.38E-04	3.92E-05 ± 1.85E-03	23.42	1.26	18.51
PFN1	NP_035202.1	profilin 1	264	4	42	276	5	43	158	2	24	5.04E-03 ± 2.12E-03	4.38E-03 ± 1.63E-03	3.76E-03 ± 1.62E-01	0.87	0.74	1.17
PDCD5	NP_062720.1	programmed cell death 5	169	3	29	72	1	10	57	1	10	2.63E-04 ± 1.86E-04	5.13E-04 ± -	1.11E-04 ± -	1.95	0.42	4.63
PDCD6IP	NP_035182.1	programmed cell death 6 interacting protein	224	3	6	211	4	6	100	2	5	1.24E-04 ± 1.21E-04	9.29E-05 ± 2.63E-05	4.82E-05 ± 2.27E-03	0.75	0.39	1.93
D11 (includes EG:22	NP_035183.2	programmed cell death protein 11	208	4	3	138	5	4	153	4	3	5.04E-05 ± 3.37E-05	8.67E-05 ± -	1.87E-05 ± 1.59E-03	1.72	0.37	4.63
PA2G4	NP_035249.1	proliferation-associated 2G4	365	7	26	122	2	6	242	3	13	1.29E-03 ± 3.77E-04	3.28E-04 ± -	2.50E-03 ± 2.84E-02	0.25	1.94	0.13
P4HB	NP_035162.1	prolyl 4-hydroxylase, beta polypeptide	523	11	26	1184	18	42	446	8	16	6.50E-04 ± 1.30E-04	6.09E-03 ± 4.40E-03	9.96E-04 ± 2.09E-02	9.37	1.53	6.11
PSMC5	NP_032976.1	protease (prosome, macropain) 26S subunit, ATPase 5	310	5	16	324	5	20	413	7	23	2.45E-04 ± 1.15E-04	5.30E-04 ± 1.66E-04	4.13E-04 ± 9.72E-03	2.17	1.69	1.29
PSMC2	NP_035318.1	proteasome (prosome, macropain) 26S subunit, ATPase 2	286	4	16	173	4	13	336	5	16	3.83E-04 ± 1.39E-04	1.36E-04 ± -	1.57E-04 ± 6.12E-03	0.35	0.41	0.87
PSMC3	NP_032974.1	proteasome (prosome, macropain) 26S subunit, ATPase 3	284	6	20	61	1	4	96	2	6	1.87E-04 ± 1.50E-04	7.31E-05 ± -	3.16E-05 ± -	0.39	0.17	2.31
PSMD7	NP_034947.1	proteasome (prosome, macropain) 26S subunit, non-ATPase, 7	151	2	13	82	1	5	145	2	12	1.03E-04 ± 0.00E+00	1.51E-04 ± 7.12E-05	2.39E-04 ± 1.54E-02	1.46	2.32	0.63
PSMA7	NP_036099.1	proteasome (prosome, macropain) subunit, alpha type 7	243	3	19	181	2	12	331	5	31	4.23E-04 ± 3.85E-05	1.43E-03 ± -	9.57E-04 ± 1.69E-02	3.39	2.26	1.50
PSMB1	NP_035315.1	proteasome (prosome, macropain) subunit, beta type 1	121	2	16	67	1	14	140	4	22	1.61E-04 ± 1.05E-04	1.35E-04 ± -	5.82E-04 ± -	0.84	3.61	0.23
PSMC6	NP_080235.2	proteasome 26S ATPase subunit 6	138	2	7	289	4	10	149	2	7	1.91E-04 ± 1.50E-04	3.74E-04 ± 2.94E-04	5.02E-04 ± 2.54E-02	1.95	2.62	0.74
PSMD1	NP_081633.1	proteasome 26S non-ATPase subunit 1	439	8	14	59	1	1	395	6	12	2.49E-04 ± 1.87E-04	1.02E-04 ± -	4.83E-04 ± 8.91E-03	0.41	1.94	0.21
PSMD11	NP_848731.1	proteasome 26S non-ATPase subunit 11	354	5	15	57	1	3	342	5	15	5.88E-04 ± 1.71E-04	3.83E-04 ± -	6.28E-04 ± 2.01E-02	0.65	1.07	0.61
PDIA3	NP_031978.1	protein disulfide isomerase associated 3	479	8	20	1083	19	42	312	7	16	9.29E-04 ± 4.18E-04	3.82E-03 ± 3.52E-04	3.59E-04 ± 7.31E-03	4.11	0.39	10.62

Proteins in common with embryonic stem cells, early primitive ectoderm-like cells and embryoid bodies, continued.

			Embryonic stem cells			Early Primitive Ectoderm-like			Embryoid Bodies								
Gene	NCBI Accession	Protein Description	Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Total Mascot Score	Total No. Peptides	Total % Protein Coverage	Embryonic stem cells AVG N_SC ± STDEV	Early primitive ectoderm-like cells AVG N_SC ± STDEV	Embryoid Bodies AVG N_SC ± STDEV	EPL/ES Ratio	EB/ES Ratio	EPL/EB Ratio
PDIA6	NP_082235.1	protein disulfide isomerase-associated 6	523	7	27	558	10	32	605	8	28	7.81E-04 ± 2.81E-04	2.15E-03 ± 1.42E-03	9.93E-04 ± 2.54E-02	2.76	1.27	2.17
PKM2	NP_035229.2	pyruvate kinase 3	1682	28	57	1271	22	45	1552	25	56	6.08E-03 ± 4.32E-04	9.41E-03 ± 2.13E-03	1.27E-02 ± 8.28E-01	1.55	2.09	0.74
RDX	NP_033067.1	radixin	381	5	9	128	2	3	144	3	7	1.70E-04 ± 5.68E-05	5.54E-05 ± -	1.68E-04 ± 6.77E-03	0.33	0.98	0.33
RANBP2	NP_035370.1	RAN binding protein 2	290	9	3	55	1	1	220	5	2	4.16E-05 ± 4.38E-05	1.06E-05 ± -	3.66E-05 ± 1.94E-03	0.25	0.88	0.29
ARHGDIA	NP_598557.3	Rho GDP dissociation inhibitor (GDI) alpha	153	2	20	135	2	16	202	2	20	1.76E-03 ± 3.83E-04	1.90E-03 ± -	2.36E-03 ± 5.32E-02	1.08	1.34	0.80
RNH1	NP_660117.2	ribonuclease/angiogenin inhibitor 1	91	1	4	110	2	9	126	2	8	1.81E-04 ± 7.26E-05	3.54E-04 ± 3.01E-04	9.18E-05 ± 8.66E-03	1.95	0.51	3.86
RPL10	NP_443067.1	ribosomal protein 10	381	8	43	51	1	13	302	6	36	1.29E-03 ± 3.57E-04	3.02E-04 ± -	3.91E-04 ± 1.73E-02	0.23	0.30	0.77
RPL10A	NP_035417.1	ribosomal protein L10A	422	7	31	287	6	29	285	5	24	3.64E-03 ± 1.04E-03	1.30E-02 ± 7.93E-03	9.11E-03 ± 1.69E-01	3.58	2.51	1.43
RPL11	NP_080195.1	ribosomal protein L11	290	4	21	316	7	38	277	4	28	5.24E-03 ± 2.09E-03	2.12E-03 ± 1.03E-03	1.91E-03 ± 4.32E-02	0.40	0.36	1.11
RPL14	NP_080250.1	ribosomal protein L14	226	3	12	218	3	12	264	4	17	3.46E-03 ± 9.66E-04	5.71E-03 ± 8.47E-03	7.50E-04 ± 4.52E-02	1.65	0.22	7.60
RPL17	NP_001002239.1	ribosomal protein L17	307	5	27	153	3	16	170	3	22	1.03E-02 ± 0.00E+00	2.40E-03 ± 2.12E-03	3.92E-03 ± 1.78E-01	0.23	0.38	0.61
RPL18A	NP_084027.1	Ribosomal protein L18A	224	4	24	90	2	12	248	4	24	4.11E-03 ± 1.46E-03	1.10E-03 ± -	9.28E-03 ± 4.38E-01	0.27	2.26	0.12
RPL22	NP_033105.1	ribosomal protein L22	274	4	51	133	2	22	205	3	41	2.46E-03 ± 2.14E-03	2.94E-03 ± 2.70E-03	5.09E-03 ± 5.60E-02	1.20	2.07	0.58
RPL23	NP_075029.1	ribosomal protein L23	274	6	56	163	3	31	227	5	39	3.23E-03 ± 4.78E-04	1.73E-03 ± 1.63E-04	3.09E-03 ± 3.59E-02	0.54	0.96	0.56
RPL35	NP_079868.1	ribosomal protein L35	65	1	11	94	2	16	59	1	11	8.07E-04 ± -	1.58E-03 ± -	1.13E-04 ± -	1.95	0.14	13.89
RPL37A	NP_033110.1	ribosomal protein L37a	124	2	29	135	2	29	79	1	20	7.08E-03 ± 4.22E-03	4.92E-03 ± 1.99E-03	4.10E-03 ± 2.15E-02	0.69	0.58	1.20
RPL4	NP_077174.1	ribosomal protein L4	776	14	31	449	10	24	661	11	25	1.72E-03 ± 2.38E-04	1.44E-03 ± 1.02E-03	1.64E-03 ± 6.81E-02	0.83	0.95	0.88
RPL7	NP_035421.2	ribosomal protein L7	334	8	27	148	4	17	195	3	13	9.20E-04 ± 2.67E-04	4.79E-04 ± -	3.36E-04 ± 4.02E-02	0.52	0.37	1.42
RPL8	NP_036183.1	ribosomal protein L8	485	8	28	325	5	19	438	7	31	3.03E-03 ± 7.59E-04	8.80E-04 ± 1.78E-04	1.65E-03 ± 4.08E-02	0.29	0.54	0.53
RPS11	NP_038753.1	ribosomal protein S11	201	4	25	139	2	16	204	4	26	4.36E-03 ± 9.51E-04	7.22E-03 ± 1.59E-03	5.21E-03 ± 1.83E-01	1.65	1.19	1.39
RPS16	NP_038675.1	ribosomal protein S16	214	4	22	95	2	8	207	4	22	2.02E-03 ± 8.87E-04	1.11E-03 ± -	1.32E-03 ± 3.64E-02	0.55	0.65	0.85
RPS3	NP_036182.1	ribosomal protein S3	497	9	47	241	6	28	454	7	39	2.66E-03 ± 1.23E-03	1.13E-03 ± 1.03E-03	2.41E-03 ± 9.24E-02	0.43	0.91	0.47
RPS5	NP_033121.1	ribosomal protein S5	249	3	17	119	2	16	111	1	10	1.46E-03 ± 2.81E-04	1.79E-03 ± 1.42E-03	2.10E-03 ± 9.97E-02	1.23	1.44	0.86
RPS8	NP_033124.1	ribosomal protein S8	416	7	41	267	5	35	342	6	34	3.08E-03 ± 1.39E-03	1.92E-03 ± 1.19E-03	1.70E-03 ± 2.79E-02	0.62	0.55	1.13
RPLP2	NP_080296.3	ribosomal protein, large P2	664	8	74	263	4	70	700	9	74	1.32E-02 ± 7.82E-03	4.40E-03 ± 4.48E-03	8.17E-03 ± 2.96E-01	0.33	0.62	0.54
RPS3A	NP_058655.2	ribosomal protein, large P2	388	8	32	388	7	30	288	5	23	5.75E-03 ± 3.41E-03	2.94E-03 ± 2.60E-03	4.44E-03 ± 2.11E-01	0.51	0.77	0.66
RPLP1	NP_061341.1	ribosomal protein, large, P1	234	3	57	149	2	56	222	3	57	3.82E-03 ± 2.08E-03	3.40E-03 ± 2.80E-03	7.75E-04 ± 2.55E-02	0.89	0.20	4.38
SUB1	NP_035424.1	RNA polymerase II transcriptional coactivator	64	1	9	54	1	9	63	1	9	1.78E-03 ± 1.81E-03	2.54E-04 ± -	1.91E-03 ± 6.44E-02	0.14	1.07	0.13
RNPC2	NP_573505.1	RNA-binding region containing protein 2	203	3	8	57	1	3	180	3	8	1.98E-04 ± 1.60E-04	6.09E-05 ± -	3.16E-04 ± -	0.31	1.60	0.19
SCFD1	NP_084101.1	sec1 family domain containing 1	196	3	7	61	1	2	325	4	11	1.38E-04 ± 5.39E-05	1.01E-04 ± -	1.24E-04 ± 5.04E-03	0.73	0.90	0.82
SEPHS1	NP_780609.3	selenophosphate synthetase 1	127	2	9	87	1	5	84	1	5	1.69E-04 ± 5.97E-05	4.94E-04 ± -	1.60E-04 ± 7.55E-03	2.93	0.95	3.09
2-Sep	NP_035021.1	septin 2	116	2	9	61	1	6	98	1	4	8.25E-04 ± 2.59E-04	5.37E-04 ± -	3.13E-03 ± 5.41E-02	0.65	3.79	0.17
SERPINH1	NP_033955.1	serine (or cysteine) proteinase inhibitor, clade H, member 1	475	7	22	1231	19	49	486	8	24	5.16E-04 ± 1.73E-04	5.84E-03 ± 2.78E-03	4.80E-04 ± 1.02E-02	11.31	0.93	12.16
SUGT1	NP_080750.1	SGT1, suppressor of G2 allele of SKP1	269	3	15	112	1	7	246	3	15	6.08E-04 ± 1.50E-04	3.20E-04 ± 3.09E-04	1.66E-04 ± 4.15E-03	0.53	0.27	1.93
SLC25A5	NP_031477.1	solute carrier family 25, member 5	111	2	13	78	2	13	124	2	8	1.06E-03 ± 5.64E-04	7.26E-03 ± -	9.13E-04 ± 9.94E-03	6.88	0.87	7.95
SLC3A2	NP_032603.2	solute carrier family 3 (activators of dibasic and neutral amino acid transport), member 2	110	2	7	77	1	4	173	3	10	9.44E-05 ± 8.90E-05	2.46E-04 ± -	3.98E-05 ± 1.88E-03	2.60	0.42	6.17
SPTBN1	NP_033286.1	spectrin beta 2 isoform 2	246	6	4	188	4	3	208	6	3	3.84E-05 ± 1.09E-05	4.50E-05 ± -	1.73E-05 ± 9.90E-04	1.17	0.45	2.60
SKP1A	NP_035673.2	S-phase kinase-associated protein 1A	298	4	29	85	1	10	234	3	19	8.46E-03 ± 5.19E-03	8.92E-04 ± 9.81E-04	4.77E-03 ± 1.72E-01	0.11	0.56	0.19
SND1	NP_062750.2	staphylococcal nuclease domain containing 1	106	2	3	89	1	2	427	7	15	1.00E-04 ± 3.86E-05	5.32E-05 ± 2.51E-05	2.81E-04 ± 2.22E-02	0.53	2.81	0.19
STOML2	NP_075720.1	stomatin-like protein 2	181	2	11	100	1	6	124	1	6	2.34E-04 ± -	1.83E-04 ± -	1.58E-04 ± -	0.78	0.67	1.16
STIP1	NP_058017.1	stress-induced phosphoprotein 1	555	10	22	214	5	13	298	6	13	9.65E-04 ± 4.58E-04	5.95E-05 ± 0.00E+00	1.97E-04 ± 1.21E-02	0.06	0.20	0.30
SOD1	NP_035564.1	superoxide dismutase 1, soluble	332	6	34	132	3	30	265	4	32	4.66E-03 ± 1.96E-03	1.15E-03 ± 7.42E-04	4.68E-03 ± 3.39E-01	0.25	1.01	0.25
ST13	NP_598487.1	suppression of tumorigenicity 13	307	5	18	198	3	8	190	3	11	4.61E-04 ± 2.46E-04	1.74E-04 ± 1.23E-04	2.63E-04 ± 1.30E-02	0.38	0.57	0.66
SMARCA4	NP_035547.2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4	163	2	2	71	1	1	263	3	3	2.39E-05 ± 1.57E-05	2.00E-05 ± -	8.65E-05 ± 2.29E-03	0.84	3.61	0.23
SMARCA5	NP_444354.2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 5	190	2	3	70	1	1	162	3	3	1.73E-04 ± 1.36E-04	1.08E-04 ± 2.17E-05	9.30E-05 ± 3.51E-03	0.62	0.54	1.16
TLN1	NP_035732.1	talin 1	520	8	5	660	12	9	217	4	2	5.21E-05 ± 1.30E-05	1.74E-04 ± 7.00E-05	1.65E-05 ± -	3.33	0.32	10.54
TARDBP	NP_663531.1	TAR DNA binding protein isoform 1	138	2	9	120	3	15	360	6	27	4.00E-04 ± -	3.12E-04 ± -	5.17E-04 ± 1.28E-02	0.78	1.29	0.60
TCP1	NP_038714.1	t-complex protein 1	184	3	9	307	6	15	294	5	13	2.98E-04 ± 1.30E-04	5.52E-04 ± 6.98E-04	3.43E-04 ± 1.82E-02	1.85	1.15	1.61
THOP1	NP_073144.2	thimet oligopeptidase 1	113	2	4	80	2	4	92	2	4	4.82E-05 ± 0.00E+00	4.70E-04 ± -	6.10E-05 ± 2.87E-03	9.76	1.26	7.71
TXN	NP_035790.1	thioredoxin 1	349	6	71	173	4	46	349	4	60	1.12E-02 ± 2.89E-03	2.77E-03 ± 8.70E-04	2.21E-02 ± 5.72E-01	0.25	1.98	0.13
THOC4	NP_035698.1	THO complex 4	113	1	7	55	1	7	75	1	7	2.60E-04 ± -	5.07E-04 ± -	1.09E-04 ± -	1.95	0.42	4.63
TTN	NP_035782.3	titin isoform N2-A	210	8	0	151	6	0	260	10	0	7.75E-06 ± 1.87E-06	1.54E-05 ± 8.19E-06	6.67E-06 ± 2.89E-04	1.99	0.86	2.31
TRAP1	NP_080784.1	TNF receptor-associated protein 1	367	6	12	98	1	2	357	5	11	1.43E-03 ± 4.45E-04	1.83E-04 ± 4.58E-05	1.58E-03 ± 7.13E-02	0.13	1.11	0.12
TKT	NP_033414.1	transketolase	376	7	16	322	6	13	197	5	16	3.81E-04 ± 1.01E-04	3.28E-04 ± 2.56E-04	2.17E-04 ± 6.47E-03	0.86	0.57	1.52
TPR	NP_598541.2	translocated promoter region protein	397	8	5	171	5	3	226	4	3	5.62E-05 ± 1.86E-05	6.17E-05 ± 6.78E-05	3.55E-05 ± 4.19E-03	1.10	0.63	1.74
TMEM4	NP_064337.1	transmembrane protein 4	64	1	8	107	2	15	86	1	8	1.82E-04 ± -	5.68E-03 ± -	2.30E-04 ± 2.03E-02	31.23	1.26	24.68
TPI1	NP_033441.1	triosephosphate isomerase 1	393	7	46	445	6	37	346	6	42	1.88E-03 ± 2.77E-04	3.03E-03 ± 8.34E-04	1.01E-03 ± 4.59E-02	1.61	0.54	3.00
TRIM28	NP_035718.1	tripartite motif protein 28	931	15	25	313	6	12	795	13	25	1.86E-03 ± 6.79E-04	2.58E-04 ± 1.83E-04	1.91E-03 ± 7.75E-02	0.14	1.03	0.13
TPM3	NP_071709.2	tropomyosin 3, gamma	138	2	7	117	2	11	119	2	7	1.46E-04 ± 4.12E-05	1.71E-04 ± 8.04E-05	7.37E-05 ± 3.48E-03	1.17	0.51	2.31
TPM4	NP_001001491.1	tropomyosin 4	96	1	7	429	8	43	81	1	7	1.33E-04 ± 1.16E-04	5.64E-04 ± 7.52E-05	2.81E-04 ± -	4.23	2.11	2.01
TUFM	NP_766333.1	Tu translation elongation factor, mitochondrial	295	5	19	56	1	3	202	3	9	2.81E-04 ± 3.59E-04	7.15E-05 ± -	2.16E-04 ± 8.17E-03	0.25	0.77	0.33
TUBA2	NP_035784.1	tubulin, alpha 2	942	14	47	725	12	40	1030	14	49	7.79E-03 ± 8.50E-04	5.44E-03 ± 3.06E-03	9.99E-03 ± 2.92E-01	0.70	1.28	0.55
TUBB3	NP_075768.1	tubulin, beta 3	821	13	34	382	5	13	765	12	33	4.35E-03 ± 1.64E-03	1.48E-03 ± 1.14E-03	2.98E-03 ± 1.52E-01	0.34	0.68	0.50
TUBB5	NP_035785.1	tubulin, beta 5	1612	24	76	1079	15	54	1647	24	75	8.80E-03 ± 2.45E-03	4.97E-03 ± 4.84E-04	8.72E-03 ± 1.19E-01	0.57	0.99	0.57
TUBB2	NP_666228.1	tubulin, beta, 2	1569	24	72	842	12	46	1335	21	69	6.32E-03 ± 1.37E-03	4.48E-03 ± 5.45E-04	6.57E-03 ± 1.48E-01	0.71	1.04	0.68
TPT1	NP_033455.1	tumor protein, translationally-controlled 1	270	3	17	247	3	17	281	3	17	3.95E-03 ± 6.94E-04	2.88E-03 ± 2.01E-03	9.68E-03 ± 2.54E-01	0.73	2.45	0.30
HSP90B1	NP_035761.1	tumor rejection antigen gp96	496	8	11	736	14	23	640	10	14	5.71E-04 ± 2.94E-04	2.69E-03 ± 1.04E-03	1.89E-03 ± 2.04E-01	4.70	3.31	1.42
YWHAE	NP_033562.2	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, epsilon polypeptide	492	8	32	443	6	27	654	10	35	8.91E-03 ± 1.52E-03	2.70E-03 ± 1.02E-03	1.35E-02 ± 6.15E-02	0.30	1.51	0.20
YWHAH	NP_035868.1	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, eta polypeptide	497	8	34	199	4	15	484	8	36	2.22E-03 ± 8.59E-04	1.23E-03 ± 5.92E-04	2.84E-03 ± 1.80E-01	0.55	1.28	0.43
YWHAQ	NP_035869.1	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, theta polypeptide	383	7	27	129	3	11	575	9	40	2.39E-03 ± 9.07E-04	9.23E-04 ± 2.64E-04	7.77E-03 ± 9.81E-02	0.39	3.25	0.12
YWHAZ	NP_035870.1	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide	749	11	51	400	6	35	743	11	47	5.63E-03 ± 1.97E-03	1.36E-03 ± 5.95E-04	8.96E-03 ± 2.08E-01	0.24	1.59	0.15
U2AF2	NP_598432.1	U2 small nuclear ribonucleoprotein auxiliary factor (U2AF) 2	244	4	18	101	2	13	333	5	22	8.65E-04 ± 4.22E-04	3.69E-04 ± 2.24E-04	1.00E-03 ± 2.74E-02	0.43	1.16	0.37
UQCRH	NP_079917.1	ubiquinol-cytochrome c reductase hinge protein	80	1	20	99	1	20	93	1	20	5.58E-04 ± 0.00E+00	3.63E-04 ± -	4.70E-04 ± -	0.65	0.84	0.77
UBE1	NP_033483.1	ubiquitin-activating enzyme E1, Chr X	512	9	12	360	5	8	771	13	17	4.22E-04 ± 1.34E-04	1.63E-04 ± 9.33E-05	4.13E-04 ± 2.71E-02	0.39	0.98	0.39
UBE2L3	NP_033482.1	ubiquitin-conjugating enzyme E2L 3	113	2	24	59	1	21	88	1	14	3.94E-04 ± 2.24E-04	2.10E-04 ± 0.00E+00	2.72E-04 ± -	0.53	0.69	0.77
UBA2	NP_057891.1	ubiquitin-like 1 (sentrin) activating enzyme E1B	101	1	2	87	1	2	262	4	10	1.17E-04 ± 1.83E-05	2.03E-04 ± -	5.36E-04 ± 1.39E-02	1.73	4.59	0.38
SAE1	NP_062722.1	ubiquitin-like 1 (sentrin) activating enzyme subunit 1	332	5	19	177	3	12	292	4	18	2.05E-04 ± 1.66E-04	2.31E-04 ± 1.96E-04	3.06E-04 ± 1.15E-02	1.13	1.49	0.75
UMPS	NP_033497.1	uridine monophosphate synthetase	164	3	10	80	1	3	255	4	9	3.10E-04 ± -	6.72E-05 ± 0.00E+00	1.31E-04 ± 1.44E-02	0.22	0.42	0.51
VDAC1	NP_035824.1	voltage-dependent anion channel 1	318	5	32	574	9	49	652	9	49	7.99E-04 ± 2.43E-04	3.54E-03 ± 1.19E-03	6.89E-03 ± 7.53E-03	4.43	8.62	0.51
XPNPEP1	NP_573479.2	X-prolyl aminopeptidase (aminopeptidase P) 1, soluble	61	1	3	89	2	4	173	3	8	6.64E-05 ± 1.88E-05	1.56E-04 ± -	8.96E-05 ± 8.08E-03	2.34	1.35	1.74

CONCLUSIONS

The overall purpose of this work was to improve and develop methods for quantitative proteomics with the goal to quantitatively compare the proteomes of differentiating embryonic stem cells.

Chapter 3: Stable isotopic labeling by tryptic incorporation of two ^{18}O atoms into the C-termini of peptides is one of the more popular techniques used for comparative proteomic studies that elect to use a stable isotopic labeling approach. We investigated reasons for loss of one or two ^{18}O labels and found that the major source of the phenomenon is enzymatically driven. We explored ways to eliminate this event, concluding that complete removal of trypsin from solution is the best approach to solve this dilemma.

Chapter 4: We report that ^{18}O incorporation facilitated by trypsin leads to false identification of sites of N-linked glycosylation during an ^{18}O N-linked glycosylated site mapping strategy. We proved that tryptic incorporation into the C-termini of peptides was the major cause of falsely identified sites when using this technique. We investigated ways to prevent this problem, concluding that trypsin must be either 1) removed completely from solution before release of the glycan or 2) the sample may be dried after glycan release and resuspended in ^{16}O water for trypsin-facilitated removal of the ^{18}O atoms. This study has been helpful in mapping sites of N-linked glycosylation for embryonic stem cells.

Chapter 5: A new method for stable isotopic labeling was developed. This method was driven by the observation that dissolving samples in urea causes partial carbamylation of primary amines. We developed a protocol that promoted full carbamylation and used this protocol with stable isotopic derivatives of urea to universally label peptides. This method had a disadvantage in that the intensity of carbamylated lysines had a reduced intensity during MALDI-TOF analysis. This method may prove useful for difficult to dissolve samples where the use of urea (and thus partial carbamylation) is unavoidable.

Chapter 6: We analyzed embryonic stem cells, early primitive ectoderm-like cells, and embryoid bodies by two dimensional reverse phase, quantifying protein expression by spectral count. Spectral counting, a label free method, was used instead of stable isotopic labeling due to the simplicity in obtaining and reporting this quantitative figure of merit. The quantitative proteomic study of

differentiating stem cell identified 1498 proteins that were distributed among the three populations. Approximately 22% of this population was composed of predicted and hypothetical proteins. Another 15% of all proteins were up or down regulated more that three fold, as determined by non-parametric statistical methods. Proteins identified in the comparative or single populations could be linked to embryogenetic events. Regulation of identified proteins was comparable versus protein expression that had been previously studied by knockout or *in situ* hybridization techniques. The observations on the regulation and functionality of unique and shared populations have indicated several candidates and topics useful for future embryogenetic studies.