DEVELOPMENT OF BIOMARKERS FOR HUMAN EMBRYONIC STEM CELL-DERIVED NEURAL PROGENITOR CELLS AND THEIR DERIVATIVES

by

BRIAN ANDREW GERWE

(Under the Direction of Steven L. Stice)

ABSTRACT

The directed differentiation and isolation of human embryonic stem cell (hESC)-derived astrocytes would provide a cell source to further study a wide range of neurological disorders. Originally, morphology was used to distinguish astrocytes; however, the high content of intermediate filaments, particularly glial fibrillary acidic protein (GFAP), is now used to distinguish astrocytes. Herein, we study the temporal expression of GFAP, a developmental glial biomarker, to further elucidate the differentiation process of human neural progenitor (hNP) cells towards an astrocyte-like cell population. However, the temporal onset of neurogenesis precedes gliogenesis and is therefore difficult to study in culture. Moreover, the in vitro isolation of astrocytes from hESC sources still remains elusive but, once accomplished, would provide further characterization of astrocyte-mediated diseases. The current study utilizes lentiviral technology to stably integrate a tissue-specific reporter construct into hNP cells, thus enabling selective monitoring for the appearance of glial-like cells in differentiation cultures. The study demonstrated a dramatic transcriptional up regulation in GFAP expression when differentiation medium was supplemented with five and ten percent fetal bovine serum (FBS), illustrating a

clear treatment effect. In addition, the scope of biomarker development was broadened with the application of high-throughput proteomics encompassing membrane fractionation and liquid-chromatography tandem mass spectrometry (LC-MS/MS) analysis to address custom biomarker screening (e.g. hESC line distinction). The membrane fraction of the human derived cell lines WA09, BG01, WA09-derived neural progenitor (hNP) cells and abBG02 (trisomy 12, 14, 17 and an extra copy of X chromosome) was successfully scanned, identifying cell surface markers unique to each line. This provided an important proof of concept, demonstrating the potential ability to delineate differentiated hESC lines and their derivatives. Cell surface marker discrimination based on high-throughput proteomics may provide a highly efficient technique to produce homogeneous populations. Therefore, it is of great importance to obtain new biomarkers that lead to basic research and clinical breakthroughs, characterization of disease pathologies and the development of novel diagnostics tools.

INDEX WORDS: GFAP, Neural Stem Cell, Neural Progenitor Cells, Differentiation, In Vitro Culture, Lentivirus, Biomarker, Proteomics

DEVELOPMENT OF BIOMARKERS FOR HUMAN EMBRYONIC STEM CELL-DERIVED NEURAL PROGENITOR CELLS AND THEIR DERIVATIVES

by

BRIAN ANDREW GERWE

B.S., The University of Georgia, 2002

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

DOCTOR OF PHILSOLOHY

ATHENS, GEORGIA

2008

DEVELOPMENT OF BIOMARKERS FOR HUMAN EMBRYONIC STEM CELL-DERIVED NEURAL PROGENITOR CELLS AND THEIR DERIVATIVES

by

BRIAN ANDREW GERWE

Major Professor: Steven L. Stice

Committee: Stephen Dalton
James D. Lauderdale
Alan E. Przybyla

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2008

DEDICATION

I would like to dedicate this to my wife for her endless support and understanding including her patience and tolerance of my moodiness throughout my years of schooling. I will always be grateful for her company and assistance while performing assays and for proof-reading my grammatically poor dissertation. We were happily married on May 19, 2007 and it has been an exciting and fulfilling journey ever since.

I am truly grateful for my loving parents who have supported my decisions throughout the years, taken pride in my accomplishments and believed in me through all my endeavors. I would also like to thank my brother for being more than a brother but a best friend as well.

ACKNOWLEDGEMENTS

I would to like to extend my deepest gratitude to Dr. Steven L. Stice for his guidance and continued support as my major professor during my graduate schooling tenure. He made the transition to his lab seamless and personally rewarding. I would like to thank the members of my committee, Dr. Stephen Dalton, Dr. James D. Lauderdale and Dr. Alan E. Przybyla for their expert assistance, insightful suggestions and endless motivation. Additionally, I would like to thank Dr. Michael W.W. Adams, Dr. Francis E. Jenney, Dr. Frank J. Sugar and Dr. Angeli Lal Menon for their guidance and instruction during my residence in the Adams' lab.

To the Stice lab, I am very fortunate and grateful to work and associate with great scientists within the lab and remarkable people outside the work environment. It goes without saying that friendship is one of the truly undervalued concepts, but one of the most appreciated when it is needed. Thus, I am indebted to my close friends from childhood, to my college roommates and friends that became a second family to me - now referred to collectively as "the Group,"- and to all the lab members and people I have grown to know during my undergraduate and graduate tenure at The University of Georgia.

TABLE OF CONTENTS

Chapter 1

Introduction and Literature Review

Publication of the first whole genome sequence of *Haemophilus influenzae* by The Institute for Genomic Research (TIGR) in 1995 spawned the genomic era of science. Yet, the current “post-genomic” era may yield equally exciting developments and discoveries in the areas of biological and medical research. The wealth and magnitude of information generated from these organism-wide comparisons present a unique opportunity of discovery for biomarker identification of specific cell types, diseased or altered tissue states, and possibly functional (gene or protein) states.

The term biomarker is officially defined by the National Institute of Health as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" (1). Interestingly, it is from this relatively vague definition that individual research areas have further defined the term to suit a particular field of study. Herein, we define biomarker, within the context of an indicator of normal biological processes, as a molecule that allows for the detection and isolation of a particular cell type (e.g. glial fibillary acidic protein [GFAP] is commonly used as a biomarker to identify the glial lineage, more specifically astrocytes).

In a research context, biomarker identification serves to further characterize and discern cell lineages and their functional capacities; however, these biomarkers are rarely clinically applicable. The clinical definition states that the marker is indicative for

the presence of a disease (30). Unfortunately, clinical biomarkers are often difficult to discover, as in the case for neurological diseases. Major challenges include the availability and quality of diseased brain tissue. In most cases, tissue samples are obtained post-mortem and are low quality samples from the large affected regions and not from brain biopsies due to cost and the risk involved. In addition, there are poor clinical diagnostic exams to determine extent or progression of a particular disease until prognosis (30). Therefore, it is of great importance to obtain new biomarkers that lead to basic research and clinical breakthroughs enabling the further characterization of disease pathology and, ultimately, leading to the development of novel methods of diagnostics and/ or therapeutics.

Neural Development and Network Formation

The neural plate is the developmental precursor of the nervous system and is formed from the developing epiblast. During the third week of development, gastrulation takes place with the formation of the primitive streak on the dorsal side of the epiblast (55). Cells from the epiblast migrate through the primitive streak forming the intraembyronic mesoderm and endoderm. The remaining cells in the epiblast become the ectoderm, while those migrating laterally through the primitive node develop into the notochord.

Signaling molecules from the bone morphogenic protein (BMP) family specify the ectoderm to become epidermis (for review (11)), consequently blocking the specification to the neuroectoderm. The dogma for specification of the neuroectoderm is that the absence of BMP signaling results in the ectoderm following a default pathway to the neuroectoderm lineage (68). Therefore, the presence of BMP signaling molecules

negatively influences the neural lineage specification through its interactions with smad proteins. Current evidence also supports positive inductive signaling of the neuroectoderm via correlation studies. It was demonstrated that diminished signaling through basic fibroblast growth factor (bFGF or FGF2) and Wnt (wingless) pathways prevented neurogenesis (105). Opponents argue that the involvement of bFGF and of Wnt in neurogenesis could be indirect measures of BMP signaling because of the number of pathways involving bFGF and Wnt signaling (86).

Once specified, the neuroectoderm thickens and forms the neural plate, thereby committing certain regions to particular parts of the nervous system. This is followed by primary and secondary neurulation, where the neural plate bends around the notochord fusing at the midline to form the neural tube. Fusion begins at the cervical regions and proceeds caudally and rostrally with complete fusion 7-10 days later (55). Upon complete caudal fusion, the cranial end proliferates and differentiates, eventually leading to the formation of the entire brain and the spinal cord. There are many intricacies involved in the formation of the (central) nervous system that were either glossed over or not mentioned because they are not directly relevant to the topic at hand. It should be noted that neural crest cells form in the region where the neural tube connects to the epidermis. It is these cells that migrate throughout the body to generate peripheral neurons and glia, pigment skin cells and several other cell types (40).

The cells proximal to the neural plate fuse, forming the neuroepithelium mediated by cell adhesion molecules. Mitotic proliferation and thickening of the neuroepithelium follows. Within the neuroepithelium, the cells display nuclear migration depending on the stage within the cell cycle; DNA synthesis occurs when the nucleus is distally

located to the lumen, near the basement membrane at the outer surface nuclei and migrates luminally as the cell undergoes mitosis (87). As the neuroepithelial cells exit the cell cycle and initiate differentiation, they move away from the lumen. Within the neural tube, two different cell types are present: neuroepithelial cells produce neurons while radial glia are both neurogenic and gliogenic in nature (48, 67). Additionally, the extended processes of radial glia appear to guide newly matured neurons away from the luminal region of the neural tube. The presence of several morphogens cause the wide diversity of functional neurons generated from the neuroepithelial cells. The morphogens elicit a concentration-dependent response in those target cells, whereby patterning the axes and the eventual developmental potential and ultimately the cells' fate (40). All neurons and glia within the central nervous system (CNS) are produced from the neuroepithelium of the neural tube; yet, the differentiation is not random (see Figure 1.1). Neuroepithelial cells must first proliferate to sustain normal development; otherwise, differentiation will produce too few neurons and glia. Neurogenesis is then initialed and then switched to glial differentiation through one or more mechanistic pathways not well established. One prevailing theory is that Notch signaling is directly involved in the inhibition of neurogenesis early in development, yet, initiates glial differentiation later in development.

Recent evidence indicates that the differentiation process is closely monitored through lateral inhibition mediated by Notch signaling (10, 56). Differentiated neurons leaving the neuroepithelium are expressing Notch ligand genes such as *Delta* and *Serrate/ Jagged* that bind to their specific receptors found on neuroepithelial cells, thus preventing further neuronal differentiation of the remaining neuroepithelial cells. Chitnis

and colleagues (1995) reported that ectopic expression of Notch signal activation inhibits the production of primary neurons in *Xenopus*, while its inhibition leads to the loss of the undifferentiated neuroepithelial cells and the over expression of primary neurons. It was concluded that the inability to maintain an undifferentiated pool of neuroepithelial cells, indirectly, leads to the loss of glial cells through the depletion of the necessary number of "neural stem cells" needed to produce neurons and glial cells (29). Furthermore, Beckers and colleagues (1999) detected strong *Delta1* mRNA levels in specific regions of the developing brain and spinal cord within mice; more specifically regions that are now known to contain subpopulations of neural stem cells, such as the subventricular zone (16). These findings are similar to those found in chicken (44) and zebrafish (5).

Furthermore, the role of Notch signaling has been directly linked to vertebrate gliogenesis. Nye *et al*. (1994) provided in vitro evidence that Notch signaling could inhibit neuronal differentiation on embryonic carcinoma cell line P19. The researchers determined that under specific conditions, the cells could differentiate into neurons, astrocytes and myoblasts; where upon Notch activation, the embryonic carcinoma cells were differentiated into astrocytes (71). Moreover, GFAP expression is increased when forced activation of Notch occurs in the dorsal regions of the neuroepithelium (95). Interestingly, Notch signaling only promotes astrocyte differentiation and not oligodendrocyte differentiation from the glial lineage.

When immature oligodendrocyte precursor cells (OPCs) were purified from rat optic nerve and exposed to Notch ligands, they remained undifferentiated as compared to the controls (103). However, several groups have reported that the same pMN

domain of the neuroepithelium that generates motor neurons will generate oligodendrocytes (77, 109). Park and Appel (2003) demonstrated that premature neurogenesis is caused by the inhibition of Notch signaling and results in the depletion of neuroepithelial cells in mutant zebrafish. Yet, constitutive expression of Notch signaling increases OPCs from the same domain in correlation to the normal developmental period and place when oligodendrocytes are formed (72).

Astrocytes

Glial cells outnumber neurons 10:1 within the nervous system (3). The term glia originated from the Greek word for "glue" because the original discovery and characterization observed the enveloping of other cell bodies and neuronal axons within the nervous system. There are two major classes: microgila and macrogila, with macroglia further sub-divided into oligodenrocytes, schwann cells, radial glia and astrocytes. Astrocytes are restricted to the central nervous system (CNS), and are found in both white and gray matter despite variations in morphology such as distinctive "fibrous" or long unbranched cellular processes and "protoplasmic" that exhibit short and highly branched cellular processes (53). Originally, morphology was used to distinguish astrocytes; however, astrocytes are now characterized by their high content of intermediate filaments, particularly glial fibrillary acidic protein (GFAP) (53). Further studies determined that mature astrocytes express astrocyte-specific glutamate transporter GLAST, S100β and glutamate synthase (GS) with "reactive" or activated astrocytes expressing GLAST, S100β, GS, vimentin, and tenascin-C (41, 53, 99). It is interesting to note that while these markers are found on mature astrocytes, a multitude of them are also expressed in radial glial and type B cells (astrocytes within the adult

subventricular zone acting as neural stem cells) (8, 66). There is no molecule known thus far that would distinctly identify astrocytes across all species.

Astrocytes express numerous receptors enabling them to participate in a wide variety of signal processes in the CNS. They provide the structural support for the brain while maintaining chemical homeostasis throughout neurotransmission. To prevent the over stimulus of neurons during the firing of an action potential, astrocytes reuptake and release neurotransmitters from the synaptic zones, therefore allowing equilibrium to return to the synaptic cleft, preventing the over stimulus (saturation) of neurons during neurotransmission. The removal of several key neurotransmitters (e.g. Glutamate, ATP and GABA) is performed through the expression of their respective membrane transporters on the plasma membrane surface (31, 81). More recently, astrocytes were shown to release glutamate or ATP in a Ca^{2+}-dependent manner (83). In addition, astrocytes regulate the ion concentration within the extracellular space between neurons. During the repolarization phase after neurotransmission, potassium is released into the extracellular space. The high permeability of astrocytes to potassium permits its rapid clearance and prevents its excess accumulation (88, 92).

It has been well documented that an imbalance of the CNS causes astrocytes to increase their metabolic activity, as well as their production of growth and trophic factors (61). The elevated concentration of astrocyte-secreted factors promotes the return to homeostasis through the protection of brain cells from energy depletion, free reactive oxygen species and perhaps calcium saturation. Astrocytes establish and maintain the CNS boundaries, which include the blood–brain barrier. This barrier is accomplished through interactions with endothelial and meningeal cells. Upon trauma or insult to the

CNS, there is a breakdown of the blood-brain barrier enabling the translocation of plasma-derived molecules into the brain, such as activated immune cells, lymphocytes and macrophages (13). The final response is astrocyte activation. Upon injury, normal astrocytes undergo a cellular transformation and adopt a “reactive astrocyte” phenotype. This transformation increases cellular levels of GFAP that may extend far from the site of injury. The upregulation of the glial specific, predominately astrocytic, intermediate filament causes a proliferation event of more reactive astrocytes that will, ultimately, lead to glial scaring of the injured site. Furthermore, it is postulated that the reactive astrocytes establish a barrier between the healthy and damaged cells, thus, recapitulating the blood-brain barrier (13).

In addition to their role in homeostasis, astrocytes may aid in vasomodulation as intermediates for neuronal regulation of blood flow (73) and in the promotion of oligodendrocyte myelination (47). Ishibashi and colleagues (2006) demonstrated that the neuronal release of ATP during neurotransmission stimulates astrocytes to secrete leukemia inhibitory factor (LIF). The secretion of the cytokine, LIF, promoted the myelinating activity of oligodendrocytes, suggesting that astrocytes contribute to the myelination of the neurons within the CNS.

The integral involvement astrocytes have in CNS homeostasis illustrates their importance. It is no wonder that astrocytes have been implicated in the several neurological disorders such as Alzheimer's (37), Alexander's Disease (22, 79). Amyotrophic Lateral Sclerosis (ALS, Lou Gehrig's Disease) (33, 69), and the formation of specific brain tumors (e.g. Glioblastomas) (21).

Glial fibrillary acidic protein (GFAP) and Its Importance

The 50 kD cytoplasmic protein is expressed by the *Gfa* gene located on band q21 of chromosome 17 (46). It is a member of the intermediate filament family that provides support and strength to cells. Several molecules of the GFAP protein bind together to form the main intermediate filament of the cytoskeleton in radial glia and astrocytes in the CNS, and, to a lesser degree, in mueller cells of the retina and non-myelinating schwann cells of the peripheral nervous system (7, 101, 110). It has been well documented that an imbalance or trauma to the CNS elicits astrocytes to increase their metabolic activity, as well as their production of growth and trophic factors, including the production of glial fibrillary acidic protein (GFAP) (61).

The promoter region is located 2.2kB upstream from *Gfa* and is capable of directing astrocyte-specific expression. Within this region is a basal promoter consisting of a TATA box, a CAAT box, and several enhancer sequences (17). Initial characterization of hGFAP was in mice utilizing the bacterial *lacZ* reporter system (23).

Recent research has established that all three forms of Alexander disease (infantile, juvenile, and adult) are caused by mutations in the GFAP protein. Alexander disease is a rare, degenerative disorder, belonging to a group of disorders known as leukodystrophies, which affects the CNS stemming from the improper development of the myelin sheath growth (2). More than 20 mutations have been identified leading to single amino acid changes in the transcribed protein, most frequently mutations of arginine to cysteine or histidine at positions 79 and 239 (22, 79). The direct mechanism leading to Alexander disease still remains elusive, but it is speculated that mutations in the GFAP protein alter normal intermediate filament assembly, leading to an intra-

cellular aggregation of proteins within the cell, termed Rosenthal fibers (79). These fibers may interfere with normal interactions and/or prevent astrocytes from maintaining the blood-brain barrier. Early onset (6 months of age) of the disease is common and predominantly affects males with symptoms ranging from mental and physical retardation to dementia and seizures (2).

[Embryonic] Stem Cell (ESC) Derivation

"…The isolation of the first human ES (hES) cells (96) , rhesus monkey (97) and marmoset (98) was remarkably similar to the [mouse] ES protocol/media used 17 years previously. Within the same year that hESCs were isolated, another group derived pluripotent human embryonic germ (hEG) cell lines from fetal gonads 5-9 week post fertilization (90).

Regardless of species, ESC isolation requires the selection away from or removal of the trophectoderm (external cell layer surrounding the inner cell mass; ICM) that gives rise to a portion of placenta). The ICM is often plated on a mouse embryonic fibroblast feeder cell layer (mEFs) and allowed to grow and expand [through continuous passaging and subculture]. The doubling time for a hESC is between 35-40 hours (80).

The starting human embryos were donated by *in vitro* fertilization (IVF) clinics to the research community, and were either in excess of what was needed for implantation or were not of the quality necessary for implantation. However one can further delineate these excess embryos into yet two more categories. The first subgroup is the excess embryos that had formed adequate cell numbers in both the trophoblast and ICM. These higher quality embryos are usually frozen to be thawed later for transfer into the reproductive tract, donated to infertile couples or donated for research. The

second subgroup contains excess embryos determined not to be viable enough to either initiate a pregnancy or to freeze for later transfer. This second subgroup of non viable embryos is usually discarded. We were able to use these embryos to derive three of the most highly investigated embryonic stem cells on the NIH registry (BG01, 02 and 03; (62). No publications to date have shown that these hESC lines differ in their usefulness in regenerative therapies when compared to other hESC lines. For the list of NIH (National Institute of Health) approved ESC lines, see the NIH stem cell registry website (http://stemcells.nih.gov/research/registry/eligibilityCriteria.asp).

[Human Embryonic Stem Cell (hESC)] Chromosomal Stability and Gene Expression

Chromosomal number is a genetic difference that could lead to alterations in the genetic expression patterns of certain hESC lines. Karyotyping most often involves a procedure called G-banding to count chromosomal number and/or chromosomal rearrangements within a cell. In terms of the established hESC lines, it has been reported that lines can maintain a normal karyotype in continuous culture under certain conditions. hES1-6 cell lines have a normal karyotype (46 chromosomes, XY) from 24 – 140 passages of continuous culture (26), and H1, H7, and H9 cell lines from passage number 60 – 100 (80). However reports of abnormal karyotype similar to those found in embryonic carcinoma (EC) cell lines of trisomy 17 and 21 have recently surfaced in the BG01 and BG02 NIH-registered hESC lines after bulk hES cell passaging compromised the genetic integrity of cells as quickly as 23 – 25 passages after manual passaging techniques had ceased (63).

Anueoploidy has been associated with abnormal gene expression. Theoretically, this abnormal gene expression may be a risk factor when used in cell transplant studies. Therefore abnormal karyotype may up regulate genes associated with teratoma formation. Our lab found that manually passaged cells could maintain a normal karyotype but when using enzymatic or even non-enzymatic techniques over 20 passages trisomies at 12,17, 14, 20 and X in BG01 and BG02 lines eventually developed. Short-term bulk passaging, typical of methods used in other stem cell types including mouse ES cells, was successful in maintaining a normal karyotype ranging from 13 passages and 15 passages for the non-enzymatic method using cell dissociation buffer and collagenase/ trypsin, respectively (63). By quantitative real–time (RT-) PCR, the majority of over 80 genes associated with pluripotency and early differentiation were significantly upregulated in aneuploid cells (22). Brimble and colleagues (24) state that abnormal karyotypes, especially trisomy 17 and 21, can be acquired through various single cell disaggregating passage techniques via positively selecting for aneuploidy. They confirmed manually passaged BG01 and BG02 hESC lines maintained a long-term normal karyotype (only tested up to 52 passages for a time period of six months) (24). Together with our study, these reports suggest that bulk hESC passaging will compromise the genetic integrity of cells as quickly as 23 – 25 passages after manual passaging techniques have ceased.

With all the above issues surrounding each hES stem cell line, a standard protocol for the derivation, culture, maintenance and characterization is needed to ensure the validity of all lines and their uses. An international initiative named The International Stem Cell Forum (http://www.stemcellforum.org.uk) was founded in

January 2003 constituting 15 countries with the intention of establishing a set of standards for the characterization and culturing of all the reported hESC lines. Laboratories in 11 of the 15 participating countries contributed hESCs (totaling 75 different lines) for the purpose of flow cytometric analysis of 17 surface antigens, quantitative RT-PCR and microarray analysis of 100 selected gene transcripts, and epigenetic studies (9). To date (January 2006), there is not a specific gene or surface marker that can unanimously define ESC status, but rather only a series of trends in gene expression levels as an ESC begins to differentiate.

[Isolation of Human Embryonic Stem Cell (hESC)–derived Neural Progenitor (hNP) Cells]

Different strategies have been used to induce neural differentiation in [hESC]. The most routine method of generating differentiated cell types has been through the three-dimensional structure of the embryoid body (EB). Embryonic stem cells in this agglomerate start spontaneous differentiation to form a sphere in suspension culture. Differentiated EBs contain neural stem cells, the proportions of which increase with retinoic acid exposure. Reubinoff and colleagues cultured [hESC] until spontaneous differentiation occurred, then isolated a subpopulation to make neurospheres (76). Pera *et al*. (2004) introduced a BMP inhibitor of noggin into this prolonged culture system (74). However, EB culture has disadvantages compared with adherent culture in that phenotype observation within the sphere is not possible with standard microscopy. In addition, stochastic differentiation yielded multiple cell lineages and limited the overall yield of the desired cells. Ying and colleagues developed a monolayer differentiation method to obtain efficient neural induction of [mouse Embryonic Stem Cells (mESC)].

When differentiation was triggered by the withdrawal of [Leukemia inhibitory factor (LIF)], the [mESC] monolayer chose a neural fate in serum deprived medium (107). This efficient neural determination did not occur in medium containing serum. Though it was uncertain whether [hESC] would behave similarly to their mouse counterparts, adherent differentiation in defined culture is an attractive strategy. We developed an adherent [hESC] differentiation in defined culture conditions (91). The differentiation process was followed and examined immunohistochemically, suggesting a gradual but uniform differentiation to [hNP] cells when serum was removed and FGF2 and LIF were added. Derived [hNP] cells were characterised both genetically and by their differentiation capacity. Subsequently, [hNP] cells were propagated using defined medium for 6 months while maintaining their differentiation potential..."

(section from Stice, S. L., N. L. Boyd, S. K. Dhara, **B. A. Gerwe**, D. W. Machacek, and S. Shin. 2006. Human embryonic stem cells: challenges and opportunities. Reprod Fertil Dev 18:839-46) (94).

Mesenchymal Stem Cells

Human bone marrow-derived mesenchymal stem cells (hMSCs) are self-renewing, non-hematopoietic multipotent cells present in the bone marrow stroma that are capable of multilineage differentiation into osteoblasts (bone-forming), adipocytes (fat tissue) and chondrocytes (cartilage-producing) (18, 75). In addition, several researchers have recently demonstrated that MSCs can transdifferentiate into neuronal cells, schwann cell-like myelinating cells, tendocytes, and hepatocytes (12, 32, 52, 57).

Traditionally, hMSC are isolated from bone marrow but have also been recognized and/ or isolated from virtually all post-natal tissues including amniotic fluid,

placenta, umbilical cord blood, adipose tissue, liver, lung, muscle, and synovial membrane (14, 50, 59). Unfortunately, hMSC identification can not be accomplished through the presence or absence of a single cell surface marker, but rather via an array of them. Differences exist between publications and can be explained by variations in culture methods and/or differentiation stage of the cells. In general, the cells are positive for a variety of markers including Stro-1, CD29, CD 73, CD90, CD105, CD166 and CD44, and negative for the following panel of hematpoietic surface markers: CD14, CD19, CD31, CD34, CD45, CD79 and HLA DR (4, 34, 38, 54). Attempts have been made to enrich cultures for multipotent hMSCs utilizing cell surface antibodies in hopes of achieving a homogeneous culture with a defined phenotype. Initially, antibody STRO-1 was used to enrich for hMSCs targeting an uncharacterized cell surface epitope (42) or in combination with selected markers stated above, including SSEA-4 (39). Yet, researchers can not distinguish multipotent MSC from other cells present within the population, and more importantly, there are no significant reported differences between the biological characteristics of the isolated cell populations using these methods. Therefore, hMSCs can only be classified in functional terms based on in vitro (directed differentiation into derived cell lineages; e.g. osteoblasts, adipocytes and chondrocytes) and in vivo functional assays, such as ectopic bone formation in SCID (severe combined immunodeficiency disease) mice (49, 75).

Lentiviral Transduction

Viral vector technology as a means for gene therapy has been in development for over 20 years, but with little success until recently. In 1990, modified retroviral vectors were subjected to the first human clinical gene therapy trial to correct adenosine

deaminase (ADA) deficiency (19). In the study, white blood cells were isolated from two patients and transduced with murine leukemia virus (MLV)-based vector expressing ADA and a neomycin selection marker. Neomycin-resistant cells were sub-cultured and re-introduced back into the respective patients. Gene therapy was discontinued after 2 years, yet the integrated vector remained stable and heterologous expression of ADA continued within the T cells for at least two more years.

The first major success story was a clinical study in France (2000) using MLV-based vectors for the treatment of human severe combined immune deficiency (SCID)-X1 disease (27). Cavazzana-Calvo and colleagues used a modified and defective retrovirus to confer gene transfer of the γc protein. In patients, a mutated form of the protein found on the X chromosome is expressed and causes a cytokine receptor deficiency leading to an early block in T and NK lymphocyte differentiation and further maturation (27, 28). Unfortunately, it was discovered that two of the eleven children in the clinical study developed a leukemia-like condition after gene therapy; it should be noted that nine of the patients were cured of the SCID-X1 disease after treatment (28, 43). The onset of the leukemia-like condition was linked to the MLV-based vector inserting itself close to a known T-leukaemia oncogene in the abnormally proliferating T cells (43). Subsequently, it was established that murine retroviruses insert preferentially into promoter regions of active genes (15). With the recent advent of stem cell research and their potential therapeutic uses, viral vectors have again been employed as possible means of correcting mutated or deficient gene products and in select cases regulate the over-expression of others. There is no universal viral vector construct;

therefore, depending on the cell type and application, several viral vectors could be utilized.

Retroviruses were the first set of viruses posed for *in vivo* research, pioneering the way for viral vector technology currently used today. The key advantage to retroviral vector systems is stable integration of genetic cargo into host genome allowing for sustained heterologous gene expression over long periods (58). Additionally, retroviral vectors contain a medium packaging capacity (8-9kb) and do not transfer sequences that encode for viral packaging proteins, reducing the risk of a host immune response against the transduced cells. The initial limitations to the system included transduction/ infection of only dividing cells and several biosafety issues associated with virulence because of its derivation from oncoretroviruses (i.e. MLV) (114).

Simple retroviruses like MLV require cell division and nuclear membrane dissolution to integrate into the target cell genome. This requirement makes the introduction of useful expression vectors into quiescent cells quite challenging. To overcome these limitations, retroviruses were "retro-fitted" with envelope proteins from other viruses such as the G glycoprotein from vesicular stomatitis virus (VSV) (25), and therefore, permitted a broader host range than before, which included non-dividing cells. Further modifications included deleting several virulence causing genes, adding self-activating or self-inactivating sequences, host cell targeting via pseudotyping and the introduction of tissue-specific promoters (45).

Lentiviruses (*lenti*-, Latin for "slow") are cataloged within a genus of the *Retroviridae* family, characterized by the long replication cycle (see Figure 1.2). In contrast to conventional retroviruses, lentiviruses can infect non-dividing cells with a

broad range of hosts including neurons and glial cells in tissue culture and *in vivo* (20, 64, 70). Various promoters including human cytomegalovirus (CMV), human phosphoglycerate kinase (PGK) and the mouse albumin promoter have been introduced into the HIV-1-based vector. Current limitations to the lentiviral vector system are insufficient methods of generating high-titer virus stocks and the biosafety concerns stemming from the HIV-1 backbone.

A typical lentiviral particle is comprised of an RNA genome that carries *cis*-acting sequences necessary for packaging, reverse transcription and integration. It is the assembly of these components that lead to the budding virion from the producer cell and to the fusion and integration into the host cell. In general terms, the replication cycle of lentiviruses, including HIV-1, closely resembles that of other retroviruses. However, lentiviruses do encode a number of regulatory and accessory proteins not encoded by the genomes of the prototypical “simple” retroviruses.

The general replication cycle for all retroviruses, including lentiviruses (HIV-1) begins with the virion making contact with the host cell. The glycoproteins on the viral surface bind to specific host receptors, which is followed by fusion of the viral envelope with the host membrane leading to release of the viral core into the cell. Viral reverse transcriptase then uses the viral RNA as a template creating double-stranded cDNA. The cDNA integrates itself into the host genome via a viral integrase. At that point, the integrated DNA is now considered a provirus. Subsequent host RNA polymerase II transcribes the provirus into mRNA that later undergoes translation into viral proteins. The viral proteins and RNA self-assemble within the host to produce new virus particles primed for budding and further infection of a new cellular host (108).

In order to produce lentiviral stocks, it is a mandatory biosafety concern to avoid the production of replication-component recombinants. The issue being that in a retroviral genome, a single RNA molecule contains all critical cis-acting or coding sequences necessary for virion production and replication. Biosafety is then best achieved via distribution of the genome over several plasmids to maximize the number of complex recombination events in order to recapitulate a replication-component virus (114). The current (3rd) generation of lentiviral vectors segregates the coding sequences over three or four plasmids.

Along with replication of the virus, another biosafety concern is the virulence of the system. The very nature of HIV being a human pathogen comes from 9 of the 15 RNA-encoded genes contributing to its pathogenesis (114). Progression of lentiviral vector systems has led to the deletion of 6 of the 9 virulence RNA-encoding genes. Furthermore, the distribution of the remaining genes across several plasmids makes it virtually impossible to constitute the parental virus; thus, resulting in a level of biosafety that matches or exceeds that of viral vectors currently used for clinical use (114). The first generation HIV-1 vectors contained all genes encoding for HIV-1 proteins except Env (70). The second generation to be engineered further excluded *vif*, *vpr*, *vpu* and *nef* accessory genes conferring the majority of the virulence associated with HIV-1 (113, 114). The current version contains only the Gag, Pol and Rev proteins from the parent strain along with the introduction of replication-defective or (self-inactivating) aspects (36).

Self-inactivating vectors have long terminal repeats (LTRs) deleted in the U3 region. The U3 region is essential for the replication of the wild-type retrovirus, but

dispensable for the replication-defective vector. Therefore, self-inactivating vectors are safer than their counterparts with full-length LTRs (112). Unlike oncoretroviral and other retroviral vectors, self-inactivating (SIN) vectors usually have a weaker or cell-type specific internal promoter than the original viral promoter; therefore, they should give lower oncogene activation (111). Furthermore, HIV vectors were found to insert preferentially into gene ends rather than promoter regions (15) unlike MLV-based vector described previously. In addition to activation, it has been theorized that insertion would inactivate tumor suppressor genes. Addressing this issue is the observation that natural HIV infection does not directly cause HIV-related tumors (111). Additionally, within a tumor-prone, mutant mouse model, transduction of a haematopoietic stem cell line with a retroviral vector, but not with HIV vector, enhanced tumorigenesis (65).

Mass Spectrometry with Respect to Membrane Protein Identification

Proteomics is the large-scale study of proteins at the structural and functional level (6). The investigation of these complex protein samples can be effectively analyzed utilizing the capabilities of mass spectrometry coupled with the availability of massive gene/ genome and protein databases in a high-throughput manner. Since membrane proteins represent one-third of the total proteins encoded by the human genome (102), defining membrane proteomes will provide a crucial understanding to membrane function in biological processes, provide protein targets for drug development, and enable cell line distinction (e.g. hESC lines and derivatives) through the use of cell surface biomarkers (51).

Two techniques currently used to detail the membrane proteome of embryonic stem cells are two-dimensional gel electrophoresis (78) and liquid chromatography

coupled to tandem mass spectrometry (LC-MS/MS) (104, 115). Traditional methods of membrane protein identification involved two-dimensional SDS PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) followed by in-gel tryptic digestion and analysis of the resolved proteins via mass spectrometry (84, 85). This technique, however, is significantly limited by the amphipathic nature of membrane proteins. The hydrophobic transmembrane-spanning domains are well buried within the lipid bilayer and are generally not solublilized in the non-detergent isoelectric buffers used during electrophoresis (106).

High-throughput methods for membrane identification enrich for insoluble proteins and separate the proteins using (SDS)-PAGE (60, 89, 93, 106). The gel lane is cut into fractions and each fraction is tryptically digested to produce peptides (93). Peptides are extracted from the gel and analyzed via reverse phase chromatography separation coupled to tandem mass spectrometry (LC-MS/MS). Each intact peptide is scanned, recorded, and further fragmented through predictable mechanisms and produces a series of mass ions correlating to the amino acid content of the peptide (82). The order of the fragmentation pattern depicts the sequence of the amino acids in the peptide. Sequences are then searched against a well established protein database to identify the protein from which the peptide originated.

High-throughput proteomics is well suited to cell culture and has been recently used to establish unique membrane signatures for a variety of cell types including hESC versus human embryonal carcinoma cells (35), human epithelial cells during cancer progression (100), and mouse astrocytes versus mouse astrocytomas (89). The most notable advantage to a LC-MS/MS proteomic approach over trial and error of current

markers (e.g. cluster of differentiation (CD) markers) is that unique surface markers, from a custom matrix of cell types, can be identified devoid of prior functional information.

Specific Aims

The directed differentiation and isolation of human embryonic stem cell (hESC)-derived astrocytes would provide an in vitro cell source for the further study of a wide range of neurological disorders such as Alzheimer's, Alexander's Disease, Amyotrophic Lateral Sclerosis (ALS, Lou Gehrig's Disease), glial tumors and central nervous system inflammation. Since astrocyte morphology is not static and can be dependent on location within the central nervous system (CNS), the high content of intermediate filaments, particularly glial fibrillary acidic protein (GFAP), is used to characterize astrocytes. Herein, we examine the temporal expression of GFAP as a biomarker to better elucidate the differentiation process of human neural progenitor (hNP) cells towards an astrocyte-like cell population. Astrocyte differentiation is difficult to study because the temporal onset of gliogenesis is preceded by neurogenesis. Furthermore, the in vitro isolation of astrocytes from human stem cell sources still remains elusive and would provide further characterization of astrocyte-mediated disease pathologies.

The goal of this dissertation is to characterize the temporal expression of glial fibrillary acidic protein (GFAP) employing a glial fluorescent reporter system. This characterization will potentially lead to the isolation of a glial-like cell type from human embryonic stem cell (hESC) derived neural progenitor (hNP).

SPECIFIC AIM 1 (SA1): Develop a tissue-specific promoter (GFAP)/ fluorescent reporter construct that will stably integrate into hNP cells utilizing lentiviral technology. This will provide the ability to selectively trace the appearance of glial-like cells in vitro and monitor their potential differentiation.

SPECIFIC AIM 2 (SA2): Characterize the temporal differentiation of hNP cells into a $GFAP^+$ cell type utilizing the developmental biomarker construct derived from SA1. Furthermore, does the derived $GFAP^+$ cell population change in GFAP expression over time.

SPECIFIC AIM 3 (SA3): Determine if membrane fractionation and liquid-chromatography tandem mass spectrometry (LC-MS/MS) techniques can be employed to develop novel biomarkers capable of distinguishing between different hESC lines and their derivatives.

If the aims are proven successful, then characterization of an astrocyte-like cell type from hESC derived neural progenitor cells will have been achieved, leading to a more complete *in vitro* model for neural development. In addition, a proof of concept applying a high-throughput proteomic approach will demonstrate the ability to delineate different hESC lines and their derivatives.

References

1. 2001. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther **69:**89-95.
2. 2008, posting date. National Institute Neurological Disorders and Stroke (NINDS): Alexander Disease Information Page. [Online.]
3. 2008, posting date. Society for Neuroscience: Astrocytes. [Online.]
4. **Abdallah, B. M., M. Haack-Sorensen, J. S. Burns, B. Elsnab, F. Jakob, P. Hokland, and M. Kassem.** 2005. Maintenance of differentiation potential of human bone marrow mesenchymal stem cells immortalized by human telomerase reverse transcriptase gene despite [corrected] extensive proliferation. Biochem Biophys Res Commun **326:**527-38.
5. **Adam, J., A. Myat, I. Le Roux, M. Eddison, D. Henrique, D. Ish-Horowicz, and J. Lewis.** 1998. Cell fate choices and the expression of Notch, Delta and Serrate homologues in the chick inner ear: parallels with Drosophila sense-organ development, p. 4645-4654, vol. 125.
6. **Aebersold, R., and M. Mann.** 2003. Mass spectrometry-based proteomics. Nature **422:**198-207.
7. **Alvarez-Buylla, A., D. R. Buskirk, and F. Nottebohm.** 1987. Monoclonal antibody reveals radial glia in adult avian brain. J Comp Neurol **264:**159-70.
8. **Alvarez-Buylla, A., J. M. Garcia-Verdugo, and A. D. Tramontin.** 2001. A unified hypothesis on the lineage of neural stem cells. Nat Rev Neurosci **2:**287-93.

9. **Andrews, P. W., N. Benvenisty, R. McKay, M. F. Pera, J. Rossant, H. Semb, G. N. Stacey, and S. C. I. S. C. Initia.** 2005. The International Stem Cell initiative: toward benchmarks for human embryonic stem cell research. Nature Biotechnology **23:**795-797.

10. **Artavanis-Tsakonas, S., M. D. Rand, and R. J. Lake.** 1999. Notch signaling: Cell fate control and signal integration in development. Science **284:**770-776.

11. **Attisano, L., and J. L. Wrana.** 2002. Signal Transduction by the TGF-beta Superfamily. Science **296:**1646-1647.

12. **Awad, H. A., D. L. Butler, G. P. Boivin, F. N. Smith, P. Malaviya, B. Huibregtse, and A. I. Caplan.** 1999. Autologous mesenchymal stem cell-mediated repair of tendon. Tissue Eng **5:**267-77.

13. **Ballabh, P., A. Braun, and M. Nedergaard.** 2004. The blood-brain barrier: an overview: structure, regulation, and clinical implications. Neurobiol Dis **16:**1-13.

14. **Barrilleaux, B., D. G. Phinney, D. J. Prockop, and K. C. O'connor.** 2006. Review: Ex Vivo Engineering of Living Tissues with Adult Stem Cells. Tissue Engineering **12:**3007-3019.

15. **Baum, C., O. Kustikova, U. Modlich, Z. X. Li, and B. Fehse.** 2006. Mutagenesis and oncogenesis by chromosomal insertion of gene transfer vectors. Human Gene Therapy **17:**253-263.

16. **Beckers, J., A. Clark, K. Wunsch, M. H. De Angelis, and A. Gossler.** 1999. Expression of the mouse Delta1 gene during organogenesis and fetal development. Mechanisms of Development **84:**165-168.

17. **Besnard, F., M. Brenner, Y. Nakatani, R. Chao, H. J. Purohit, and E. Freese.** 1991. Multiple Interacting Sites Regulate Astrocyte-Specific Transcription of the Human Gene for Glial Fibrillary Acidic Protein. Journal of Biological Chemistry **266:**18877-18883.

18. **Bianco, P., M. Riminucci, S. Gronthos, and P. G. Robey.** 2001. Bone marrow stromal stem cells: nature, biology, and potential applications. Stem Cells **19:**180-92.

19. **Blaese, R. M., K. W. Culver, A. D. Miller, C. S. Carter, T. Fleisher, M. Clerici, G. Shearer, L. Chang, Y. Chiang, P. Tolstoshev, J. J. Greenblatt, S. A. Rosenberg, H. Klein, M. Berger, C. A. Mullen, W. J. Ramsey, L. Muul, R. A. Morgan, and W. F. Anderson.** 1995. T Lymphocyte-Directed Gene Therapy for ADA- SCID: Initial Trial Results After 4 Years, p. 475-480, vol. 270.

20. **Blomer, U., L. Naldini, T. Kafri, D. Trono, I. M. Verma, and F. H. Gage.** 1997. Highly efficient and sustained gene transfer in adult neurons with a lentivirus vector. Journal of Virology **71:**6641-6649.

21. **Brandes, A. A., A. Tosoni, E. Franceschi, M. Reni, G. Gatta, and C. Vecht.** 2008. Glioblastoma in adults. Crit Rev Oncol Hematol **67:**139-52.

22. **Brenner, M., A. B. Johnson, O. Boespflug-Tanguy, D. Rodriguez, J. E. Goldman, and A. Messing.** 2001. Mutations in GFAP, encoding glial fibrillary acidic protein, are associated with Alexander disease. Nat Genet **27:**117-120.

23. **Brenner, M., W. C. Kisseberth, Y. Su, F. Besnard, and A. Messing.** 1994. Gfap Promoter Directs Astrocyte-Specific Expression in Transgenic Mice. Journal of Neuroscience **14:**1030-1037.

24. **Brimble, S. N., X. M. Zeng, D. A. Weiler, Y. Q. Luo, Y. Liu, I. G. Lyons, W. J. Freed, A. J. Robins, M. S. Rao, and T. C. Schulz.** 2004. Karyotypic stability, genotyping, differentiation, feeder-free maintenance, and gene expression sampling in three human embryonic stem cell lines derived prior to August 9, 2001. Stem Cells and Development **13:**585-597.

25. **Burns, J. C., T. Friedmann, W. Driever, M. Burrascano, and J. K. Yee.** 1993. Vesicular Stomatitis-Virus G Glycoprotein Pseudotyped Retroviral Vectors - Concentration to Very High-Titer and Efficient Gene-Transfer into Mammalian and Nonmammalian Cells. Proceedings of the National Academy of Sciences of the United States of America **90:**8033-8037.

26. **Buzzard, J. J., N. M. Gough, J. M. Crook, and A. Colman.** 2004. Karyotype of human ES cells during extended culture. Nature Biotechnology **22:**381-382.

27. **Cavazzana-Calvo, M., S. Hacein-Bey, C. D. Basile, F. Gross, E. Yvon, P. Nusbaum, F. Selz, C. Hue, S. Certain, J. L. Casanova, P. Bousso, F. Le Deist, and A. Fischer.** 2000. Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. Science **288:**669-672.

28. **Check, E.** 2002. Gene therapy: A tragic setback. Nature **420:**116-118.

29. **Chitnis, A., D. Henrique, J. Lewis, D. Ish-Horowicz, and C. Kintner.** 1995. Primary neurogenesis in Xenopus embryos regulated by a homologue of the Drosophila neurogenic gene Delta. Nature **375:**761-6.

30. **Coon, K. D., T. Dunckley, and D. A. Stephan.** 2004. Biomarker identification in neurologic diseases: improving diagnostics and therapeutics. Expert Rev Mol Diagn **4:**361-75.

31. **Danbolt, N. C.** 2001. Glutamate uptake. Prog Neurobiol **65:**1-105.

32. **Dezawa, M., H. Kanno, M. Hoshino, H. Cho, N. Matsumoto, Y. Itokazu, N. Tajima, H. Yamada, H. Sawada, H. Ishikawa, T. Mimura, M. Kitada, Y. Suzuki, and C. Ide.** 2004. Specific induction of neuronal cells from bone marrow stromal cells and application for autologous transplantation. J Clin Invest **113:**1701-10.

33. **Di Giorgio, F. P., M. A. Carrasco, M. C. Siao, T. Maniatis, and K. Eggan.** 2007. Non-cell autonomous effect of glia on motor neurons in an embryonic stem cellâ€“based ALS model. Nature Neuroscience **10:**608-614.

34. **Dominici, M., K. Le Blanc, I. Mueller, I. Slaper-Cortenbach, F. Marini, D. Krause, R. Deans, A. Keating, D. Prockop, and E. Horwitz.** 2006. Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. Cytotherapy **8:**315-7.

35. **Dormeyer, W., D. van Hoof, S. R. Braam, A. J. Heck, C. L. Mummery, and J. Krijgsveld.** 2008. Plasma membrane proteomics of human embryonic stem cells and human embryonal carcinoma cells. J Proteome Res **7:**2936-51.

36. **Dull, T., R. Zufferey, M. Kelly, R. J. Mandel, M. Nguyen, D. Trono, and L. Naldini.** 1998. A third-generation lentivirus vector with a conditional packaging system. Journal of Virology **72:**8463-8471.

37. **Edwards, M. M., and S. R. Robinson.** 2006. TNF alpha affects the expression of GFAP and S100B: implications for Alzheimer's disease. J Neural Transm **113:**1709-15.

38. **Foster, L. J., P. A. Zeemann, C. Li, M. Mann, O. N. Jensen, and M. Kassem.** 2005. Differential expression profiling of membrane proteins by quantitative

proteomics in a human mesenchymal stem cell line undergoing osteoblast differentiation. Stem Cells **23:**1367-77.

39. **Gang, E. J., D. Bosnakovski, C. A. Figueiredo, J. W. Visser, and R. C. Perlingeiro.** 2007. SSEA-4 identifies mesenchymal stem cells from bone marrow. Blood **109:**1743-51.

40. **Gilbert, S. F.** 2003. Developmental Biology, Seventh Edition ed. Sinauer Assoicates, Inc., Sunderland, Massachusetts.

41. **Gotz, M., and Y. A. Barde.** 2005. Radial glial cells defined and major intermediates between embryonic stem cells and CNS neurons. Neuron **46:**369-72.

42. **Gronthos, S., S. E. Graves, S. Ohta, and P. J. Simmons.** 1994. The STRO-1+ fraction of adult human bone marrow contains the osteogenic precursors. Blood **84:**4164-73.

43. **Hacein-Bey-Abina, S., C. Von Kalle, M. Schmidt, M. P. McCcormack, N. Wulffraat, P. Leboulch, A. Lim, C. S. Osborne, R. Pawliuk, E. Morillon, R. Sorensen, A. Forster, P. Fraser, J. I. Cohen, G. de Saint Basile, I. Alexander, U. Wintergerst, T. Frebourg, A. Aurias, D. Stoppa-Lyonnet, S. Romana, I. Radford-Weiss, F. Gross, F. Valensi, E. Delabesse, E. Macintyre, F. Sigaux, J. Soulier, L. E. Leiva, M. Wissler, C. Prinz, T. H. Rabbitts, F. Le Deist, A. Fischer, and M. Cavazzana-Calvo.** 2003. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science **302:**415-419.

44. **Haddon, C., Y. J. Jiang, L. Smithers, and J. Lewis.** 1998. Delta-Notch signalling and the patterning of sensory cell differentiation in the zebrafish ear: evidence from the mind bomb mutant, p. 4637-4644, vol. 125.

45. **Hu, W. S., and V. K. Pathak.** 2000. Design of retroviral vectors and helper cells for gene therapy. Pharmacological Reviews **52:**493-511.

46. **Isaacs, A., M. Baker, F. Wavrant-De Vrieze, and M. Hutton.** 1998. Determination of the Gene Structure of Human GFAP and Absence of Coding Region Mutations Associated with Frontotemporal Dementia with Parkinsonism Linked to Chromosome 17. Genomics **51:**152-154.

47. **Ishibashi, T., K. A. Dakin, B. Stevens, P. R. Lee, S. V. Kozlov, C. L. Stewart, and R. D. Fields.** 2006. Astrocytes promote myelination in response to electrical impulses. Neuron **49:**823-32.

48. **Jacobson, M.** 1968. Cessation of DNA Synthesis in Retinal Ganglion Cells Correlated with Time of Specification of Their Central Connections. Developmental Biology **17:**219-&.

49. **Johnstone, B., T. M. Hering, A. I. Caplan, V. M. Goldberg, and J. U. Yoo.** 1998. In vitro chondrogenesis of bone marrow-derived mesenchymal progenitor cells. Exp Cell Res **238:**265-72.

50. **Jorgensen, C., D. Noel, and G. Gross.** 2002. Could inflammatory arthritis be triggered by progenitor cells in the joints? Ann Rheum Dis **61:**6-9.

51. **Josic, D., and J. G. Clifton.** 2007. Mammalian plasma membrane proteomics. Proteomics **7:**3010-29.

52. **Keilhoff, G., A. Goihl, K. Langnase, H. Fansa, and G. Wolf.** 2006. Transdifferentiation of mesenchymal stem cells into Schwann cell-like myelinating cells. Eur J Cell Biol **85:**11-24.

53. **Kessaris, N., N. Pringle, and W. D. Richardson.** 2008. Specification of CNS glia from neural stem cells in the embryonic neuroepithelium. Philos Trans R Soc Lond B Biol Sci **363:**71-85.

54. **Kumar, S., D. Chanda, and S. Ponnazhagan.** 2008. Therapeutic potential of genetically modified mesenchymal stem cells. Gene Ther **15:**711-5.

55. **Lauder, J. M.** 1995. Chapter 8: Neuroembryology and Differentiation. *In* M. P. Conn (ed.), Neuroscience in Medicine. J. B. Lippincott Company, Philadelphia.

56. **Lewis, J.** 1998. Notch signalling and the control of cell fate choices in vertebrates. Seminars in Cell and Developmental Biology **9:**583-589.

57. **Luk, J. M., P. P. Wang, C. K. Lee, J. H. Wang, and S. T. Fan.** 2005. Hepatic potential of bone marrow stromal cells: development of in vitro co-culture and intra-portal transplantation models. J Immunol Methods **305:**39-47.

58. **Lundstrom, K.** 2003. Latest development in viral vectors for gene therapy. Trends in Biotechnology **21:**117-122.

59. **Meirelles, L. d. S., P. C. Chagastelles, and N. B. Nardi.** 2006. Mesenchymal stem cells reside in virtually all post-natal organs and tissues. J Cell Sci **119:**2204-2213.

60. **Miguet, L., K. Pacaud, C. Felden, B. Hugel, M. C. Martinez, J. M. Freyssinet, R. Herbrecht, N. Potier, A. van Dorsselaer, and L. Mauvieux.** 2006. Proteomic

analysis of malignant lymphocyte membrane microparticles using double ionization coverage optimization. Proteomics **6:**153-71.

61. **Min, K.-J., M.-s. Yang, S.-U. Kim, I. Jou, and E.-h. Joe.** 2006. Astrocytes Induce Hemeoxygenase-1 Expression in Microglia: A Feasible Mechanism for Preventing Excessive Brain Inflammation. J. Neurosci. **26:**1880-1887.

62. **Mitalipova, M., J. Calhoun, S. J. Shin, D. Wininger, T. Schulz, S. Noggle, A. Venable, I. Lyons, A. Robins, and S. Stice.** 2003. Human embryonic stem cell lines derived from discarded embryos. Stem Cells **21:**521-526.

63. **Mitalipova, M. M., R. R. Rao, D. M. Hoyer, J. A. Johnson, L. F. Meisner, K. L. Jones, S. Dalton, and S. L. Stice.** 2005. Preserving the genetic integrity of human embryonic stem cells. Nature Biotechnology **23:**19-20.

64. **Miyoshi, H., M. Takahashi, F. H. Gage, and I. M. Verma.** 1997. Stable and efficient gene transfer into the retina using an HIV-based lentiviral vector. Proceedings of the National Academy of Sciences of the United States of America **94:**10319-10323.

65. **Montini, E., D. Cesana, M. Schmidt, F. Sanvito, M. Ponzoni, C. Bartholomae, L. S. Sergi, F. Benedicenti, A. Ambrosi, C. Di Serio, C. Doglioni, C. von Kalle, and L. Naldini.** 2006. Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration. Nature Biotechnology **24:**687-696.

66. **Mori, T., A. Buffo, and M. Gotz.** 2005. The novel roles of glial cells revisited: the contribution of radial glia and astrocytes to neurogenesis. Curr Top Dev Biol **69:**67-99.

67. **Mori, T., A. Buffo, M. Gotz, and P. S. Gerald.** 2005. The Novel Roles of Glial Cells Revisited: The Contribution of Radial Glia and Astrocytes to Neurogenesis, p. 67-99, Current Topics in Developmental Biology, vol. Volume 69. Academic Press.

68. **Munoz-Sanjuan, I., and A. H. Brivanlou.** 2002. Neural induction, the default model and embryonic stem cells. Nature Reviews Neuroscience **3:**271-280.

69. **Nagai, M., D. B. Re, T. Nagata, A. n. Chalazonitis, T. M. Jessell, H. Wichterle, and S. Przedborski.** 2007. Astrocytes expressing ALS-linked mutated SOD1 release factors selectively toxic to motor neurons. Nature Neuroscience **10:**615-622.

70. **Naldini, L., U. Blomer, P. Gallay, D. Ory, R. Mulligan, F. H. Gage, I. M. Verma, and D. Trono.** 1996. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science **272:**263-267.

71. **Nye, J. S., R. Kopan, and R. Axel.** 1994. An Activated Notch Suppresses Neurogenesis and Myogenesis but Not Gliogenesis in Mammalian-Cells. Development **120:**2421-2430.

72. **Park, H.-C., and B. Appel.** 2003. Delta-Notch signaling regulates oligodendrocyte specification, p. 3747-3755, vol. 130.

73. **Parri, R., and V. Crunelli.** 2003. An astrocyte bridge from synapse to blood flow. Nat Neurosci **6:**5-6.

74. **Pera, M. F., J. Andrade, S. Houssami, B. Reubinoff, A. Trounson, E. G. Stanley, D. Ward-van Oostwaard, and C. Mummery.** 2004. Regulation of

human embryonic stem cell differentiation by BMP-2 and its antagonist noggin. Journal of Cell Science **117:**1269-1280.

75. **Pittenger, M. F., A. M. Mackay, S. C. Beck, R. K. Jaiswal, R. Douglas, J. D. Mosca, M. A. Moorman, D. W. Simonetti, S. Craig, and D. R. Marshak.** 1999. Multilineage potential of adult human mesenchymal stem cells. Science **284:**143-7.

76. **Reubinoff, B. E., P. Itsykson, T. Turetsky, M. F. Pera, E. Reinhartz, A. Itzik, and T. Ben-Hur.** 2001. Neural progenitors from human embryonic stem cells. Nat Biotechnol **19:**1134-40.

77. **Richardson, W. D., H. K. Smith, T. Sun, N. P. Pringle, A. Hall, and R. Woodruff.** 2000. Oligodendrocyte lineage and the motor neuron connection. Glia **29:**136-142.

78. **Robertson, E. J.** 1987. Embryo-derived stem cell lines., p. 71-112. *In* E. J. Robertson (ed.), Teratocarcinomas and Embryonic Stem Cells: A Practical Approach. IRL Press, Oxford.

79. **Rodriguez.** 2001. Infantile Alexander disease: Spectrum of GFAP mutations and genotype-phenotype correlation (vol 69, pg 1134, 2001). American Journal of Human Genetics **69:**1413-1413.

80. **Rosler, E. S., G. J. Fisk, X. Ares, J. Irving, T. Miura, M. S. Rao, and M. K. Carpenter.** 2004. Long-term culture of human embryonic stem cells in feeder-free conditions. Developmental Dynamics **229:**259-274.

81. **Rothstein, J. D., M. Dykes-Hoberg, C. A. Pardo, L. A. Bristol, L. Jin, R. W. Kuncl, Y. Kanai, M. A. Hediger, Y. Wang, J. P. Schielke, and D. F. Welty.**

1996. Knockout of glutamate transporters reveals a major role for astroglial transport in excitotoxicity and clearance of glutamate. Neuron **16:**675-86.

82. **S.A. Martin, R. S. J., C.E. Costello, K. Bieman.** 1988. The Analysis of Peptides and Proteins by Mass Spectrometry, Proceedings of the Fourth Texas Symposium on Mass Spectrometry, p. 135-150. *In* C. J. McNeal (ed.). John Wiley & Sons, New York.

83. **Santello, M., and A. Volterra.** 2008. Synaptic modulation by astrocytes via Ca(2+)-dependent glutamate release. Neuroscience.

84. **Santoni, V., S. Kieffer, D. Desclaux, F. Masson, and T. Rabilloud.** 2000. Membrane proteomics: use of additive main effects with multiplicative interaction model to classify plasma membrane proteins according to their solubility and electrophoretic properties. Electrophoresis **21:**3329-44.

85. **Santoni, V., M. Molloy, and T. Rabilloud.** 2000. Membrane proteins and proteomics: un amour impossible? Electrophoresis **21:**1054-70.

86. **Sasai, Y.** 2005. Developmental biology - A blank canvas no more. Nature **435:**433-434.

87. **Sauer, F. C.** 1935. Mitosis in the neural tube. J. Comp. Neurol. **62:**377-405.

88. **Schousboe, A., A. Sarup, L. K. Bak, H. S. Waagepetersen, and O. M. Larsson.** 2004. Role of astrocytic transport processes in glutamatergic and GABAergic neurotransmission. Neurochem Int **45:**521-7.

89. **Seyfried, N. T., L. C. Huysentruyt, J. A. Atwood, 3rd, Q. Xia, T. N. Seyfried, and R. Orlando.** 2008. Up-regulation of NG2 proteoglycan and interferon-

induced transmembrane proteins 1 and 3 in mouse astrocytoma: a membrane proteomics approach. Cancer Lett **263:**243-52.

90. **Shamblott, M. J., J. Axelman, S. P. Wang, E. M. Bugg, J. W. Littlefield, P. J. Donovan, P. D. Blumenthal, G. R. Huggins, and J. D. Gearhart.** 1998. Derivation of pluripotent stem cells horn cultured human primordial germ cells. Proceedings of the National Academy of Sciences of the United States of America **95:**13726-13731.

91. **Shin, S., M. Mitalipova, S. Noggle, D. Tibbitts, A. Venable, R. Rao, and S. L. Stice.** 2006. Long-Term Proliferation of Human Embryonic Stem Cell-Derived Neuroepithelial Cells Using Defined Adherent Culture Conditions. Stem Cells **24:**125-138.

92. **Simard, M., and M. Nedergaard.** 2004. The neurobiology of glia in the context of water and ion homeostasis. Neuroscience **129:**877-96.

93. **Simpson, R. J., L. M. Connolly, J. S. Eddes, J. J. Pereira, R. L. Moritz, and G. E. Reid.** 2000. Proteomic analysis of the human colon carcinoma cell line (LIM 1215): development of a membrane protein database. Electrophoresis **21:**1707-32.

94. **Stice, S. L., N. L. Boyd, S. K. Dhara, B. A. Gerwe, D. W. Machacek, and S. Shin.** 2006. Human embryonic stem cells: challenges and opportunities. Reprod Fertil Dev **18:**839-46.

95. **Tanigaki, K., F. Nogaki, J. Takahashi, K. Tashiro, H. Kurooka, and T. Honjo.** 2001. Notch1 and Notch3 instructively restrict bFGF-responsive multipotent neural progenitor cells to an astroglial fate. Neuron **29:**45-55.

96. **Thomson, J. A., J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall, and J. M. Jones.** 1998. Embryonic stem cell lines derived from human blastocysts. Science **282:**1145-1147.

97. **Thomson, J. A., J. Kalishman, T. G. Golos, M. Durning, C. P. Harris, R. A. Becker, and J. P. Hearn.** 1995. Isolation of a Primate Embryonic Stem-Cell Line. Proceedings of the National Academy of Sciences of the United States of America **92:**7844-7848.

98. **Thomson, J. A., J. Kalishman, T. G. Golos, M. Durning, C. P. Harris, and J. P. Hearn.** 1996. Pluripotent cell lines derived from common marmoset (Callithrix jacchus) blastocysts. Biology of Reproduction **55:**254-259.

99. **Vaccarino, F. M., D. M. Fagel, Y. Ganat, M. E. Maragnoli, L. R. Ment, Y. Ohkubo, M. L. Schwartz, J. Silbereis, and K. M. Smith.** 2007. Astroglial cells in development, regeneration, and repair. Neuroscientist **13:**173-85.

100. **Vercoutter-Edouart, A. S., M. C. Slomianny, O. Dekeyzer-Beseme, J. F. Haeuw, and J. C. Michalski.** 2008. Glycoproteomics and glycomics investigation of membrane N-glycosylproteins from human colon carcinoma cells. Proteomics **8:**3236-56.

101. **Verderber, L., W. Johnson, L. Mucke, and V. Sarthy.** 1995. Differential regulation of a glial fibrillary acidic protein-LacZ transgene in retinal astrocytes and Muller cells. Invest Ophthalmol Vis Sci **36:**1137-43.

102. **Wallin, E., and G. von Heijne.** 1998. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. Protein Sci **7:**1029-38.

103. **Wang, S. L., A. D. Sdrulla, G. diSibio, G. Bush, D. Nofziger, C. Hicks, G. Weinmaster, and B. A. Barres.** 1998. Notch receptor activation inhibits oligodendrocyte differentiation. Neuron **21:**63-75.

104. **Weatherly, D. B., J. A. Atwood, 3rd, T. A. Minning, C. Cavola, R. L. Tarleton, and R. Orlando.** 2005. A Heuristic method for assigning a false-discovery rate for protein identifications from Mascot database search results. Mol Cell Proteomics **4:**762-72.

105. **Wilson, S. I., A. Rydstrom, T. Trimborn, K. Willert, R. Nusse, T. M. Jessell, and T. Edlund.** 2001. The status of Wnt signalling regulates neural and epidermal fates in the chick embryo. Nature **411:**325-330.

106. **Wu, C. C., M. J. MacCoss, K. E. Howell, and J. R. Yates, 3rd.** 2003. A method for the comprehensive proteomic analysis of membrane proteins. Nat Biotechnol **21:**532-8.

107. **Ying, Q. L., M. Stavridis, D. Griffiths, M. Li, and A. Smith.** 2003. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat Biotechnol **21:**183-6.

108. **Zhang, X. J., and W. T. Godbey.** 2006. Viral vectors for gene delivery in tissue engineering. Advanced Drug Delivery Reviews **58:**515-534.

109. **Zhou, Q., and D. J. Anderson.** 2002. The bHLH transcription factors OLIG2 and OLIG1 couple neuronal and glial subtype specification. Cell **109:**61-73.

110. **Zhuo, L., B. Sun, C.-L. Zhang, A. Fine, S.-Y. Chiu, and A. Messing.** 1997. Live Astrocytes Visualized by Green Fluorescent Protein in Transgenic Mice. Developmental Biology **187:**36-42.

111. **Zubler, R. H.** 2006. Ex vivo expansion of hematopoietic stem cells and gene therapy development. Swiss Med Wkly **136:**795-9.

112. **Zufferey, R., T. Dull, R. J. Mandel, A. Bukovsky, D. Quiroz, L. Naldini, and D. Trono.** 1998. Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. Journal of Virology **72:**9873-9880.

113. **Zufferey, R., D. Nagy, R. J. Mandel, L. Naldini, and D. Trono.** 1997. Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. Nature Biotechnology **15:**871-875.

114. **Zufferey, R., Trono, D.** 2000. Gene Cloning, Expression, and Mutagenesis, p. 21.1 - 21.12. *In* D. R. e. a. Sibley (ed.), Current Protocols in Neuroscience. John Wiley and Sons, New York, NY.

115. **Zybailov, B., A. L. Mosley, M. E. Sardiu, M. K. Coleman, L. Florens, and M. P. Washburn.** 2006. Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. J Proteome Res **5:**2339-47.

Figure 1.1. Order of In Vivo Mammalian Differentiation of Neuroepithelial (NE) Cells. **(A)** The NE cells undergo symmetric division establishing the neuroepithelial pool; then asymmetric dividing to form the radial glial pool. Upon completion of both pools, the NE cells differentiate into the terminal cell types of the CNS. The temporal onset of gliogenesis precedes neurogenesis. For comparison, temporal differentiation within the mouse is illustrated in **(B)**.

Adapted from: Gotz, M., and Y. A. Barde. 2005. Radial glial cells: Defined and major intermediates between embryonic, stem cells and CNS neurons. Neuron 46:369-372; Glibert (2003) Developmental Biology, 4th Edition

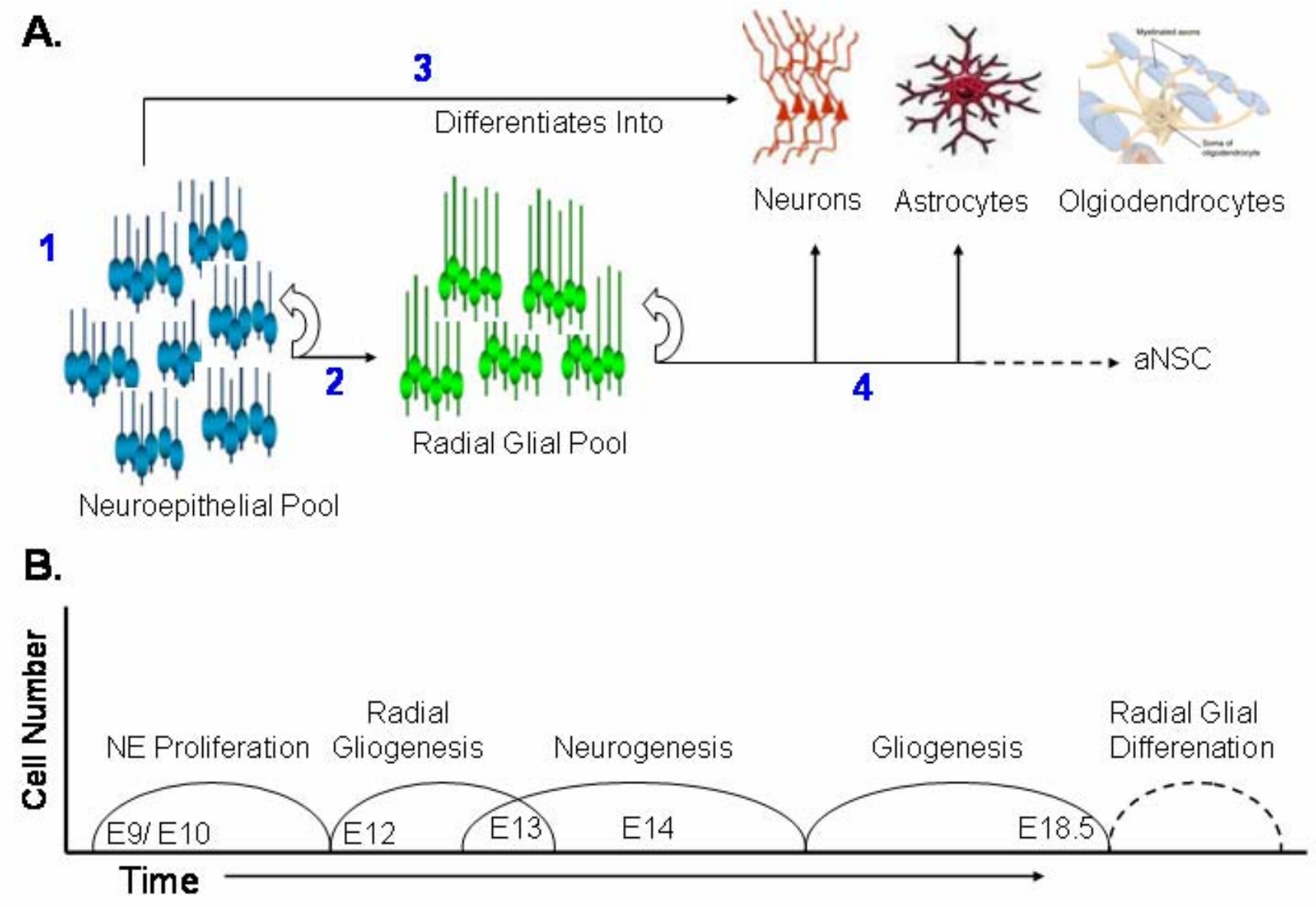
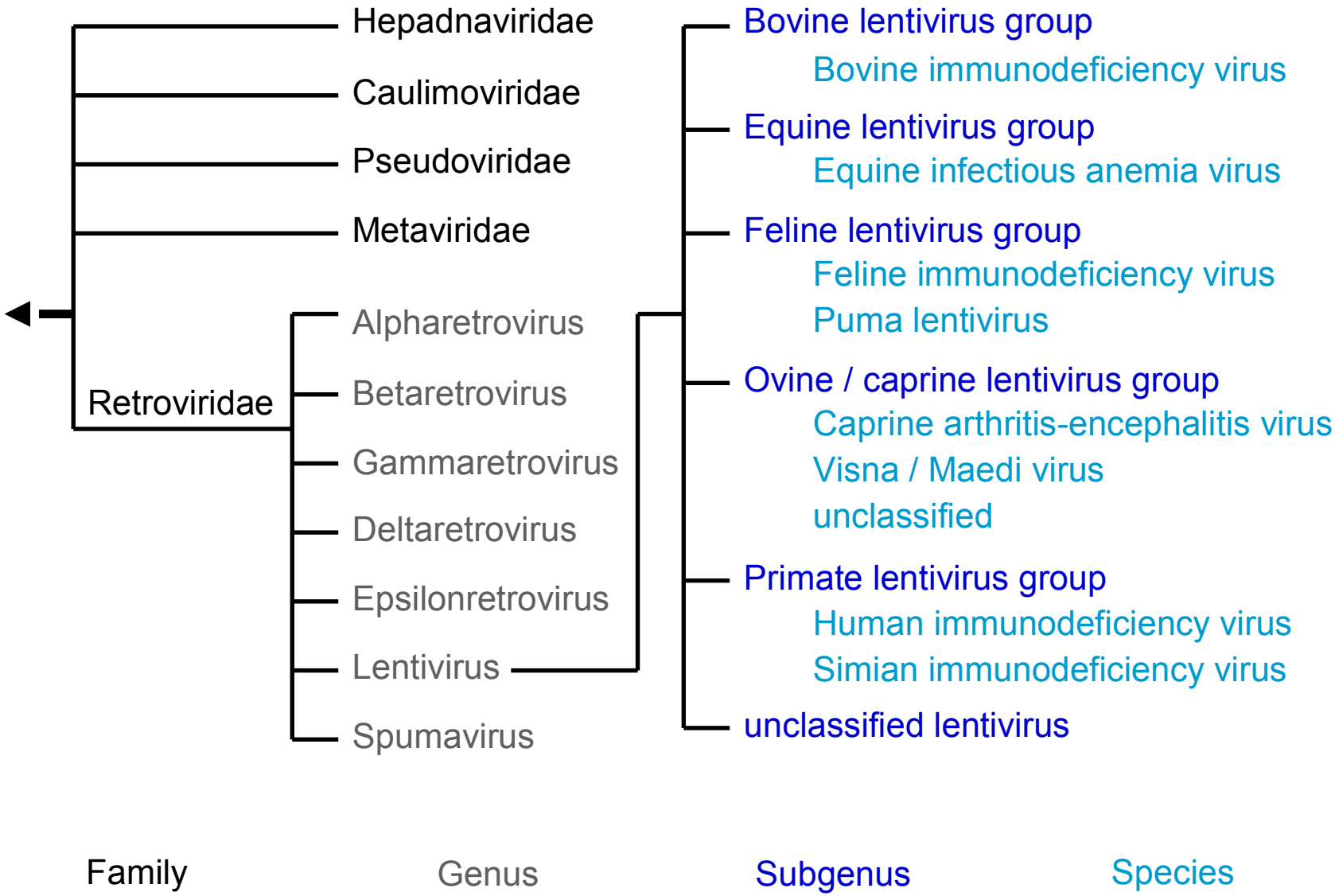

A.
3
Differentiates Into
Neurons
Astrocytes
Olgiodendrocytes
1
2
4
aNSC
Radial Glial Pool
Neuroepithelial Pool
B.
Cell Number
NE Proliferation
Radial Gliogenesis
Neurogenesis
Gliogenesis
Radial Glial Differenation
E9/ E10
E12
E13
E14
E18.5
Time

Figure 1.2. Taxonomy Classification of the Lentivirus Genus. This particular genus stems from the Retroviridae family ultimately under the DNA-RNA Reverse Transcribing Viruses classification. As shown, the Lentiviral genus is further partitioned by means of the infected host species. The major lentiviral species are shown.

Adapted from: (Tree of Life Web Project. 2005. DNA-RNA Reverse Transcribing Viruses. Version 22 December 2005 (temporary). http://tolweb.org/DNA-RNA_Reverse_Transcribing_Viruses/21831/2005.12.22 *in* The Tree of Life Web Project, http://tolweb.org);NCBI Taxonomy Browser (http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Undef&id=35268&lvl=3&lin=s&keep=1&srchmode=1&unlock)


Hepadnaviridae
Caulimoviridae
Pseudoviridae
Metaviridae
Retroviridae
Alpharetrovirus
Betaretrovirus
Gammaretrovirus
Deltaretrovirus
Epsilonretrovirus
Lentivirus
Spumavirus
Bovine lentivirus group
Bovine immunodeficiency virus
Equine lentivirus group
Equine infectious anemia virus
Feline lentivirus group
Feline immunodeficiency virus
Puma lentivirus
Ovine / caprine lentivirus group
Caprine arthritis-encephalitis virus
Visna / Maedi virus
unclassified
Primate lentivirus group
Human immunodeficiency virus
Simian immunodeficiency virus
unclassified lentivirus
Family
Genus
Subgenus
Species

CHAPTER 2

GENETIC MANIPULATION OF NEURAL PROGENITORS DERIVED FROM HUMAN EMBRYONIC STEM CELLS[1]

Brian A. Gerwe*, Sujoy K. Dhara*, Anirban Majumder, Mahesh C. Dodla, Nolan L. Boyd, Kowser Hasneen and Steven L. Stice.

*Authors contributed equally to this work

To be submitted to PNAS.

Abstract

Human embryonic stem cell (hESC) derived Neural progenitors (NP) are a useful tool to further our understanding of development biology. However, inability to monitor the cells and their gene expression in vitro and in vivo presents significant obstacles. To overcome this we have developed techniques to express genetically encoded fluorescent reporter proteins in NP cells. We examined three different methods (transduction, transfection and nucleofection) of gene delivery using expression of enhanced green fluorescent protein (EGFP) under the control of the mouse ubiquitin C (Ubc) promoter to compare the efficiency of each technique. Considering different parameters (percentage cells expressing transgene, viability and long term transgene expression), transduction was observed to be the best method of genetic manipulation in which about 37% NP cells express EGFP. Transduced cells retained their differentiation potential as demonstrated by co-expression of EGFP with neural progenitor (NESTIN, SOX2, MUSASHI1) and neuronal markers (GABA, TH and TuJ1). Further, we generated a plasmid useful for tracking astrocyte lineage utilizing the GFAP (glial fibrillary acidic protein) promoter with EGFP and tomato Red (tmRed) reporter proteins. Herein, we demonstrate the effective detection of the $GFAP^{+}$ lineage on transduced NP-derived GFAP expressing cells. Finally, we compared the efficiency of three constitutively expressing promoters (CMV, EF1α and Ubc) driving EGFP expression in NP cells and studied the silencing (in 20-30% cells) associated with these promoters. These results confirm the long-term in vitro transgene expression of genetically modified NP cells suggesting lentiviral transduction methods could enhance

transplantation studies through cell tracking capabilities and provide long-term gene expression vital to clinical applications.

Introduction

Human embryonic stem cell (hESC) derived neural progenitors (NP) are unique self-renewing cells that can differentiate into all cell types found within the central nervous system, which include neurons, glial cells and oligodendrocytes (7, 27, 28). They represent specific facets of embryonic development and thus, present an opportunity to study neural development and differentiation in vitro (28). These cells, upon transplantation, are expected to migrate to different niches and integrate into functional structures in the central nervous system. This makes them possible candidates for in vivo delivery and expression of therapeutic genes. In addition, NP cells have potential to serve as an abundant cell source for basic research and toxicity screens, as well as further our understanding of neurodegenerative diseases, that ultimately may lead to cell therapies (28).

The majority of multipotent neural stem cells are isolated from fetal or adult tissues, but lack molecular characteristics of NP cells of embryonic origin. hESC derived NP cells can overcome this limitation and effectively utilized for developmental disease models (1). However, our ability to characterize and monitor live cells in culture or in the recipient animal models is limited. Expression of fluorescent reporter proteins in NP cells, achieved by genetic manipulation, can overcome this challenge allowing for broad applications in research and therapy (14). Additionally, methods that result in the successful expression of reporter fluorescent proteins within NP cells can potentially serve as a general guideline for genetic modification with other genes in NP cells.

Therefore, it is important to explore methods to genetically modify NP cells with exogenous marker genes.

Human NP cells can be genetically manipulated by two approaches. Since NP cells are derived from hESC, one can generate stable hESC clones that express a transgene of interest, and then differentiate the clones into NP cells. Alternatively, hESCs can be differentiated into NP cells first, and the transgene delivered afterwards. Previous attempts by other groups had varying success in generating stable hESC transfectants due, primarily, to a very poor rate of stable transfection in hESCs (~10^{-6}) (10). Therefore, the generation of hESC clones expressing transgenes still presents a great challenge. Additionally, in spite of stable integration of transgene into the host genome, expression, especially that for fluorescent proteins, is rarely maintained for prolonged period in transfected cells.

Significant efforts have been made to improve gene delivery systems including the use of various promoters. The human cytomegalovirus immediate-early (CMV) promoter is a popular candidate and has been used in many studies to drive the expression of transgenes. However, viral promoters (e.g. CMV) tend to have a propensity towards silencing (19). To circumvent the issue, human or non-human mammalian promoters like mouse Rex1 (10), Ubiquitin C (Ubc) (19), Ef1□ (16) and Rosa26 (19) were considered.

Various delivery systems, polyethylene compounds such as ExGen 500 (Fermentas, Inc) (10, 19), electroporation (33) and viral based delivery systems (25) are currently used in the field. Combination of proprietary transfecting reagent and electroporation (Nucleofection by Amaxa, Inc) has also produced superior results to

using either electroporation or chemical transfection (18, 29). Nucleofection followed by transduction in a two stage technique with retro/lenti-virus resulted in an increased efficiency of stable gene expression varying between 20-33% hESC. The stable hESC line retained expression of fluorescent protein upon differentiation to neural cells (16, 23) and allowed clonal isolation of NP cells in vitro (11).

Various genetic manipulation methods have been attempted in neural stem or progenitor cells from fetal or adult sources. Blits et al. (2005) used a lentiviral delivery system utilizing the CMV promoter for long-term expression of GFP or neurotrophic factors (BDNF, CNTF, D15A, GDNF, MNT and NT-3) in fetal rat NP cells for transplantation to spinal cord or brain (5). A similar approach has been tried in spinal cord model using retrovirus delivery system (24). However, so far no report is available on genetic modifications of NP cells derived from hESC.

Against this background, the objective of our study was to generate hESC derived NP cells expressing fluorescent proteins. To achieve that goal, we followed both approaches discussed above to produce NP cells expressing enhanced GFP (EGFP) fluorescent protein. Following the first approach, we produced an EGFP-positive stable hESC line. These cells were then differentiated down the neural lineage, and NP cells were examined to verify continued expression of fluorescent proteins. In the second approach, we generated three different NP cell lines from hESC cells. The efficiency of the three delivery systems was compared using a constitutive Ubc promoter driving EGFP, which showed that the lentivirus system had the optimal insertion efficiency and cell viability. Based on the results from both approaches, we conclude that the lentiviral system as a vehicle for gene delivery is the optimal method

to genetically modify hESC-derived NP cells and a differentiated derivative (e.g. glial-like cells). In addition, using the same lentiviral vehicle, we compared three different promoters (CMV, EF1α, and Ubc) for silencing and determined 70-80% of the transduced cell maintained flourescent protein expression over an extended period of culture.

Materials and Methods

hESC Culture Conditions

WA09 hESC with normal karyotype were cultured on ICR mouse (Harlan) embryonic fibroblast (MEF) feeders inactivated by mitomycin C (Sigma-Aldrich). The cells were cultured in 20% KSR stem cell media, which consists of Dulbecco's modified Eagle medium (DMEM)/F12 supplemented with 20% knockout serum replacement (KSR), 2 mM L-glutamine, 0.1 mM non-essential amino acids, 50 units/ml penicillin/50 μg/ml streptomycin (Gibco), 0.1mM β-mercaptoethanol (Sigma-Aldrich) and 4 ng/ml FGF2 (Sigma-Aldrich and R & D Systems). hESCs were maintained in 5% CO_2 and at 37° C. Cells were passaged as previously described by our laboratory (22).

Human Neural Progenitor (NP) Culture Conditions

NP were derived from WA09 human embryonic-stem cells (hESC) with normal karyotype as described previously by our laboratory (28). Cells were propagated and maintained in proliferation media consisting of neurobasal medium (Gibco) supplemented with 2mM L-glutamine, 1x penicillin/ streptomycin, 1x B27 (Gibco), 20ng/ml of FGF2 (Sigma-Aldrich) and 10ng/ml of Leukemia inhibitory factor (LIF, Chemicon) under conditions of 5% CO_2 at 37°C. Cells were passaged 1:3 or 1:4 by

mechanical dissociation via trituration and re-plated on polyornithine (20 µg/ml) and laminin (5 µg/ml) coated plates with media changes every other day.

Human Astrocytoma Cell Culture

Astrocytoma cells were purchased from ATCC (CRL-1718) and maintained in tissue culture according to supplier's instruction. Cells were sub-cultured on tissue culture grade dishes or tissue culture flasks in RPMI-1640 medium (ATCC # 30-2001) supplemented with 10% fetal bovine serum (FBS, Hyclone). Production of astrocytoma-conditioned medium (ACM) was achieved by adding differentiation media (FGF2 withdrawn from NP proliferation media) on CRL-1718 human astrocytoma cells for a period of 24 hours, after which, it was harvested and used for the differentiation of NP cells toward a $GFAP^+$ cell fate.

Neural Progenitor Differentiation Towards $GFAP^+$ Lineage

$GFAP^+$ cells were enriched in an adherent culture system on polyornithine (20 µg/ml) and laminin (5 µg/ml) coated plates in 1:1 ratio of differentiation media and ACM media for 24 days under conditions of 5% CO_2 atmosphere at 37°C. Media was replaced every third day.

pFUGW GFAP:EGFP and GFAP: tdTomatoRed Packaging Vector Construction

pFUGW lentiviral packaging vector was a gift from Dr. James Lah at Emory University. The pGfa2-cLac plasmid (6) carrying the hGFAP promoter was a gift from Dr. Michael Brenner at University of Alabama. The hGFAP promoter was excised from pGfa2-cLac plasmid using *Bgl* II and *Sal* I and inserted into *Bgl* II and *Sal* I digested shuttle vector DsRed Express-1 (Clontech). The shuttle vector was digested with *Afe* I and *Bam* HI and the promoter was excised and inserted into like digested pFUGW,

yielding pFUGW GFAP:EGFP. Neomycin resistance was conferred through digestion of DsRed Express-1 (Clontech) with *Dra* I and *Dra* III and excision of the neomycin-kanamycin resistance gene cassette including the SV40 promoter regions. The *Dra* III end of the fragment was blunted (NEB Quick Blunting Kit, cat. E1201S) and ligated into *Afe* I digested pFUGW GFAP:EGFP creating pFUGW GFAP:EGFP :: SV40:NeoR. The insertions removed the ubiquitin C (Ubc) sequence from the pFUGW plasmid.

In addition, EGFP was replaced by tandem dimer Tomato Red fluorescent protein (tdTomatoRed, GenBank Id: 55420623) to generate GFAP driving tdTomatoRed fluorescent proteins. For this, the digested *BamH* I – *Not* I fragment containing tdTomatoRed sequence from pRSET-beta tdTomato plasmid (a gift from Dr. Roger Tsien) was inserted into the shuttle pZsGreen1N1 (Clonetech) plasmid with like *BamH* I – *Not* I opened sites. The fragment was recovered via digestion with *Afl* II (blunted after digestion) and then *BamH* I. Replacement of the EGFP by this fragment within, pFUGW GFAP:EGFP :: SV40:NeoR was accomplished by digestion with *EcoR* I (subsequently blunted) and then with *BamH* I and the tdTomatoRed fragment was ligated into this site. The newly constructed plasmid is called pFUGW GFAP:tdTomatoRed :: SV40:NeoR. EGFP and tdTomatoRed sequences were confirmed by sequencing using primers located on the GFAP promoter sequence (4).

Lentiviral Production and Transduction of NP Cells

Lentiviral particles were produced and titered in the laboratory of Dr. James Lah (Emory University). Vector particles were produced in HEK293T cells by transient co-transfection with the pFUGW transfer vector, the HIV-1 packaging vector ΔR8.9, and the VSV-G (Vesicular Somatitis Virus- G) envelope glycoprotein. Media was replaced the

morning post-transduction and then placed back in the incubator. After 28 hours, the virus-containing supernatant was removed, filtered through a 0.45μM filter unit, concentrated by centrifugation, aliquoted and frozen at -80^{o}C. Viral titers were calculated using a serial dilution method.

Two days prior to transduction, a confluent 35 mm dish of NP cells was prepared and maintained as described above. The following day, the dish was split 1:2 with fresh proliferation media added and incubated overnight under conditions of 5% CO_2 atmosphere at 37°C. With NP cells at ~60% confluence, viral particle aliquots were combined with 2 ml of warmed proliferation media to a multiplicity of infection (MOI) of 5 or 10 and added to 1 x 35 mm plate of NP cells as if normal media change. Media was also changed in the remaining plate and used as the control. Both plates were centrifuged (250g for 5 minutes at 23°C) and incubated overnight (37°C with a 5% CO_2 atmosphere). Twenty-four hour post-transduction, the media was aspirated off, washed with PBS$^{+/+}$ (phosphate buffered saline with Ca^{+2} and Mg^{+2}) and replaced with 2 ml of fresh NP proliferation media.

Transfection

NP cells were transfected using ExGen500 (MBI Fermentas) reagent following a protocol adapted for hESC transfection (8). Briefly, 1 x 10^6 NP cells were plated overnight in 35 mm tissue culture dish. Immediately before transfection, cells were replenished with fresh 1 ml media. 2 μg DNA were diluted in 100 μl of 150mM NaCl with 3.3 μl of ExGen 500 then added drop by drop. The solution was vortexed for 10 seconds and spun briefly. The DNA–reagent solution was incubated at room temperature for 10 minutes. After incubation, DNA was added to the NP culture, gently

mixed and centrifuged (280g x 5 minutes). The culture was incubated (37°C with a 5% CO_2 atmosphere) for 30 min, after which the media containing the DNA–reagent solution was aspirated, the cells were washed with PBS and fresh 2ml media was added. The culture was returned to incubator (37°C with a 5% CO_2 atmosphere).

For stable transfection of NP cells by nucleofection, pZsGreen1N1 plasmid (Clontech) was linearized by ApaLI restriction enzyme (NEB). Cells were propagated 24 hours post nucleofection with G418 selection (200 µg/ml) 48 hours after re-plating. G418 selection was continued for two weeks.

Nucleofection

A total of 1 x 10^6 cells were resuspended in 100 µl of Nucleofector Solution Kit V at room temperature. Plasmid DNA (5 µg), diluted in TE buffer, was transferred into the nucleofection cuvette. The cell solution was added and nucleofection was performed as per the manufacturer's settings (electrical setting B-16). Immediately after pulsing, the cells were transferred to pre-warmed neural proliferation media and plated on polyornithine (20 µg/ml) and laminin (5 µg/ml) coated six-well plates.

For stable transfection of hESC by nucleofection, pZsGreen1N1 plasmid (Clontech) was linearized by ApaLI restriction enzyme (NEB). Cells were propagated 24 hours post nucleofection with G418 selection (200 µg/ml) 48 hours after re-plating. G418 selection was continued for two weeks.

Immunocytochemistry

Immunostaining was performed as described previously (7). In brief, cells were grown on permanox (NUNC) 4-chamber well slides, fixed in 4% paraformaldehyde (Sigma-Aldrich) for 15 minutes in room temperature. Antibodies were directed against

SOX2 (1:200, R&D Systems), Musashi1 (1:100, Neuromics), Nestin (1:100, Neuromics), TuJ1 (1:500, Neuromics), TH (1:50, Neuromics), and GABA (1:200, Neuromics). Primary antibodies were detected using appropriate fluorescent secondary antibodies (1:1,000 Molecular Probes-Invitrogen). All wells were counter-stained with DAPI. Picture acquisition was performed using an Olympus Ix81 with Disc-Spinning Unit and Slide Book Software (Intelligent Imaging Innovations) and analyzed by way of Image-Pro Plus (Media Cybernetics).

Reverse-Transcription Polymerase Chain Reaction (RT-PCR)

RNA was isolated from NP cells using RNeasy plus kit (Qiagen cat# 74134) as described previously (7). 1 µg RNA was reverse transcribed using Advantage RT-for-PCR kit (Clontech, Cat# 639505). PCR was performed using GoTaq Green Master Mix (Promega Cat# M7121) 60°C annealing temperature for 30 cycles. A list of primers and sequences is provided in Table 1.

Flow Cytometry and Cell Sorting (fluorescence-activated cell sorting, FACS)

Transduced NP cells were washed twice with $PBS^{+/+}$ to remove debris prior to collection. Cells were harvested by mechanical dissociation via trituration, centrifuged at 1,000 g at 23°C for 4 minutes. Pellet was gently resuspended in 0.5 ml of fresh pre-warmed neural proliferation media and stored on ice until visualized or sorted.

Flow cytometry was performed using FACS Calibur system (Becton Dickinson-BD Biosciences) and the data analyzed using FlowJo Cytometry analysis software (Tree Star). Propidium iodide was added to each sample to determine dead cells prior to run them on flow-cytometer. Forward- and side-scatter plots were used to exclude

dead cells and debris from the histogram analysis. For this experiment, GFP was excited at 488 and the fluorescence analyzed.

Cell sorting was performed on a MoFlo (Beckman Coulter, Hialeah, Florida) using a 100μm nozzle. GFP was excited at 488 and the fluorescence was collected through a 530/40 BP filter. Post sort, cells were collected in fresh pre-warmed neural proliferation media, centrifuged and re-plated on polyornithine (20 μg/ml) and laminin (5 μg/ml) coated plates with media changes every other day.

Statistical Analysis and graphics

Ananlysis of variance (ANOVA) and comparison of means (Tukey's test) were performed using GLM procedure by SAS 8.01 (SAS Institute, Cary, NC).

Results

Silencing of Stably Transfected Human Embryonic Stem Cells (hESC)

Our main objective was to generate self-renewing human neural progenitors (NP) expressing a fluorescent reporter protein. We first generated EGFP expressing stable hESC clones using a linear plasmid containing EGFP reporter under the CMV promoter (pZsGreen1N1). In transient transfection, delivery efficiency was higher with nucleofection (20-25% EGFP+ cells) than transfection with ExGen500 (~5%). Therefore, we delivered the linear construct using nucleofection and obtained ~10 stable clones per million transfected cells (Figure 2.1.A). Selection for stable clones was carried out by neomycin resistance (G418 selection, 200μg/ml) driven by the SV40 promoter. Only one of every five G418 resistant stable hESC clones generated by nucleofection expressed EGFP. These clones were continuously maintained and sub-cultured in the presence of G418.

After ten to twelve days of hESC differentiation, early neural rosette structures (28)) appeared and they retained EGFP expression like hESC clone. However, EGFP expression is lost before completion of the neural differentiation process. Loss of EGFP expression was also noticed in hESC clones within two months of their propagation as hESC in culture (10-15 passages), though they continued to be resistant to G418. Thus, the hESC clones were $EGFP^{-}/G418^{+}$ (Figure 2.1.A) and failed to generate $EGFP^{+}/G418^{+}$ neural rosettes. Due to these results, we attempted to genetically manipulate the NP cells derived from wild type hESCs.

Stable Transfection of Neural Progenitors Derived from Human ES cells

Due to silencing of the transfected reporter system during the differentiation of hESC into NP cells, we tried the second approach of transfecting NPs after differentiation. NPs were derived from WA09 hESCs as before (28) and were then transfected using the same linear construct and techniques described in the previous section. Similar to results observed with hESC, nucleofection produced more (15-20%; Figure 2.1.B) EGFP expressing cells than ExGen500 (5%) (not shown) after 24 hours of transfection. After more than 2 weeks of G418 selection, NP cells transfected with ExGen500 did not demonstrate resistance, while a few (<1 from every 5 million) NP cells that underwent nucleofection with linear construct were maintained in culture. However, none of the G418-resistant NP cells displayed EGFP expression. These results indicate that the delivery efficiency is a significant component to successful stable reporter expression of NP cells. Therefore, we further characterized the ability of several delivery systems with the aim to generate stably transfected NP cells that do not undergo silencing.

Neural Progenitors can be Genetically Marked by Different Methods

To determine the optimum delivery system for genetic manipulation, cells from three NP cell lines were delivered with a mouse Ubiquitin-C (Ubc) promoter EGFP construct (Figure 2.2). This was done using three different delivery methods: electroporation with the nucleofection kit (Amaxa), transfection with ExGen500 (MBI Fermentas) and a lentiviral transduction (5 MOI, multiplicity of infection) delivery system. The gene delivery efficiency of each method was determined by measuring the percentage of cells expressing EGFP and viability after 48 hours. ExGen500 transfection produced the least (<3%) EGFP-positive cells, whereas 15% of the nucleofected cells expressed EGFP (Figure 2.2). However, lentiviral transduction resulted in the maximum EGFP-positive cells (35 - 40%). Overall, viability of NP cells were highly compromised after genetic manipulation with only 6% of cells being alive after nucleofection, followed by 12.7% live cells after ExGen500 transfection and a significantly higher proportion of 37% ($P<0.05$) after lentiviral transduction (Figure 2.2). Results indicate all three lines showed similar viability and EGFP expression. Considering both viability and efficiency for delivering a transgene, lentiviral transduction appears to be the best method for genetic manipulation of hESC-derived NP cells of the methods evaluated.

Lentiviral Transduced Cells Express Neural Markers

We further tested if genetic manipulation by lentiviral transduction negatively affected the differentiation potential of NP cells. We differentiated the transduced NPs using differentiation media for 14 days. Transduced NP and differentiated neurons were then immunostained for known neural markers (Figure 2.3). The results demonstrated

that EGFP$^+$ NP cells expressed MUSASHI1, NESTIN, SOX2 (immunostaining, Figure 2.3.A), PAX6 and NCAM1 (RT-PCR, Figure 2.3.B). When transduced NPs were randomly differentiated into neurons, they maintained EGFP expression similar to NP cells. In addition, these neurons expressed beta III tubulin (TuJ1), gamma aminobutyric acid (GABA) and tyrosine hydroxylase (TH) (Figure 2.4). Expression of both neural progenitor markers (MUSASHI1, NESTIN, SOX2, PAX6 and NCAM1) before differentiation and neuronal markers (TuJ1, GABA and TH) after differentiation indicated that genetic manipulation did not alter the differentiation potential of hESC-derived NP cells.

Transduced Neural Porgenitors Provide Glial Lineage-Specific Expression

In order to extend the application of the lentiviral based delivery methods for studying NP differentiation, we examined the ability of these cells to express fluorescent proteins under the glial specific promoter glial fibrillary acidic protein (GFAP) after glial differentiation. For this purpose, the lentiviral plasmid was modified by replacement of the constitutive Ubc promoter with the hGFAP promoter (4, 6) expressing EGFP or tomato-Red (tdTomatoRed) fluorescent reporter proteins. Validation of the construct was performed using astrocytoma cells (ATCC # CRL-1718) that endogenously express GFAP. Transduction with the GFAP:EGFP lentivirus resulted in the expression of fluorescent protein and its co-localization with the GFAP antibody upon immunostaining (Figure 2.5.A). To generate glial-like cells from hESC-derived NPs, we differentiated them in the presence of combined ACM- Diffeerentiation Media (for detail, see Materials and Methods) for 24 days and then transduced the differentiated cells. Three days post-transduction, approximately 2-3% of the differentiated cells expressed EGFP,

indicating active indigenous GFAP promoter activity (Figure 2.5.B). This suggests that the lentiviral system could be optimized for genetically expressing exogenous genes in hESC-derived NP cells and further differentiated lineages.

Lentiviral System is Prone to Transgene Silencing in Neural Progenitor Cells

Lentiviral transduced NP cells showed varied EGFP expression under all promoters we examined. For a sustained transgene expression in NP cells, it was vital that the inserted promoter could evade silencing during propagation of NP cells in culture. For this purpose, we compared three different constitutive promoters, CMV, EF1α and Ubc for their efficiency at avoiding transgene silencing. Forty eight hours after gene delivery (10 MOI, multiplicity of infection), we quantified $EGFP^+$ cells and viability by flow cytometry (Figure 2.6.B). Transducing the cells using the CMV promoter resulted in a population where 37% of the cells were EGFP-positive. This was significantly ($p< 0.05$) higher than the percentage of cells that were positive using the mammalian promoters EF1α (27%) and Ubc (15%) (Figure 2.6.B,C). We observed higher level (~3.5% cells) of auto-fluorescence in culture with cells transduced with CMV promoter than other two promoters ($P<0.05$) (Figure 2.6.B). EGFP signals were more readily detected with CMV based transduction, suggesting the CMV promoter is more efficient in transcription of EGFP over the other two promoters.

Transduced cells were further sorted to obtain a stock of 100% $EGFP^+$ NP cells (at this point, called passage P0) and re-plated for subculture. About 1 month later, at P5 we performed flow cytometry analysis to quantify the percentage of $EGFP^+$ cells. This was repeated at P10 approximately 2 months after sorting. Both passage numbers and promoters significantly affected the number of overall $EGFP^+$ cells ($P<0.0004$).

Irrespective of promoters used, we noticed a loss in the percentage of $EGFP^+$ cells at both P5 and P10. At passage 5, we found 68% $EGFP^+$ cells for both promoters CMV and Ubc, while 78% of the cells under the EF1α promoter remained EGFP-positive (Figure 2.6.C). However, at P10, all three transduced cell populations showed about 80% EGFP-positive cells. The percent of $EGFP^+$ cells generated with EF1α promoter did not differ between passage 5 and 10. For the other two promoters, percent of $EGFP^+$ P5 and P10 cells differed significantly ($P<0.05$).

We further quantified median relative fluorescent intensity (RFI) in transduced and non-transduced NP cells (negative control). The RFI median values varied between ~70 to ~90 channel numbers with the Ubc promoter at P10 producing the highest channel number and the CMV promoter producing the lowest channel number at P5. Similar to the percentage of GFP positive cells, RFI measures were also affected by passage number and promoter significantly ($P<0.0001$). After transduction, CMV driven EGFP expression appeared to have the brightest fluorescent expression (Figure 2.6.A). However, after propagation of the NP cells, the average RFI median values did not differ amongst all promoter groups within each passage time point (P5 or P10). There was an increase in RFI value when cells were propagated from P5 to P10. Therefore, overall there was evidence of silencing of all promoters we tested and the percentage of cells that underwent silencing varied between 20-30% over a period of two months in culture.

Discussion

Human embryonic stem cell (hESC) derived neural progenitors (NP) are potentially an important material to further our understanding of neural development and

differentiation. We demonstrate that NP cell lines can be genetically manipulated with high efficiency and reliably via a lentiviral delivery system. Additionally, this system can be used for tracing both constitutive and lineage specific genes in these NP cells. The lentiviral system did not affect the differentiation potential of NP cells. Furthermore, we demonstrated that cell sorting based on reporter expression can effectively be used for enriching transgene expressing cells. For long-term expression, both viral and eukarytoic promoters tested could be useful. NPs cells transduced with eukarytoic promoter EF1α produced a highly enriched and stable population earlier (<5 passages) than other promoters. Previous attempts have been made to generate neural tissues from hESC that express reporter proteins (17, 30); however, to the best of our knowledge, we are the first to report a hESC-derived self-renewing cell population that was integrated with a permanent reporter system for continuous cell tracking capabilities.

A number of laboratories have produced stable hESC clones (10) and demonstrated that stable lines could generate cell types belonging to three germ layers namely, ectoderm, mesoderm and endoderm (12, 20). Of different methodologies used for transfection, nucleofector reagent (Amaxa) was also used for generating stable hESC clones (16, 18, 29). Using nucleofection kit V and B16 program, Siemen et al. demonstrated a population where 66% of cells were positive for the introduced transgene (29). Therefore, we used this same technique to generate stable hESC clones and NP lines. We succeeded in producing stable drug resistant hESC clones, but not NP populations (Figure 2.1). We obtained about one EGFP expressing clone for every five drug-resistant hESC clones which was almost twice the reported efficiency

(29). We could expand those hESC clones (tested for 10 passages) and freeze/thaw without affecting the colony morphology or EGFP expression. However, nucleofection with NP cells did not yield any of stable cell lines expressing drug resistance or EGFP possibly due to poor clonal propagation (11). Additionally, efficiency in gene delivery might be another issue that has adversely affected the outcome of this study.

This led us to examine different gene delivery systems with three different NP lines derived from WA09 hESC. Previously Ubc promoter was identified to perform better compared to viral promoters in hESC (19). To examine gene delivery systems, we therefore used a plasmid that would constitutively express EGFP by mouse-Ubc promoter. We choose to employ nucleofection and two other packaging systems, ExGen500 (MBI Fermentas) and lentiviral transduction. ExGen500 was the first reagent that resulted in superior transfection of hESC (10) and a number of laboratories have successfully used vesicular somatitis virus G (VSV-G) pseudotyped lentiviral transduction to transduce a variety of hESC lines (3, 13, 15, 16, 21, 25, 32). When VSV-G pseudotyped lentivirus used for transduction, feeder cells MEF were also transduced in hESC culture, reducing efficiency of hESC transduction (15). This factor did not affect NP transduction for they do not require any feeder support and can be grown on laminine-poly-ornithine matrices.

Unlike transfection of hESC by nucleofection and ExGen500, where cell viability is minimally affected (survival >70%) (9, 29), viability of NP cell lines were adversely affected by nucleofection and ExGen500 suggesting a cytotoxic effect to the transfection solutions or an effect of the technique itself. Conversely, cell survival was not affected by lentiviral transduction (Figure 2.2). It is important to note that viability of

wild type NP cells also suffered substantially during flow-cytometry experiment. We achieved similar transduction efficiencies with WA09 hESC-derived NP cells under mouse Ubc promoter (Figure 2.2) as other groups have with hESC 38-87% $EGFP^+$ cells (13, 21). The results also suggest efficiency of gene transfer varies greatly between the delivery systems.

For an efficient gene delivery system, maintenance of stem cell properties is critical. Our results demonstrate that lentivirally transduced NP cells do express neural progenitor markers (Figure 2.3) (7, 28) and maintain their ability to differentiate as indicated by the expression of TuJ1, GABA, TH while retaining the EGFP expression (Figure 2.3 and 2.4). These transduced cells have been maintained in culture for over 30 passages. We further tested the ability of this system to monitor the onset of GFAP expression in differentiating NP cells, denoting a glial-like cell fate. EGFP expression could be detected in glial-like cells after 24 days of differentiation. Considering viability, efficiency, and long-term reporter expression, the lentiviral based gene delivery system serves as the best option for transgenesis in hESC-derived NP cells. Additionly, it is capable of accomplishing efficient transduction without compromising the broad cellular features, flexibility in employing ubiquitous and lineage-specific promoter in NP cells.

Important applications of NP cell transgenesis include animal transplantation studies and ex vivo cell based assays. Cell tracking capabilities of these manipulated cells is imperative and allows for the constant monitoring as differentiation occurs in vivo or ex vivo. Often viral vectors are noted for their inability to sustain transgene expression (19) and this is specifically well known for cells that divide rapidly (e.g. hESC). Recently it has been shown that expression of transgenes delivered by

lentiviral vectors were suppressed in a promoter-dependant manner specifically in hESCs, but not in mouse ESCs (31). In the absence of data from other laboratories on transgenesis in NP cells, we discuss our results in the context of hESC cells. Of different promoters we examined, we had higher percentage of $EGFP^+$ NP cells with CMV (37% versus 1.1% in hESC) and EF1α (27% in NP versus 18.6% in hESC) promoter. We used 10 MOI for achieving higher transduction efficiency than reported (31). Koch et al. (2006) obtained similar transduction efficiency (26% $EGFP^+$ hESC cells) with EF1α promoter in hESC (16). To enhance $EGFP^+$ cells in culture we sorted and maintained these cells over an extended period in culture (more than two months), and during this time, tested the reporter expression by flow cytometry at passage 5 and 10. In contrast to results in hESC described by Xia et al. (31), we did not observe any preferential silencing among the three promoters (CMV, Ubc and EF1α) driving EGFP expression. After two months in culture, 70-80% of the total cells were still $EGFP^+$. Taking into account the temporal effect on EGFP expression, among the three promoters, EF1α varied the least and had higher fluorescence intensity during the course of the study (Figure 2.6.C). Considering the duration of reporter expression in culture, it is expected that these transduced NP cells driven by all three promoters could be useful for in vivo transplantation studies. It should be noted that transduced NP cells described here were not subjected to clonal analysis. Also, we did not analyze insert copy number and integration sites. Therefore, for transplantation purposes, the risks associated with specific or random integration are not apparent.

In our study, silencing was minimal (about 20%) in NP cells transduced with lentiviral vectors. However, another possible technique to further reduce the observed

silencing could be the use of an IRES (internal ribosome entry site) element linked to drug resistant genes under the same promoter in the same vector. This strategy was successfully reduced the duration of selection from 14 days to 3 days in hESC (16).

hESC-derived NP cells better suited for maintenance of transgene expression compared to highly proliferating hESC. They provide a potentially unlimited cell source for basic and therapeutic applications devoid of immortalization techniques unlike primary cell lines from adult tissue. The coupling of these cells with lentiviral transduction may allow for the controlled release of therapeutic molecules including neurotransmitters or immuno-suppressive proteins that would reduce tissue rejection from heterologous origin. Transduced cells even may serve as carriers of shRNA molecules for targeting aberrant gene expression in disease models (2, 26).

References

1. **Aharonowiz, M., O. Einstein, N. Fainstein, H. Lassmann, B. Reubinoff, and T. Ben-Hur.** 2008. Neuroprotective effect of transplanted human embryonic stem cell-derived neural precursors in an animal model of multiple sclerosis. PLoS ONE **3:**e3145.
2. **An, D. S., Y. Xie, S. H. Mao, K. Morizono, S. K. Kung, and I. S. Chen.** 2003. Efficient lentiviral vectors for short hairpin RNA delivery into human cells. Hum Gene Ther **14:**1207-12.
3. **Ben-Dor, I., P. Itsykson, D. Goldenberg, E. Galun, and B. E. Reubinoff.** 2006. Lentiviral vectors harboring a dual-gene system allow high and homogeneous transgene expression in selected polyclonal human embryonic stem cells. Mol Ther **14:**255-67.
4. **Besnard, F., M. Brenner, Y. Nakatani, R. Chao, H. J. Purohit, and E. Freese.** 1991. Multiple Interacting Sites Regulate Astrocyte-Specific Transcription of the Human Gene for Glial Fibrillary Acidic Protein. Journal of Biological Chemistry **266:**18877-18883.
5. **Blits, B., B. M. Kitay, A. Farahvar, C. V. Caperton, W. D. Dietrich, and M. B. Bunge.** 2005. Lentiviral vector-mediated transduction of neural progenitor cells before implantation into injured spinal cord and brain to detect their migration, deliver neurotrophic factors and repair tissue. Restor Neurol Neurosci **23:**313-24.
6. **Brenner, M., W. C. Kisseberth, Y. Su, F. Besnard, and A. Messing.** 1994. Gfap Promoter Directs Astrocyte-Specific Expression in Transgenic Mice. Journal of Neuroscience **14:**1030-1037.

7. **Dhara, S. K., K. Hasneen, D. W. Machacek, N. L. Boyd, R. R. Rao, and S. L. Stice.** 2008. Human neural progenitor cells derived from embryonic stem cells in feeder-free cultures. Differentiation **76:**454-64.

8. **Drukker, M., S. K. Dhara, and N. Benvenisty.** 2005. Genetic engineering of human embryonic stem cells p. 215-230. *In* J. S. Odorico, S.-C. Zhang, and R. A. Pedersen (ed.), Human embryonic stem cells. Garland Science/BIOS Scientific Publishers, Oxford.

9. **Eiges, R.** 2006. Genetic manipulation of human embryonic stem cells by transfection. Methods Mol Biol **331:**221-39.

10. **Eiges, R., M. Schuldiner, M. Drukker, O. Yanuka, J. Itskovitz-Eldor, and N. Benvenisty.** 2001. Establishment of human embryonic stem cell-transfected clones carrying a marker for undifferentiated cells. Curr Biol **11:**514-8.

11. **Elkabetz, Y., G. Panagiotakos, G. Al Shamy, N. D. Socci, V. Tabar, and L. Studer.** 2008. Human ES cell-derived neural rosettes reveal a functionally distinct early neural stem cell stage. Genes Dev **22:**152-65.

12. **Gerrard, L., D. Zhao, A. J. Clark, and W. Cui.** 2005. Stably transfected human embryonic stem cell clones express OCT4-specific green fluorescent protein and maintain self-renewal and pluripotency. Stem Cells **23:**124-33.

13. **Gropp, M., P. Itsykson, O. Singer, T. Ben-Hur, E. Reinhartz, E. Galun, and B. E. Reubinoff.** 2003. Stable genetic modification of human embryonic stem cells by lentiviral vectors. Mol Ther **7:**281-7.

14. **Jakel, R. J., B. L. Schneider, and C. N. Svendsen.** 2004. Using human neural stem cells to model neurological disease. Nat Rev Genet **5:**136-44.

15. **Jang, J. E., K. Shaw, X. J. Yu, D. Petersen, K. Pepper, C. Lutzko, and D. B. Kohn.** 2006. Specific and stable gene transfer to human embryonic stem cells using pseudotyped lentiviral vectors. Stem Cells Dev **15:**109-17.

16. **Koch, P., H. Siemen, A. Biegler, J. Itskovitz-Eldor, and O. Brustle.** 2006. Transduction of human embryonic stem cells by ecotropic retroviral vectors. Nucleic Acids Res **34:**e120.

17. **Ladewig, J., P. Koch, E. Endl, B. Meiners, T. Opitz, S. Couillard-Despres, L. Aigner, and O. Brustle.** 2008. Lineage selection of functional and cryopreservable human embryonic stem cell-derived neurons. Stem Cells **26:**1705-12.

18. **Lakshmipathy, U., B. Pelacho, K. Sudo, J. L. Linehan, E. Coucouvanis, D. S. Kaufman, and C. M. Verfaillie.** 2004. Efficient transfection of embryonic and adult stem cells. Stem Cells **22:**531-43.

19. **Liew, C. G., J. S. Draper, J. Walsh, H. Moore, and P. W. Andrews.** 2007. Transient and stable transgene expression in human embryonic stem cells. Stem Cells **25:**1521-8.

20. **Liu, Y. P., O. V. Dovzhenko, M. A. Garthwaite, S. V. Dambaeva, M. Durning, L. M. Pollastrini, and T. G. Golos.** 2004. Maintenance of pluripotency in human embryonic stem cells stably over-expressing enhanced green fluorescent protein. Stem Cells Dev **13:**636-45.

21. **Ma, Y., A. Ramezani, R. Lewis, R. G. Hawley, and J. A. Thomson.** 2003. High-level sustained transgene expression in human embryonic stem cells using lentiviral vectors. Stem Cells **21:**111-7.

22. **Mitalipova, M., J. Calhoun, S. J. Shin, D. Wininger, T. Schulz, S. Noggle, A. Venable, I. Lyons, A. Robins, and S. Stice.** 2003. Human embryonic stem cell lines derived from discarded embryos. Stem Cells **21:**521-526.

23. **Nolden, L., F. Edenhofer, S. Haupt, P. Koch, F. T. Wunderlich, H. Siemen, and O. Brustle.** 2006. Site-specific recombination in human embryonic stem cells induced by cell-permeant Cre recombinase. Nat Methods **3:**461-7.

24. **Ohori, Y., S. Yamamoto, M. Nagao, M. Sugimori, N. Yamamoto, K. Nakamura, and M. Nakafuku.** 2006. Growth factor treatment and genetic manipulation stimulate neurogenesis and oligodendrogenesis by endogenous neural progenitors in the injured adult spinal cord. J Neurosci **26:**11948-60.

25. **Pfeifer, A., M. Ikawa, Y. Dayn, and I. M. Verma.** 2002. Transgenesis by lentiviral vectors: lack of gene silencing in mammalian embryonic stem cells and preimplantation embryos. Proc Natl Acad Sci U S A **99:**2140-5.

26. **Sapru, M. K., J. W. Yates, S. Hogan, L. Jiang, J. Halter, and M. C. Bohn.** 2006. Silencing of human alpha-synuclein in vitro and in rat brain using lentiviral-mediated RNAi. Exp Neurol **198:**382-90.

27. **Shin, S., S. Dalton, and S. L. Stice.** 2005. Human motor neuron differentiation from human embryonic stem cells. Stem Cells Dev **14:**266-9.

28. **Shin, S., M. Mitalipova, S. Noggle, D. Tibbitts, A. Venable, R. Rao, and S. L. Stice.** 2006. Long-term proliferation of human embryonic stem cell-derived neuroepithelial cells using defined adherent culture conditions. Stem Cells **24:**125-38.

29. **Siemen, H., M. Nix, E. Endl, P. Koch, J. Itskovitz-Eldor, and O. Brustle.** 2005. Nucleofection of human embryonic stem cells. Stem Cells Dev **14:**378-83.

30. **Singh Roy, N., T. Nakano, L. Xuing, J. Kang, M. Nedergaard, and S. A. Goldman.** 2005. Enhancer-specified GFP-based FACS purification of human spinal motor neurons from embryonic stem cells. Exp Neurol **196:**224-34.

31. **Xia, X., Y. Zhang, C. R. Zieth, and S. C. Zhang.** 2007. Transgenes delivered by lentiviral vector are suppressed in human embryonic stem cells in a promoter-dependent manner. Stem Cells Dev **16:**167-76.

32. **Xue, T., H. C. Cho, F. G. Akar, S. Y. Tsang, S. P. Jones, E. Marban, G. F. Tomaselli, and R. A. Li.** 2005. Functional integration of electrically active cardiac derivatives from genetically engineered human embryonic stem cells with quiescent recipient ventricular cardiomyocytes: insights into the development of cell-based pacemakers. Circulation **111:**11-20.

33. **Zwaka, T. P., and J. A. Thomson.** 2003. Homologous recombination in human embryonic stem cells. Nat Biotechnol **21:**319-21.

Figure 2.1. Genetic Manipulation in Human Embryonic Stem Cells (hESC) and Neural Progenitors (NP) Derived from hESC

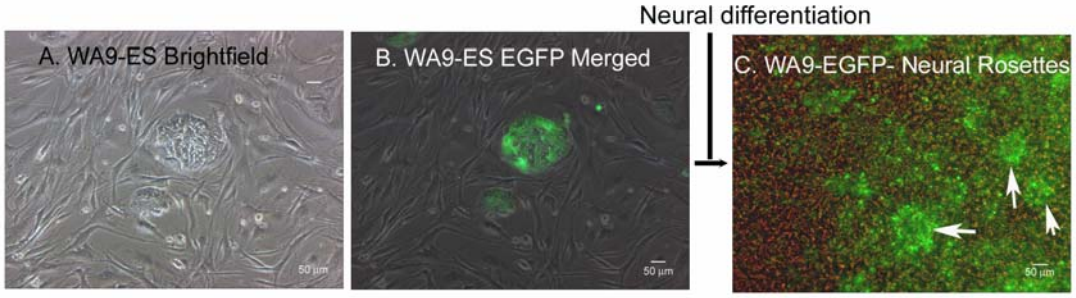
I. Schematic diagram of the construct

II.A, B. Stable hESC clone (WA09 line) expressing EGFP generated using nucleofection (Amaxa kit). This clone, differentiated to the neural lineage and formed neural rosette structures expressing EGFP (arrow heads in (**C**). The cells from these neural rosettes do not, however, maintain EGFP expression in prolonged culture. In an alternate approach, NP cells were first derived from hESC cells, and then stably transfected with the same construct. Shown here bright field (**D**), EGFP (**E**) and merged (**F**) images of NP cells. Though many cells received the construct and initially expressed the reporter protein, they did not survive in the presence of G418 (neomycin sensitive). Continued expression in long term culture was not achieved with either approach.

I. EGFP construct used for stable transfection

II. Stable transfection in hESC and NP cells

a) Stable trasfection of WA9 ES cells and neural differentiation



b) Stable transfection of NP derived from wild type WA09 ES cells

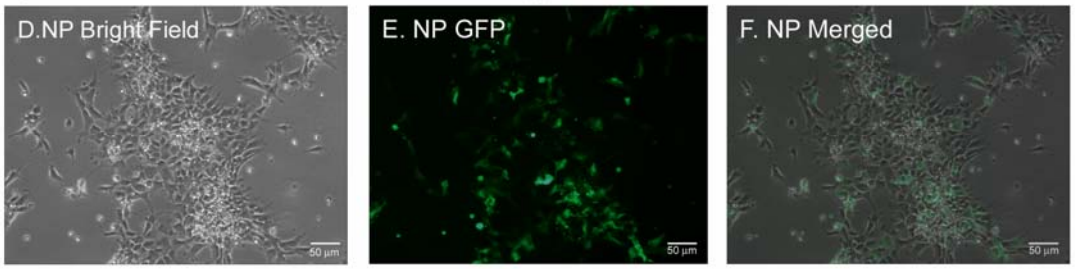


Figure 2.2. Comparison of Efficiency and Viability of NP cells for Three Different Methods of Gene Delivery. NP lines derived from WA9 hESC were transfected with EGFP under Ubiquitin C (Ubc) promoter (construct shown in Figure 2.3.A) using 3 different delivery methods: Nucleofection, Transfection with ExGen 500 and lentiviral transduction. Proportion of cells expressing EGFP and viability were determined using flow-cytometry. Mean and standard errors are shown for percentage of live cells and EGFP expression for each delivery method. Superscripts (α, β, γ for viability and a, b, c for EGFP expression) indicate statistical difference at $P<0.05$ significance level.

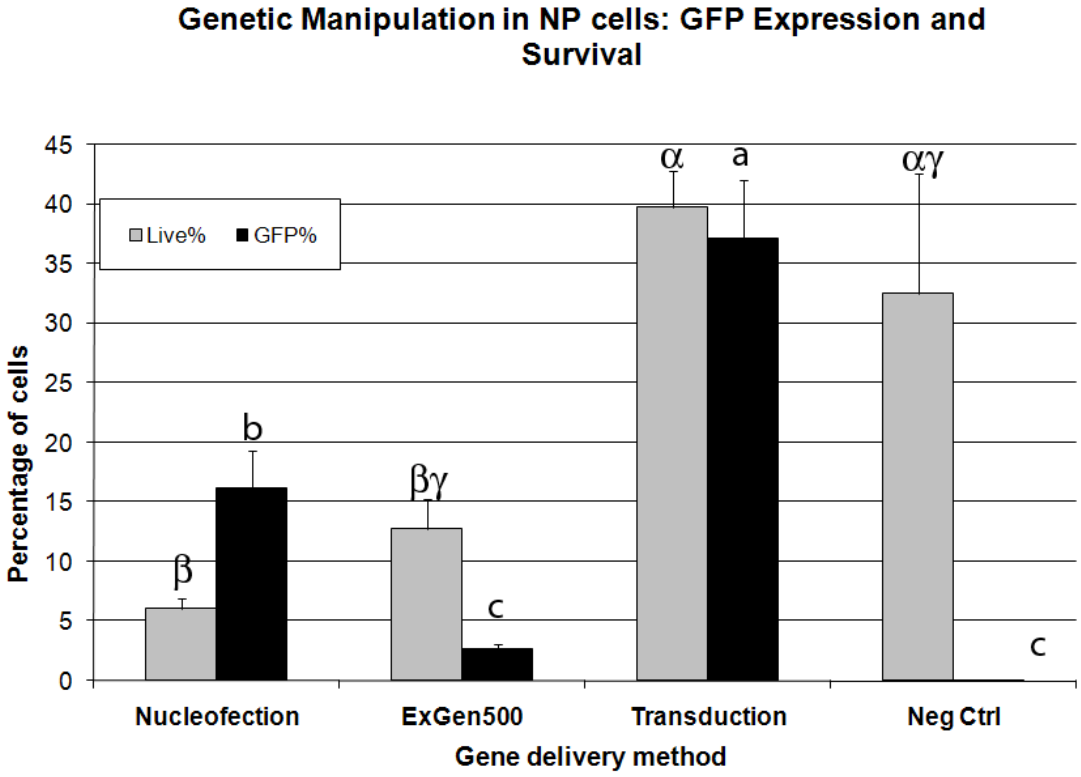

Genetic Manipulation in NP cells: GFP Expression and Survival
Live%
GFP%
Percentage of cells
45
40
35
30
25
20
15
10
5
0
β
b
βγ
c
α
a
αγ
c
Nucleofection
ExGen500
Transduction
Neg Ctrl
Gene delivery method

Figure 2.3.A. <u>NP cells Transduced using lenti-viral system maintain expression of neural progenitor markers</u>

I. Schematic diagram of lentiviral construct used for transduction of NP cells.

II. Shown are both live (A-C, top row) and immunostained images (**D-L**)of NP cells. Ubc-EGFP transduced NP cells express EGFP (**A, D, G** and **J**). In each row, for every EGFP field, shown bright field (**B**), or immunostained (red) images of neural progenitor markers, Musashi1 (**E**), Nestin (**H**) and SOX2 (**K**). The rightmost column shows the merged images (**C, F, I, L**) of EGFP field with bright field or immunostained images presented in first two columns (**A-B, D-E, G-H, J-K**). Nuclei were stained with DAPI (blue).

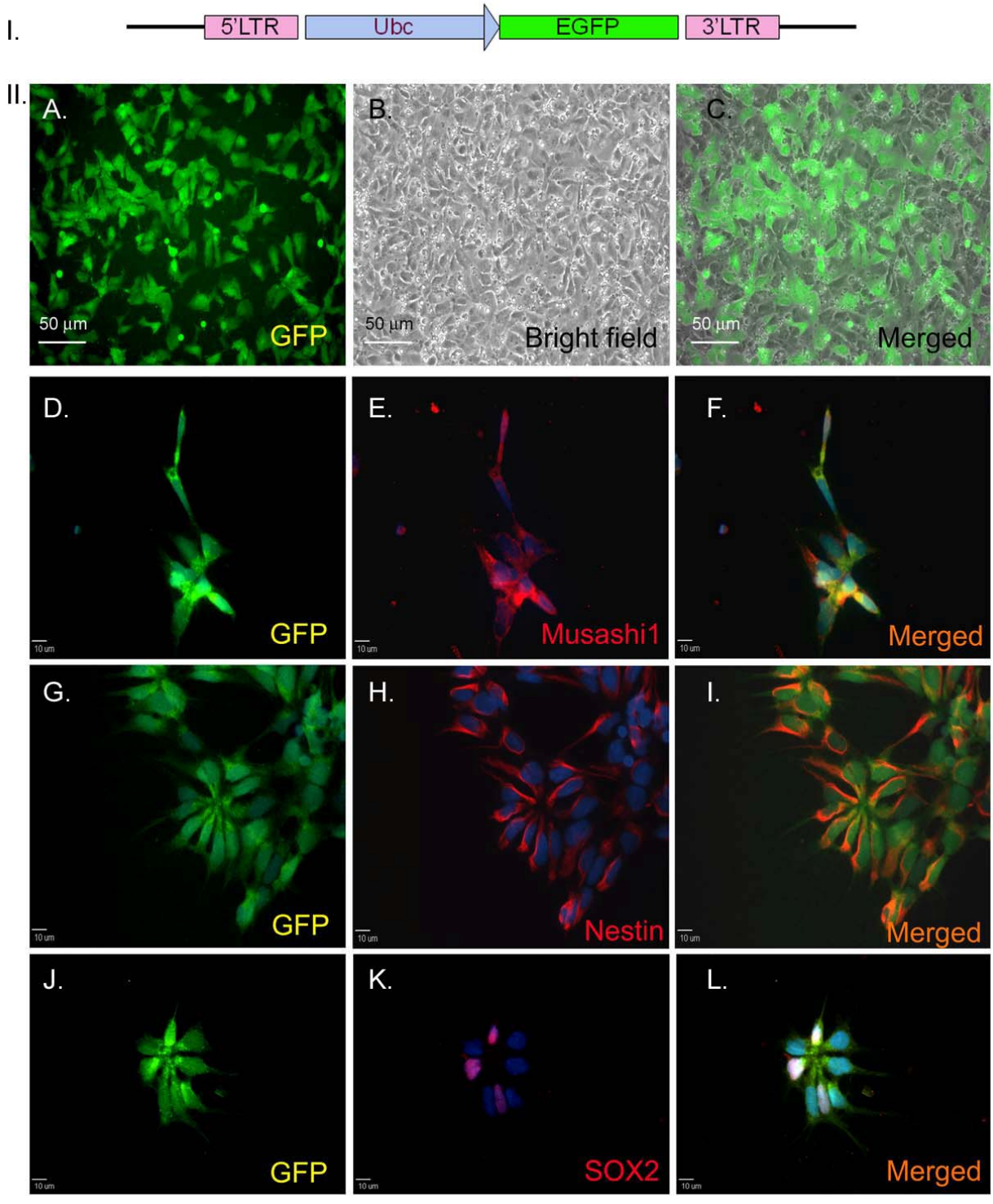

I.
5'LTR
Ubc
EGFP
3'LTR
II.
A.
50 µm
GFP
B.
50 µm
Bright field
C.
50 µm
Merged
D.
GFP
E.
Musashi1
F.
Merged
G.
GFP
H.
Nestin
I.
Merged
J.
GFP
K.
SOX2
L.
Merged

Figure 2.3.B. **NP cells Transduced using Lentiviral System Maintained Expression of Neural Progenitor Marker Transcripts.** RT-PCR was used to detect expression of neural progenitor markers in NP cells transduced with EGFP under 3 different promoters, using lentiviral delivery and cultured for a prolonged period (>4 months). Transduced cells express neural progenitor markers (PAX6, NESTIN, NCAM1) as well as EGFP. Non-transduced NP cells are shown as a negative control for this experiment.

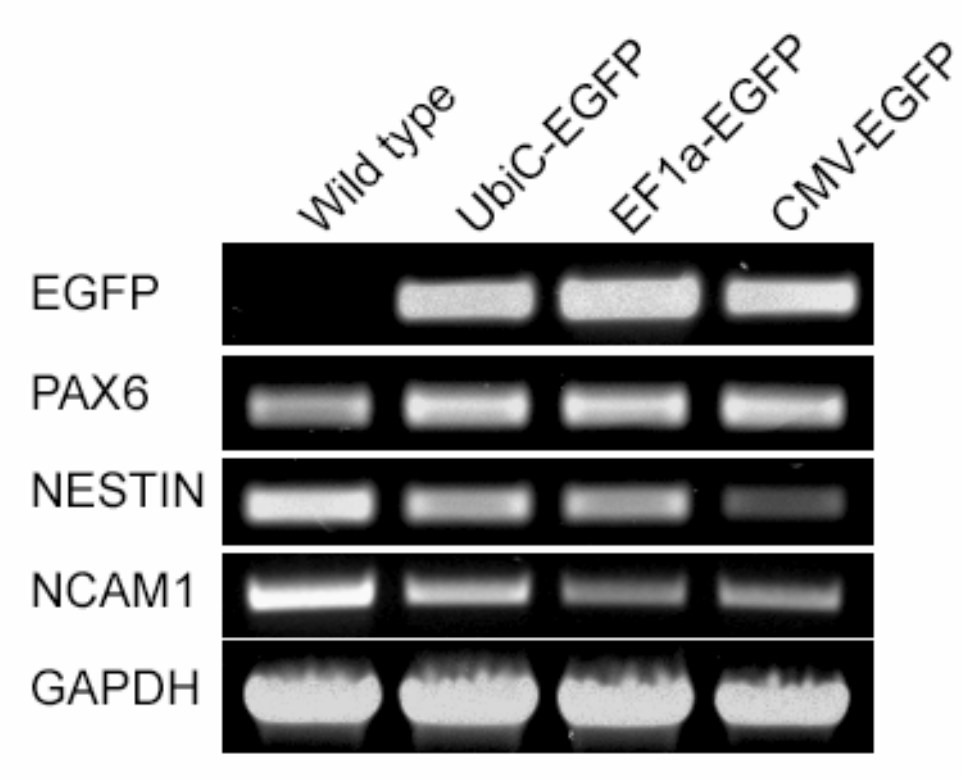

Wild type
UbiC-EGFP
EF1a-EGFP
CMV-EGFP
EGFP
PAX6
NESTIN
NCAM1
GAPDH

Figure 2.4. NP cells Transduced by Lentiviral System Maintained their Differentiation Potential As Well As the Transgene Expression. NP cells were transduced with Ubc-EGFP construct (shown in Figure 2.3.A) and differentiated for two weeks by withdrawing growth factor FGF2 from the proliferation medium. Live cell images are shown in (A) (EGFP), (**B**) (corresponding bright field) and (**C**) (merged of both fields). Differentiated cells were immunostained with antibodies for different neuronal markers: (**E**) TuJ1 (neuronal beta III tubulin), (**H**) GABA (gamma amino butyric acid, an inhibitory neuronal marker) and (**K**) TH (tyrosine hydroxylase, an excitatory neuronal marker). These markers co-localized with GFP expression, as shown. Nuclei were stained with DAPI.

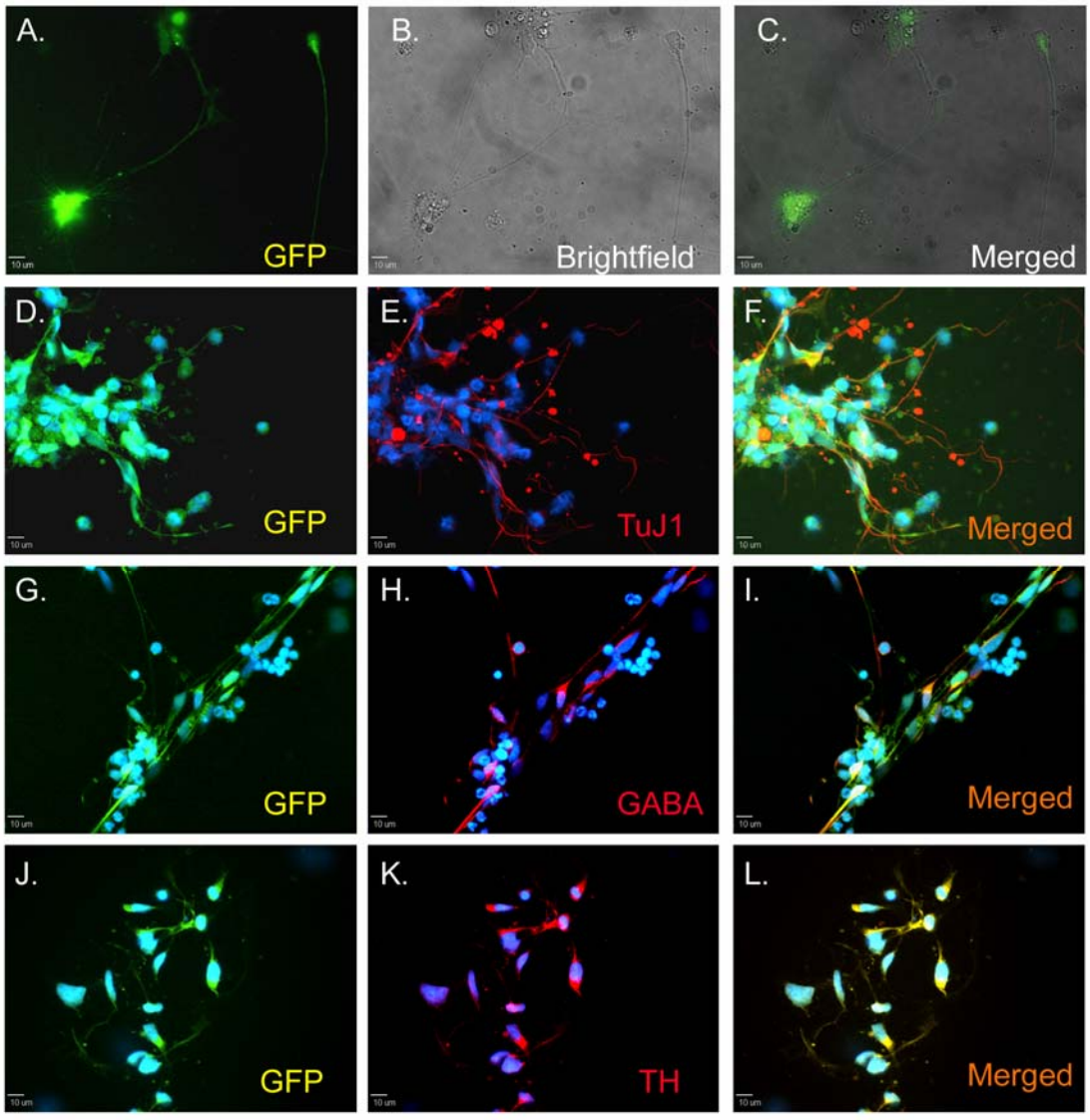

A.
GFP
B.
Brightfield
C.
Merged
D.
GFP
E.
TuJ1
F.
Merged
G.
GFP
H.
GABA
I.
Merged
J.
GFP
K.
TH
L.
Merged

Figure 2.5. Lentiviral Transduction can Track Astrocyte-like Cells Expressing Glial Fibrillary Acidic Protein (GFAP) in Culture

I. Schematic diagrams of lentiviral constructs

II. For validation of the construct, human astrocytoma (ATCC# CRL-1718) cells were transduced (**A-F**) with the lentiviral construct. Shown here are expression of (**A**) EGFP, (**B**) corresponding bright field and (**C**) merged images for the GFAP:GFP lentiviral construct. Transduction of human astrocytoma cells with GFAP:tdTomato RFP viral particles were used for further validation using immunostaining. Tomato-Red fluorescent protein expressing human astrocytoma cells (**D**) also express GFAP protein (immunostained in **E**, green). Endogenous GFAP co-localized with reporter protein (**F**). Nuclei were stained with DAPI (blue).

NP cells were differentiated for 24 days in ACM-Differentiation media and transduced with GFAP:GFP virus. Upon differentiation, these cells appeared morphologically similar to that of astrocytes (bright field, **H**) and several of them express EGFP (**G**) 48 hours post-transduction. The merged image of bright field and EGFP field (**I**) identifies cells expressing EGFP (<5%).

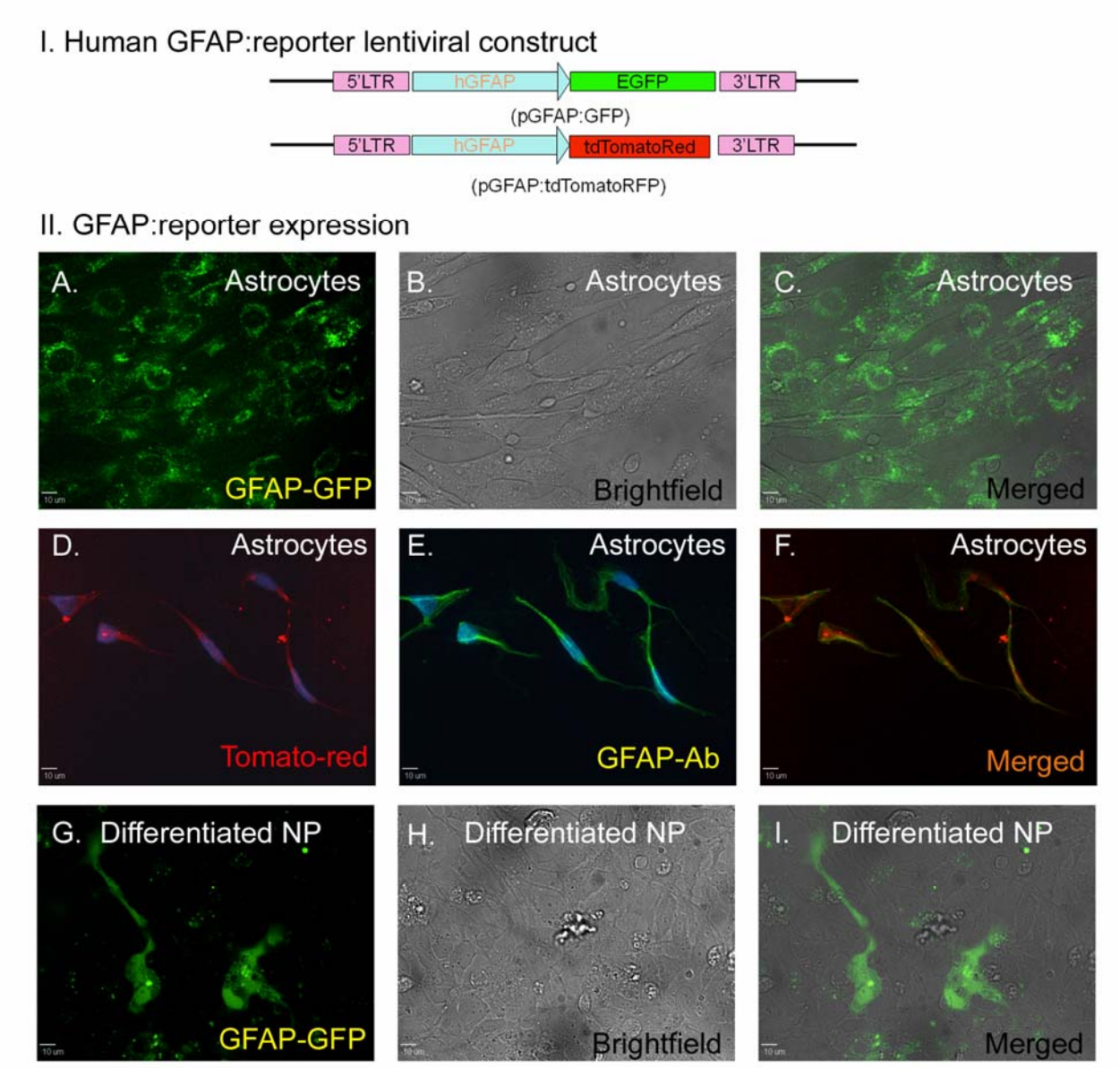
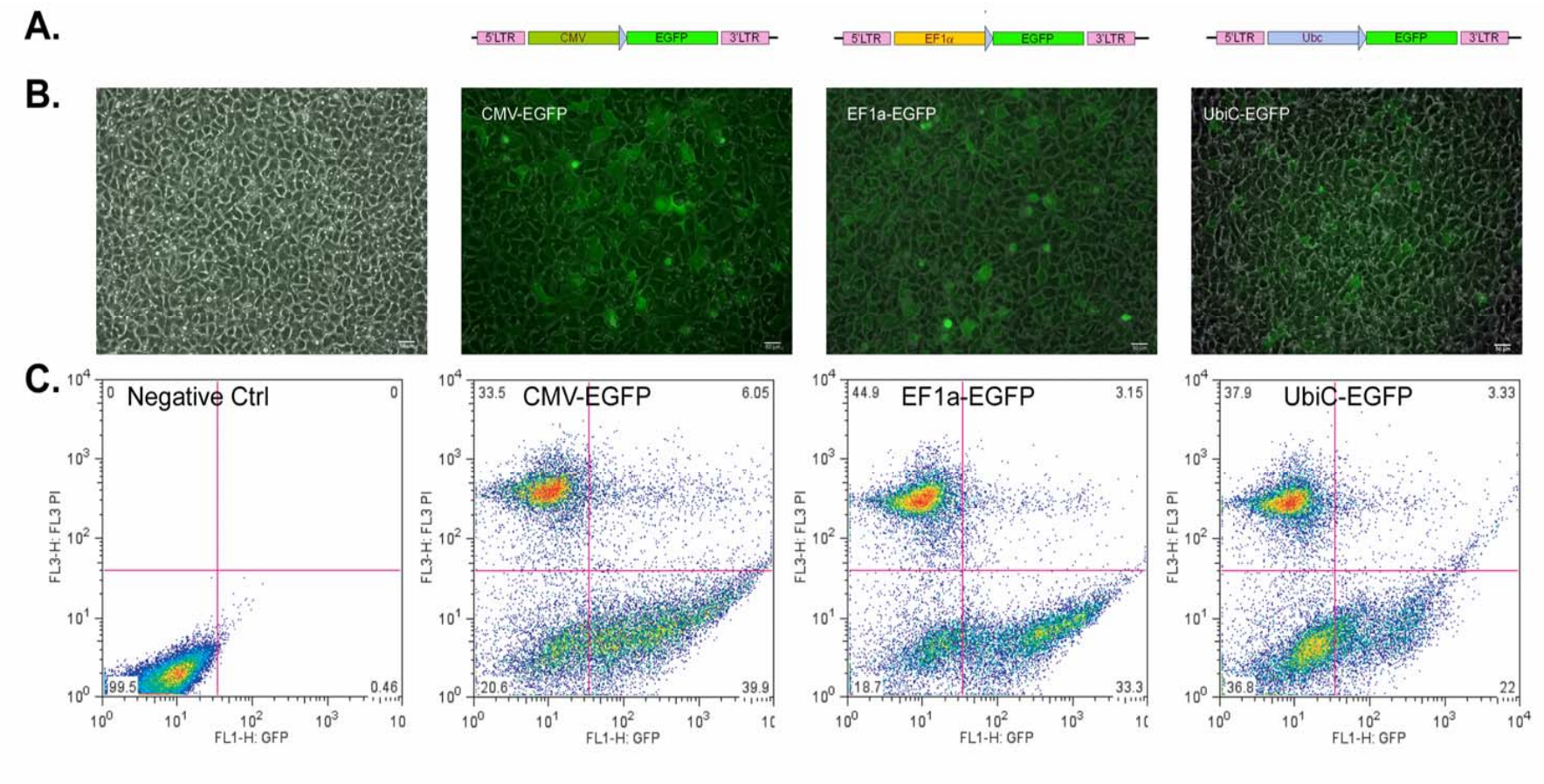
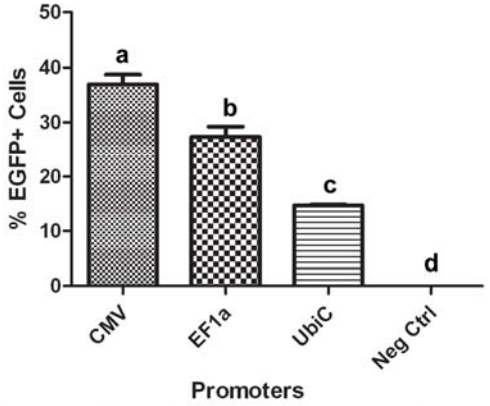

I. Human GFAP:reporter lentiviral construct
5'LTR
hGFAP
EGFP
3'LTR
(pGFAP:GFP)
5'LTR
hGFAP
tdTomatoRed
3'LTR
(pGFAP:tdTomatoRFP)
II. GFAP:reporter expression
A. Astrocytes
GFAP-GFP
B. Astrocytes
Brightfield
C. Astrocytes
Merged
D. Astrocytes
Tomato-red
E. Astrocytes
GFAP-Ab
F. Astrocytes
Merged
G. Differentiated NP
GFAP-GFP
H. Differentiated NP
Brightfield
I. Differentiated NP
Merged

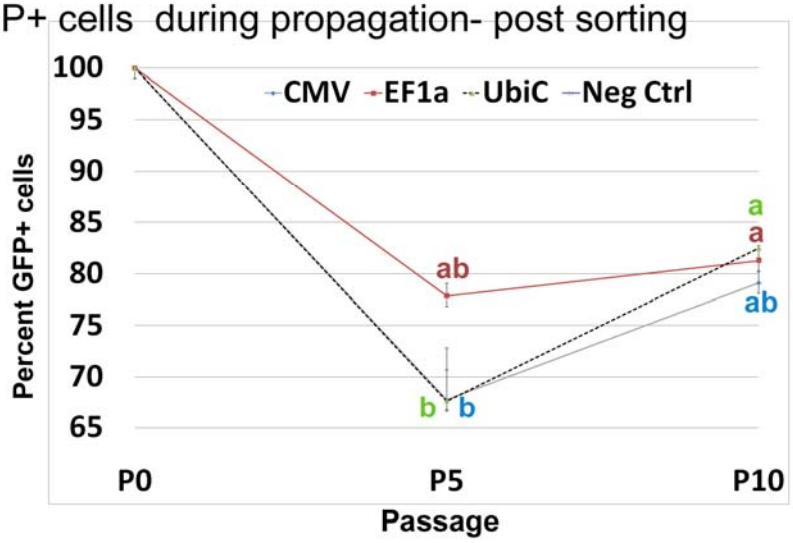
Figure 2.6. Comparison of Transgene Expression and Silencing for Three Different Ubiquitous Promoters. NP cells were transduced using lentiviral delivery system driving EGFP expression under CMV, EF1α and UbC promoters (Schematic diagram of constructs are shown in **A**). (**B**) Merged images of EGFP and bright fields of transduced cells (**C**) Flow-cytometry analysis quantifying percentage of EGFP+ cells before sorting (representative flowcytometry diagram) (**D**) Mean and standard errors of $EGFP^+$ cells from three independent experiments (before sorting). (**E**) After sorting, the transduced cells were propagated and $EGFP^+$ cells were quantified at passage 5 and 10. (**F**) Further, relative fluorescent intensity was measured (as median channel numbers) and compared for all three promoters. The experiment was replicated independently three time (n=3) and all statistics are shown at ($P<0.05$) significance level. Significant differences are shown by different superscripts.


A.
5'LTR
CMV
EGFP
3'LTR
5'LTR
EF1α
EGFP
3'LTR
5'LTR
Ubc
EGFP
3'LTR
B.
CMV-EGFP
EF1a-EGFP
UbiC-EGFP
C.
Negative Ctrl
0
0
99.5
0.46
CMV-EGFP
33.5
6.05
20.6
39.9
EF1a-EGFP
44.9
3.15
18.7
33.3
UbiC-EGFP
37.9
3.33
36.8
22
FL3-H: FL3 PI
FL1-H: GFP

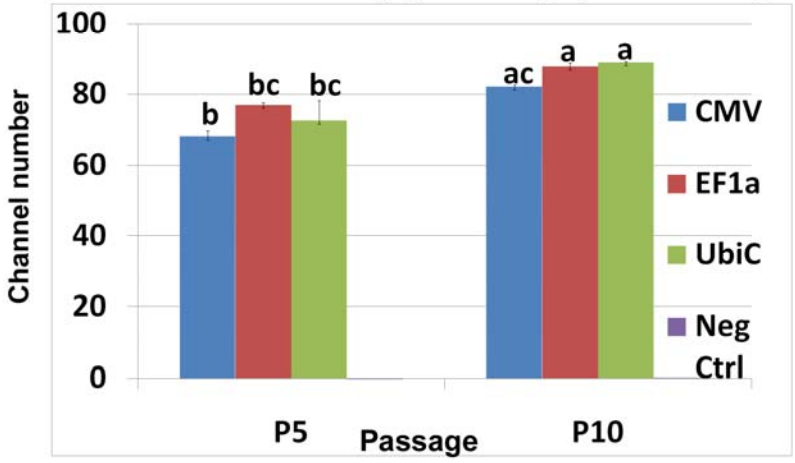
D. Quantification GFP+ cells under three promoters- pre sorting



E. EGFP+ cells during propagation- post sorting



F. Relative fluorescent intensity (Median)- post sorting



CHAPTER 3

TEMPORAL GLIAL FIBRILLARY ACIDIC PROTEIN EXPRESSION DURING THE DIFFERENTIATION OF HUMAN EMBRYONICALLY-DERIVED NEURAL PROGENITOR CELLS[1]

Brian A. Gerwe*, Sujoy K. Dhara*, Tulsi Patel, David W. Machacek and Steven L. Stice.

*Authors contributed equally to this work

To be submitted to Glia

Abstract

The directed differentiation and isolation of human embryonic stem cell (hESC)-derived astrocytes would provide an in vitro cell source for the further study of a wide range of neurological disorders such as Alzheimer's, Alexander's Disease, Amyotrophic Lateral Sclerosis (ALS, Lou Gehrig's Disease), glial tumors and central nervous system inflammation. Astrocytes have a dynamic morphology that can be dependent on their location within the central nervous system (CNS). For this reason, a high expression of intermediate filaments, particularly glial fibrillary acidic protein (GFAP), is used to characterize astrocytes. Astrocyte differentiation is difficult to study because the temporal onset of gliogenesis is preceded by neurogenesis. Furthermore, the in vitro isolation of astrocytes from a human stem cell source still remains elusive. Therefore, in this study, we examine the temporal expression of GFAP to better elucidate the differentiation process of human neural progenitor (hNP) cells towards an astrocyte-like cell population. During the directed differentiation, we followed temporal changes in GFAP expression via a lentiviral transduction reporter scheme and RT-PCR assays. We found that the addition of 5% fetal bovine serum (FBS) to the differentiation medium significantly up-regulated ($P<0.05$) GFAP expression within differentiating hNP cultures. Thus, we were able to conclude that 5% FBS directs the differentiation of hESC derived hNP cells toward the astrocytic lineage, based on the increase in GFAP expression.

Introduction

It is anticipated that neural cell based therapies will offer new avenues of treatment for diverse neural degenerative diseases and further the development of novel diagnostics and therapeutics. It has been well documented that an imbalance in

the CNS increases metabolic activity in astrocytes and causes them to increase production of growth and trophic factors (16). Upon injury, normal astrocytes undergo a cellular transformation and adopt a “reactive astrocyte” phenotype. This transformation increases cellular levels of GFAP that may extend far from the site of injury. Glial Fibrilary Acidic Protien (GFAP) is a glial specific and predominantly astrocytic intermediate filament. The up regulation of GFAP initiates proliferation of more reactive astrocytes that ultimately lead to glial scaring of the injured site. Furthermore, it is postulated that reactive astrocytes establish a barrier between healthy and damaged cells, thus recapitulating the blood-brain barrier (4). The elevated concentration of astrocyte-secreted factors promotes the return to homeostasis through the protection of brain cells from energy depletion, free reactive oxygen species and perhaps calcium saturation.

Astrocytes are restricted to the CNS, and are found in both white and gray matter with variations in morphology that include distinctive “fibrous” cells that have long unbranched cellular processes and “protoplasmic” cells that exhibit short and highly branched cellular processes (12). Because of the non-uniform morphology, astrocytes are now characterized by their high content of GFAP intermediate filaments (12). While this is their hallmark characteristic, it should be noted that GFAP is not unique of astrocytes. Serving as the main intermediate filament, GFAP provides strength and support to the cytoskeleton of radial glia and astrocytes in the CNS, and, to a lesser degree, to mueller cells of the retina and non-myelinating schwann cells of the peripheral nervous system (1, 23, 26).

Astrocytes within the central nervous system (CNS) are produced from the neuroepithelium of the neural tube; yet, the differentiation is not random (Figure 3.1). Neuroepithelial cells must first proliferate to sustain normal development; otherwise, differentiation will produce too few neurons and glia. Once there is a large enough population of neuropethelial cells, neurogenesis is initialed. Then, a switch to glial differentiation occurs through one or more mechanistic pathways that are not well established. One prevailing theory is that Notch signaling is directly involved in the inhibition of neurogenesis early in development, and in the initiation of glial differentiation later in development. Recent evidence indicates that the differentiation process is closely monitored through lateral inhibition mediated by Notch signaling (2, 13). Because the temporal onset of neurogenesis precedes gliogenesis, it is difficult to study gliogenesis in culture.

To enhance our ability to monitor the onset of glial differentiation with hNP cells, we modified the lentiviral packaging vector pFUGW (15) to include a glial lineage reporting system. The new construct, which is named pFUGW GFAP:EGFP :: SV40:NeoR, comprises of a conditional reporter (hGFAP:EGFP) with constitutive neomycin resistance for selection. The human GFAP promoter region is a 2.2kB region found upstream of *Gfa (GFAP gene)*. Within this region is a basal promoter consisting of a TATA box and a CAAT box and several enhancer sequences (5). Initial characterization of hGFAP was performed in mice with the bacterial *lacZ* reporter system (8). We chose lentiviral intergration (derived from HIV-1 backbone) over other genetic manipulation approaches, such as adenovirus vectors and murine leukemia virus (MLV)-based retrovirus, due to several key advantages including stable DNA

intregration (19), infection of non-dividing cells with a broad range of hosts including neurons and glial cells in tissue culture and *in vivo* (6, 18, 20), and self-inactivating vectors (SIN) having the long terminal repeats (LTRs) deleted in the U3 region. The U3 region is essential for the replication of the wild-type retrovirus, but dispensable for the replication-defective vector. Therefore, self-inactivating vectors are safer than their counterparts with full-length LTRs (27).

We establish a quantitative real time-PCR expression profile for hNP cells grown for seven weeks in differentiation media. The expression profile illustrates no change (<1 relative fold, $P<0.05$) in GFAP expression over the seven week time period. Upon the addition of 5% fetal bovine serum (FBS) to this medium, significant up-regulation ($P<0.05$) of GFAP, relative to treatments without FBS, is observed. Modifications to the differentiation media did not shorten the temporal onset of GFAP expression, although relative expression levels were altered. In addition, we transduced hNP cells to allow for the selection of GFAP expressing cells through a GFP reporter. The results demonstrate the clear ability to track astrocyte differentiation in vitro, which will allow further isolation of pure hESC-derived astrocyte populations and can further serve to define the mechanisms of astrocyte specification and subsequent development within the CNS. A study of astrocyte differentiation will lead to insights into the complexities of the blood brain-barrier and provide a human tissue source for studying several debilitating CNS pathologies.

Materials and Methods

Human Neural Progenitor (hNP) Culture Conditions

hNP were derived from WA09 human embryonic-stem cells with normal karyotype as described in past reports by our laboratory (17). Cells were propagated and maintained in proliferation media consisting of neurobasal medium (Gibco) supplemented with 2mM L-glutamine, 1x penicillin/ streptomycin, 1x B27 (Gibco), 20ng/ml of bFGF (Sigma-Aldrich) and 10ng/ml of Leukemia inhibitory factor (LIF, Chemicon) under conditions of 5% CO_2 atmosphere at 37°C. Cells were passaged 1:3 or 1:4 by mechanical dissociation via trituration and re-plated on polyornithine (20 µg/ml) and laminin (5 µg/ml) coated plates with media changes every other day. hNP cultures in decreased levels of the B27 with or without 1 X N2 supplementation were continuously maintained and passaged as described above.

Neural Progenitor Differentiation Towards GFAP$^+$ Lineage

hNP cells were differentiated on polyornithine (20 µg/ml) and laminin (5 µg/ml) coated plates in differentiation media (bFGF withdrawn from proliferation media) with or without 5 or 10% fetal bovine serum (FBS, Hyclone) and with or without astrocyte conditioned media (ACM) for seven weeks (DAY0 – DAY49) under conditions of 5% CO_2 atmosphere at 37°C. ACM was added to –bFGF differentiation media (1:1 ratio) to prevent the depletion of growth factors during differentiation. Media was replaced every third day.

Human Astrocytoma Cell Culture

Human astrocytoma cells were purchased from ATCC (CRL-1718) and maintained in tissue culture according to supplier's instruction. Cells were sub-cultured

on tissue culture grade dishes or tissue culture flasks in RPMI-1640 medium (ATCC # 30-2001) supplemented with 10% fetal bovine serum (FBS, Hyclone). Production of astrocytoma-conditioned medium (ACM) was achieved by adding –bFGF differentiation media supplemented with 5 or 10% FBS on human astrocytoma cells for a period of 24 hours, after which, it was harvested and used for the differentiation of hNP cells toward a $GFAP^+$ cell fate.

pFUGW GFAP:EGFP :: SV40:NeoR Packaging Vector Construction

pFUGW lentiviral packaging vector was a gift from Dr. James Lah at Emory University. The pGfa2-cLac plasmid (8) carrying the hGFAP promoter was a gift from Dr. Michael Brenner at University of Alabama. The hGFAP promoter was excised from pGfa2-cLac plasmid using *Bgl* II and *Sal* I and inserted into *Bgl* II and *Sal* I digested shuttle vector DsRed Express-1 (Clontech). The shuttle vector was digested with *Afe* I and *Bam* HI and the promoter was excised and inserted into like digested pFUGW, yielding pFUGW GFAP:EGFP. Neomycin resistance was conferred through digestion of DsRed Express-1 (Clontech) with *Dra* I and *Dra* III and excision of the neomycin-kanamycin resistance gene cassette including the SV40 promoter regions. The Dra III end of the fragment was blunted (NEB Quick Blunting Kit, cat. E1201S) and ligated into *Afe* I digested pFUGW GFAP:EGFP creating pFUGW GFAP:EGFP :: SV40:NeoR.

Lentiviral Transduction of hNP Cells

Lentiviral particles were produced and titered in the laboratory of Dr. James Lah (Emory University). Vector particles were produced in HEK293T cells by transient co-transfection with the pFUGW transfer vector, the HIV-1 packaging vector ΔR8.9, and the VSVG envelope glycoprotein. Media was replaced the morning post-transduction and

then placed back in the incubator. 28 hours later, the virus-containing supernatant was removed, filtered through a 0.45μM filter unit, concentrated by centrifugation, aliquoted and frozen at -80^{o}C. Viral titers were calculated using a serial dilution method.

Two days prior to transduction, a confluent 35 mm dish of hNP cells was prepared and maintained as described above. The following day, the dish was split 1:2 with fresh proliferation media added and incubated overnight under conditions of 5% CO_2 atmosphere at 37°C. With hNP cells at ~60% confluence, viral particle aliquots were combined with 2 ml of warmed proliferation media to a multiplicity of infection (MOI) of 5 and added to 1 x 35 mm plate of hNP cells as if normal media change. Media was also changed in the remaining plate and used as the control. Both plates were centrifuged (250g for 5 minutes at 23°C) and incubated overnight (37°C with a 5% CO_2 atmosphere). Twenty-four hour post-transduction, the media was aspirated off and replaced with 2 ml of fresh media. G418 selection (200 ug/ml) for transduced cells was introduced five days post-transduction. Media was changed every other day and cells were passaged as needed.

Quantitative Real time-polymerase Chain Reaction (RT-PCR)

mRNA was extracted using the Qiashredder and RNeasy Mini kits (Qiagen) as per manufacturer's protocol (version October 2005). RNA quality and quantity was verified using a RNA 600 Nano Assay and the Agilent 2100 Bioanalyzer (Agilent Technologies). Total RNA (500 ng) was reverse-transcribed using the cDNA Archive Kit (Applied Biosystems Inc.) as per manufacturer's protocols. One micro liter of the cDNA sample (diluted 1:7.5) was added to 1X primer/ probe solution (stock 20X, Applied BioSystems) along with 2X Taqman PCR Master Mix into 384 well plate. Sterile water

was added (v_f = 10µl). Real-time PCR and relative quantification was carried out on the ABI PRISM 7900 Sequence Detection System (Applied Biosystems Inc.) under standard cycle conditions. Experiments were replicated in quadruplicate, unless otherwise stated. All failed (undetermined) reactions were excluded when ΔCt values were calculated. For calculation of relative fold change values, samples were normalized against their respective endogenous 18S ribosomal RNA utilizing the ΔΔCT method of quantification (Applied Biosystems Inc.) (14). Then, the average fold changes from four independent runs were calculated as $2^{-\Delta\Delta CT}$. Statistical significance was determined via 2-way ANOVA and Tukey's Pair-Wise (SAS) comparisons focusing on temporal GFAP expression.

Qualitative Reverse Transcription-polymerase Chain Reaction (RT-PCR)

mRNA was extracted using the Qiashredder and RNeasy Mini kits (Qiagen) as per manufacturer's protocol (version October 2005). RNA quality and quantity was verified using a RNA 600 Nano Assay and the Agilent 2100 Bioanalyzer (Agilent Technologies). For cDNA synthesis, total RNA (500 ng) was reverse-transcribed using the cDNA Archive Kit (Applied Biosystems Inc.) as per manufacturer's protocols. Primer sequences used for PCR of Nestin, forward 5'-CAGGAGAAACAGGGCCTACA-3' and reverse 5'- TAAGAAAGGCTGGCACAGGT-3' (313bp, GeneID: 10763); PAX6, forward 5'- CCGGCAGAAGATTGTAGAGC-3', reverse 5'- CTAGCCAGGTTGCGAAGAAC-3' (363bp, GeneID: 5080); CD44, forward 5'-CCCAGATGGAGAAAGCTCTG-3', reverse 5'-CTGTCTGTGCTGTCGGTGAT-3' (463bp, GeneID: 960); VIM, forward 5'-CCTTGAACGCAAAGTGGAAT-3', reverse 5'-GCTTCAACGGCAAAGTTCTC-3' (408bp, GeneID: 7431); S100β, forward 5'-GCCCTCATCGACGTTTTCCA-3', reverse 5'-

AAGAGTCCCTGGGGCCAGTC-3' (491bp, GeneID: 6285); CNTFR, forward 5'-ACCATTGTGAAGCCTGATCC-3', reverse 5'-GGAGGAGAAATCGGATGTGA-3' (864bp, GeneID: 1271); IL6R, forward 5'-CTCCTGCCAGTTAGCAGTCC-3', reverse 5'-TGTCGCATTTGCAGAATCTC-3' (535bp, GeneID: 3570); SC1, forward 5'-TGTGACGTCTGGTTCCATGT-3', reverse 5'-GTGATGCCACCAAGAACCTT-3' (482bp, GeneID: 6941); BLBP, forward 5'-AAGGATGGTGGAGGCTTTCT-3', reverse 5'-ACAGCAACCACATCACCAAA-3' (381bp, GeneID: 2173); GLAST (SLC1A3), forward 5'-TGCTGGGGAATTCACCTCGT-3', reverse 5'-CGCATTCCCATCTTCCCTGA-3' (520bp, GeneID: 6507); GLUL, forward 5'-CCCTGCCTCAGGGTGAGAAA-3', reverse 5'-TGGCGCTACGATTGGCTACA-3' (909bp, GeneID: 2752); AQP4, forward 5'-AGCCTGGGATCCACCATCAA-3', reverse 5'-CTCCCCGGTCAACGTCAATC-3' (769bp, GeneID: 361); GAPDH, forward 5'-GAGTCAACGGATTTGGTCGT-3', reverse 5'-TTGATTTTGGAGGGATCTCG-3' (238bp, GeneID: 2597). The Qualitative RT-PCR conditions were initiated with denaturing at 95° C for 4 minutes followed by 30 cycles at 94° C for 1 minute, 60° C for 1 minute, 72° C for 1 minute with a final extension at 72° C for 10 minutes. PCR Products were separated on 2% agarose gels containing ethidium bromide for 60 minutes at 100 volts, visualized by UV light, and analyzed using AlphaEaseFC software on FlourChem 8000 (Alpha Innotech Corporation). Total RNA not subjected to reverse transcription (RT) was accessed for possible DNA contamination for each sample.

Immunocytochemistry

Cells were plated at DAY0 onto permanox (NUNC) 4-chamber well slides and differentiated as described above. At appropriate time point, the slide was fixed with 4%

paraformaldehyde for 15 minutes. Antibodies were directed against GFAP (Sigma-Aldrich, 1:500), Nestin (Neuromics, 1:100) and beta III tubulin (Neuromics, 1:500). Primary antibodies were detected using secondary antibodies conjugated to Alexa Flour 488 (Molecular Probes, 1:1000), Texas Red (Jackson ImmunoResearch Laboratories, Inc., 1:250), and Cy-5 (Jackson ImmunoResearch Laboratories, Inc., 1:250). All wells were counter-stained with DAPI. Picture acquisition was performed using an Olympus Ix81 with Disc-Spinning Unit and Slide Book Software (Intelligent Imaging Innovations) and analyzed by way of Image-Pro Plus (Media Cybernetics).

Results

Temporal GFAP Expression during Directed Differentiation of a hESC-derived hNP Cells under Media Conditions Supplemented with 5% FBS

We used quantitative real time (RT)-PCR analysis to identify the temporal expression pattern for Nestin, beta III Tubulin, and GFAP in vitro with or without 5% FBS supplementation to the –bFGF differentiation media. Nestin is an intermediate filament commonly used to denote neural stem cell populations (22), while beta III tubulin and GFAP are intermediate filament markers expressed in differentiated neurons (21) and glial or astroctyic cell types (7, 10), respectively. hNP cells were differentiated in –bFGF media in the presence of absence of 5% FBS over a 49 day (seven-week) period. Media was replenished every third day to promote cell-cell signaling during differentiation. Nestin expression (Figure 3.2.A) did not change relative to the DAY0 control under either condition, leading us to suspect that a population within the culture remains proliferative even after extended periods of differentiation. This can be further

deduced by the relatively low increase (<3.3 fold change) of beta III Tubulin expression (Figure 3.2.A).

The expression profile (Figure 3.2.B) illustrates no change (<1 relative fold, $P<0.05$) in GFAP expression over the seven-week time period using differentiation media alone. However, upon the addition of 5% FBS to this media, significant up-regulation ($P<0.05$) of GFAP was observed, relative to treatments without FBS. After the onset of GFAP expression, we noted a significant increasing trend in expression between DAY21 (22.3 +/- 11.9 fold change) and DAY49 (>662.6 +/- 100.1 fold). Based on 2-way ANOVA and Tukey's Pair-Wise (SAS) comparisons, we detected a day effect between DAY42 and DAY49 upon the addition of FBS to the differentiation media. GFAP mRNA transcripts were undetectable for DAY14 under both media conditions and DAY21–DAY35 under –bFGF differentiation conditions. Nonetheless, these results demonstrate a clear treatment effect when differentiation medium is supplemented with FBS.

Evaluation of the quantitative RT-PCR expression profile was further confirmed by immunocytochemistry (Figure 3.2.C) demonstrating enhanced GFAP expression in a temporal manner in the presence of 5% FBS. Thus, quantitative RT-PCR and immunocytochemistry showed enhanced glial-like marker expression under -bFGF media supplemented with 5% FBS conditions over a seven-week interval.

Up Regulation of CD44, AQP4 and IL6R when 5% FBS is Supplemented to Differentiation Media

Further characterization of the $GFAP^+$ enriched population led to qualitative reverse transcription-PCR of glial progenitor and glial associated genes (Figure 3.3). Of

the three glial-progenitor genes examined, CD44 showed differential expression between the two differentiation media conditions. CD44 was expressed at very low levels at DAY0 with continued low-level expression until DAY35 in plain differentiation media. On the other hand, the addition of 5% FBS caused up regulation of CD44 at DAY14 and increased expression from DAY21 to DAY49. Consistent with above results, quantitative RT-PCR illustrated that the presence of FBS maintained expression of CD44 at DAY14 (-0.83 +/- 2.28 fold change) rather than its down regulation (-14.35 +/- 1.28) relative to DAY0. Furthermore, supplemented FBS demonstrated substantial up regulation (235.88 +/- 76.81 fold change) at DAY49 relative to the control as compared to the plain differentiation media (12.8 +/- 2) of the two remaining glial-progenitor genes. There was no difference in Nestin and Pax6 expression regardless of condition or time interval demonstrating the continual proliferative nature of the culture.

Additional reverse transcription-PCR analysis of glial-associated genes revealed up regulation of IL6R (interleukin-6 receptor) and AQP4 (Aquaporin 4) genes when 5% FBS was supplemented to differentiation media. While IL6R was not expressed in undifferentiated cells (DAY0) or in cells growing in plain differentiation media over a 7-week period, up regulation occurred by DAY28 in FBS supplemented media and continued throughout the remaining period of differentiation. A similar pattern was observed for AQP4 with only very low expression at DAY49 in the absence of FBS, but a continual increase of expression in the presence of 5% FBS over the 7-week period. Quantitative real time-PCR performed on AQP4 showed a significant ($P<0.05$) up regulation (5.23 +/- 1.8) at DAY14 of differentiation with further up regulation at DAY49 in differentiation media with FBS; in addition, a treatment effect was observed at DAY14

when FBS was supplemented to the differentiation media. Statistical significance was not reported for DAY49 due to variation in microenvironments during differentiation. The remaining glial genes tested did not show a differential expression pattern under either media condition. It should be noted that S100β and CNTFR were not expressed until after DAY21 in –bFGF differentiation media, while expression was observed by DAY14 when 5% FBS was supplemented to the medium. Thus, a differential up regulation of CD44, AQP4, and IL6R in –bFGF media supplemented with 5% FBS was observed by qualitative reverse transcription-PCR with the expression pattern of CD44 and AQP4 further validated by quantitative real time (RT)-PCR. It should be noted that the DAY14 analysis of –bFGF differentiation media alone was performed in triplicate versus quadruplicate due to degraded RNA of one sample.

Lentiviral Transduction of the Glial Promoter GFAP Enables the Selective Monitoring for GFAP$^+$ Enriched cells In Vitro

We have previously demonstrated the selective expression of GFAP:GFP within transduced human astrocytoma (ATCC CRL-1718) cells, while absent in transduced undifferentiated hNP cells (Chapter 2 of dissertation; to be submitted, Dhara et al., 2008). hNP cells were transduced with a GFAP:GFP reporter construct capable of G418 selection (Neomycin resistance). Three days post transduction, hNP cells were screened for positive transformants using G418 selection (200ug/ml) (Figure 3.4.A). Selection, maintenance, and low density passaging (1:5) were continued for 30 days to ensure purity of transduced cell population. hNP cells were differentiated in –bFGF differentiation media in the presence of 5% or 10% FBS with or without astrocyte

conditioned media (ACM) over a 49 day interval. Media was changed every third day to promote cell-cell signaling during differentiation.

Immunocytochemistry confirmed the selective ability of the GFAP:GFP construct to only GFAP expressing cells (Figure 3.4.B). A representative image from DAY35 with 10% FBS and ACM illustrates the selective nature of the construct via the absence of GFP expression in beta III Tubulin (TuJ-1) expressing cells. No GFAP or GFP expression was observed at DAY0. Even though cell morphology is no longer a valid method to distinguish astrocytes in culture, it was interesting to note the elongated and flatten cell morphology common to unbranched astrocytes found within the CNS (Figure 3.4C).

Quantitative real time-PCR of the transduced cells differentiated over 35 days in the presence of 5% or 10% FBS with or without astrocyte conditioned media (ACM) was performed to investigate whether GFAP expression could be manipulated via decreased time for onset of expression or increased expression. The presence of 10% FBS supplemented to the –bFGF differentiation media demonstrated a significant ($p<0.05$) difference at DAY21 compared to 5% FBS –bFGF differentiation cultures, regardless of the presence of ACM (Figure 3.4.D). Furthermore, GFAP expression was significantly ($p<0.05$) down regulated in 5% FBS cultures at DAY21. The results suggest that 10% FBS may decrease the time interval needed for the onset of GFAP expression. Following the onset of GFAP expression, significant ($p<0.05$) up regulation of GFAP was observed in the 5% FBS without ACM (4 +/- 0.95 relative fold change) and in the 10% FBS + ACM culture (12.6 +/- 1.1 fold) at DAY28. By DAY35, the 5% and 10% FBS without ACM cultures demonstrated a significant increase in GFAP expression relative

to DAY0 with 55.6 +/- 10.2 and 48.9 +/- 13 fold change, respectively. Interestingly, the addition of ACM to the differentiated media increased the average relative GFAP expression (60.5 +/- 38.1 and 99.1 +/- 49.5 fold change for 5% FBS + ACM and 10% FBS + ACM, respectively) above that of FBS alone, regardless of concentration; yet, significance was not achieved because of the variability of the time points. This intriguing result was first observed with significance ($p<0.05$) at DAY28 between 10% FBS with (1.5 +/- 0.15) or without ACM (12.6 +/- 1.1) differentiated cultures. It should be noted that DAY21 supplemented with 5% FBS + ACM was assessed based three replicates, not four as all other conditions and time points because GFAP mRNA fell below detectable limits for one replicate.

Partial Withdrawal of B27 from hNP Proliferation and –bFGF Differentiation Media Increased GFAP Expression

Despite the successful onset of GFAP expression from hNP cells, the culture system is still ill-defined and densely populated with neuronal lineage-fate restricted cells. This suggested that hNP cells had a propensity to differentiate to the neuronal lineage because of possible media components used during hNP maintenance, proliferation and standard differentiation. One component in our culture and differentiation media is B27 (media supplement) that has been routinely added to neuronal differentiation or maintenance media formulations (9). Wu and colleagues (2002) primed fetal neural stem cells on laminin coated plates with bFGF and heparin, followed by differentiation with the B27 supplement and were able to induce multiple subtypes of neurons, including cholinergic neurons, for further study and transplantation

(24). Therefore, we hypothesized that the withdrawal of B27 may decrease neuronal differentiation and enhance glial differentiation.

Upon complete withdrawal of B27, hNP cells did not survive (data not published). Therefore, B27 was incrementally decreased from proliferating hNP populations with or without 1 X N2 media supplement (a component of B27). Cells were maintained and passaged with varying concentrations of B27 (0 X, 0.125 X, 0.25 X, 0.5X and 1 X [control]) in the presence or absence of the 1 X N2 media supplement. Cells were successfully maintained and passaged for two months in 0.125 X, 0.25 X, 0.5X, 1 X B27, and 0.5 X B27 + 1 X N2 culturing conditions in order to properly acclimate the cells to the new media composition and verify that cultures did not senesce or differentiate. These conditions were analyzed for A2B5 (glial precursor marker), GFAP (mature astrocyte marker) and Nestin (neural progenitor marker) via immunostaining. Results for 0.125 X B27 and 0.5 X B27 + 1 X N2 are reported here (Figure 3.5.A). All cultures retained Nestin expression with low a percentage (visually observed) of the culture becoming immuno-reactive to A2B5 indicating cells may be in transition of adopting a glial cell fate. After two months of continued maintenance, the cells were differentiated in –bFGF differentiation media supplemented with 5% FBS.

Quantitative real (RT)-PCR was performed on 0.125 X B27 and 0.5 X B27 + 1 X N2 (Figure 3.5.B). These results were compared against –bFGF supplemented with 5% FBS results previously used in Figure 3.3. Differentiation experiments illustrated in Figure 3.3 and 3.4 were all performed in parallel, thus allowing direct comparison between time points and conditions. All DAY21 treatments were significant ($p<0.05$) when compared to DAY0 undifferentiated hNP cells with 3.6 +/- 1 relative fold change

for 0.5 X B27 + 1 X N2 and 6.3 +/- 0.4 for 0.125 X B27 treatments. Not only were all treatments at DAY28 significant from DAY0 of undifferentiated hNP cells, but all treatments demonstrated a day effect when compared to DAY21. There was a ~10.5 fold difference between DAY28 hNP cells differentiated in –bFGF media with 5% FBS (4.35 +/- 1 relative fold change) and the cells proliferated in 0.125 X B27 and then differentiated with in –bFGF media with 5% FBS (14.84 +/- 2).

Discussion

We demonstrate that the temporal expression of GFAP can be utilized as a developmental biomarker to elucidate the differentiation process of human neural progenitor (hNP) cells towards an astrocyte-like cell population. Quantitative RT-PCR determined that culturing hNP cells in –bFGF differentiation media supplemented with FBS increased the expression of GFAP by DAY21 (22.3 +/- 11.9 fold change) and substantial up regulation continued to DAY49 (663 +/- 100 fold change) relative to the undifferentiated DAY0 control. In the absence of FBS, an increase in GFAP expression was not observed until DAY49 (1.60 +/- 0.44 relative fold change).

Increased FBS concentration (from 5% to 10%) showed a significant ($p<0.05$) difference at DAY21 compared to 5% FBS –bFGF differentiation cultures, which were down regulated regardless of the presence of ACM. Further analysis is needed to determine if 10% FBS alone could shorten the length of time needed for the onset of GFAP expression under current culture system. DAY35 treatments in 5% and 10% FBS showed significant increases in GFAP expression relative to DAY0. Intriguingly, the addition of ACM to the –bFGF differentiated media increased the average relative

GFAP expression at DAY35 above that of FBS alone, however, a larger viability was demonstrated within the time point.

Enhancement of glial differentiation through the withdrawal of B27 media supplement from the differentiation medium was investigated. B27 has been well established as a neuronal differentiation supplement (9). Our results suggest that the partial removal of B27 from standard culturing conditions may affect downstream differentiation efforts. Intriguingly, the complete removal of B27 is detrimental to cell proliferation and 1 X N2 replacement is not sufficient to promote cell survival. We acclimatized hNP cells to modified concentrations of B27 with standard maintenance and proliferation media conditions, whereby a low percentage of cells were A2B5 immuno-reactive. Attempts to identify $A2B5^{+}$ cells under standard proliferation conditions failed, alluding to a possible transition of cell-fate to a more glial-restricted lineage over a period of two months in decreased levels of the media supplement B27. Yet, Nestin expression was not altered within the cultures, which verifies retained neural progenitor status for the vast majority of hNP cells. The hNP cells were differentiated in –bFGF differentiation media supplemented with 5% FBS after two months of continued culture with varied B27 concentrations.

Quantitative real (RT)-PCR was performed on 0.125 X B27 and 0.5 X B27 + 1 X N2 cells differentiated in –bFGF supplemented with 5% FBS. All DAY28 treatments were significant ($p<0.05$) relative to DAY0 (undifferentiated hNP cells). In addition, the treatments demonstrated a day effect with a significant increase in GFAP expression from DAY21 to DAY28.

After the establishment of $GFAP^+$ cells upon differentiation, we transduced hNP cells with the lentiviral GFAP:GFP :: SV40:NeoR construct and demonstrated the selective expression of GFP to only GFAP expressing cells. The successful conditional expression of the GFAP:GFP lentiviral construct allowed for the further characterization of a $GFAP^+$ cell type.

Additional studies are needed to further characterize the temporal expression of GFAP, particularly its onset, during the directed differentiation of hNP cells. As noted above, the presence of FBS within the differentiation media up regulates GFAP expression of hNP cells towards a glial-like cell fate. However, the signaling molecules present in the FBS that stimulate or permit this differentiation are still unknown. Fetal bovine serum is a vast cocktail of macromolecules, carrier proteins, carbohydrates, amino acids, supplements and trace elements, hormones and growth factors (11, 25) with only several of the major known biological components important for cell growth are albumin, antichymotrpsin, apolipoproteins, biotin, and growth supporting factors (3). In 2006, a proteomic study identified a list of components found in fetal bovine serum (25). Interestingly, the researchers identified a varying number of proteins and at different total protein concentrations between the three sera examined. Two of the three sera were obtained from the same company, yet different lots. Thus, fetal bovine serum remains ill-defined with varying biological components that will ultimately harbor efforts to parse out cell-cell interactions and signaling molecules required for the onset of mammalian gliogenesis.

In summary, characterization of an astrocyte-like cell type from hESC-derived neural progenitor cells has been achieved and will, hopefully, lead to a more complete

in vitro model for astrocyte specification (e.g gliogenesis). The future isolation of a homogenous population may provide a human tissue source for studying neural stem cell-based therapies and debilitating central nervous system pathologies.

References

1. **Alvarez-Buylla, A., D. R. Buskirk, and F. Nottebohm.** 1987. Monoclonal antibody reveals radial glia in adult avian brain. J Comp Neurol **264:**159-70.
2. **Artavanis-Tsakonas, S., M. D. Rand, and R. J. Lake.** 1999. Notch signaling: Cell fate control and signal integration in development. Science **284:**770-776.
3. **Baker, H., B. DeAngelis, and O. Frank.** 1988. Vitamins and other metabolites in various sera commonly used for cell culturing. Experientia **44:**1007-10.
4. **Ballabh, P., A. Braun, and M. Nedergaard.** 2004. The blood-brain barrier: an overview: structure, regulation, and clinical implications. Neurobiol Dis **16:**1-13.
5. **Besnard, F., M. Brenner, Y. Nakatani, R. Chao, H. J. Purohit, and E. Freese.** 1991. Multiple Interacting Sites Regulate Astrocyte-Specific Transcription of the Human Gene for Glial Fibrillary Acidic Protein. Journal of Biological Chemistry **266:**18877-18883.
6. **Blomer, U., L. Naldini, T. Kafri, D. Trono, I. M. Verma, and F. H. Gage.** 1997. Highly efficient and sustained gene transfer in adult neurons with a lentivirus vector. Journal of Virology **71:**6641-6649.
7. **Bovolenta, P., R. K. Liem, and C. A. Mason.** 1984. Development of cerebellar astroglia: transitions in form and cytoskeletal content. Dev Biol **102:**248-59.
8. **Brenner, M., W. C. Kisseberth, Y. Su, F. Besnard, and A. Messing.** 1994. Gfap Promoter Directs Astrocyte-Specific Expression in Transgenic Mice. Journal of Neuroscience **14:**1030-1037.

9. **Chen, Y., B. Stevens, J. Chang, J. Milbrandt, B. A. Barres, and J. W. Hell.** 2008. NS21: re-defined and modified supplement B27 for neuronal cultures. J Neurosci Methods **171:**239-47.

10. **Dahl, D.** 1981. The vimentin-GFA protein transition in rat neuroglia cytoskeleton occurs at the time of myelination. J Neurosci Res **6:**741-8.

11. **Gstraunthaler, G.** 2003. Alternatives to the use of fetal bovine serum: serum-free cell culture. ALTEX **20:**275-81.

12. **Kessaris, N., N. Pringle, and W. D. Richardson.** 2008. Specification of CNS glia from neural stem cells in the embryonic neuroepithelium. Philos Trans R Soc Lond B Biol Sci **363:**71-85.

13. **Lewis, J.** 1998. Notch signalling and the control of cell fate choices in vertebrates. Seminars in Cell and Developmental Biology **9:**583-589.

14. **Livak, K. J., and T. D. Schmittgen.** 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods **25:**402-8.

15. **Lois, C., E. J. Hong, S. Pease, E. J. Brown, and D. Baltimore.** 2002. Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. Science **295:**868-72.

16. **Min, K.-J., M.-s. Yang, S.-U. Kim, I. Jou, and E.-h. Joe.** 2006. Astrocytes Induce Hemeoxygenase-1 Expression in Microglia: A Feasible Mechanism for Preventing Excessive Brain Inflammation. J. Neurosci. **26:**1880-1887.

17. **Mitalipova, M., J. Calhoun, S. J. Shin, D. Wininger, T. Schulz, S. Noggle, A. Venable, I. Lyons, A. Robins, and S. Stice.** 2003. Human embryonic stem cell lines derived from discarded embryos. Stem Cells **21:**521-526.

18. **Miyoshi, H., M. Takahashi, F. H. Gage, and I. M. Verma.** 1997. Stable and efficient gene transfer into the retina using an HIV-based lentiviral vector. Proceedings of the National Academy of Sciences of the United States of America **94:**10319-10323.

19. **Naldini, L., U. Blomer, F. H. Gage, D. Trono, and I. M. Verma.** 1996. Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector. Proc Natl Acad Sci U S A **93:**11382-8.

20. **Naldini, L., U. Blomer, P. Gallay, D. Ory, R. Mulligan, F. H. Gage, I. M. Verma, and D. Trono.** 1996. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science **272:**263-267.

21. **Shin, S., S. Dalton, and S. L. Stice.** 2005. Human motor neuron differentiation from human embryonic stem cells. Stem Cells and Development **14:**266-269.

22. **Shin, S., M. Mitalipova, S. Noggle, D. Tibbitts, A. Venable, R. Rao, and S. L. Stice.** 2006. Long-Term Proliferation of Human Embryonic Stem Cell-Derived Neuroepithelial Cells Using Defined Adherent Culture Conditions. Stem Cells **24:**125-138.

23. **Verderber, L., W. Johnson, L. Mucke, and V. Sarthy.** 1995. Differential regulation of a glial fibrillary acidic protein-LacZ transgene in retinal astrocytes and Muller cells. Invest Ophthalmol Vis Sci **36:**1137-43.

24. **Wu, P., Y. I. Tarasenko, Y. Gu, L. Y. Huang, R. E. Coggeshall, and Y. Yu.** 2002. Region-specific generation of cholinergic neurons from fetal human neural stem cells grafted in adult rat. Nat Neurosci **5:**1271-8.

25. **Zheng, X., H. Baker, W. S. Hancock, F. Fawaz, M. McCaman, and E. Pungor, Jr.** 2006. Proteomic analysis for the assessment of different lots of fetal bovine serum as a raw material for cell culture. Part IV. Application of proteomics to the manufacture of biological drugs. Biotechnol Prog **22:**1294-300.

26. **Zhuo, L., B. Sun, C.-L. Zhang, A. Fine, S.-Y. Chiu, and A. Messing.** 1997. Live Astrocytes Visualized by Green Fluorescent Protein in Transgenic Mice. Developmental Biology **187:**36-42.

27. **Zufferey, R., T. Dull, R. J. Mandel, A. Bukovsky, D. Quiroz, L. Naldini, and D. Trono.** 1998. Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. Journal of Virology **72:**9873-9880.

Figure 3.1. Order of in vivo mammalian differentiation of neuroepithelial (NE) cells. (A) The NE cells undergo symmetric division establishing the neuroepithelial pool; then asymmetric dividing to form the radial glial pool. Upon completion of both pools, the NE cells differentiate into the terminal cell types of the CNS. The temporal onset of gliogenesis precedes neurogenesis. For comparison, temporal differentiation within the mouse is illustrated in **(B)**.

Adapted from: Gotz, M., and Y. A. Barde. 2005. Radial glial cells: Defined and major intermediates between embryonic, stem cells and CNS neurons. Neuron 46:369-372; Glibert (2003) Developmental Biology, 4th Edition

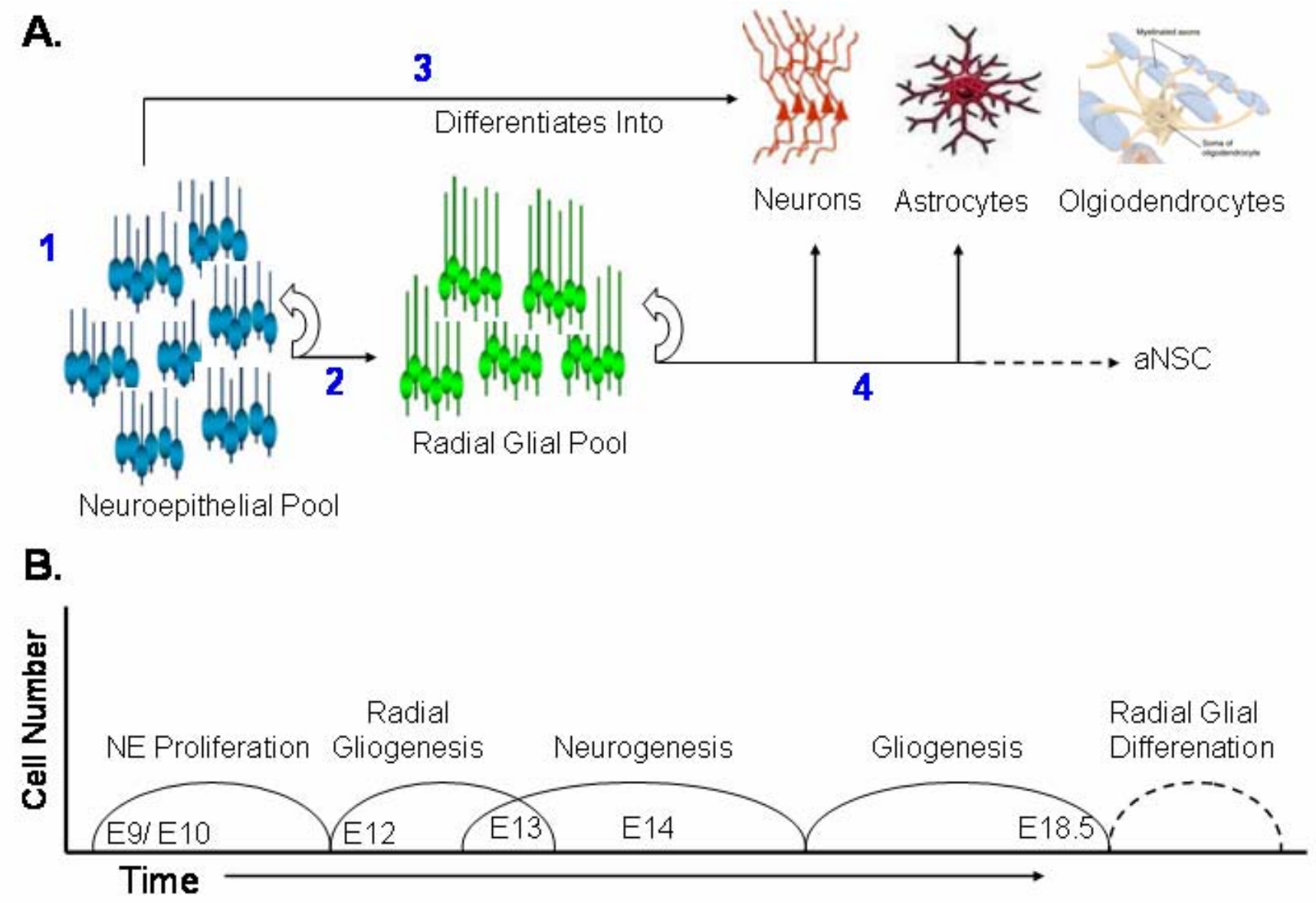

A.
3
Differentiates Into
Neurons
Astrocytes
Olgiodendrocytes
1
2
4
aNSC
Radial Glial Pool
Neuroepithelial Pool
B.
Cell Number
NE Proliferation
Radial Gliogenesis
Neurogenesis
Gliogenesis
Radial Glial Differenation
E9/ E10
E12
E13
E14
E18.5
Time

Figure 3.2. **Fetal Bovine Serum Induced Expression of Neural Genes.** hNP cells were cultured on polyornithine and laminin coated plates in –bFGF differentiation media with or without 5% FBS for two to seven weeks. Nestin and beta III Tubulin quantitative real time-PCR expression profile **(A)**. Nestin expression remains statistically insignificant throughout each respective differentiation culturing condition. Beta III Tubulin expression remained relatively low (<3.3 fold change) irrespective on culturing condition and to DAY0. The quantitative real time-PCR expression profile for GFAP **(B)** indicates no change (<1 relative fold, $P<0.05$) in over the seven week time period using –bFBF differentiation media alone. Conversely, upon the addition to 5% FBS to the medium, significant up-regulation ($P<0.05$) of GFAP is observed. A day effect ($P<0.05$) was observed on DAY42 – DAY49. **(C)** Immunocytochemistry demonstrating cell morphology and protein localization, respectively. ***** denotes significance ($p<0.05$) from DAY0 and ** denoting the observed day effect. Abbreviations: GFAP, glial fibrillary acidic protein.

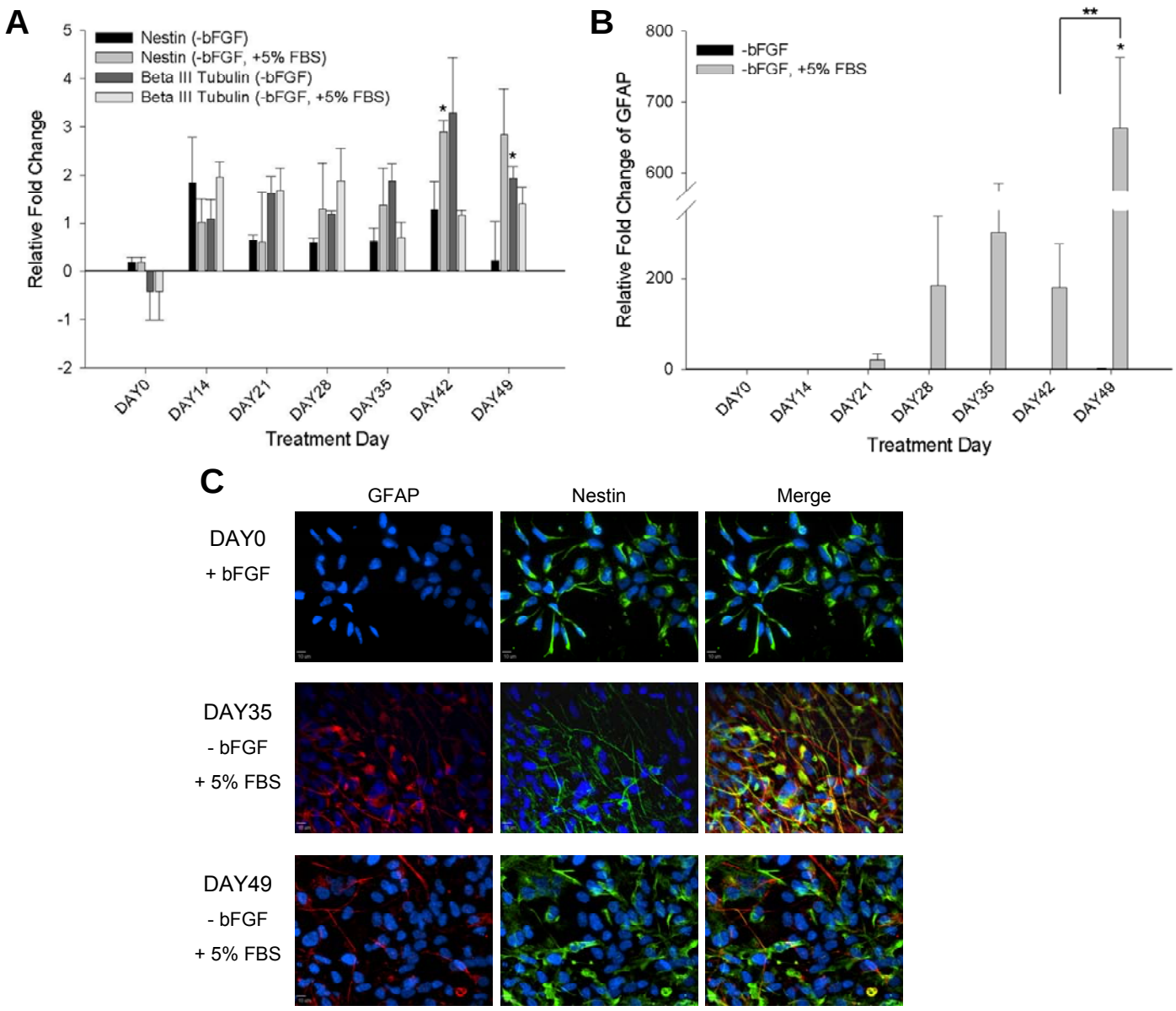

A
Nestin (-bFGF)
Nestin (-bFGF, +5% FBS)
Beta III Tubulin (-bFGF)
Beta III Tubulin (-bFGF, +5% FBS)
Relative Fold Change
5
4
3
2
1
0
-1
-2
DAY0
DAY14
DAY21
DAY28
DAY35
DAY42
DAY49
Treatment Day
B
-bFGF
-bFGF, +5% FBS
Relative Fold Change of GFAP
800
700
600
200
0
**
*
Treatment Day
C
GFAP
Nestin
Merge
DAY0
+ bFGF
DAY35
- bFGF
+ 5% FBS
DAY49
- bFGF
+ 5% FBS

Figure 3.3. Temporal Expression of Glial Progenitor and Glial-associated Genes.

A time-course reverse transcriptase-PCR analysis for the expression of **(A)** glial precursor markers and **(C)** glial-associated markers of hNP cells differentiated over 49 days in the presence of absence of 5% FBS. GAPDH was used as PCR internal control. Up Regulation of CD44, AQP4 and IL6R when 5% FBS is Supplemented to Differentiation Media was observed. Subsequent quantitative real time (RT)-PCR investigation of CD44 **(B)** and AQP4 **(B)** followed. ***** denotes significance ($p<0.05$) from DAY0; ** denotes a treatment effect. Abbreviations: NES, Nestin; VIM, Vimentin; CNTFR, ciliary neurotrophic factor receptor; IL6R, Interleukin 6 receptor; SC1, Schwann Cell Factor 1; BLBP, Brain Lipid Binding Protein; GLAST, Glutamate Transporter; GLUL, Glutamate-ammonia Ligase (Glutamine Synthetase); AQP4, Aquaporin 4; GAPDH, Glyceraldehyde 3-phosphate dehydrogenase

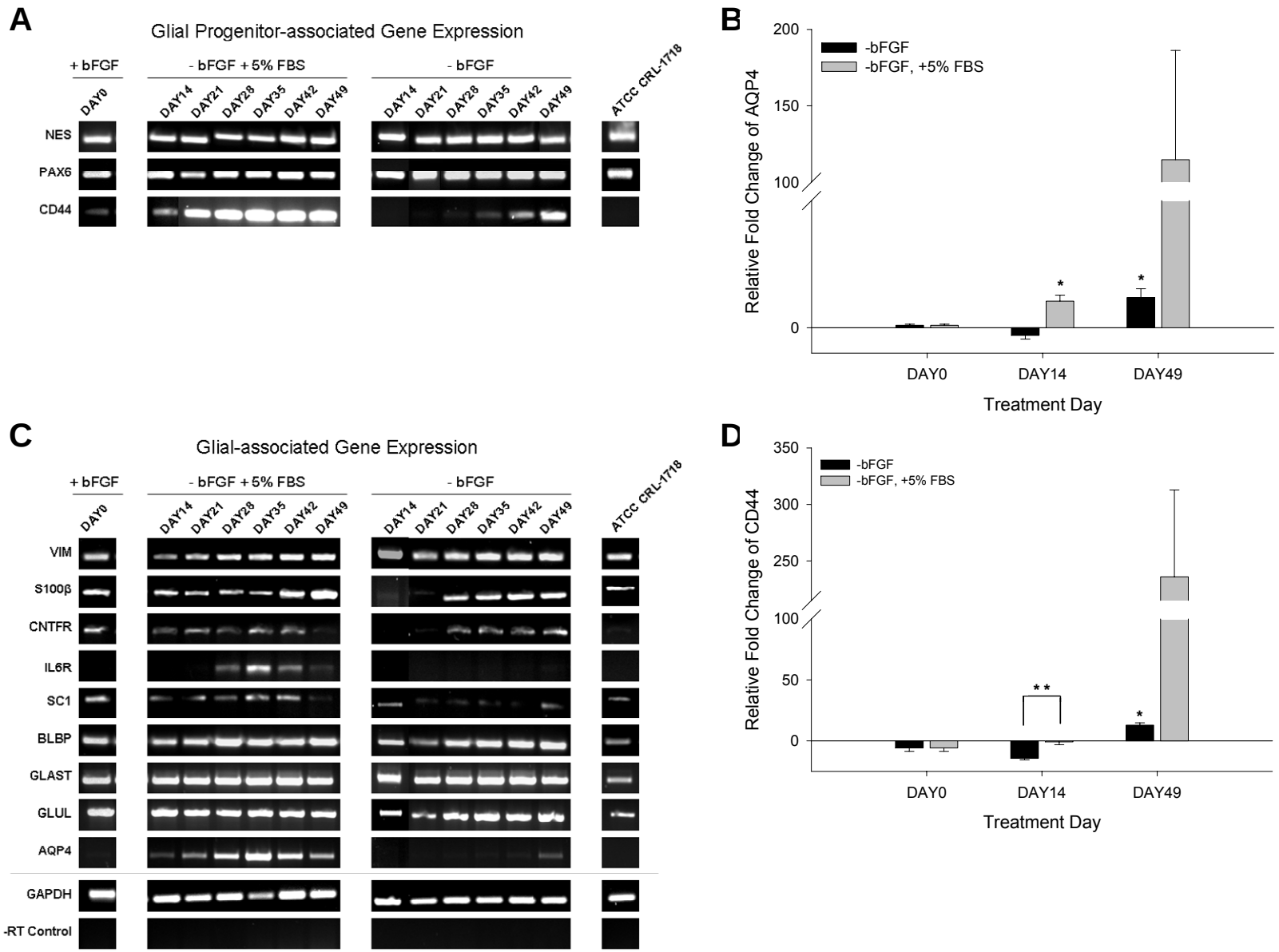

A
Glial Progenitor-associated Gene Expression
+ bFGF
- bFGF + 5% FBS
- bFGF
ATCC CRL-1718
DAY0
DAY14
DAY21
DAY28
DAY35
DAY42
DAY49
NES
PAX6
CD44
B
-bFGF
-bFGF, +5% FBS
Relative Fold Change of AQP4
200
150
100
0
DAY0
DAY14
DAY49
Treatment Day
C
Glial-associated Gene Expression
VIM
S100β
CNTFR
IL6R
SC1
BLBP
GLAST
GLUL
AQP4
GAPDH
-RT Control
D
Relative Fold Change of CD44
350
300
250
100
50
0

Figure 3.4. Lentiviral Transduced hNP cells Differentiated Towards GFAP$^+$ Lineage. **(A)** Scheme for positive selection of transduced hNP cells. DAY3 post-transduction, hNPs cells were selected for neomycin resistance. Cells were continuously cultured at low densities in the presence of G418 (200ug/ml) for 1 month, after which cells were differentiated in –bFGF media supplemented with FBS and the presence or absence of ACM while still under antibiotic selection. Immunocytochemistry confirmed the selective ability of the GFAP:GFP construct to only GFAP expressing cells **(B)**. DAY0 was devoid of GFAP and GFP fluorescence. Representative image from DAY35 of 10% FBS with ACM illustrates the selective nature of the construct via overlap of GFAP and GFP expression and the absence of GFP overlap in beta III Tubulin (TuJ-1) expressing cells. **(C)** Cell morphology of cells found within the culture. Comparison of Brightfield images depict hNP cells before differentiation, in –bFGF differentiation media without FBS and cell type present in FBS cultures. **(D)** Quantitative real time-PCR of the transduced cells differentiated over 35 days in the presence of 5% or 10% FBS with or without astrocyte conditioned media (ACM) was performed to investigate whether GFAP expression could be manipulated via decreased time for onset of expression or increased expression. The presence of 10% FBS supplemented to the –bFGF differentiation media demonstrated a significant ($p<0.05$) difference at DAY21 compared to 5% FBS –bFGF differentiation cultures, regardless of ACM. This suggests 10% FBS supplementation may decrease the time interval needed for the onset of GFAP expression. ***** denotes significance ($p<0.05$) from DAY0; ** denotes significance ($p<0.05$) between 5% and 10% FBS cultures regardless of addition of ACM at DAY21.

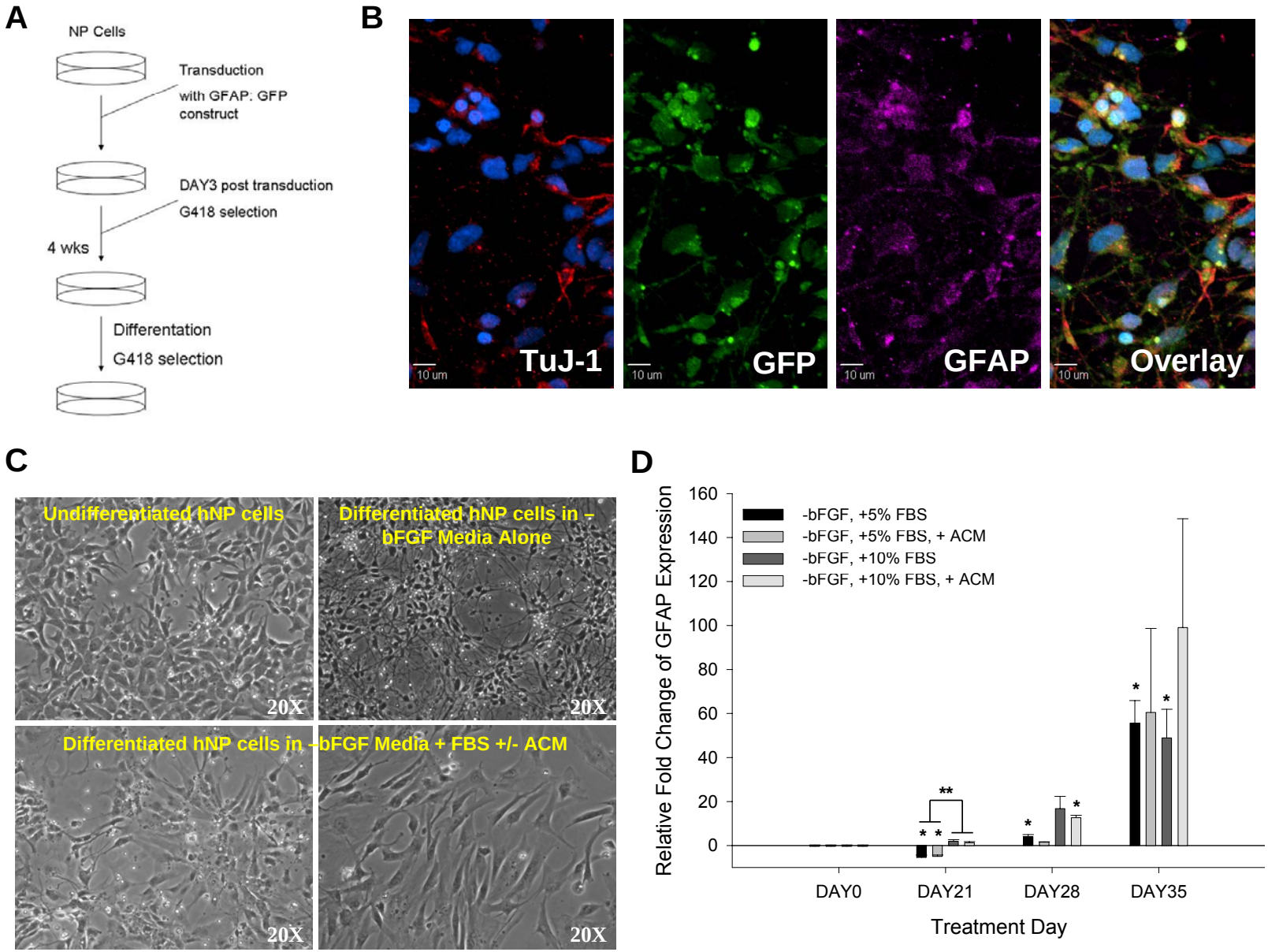
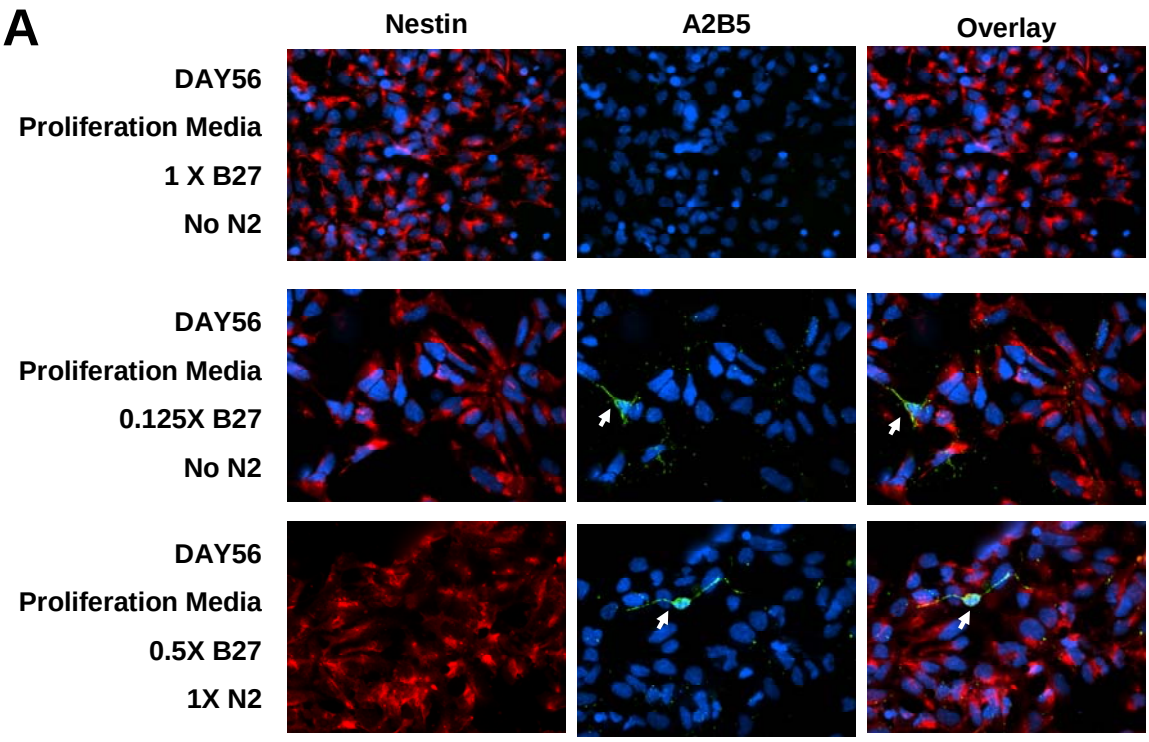
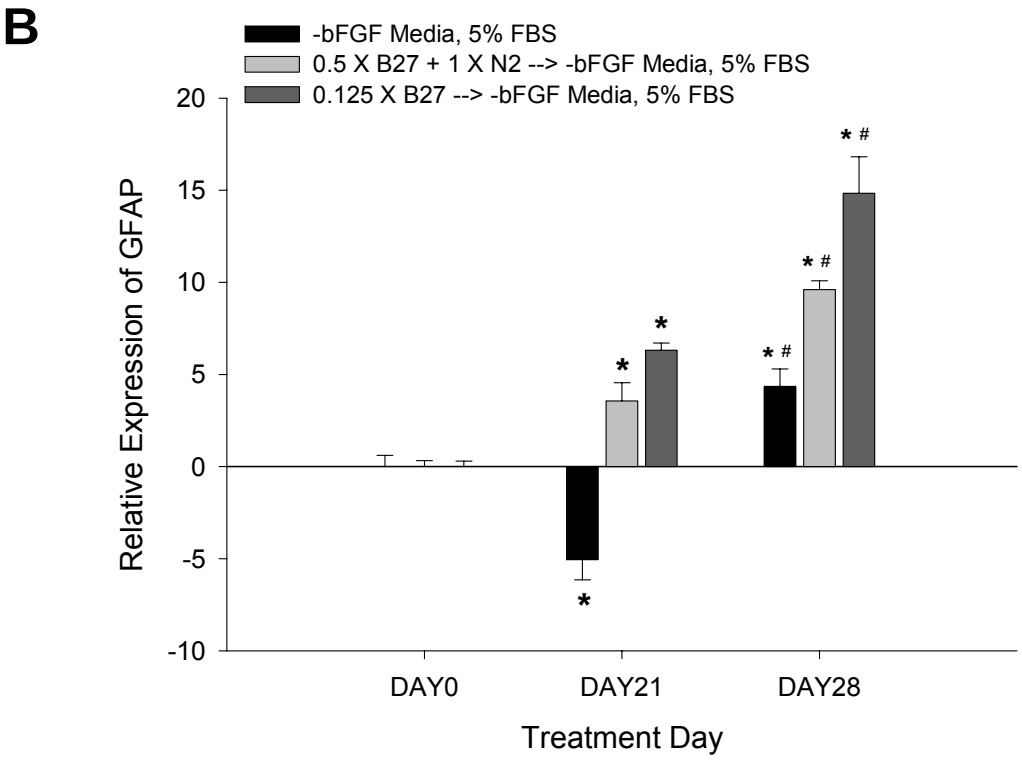

A
NP Cells
Transduction
with GFAP: GFP
construct
DAY3 post transduction
G418 selection
4 wks
Differentation
G418 selection
B
TuJ-1
GFP
GFAP
Overlay
10 um
C
Undifferentiated hNP cells
Differentiated hNP cells in –
bFGF Media Alone
Differentiated hNP cells in –bFGF Media + FBS +/- ACM
20X
D
Relative Fold Change of GFAP Expression
-bFGF, +5% FBS
-bFGF, +5% FBS, + ACM
-bFGF, +10% FBS
-bFGF, +10% FBS, + ACM
DAY0
DAY21
DAY28
DAY35
Treatment Day

Figure 3.5. Partial Withdrawal of B27 during hNP –bFGF Differentiation Increased GFAP Expression. **(A)** Varying concentrations of B27 (0 X, 0.125 X, 0.25 X, 0.5X and 1 X [control]) in the presence of absence of 1 X N2 media supplement were investigated for possible enhancement of glial differentiation. Cells were successfully maintained for two months at all B27 concentrations examined including 0.5 X B27 + 1 X N2 with the exception of 0 X B27. **(A)** These conditions were analyzed for A2B5 (glial precursor marker), GFAP (mature astrocyte marker) and Nestin (neural progenitor marker) via immunostaining with 0.125 X B27 and 0.5 X B27 + 1 X N2 reported here **(B)**. At DAY56, the cells were differentiated in –bFGF differentiation media supplemented with 5% FBS. Cultures retained Nestin expression and with a small percentage of cells adopting a glial precursor cell fate. **(C)** Quantitative real-PCR on 0.125 X B27 and 0.5 X B27 + 1 X N2 and compared against –bFGF differentiation supplemented with 5% FBS results from Figure 3.3. DAY21 treatments were significant ($p<0.05$) when compared to DAY0 hNP cells with 3.6 +/- 1 relative fold change for 0.5 X B27 + 1 X N2 and 6.3 +/- 0.4 for 0.125 X B27 conditions. Treatments at DAY28 were not only significant from DAY0, but all treatments demonstrated a day effect (denoted by #) when compared to DAY21. ***** denotes significance ($p<0.05$) from DAY0.


A
Nestin
A2B5
Overlay
DAY56
Proliferation Media
1 X B27
No N2
DAY56
Proliferation Media
0.125X B27
No N2
DAY56
Proliferation Media
0.5X B27
1X N2


B
-bFGF Media, 5% FBS
0.5 X B27 + 1 X N2 --> -bFGF Media, 5% FBS
0.125 X B27 --> -bFGF Media, 5% FBS
Relative Expression of GFAP
20
15
10
5
0
-5
-10
DAY0
DAY21
DAY28
Treatment Day

CHAPTER 4

DISTINCTION OF HUMAN EMBRYONIC STEM CELL LINES AND DERIVED CELL TYPES UTILIZING HIGH-THROUGHPUT PROTEOMICS AND FLOW CYTOMETRY[1]

---

[1]Brian A. Gerwe*, Peggi M. Angel*, Frank D. West III, Kowser Hasneen, Ron Orlando and Steven L. Stice.

*Authors contributed equally to this work

To be submitted to *Cell Stem Cell*

Abstract

Human embryonic stem cells (hESCs) present an invaluable resource within regenerative medicine and developmental research due to their self-renewal and differentiation potential. The unique proteome of each hESC line and their differentiated derivatives can be exploited for the identification of specific cell line biomarkers. Furthermore, the delineation of cell lineage based on cell surface markers will only add to the potential of hESCs whereby isolation of specific populations are possible for further study. Here, we applied a high-throughput proteomic approach to discover cell surface markers enabling the distinction between human derived cell lines WA09 (H9), BG01, WA09-derived neural progenitor (hNP) cells and abBG02 (trisomy 12, 14, 17 and an additional X chromosome copy). We discovered proteins unique to each cell line utilizing liquid-chromatography coupled to tandem mass spectrometry (LC-MS/MS). Flow cytometry was used to investigate select proteins from the proteomic dataset. We found that junctional adhesion molecule 1 (JAM-1) was unique to the BG01, BG02 and abBG02 hESC lines, while dysferlin (DYSF) expression was specific to the WA09 hESC line. In addition, ciliary neurotrophic factor receptor alpha (CNTFRα) expression was able to distinguish the WA09-derived hNP cells from the parent hESC population. The current study provides proof of concept that a combination of proteomics and traditional cell defining methods are an efficient and sensitive approach to identify new surface candidates for cell line and lineage distinctions.

Introduction

Embryonic stem cells (ESCs) are unique cells with the capacity to undergo self-renewal and differentiation into all adult cell types (13, 22, 41). When stably maintained in vitro, human ESCs (hESCs) provide a potentially unlimited cell source for clinical applications such as regenerative medicine and cell therapy. When differentiated, hESCs hold potential for drug discovery, toxicology screens and serve as a model for studying basic human development.

Embryonic cells and their differentiated counterparts are commonly distinguished by a collection of a few markers. These markers denote pluripotency (e.g. SSEA-3 and -4 (34, 41)) or a differentiated cell status. Although it is well known that there are distinct genetic differences in human embryonic stem cells (1, 2), few markers, if any, are able to provide a distinction between hESC lines. Additionally, chromosomal instability of hESCs has been demonstrated with long-term enzymatic bulk passaging techniques (10, 12, 26, 48). While the causes of this genetic instability are poorly understood, the ability to identify karyotypically abnormal hESCs could provide a screen to detect abnormal cells, which is essential for developing clinical therapies. Thus, development of markers to better characterize and distinguish particular hESC populations has high potential for providing an invaluable resource to culturing and differentiating hESCs.

Multipotent progenitor cells, such as neural progenitor and mesenchymal stem cells, represent an initial level of hESC differentiation that produce cells that are unipotent or lineage restricted (45). With respect to human neural progenitor (hNP) cells, Shin et al. (36, 37) demonstrated differentiation into all three neural lineages found within the central nervous system. The unique protein expression pattern displayed at

various stages of differentiation has enabled additional researchers to define (neural) fate decisions (6, 17, 32). However, the usefulness of this system is somewhat limited when attempting to address questions about a specific neural cell type, as these cultures give rise to large heterogeneous populations (29). Heterogeneity is proposed to result from the cellular clumping methods employed during isolation due to increased cell survival over single cell dissociation methods (29). The issue of heterogeneity is further compounded by the lack of cell surface markers needed for isolation of pure neural populations, which will be essential for detailed study and clinical applications.

Previous studies have shown differential expression of proteins between normal hESC lines and karyotypically abnormal (16) or embryonic carcinoma lines (11). These unique protein expression patterns have enabled researchers to explore further lineage fate decisions. Continued investigation of global surface protein expression relative to differentiated cell types could potentially enhance our understanding of the regulatory mechanisms that permit or prohibit the differentiation of hESCs into progenitor cells and, ultimately, progenitor cells into their mature differentiated cell type.

Given that hESC lines have been established through various methods and with different source material, each line encapsulates a unique proteome that can be exploited for the identification of hESC line specific biomarkers. A high-throughput membrane proteomics approach is an ideal way to identify novel or unique surface marker candidates for cell lines (19, 46). A membrane proteomics study investigates the comprehensive membrane protein expression of a cell type or tissue under defined biological conditions. This strategy strives to enrich for and capture the expression patterns of hundreds of insoluble proteins. Within this collection of hundreds of

proteins, individual proteins provide highly informative data that includes amino acid sequence information. Amino acid sequence information may be used to predict subcellular localization, number of transmembrane passes, or hydrophilic/hydrophobic regions of the protein thereby annotating a protein as membrane bound or associated. By comparing membrane proteomic datasets from different cell types, a population of membrane bound proteins may be isolated with a high potential of uniqueness to a specified cell type. Proteomic approaches have been recently used to establish unique comparative membrane signatures of a variety of cell types including human embryonic stem cells versus human embryonal carcinoma cells (11), human epithelial cells during cancer progression (43), and mouse astrocytes versus mouse astrocytomas (33). For the stem cell field, this presents a powerful methodology to produce candidate surface markers that define a particular cell line or denote cells of a differentiated status.

In this study, we apply a high-throughput proteomics approach employing membrane fractionation and LC-MS/MS analysis to identify membrane protein signatures unique to hESC lines and their differentiated cell types. Membrane fractions of the human derived cell lines WA09, BG01, WA09-derived neural progenitor (hNP) cells and abBG02 (a chromosomal aberrant BG02 embryonic stem cell line having trisomy 12, 14, 17 and an additional X chromosome copy) were isolated and analyzed to identify novel cell surface markers. From the proteomic analysis, several biomarker candidates were validated with flow cytometry. We establish that junctional adhesion molecule 1 (JAM-1 also referred to as F11R or CD321) is unique to the BG01, BG02 and abBG02 hESC lines, while dysferlin (DYSF) and Frizzled 7 (FZD7) expression were specific to the WA09 hESC line. Furthermore, we demonstrate proof of concept for

uniquely delineating a parental hESC population from their differentiated early derivatives. Using ciliary neurotrophic factor receptor alpha (CNTFRα) expression, hNP cells could be distinguished from the WA09 hESC parent population. Cell surface markers derived from high-throughput proteomics provide a highly efficient technique to produce new candidates for cell sorting.

Materials and Methods

hESC Culture Conditions

BG01, BG02 and WA09 hESC with normal karyotype were cultured on ICR mouse (Harlan) embryonic fibroblast (MEF) feeders inactivated by mitomycin C (Sigma-Aldrich) or on polyornithine (20 µg /ml) and laminin (5 µg/ml) coated (feeder free) plates. The cells were cultured in 20% KSR stem cell media, which consists of Dulbecco's modified Eagle medium (DMEM)/F12 supplemented with 20% knockout serum replacement (KSR), 2 mM L-glutamine, 0.1 mM non-essential amino acids, 50 units/ml penicillin/50 µg/ml streptomycin (Gibco), 0.1mM β-mercaptoethanol (Sigma-Aldrich) and 4 ng/ml bFGF (Sigma-Aldrich and R & D Systems). Media prepared for feeder free cultures were conditioned by exposing them to MEFs for 24 hours and supplemented with an additional 4ng/ml of bFGF. hESCs were maintained in 5% CO_2 and at 37° C. Cells were passaged every 3 days by mechanical dissociation and re-plated to prevent undirected differentiation with daily media changes as previously described by our laboratory (25). Bulk passaging of embryonic stem cells was performed using 0.05% trypsin-EDTA (Gibco). BG02 cells with abnormal karyotype, termed abBG02, were also cultured as described above.

Human Neural Progenitor (hNP) Culture Conditions

hNP were derived from WA09 human embryonic-stem cells with normal karyotype as described in past reports by our laboratory (25). Cells were propagated and maintained in proliferation media consisting of neurobasal medium (Gibco) supplemented with 2mM L-glutamine, 1x penicillin/ streptomycin, 1x B27 (Gibco), 20ng/ml of bFGF (Sigma-Aldrich) and 10ng/ml of Leukemia inhibitory factor (LIF, Chemicon) under conditions of 5% CO_2 atmosphere at 37°C. Cells were passaged 1:3 or 1:4 by mechanical dissociation via trituration and re-plated on polyornithine (20 µg/ml) and laminin (5 µg/ml) coated plates with media changes every other day.

*Human Mesenchymal Stem Cell (*hMSC*) Culture Conditions*

WA09-derived hMSCs (ArunA Biomedical) were derived from WA09 human embryonic stem cells of a normal karyotype. Cells were maintained in alpha MEM (Gibco), 10% fetal bovine serum (FBS, Hyclone), 2mM L-glutamine (Gibco) and 1X penicillin / streptomycin (Gibco) at 5% CO_2 atmosphere and at 37°C. Cells were maintained and passaged 1:2 by enzymatic dissociation (0.05% trypsin-EDTA, Gibco) and re-plated with media changes every third day.

Membrane Isolation

Membrane proteins were isolated as described previously (24, 33, 38, 46). In brief, cells were washed three times in ice cold PBS, collected by centrifugation at 1500 g. Cells were suspended in 3 mL lysis buffer (10 mM HEPES, 1 mM EDTA, pH 7.2) with protease inhibitors (Sigma-Aldrich) and incubated on ice for 30 minutes. Cell membranes were disrupted using 20 strokes of a chilled 7 mL Dounce homogenizer. A buffered sucrose solution (500 mM sucrose, 1 mM EDTA, pH 7.2) was added in equal volume to the homogenate followed by an additional 20 strokes of the Dounce

homogenizer. Cell debris was removed from the homogenate by pelleting at 7000 g for 10 minutes at 4˚C. The supernatant was collected, placed in a pre-chilled swinging rotor bucket, and topped with 6 mL of the 50/50 lysis/sucrose solution. Membranes were isolated by centrifugation at 120,000 g for 1.5 hour at 4˚C. The supernatant was removed and the pellet was dissociated in 100 mM Na_2CO_3, pH 11.3. A high speed centrifugation of 120,000 g was applied for 1.5 hour at 4˚C to collect the pellet. The pellet was immediately dissolved in a 5% SDS, 15 mM $NaCO_2$. Protein concentration was determined using a BCA assay, following the manufacturer's instructions.

In-gel Digestion

An equal amount of lyophilized protein (70 µg) from each cell type was used for analysis. Dried protein was dissolved in Laemmli buffer, boiled for 15 minutes, cooled on ice and loaded onto a 4-12% Bis-Tris minigel (Invitrogen). Electrophoresis was performed for 45 minutes at 150 volts using a MOPS-SDS running buffer with an antioxidant additive (Invitrogen). Gels were washed in ice cold water for 10 minutes and cut into nine fractions. Each fraction was minced into 1 mm cubes and washed three times in 50 mM NH_4HCO_3. Bands were dehydrated three times by 20 minute incubations in 50/50 (50 mM NH_4HCO_3) / acetonitrile followed by 100% acetonitrile and drying under vacuum for 25 minutes. Freshly prepared 20 mM DTT/ 50 mM NH_4HCO_3 was added to each fraction and incubated for 2 hours at 65˚C. The DTT solution was replaced with 40 mM iodoacetamide and incubated for 1 hour at room temperature. Gel pieces were then washed and dehydrated as described above. Proteins were digested by adding 100 µL of a 0.02 µg/µL trypsin solution to the pieces and incubating at 37˚C overnight. Peptides were extracted once with 50/50 (5% formic acid)/ acetonitrile to halt

digestion and twice with a 50/50 (50 mM NH_4HCO_3) /acetonitrile solution. Peptide solutions were dried to reduce volume and filtered with a 30,000 MWCO filter to remove trypsin.

LC-MS/MS

Peptides from gel fractions were pressure loaded onto a 75 μm I.D. microcapillary column hand packed with 8.5 cm of 5 μm, 300 Å C18 (Rainin). Peptide separation was performed via a 0.75% B/min gradient from 5% B to 60% B at ~200 nL/min using an Agilent 1100 series HPLC. Mobile phase A was 0.1% formic acid in water, mobile phase B was 0.1% formic acid in acetonitrile. Peptides eluted directly from the column into a hybrid linear ion trap coupled to a Fourier Transform Ion Cyclotron Resonance (FT-ICR) mass spectrometer (LTQ-FT, ThermoFinnegan). Source voltage was set to 2.2. kV. Each full MS scan (m/z range 450 -2000) was followed by nine MS/MS events on the nine most abundant peaks. Dynamic exclusion was enabled and utilized a maximum exclusion list size of 50 with an exclusion duration of 160 s.

Protein Identification

Raw tandem mass spectra were converted into mzXML format and then into peak list using ReAdW followed by mzMXL2Other maintaining the default parameters (28). Peak Lists (PKL files) were searched using Mascot version 1.9.0 (www.matrixscience.com) against an NCBI non-redundant human database (34,180 entries) downloaded 3/30/08. Search parameters were set to trypsin as the enzyme specification, 2 missed cleavages, 50 ppm for the precursor and 0.6 Da for the fragment masses. Variable modifications used were carbamidomethylation of cysteines,

deamidation of asparagines, and methoxidation (expected experimental modifications). A decoy database strategy was used to limit random matches. Peptides identified from the database search were clustered using ProteoIQ version 1.06. The initial protein list was produced using a false discovery rate ≤ 5.0, a protein probability of 0.1, minimum ion score of 10 and maximum peptide length of 4. The combined use of false discovery rate and protein probability results in a much lower FDR, as discussed previously (8). For this analysis, the lowest scoring protein was calculated to have a Mascot score of 51, corresponding to a protein FDR of 1.02% and a peptide FDR of 0.028%.

Prediction of Protein Subcellular Location

Prediction of plasma membrane and transmembrane spanning regions were made using WoLF PSORT (18) and TMHMM v2.0 (www. http://www.cbs.dtu.dk). Manual verification of candidate location by literature search was employed to ensure cell surface location.

Flow Cytometry

Cells were harvested, washed in $PBS^{-/-}$ and fixed with 2% PFA/ $PBS^{+/+}$ for 10 minutes at room temperature. Cells were then washed 3 times in $PBS^{-/-}$ and blocked in 6% donkey or goat serum for 1 hour. Antibodies were directed against Dysferlin (Novus Biologicals, 1:60) Frizzled 7 (R&D Systems; 1:100), JAM-1 (BD Bioscience, 1:200) and CNTFRa (Santa Cruz Biotechnology, 1:100). Primary antibodies were detected using secondary antibodies conjugated to Alexa Flour 488 (Molecular Probes, 1:1000). Secondary Antibody only samples served as negative controls. Cells were sorted and analyzed using Dakocytomation Cyan (DakoCytomation) and FlowJo Cytometry analysis software (Tree Star).

Results and Discussion

Surface membrane proteins are key players in a cellular microenvironment that includes cell-to-cell and cell-matrix mediated signaling. Expression of cell surface proteins may be altered due to external signaling stimului (e.g. response to cell media) and/or intracellular signaling (e.g. native differentiation), often a cyclical response (9, 30, 47). Changes in protein expression on the membrane surface reflect functional changes of the cell; for embryonic stem cells, these alterations are linked to genetic variation or lineage differentiation.

Membrane Proteomics

A membrane proteomics approach was employed to identify new candidates of cell surface expression by comparing normal embryonic stem cell lines (BG01 and WA09), genetically abnormal embryonic stem cell lines (abBG02), and a neural progenitor line (hNP, WA09-derived) against each other. Insoluble proteins from each cell type were isolated using a sucrose gradient, ultra high speed centrifugation and analyzed by SDS-PAGE. Gel lanes were cut into fractions and tryptically digested. Tryptic peptides extracted from the SDS-PAGE separated proteins were analyzed by microcapillary reverse phase chromatography coupled to an LTQ-FT mass spectrometer, selected for rapid mass scanning and high mass accuracy capability. Peptide fragmentation patterns were compared to an *in silico* digestion of the human proteome using the MASCOT algorithm. Statistical clustering of identified peptides was performed using ProteoIQ, employing both false discovery rate and protein probability to limit random matches and eliminate redundant peptides. Software programs Wolf Psort and TMHMM were used to predict localization of proteins to the plasma membrane or

transmembrane spanning regions, respectively, based upon total amino acid sequence of the identified proteins. The resulting list of proteins was manually verified to confirm membrane status.

A total of 779 membrane proteins were found across all proteomes, an approximate 40% overall enrichment (Figure 4.1). Analysis of crude enrichments of insoluble proteins showed that embryonic stem cell proteomes were enriched to an approximate 37%, while neural progenitors were enriched to 63%. The larger percentage of predicted membrane proteins for the hNPs is likely due to the more differentiated status of this cell type. Comparisons were made between membrane proteins predicted by TMHMM, which predicts transmembrane spanning regions, and Wolf Psort, which predicts subcellular location. The majority of isolated insoluble proteins were predicted to have transmembrane spanning regions (Figure 4.1). Only a small percentage of insoluble proteins (7.7% BG01, 5.0% abBG02, 5.3% WA09, 6.4% hNP) were not predicted with transmembrane spanning regions, defined only by Wolf Psort as having a high chance of localization to the plasma membrane. Thus, the isolation method had a high degree of bias towards transmembrane spanning proteins, proteins likely to be candidates for cell surface markers.

Venn Diagram Analysis

Analysis of membrane proteins by Venn diagram (Figure 4.2) illustrates the distribution of membrane proteins among the four cell lines. Unique expression is defined as expression seen exclusively in one population versus the others, but does not rule out false negatives due to mass spectrometry detection limits. Therefore,

proteins should not be considered biomarkers of a particular cell line based solely on proteomics analysis. All 779 proteins are identified with false discovery rates ≤ 1%.

As shown by the Venn diagram, hNP have the highest percentage of unique membrane proteins (18.8%) compared to membrane proteins across all lines. Each embryonic stem cell line had a smaller population (<10%) of membrane proteins identified exclusive to each cell line, including the abnormal BG02 line. The larger percent of proteins linked only to the hNP may be expected as these cells are further down the path of differentiation. However, the significant overlap between hNPs with hESCs indicates a conservation of expression supporting selective regulation of cell fate decisions (7).

Overall, 19% of all proteins overlapped between all populations and a majority of this shared expression included the neural progenitor line. No overlap was soley observed between the WA09 and hNP or between abBG02 and hNP in this study. It should be noted that quantitative changes in protein expression were not examined in this investigation; however, there is high potential for proteins shared between cell populations to be differentially regulated. For instance, differences in sequence coverage, used as a semi-quantitative measurement of regulation (14), were observed in the shared BG01/ hNP population. Eighteen of the 56 protein identified in this shared population had over 5% greater sequence coverage in one cell population versus the other. While this study focused on surface candidates for cell line discrimination, the proteomics approach was useful in providing a fingerprint of protein expression for each cell type. Further investigation of this information could be informative in understanding subtle differences between the chosen cell lines.

Protein Expression

Although shared proteins were not selected as candidates for surface markers, shared populations were examined for key protein groups. Interestingly, over 5% of the shared population were proteins associated with glycosylation (e.g., mannosylation, asparagine-linked glycosylation, and glycosyltransferases) or were known glycoproteins (e.g., basigin 1 (all cells) and basigin 3 (BG01, abBG02), heparin sulfate proteoglycan 2 (WA09), and synaptic 2 glycoprotein (hNP). Glycoproteins are known to be differentially expressed on hESC surfaces and have been previously shown to be robust markers for pluripotency (31, 42). Identification of this post-translational activity and related proteins show the strength of the proteomics approach in defining the functional landscape of undifferentiated and lineage committed human embryonic stem cells.

Growth factors were another protein group highlighted by the proteomic analysis. Although growth factors are generally soluble proteins, they bind to surface receptors and are expected to be found in close association with the membrane population. This analysis found 13 growth factors related to the cell populations, including FGF2 and FGF binding protein 3. Nearly half (46%) of the growth factors were part of the insulin or insulin-like growth factor family, either binding or receptor related, important to growth and differentiation (4, 20). These proteins interact with integrins, another protein family highly expressed. Here, the proteomic strategy identified twelve integrins and one integrin binder (calmyrin, WA09). Eight integrins were observed only in embryonic stem cell populations, while four integrins were identified across all cell populations. It is common knowledge that integrins play an important role in focal adhesion, cell-cell

adhesion, and cell-matrix signaling (15) and are thus critical for stem cell survival and differentiation.

Protein expression belonging to the abnormal BG02 line was investigated. Since the abnormal BG02 line is characterized by trisomy 12, 14, 17, and X, membrane proteins assigned to this population were annotated by gene location. Of the 61 proteins identified only in the BG02 population, thirteen proteins were derived from aberrant chromosome locations. Dishevelled-associated activator of morphogenesis 1, DAAM1, was an especially interesting protein observed in this group. Database annotation shows that DAAM1 is involved in cell functions of motility, adhesion, cytokinesis and is implicated in cell polarity. DAAM1 is reported to be an essential component of the WNT signaling pathway and plays important roles in development including directing migration of individual cells, and in cardiac and neuronal development (44). DAAM1 is intracellularly membrane bound and therefore not particularly useful for sorting abnormal cells. However, since this protein mediates at least three signaling cascades to act as a transducer from frizzled receptors of the WNT pathway to RHO and RAC (44), this protein has potential to denote chromosomal abnormalities by immunohistochemistry approaches making it further intriguing. The identification of this protein and others led to a high percentage (21%) of membrane proteins relevant to trisomy abnormalities. This demonstrates the sensitivity of the proteomic approach towards characterizing protein expression resulting from abnormal genetic expression.

On the whole, investigation of protein populations obtained by membrane proteomic analysis of the four selected cell lines resulted in rich and appropriate

information describing protein populations from embryonic stem cell lines, progenitor lines, and the abnormal embryonic stem cell line.

Validation of High-Throughput Proteomics Approach Using Flow Cytometry

Candidates for surface marker expression were selected from protein populations unique to cell types (Table 1). Surface candidate proteins had functions ranging from cell adhesion to involvement in muscle contraction to cytokine receptors. Flow cytometry on top candidates with cell surface antibodies was used to validate the LC-MS/MS proteomics. Dysferlin, a novel molecule involved in muscle contraction with no mammalian protein homology (3), was a top candidate to identify the WA09 hESC line. Consistent with the proteomic data, Dysferlin was differentiately expressed between analyzed cell lines (Figure 4.3) demonstrating higher expression (79.2% positive) when compared to BG01 (12.3%), abBG02 (27.6%) and WA09 hESC-dervied hNP cells (26.4%).

The identification of CNTFRα in the proteomics analysis enabled an effective distinction between the parental WA09 hESC line and its derived hNP lineage via flow cytometry (Figure 4.3). This protein is the alpha subunit of the CNTF complex that confers specificity, highly conserved across species, and predominantly localized to cells of the nervous system and skeletal muscle (39). hNP cells were found to be 58.3% positive for CNTFRα, while the parental WA09 stem cell population was only 1.6% positive. Shimazaki and colleagues (2001) showed that CNTF, which recruits and activates the leukemia inhibitory factor receptor (LIFR) /gp130 complex through CNTFR, supports the self-renewal capability of EGF-responsive neural stem cells through the suppression of glial lineage restriction progenitor cells (35). Strengthening the argument

for the high degree of selectivity for the cell surface markers chosen, a WA09-derived mesenchymal stem cell line (hMSC) was additionally stained and quantified (Figure 4.3). Flow cytometry resulted in low expression levels of dysferlin (1.8% positive cells) and CNTFRα (1.0% positive cells). It should be noted that other cell lines studied were found to contain low expression levels of CNTFRα, 9.7% and 7.0% positive for cells of the BG01 and the abnormally karyotyped BG02 hESC lines, respectively.

Two additional cell surface proteins identified and validated with flow cytometry are Frizzled 7 (FZD7) and junctional adhesion molecule 1 (JAM-1), which demonstrated uniqueness in proteomic analysis for BG01 and karyotically abnormal BG02 cells, respectively. Frizzled 7 is a 7-transmembrane domain protein acting as a receptor for WNT signaling proteins, which has recently been implicated to contributing to the self-renewal capability of hESCs (23). Even though Frizzled 7 was predicted to be uniquely expressed by the BG01 hESC line, it was only expressed at marginal levels (0.6% positive cells). Interestingly, 79.5% of the cells within the WA09 population were positive to the Frizzled 7 antibody. All other cell lines examined by flow cytometry reported less than 24% positive cell levels. It is possible that Frizzled 7 was not expressed on the surface of BG01 hESC line, but could have been localized to intracellular membranous compartments that were co-purified during insoluble protein isolation. Interestingly, Dromeyer and collegues (2008) concluded that it was uniquely expressed in the HUES-7 hESC line as compared to the NT2/D1 ECC cell line (11).

JAM-1 has been reported as a novel cell surface marker for hematopoietic stem cells (40) and as a marker for cells of the endothelial lineage (21, 27). Of the cell lines analyzed, BG01 (97.1%) and abBG02 cells (97.7%) (Figure 4.4) were positive relative

to 9.1% for WA09 and 23.9% for the WA09-derived hNP cells (Figure 4.3). With hopes of identifying an abnormal karotype, karyotypically normal BG02 cells were additionally quantified using flow cytometry (Figure 4.4). BG02 cells of a normal karyotype produced similar results for JAM-1 (94.4% positive cells) relative to those of the abnormal karyotype. These results suggest that JAM-1 is a cell surface marker that could identify BG01 and karotypically abnormal and normal BG02 hESC lines. Interestingly, it was recently reported that BG02 cells were directed down the endothelial lineage to form primitive vascular networks under the effects of bone morphogenetic protein-4 (BMP4), which is known to be required for vascular/hematopoietic specification (5). Therefore, the high expression of JAM-1 as compared to the other hESC lines analyzed may indicate a propensity for BG01 and BG02 hESC lines to differentiate to an endothelial fate.

Conclusion

Using a high-throughput proteomics approach, we identified 779 membrane proteins from the BG01, WA09, abBG02 and WA09-derived neural progenitors. This demonstrates that examination of global membrane protein expression has high potential for identification of new surface candidates that can discriminate between normal, abnormal, and progenitor cell lines. The proteomic study highlights the importance of membrane proteins and their functional role in crucial fate-determining processes such as cell-cell signaling and cell-matrix interactions. These findings were further supported by flow cytometry. We showed junctional adhesion molecule 1 (JAM-1) is unique to BG01 and BG02 hESC lines, regardless of karyotype. Interestingly, a recent publication used similar methods and demonstrated that JAM-1 and Dysferlin

were expressed at similar levels within the HUES-7 hESC line and NT2/D1 ECC cell line (11). This emphasizes that cell surface markers used to delineate one particular hESC line from another does not ensure universal application. A high-throughput proteomic approach is advantageous and possibly necessary via the customizability of the number of cell lines screened. We additionally illustrated proof of concept for uniquely distinguishing a parental hESC population from differentiated early derivatives. Dysferlin (DYSF) was specific for the WA09 hESC line, while ciliary neurotrophic factor receptor alpha (CNTFRα) was able to delineate the derived hNP cell population. Thus, we provide evidence that unique identification of cell surface markers based on high-throughput proteomics can be achieved where possible cell discrimination can be employed to limit heterogeneous cell populations. This study allows new insight into functional differences between embryonic stem cell and progenitor lineages.

References

1. **Abeyta, M. J., A. T. Clark, R. T. Rodriguez, M. S. Bodnar, R. A. Pera, and M. T. Firpo.** 2004. Unique gene expression signatures of independently-derived human embryonic stem cell lines. Hum Mol Genet **13:**601-8.

2. **Adewumi, O., B. Aflatoonian, L. Ahrlund-Richter, M. Amit, P. W. Andrews, G. Beighton, P. A. Bello, N. Benvenisty, L. S. Berry, S. Bevan, B. Blum, J. Brooking, K. G. Chen, A. B. Choo, G. A. Churchill, M. Corbel, I. Damjanov, J. S. Draper, P. Dvorak, K. Emanuelsson, R. A. Fleck, A. Ford, K. Gertow, M. Gertsenstein, P. J. Gokhale, R. S. Hamilton, A. Hampl, L. E. Healy, O. Hovatta, J. Hyllner, M. P. Imreh, J. Itskovitz-Eldor, J. Jackson, J. L. Johnson, M. Jones, K. Kee, B. L. King, B. B. Knowles, M. Lako, F. Lebrin, B. S. Mallon, D. Manning, Y. Mayshar, R. D. McKay, A. E. Michalska, M. Mikkola, M. Mileikovsky, S. L. Minger, H. D. Moore, C. L. Mummery, A. Nagy, N. Nakatsuji, C. M. O'Brien, S. K. Oh, C. Olsson, T. Otonkoski, K. Y. Park, R. Passier, H. Patel, M. Patel, R. Pedersen, M. F. Pera, M. S. Piekarczyk, R. A. Pera, B. E. Reubinoff, A. J. Robins, J. Rossant, P. Rugg-Gunn, T. C. Schulz, H. Semb, E. S. Sherrer, H. Siemen, G. N. Stacey, M. Stojkovic, H. Suemori, J. Szatkiewicz, T. Turetsky, T. Tuuri, S. van den Brink, K. Vintersten, S. Vuoristo, D. Ward, T. A. Weaver, L. A. Young, and W. Zhang.** 2007. Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. Nat Biotechnol **25:**803-16.

3. **Anderson, L. V., K. Davison, J. A. Moss, C. Young, M. J. Cullen, J. Walsh, M. A. Johnson, R. Bashir, S. Britton, S. Keers, Z. Argov, I. Mahjneh, F.**

Fougerousse, J. S. Beckmann, and K. M. Bushby. 1999. Dysferlin is a plasma membrane protein and is expressed early in human development. Hum Mol Genet **8:**855-61.

4. **Baxter, R. C.** 2000. Insulin-like growth factor (IGF)-binding proteins: interactions with IGFs and intrinsic bioactivities. Am J Physiol Endocrinol Metab **278:**E967-76.

5. **Boyd, N. L., S. K. Dhara, R. Rekaya, E. A. Godbey, K. Hasneen, R. R. Rao, F. D. West, 3rd, B. A. Gerwe, and S. L. Stice.** 2007. BMP4 promotes formation of primitive vascular networks in human embryonic stem cell-derived embryoid bodies. Exp Biol Med (Maywood) **232:**833-43.

6. **Carpenter, M. K., M. S. Inokuma, J. Denham, T. Mujtaba, C. P. Chiu, and M. S. Rao.** 2001. Enrichment of neurons and neural precursors from human embryonic stem cells. Exp Neurol **172:**383-97.

7. **Chang, H. H., M. Hemberg, M. Barahona, D. E. Ingber, and S. Huang.** 2008. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature **453:**544-7.

8. **Choi, H., and A. I. Nesvizhskii.** 2008. Semisupervised model-based validation of peptide identifications in mass spectrometry-based proteomics. J Proteome Res **7:**254-65.

9. **Conley, B. J., J. C. Young, A. O. Trounson, and R. Mollard.** 2004. Derivation, propagation and differentiation of human embryonic stem cells. Int J Biochem Cell Biol **36:**555-67.

10. **Cowan, C. A., I. Klimanskaya, J. McMahon, J. Atienza, J. Witmyer, J. P. Zucker, S. Wang, C. C. Morton, A. P. McMahon, D. Powers, and D. A. Melton.** 2004. Derivation of embryonic stem-cell lines from human blastocysts. N Engl J Med **350:**1353-6.

11. **Dormeyer, W., D. van Hoof, S. R. Braam, A. J. Heck, C. L. Mummery, and J. Krijgsveld.** 2008. Plasma membrane proteomics of human embryonic stem cells and human embryonal carcinoma cells. J Proteome Res **7:**2936-51.

12. **Draper, J. S., K. Smith, P. Gokhale, H. D. Moore, E. Maltby, J. Johnson, L. Meisner, T. P. Zwaka, J. A. Thomson, and P. W. Andrews.** 2004. Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. Nat Biotechnol **22:**53-4.

13. **Evans, M.** 1981. Origin of Mouse Embryonal Carcinoma-Cells and the Possibility of Their Direct Isolation into Tissue-Culture. Journal of Reproduction and Fertility **62:**625-&.

14. **Florens, L., M. P. Washburn, J. D. Raine, R. M. Anthony, M. Grainger, J. D. Haynes, J. K. Moch, N. Muster, J. B. Sacci, D. L. Tabb, A. A. Witney, D. Wolters, Y. Wu, M. J. Gardner, A. A. Holder, R. E. Sinden, J. R. Yates, and D. J. Carucci.** 2002. A proteomic view of the Plasmodium falciparum life cycle. Nature **419:**520-6.

15. **Giancotti, F. G., and E. Ruoslahti.** 1999. Integrin signaling. Science **285:**1028-32.

16. **Herszfeld, D., E. Wolvetang, E. Langton-Bunker, T. L. Chung, A. A. Filipczyk, S. Houssami, P. Jamshidi, K. Koh, A. L. Laslett, A. Michalska, L.**

Nguyen, B. E. Reubinoff, I. Tellis, J. M. Auerbach, C. J. Ording, L. H. Looijenga, and M. F. Pera. 2006. CD30 is a survival factor and a biomarker for transformed human pluripotent stem cells. Nat Biotechnol **24:**351-7.

17. **Hong, S., U. J. Kang, O. Isacson, and K. S. Kim.** 2008. Neural precursors derived from human embryonic stem cells maintain long-term proliferation without losing the potential to differentiate into all three neural lineages, including dopaminergic neurons. J Neurochem **104:**316-24.

18. **Horton, P., K. J. Park, T. Obayashi, N. Fujita, H. Harada, C. J. Adams-Collier, and K. Nakai.** 2007. WoLF PSORT: protein localization predictor. Nucleic Acids Res **35:**W585-7.

19. **Josic, D., and J. G. Clifton.** 2007. Mammalian plasma membrane proteomics. Proteomics **7:**3010-29.

20. **Kelley, K. M., Y. Oh, S. E. Gargosky, Z. Gucev, T. Matsumoto, V. Hwa, L. Ng, D. M. Simpson, and R. G. Rosenfeld.** 1996. Insulin-like growth factor-binding proteins (IGFBPs) and their regulatory dynamics. Int J Biochem Cell Biol **28:**619-37.

21. **Malergue, F., F. Galland, F. Martin, P. Mansuelle, M. Aurrand-Lions, and P. Naquet.** 1998. A novel immunoglobulin superfamily junctional molecule expressed by antigen presenting cells, endothelial cells and platelets. Mol Immunol **35:**1111-9.

22. **Martin, G. R.** 1981. Isolation of a Pluripotent Cell-Line from Early Mouse Embryos Cultured in Medium Conditioned by Teratocarcinoma Stem-Cells.

Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences **78:**7634-7638.

23. **Melchior, K., J. Weiss, H. Zaehres, Y. M. Kim, C. Lutzko, N. Roosta, J. Hescheler, and M. Muschen.** 2008. The WNT receptor FZD7 contributes to self-renewal signaling of human embryonic stem cells. Biol Chem **389:**897-903.

24. **Miguet, L., K. Pacaud, C. Felden, B. Hugel, M. C. Martinez, J. M. Freyssinet, R. Herbrecht, N. Potier, A. van Dorsselaer, and L. Mauvieux.** 2006. Proteomic analysis of malignant lymphocyte membrane microparticles using double ionization coverage optimization. Proteomics **6:**153-71.

25. **Mitalipova, M., J. Calhoun, S. J. Shin, D. Wininger, T. Schulz, S. Noggle, A. Venable, I. Lyons, A. Robins, and S. Stice.** 2003. Human embryonic stem cell lines derived from discarded embryos. Stem Cells **21:**521-526.

26. **Mitalipova, M. M., R. R. Rao, D. M. Hoyer, J. A. Johnson, L. F. Meisner, K. L. Jones, S. Dalton, and S. L. Stice.** 2005. Preserving the genetic integrity of human embryonic stem cells. Nat Biotechnol **23:**19-20.

27. **Palmeri, D., A. van Zante, C. C. Huang, S. Hemmerich, and S. D. Rosen.** 2000. Vascular endothelial junction-associated molecule, a novel member of the immunoglobulin superfamily, is localized to intercellular boundaries of endothelial cells. J Biol Chem **275:**19139-45.

28. **Pedrioli, P. G., J. K. Eng, R. Hubley, M. Vogelzang, E. W. Deutsch, B. Raught, B. Pratt, E. Nilsson, R. H. Angeletti, R. Apweiler, K. Cheung, C. E. Costello, H. Hermjakob, S. Huang, R. K. Julian, E. Kapp, M. E. McComb, S. G. Oliver, G. Omenn, N. W. Paton, R. Simpson, R. Smith, C. F. Taylor, W.**

Zhu, and R. Aebersold. 2004. A common open representation of mass spectrometry data and its application to proteomics research. Nat Biotechnol **22:**1459-66.

29. **Peh, G. S., R. Lang, M. Pera, and S. Hawes.** 2008. CD133 expression by neural progenitors derived from human embryonic stem cells and its use for their prospective isolation. Stem Cells Dev.

30. **Pelton, T. A., M. D. Bettess, J. Lake, J. Rathjen, and P. D. Rathjen.** 1998. Developmental complexity of early mammalian pluripotent cell populations in vivo and in vitro. Reprod Fertil Dev **10:**535-49.

31. **Rao, R. R., A. V. Johnson, and S. L. Stice.** 2007. Cell surface markers in human embryonic stem cells. Methods Mol Biol **407:**51-61.

32. **Rao, R. R., and S. L. Stice.** 2004. Gene expression profiling of embryonic stem cells leads to greater understanding of pluripotency and early developmental events. Biol Reprod **71:**1772-8.

33. **Seyfried, N. T., L. C. Huysentruyt, J. A. Atwood, 3rd, Q. Xia, T. N. Seyfried, and R. Orlando.** 2008. Up-regulation of NG2 proteoglycan and interferon-induced transmembrane proteins 1 and 3 in mouse astrocytoma: a membrane proteomics approach. Cancer Lett **263:**243-52.

34. **Shevinsky, L. H., B. B. Knowles, I. Damjanov, and D. Solter.** 1982. Monoclonal antibody to murine embryos defines a stage-specific embryonic antigen expressed on mouse embryos and human teratocarcinoma cells. Cell **30:**697-705.

35. **Shimazaki, T., T. Shingo, and S. Weiss.** 2001. The ciliary neurotrophic factor/leukemia inhibitory factor/gp130 receptor complex operates in the maintenance of mammalian forebrain neural stem cells. J Neurosci **21:**7642-53.

36. **Shin, S., S. Dalton, and S. L. Stice.** 2005. Human motor neuron differentiation from human embryonic stem cells. Stem Cells and Development **14:**266-269.

37. **Shin, S., M. Mitalipova, S. Noggle, D. Tibbitts, A. Venable, R. Rao, and S. L. Stice.** 2006. Long-Term Proliferation of Human Embryonic Stem Cell-Derived Neuroepithelial Cells Using Defined Adherent Culture Conditions. Stem Cells **24:**125-138.

38. **Simpson, R. J., L. M. Connolly, J. S. Eddes, J. J. Pereira, R. L. Moritz, and G. E. Reid.** 2000. Proteomic analysis of the human colon carcinoma cell line (LIM 1215): development of a membrane protein database. Electrophoresis **21:**1707-32.

39. **Sleeman, M. W., K. D. Anderson, P. D. Lambert, G. D. Yancopoulos, and S. J. Wiegand.** 2000. The ciliary neurotrophic factor and its receptor, CNTFR alpha. Pharm Acta Helv **74:**265-72.

40. **Sugano, Y., M. Takeuchi, A. Hirata, H. Matsushita, T. Kitamura, M. Tanaka, and A. Miyajima.** 2008. Junctional adhesion molecule-A, JAM-A, is a novel cell-surface marker for long-term repopulating hematopoietic stem cells. Blood **111:**1167-72.

41. **Thomson, J. A., J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall, and J. M. Jones.** 1998. Embryonic stem cell lines derived from human blastocysts. Science **282:**1145-1147.

42. **Venable, A., M. Mitalipova, I. Lyons, K. Jones, S. Shin, M. Pierce, and S. Stice.** 2005. Lectin binding profiles of SSEA-4 enriched, pluripotent human embryonic stem cell surfaces. BMC Dev Biol **5:**15.

43. **Vercoutter-Edouart, A. S., M. C. Slomianny, O. Dekeyzer-Beseme, J. F. Haeuw, and J. C. Michalski.** 2008. Glycoproteomics and glycomics investigation of membrane N-glycosylproteins from human colon carcinoma cells. Proteomics **8:**3236-56.

44. **Wallingford, J. B., and R. Habas.** 2005. The developmental biology of Dishevelled: an enigmatic protein governing cell fate and cell polarity. Development **132:**4421-36.

45. **Weiss, S., B. A. Reynolds, A. L. Vescovi, C. Morshead, C. G. Craig, and D. van der Kooy.** 1996. Is there a neural stem cell in the mammalian forebrain? Trends Neurosci **19:**387-93.

46. **Wu, C. C., M. J. MacCoss, K. E. Howell, and J. R. Yates, 3rd.** 2003. A method for the comprehensive proteomic analysis of membrane proteins. Nat Biotechnol **21:**532-8.

47. **Yu, J., and J. A. Thomson.** 2008. Pluripotent stem cell lines. Genes Dev **22:**1987-97.

48. **Zeng, X., J. Chen, Y. Liu, Y. Luo, T. C. Schulz, A. J. Robins, M. S. Rao, and W. J. Freed.** 2004. BG01V: a variant human embryonic stem cell line which exhibits rapid growth after passaging and reliable dopaminergic differentiation. Restor Neurol Neurosci **22:**421-8.

Figure 4.1. Enrichment of Membrane Proteins From Duplicate Analysis of Each Population. The programs Wolf PSort (predicts subcellular localization) or TMHMM (predicts transmembrane spanning regions) were used to calculate membrane status of the identified proteins. Proteins were annotated as soluble if either Wolf PSort or TMHMM did not predict the proteins as plasma membrane or transmembrane spanning. WA09-derived neural progrenitors had the highest amount of enrichment for membrane proteins (64.1%, 471 predicted membrane proteins). Embryonic stem cell populations showed similar enrichment for plasma membrane or transmembrane spanning regions (WA09, 244 membrane proteins, 32.7% enrichment; BGO1, 508 membrane proteins, 40.3% enrichment; abBG02, 421 membrane proteins, 37.0% enrichment). For all four cell lines, the insoluble fraction contained a high amount transmembrane spanning regions, A total of 231 transmembrane proteins (94.7%) were identified from the WA09 line, 469 (92.3%) from the BGO1, 400 (95.0%) from the abBGO2 line, and 441 (93.6%) from the hNP line. Transmembrane spanning proteins have high potential as candidates for surface marker of cell type.

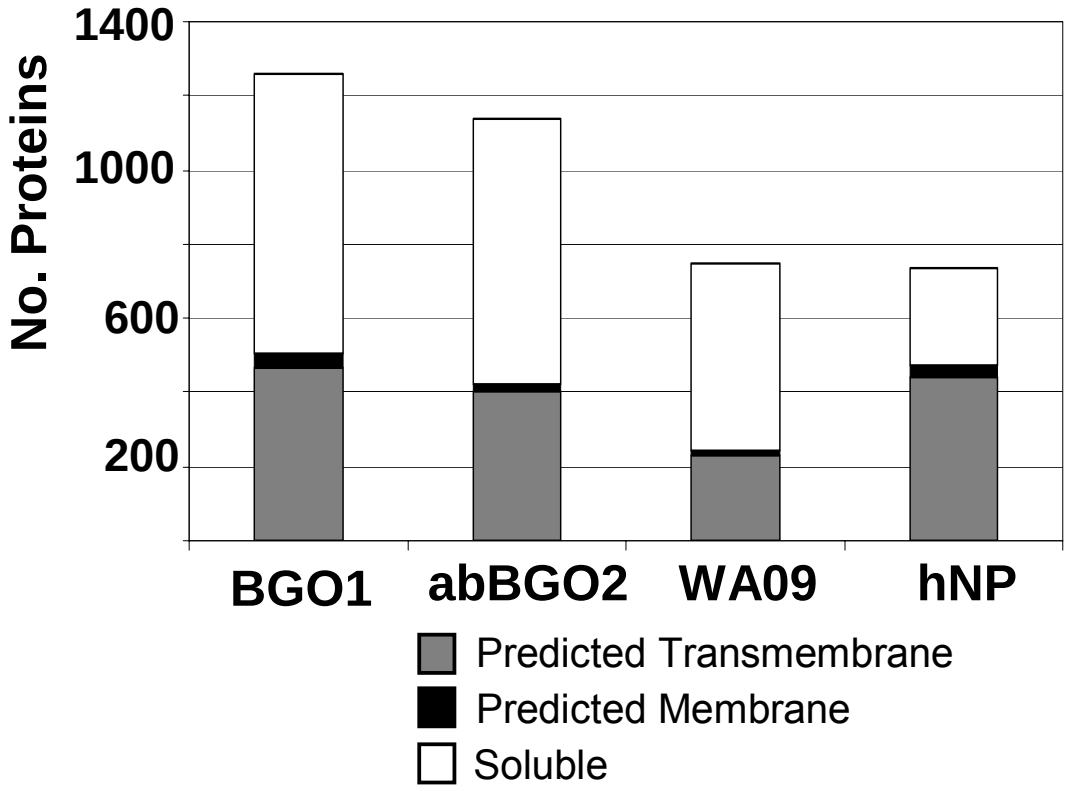

1400
1000
600
200
No. Proteins
BGO1
abBGO2
WA09
hNP
Predicted Transmembrane
Predicted Membrane
Soluble

Figure 4.2. Venn Diagram Demonstrating Membrane Protein Differences and Overlap Between Cell Types. Proteins annotated as localized to the plasma membrane or having transmembrane spanning regions were compared across all cell populations. A total of 19% (149) proteins were observed in all four cell lines. The WA09-derived neural progenitors had the highest number of unique proteins (17.7%, 138 proteins). Each embryonic stem cell line had a smaller population of proteins unique to each population (WA09, 7.19%, 56 proteins; BGO1 10.4%, 81 proteins, abBGO2, 7.83%, 61 proteins).

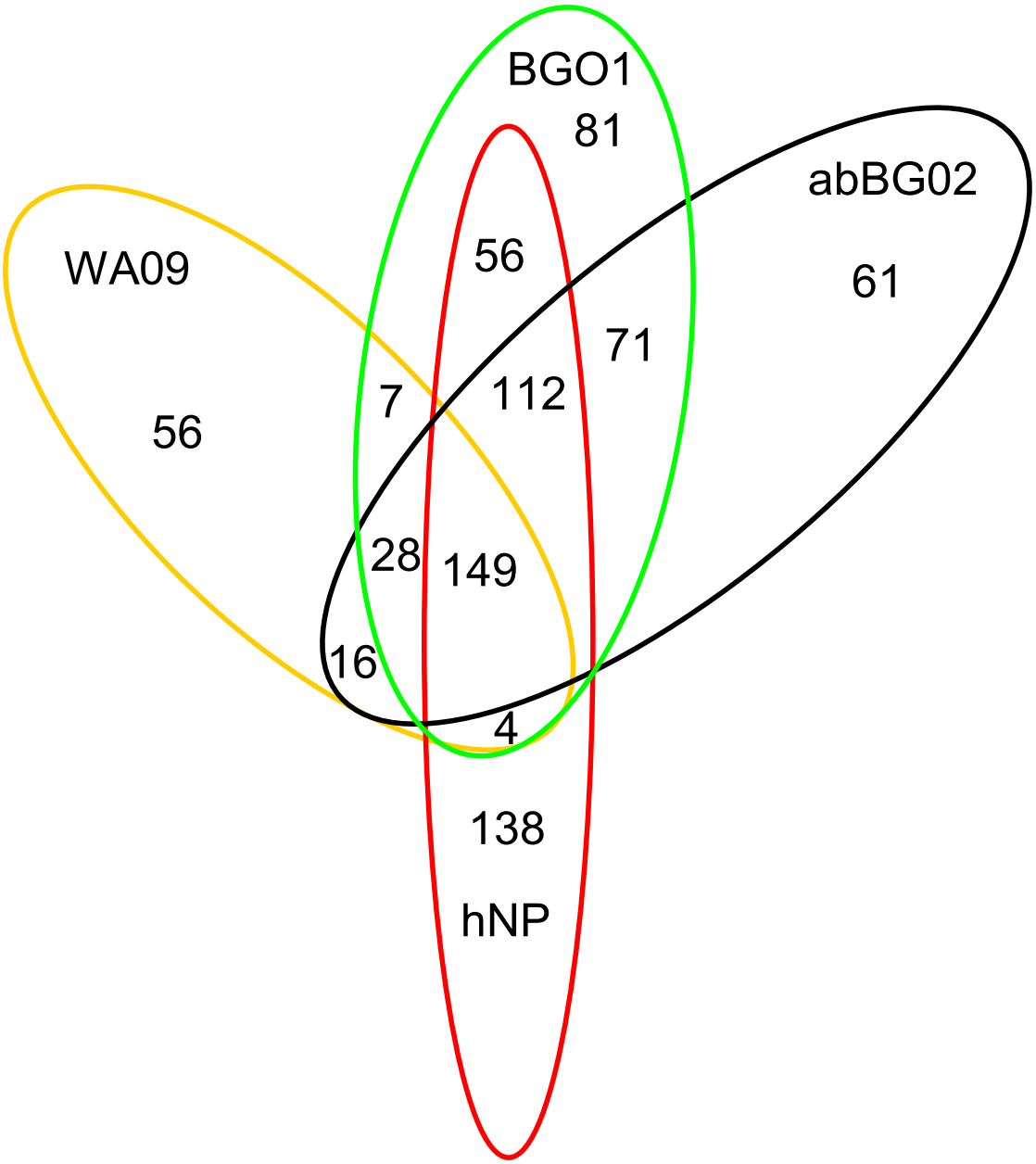

BGO1
81
abBG02
61
WA09
56
56
71
7
112
28
149
16
4
138
hNP

Figure 4.3. **Flow Cytometry Validation of Proteomic Analysis for Potential Biomakers Exhibit Differential Expression between WA09 hESC and WA09-derived hNP cells.** **(A)** WA09 exhibited differential expression of Dysferlin and Frizzled 7 with a higher percentage of positive cells relative to hNPs and hMSCs. However, hNPs demonstrated a higher percentage of CNTFRα positive cells. JAM-1 did not display preferential expression between analyzed cell types. Gray boxes within the table depict cell surface markers that were predicted through the proteomic analysis to be unique to each cell line. **(B)** Flow cytometry histograms were gated on n = 10,000. Abbreviations: hESC – human embryonic stem cell; hNP – human neural progenitor cell stem cell; hMESC – human mesenchymal embryonic stem cell; JAM-1 – junctional adhesion molecule 1; CNTFRα – ciliary neurotrophic factor receptor alpha

A

	WA09	WA09-derived hNP cells	WA09-derived hMESC
Dysferlin	79.2%	26.4%	1.8%
Frizzled 7	79.5%	8.8%	1.2%
JAM-1	9.1%	23.9%	3.8%
CNTFRα	1.6%	58.3%	1.0%

B

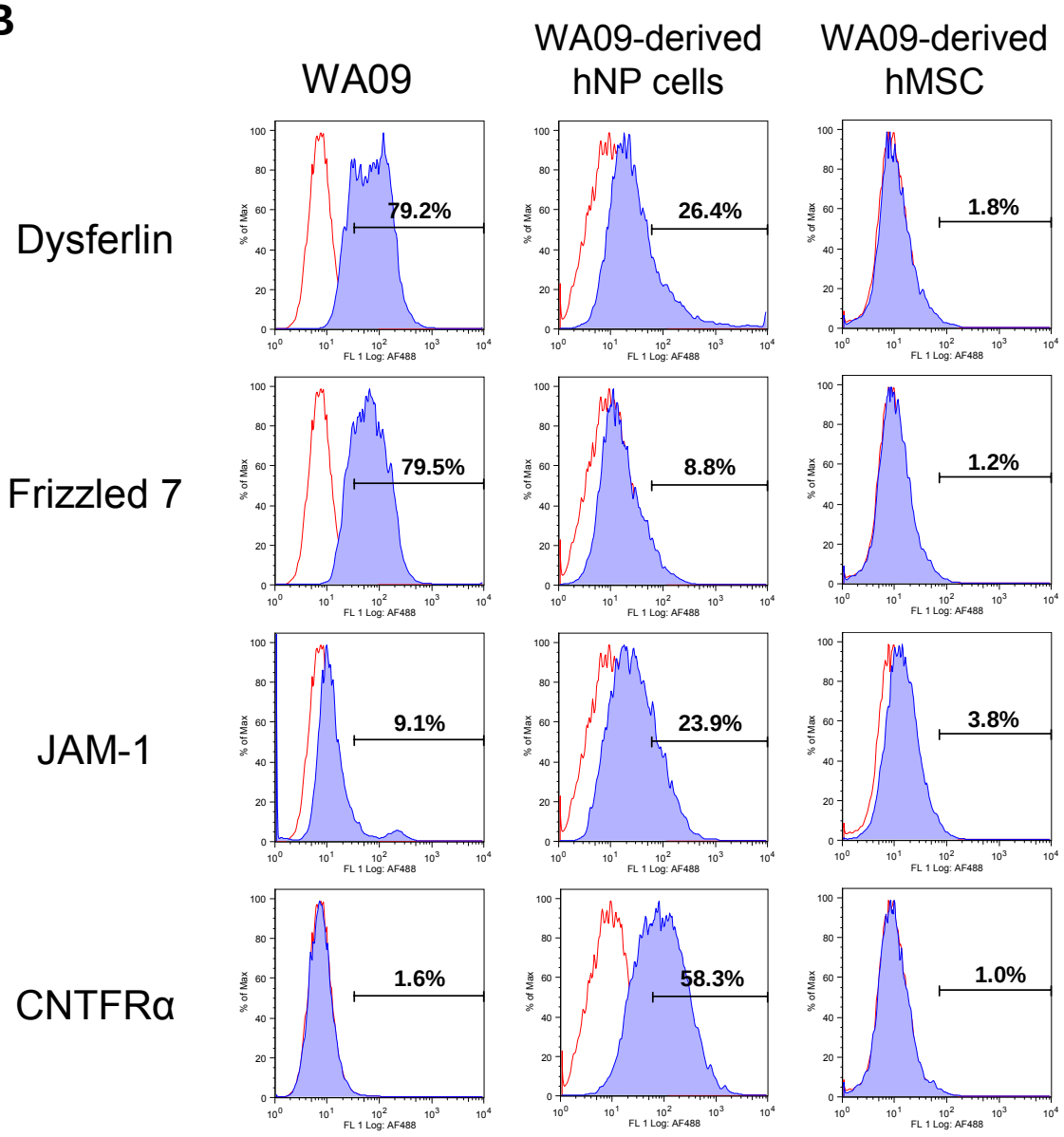


Figure 4.4. Flow Cytometry Validation of Proteomic Analysis for Potential Biomakers Demonstrate Preferential Expression of JAM-1 for BG01 and BG02 hESC lines. **(A)** Table represents percentage of positive cells for biomarkers examined. BG01 and karoytipically abnormal and normal BG02 hESC lines displayed a substantial (>94%) number of positive cells for JAM-1. Dysferlin, Frizzled 7, and CNTFRα did not display preferential expression between BG01 and BG02 hESC lines analyzed. Gray boxes within the table depict cell surface markers that were predicted through the proteomic analysis to be unique to each cell line. Percentages account for a 1% false positive rate. **(B)** Flow cytometry histograms were gated on n = 10,000. Abbreviations: JAM-1 – junctional adhesion molecule 1; CNTFRα – ciliary neurotrophic factor receptor alpha

A

	BG01	BG02	AbBG02
Dysferlin	12.3%	32.1%	27.6%
Frizzled 7	0.6%	2.4%	23.9%
JAM-1	97.1%	94.8%	97.7%
CNTFRα	9.7%	6.1%	7.0%

B

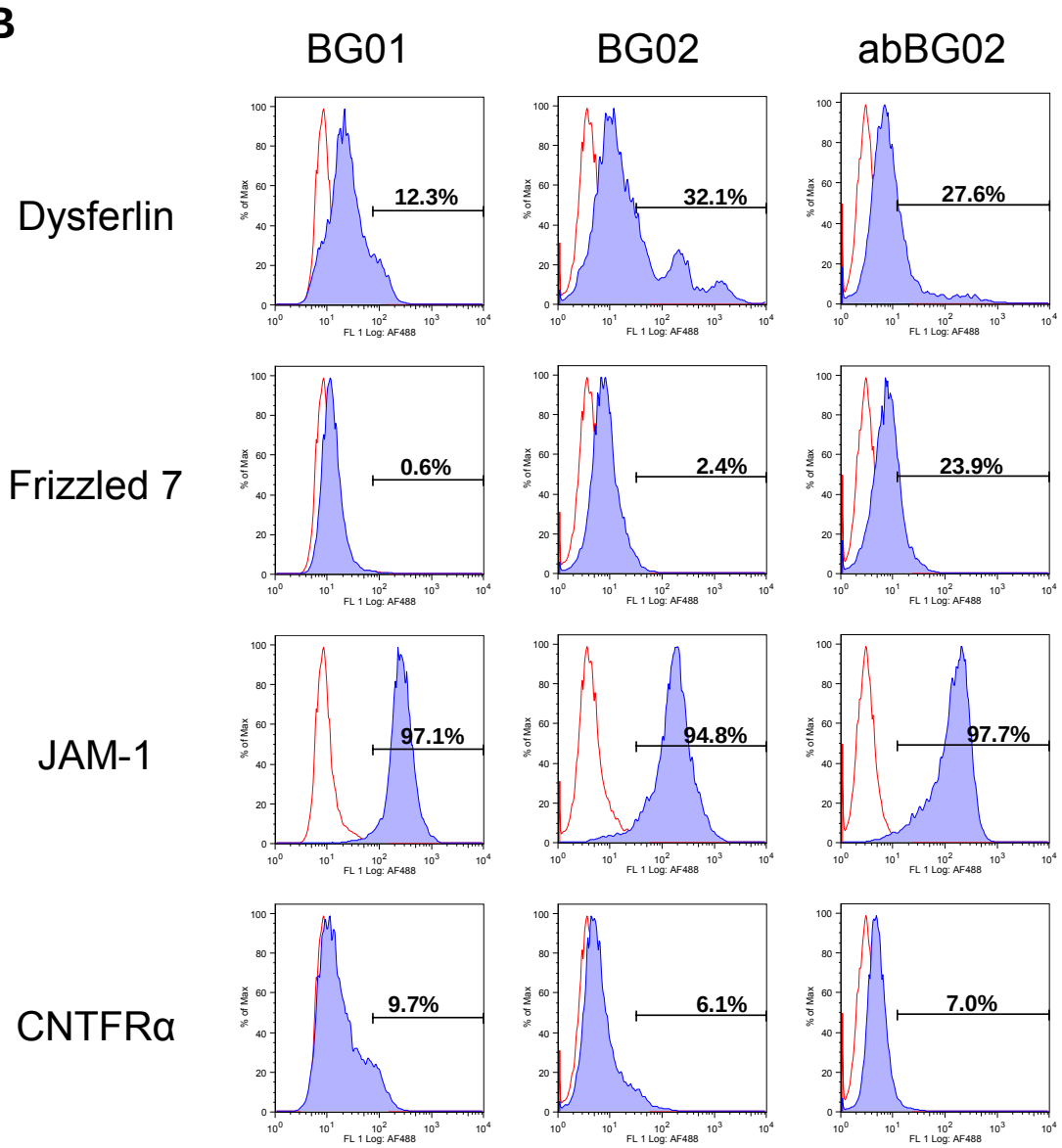


Table 4.1. Top Surface Candidate Markers Identified By the Proteomics Study

Accession	Description	Function[a]	Gene Location[a]	Mascot Score[b]	Protein Probability[c]	No. Peps[d]	Sequence Coverage	Predicted TMSR/PP [e]
Cell Type: BGO1								
NP_003498.1	Frizzled 7	WNT signaling	2q33.1	227.5	0.984	5	2.7	7/29
NP_002664.2	Plexin B1	Axon guidance	3p21.31	92.5	0.771	3	1.78	-/9
NP_002578.2	Protocadherin 1, isoform 1 precursor	Neural cell adhesion	5q31.3	75.6	0.449	2	3.11	1/19
NP_004326.1	Bone marrow stromal cell antigen 2	B-cell growth and development	19p13.2	56.7	0.931	1	8.89	1/29
NP_08915.1	Claudin 10 isoform b	Cell adhesion	13q31	56.0	0.937	1	6.14	4/31
Cell Type: WA09								
NP_003485.1	Dysferlin	Membrane fusion	2p13.3	847.4	1.000	17	11.15	1/-
NP_001926.2	Dipeptidylpeptidase	T-cell activation	2q24.3	702.5	1.000	15	23.37	1/-
NP_932766.1	Aquaporin	Water channel protein	7p14	231.7	1.000	2	13.76	6/31
NP_002297.1	Galectin 3	Binds to T cell receptors	14q22.3	190.5	1.000	3	15.2	MC
NP_000879.2	Integrin, beta 6	Focal adhesion	2q24.2	298.7	1.000	5	9.52	1/-
NP_000349.1	Transforming growth factor, β-induced, 68kDa	Cell-collagen interactions	5q31	363.8	1.000	9	14.9.93	1/-
Cell Type: abBGO2								
NP_653086.1	F11 receptor precursor	Junctional adhesion	1q21.2	620.9	0.909	12	41.14	2/7
NP_059992.3	Apolipoprotein E receptor (LRP8)	APOE Receptor	1p32.3	123.0	0.333	2	4.29	2/5.5
NP_004702.2	Synaptogyrin 1 isoform 1a	Synaptic plasticity	22q13.1	108.9	0.656	2	11.2	4/32
NP_002635.2	Polymeric immunoglobulin receptor	Immunoglobin transport	1q31	93.9	0.997	1	2.23	1/11
NP_005630.1	Synaptotagmin	Neuro-calcium binder	12cen-q21	90.2	0.596	2	6.16	MC
NP_620409.1	Vang-like 1	WNT signaling	1p13.1	84.9	0.993	1	3.24	4/27
NP_006496.2	Poliovirus receptor	Cell adhesion/	19q13.2	76.8	0.653	2	6.24	1/18
Cell Type: hNP								
NP_036434.1	Latrophilin 2	Exocytosis regulatory	1p13.1	671.3	1.000	12	12.12	7/32

	precursor							
NP_004654.1	Ras-related protein Rab-11A	Membrane trafficking	15q22.31	387.3	0.905	9	36.57	MC
NP_000192.1	Intercellular adhesion molecule 1 precursor	Cell adhesion	19p13.3	305.2	1.000	5	14.85	1/10
NP_671693.1	Ciliary Neurotrophic factor receptor preprotein	Jak-STAT Signaling	9p13	237.1	0.922	4	24.46	0/13
NP_000866.1	Insulin-like growth factor 1 receptor precursor	Focal adhesion	15q26.3	233.8	0.984	6	5.34	1/10
NP_001026872.2	Semaphorin 5B isoform 1	Axon guidance	3q21.1	180.6	0.998	4	4.87	1/-
NP_001768.1	CD47 precursor	ECM-Cell adhesion	3q13.1	107.3	0.656	3	9.60	6/32
NP_000255.1	Patched	Hedgehog signaling	9q22.1	98.2	0.999	1	1.04	12/32

-a Pathway annotation and protein coding gene location are derived from KEGG Genes Database, species homo sapiens. Function refers to the top listed function.
-b Score is the total protein Mascot Score. A Mascot score of >51 indicates a false discovery rate of less than one percent.
-c Protein probability is the probability that the protein is present in the dataset, based proportion of peptides unique to that protein. A protein probability of 1.000 indicates that the protein is uniquely identified in the dataset.
- No. Pep refers to the number of tryptic peptides clustered to each protein.
- Predicted transmembrane spanning regions (TMSR) using TMHMM software and PP, predicted plasma (PP) membrane using Wolf Psort and reported as percent localization to that subcellular localization.
MC – Manual curation from the Kegg Genes and Entrez Gene databases.

Chapter 5

Conclusion

The field of human embryonic stem cell (hESC) biology has rapidly advanced since its inception in 1998 (7). In vitro studies have shown that hESCs have limitless potential for clinical applications (e.g. regenerative medicine and cell therapy), while differentiated, hESCs prove promising for drug discovery, toxicology screens and serve as a model for studying basic human development. Therefore, identifying unique [bio]markers has the potential to better characterize and classify hESC populations and their derivatives with greater culturing and differentiation potential.

Herein, I define the use of the molecule (glial fibillary acidic protein [GFAP]) as a developmental biomarker to identify and isolate the glial lineage, more specifically astrocytes. This definition is within the context established by the National Institute of Health which defines a biomarker as "…a characteristic that is objectively measured and evaluated as an indicator of normal biological processes…" (1). The goal of this dissertation was to characterize the temporal expression of glial fibrillary acidic protein (GFAP) utilizing a fluorescent reporter system. This characterization will potentially lead to the eventual isolation of a glial-like cell type from hESC-derived neural progenitor (hNP) cells.

The objective for the first study was to develop a fluorescent reporter construct utilizing the glial-specific GFAP promoter that would stably integrate into hNP cells utilizing lentiviral technology. The pFUGW lentiviral packgaging vector was modified to

contain the hGFAP promoter (2.2 kB) driving the conditional expression of EGFP or tomato Red (tmRed). Furthermore, the addition of neomycin resistance under constitutive expression from the SV40 promoter enabled the positive selection of a relatively pure population of transduced cells. I was able demonstrate the selective expression of GFAP:GFP within transduced human astrocytoma (ATCC # CRL-1718) cells with no expression in transduced undifferentiated hNP cells. The results suggest the ability to selectively trace the appearance of glial-like cells in vitro and monitor their potential differentiation devoid of a "leaky" expression system leading to false-positives. In addition, the low percentage of silenced expression of EGFP after eight weeks in culture suggest these cells may offer an efficient tool for in vivo implantation studies. Cells capable of maintained fluorescent expression such as these provide a significant tool to study differentiation, neural development and may ultimately lead to cell therapies.

The successful conditional expression of the GFAP:GFP lentiviral construct allowed for the characterization of a $GFAP^+$ cell type. I examined the temporal expression of GFAP as a biomarker to better elucidate the differentiation process of human neural progenitor (hNP) cells towards an astrocyte-like cell population. I determined that hNP cells in –bFGF differentiation media supplemented with FBS will permit the increased expression of GFAP by DAY21 (22.3 +/- 11.9 fold change) and substantial up regulation continued to DAY49 (663 +/- 100 fold change) relative to the DAY0 control. In the absence of FBS, an increase in GFAP expression was not observed until DAY49 (1.60 +/- 0.44 relative fold change). After the establishment of $GFAP^+$ cells upon differentiation, I transduced hNP cells with the lentiviral GFAP:GFP ::

SV40:NeoR construct and demonstrated the selective expression of GFP to only GFAP expressing cells. Additional experiments assessed whether the temporal onset of GFAP expression could be decreased or its expression enhanced. An increase of FBS concentration (from 5% to 10%) showed a significant ($p<0.05$) difference at DAY21 compared to 5% FBS –bFGF differentiation cultures, which were down regulated regardless of the presence of ACM. These results suggest that 10% FBS alone could shorten the length of time needed for the onset of GFAP expression under current culture system. Interestingly, DAY35 treatments in 5% and 10% FBS showed significant increases in GFAP expression relative to DAY0, yet the addition of ACM to the –bFGF differentiated media increased the average relative GFAP expression at DAY35 above that of FBS alone; However, there was larger viability at the DAY35 time point.

Enhancement of glial differentiation through the withdrawal of B27 media supplement from the differentiation medium was investigated. Our results suggest that the partial removal of B27 from standard culturing conditions may affect downstream differentiation efforts. Intriguingly, the complete removal of B27 is detrimental to cell proliferation and 1 X N2 replacement is not sufficient to promote cell survival. hNP cells were acclimated to modified concentrations of B27 for a period of two months and then differentiated in –bFGF differentiation media supplemented with 5% FBS. Quantitative real (RT)-PCR performed on 0.125 X B27 and 0.5 X B27 + 1 X N2 cells showed significant ($p<0.05$) up regulation at DAY28 relative to DAY0 (undifferentiated hNP cells) and a day effect when compared to DAY21 with a significant increase in GFAP

expression. Thus, the directed differentiation and isolation of human embryonic stem cell (hESC)-derived astrocytes still remains elusive, however the results are promising.

In the last study, I determined that membrane fractionation and liquid-chromatography tandem mass spectrometry (LC-MS/MS) techniques could be used to develop novel biomarkers capable of distinguishing between different hESC lines and their derivatives, hNP cells, thereby overcoming the vast heterogeneity found within differentiating hESC culture systems, which is vital to further study and clinical application (6). The screening for distinct cell surface markers enabled the establishment of a plasma membrane profile for each cell line, which in turn allowed for the identification of hESC line specific biomarkers. I was able to distinguish the WA09 hESCs from the BresaGen derived cell lines utilizing flow cytometry with Dysferlin (DYSF) and junctional adhesion molecule 1 (JAM-1, F11R, CD321), respectively. Furthermore, through the exploitation of unique protein signature patterns displayed at various stages of differentiation or maturation, I was able to delineate [neural] fate decisions. This was achieved by uniquely distinguishing the parental hESC population from differentiated early derivatives. Dysferlin (DYSF) was specific for the WA09 hESC line, while ciliary neurotrophic factor receptor alpha (CNTFRα) was able to delineate the derived hNP cell population. Therefore, I provided evidence that uniquely identifying cell surface markers based on high-throughput proteomics can be achieved where possible cell discrimination can be employed to produce homogeneous populations.

Future Studies

Even with the significant accomplishments documented within this dissertation to characterize the temporal expression of GFAP during the directed differentiation of hNP

cells, several questions still remain. As noted above, the presence of FBS within the differentiation media up regulates GFAP expression of hNP cells towards a glial-like cell fate. Which signaling molecules present in the FBS stimulate or permit this differentiation not observed when cells are cultured in its absence? Furthermore, differentiation of hNP cells generates a highly heterogeneous population of varying lineages and degrees of maturation. How could one isolate the $GFAP^+$ glial-like cell population from the remaining culture during differentiation to yield a more homogenous population?

Fetal bovine serum is a vast cocktail of macromolecules, carrier proteins, carbohydrates, amino acids, supplements and trace elements, hormones and growth factors (4, 8). Several of the major known biological components important for cell growth are albumin, antichymotrpsin, apolipoproteins, biotin, and growth supporting factors (2). In 2006, a proteomic study identified a list of components found in fetal bovine serum (8). Interestingly, the researchers identified a varying number of proteins and at different total protein concentrations between the three sera examined. Two of the three sera were obtained from the same company, yet different lots. Thus, fetal bovine serum remains ill-defined with varying biological components between lots and company suppliers.

To identify which biological components within FBS stimulate the onset of GFAP expression (gliogenesis), I would apply the same basic strategy as the proteomic study. FBS would be filtered through membranes of various molecular weight cut-offs. hNP cells would be differentiated using –bFGF differentiation medium supplemented with unique fractionated sera. Immunocytochemistry and real time-PCR would be performed

to identify which fractionations yielded $GFAP^+$ cells. SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) of the successful fractions would allow for preliminary assessment of purity and total proteins present. If FBS fraction is still too complex for LC-MS/MS analysis, FPLC (Fast Protein Liquid Chromatography) utilizing size exclusion and other protein purification columns would be employed to further fractionate the sample for analysis. After protein identification of particular samples has occurred, hNP cells can then be differentiated in –bFGF differentiation medium spiked with identified proteins and accessed for the onset and continual expression of GFAP relative to the control (complete FBS supplemented differentiation media).

Within the central nervous system, astrocytes are found in both white and gray matter despite variations in morphology such as distinctive “fibrous” or long unbranched cellular processes and “protoplasmic” that exhibit short and highly branched cellular processes (5). This diverse morphology led researchers to characterize astrocytes by their high content of intermediate filaments, particularly glial fibrillary acidic protein (GFAP) (5). Therefore, standard isolation techniques (e.g. flow cytometry and fluorescence-activated cell sorting [FACS]) that sort based on cell shape and size determinants (3) would be ineffective. Moreover, the branched nature the of $GFAP^+$ glial-like cell population observed in these studies lends itself to cell shearing and breaking off of the branched membranes from the cell body, causing a significant decrease in post-sorting viability (unpublished data). This may indicate that the isolation process may need to coincide with the onset of GFAP expression rather than after maturation of these glial-like cells.

The issue with screening against the other heterogeneous cells types observed upon differentiation in the presence of FBS in vitro is that neurogenesis precedes gliogenesis. If all neuronal cell types (beta III Tubulin positive) are removed before gliogenesis, glia differentiation may not occur due to important neural glia cell-cell signaling interactions within the culture. Results from my studies indicate that beta III tubulin expression is increased within 14 days of differentiation and remains constant afterward. However, the onset of GFAP transcriptional expression occurs by DAY21 and substantially increases at subsequent time points (e.g. >633 in relative fold change at DAY49). Building on the pFUGW GFAP:GFP :: SV40:NeoR lentiviral system developed within this dissertation, I would modify the reporter system to include another mammalian antibiotic resistance gene fused to GFP to select against non-GFAP expressing cells. In the proposed case, neomycin would still selectively screen for transduced hNP cells before differentiation. Upon glial-like cell differentiation, the onset of GFAP expression would drive a mammalian selectable resistance gene (e.g. Blasticidin, Hygromycin B, Mycophenolic Acid or Zeocin) florescent reporter fusion allowing for positive selection of GFAP expressing cells. The fusion protein will have to be assessed to determine if there is a loss of antibiotic resistance as compared to the selectable gene alone. Thus, I will have enhanced the existing system to negatively select against all non-GFAP expressing cells with the hope of establishing a more homogenous population needed for further basic research and clinical studies.

In summary, characterization of an astrocyte-like cell type from hESC-derived neural progenitor cells has been achieved and will, hopefully, lead to a more complete in vitro model for neural development. In addition, the application of a high-throughput

proteomic approach illustrated the distinction between different hESC lines and the further delineation from their derived cell types (e.g. hNP cells). This proof of concept will enhance the field's ability to acquire a more homogenous population needed for clinical application. It is hoped that the results from these proposed studies will help to further the understanding of gliogenesis and advance potential neural stem cell-based therapies.

References

1. 2001. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther **69:**89-95.
2. **Baker, H., B. DeAngelis, and O. Frank.** 1988. Vitamins and other metabolites in various sera commonly used for cell culturing. Experientia **44:**1007-10.
3. **Brown, M., and C. Wittwer.** 2000. Flow cytometry: principles and clinical applications in hematology. Clin Chem **46:**1221-9.
4. **Gstraunthaler, G.** 2003. Alternatives to the use of fetal bovine serum: serum-free cell culture. ALTEX **20:**275-81.
5. **Kessaris, N., N. Pringle, and W. D. Richardson.** 2008. Specification of CNS glia from neural stem cells in the embryonic neuroepithelium. Philos Trans R Soc Lond B Biol Sci **363:**71-85.
6. **Peh, G. S., R. Lang, M. Pera, and S. Hawes.** 2008. CD133 expression by neural progenitors derived from human embryonic stem cells and its use for their prospective isolation. Stem Cells Dev.
7. **Thomson, J. A., J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall, and J. M. Jones.** 1998. Embryonic stem cell lines derived from human blastocysts. Science **282:**1145-1147.
8. **Zheng, X., H. Baker, W. S. Hancock, F. Fawaz, M. McCaman, and E. Pungor, Jr.** 2006. Proteomic analysis for the assessment of different lots of fetal bovine serum as a raw material for cell culture. Part IV. Application of proteomics to the manufacture of biological drugs. Biotechnol Prog **22:**1294-300.

APPENDICES

CSIRO PUBLISHING

www.publish.csiro.au/journals/rfd *Reproduction, Fertility and Development*, 2006, **18**, 839–846

Human embryonic stem cells: challenges and opportunities

Steven L. Stice[A,C], *Nolan L. Boyd*[A], *Sujoy K. Dhara*[A], *Brian A. Gerwe*[A], *David W. Machacek*[A] *and Soojung Shin*[B]

[A]Regenerative Bioscience Center, ADS Department, University of Georgia, Athens, GA 30606, USA.
[B]Invitrogen Corporation, Carlsbad, CA 92008, USA.
[C]Corresponding author. Email: sstice@arches.uga.edu

Abstract. Human and non-human primate embryonic stem (ES) cells are invaluable resources for developmental studies, pharmaceutical research and a better understanding of human disease and replacement therapies. In 1998, subsequent to the establishment of the first monkey ES cell line in 1995, the first human ES cell line was developed. Later, three of the National Institute of Health (NIH) lines (BG01, BG02 and BG03) were derived from embryos that would have been discarded because of their poor quality. A major challenge to research in this area is maintaining the unique characteristics and a normal karyotype in the NIH-registered human ES cell lines. A normal karyotype can be maintained under certain culture conditions. In addition, a major goal in stem cell research is to direct ES cells towards a limited cell fate, with research progressing towards the derivation of a variety of cell types. We and others have built on findings in vertebrate (frog, chicken and mouse) neural development and from mouse ES cell research to derive neural stem cells from human ES cells. We have directed these derived human neural stem cells to differentiate into motoneurons using a combination of developmental cues (growth factors) that are spatially and temporally defined. These and other human ES cell derivatives will be used to screen new compounds and develop innovative cell therapies for degenerative diseases.

Extra keywords: culture, karyotype, neural differentiation.

Human and non-human primate embryonic stem cells

In this review, we will describe some of the unique qualities of non-human primate and human embryonic stem cells, the challenges in maintaining these unique phenotypes and the opportunities associated with differentiation towards neural stem cells and specific phenotypes such as motoneurons. The advances are based on previous discoveries in developmental biology.

Stem cell derivation

The isolation of the first human embryonic stem (hES) cells (Thomson *et al*. 1998), as well as primate embryonic stem cells from rhesus monkey (Thomson *et al*. 1995) and marmoset (Thomson *et al*. 1996), was remarkably similar to the mouse embryonic stem (mES) cell protocol/media used 17 years previously. Within the same year that hES cells were isolated, another research group derived pluripotent human embryonic germ (hEG) cell lines from fetal gonads 5–9 weeks post fertilisation (Shamblott *et al*. 1998).

Regardless of species, ES cell isolation requires the selection away from, or removal of, the trophectoderm (external cell layer surrounding the inner cell mass (ICM) that gives rise to a portion of the placenta). The ICM is often plated on a mouse embryonic fibroblast feeder cell layer (mEFs) and allowed to grow and expand. This culturing system is maintained by means of continuous passaging and subculture. The doubling time for a hES cell line is between 35 and 40 h (Rosler *et al*. 2004).

The human embryos used to start ES cell derivation were donated by *in vitro* fertilisation (IVF) clinics to the research community, and were either in excess of what was needed for implantation or were not of sufficient quality. However, one can further delineate these excess embryos into two more categories (Fig. 1). The first subgroup is the excess embryos that had formed adequate cell numbers in both the trophoblast and ICM. These higher quality embryos are usually frozen to be thawed later for transfer into the reproductive tract, donated to infertile couples or donated for research. The second subgroup contains excess embryos determined not to be viable enough to either initiate a pregnancy or to freeze for later transfer. This second subgroup of non-viable embryos is usually discarded. We were able to use these embryos to derive three of the most highly investigated embryonic stem cells in the NIH registry (BG01, BG02 and BG03; Mitalipova *et al*. 2003). No publications to date have shown that these hES cell lines differ in their usefulness in regenerative therapies when compared with other hES

 10.1071/RD06113 1031-3613/06/080839

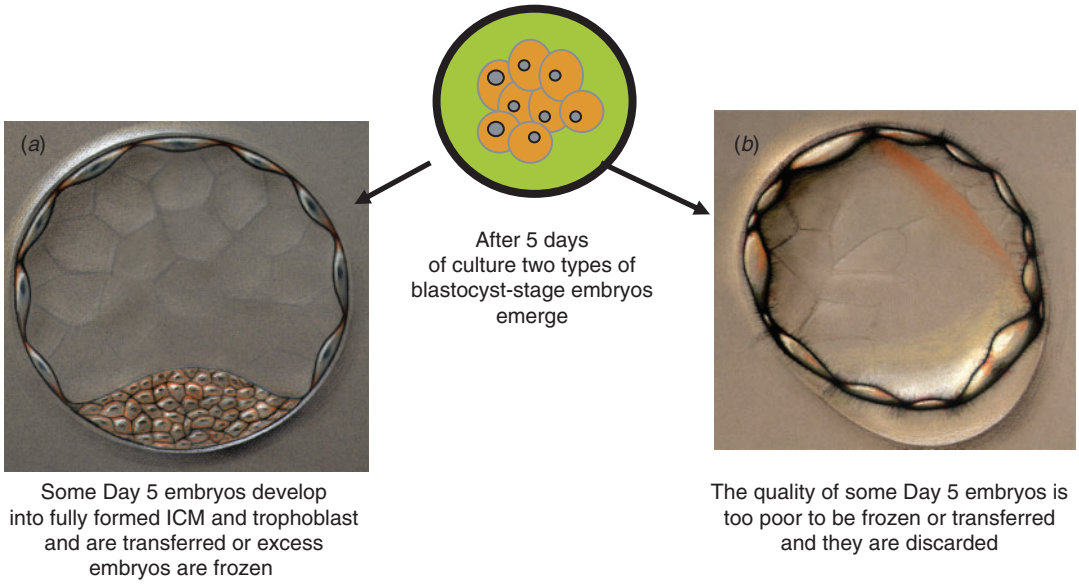


Fig. 1. (*a*) Type of human embryos used to derive embryonic stem cells. (*b*) Poor-quality embryos that are discarded are available and were used to make three of the 21 human embryonic stem cell lines on the National Institute of Health registry. ICM, inner cell mass.

cell lines. For the list of NIH (National Institute of Health)-approved ES cell lines, see the NIH stem cell registry website (http://stemcells.nih.gov/research/registry/eligibilityCriteria.asp, verified September 2006).

Human embryonic stem cell chromosomal stability and gene expression

Variations in chromosomal number could lead to alterations in the genetic expression patterns of certain hES cell lines. Karyotyping most often involves a procedure called G-banding to count chromosomal number and/or chromosomal rearrangements within a cell. In terms of the established hES cell lines, it has been reported that lines can maintain a normal karyotype in continuous culture under certain conditions. For example, hES1–6 cell lines have a normal karyotype (46 chromosomes, XY) during 24–140 passages of continuous culture (Buzzard *et al*. 2004), and H1, H7 and H9 cell lines during passage number 60–100 (Rosler *et al*. 2004). However, reports of abnormal karyotypes similar to those found in embryonic carcinoma (EC) cell lines of trisomy 17 and 21 have recently surfaced in the BG01 and BG02 NIH-registered hES cell lines after bulk hES cell passaging compromised the genetic integrity of cells as quickly as 23–25 passages after manual passaging techniques had ceased (Mitalipova *et al*. 2005).

Aneuploidy has been associated with abnormal gene expression. Theoretically, abnormal gene expression may be a risk factor when used in cell transplant studies. Therefore, abnormal karyotype may up-regulate genes associated with teratoma formation. Our laboratory found that manually passaged cells could maintain a normal karyotype, but when using enzymatic or even non-enzymatic techniques over 20 passages, trisomies at 12, 17, 14, 20 and X in BG01 and BG02 lines eventually developed. Short-term bulk passaging, typical of methods used in other stem cell types including mES cells, was successful in maintaining a normal karyotype from 13 passages and 15 passages for the non-enzymatic method using cell dissociation buffer and collagenase/trypsin, respectively (Mitalipova *et al*. 2005). By quantitative real-time (RT) polymerase chain reaction (PCR), the majority of over 80 genes associated with pluripotency and early differentiation were significantly up-regulated in aneuploid cells (Mitalipova *et al*. 2005). Brimble and colleagues (2004) stated that abnormal karyotypes, especially trisomy 17 and 21, can be acquired through various single-cell disaggregating passage techniques by positively selecting for aneuploidy. They confirmed that manually passaged BG01 and BG02 hES cell lines maintained a long-term normal karyotype (tested only up to 52 passages for a period of 6 months) (Brimble *et al*. 2004). Together with our study, these reports suggest that bulk hES cell passaging will compromise the genetic integrity of cells as quickly as 23–25 passages after manual passaging techniques have ceased.

With all the above issues surrounding each hES stem cell line, a standard protocol for their derivation, culture, maintenance and characterisation is needed to ensure the validity of all lines and their uses. An international initiative named The International Stem Cell Forum (http://www.stemcellforum.org.uk, verified September 2006) was founded in January 2003 comprising 15 countries with the intention of establishing a set of standards for the characterisation and culturing of all the reported hES cell lines.

Table 1. Phenotype marker expression for among-model embryonic stem (ES)-cell lines
Modified table from Pera *et al.* 2000. EC, embryonic carinoma; EG, embryonic germ; ES, embryonic stem; hTERT, telomerase reverse transcriptase; N/A, antibodies do not react with mouse cells. It is unknown whether this is due to lack of expression or species specificity of the antibody; N/D, data not found; SSEA, stage-specific embryonic antigen; TRA-1-60, TRA-1-81 and GCTM-2, extracellular antigenic epitopes

Marker	Human ES cells	Human EG cells	Human EC cells	Primate ES cells	Mouse ES, EG and EC cells
SSEA-1	−	+	−	−	+
SSEA-3	+	+	+	+	−
SSEA-4	+	+	+	+	−
TRA-1-60	+	+	+	+	N/A
TRA-1-81	+	+	+	+	N/A
GCTM-2	+	N/D	+	+	N/A
Alkaline phosphatase	+	+	+	+	+
Nanog	+	+	N/D	N/D	N/D; + (ES)
Oct-4	+	+	+	+	+
hTERT	+	N/D	N/D	N/D	N/D; + (ES)

Laboratories in 11 of the 15 participating countries contributed hES cells (totalling 75 different lines) for the purpose of flow cytometric analysis of 17 surface antigens, quantitative RT PCR and microarray analysis of 100 selected gene transcripts, and for epigenetic studies (Andrews *et al*. 2005). To date (January 2006), there is not a specific gene or surface marker that can unequivocally define ES cell status, but rather only a series of trends in gene expression levels as an ES cell line begins to differentiate.

Morphology and marker expression

Morphological distinctions between mES, hES and non-human primate ES cells exist in terms of phenotype and marker expression. Owing to their evolutionary closeness, hES and non-human primate ES cells share similar phenotypic properties of flat monolayer colony growth with distinct cell borders in culture, whereas mES cells have a more rounded, clumped appearance with indistinguishable cell borders. Beyond phenotypical differences between species ES cells, gene-marker expression also differs among the model lines as illustrated in Table 1. Several studies report that cell-surface markers specific to the ICM lineage of the blastocyst have been maintained during prolonged hES cultures. It should be noted that the markers shown in Table 1 are not completely stage specific and have been shown to be expressed in other tissue types.

Expression marker description

Cell origins are often defined by one or more cell surface and or intracellular epitopes unique to that particular cell type. Stage-specific embryonic antigen (SSEA) markers are used to distinguish early stages of cell development, denoting pluripotency. These markers are globo-series glycolipids and are recognised by monoclonal antibodies. The SSEA-4 epitope is the globo-series glycolipid GL7. It is produced by the addition of sialic acid to the globo-series glycolipid Gb5, which carries the SSEA-3 epitope (Kannagi *et al*. 1983*a*, 1983*b*). It has been demonstrated that GL7 can react with antibodies to both SSEA-3 and SSEA-4 (Kannagi *et al*. 1983*a*, 1983*b*). Human and non-human primate ES cells will express SSEA-3 and -4 during pluripotency and only SSEA-1 upon differentiation (Andrews *et al*. 1996; Thomson and Marshall 1998; Thomson *et al*. 1998; Reubinoff *et al*. 2000). Interestingly, mES cells have reversed SSEA marker expression patterns compared with primates (Kannagi *et al*. 1983*a*).

Other cell-surface antigens are coupled to the pericellular matrix, more specifically a series of related keratin sulfate proteoglycans within the matrix. The TRA-1-60 epitope adheres to a particular epitope of the proteoglycan and is sialidase sensitive, whereas antibody TRA-1-81 reacts with another unknown epitope of the same core proteoglycan molecule. Both antibodies recognise proteoglycans found on all human and non-human primate stem cell types. The reason is unclear why these antibodies do not label mES or EC cells, but it is postulated that either mouse cells lack that particular antigen or the antibodies are species specific. The antibody GCTM2 has been reported to recognise an epitope associated with a cell-surface keratin sulfate, but its relative closeness to TRA-1-60 and TRA-1-81 has not been ascertained (Cooper *et al*. 1992).

All ES, EC and EG cells express alkaline phosphatase activity. In humans, there are four isoforms of the enzyme and it is not clear which isoform hES cells express. Human EC cells express the non-specific tissue form, along with a form (antibody detectable) that will cross-react with the germ cell or placental form (Pera *et al*. 2000).

Nanog is a NK-2-type homeodomain gene thought to encode a transcription factor that is critically involved in the self-renewal of stem cells. Thus, it may possibly act to repress genes necessary for differentiation and activate

those involved in self-renewal. Lin and colleagues (2005) demonstrated that the tumour suppressor p53 binds to the promoter of Nanog, suppressing its expression after DNA damage induced by UV irradiation or chemical treatment in mES cells. Therefore, p53 can stimulate differentiation of embryonic stem cells into alternate cell types that undergo p53-dependent cell-cycle arrest and apoptosis when genetic integrity is not preserved.

Oct-4, a POU-domain transcription factor, is highly expressed in ES cells (Thomson *et al*. 1998; Reubinoff *et al*. 2000; Niwa 2001) and has been shown to be essential for maintaining pluripotency in mES cells (Niwa 2001). It has been reported that Oct-4 transcripts are nearly exclusively found in pluripotent cells *in vivo* and within culture. Oct-4 down-regulation is observed in differentiating cells (Rosner *et al*. 1990). Not only is Oct-4 necessary for the maintenance of pluripotency, but its expression level governs three cell fates once differentiation occurs (Niwa *et al*. 2000; Hay *et al*. 2004). Niwa *et al*. (2000) showed a 2-fold increase promoted mES cell differentiation into embryonic and extraembryonic cell types typically produced upon withdrawal of the cytokine leukemia inhibitory factor (LIF), whereas a reduction in the level of Oct-4 to less than 50% triggered dedifferentiation of these cells into the trophoblast. These authors hypothesise possible roles for Oct-4 as a master regulator for initiation, maintenance and differentiation of pluripotent cells along with preventing respecification and dedifferentiation into the extraembryonic ectoderm (Niwa *et al*. 2000; Niwa 2001). Several candidate genes have been reported as targets of Oct-4 based on stem cell expression patterns and immunoprecipitation, but few have been conclusively verified. In the ongoing search for the identification of pluripotent markers, Xu and colleagues have reported that the catalytic component of telomerase, telomerase reverse transcriptase or hTERT, is expressed in undifferentiated cells and down-regulated upon differentiation (Xu *et al*. 2001).

Neural stem cells derived from human embryonic stem cells

A greater understanding of mammalian cellular differentiation and cell-fate specification has, and continues to be, of intense interest to developmental biologists. With the advent of hES cells, human developmental biology has progressed in recent years. Research on cell fate-specification in the central nervous system (CNS) is of enormous interest given the therapeutic potential in neuronal repair strategies. The mammalian central nervous system is developed from the neural tube. The early neural tube is composed of a single layer of pseudostratified columnar epithelium of neural epithelial progenitor cells. These neural progenitor cells are induced to form specified neural phenotypes such as motoneurons (Sanai *et al*. 2005). Below we will discuss the stepwise differentiation of hES cells to neural progenitors and further differentiation to motoneurons.

Table 2. Phenotype profile of mouse and human neural progenitor (NP) stem cells

Mouse data from Liu *et al*. 2002, human data from Shin *et al*. 2006

Antigen	Mouse NP cells	Human NP cells
Nestin	+	+
Sox2	+	+
A2B5/4D4	–	–
GFAP/CD44	–	–
RC1/S100/Vimentin	–	–
Sox10/NG2/PDGFRα	–	?
O4/GALC	–	–
Oct-4	–	–

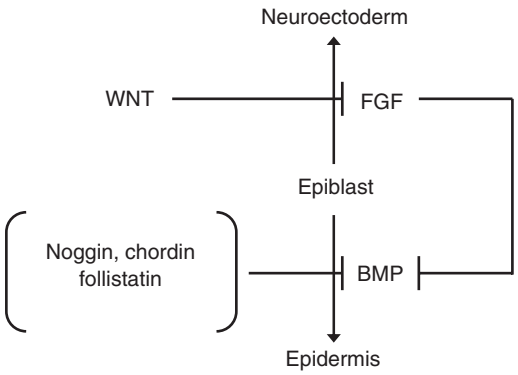


Fig. 2. Signalling pathways in neural induction.

Neural epithelial progenitors, or neural progenitor (NP) cells, are self-renewing cells that can differentiate into neurons, oligodendrocytes and astrocytes. According to Liu, who examined immunohistochemistry patterns in mouse tissue sections, the phenotype of NP cells can be characterised by several specific markers (Table 2) (Liu *et al*. 2002). Likewise, it has been shown that human NP cells have a phenotype characterisation similar to that found in mouse (Mayer-Proschel 2002). These cells expressed Nestin and Sox2, but did not express any other late stage neuronal or glial lineage markers. Neural progenitor cells give rise to all the cells that comprise the mammalian central nervous system, including various types of neurons and glial cells. However, it is not clear whether they multiply by symmetric or asymmetric cell division.

During early human development, NP cells form the neural tube during the third and fourth weeks of gestation. To acquire this NP cell induction from the epiblast, several molecules are required. Major signalling pathways involved in this induction are bone morphogenic protein (BMP), fibroblast growth factor (FGF) and Wnts (Fig. 2). Bone morphogenic protein signals block neural fate and promote epidermal fate and are excluded from prospective neural cells (see Wilson and Edlund 2001). Noggin, chordin and follistatin are known BMP inhibitors. Fibroblast growth factor has the dual role of repression of BMP expression and

promotion of a neural pathway that is independent of the BMP pathway repression, whereas Wnts block some of FGF's effect on BMP repression. Serum contains numerous undefined proteins including growth factors and molecules that could stimulate or prevent differentiation towards a particular lineage. Tropepe *et al*. (2001) proposed neural induction as a default choice of cell fate. When mES cells were dissociated and cultured in serum- and feeder-deprived conditions, colony-forming primitive neural stem cell populations could be obtained. Thus, defined culture will be beneficial not only for directed differentiation but also for the elimination of mesodermal and endodermal differentiation.

Strategies used to isolate human neural progenitor cells

Different strategies have been used to induce neural differentiation in hES cells. The most routine method of generating differentiated cell types has been through the three-dimensional structure of the embryoid body (EB). Embryonic stem cells in this agglomerate start spontaneous differentiation to form a sphere in suspension culture. Differentiated EBs contain neural stem cells, the proportions of which increase with retinoic acid exposure. Reubinoff and colleagues cultured hES cells until spontaneous differentiation occurred, then isolated a subpopulation to make neurospheres (Reubinoff *et al*. 2001). Pera *et al*. (2004) introduced a BMP inhibitor of noggin into this prolonged culture system. However, EB culture has disadvantages compared with adherent culture in that phenotype observation within the sphere is not possible with standard microscopy. In addition, stochastic differentiation yielded multiple cell lineages and limited the overall yield of the desired cells.

Ying and colleagues developed a monolayer differentiation method to obtain efficient neural induction of mES cells. When differentiation was triggered by the withdrawal of LIF, the mES cell monolayer chose a neural fate in serum-deprived medium (Ying *et al*. 2003). This efficient neural determination did not occur in medium containing serum. Though it was uncertain whether hES cells would behave similarly to their mouse counterparts, adherent differentiation in defined culture is an attractive strategy. We developed an adherent hES cell differentiation in defined culture conditions (Shin *et al*. 2006). The differentiation process was followed and examined immunohistochemically, suggesting a gradual but uniform differentiation to NP cells when serum was removed and FGF2 and LIF were added (Table 2). Derived NP cells were characterised both genetically and by their differentiation capacity. Subsequently, NPs were propagated using defined medium for 6 months while maintaining their differentiation potential.

Neural patterning in development of motoneurons

Several mature neuronal cell phenotypes have been derived from hES cells when exposed to growth factors first determined in neural patterning studies. As an example, we will describe motoneuron patterning and how it relates to deriving this phenotype *in vitro*. Cell fate determination occurs during and following neural-tube closure. Among several molecules, sonic hedgehog (Shh), retinoic acid (RA) and FGFs have been well defined molecules involved in dorsal–ventral and anterior–posterior axis formation. Here we review the role and interaction of these three factors in motoneuron development. Based on this prior understanding, we were able to derive a motoneuron phenotype from hES cells in 2005 (Shin *et al*. 2005).

Roles of sonic hedgehog

Dorsoventral axis formation involves the action of two opposing signalling pathways. Sonic hedgehog (Shh) originates ventrally from the notochord and later from the floor plate, whereas BMP diffuses dorsally from the boundary of neural and non-neural ectoderm and later from the roof plate. The notochord is the source of signals involved in the specification of the floor plate and secondarily to the formation of motoneurons and ventral interneurons with Shh being the major regulator of this signal. Loss- and gain-of-function studies of Shh have suggested that it is both necessary and sufficient to induce the floor plate (for review see Wilson and Maden 2005). Briefly, Shh-neutralising antibodies inhibited induction of ventral types in the developing notochord. Also, Shh-knockout mice lack floor plate and motoneuron development, although four other classes of ventral neurons still develop. In contrast, when the explant was exposed to Shh, an ectopic floor plate and motoneurons developed. Sonic hedgehog has a membrane-bound and non-membrane-bound form. It is presumed that the bound-protein form is involved in floor-plate induction and the soluble secreted form is in charge of motoneuron specification. Soluble secreted Shh has been shown to also act as a morphogen, eliciting different cell fates at varying concentration thresholds. Graded Shh activity directs neural identity through a set of homeodomain proteins that exhibit mutual cross repressive interactions.

Within the ventral spinal cord, five progenitor domains can be identified by unique combinations of transcription factors. The patterns of gene regulation in the progenitor domains are established by the high-ventral to low-dorsal gradient of Shh. Sonic hedgehog can either induce or repress the expression of the transcription factors within progenitor cells. These graded responses (either positive or negative) to Shh lead to the patterned expression of unique combinations of factors in each progenitor cell domain. A second level of transcriptional regulation is also in place involving homeodomain (HD) transcription factors, termed class I (Pax7, Pax6, Dbx1/2, and Irx3) and class II (the Nkx6 and Nkx2) proteins. Sonic hedgehog signalling regulates the expression of these HD class I and class II factors that have an opposing cross-repressive interaction (for review see Wilson and Maden 2005).

Roles of retinoic acid

Retinoic acid (RA) is the biologically active derivative of vitamin A and it induces a variety of embryonic carcinoma and neuroblastoma cell lines to differentiate into neurons. Retinoic acid acts through at least two sets of receptors, retinoic acid receptors (RARs) and retinoid X receptors (RXRs), which can interact with multiple putative coactivators and corepressors to yield a complex molecular pathway with a variety of pleiotropic effects. Retinoids are thought to function as morphogens during anterior–posterior patterning *in vivo*. In fact, endogenous retinoid expression has been documented at very low levels in the forebrain and midbrain. However, the hindbrain contains localised gradients of RA, with high expression observed in the spinal cord (Horton and Maden 1995). Ectopic expression of RA within whole *Xenopus* embryos led to increased expression in the hindbrain and spinal cord with a corresponding decrease in the forebrain. Retinoic acid affects the development of the head structure in a concentration-dependent manner while possessing the ability to alter tail structures as well. Within the hindbrain, gradually increased RA concentrations generate a stepwise gene activation effect in sequentially more posterior segments. In addition, RA is implicated in establishing regional identity within the spinal cord itself. It appears to act at sequential developmental stages to impose different rostrocaudal positional values.

Roles of fibroblast growth factor

Although FGF has been shown to play a major role as a repressor of differentiation, it has recently been shown to contribute to neural patterning. Fibroblast growth factor (including -2, -4, and -8) inhibits differentiation of adjacent neural tissue that opposes the effect of RA. Fibroblast growth factors 2, 4 and 8 are secreted by presomitic mesoderm and caudal cells, inhibiting neural tissue from differentiation and maintaining the caudal region as a stem zone (for review see Wilson and Maden 2005). They also affect patterning of the ventral spinal cord by differential inhibitory action on basic helix-loop-helix (bHLH) (e.g. Olig2) and homeobox transcription factors. Novitch and colleagues demonstrated that forced expression of FGF within *in vivo* neural cells resulted in marked repression of class I proteins and limited repression of class II proteins (Novitch *et al*. 2003). They also observed a joint action between FGF and RA causing an induction of Olig2 expression in the absence of Shh. This suggests that there is an Shh-independent pathway in ventral neural pattern formation. The independent role of FGF was also observed in oligodendrocyte progenitor induction from neocortical precursors in culture (Kessaris *et al*. 2004). This FGF activity was not affected by cyclopamine, a Shh pathway inhibitor. In contrast, PD173074, an inhibitor of FGF receptor (FGFR), blocked Shh activity. It has been demonstrated that constitutive activity of FGFR maintains a basal level of phosphorylated mitogen activated protein kinase (MAPK), and that Shh depends on MAPK for Olig2 induction.

Human neural progenitor to motoneuron differentiation using FGF2, RA and Shh

Recently, we demonstrated that human NP cells could form a motoneuron phenotype (Shin *et al*. 2005). The effect of inductive signals from FGF2, RA and Shh were examined. Gene expression changes in motoneuron progenitors (Olig2 expressing) of the motoneuron specific gene (*HB9*) were monitored using quantitative RT PCR. We found that all three factors had an additive effect, but individually, FGF2 was the most potent inducer of motoneuron-associated gene and marker expression. In the end, when these inducing factors were used on our progenitors, 10–20% of the cells in these cultures contained a motoneuron phenotype.

Future research to further enrich motoneuron phenotypes will require additional spatial and temporal cues. Motoneurons are derived from the same lineage as oligodendrocytes and it is possible that motoneuron progenitors could follow the same pathway. Also, motoneurons may require neurotrophic support, so although more progenitors are produced, the mature motoneurons might not survive. In the developing vertebrate, *in vivo* motoneuron survival depends on neurotrophic support. The predominant signalling molecules required include glial-cell-line-derived neurotrophic factor (GDNF), brain-derived neurotrophic factor (BDNF) and cardiotrophin-1 (CT-1) (Li *et al*. 2005). In this same report, optimal motoneuron development was obtained when early-stage NP cells from hES cells were induced using Shh and RA.

Conclusions

Previous studies have indicated that neuronal differentiation can be achieved using mouse and human ES cells. However, research towards uniform directed differentiation and isolation of desired cell types remains an ongoing effort in many laboratories. It has been shown that defined medium prevents mesodermal differentiation of mES cells and adherent differentiation was introduced for efficient homogeneous differentiation. In addition, inductive signals during *in vivo* development have been shown to induce motoneuron differentiation in mouse and non-human primate ES cells.

In this review, we have described progress towards culturing stable euploid hES cells that can be maintained and propagated. These hES cells could theoretically then be used in an adherent differentiation with defined medium process to establish homogeneous populations of NP cells. Finally, morphogens can be used to direct differentiation towards a specific phenotype. We have described the roles of FGF2, RA and Shh in the differentiation of a hES cell into a motoneuron.

References

Andrews, P. W., Casper, J., Damjanov, I., Duggan-Keen, M., Giwercman, A., *et al.* (1996). Comparative analysis of cell surface antigens expressed by cell lines derived from human germ cell tumours. *Int. J. Cancer* **66**, 806–816. doi:10.1002/(SICI)1097-0215(19960611)66:6<806::AID-IJC17>3.0.CO;2-0

Andrews, P. W., Benvenisty, N., McKay, R., Pera, M. F., Rossant, J., Semb, H., Stacey, G. N., and Initia, S. C. I. S. C. (2005). The International Stem Cell initiative: toward benchmarks for human embryonic stem cell research. *Nat. Biotechnol.* **23**, 795–797. doi:10.1038/NBT0705-795

Brimble, S. N., Zeng, X. M., Weiler, D. A., Luo, Y., Liu, Y., *et al.* (2004). Karyotypic stability, genotyping, differentiation, feeder-free maintenance, and gene expression sampling in three human embryonic stem cell lines derived prior to August 9, 2001. *Stem Cells Dev.* **13**, 585–597. doi:10.1089/SCD.2004.13.585

Buzzard, J. J., Gough, N. M., Crook, J. M., and Colman, A. (2004). Karyotype of human ES cells during extended culture. *Nat. Biotechnol.* **22**, 381–382. doi:10.1038/NBT0404-381

Cooper, S., Pera, M. F., Bennett, W., and Finch, J. T. (1992). A novel keratan sulfate proteoglycan from a human embryonal carcinoma cell-line. *Biochem. J.* **286**, 959–966.

Hay, D. C., Sutherland, L., Clark, J., and Burdon, T. (2004). Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. *Stem Cells* **22**, 225–235. doi:10.1634/STEMCELLS.22-2-225

Horton, C., and Maden, M. (1995). Endogenous distribution of retinoids during normal development and teratogenesis in the mouse embryo. *Dev. Dyn.* **202**, 312–323.

Kannagi, R., Cochran, N. A., Ishigami, F., Hakomori, S., Andrews, P. W., Knowles, B. B., and Solter, D. (1983*a*). Stage-specific embryonic antigens (Ssea-3 and Ssea-4) are epitopes of a unique globo-series ganglioside isolated from human teratocarcinoma cells. *EMBO J.* **2**, 2355–2361.

Kannagi, R., Levery, S. B., Ishigami, F., Hakomori, S., Shevinsky, L. H., Knowles, B. B., and Solter, D. (1983*b*). New globoseries glycosphingolipids in human teratocarcinoma reactive with the monoclonal antibody directed to a developmentally regulated antigen, stage-specific embryonic antigen 3. *J. Biol. Chem.* **258**, 8934–8942.

Kessaris, N., Jamen, F., Rubin, L. L., and Richardson, W. D. (2004). Cooperation between sonic hedgehog and fibroblast growth factor/MAPK signalling pathways in neocortical precursors. *Development* **131**, 1289–1298. doi:10.1242/DEV.01027

Li, X.-J., Du, Z.-W., Zarnowska, E. D., Pankratz, M., Hansen, L. O., Pearce, R. A., and Zhang, S.-C. (2005). Specification of motoneurons from human embryonic stem cells. *Nat. Biotechnol.* **23**, 215–221. doi:10.1038/NBT1063

Lin, T., Chao, C., Saito, S., Mazur, S. J., Murphy, M. E., Appella, E., and Xu, Y. (2005). p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression. *Nat. Cell Biol.* **7**, 165–171. doi:10.1038/NCB1211

Liu, Y., Wu, Y., Lee, J. C., Xue, H., Pevny, L. H., Kaprielian, Z., and Rao, M. S. (2002). Oligodendrocyte and astrocyte development in rodents: an *in situ* and immunohistological analysis during embryonic development. *Glia* **40**, 25–43. doi:10.1002/GLIA.10111

Mayer-Proschel, M. (2002). Human neural precursor cells – an *in vitro* characterization. *Clin. Neurosci. Res.* **2**, 58–69. doi:10.1016/S1566-2772(02)00007-5

Mitalipova, M., Calhoun, J., Shin, S., Wininger, D., Schulz, T., *et al.* (2003). Human embryonic stem cell lines derived from discarded embryos. *Stem Cells* **21**, 521–526. doi:10.1634/STEMCELLS.21-5-521

Mitalipova, M. M., Rao, R. R., Hoyer, D. M., Johnson, J. A., Meisner, L. F., Jones, K. L., Dalton, S., and Stice, S. L. (2005). Preserving the genetic integrity of human embryonic stem cells. *Nat. Biotechnol.* **23**, 19–20. doi:10.1038/NBT0105-19

Niwa, H. (2001). Molecular mechanism to maintain stem cell renewal of ES cells. *Cell Struct. Funct.* **26**, 137–148. doi:10.1247/CSF.26.137

Niwa, H., Miyazaki, J., and Smith, A. G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genet.* **24**, 372–376. doi:10.1038/74199

Novitch, B. G., Wichterle, H., Jessell, T. M., and Sockanathan, S. (2003). A requirement for retinoic acid-mediated transcriptional activation in ventral neural patterning and motor neuron specification. *Neuron* **40**, 81–95. doi:10.1016/J.NEURON.2003.08.006

Pera, M. F., Reubinoff, B., and Trounson, A. (2000). Human embryonic stem cells. *J. Cell Sci.* **113**, 5–10.

Pera, M. F., Andrade, J., Houssami, S., Reubinoff, B., Trounson, A., Stanley, E. G., Ward-van Oostwaard, D., and Mummery, C. (2004). Regulation of human embryonic stem cell differentiation by BMP-2 and its antagonist noggin. *J. Cell Sci.* **117**, 1269–1280. doi:10.1242/JCS.00970

Reubinoff, B. E., Pera, M. F., Fong, C. Y., Trounson, A., and Bongso, A. (2000). Embryonic stem cell lines from human blastocysts: somatic differentiation *in vitro*. *Nat. Biotechnol.* **18**, 399–404. doi:10.1038/74447

Reubinoff, B. E., Itsykson, P., Turetsky, T., Pera, M. F., Reinhartz, E., Itzik, A., and Ben-Hur, T. (2001). Neural progenitors from human embryonic stem cells. *Nat. Biotechnol.* **19**, 1134–1140. doi:10.1038/NBT1201-1134

Rosler, E. S., Fisk, G. J., Ares, X., Irving, J., Miura, T., Rao, M. S., and Carpenter, M. K. (2004). Long-term culture of human embryonic stem cells in feeder-free conditions. *Dev. Dyn.* **229**, 259–274. doi:10.1002/DVDY.10430

Rosner, M. H., Vigano, M. A., Ozato, K., Timmons, P. M., Poirier, F., Rigby, P. W. J., and Staudt, L. M. (1990). A Pou-domain transcription factor in early stem-cells and germ-cells of the mammalian embryo. *Nature* **345**, 686–692. doi:10.1038/345686A0

Sanai, N., Alvarez-Buylla, A., and Berger, M. S. (2005). Mechanisms of disease: Neural stem cells and the origin of gliomas. *N. Engl. J. Med.* **353**, 811–822. doi:10.1056/NEJMRA043666

Shamblott, M. J., Axelman, J., Wang, S. P., Bugg, E. M., Littlefield, J. W., Donovan, P. J., Blumenthal, P. D., Huggins, G. R., and Gearhart, J. D. (1998). Derivation of pluripotent stem cells horn cultured human primordial germ cells. *Proc. Natl Acad. Sci. USA* **95**, 13 726–13 731. doi:10.1073/PNAS.95.23.13726

Shin, S., Dalton, S., and Stice, S. L. (2005). Human motor neuron differentiation from human embryonic stem cells. *Stem Cells Dev.* **14**, 266–269. doi:10.1089/SCD.2005.14.266

Shin, S., Mitalipova, M., Noggle, S., Tibbitts, D., Venable, A., Rao, R., and Stice, S. L. (2006). Long-term proliferation of human embryonic stem cell-derived neuroepithelial cells using defined adherent culture conditions. *Stem Cells* **24**, 125–138. doi:10.1634/STEMCELLS.2004-0150

Thomson, J. A., and Marshall, V. S. (1998). Primate embryonic stem cells. *Curr. Top. Dev. Biol.* **38**, 133–165.

Thomson, J. A., Kalishman, J., Golos, T. G., Durning, M., Harris, C. P., Becker, R. A., and Hearn, J. P. (1995). Isolation of a primate embryonic stem-cell line. *Proc. Natl Acad. Sci. USA* **92**, 7844–7848. doi:10.1073/PNAS.92.17.7844

Thomson, J. A., Kalishman, J., Golos, T. G., Durning, M., Harris, C. P., and Hearn, J. P. (1996). Pluripotent cell lines derived from common marmoset (*Callithrix jacchus*) blastocysts. *Biol. Reprod.* **55**, 254–259. doi:10.1095/BIOLREPROD55.2.254

Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., and Jones, J. M. (1998). Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147. doi:10.1126/SCIENCE.282.5391.1145

Tropepe, V., Hitoshi, S., Sirard, C., Mak, T. W., Rossant, J., and van der Kooy, D. (2001). Direct neural fate specification from embryonic stem cells: a primitive mammalian neural stem cell stage acquired through a default mechanism. *Neuron* **30**, 65–78. doi:10.1016/S0896-6273(01)00263-X

Wilson, L., and Maden, M. (2005). The mechanisms of dorsoventral patterning in the vertebrate neural tube. *Dev. Biol.* **282**, 1–13. doi:10.1016/J.YDBIO.2005.02.027

Wilson, S. I., and Edlund, T. (2001). Neural induction: toward a unifying mechanism. *Nat. Neurosci.* **4**, 1161–1168. doi:10.1038/NN747

Xu, C. H., Inokuma, M. S., Denham, J., Golds, K., Kundu, P., Gold, J. D., and Carpenter, M. K. (2001). Feeder-free growth of undifferentiated human embryonic stem cells. *Nat. Biotechnol.* **19**, 971–974. doi:10.1038/NBT1001-971

Ying, Q. L., Stavridis, M., Griffiths, D., Li, M., and Smith, A. (2003). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat. Biotechnol.* **21**, 183–186. doi:10.1038/NBT780

Manuscript received 24 April 2006; revised and accepted 4 September 2006.

J Struct Funct Genomics (2007) 8:1–10
DOI 10.1007/s10969-007-9026-3

Structural and transcriptional analyses of a purine nucleotide-binding protein from *Pyrococcus furiosus*: a component of a novel, membrane-bound multiprotein complex unique to this hyperthermophilic archaeon

Brian Gerwe · Laura-Lee Clancy Kelley · Bret D. Dillard · Thomas Lai · Zhi-Jie Liu · Wolfram Tempel · Lirong Chen · Jeff Habel · Doowon Lee · Francis E. Jenney Jr. · Frank J. Sugar · Jane S. Richardson · David C. Richardson · M. Gary Newton · Bi-Cheng Wang · Michael W. W. Adams · John P. Rose

Received: 18 January 2007 / Accepted: 2 September 2007 / Published online: 12 October 2007

Abstract The open-reading frame PF0895 in the genome of the hyperthermophilic archaeon, *Pyrococcus furiosus*, encodes a 206-residue protein (M_R 23,152). The structure of the recombinant protein was solved by single isomorphous replacement with anomalous scattering (SIRAS) using a mercury derivative. It has been refined to 1.70 Å with a crystallographic R and R_{free} values of 19.7% and 22.3%, respectively. The PF0895 structure is similar to those of the ATP binding cassettes observed in the ABC transporter family. However, bioinformatics and molecular analyses indicate that PF0895 is not part of the expected five-gene operon that encodes a typical prokaryotic solute-binding ABC transporter. Rather, transcriptional profiling data show that PF0895 is part of a novel four-gene operon (PF0895–PF0896–PF0897–PF0897.1) where only PF0895 has homologs in other organisms. Interestingly, from genome analysis, *P. furiosus* itself contains a second version of this complex, encoded by PF1090–PF1093. From the structural studies we can only conclude that one of the subunits of this novel membrane complex, PF0895, and its homolog PF1090, likely bind a purine nucleotide. PF0895 is therefore predicted to be part of a membrane-bound multiprotein complex unrelated to ABC transporters that is so far unique to *P. furiosus*. It appears to play a role in the stress response, as its expression is down regulated when the organism is subjected to cold-shock, where cells are transferred from 95°C, near the optimal growth temperature, to 72°C, near the minimal growth temperature. The related PF1090-containing operon is unaffected by cold-shock and is independently regulated.

Keywords Structural genomics · SECSG · ATP-binding protein · PF0895 · Subdomain structure · 1SGW

Brian Gerwe and Laura-Lee Clancy Kelley contributed equally to this work.

B. Gerwe · L.-L. C. Kelley · B. D. Dillard · T. Lai · Z.-J. Liu · W. Tempel · L. Chen · J. Habel · D. Lee · F. E. Jenney Jr. · F. J. Sugar · M. G. Newton · B.-C. Wang · M. W. W. Adams · J. P. Rose (✉)
Southeast Collaboratory for Structural Genomics (SECSG), Department of Biochemistry and Molecular Biology, University of Georgia, Room B204B, Davison Life Science Complex, Athens, GA 30602, USA
e-mail: rose@bcl4.bmb.uga.edu

Present Address:
J. Habel
The Scripps Research Institute, Jupiter, FL, USA

J. S. Richardson · D. C. Richardson
Department of Biochemistry, Duke University, Durham, NC, USA

Introduction

Three basic types of transport systems have been identified for the translocation of solutes across membranes: ABC transporter (ATP-binding cassette), TRAP (ATP-independent) and symport systems [1–4]. It has been well documented that these are intra-membrane proteins essential to many cellular processes and are involved in critical reactions carried out by a wide variety of organisms [5–11]. The ABC type transporters are ubiquitous in nature and couple ATP hydrolysis to the translocation of a solute across the cytoplasmic membrane [5–11]. The typical prokaryotic ABC transporter architecture consists of five proteins; two integral membrane permeases, two

cytoplasmically located ATPases, and an extracellular substrate-binding protein [12]. It is this unique binding domain which delivers the substrate to the membrane domains and allows for the high-binding affinity between substrates and the ABC transporter [12].

Herein we describe the structure of a protein from *Pyrococcus furiosus* [13] which is related to a subunit of ABC transporters. This organism grows optimally near 100°C and is classified within the archaeal domain of life. Its genome contains an open reading frame (ORF), PF0895, which is predicted to encode a 206 amino acid polypeptide. Sequence analyses by BLAST [14] and InterPro [15] indicate that this protein is the ATPase cassette within the ABC transporter family (IPR003439) [14, 16–18]. The structure of PF0895 is consistent with this functional assignment as a member of an ABC transporter. However, it is cotranscribed with a set of genes that do not resemble the well-described 5 member ABC transporter family. This implies that PF0895, and the other members of the cotranscribed operon, PF0896, PF0897, and PF0897.1, are part of a novel type of membrane-bound multiprotein complex, in which only one member, PF0895, resembles traditional ABC transporters. Curiously, *P. furiosus* contains a very close homolog of this same multiprotein complex, but such complexes have yet to be identified in other organisms, including other *Pyrococcus* species.

Materials and methods

Protein expression and purification

The ORF encoding PF0895 (*Pfu*–867808, see www.secsg.org) was cloned into an expression plasmid pET24dBam and expressed in the *Escherichia coli* host strain BL21 Star DE3 pRIL [19, 20]. Two 1-l batches were grown in 2.8L Fernbach flasks in auto-induction media (rich ZYP-5052 and defined PA-5052) [21] at 37°C (300 rpm) for 6 h and then incubated at 18°C for 16 h. The cells were harvested by centrifugation (5487×g, 30 min) and resuspended in 20 ml of 20 mM sodium phosphate, 500 mM NaCl, 10 mM imidazole, pH 8.1. They were stored at –80°C until needed. The recombinant protein contained an N-terminal His_6 purification tag (AHHHHHHGS—M_R 24,059) and was purified according to the high-throughput protocols established for *P. furiosus* protein production by the Southeast Collaboratory For Structural Genomics (SECSG) [19, 22]. This involved three column chromatography steps; immobilized metal (Ni) affinity (EMD Biosciences, Madison, WI), cation exchange (HiTrap S; GE Healthcare, Piscataway, NJ) and size exclusion (Superdex 75; GE Healthcare). Protein identity and metal content were assessed using electrospray ionization mass spectroscopy (ESI-MS) and by inductively coupled plasma emission spectroscopy (ICP) carried out by the Department of

Table 1 X-ray diffraction data and statistics

	Native	Derivative
Protein		
Growth medium	ZYP-5052	PA-5052
Crystallization		
Protein concentration	5.9 mg/ml	9.4 mg/ml
Crystallization conditions	15% (w/v) PEG 4000 10% (v/v) glycerol 10% (v/v) 2-propanol 100 mM Tris·HCl, pH 7.9	15% (w/v) PEG 3000 100 mM Tris·HCl, pH 8.8
Soaking conditions	None	0.2 mM Baker Dimercurial[a]
Data collection		
Wavelength (Å)	1.0660	0.9793
Distance (mm)	160.00	180.00
Oscillation width (°)	0.50	1.00
Number of images	380	180
Resolution (Å)	30.0–1.70 (1.76–1.70)[b]	50.0–1.90 (2.05–1.90)[b]
Unit Cell (a, b, c) (Å)	44.98, 67.58, 72.77	45.14, 67.83, 73.08
Space Group	P $2_12_12_1$	P $2_12_12_1$
Rsym[c]	0.041 (0.123)[b]	0.064 (0.308)[b]
Completeness (%)	99.4 (98.5)[b]	98.9 (97.1)[b]
$\langle I/\sigma_I \rangle$	35.94 (11.3)	20.25 (4.4)
Redundancy	7.3 (5.5)	6.7 (5.5)

[a] 1,4-diacetoxymercury-2,3-dimethoxy butane

[b] The statistics for the highest resolution shell are given in parentheses

[c] Rsym = $\sum |I - \langle I \rangle| / \sum (I)$, where $\langle I \rangle$ is the average intensity multiple observations of symmetry-related reflections

Table 2 Crystallographic and refinement data for PF0895

Crystal	
Volume fraction of protein[a]	53%
V_m (Å^3/dalton)[a]	2.30
Model	
Total number of residues	204 (214 total)
Total non-H atoms	1640 (1541 protein atoms, 3 hetero atoms)
Number of water molecules	96
Refinement	
Resolution range	30–1.70 Å
R^b (obs.)	19.7% (19.3%)
R_{free}	22.3% (23.4%)
Stereochemical ideality[c]	
RMSD bonds	0.011 Å
RMSD angles	1.294°
RMSD torsion angles	4.867°
<Temperature factor>	19.44 Å^2 (protein)
Ramachandran plot[d]	
Residues in most favored regions	167 (93.3%)
Residues in allowed regions	11 (6.1%)
Residues in generously allowed regions	1 (0.6%)
Residues in disallowed regions	0 (0%)

[a] Matthews' coefficient [64]

[b] $R = \sum ||F_o|-|F_c|| / \sum |F_o|$, where F_o and F_c are observed and calculated structure factor amplitude, respectively. R_{free} is calculated for a randomly chosen 3% of reflections

[c] RMSD is the root mean square deviation from ideal geometry

[c] Ramachandran plot [65]

Chemistry and Chemical Analysis Laboratory, University of Georgia, respectively. The two purification procedures yielded 3.2 and 4.1 mg of protein, which were concentrated to 5.6 and 9.4 mg/ml, respectively, in 20 mM Tris buffer containing 300 mM NaCl, 2 mM dithiothreitol, pH 8.0 for the crystallization trials. Protein concentrations were estimated using a calculated extinction coefficient of 10,204 M^{-1} cm^{-1} at 280 nm [23].

X-ray diffraction analysis

Initial crystallization conditions for the protein were determined by screening against a set of eight commercial crystallization screens using 200 nanoliter sitting drops containing equal volumes of protein and precipitant solutions as described by Liu [24]. Crystal optimization was carried out using 2 µl drops using the micro batch under oil method [24].

The structure was solved by the method of single isomorphous replacement with anomalous scattering (SIRAS) using a crystal soaked (~2 h) in its mother liquor containing 0.2 mM 1,4-diacetoxymercury-2,3-dimethoxybutane (Bakers Dimercurial). The crystals were not back soaked. For data collection, crystals were harvested using a rayon loop of appropriate size and flash frozen in liquid nitrogen [25]. Cryoprotection was achieved using a 20% (v/v) glycerol-mother liquor mixture. Data sets on both the native and heavy atom soaked crystals were collected on beamline 22-ID, Advanced Photon Source using a Mar 225 CCD detector as described in Table 1. All data were indexed, integrated, and scaled using the HKL2000 software suite [26]. The mercury substructure was determined from the derivative data using SOLVE V2.03 [27] and used to generate the initial (SIRAS) protein phases (MLPHARE [28]). Phase extension was carried out by merging the initial phases with the high-resolution structure factors from the native data set (CAD [28]). The initial model was obtained using the automatic tracing feature of ARP/wARP [29] and adjusted manually, where necessary, using XFIT [30]. Structure refinement was carried out using REFMAC5 [31], and validated using MOLPROBITY [32] and PROCHECK [33]. Both coordinates and structure factors have been deposited in the Protein Data Bank (PDB) [34] using the pdb_extract [35] tool.

Gene expression profiling

Potential signal sequences and transmembrane domains were calculated as described previously [36]. The transcriptional profiling data were obtained from published reports that used cells grown using either maltose or peptides as the carbon source at 95°C [20, 37] or cells grown at 95°C or 72°C using maltose as the carbon source [20, 37].

Results and discussion

Structural analysis

The molecular weight of the recombinant PF0895 protein was estimated to be 24,056 Da by ESI-MS analysis, which is in good agreement with the calculated value of 24,059 Da (with N-terminal His_6 affinity tag). Crystals were obtained from two different protein preparations (ZYP-5052 and PA-5052). In the initial screen both samples produced crystals from the Wizard I (www.emeraldbiosystems.com) condition 22 formulation (10% (v/v) 2-propanol, 100 mM Tris·HCl, pH 8.9). Crystals from the two preparations were then optimized to give the conditions presented in Table 1. Glycerol was added to the ZYP-5052 setups as a nucleation suppressant. Crystals

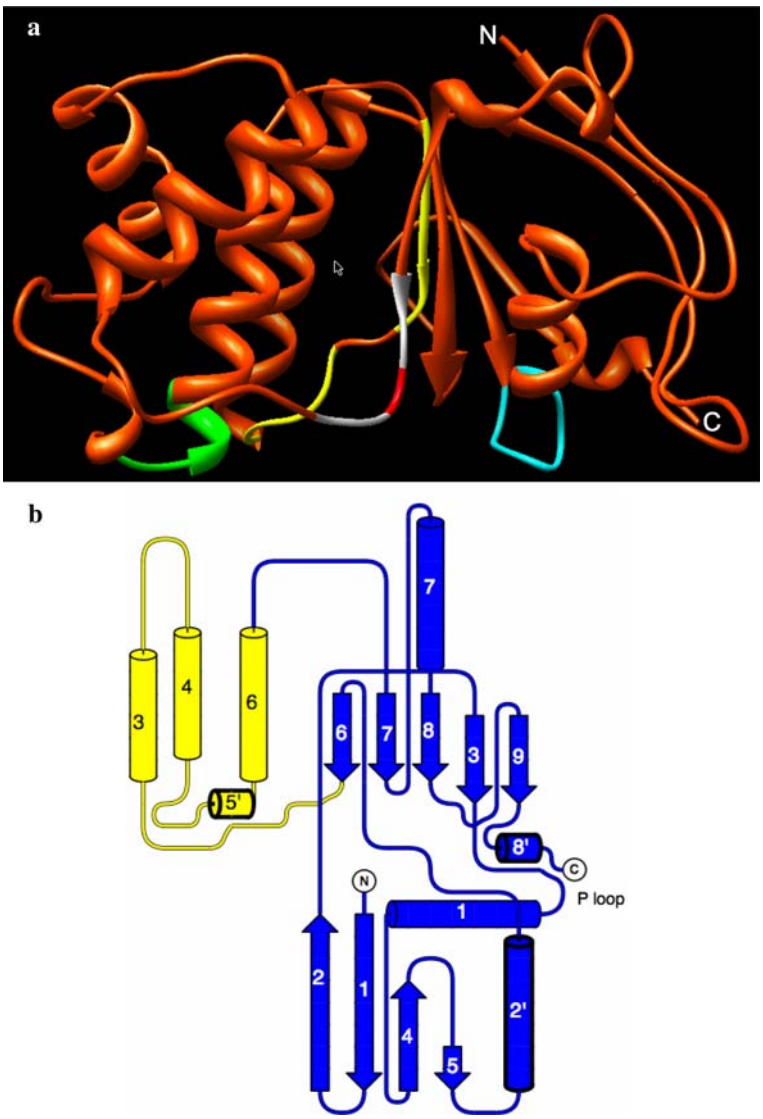


Fig. 1 (**a**) Ribbon drawing (CHIMERA [56]) of the PF0895 structure (PDB ID: 1SGW) viewed down its central β core. The molecule is colored according to sequence alignment shown in Fig. 2: cyan—the Walker A motif, gray—Q-loop, green (Gln labeled red)—signature motif, yellow—Walker B and D-loop motifs, Switch not highlighted. **(b)** A topology diagram showing the secondary structural features of PF0895. The P-loop domain and central β core (β3, β6–9) are colored blue. The helical domain is colored yellow. Helices denoted with a prime are 3_{10} helices. Image created using TopDraw [66]

generally appeared in 3–5 days and grew to usable size by week 2. Native data were collected from crystals from the ZYP-5052 preparation while crystals produced from the PA-5052 preparation were used in making the Bakers Dimercurial derivative. Table 1 summarizes the results from the X-ray analysis.

The structure was solved using the SCA2Structure pipeline [38] and refined to 1.7 Å resolution. The refined PF0895 model (PDB entry 1SGW) consists of residues 2–200 of the native protein, a sodium ion, two chloride ions and 96 solvent molecules modeled as water. Refinement of the structure converged to give an R value of 0.197 (R_{free} = 0.234) with good stereochemistry, see Table 2. Residues of the N-terminal His_6 tag used for purification and C-terminal residues 201–206 were not observed in the electron density maps and are assumed to be disordered.

The protein crystallized as a monomer of dimensions (47 × 44 × 46 Å), see Fig. 1. The PF0895 molecule has a kidney shaped structure similar to that reported for other ATP-binding cassette proteins of the ABC transporter family with overall root mean square deviations (RMSDs) (DALI [39]) ranging from 1.4 to 3.6 Å (Cα's). The molecule has a solvent accessible surface area [40] of 9438 Å [41] and is characterized by a central beta core composed of a 5- stranded (β3, β6–β9) parallel β-sheet separating the two domains of the protein; a large α/β domain (residues 1–88 and 151–214) that includes the N- and C-terminal regions of the protein and a smaller mostly helical domain (residues 89–150) [42]. The larger α/β domain resembles a typical P-loop NTPase domain [43] and consists of N-terminal strands β1, β2, β4, β5 forming a 4-stranded antiparallel β-sheet, two N-terminal helices α1, α2′ (here prime denotes a 3_{10} helix) and the short C-terminal helix α8′ (see Fig. 1b). Structural (DALI [39]) and sequence comparisons (EXPRESSO [44]) show that the PF0895 P-loop domain contains the characteristic Walker A motif or P-loop

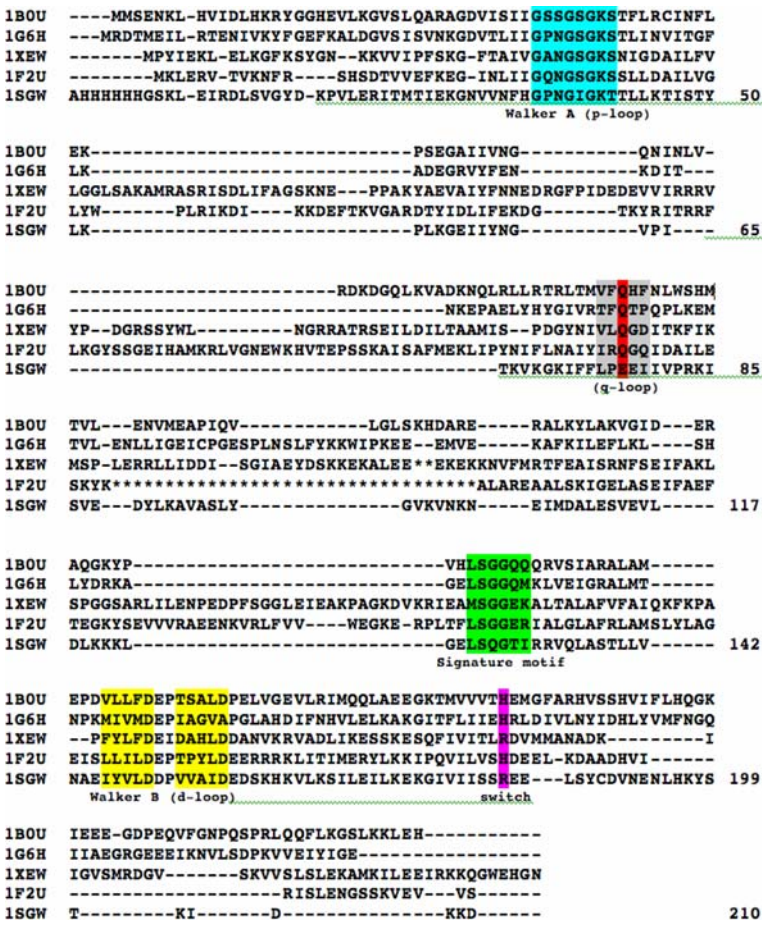


Fig. 2 A sequence alignment (EXPRESSO [44]) showing the sequence homology between PF0895 (1SGW) and similar ABC type ATPases from *Salmonella typhimurium* (1B0U), *Methanococcus jannaschii* (1G6H), *P. furiosus* SMC protein (1XEW) and *P. furiosus* Rad50 (1F2U). Common features of the ABC-ATPase superfamily are also shown: Walker A motif (cyan), Q-loop (gray), signature motif (green), Walker B and D-loop motifs (yellow) and Switch (magenta). Starred residues represent the long coiled-coiled regions in the 1XEU and 1FTU sequences (not observed in the structures) that connect the N- and C-terminal domains

(residues 34–41), Q-loop (residues 75–79), Walker B motif (residues 146–150), D-loop (residues 153–157) and switch (residue 183) nucleotide binding motifs [45] consistent with ABC type ATPases, see Fig. 2.

The smaller PF0895 helical domain (residues 89–150) again is typical of the ABC type ATPases. The PF0895 helical domain consists of helices α3, α4, α5′ and α6 with helices α3, α4, and α6 forming an anti-parallel α-helical bundle (Fig. 1b). Both sequence [44] and structural alignment [46] place the (LSGGQ) ABC signature motif or "linker sequence" (residues 126–131) in the short loop spanning the C-terminus of α5′ and the N-terminus of α6 (see Figs. 1, 2). The ABC signature motif is a unique feature of ABC transporters and is thought to play an important role in ATP hydrolysis and in solute transport [47, 48]. Within the transport complex it is believed that the two ATP binding cassettes form a "head-to-tail" dimer placing the ATP binding site and Walker A motif from one molecule in close proximity to the signature motif of its dimer partner [49, 50] as observed in the crystal structure of the Rad50 ATPase-ATP complex (PDB entry 1F2U) [51]. Mutations within the signature motif have been shown to inactivate the complex [51, 52] and to be involved in diseases such as cystic fibrosis [53]. The ABC signature sequence observed in PF0895 (LSQGT) is a variant of the (LSGGQ) sequence observed in most ABC type ATPases. The LSQGT motif has been observed in ABC transporters associated with the retina found in cow (GI:27806343) [54] and dog (GI:40556955) [55]. The sequence variant may reflect the ability of *P. furiosus* to use ADP as an energy source or a preference for another substrate; however, this hypothesis must await experimental proof.

PF0895 is structurally similar to a number of ABC type ATPases including Rad50 from *P. furiosus* that is involved in double stranded DNA break repair [51]. The Rad50 structure has been determined in both the ATP-free (PDB entry 1F2T) and ATP-bound (1FTU) forms [51], thus making it a good model for comparison. The Rad50 structures show that there is a significant conformational change in the enzyme upon ATP binding that results in the formation of the active ATP bound head-to-tail dimer. This ATP-driven conformational control of dimerization is thought to be a common feature of the ABC type ATPase superfamily [51]. Interestingly, when the Rad50 and PF0895 structures are compared [46] the ATP-free PF0895 structure more closely resembles (in terms of the position of the P-loop, Q-loop and signature motifs) the ATP-bound

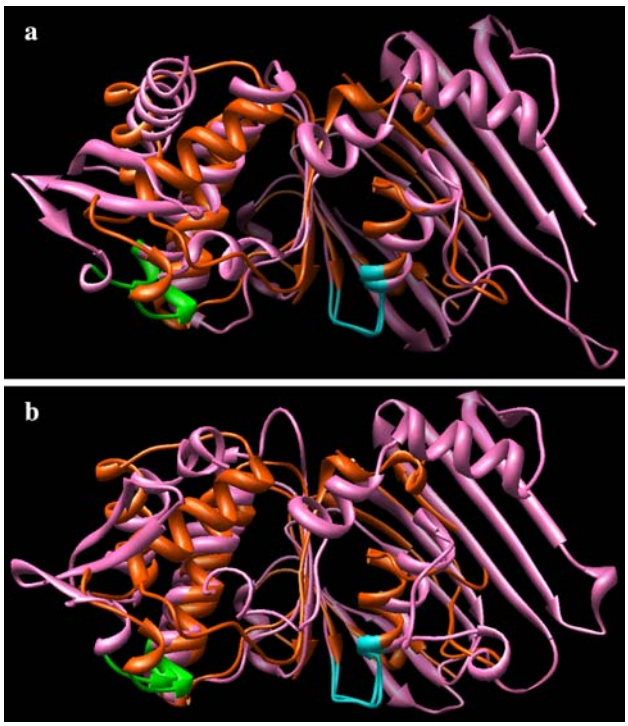


Fig. 3 The superposition [46] of the PF0895 structure (burnt orange) onto the *P. furiosus* Rad50 structures (pink). (**a**) The ligand free Rad50 structure (1F2T) RMSD 1.8 Å (35 Ca pairs) and (**b**) The ATP bound Rad50 structure (1F2u) RMSD 1.6 Å (69 Ca pairs). The P-loop (cyan) and signature motif (green) are shown

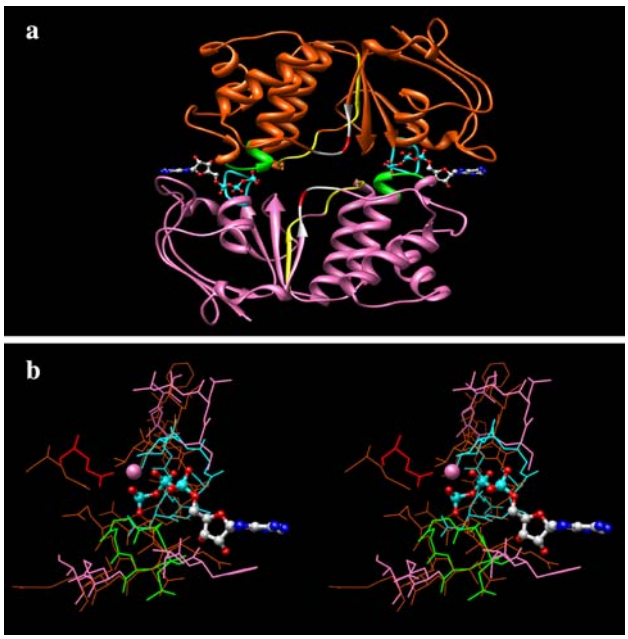


Fig. 4 (**a**) A computer model of ATP bound PF0895 dimer created by the superposition [46] of the PF0895 structure (1SGW) onto that of Rad50 ATP bound complex (1F2U). Features common to the ABC-ATPase superfamily are colored according to the scheme presented in Fig. 2. The bound nucleotide is colored as follows: carbon—grey, oxygen—red and nitrogen—blue. (**b**) The superposition (presented in stereo) of the putative PF0895 ATP binding pocket (Walker A (cyan) and signature motifs (green) modeled on the Rad50 ATP-bound structure (1FTU; pink) showing that most, if not all, protein-nucleotide interactions can be maintained in the modeled PF0895 dimer. The RMSD (Cα) between the structures is 0.70 Å. The nucleotide is colored as follows: carbon—grey, oxygen—red and nitrogen—blue. The magenta sphere is a magnesium ion hydrogen bonded to Gln 140 in the Rad50 structure that is essential for ATPase activity

Rad50 enzyme, see Fig. 3. The MatchMaker alignment tool [46] running within CHIMERA [56] used for the alignments shown in Fig. 3 uses both sequence and structural information in the alignment process. This approach is especially useful for structures having low sequence homology, such as the present case, where the homology between the PF0895 and Rad50 sequences is ~23% (excluding the Rad50 coiled-coil region).

Comparing the PF0895 structure to the structures of both ATP-free and ATP-bound Rad50 the MatchMaker alignment shows good agreement between the positions of the P-loop in PF0895 and both Rad50 structures. In addition, MatchMaker also places the side chains of Gln140 (Rad50) of the Q-loop and Glu 77 (its PF0895 counterpart) within 1 Å of each other, with Gln140 NE2 lying 0.3 Å from Glu 77 OE1. In the Rad50-ATP complex structure Gln140 is hydrogen bonded to a magnesium ion, necessary for ATPase activity, in the ATP binding site. The alignment also places Cl 201 in the PF0895 structure within 0.3 Å of the β phosphate group of the bound ATP observed the Rad50-ATP complex.

Interestingly the alignment shows that the position of the signature motif in PF0895 is more similar to the position of the signature motif observed in the ATP-bound Rad50 enzyme (Fig. 3b). For example, using the position of the alpha carbon of the signature motif leucine as a guide, the superposition of PF0895 and ATP-bound Rad50 structures gives a Cα–Cα distance of 3.3 Å while the superposition of PF0895 and ATP-free Rad50 structures gives a Cα–Cα distance of 8.5 Å. Although the structure reported here appears similar to the Rad50-ATP complex the PF0895 structure does not preclude some conformational change in the enzyme upon ATP binding as observed in Rad50 and other ABC type ATPases.

A hypothetical model [46] of the PF0895 ATP-bound dimer, generated by the superposition of the PF0895 structure onto the RAD50-ATP complex dimer (PDB entry 1F2U), is shown in Fig. 4. A close inspection of the modeled PF0895 ATP-binding pocket (Fig. 4b) suggests that with a little rearrangement of the pocket residues (primarily residues in the signature motif) ATP could be easily accommodated since a majority of the protein–phosphate interactions observed in the Rad50-ATP complex could be maintained in the PF0895 dimer. However, lacking experimental evidence it is possible that the true

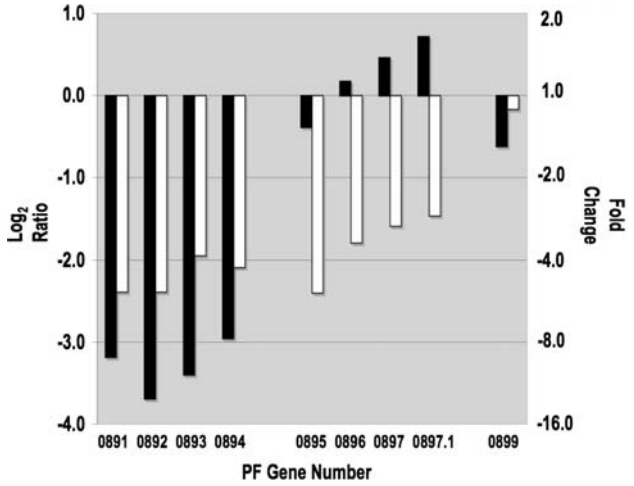


Fig. 5 Regulation of expression of the gene encoding PF0895 and of adjacent genes. The bars indicate the changes in expression of PF0895 and the proposed operon (PF0895–PF0897.1), of the operon encoding hydrogenase I (PF0891–PF0894), and of the adjacent gene encoding PF0899. Solid bars indicate the change in expression when peptides are the primary carbon source rather than maltose [37]. The open bars indicate the change in expression when cells are transferred from 95° to 72°C for 5 h [20]

substrate is ADP rather than ATP. This is relevant because some hyperthermophiles including *P. furiosus* use ADP as an energy source in the activation of sugars and sugar phosphates [57].

Genome organization and transcriptional profiling

PF0895 is classified as a member of the COG4152 family and has high similarity to the ATP-binding proteins of uncharacterized ABC-type transporters in a variety of bacteria, as well as archaea. For example, it shows approximately 60% sequence similarity to proteins in the

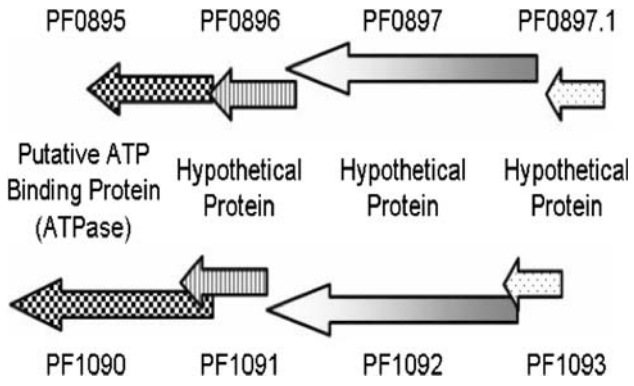


Fig. 6 Proposed operon structure containing PF0895. PF0897.1 was recently annotated [23]. The homologous operon that includes PF1090 is also shown where corresponding genes have the same shading. The degree of overlap of the genes is indicated in both operons

mesophilic bacteria, *Borrelia burgdorferi* and *Cytophaga hutchinsonii*. However, from the genome sequence of *P. furiosus*, PF0895 is not part of the expected five-gene operon encoding two permeases, two ATPases and a solute-binding protein. Instead, the gene is situated between genes encoding a hypothetical protein (PF0896) and one (PF0894) of the four subunits (PF0891–PF0894) of the cytoplasmic hydrogenase I of *P. furiosus* [58]. That PF0895 and the hydrogenase have no functional relationship is suggested by the fact that they are encoded on separate strands, and evidence for this is provided by transcriptional analyses generated by whole genome DNA microarrays [20, 37]. As shown in Fig. 5, expression of the genes encoding the four subunits of the hydrogenase is dramatically down-regulated (~8-fold) in cells grown with peptides as the primary carbon source rather than maltose. However, under the same conditions, expression of PF0895 is unaffected. The expression of this gene does respond; in this case it is down-regulated (>4-fold) when cells are grown (for 5 h) at a sub-optimal temperature (72°C), rather than at 95°C, which is much closer to the optimum. Interestingly, expression of the hydrogenase is also down-regulated in the same cells (Fig. 5), although apparently independent of the expression of PF0895.

The array data also indicate that PF0895 does not function independently. As shown in Fig. 5, PF0895 appears to be coordinately regulated with two adjacent genes, PF0896 and PF0897, as part of the so-called cold shock response of this organism, with little if any change in expression of any of these three genes when the carbon source is changed. However, both PF0896 and PF0897 do not have homologs in the genome sequences of any other organism and remain hypothetical, although the fact that their expression is regulated indicates that they are true genes. The adjacent gene, PF0898, is not expressed at a detectable level under the growth conditions tested (data not shown). PF0898 is annotated as a transposase, and is therefore not part of the putative PF0895–PF0896–PF0897 operon. In a recent revision of the annotation of the *P. furiosus* genome, the intergenic region between PF0897 and PF0898 was recognized to potentially contain a previously unannotated gene, termed PF0897.1 [23]. The corresponding protein also lacks a homolog in any other organism whose genome has been sequenced [23]. As shown in Fig. 5, the expression of this gene is also up-regulated in response to cold-shock, consistent with it being a true gene. Moreover, these transcriptional data suggest that the gene encoding PF0895 is part of a four-gene operon, as indicated in Fig. 6. All four genes are on the same DNA strand and are separated from each other by less than 10 nucleotides, consistent with an operon arrangement. The fact that these two different putative operons (PF0891–PF0894 and PF0895–PF0897.1) show

Table 3 Comparison of the proteins encoded by the putative operons containing PF0895 and PF1090

Gene	MW(kDa)	pI	TMD[a]	Signal sequence[a]	Similarity/ Identity (%)	Other homologs
PF0895	23.1	8.4	0	No	81/61	Yes
PF1090	28	9.2	0	No		Yes
PF0896	15.9	9.2	3	Yes	65/40	No
PF1091	15.9	9.3	3	Yes		No
PF0897	43.2	9.8	10	No	52/32	No
PF1092	44.5	9.4	10	Yes		No
PF0897.1	9.7	10.7	1	Yes	59/34	No
PF1093	9.9	10.3	2	Yes		No

[a] Indicates whether the corresponding protein is predicted to encode a signal sequence and transmembrane domains (TMD), and if so, the number of TMDs

dramatically different regulation under at least one set of growth conditions (peptides versus maltose) is consistent with these being two independently regulated sets of transcriptional units.

P. furiosus therefore contains a novel four-gene operon, PF0895–PF0897.1, where only the putative ATP-binding subunit (PF0895) has any similarity to any genes in other microbial genomes. However, sequence analyses using the proteins encoded by the other three genes, PF0896, PF0897 and PF0897.1, reveal that *P. furiosus* itself contains a very similar complex derived from a paralogous four-gene operon encoded by PF1090–PF1093. The transcriptional data indicate that the expression of the genes within this operon is not regulated (<2-fold change) by the carbon source [20, 37] or by cold-shock [20, 37], and therefore it is regulated independently of the PF0895 operon (data not shown). The corresponding gene properties of the products of the paralogous genes are summarized in Table 3 and Fig. 6. PF0896 and PF0897, and the corresponding PF1091 and PF1092, are all predicted to be integral membrane proteins with recognizable signal sequences (except for PF0897). PF0897.1 and its paralog, PF1093, also are predicted to contain signal sequences. They are extremely basic proteins (pI > 10.5, see Table 3) and could potentially be involved in the recognition of extracellular solutes. As indicated in Table 3, only PF0895 and PF1090 have homologs in other sequenced genomes by BLAST analyses or are part of InterPro families. That these two homologous, multiprotein membrane-bound complexes are seemingly unique to *P. furiosus* is remarkable given that genome sequences of three very closely related species, *P. horikoshii* [59], *P. abyssi* [60] and *Thermococcus kodakaraensis* [61, 62] are available (see www.genomesonline.org). Why orthologs are not present in these organisms, especially given the extremely similar metabolism and physiology of *Pyrococcus* and *Thermococcus* species, all of which are classified as hyperthermophilic archaea, is a fascinating question. Interestingly, another hyperthermophilic archaeon, *P. woesei*, is virtually identical to *P. furiosus* and was recently shown by DNA microarray hybridization analysis to lack close homologs of only a 100 or so genes found in the latter organism [63]. Among those that are present are homologs of seven of the eight genes listed in Table 3 (the array did not contain PF0897.1). Consequently, *P. woesei* is the only organism that appears to contain homologs of this novel *P. furiosus* multiprotein complex.

Acknowledgements The work described in this article was funded in part by the following organizations: the National Institutes of Health (GM62407, GM60329), the Department of Energy (FG05-95ER20175), IBM Life Sciences, the Georgia Research Alliance and the University of Georgia Research Foundation. Data were collected at 22ID, Southeast Regional Collaborative Access Team (SER-CAT, www.ser-cat.org/), Advanced Photon Source, Argonne National Laboratory. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38.

References

1. Janausch IG et al (2002) C4-dicarboxylate carriers and sensors in bacteria. Biochim Biophys Acta 1553(1–2):39–56
2. Kelly DJ, Thomas GH (2001) The tripartite ATP-independent periplasmic (TRAP) transporters of bacteria and archaea. FEMS Microbiol Rev 25(4):405–424
3. Steinberg TH (1994) Cellular transport of drugs. Clin Infect Dis 19(5):916–921
4. Winnen B, Hvorup RN, Saier MH Jr (2003) The tripartite tricarboxylate transporter (TTT) family. Res Microbiol 154(7): 457–465
5. Igarashi K, Ito K, Kashiwagi K (2001) Polyamine uptake systems in *Escherichia coli*. Res Microbiol 152(3–4):271–278
6. Jones PM, George AM (2004) The ABC transporter structure and mechanism: perspectives on recent research. Cell Mol Life Sci 61(6):682–699
7. Koster W (2001) ABC transporter-mediated uptake of iron, siderophores, heme and vitamin B12. Res Microbiol 152 (3–4):291–301
8. Schmitt L, Tampe R (2002) Structure and mechanism of ABC transporters. Curr Opin Struct Biol 12(6):754–760
9. Schneider E (2001) ABC transporters catalyzing carbohydrate uptake. Res Microbiol 152(3–4):303–310
10. Sharma R et al (2003) Energy dependent transport of xenobiotics and its relevance to multidrug resistance. Curr Cancer Drug Targets 3:89–107

11. van der Does C, Tampe R (2004) How do ABC transporters drive transport? Biol Chem 385(10):927–933
12. Albers SV et al (2004) Insights into ABC transport in archaea. J Bioenerg Biomembr 36(1):5–15
13. Fiala G, Stetter KO (1986) *Pyrococcus furiosus* sp-nov represents a novel genus of marine heterotrophic Archaebacteria growing optimally at 100-degrees C. Arch Microbiol 145:56–61
14. Altschul SF et al (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410
15. Mulder NJ et al (2005) InterPro, progress and status in 2005. Nucleic Acids Res 33(Database issue):D201–D205
16. Altschul SF et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25(17):3389–3402
17. Madden TL, Tatusov RL, Zhang J (1996) Applications of network BLAST server. Methods Enzymol 266:131–141
18. Zhang J, Madden TL (1997) PowerBLAST: a new network BLAST application for interactive or automated sequence analysis and annotation. Genome Res 7(6):649–656
19. Sugar FJ et al (2005) Comparison of small- and large-scale expression of selected *Pyrococcus furiosus* genes as an aid to high-throughput protein production. J Struct Funct Genomics 6(2–3):149–158
20. Weinberg MV et al (2005) Cold shock of a hyperthermophilic archaeon: *Pyrococcus furiosus* exhibits multiple responses to a suboptimal growth temperature with a key role for membrane-bound glycoproteins. J Bacteriol 187(1):336–348
21. Studier FW (2005) Protein production by auto-induction in high density shaking cultures. Protein Expr Purif 41(1):207–234
22. Adams MWW et al (2003) The southeast collaboratory for structural genomics: a high-throughput gene to structure factory. Accounts Chem Res 36(3):191–198
23. Poole FL 2nd et al (2005) Defining genes in the genome of the hyperthermophilic archaeon *Pyrococcus furiosus*: implications for all microbial genomes. J Bacteriol 187(21):7325–7332
24. Liu ZJ et al (2005) The high-throughput protein-to-structure pipeline at SECSG. Acta Crystallogr D Biol Crystallogr 61 (Pt 6):679–684
25. Teng TY (1990) Mounting of crystals for macromolecular crystallography in a freestanding thin-film. J Appl Crystallogr 23:387–391
26. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol A276:307–326
27. Terwilliger TC, Berendzen J (1999) Automated MAD and MIR structure solution. Acta Crystallogr D Biol Crystallogr 55 (Pt 4):849–861
28. Bailey S (1994) The CCP4 Suite – programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 50:760–763
29. Perrakis A, Morris R, Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. Nat Struct Biol 6(5):458–463
30. McRee DE (1999) XtalView/Xfit – a versatile program for manipulating atomic coordinates and electron density. J Struct Biol 125(2–3):156–165
31. Murshudov GN (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 53(Pt 3):240–255
32. Lovell SC et al (2003) Structure validation by C alpha geometry: phi, psi and C beta deviation. Proteins Struct Funct Genet 50(3):437–450
33. Laskowski BA et al (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 26:283–291
34. Berman HM et al (2000) The Protein Data Bank. Nucleic Acids Res 28(1):235–242
35. Yang H et al (2004) Automated and accurate deposition of structures solved by X-ray diffraction to the Protein Data Bank. Acta Crystallogr D Biol Crystallogr 60(Pt 10):1833–1839
36. Holden JF et al (2001) Identification of membrane proteins in the hyperthermophilic archaeon *Pyrococcus furiosus* using proteomics and prediction programs. Comp Funct Genomics 2(5):275–288
37. Schut GJ et al (2003) Whole-genome DNA microarray analysis of a hyperthermophile and an archaeon: *Pyrococcus furiosus* grown on carbohydrates or peptides. J Bacteriol 185(13):3935–3947
38. Liu ZJ et al (2005) Parameter-space screening: a powerful tool for high-throughput crystal structure determination. Acta Crystallogr D Biol Crystallogr 61(Pt 5):520–527
39. Holm L, Sander C (1993) Protein structure comparison by alignment of distance matrices. J Mol Biol 233(1):123–138
40. Gerstein M (1992) A resolution-sensitive procedure for comparing protein surfaces and its application to the comparison of antigen-combining sites. Acta Cryst A48:271–276
41. Liu Z et al (2005) Quantitative evaluation of protein-DNA interactions using an optimized knowledge-based potential. Nucleic Acids Res 33(2):546–558
42. Gibrat JF, Madej T, Bryant SH (1996) Surprising similarities in structure comparison. Curr Opin Struct Biol 6(3):377–385
43. Saraste M, Sibbald PR, Wittinghofer A (1990) The P-loop – a common motif in ATP- and GTP-binding proteins. Trends Biochem Sci 15(11):430–434
44. Armougom F et al (2006) Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. Nucleic Acids Res 34(Web Server issue):W604–W608
45. Walker JE et al (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J 1(8):945–951
46. Meng EC et al (2006) Tools for integrated sequence-structure analysis with UCSF Chimera. BMC Bioinformatics 7:339
47. Braibant M et al (1996) Identification of a second *Mycobacterium tuberculosis* gene cluster encoding proteins of an ABC phosphate transporter. FEBS Lett 394(2):206–212
48. Hewitt EW, Lehner PJ (2003) The ABC-transporter signature motif is required for peptide translocation but not peptide binding by TAP. Eur J Immunol 33(2):422–427
49. Fetsch EE, Davidson AL (2002) Vanadate-catalyzed photocleavage of the signature motif of an ATP-binding cassette (ABC) transporter. Proc Natl Acad Sci USA 99(15):9685–9690
50. Smith PC et al (2002) ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer. Mol Cell 10(1):139–149
51. Hopfner KP et al (2000) Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. Cell 101(7):789–800
52. Frelet A, Klein M (2006) Insight in eukaryotic ABC transporter function by mutation analysis. FEBS Lett 580(4):1064–1084
53. Melin P et al (2004) The cystic fibrosis mutation G1349D within the signature motif LSHGH of NBD2 abolishes the activation of CFTR chloride channels by genistein. Biochem Pharmacol 67(12):2187–2196
54. Illing M, Molday LL, Molday RS (1997) The 220-kDa rim protein of retinal rod outer segments is a member of the ABC transporter superfamily. J Biol Chem 272(15):10303–10310
55. Kijas JW et al (2004) Cloning of the canine ABCA4 gene and evaluation in canine cone-rod dystrophies and progressive retinal atrophies. Mol Vis 10:223–232
56. Pettersen EF et al (2004) UCSF Chimera – a visualization system for exploratory research and analysis. J Comput Chem 25(13):1605–1612

57. Verhees CH et al (2003) The unique features of glycolytic pathways in Archaea. Biochem J 375(Pt 2):231–246
58. Ma K, Adams MW (2001) Hydrogenases I and II from *Pyrococcus furiosus*. Methods Enzymol 331:208–216
59. Kawarabayasi Y (2001) Genome of *Pyrococcus horikoshii* OT3. Methods Enzymol 330:124–134
60. Cohen GN et al (2003) An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*. Mol Microbiol 47(6):1495–1512
61. Fukui T et al (2005) Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with Pyrococcus genomes. Genome Res 15(3): 352–363
62. Lecompte O et al (2001) Genome evolution at the genus level: comparison of three complete genomes of hyperthermophilic archaea. Genome Res 11(6):981–993
63. Hamilton-Brehm SD, Schut GJ, Adams MW (2005) Metabolic and evolutionary relationships among *Pyrococcus* species: genetic exchange within a hydrothermal vent environment. J Bacteriol 187(21):7492–7499
64. Matthews BW (1968) Solvent content of protein crystals. J Mol Biol 33(2):491–497
65. Ramachandran GN, Sasisekharan V (1968) Conformation of polypeptides and proteins. Adv Protein Chem 23:283–438
66. Bond CS (2003) TopDraw: a sketchpad for protein structure topology cartoons. Bioinformatics 19(2):311–312

JOURNAL OF BACTERIOLOGY, Nov. 2005, p. 7325–7332 Vol. 187, No. 21
0021-9193/05/$08.00+0 doi:10.1128/JB.187.21.7325–7332.2005

Defining Genes in the Genome of the Hyperthermophilic Archaeon *Pyrococcus furiosus*: Implications for All Microbial Genomes†

Farris L. Poole II, Brian A. Gerwe, Robert C. Hopkins, Gerrit J. Schut, Michael V. Weinberg, Francis E. Jenney, Jr., and Michael W. W. Adams*

Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia 30602-7229

Received 16 May 2005/Accepted 19 August 2005

The original genome annotation of the hyperthermophilic archaeon *Pyrococcus furiosus* contained 2,065 open reading frames (ORFs). The genome was subsequently automatically annotated in two public databases by the Institute for Genomic Research (TIGR) and the National Center for Biotechnology Information (NCBI). Remarkably, more than 500 of the originally annotated ORFs differ in size in the two databases, many very significantly. For example, more than 170 of the predicted proteins differ at their N termini by more than 25 amino acids. Similar discrepancies were observed in the TIGR and NCBI databases with the other archaeal and bacterial genomes examined. In addition, the two databases contain 60 (NCBI) and 221 (TIGR) ORFs not present in the original annotation of *P. furiosus*. In the present study we have experimentally assessed the validity of 88 previously unannotated ORFs. Transcriptional analyses showed that 11 of 61 ORFs examined were expressed in *P. furiosus* when grown at either 95 or 72°C. In addition, 7 of 54 ORFs examined yielded heat-stable recombinant proteins when they were expressed in *Escherichia coli*, although only one of the seven ORFs was expressed in *P. furiosus* under the growth conditions tested. It is concluded that the *P. furiosus* genome contains at least 17 ORFs not previously recognized in the original annotation. This study serves to highlight the discrepancies in the public databases and the problems of accurately defining the number and sizes of ORFs within any microbial genome.

The advent of high-throughput DNA sequencing has brought into focus the problems associated with predicting protein-encoding, open reading frames (ORFs) on a genome-wide basis. The rapid rate at which genomes are now being sequenced has made the practice of manually identifying and annotating ORFs slow and impractical. Widely used sequence databases such as the Comprehensive Microbial Resource (CMR) (28) of the Institute for Genomic Research (TIGR; http://www.tigr.org/tigr-scripts/CMR2/CMRHomePage.spl) and the Reference Sequence (RefSeq) (29) collection of the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/RefSeq/) rely heavily on software, such as Glimmer2, GeneMarkS, and BLAST2, to automatically identify ORFs (3, 6, 12). The majority of ORFs are assigned automatically based on sequence comparisons to ORFs already present in public databases.

It seems reasonable to assume that the different public resources would identify the same set of ORFs in a given genome and that the predicted start site of a given ORF does not depend on the automated analysis that was used. However, as demonstrated herein, these are false assumptions and ones that can have serious consequences. Of course, a genuine ORF in a given genome that shows no sequence similarity to ORFs identified in other genomes tends to be overlooked by automated analyses. This is an important issue because ORFs that are unique to an organism are likely to be at the root of strain- and species-specific differences (19). Conversely, in the absence of anything to compare it with, an ORF that is proposed to be unique to a given genome could also be an artifact of the automated analyses (32). In addition, genome sequencing errors and biological frameshifts could result in the misannotation of ORFs.

Despite these problems, it is imperative that a genome be annotated as accurately as possible, especially given the increasing use of genome-based experimental approaches such as DNA microarrays, protein arrays, and structural genomics. Although the results from the automated analyses in the public databases are frequently assumed to be highly accurate, a plethora of additional bioinformatics tools for ORF prediction continue to be reported, which serve to illustrate that this is not the case as well as highlighting the complexity of the problem (4, 7, 9, 16, 25, 27, 43). Comparative genomics can also play an important role in genome analysis. A prime example is the genome of the eukaryote *Saccharomyces cerevisiae*, which has been analyzed extensively (8, 38). Remarkably, a recent comparison of its genome with those of three related species led to the revision of almost 15% of the more than 6,000 annotated ORFs, including the elimination of almost 500 of them (24). While an accurate description of the ORFs in a genome must be derived from experimental as well as bioinformatic analyses, there are few examples where this has been achieved. One involves the genome of *Mycobacterium tuberculosis*, where six previously unannotated ORFs were identified using a proteomics-based approach (21).

Currently, a total of 22 archaeal genomes have been sequenced (www.ncbi.nlm.nih.gov/genomes/lproks.cgi?view = 1) and at least 30 more are in progress (http://www.genomesonline.org). The archaeal group also provides several examples of two or more

* Corresponding author. Mailing address: Department of Biochemistry and Molecular Biol., Davison Life Sciences Complex, University of Georgia, Athens, GA 30602-7229. Phone: 706 542-2060. Fax: 706 542-0229. E-mail: adams@bmb.uga.edu.

† Supplemental material for this article may be found at http://jb.asm.org/.

genome sequences from species of the same genus, which can be used to provide insight into genome annotation. These include three from species of *Pyrococcus* (10, 22), including *P. furiosus* (31). *P. furiosus* grows optimally near 100°C (13) and is one of the best studied of the hyperthermophilic archaea. For example, it has been the subject of several studies using genomewide DNA microarrays (34, 35, 37, 40) and is also the focus of a structural genomics initiative (1).

The genome of *P. furiosus* is approximately 1.9 Mb in size and the original annotation deposited in GenBank in 2002 contained 2,065 ORFs (PF0001 to PF2065) (31). Subsequently, two automated annotations have been made available by CMR at TIGR and by RefSeq at NCBI. We anticipated that these three annotations would be highly similar, if not identical. However, our analyses show that there are profound differences between them and that this phenomenon is not unique to *P. furiosus*. In addition, we provide experimental evidence that there are many ORFs that were not identified in the original genome annotation that are functional and/or encode stable proteins.

MATERIALS AND METHODS

Bioinformatic analyses. The original annotation of the *P. furiosus* genome (2,065 ORFs) deposited in GenBank in 2002 was obtained from NCBI (http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?18976372:OLD09:199672). The CMR and RefSeq annotations were obtained in January 2004 from TIGR (http://www.tigr.org/tigr-scripts/CMR2/GenomePage3.spl?database = ntpf01) and NCBI (http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?18976372:OLD13:1040042), respectively. All data sets were incorporated into a Microsoft SQL Server 2000 database for indexing and extraction. Ad hoc PERL scripts were used to determine the position and degree of ORF overlap. A Microsoft Visual Basic 6.0 program, based on the Expasy ProtParam Tool (http://www.expasy.org/tools/protparam.html), was written and used to predict the physical properties (molecular weight, isoelectric point, length, extinction coefficient, etc.) of each ORF product (41).

Transmembrane domains and signal sequences were predicted using six web-based programs as previously described (18). InterPro analysis was performed using the InterProScan tool (v. 3.3; http://www.ebi.ac.uk/interpro/), and the PFAM analysis (v. 15; http://www.sanger.ac.uk/Software/Pfam/) was conducted on data from July 2004 as described (5, 42). Ribosome binding sites were predicted using RBSfinder (http://www.tigr.org/software/). Sequence alignments were performed using the BLAST toolkit (v. 2.2.6; ftp://ftp.ncbi.nlm.nih.gov/BLAST/) (3). The TBLASTN and BLASTP programs were run against BLAST databases that were built or downloaded in February 2004 from files on NCBI's FTP site (ftp://ftp.ncbi.nlm.nih.gov). These included nonredundant nucleotide and peptide databases, and the complete genomes of *Pyrococcus horikoshii* and *Pyrococcus abyssi*. In addition, each program was run multiple times using different substitution matrices (PAM30, PAM70, and BLOSUM62), word sizes (2, 7, and 3), and filters (low complexity) in order to compensate for the various sequence lengths.

Growth conditions and microarray analyses. The transcriptional analyses of *P. furiosus* were conducted using PCR-based microarrays. The arrays were generated by spotting full-length PCR products of each gene onto glass slides as previously described by Schut et al. (34). For this study, *P. furiosus* was grown under six different conditions using previously published methods for growing cells, harvesting RNA and performing the DNA microarray analysis (34, 35, 37, 40). Two conditions involved growing *P. furiosus* cells in batch culture. These were using either peptides or maltose as the carbon source at 95°C in medium containing elemental sulfur (34, 40).

The other four growth conditions were generated from cold shock kinetic experiments. The *P. furiosus* cultures were initially grown to mid-log phase with maltose as the carbon source at 95°C in the absence of sulfur. At time zero they were rapidly cooled to 72°C, and samples for RNA analyses were removed at 0, 1, 2, and 5 h postshock (40). The batch experiments were carried out twice with two duplicates for each growth condition and were hybridized to arrays containing two copies of each ORF. This yielded eight data points per ORF per condition. The kinetic experiments were carried out in the same fashion except that the array contained three copies of each ORF, which yielded a total of 12 data points per ORF per condition. Gene expression was assessed using the minimum likely signal intensity (MLSI) value, which was calculated using the equation mean (signal intensity – background intensity) – standard deviation (signal – background).

ORFs were considered to be expressed when the value of the mean (signal intensity minus background intensity) was greater than 2,000 arbitrary fluorescence intensity units (which is twice the detection limit) and an MLSI value above 1,500.

Production of recombinant proteins. For recombinant protein production, all ORFs were cloned and expressed in *Escherichia coli*. Recombinant cells were grown on a 1-liter scale, expression of the target *P. furiosus* ORF was induced, and cells were harvested and broken as described previously (1, 20, 39), except that cell extracts were heated at 80°C for 30 min to precipitate *E. coli* proteins (or unstable recombinant proteins), cooled to 4°C, and then clarified by centrifugation (40,000 × *g*). Recombinant proteins contained an N-terminal hexa-His tag and each was purified using a column (5 ml) of Histrap Ni affinity resin controlled with an AKTA explorer (GE Healthcare, Piscataway, NJ). After applying the cell extract, the column was washed with 5 column volumes of 20 mM phosphate buffer, pH 7.0, containing 500 mM NaCl, 10 mM imidazole, 5% (vol/vol) glycerol, and 2 mM dithiothreitol. The absorbed protein was eluted with a gradient of 0 to 500 mM imidazole over 20 column volumes.

The major protein peak was collected and concentrated to 10 ml by ultrafiltration (Millipore, Bedford, MA), diluted 15-fold in 20 mM Tris buffer, pH 8.0, containing 5% (vol/vol) glycerol and 2 mM dithiothreitol, and then applied to a column (5 ml) of Q Sepharose (GE Healthcare). The column was washed with 5 column volumes of the same buffer, and the bound protein was eluted with a 0 to 2 M NaCl gradient over 20 column volumes. The major protein was concentrated to 1.5 ml and applied to a 16/60 column of Superdex75 (GE Healthcare) equilibrated with the same Tris buffer. The major protein peak was collected and concentrated to a volume of 1 ml by ultrafiltration. The protein concentration was estimated by the absorption at 280 nm using a calculated extinction coefficient (15). The purity of the recombinant protein was determined by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) (4 to 20% Criterion gels, Bio-Rad, Hercules, CA), and its identity was determined by tryptic digestion of the excised band and analysis by matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF-MS). Protein mass was determined by electrospray ionization mass spectrometry at the Department of Chemistry, University of Georgia.

RESULTS AND DISCUSSION

Original annotation of the *P. furiosus* genome. The original annotation deposited in GenBank in 2002 contained 2,065 ORFs, labeled PF0001 to PF2065. Of these, 1,041 were designated as encoding proteins of known function, meaning that they showed some sequence similarity to proteins (or genes) that had been characterized to some extent in other organisms. In addition, 869 were annotated as conserved hypothetical, showing sequence similarity to ORFs of unknown function in the genomes of other organisms. The remaining 155 ORFs were considered unique to *P. furiosus*, having no similarity to any other ORFs at the time of deposition, including those in the closely related genomes of *P. horikoshii* (23) and *P. abyssi* (10). Current analyses of the 2,065 ORFs using the InterPro and PFAM databases reveal that 473 of the 869 conserved hypothetical ORFs are of "known" function, or at least some general functional information is available for them. Similarly, 25 of the 155 unique ORFs are no longer unique to *P. furiosus* and can now be reclassified either as of known function (24 ORFs) or as conserved hypothetical (1 ORF). Thus, at present, of the original 2,065 ORFs in the genome, 1,538 (75%) are known, 397 (19%) are conserved hypothetical, and 130 (6%) are unique to *P. furiosus*.

Prior to the GenBank deposition, a draft of the annotated *P. furiosus* genome was made publicly available (http://comb5-156.umbi.umd.edu/genemate) (31) that contained an additional 127 ORFs. These ORFs were subsequently removed from the an-

TABLE 1. Comparison of three *P. furiosus* annotations[a]

Comparison	Parameter	No. of ORFs		
		CMR	RefSeq	UGa
GenBank	Annotated ORFs	2,261	2,125	2,192
	ORFs identical to GenBank[b]	1,488	1,981	2,065
	ORFs similar to GenBank[c]	552	84	0
	ORFs in common[d]	2,040	2,065	2,065
	ORFs not in GenBank (ExGB ORFs)	221	60	127
	GenBank ORFs discarded in new annotation	25	0	0
ExGB	Identical in all three annotations[b]	23	23	23
	Similar in all three annotations[c]	12	12	12
	In common	35	35	35
	Exclusive (absent in other two annotations)	127	9	45
	Remaining (in one other annotation)	59	16	47

[a] The UGa annotation is the original 2,065 GenBank ORFs plus the 127 UGa ORFs.
[b] Same start and stop sites.
[c] Different start site but same stop site.
[d] ORFs in common with GenBank; the subtotal of identical and similar ORFs.

notation that was deposited in GenBank in 2002 but have been maintained in our own database. The annotation algorithms used at the time of deposition excluded them presumably based on their lack of sequence homologs and/or small size (see below). For the purposes of distinguishing them from ORFs in other databases, they will be referred to as the 127 University of Georgia (UGa) ORFs. Thus, the UGa annotation includes the 2,065 original GenBank ORFs plus another 127 hypothetical (UGa) ORFs, for a total of 2,192 ORFs. All putative ORFs that were not present in the original GenBank annotation of 2,065 ORFs will be referred to as ExGB ORFs (Table 1), and all 127 UGa ORFs fall into this category.

Some properties of the UGa ORFs, including start and stop nucleotides and intergenic distances, are given as supplemental information (Table S1 in the supplemental material). They are numbered with reference to the lowest consecutive ORF in the GenBank annotation (e.g., PF0438) and are designated as PFxxxx0.1 (e.g., PF0438.1, and PF0438.2 if necessary). They are predicted to encode proteins ranging from 14 to 371 amino acids, where the median length is 49 amino acids. Only 13 of the 127 UGa ORFs show sequence similarity (≥80% match length, and E value $\leq 1 \times e^{-06}$) to regions in other genomes. Two of them are only found at the nucleotide level, and the remaining eleven are annotated as ORFs in those other genomes. Of those eleven, six are genus specific, as they are found only in the genomes of *P. abyssi* and *P. horikoshii*. Thus, comparative genomics provides some evidence that eleven of the 127 hypothetical UGa ORFs are true ORFs. These data are presented in the supplementary information (Table S1 in the supplemental material).

Recent annotations of the *P. furiosus* genome. After the annotated *P. furiosus* genome was deposited in GenBank, an automated analysis of the genome sequence was carried out by the CMR of TIGR using Glimmer2 (12) and TIGRFAM (17). This identified a total of 2,261 ORFs, which were numbered sequentially NT01PF0001 to NT01PF2261. Note that there is no correlation between these CMR ORF numbers and those used in the GenBank (PF0001 to PF2065). As shown in Table 1, of the original 2,065 GenBank ORFs, only 2,040 of them were recognized by CMR (such that they have an identical stop site), leaving an additional 221 ExGB ORFs in the CMR annotation. A second automated annotation of the *P. furiosus* genome has also become available, in addition to that of CMR. This was released (on 13 January 2004) by NCBI under the RefSeq project (29) and contained 2,125 ORFs. Of these, 2,065 are in common with those in GenBank (where they have at least an identical stop site). Hence, there are a total of 60 RefSeq ExGB ORFs.

Comparison of the original ORFs in the three genome annotations. The 2,065 ORFs described in the original annotation in GenBank have been widely used in subsequent studies of *P. furiosus*. Assuming that all these ORFs are indeed genuine, are they exactly the same in the subsequent automated annotations, and if not, how do they differ? As already noted, the CMR annotation does not recognize 25 of the 2,065 GenBank ORFs, or more precisely, the stop codons for those 25 ORFs are not so designated in CMR. Sixteen of these 25 discarded ORFs overlap with other CMR ORFs (see Table S2 in the supplemental material). RefSeq has not discarded any of the original GenBank ORFs. The stop sites of the remaining 2,040 ORFs in CMR do match those in GenBank and the same is true for all 2,065 GenBank ORFs identified in the RefSeq annotation (see Fig. S1 in the supplemental material). Where they differ, however, is at their start sites. These differences are likely due to the various prediction methods that are used, which range from automated analyses to more subjective manual selection (6, 12, 26, 28, 29).

Of the 2,065 original GenBank ORFs, there are 84 instances where RefSeq and GenBank disagree on the start nucleotides. A similar analysis between the GenBank and CMR annotations reveals that 552 of the 2,065 ORFs in the *P. furiosus* genome differ in their start nucleotides, and a comparison of the RefSeq to the CMR annotations reveals 589 discrepancies in the start sites. As shown diagrammatically in Fig. 1, in most cases these differences are not a matter of a codon or two. For example, more than 170 of the proteins predicted by CMR and RefSeq differ at the N termini by more than 25 amino acids. It is remarkable that 28% of the 2,065 ORFs differ in their start codons in one or more of the three publicly available annotations of the *P. furiosus* genome. Moreover, this problem does not appear to be unique to *P. furiosus*. For example, as shown in Table 2, the CMR and RefSeq annotations of the genomes of the bacterium *Clostridium perfringens* strain 13 (36) and of the archaeon *Pyrobaculum aerophilum* strain IM2 (14), both of which were chosen at random, show differences in almost 400 and 1,200 ORFs, respectively. Hence, the discrepancies in *P. furiosus* genome annotations illustrate a generic problem.

One way to mitigate differences in annotation is for databases to provide versions numbers, dates, and change logs that can be easily seen by the end user. In addition, when alternative ORFs or start sites are possible, some indication of their potential existence and even likelihood could be presented. Obviously, experimental approaches must be used to validate annotations, at least at the present time. In the case of

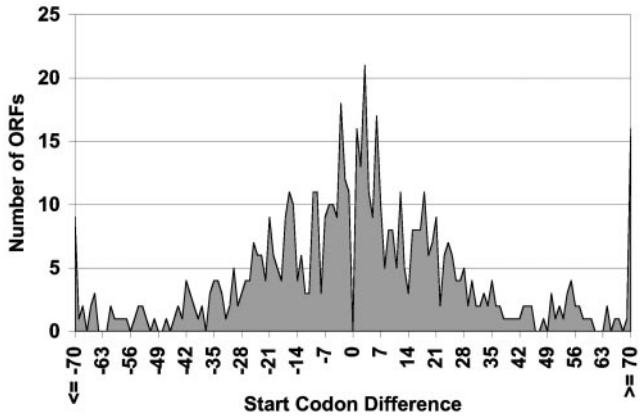


FIG. 1. Database comparisons for the *P. furiosus* genome. Differences are shown in the positions of start codons for the genes in the annotations in the RefSeq and CMR databases. For clarity, the zero value (where the two annotations agree) is not shown.

P. furiosus, we chose to experimentally assess the validity of a subset of the ORFs that were not present in original annotation of the genome.

The ORFs that were not part of the 2,065 ORFs in the original GenBank deposition (ExGB ORFs) comprise 127 UGa ORFs, 221 CMR ORFs, and 60 RefSeq ORFs. As summarized in Table 1, there are 35 ORFs in common. Of these, 23 are identical (having the same start and stop sites) and 12 are similar (having only the same stop site). Sixty-three of the 127 UGa ORFs are identical to CMR ORFs and 24 are identical to RefSeq ORFs (see Fig. S2 in the supplemental material). Conversely, 45 of the 127 UGa ORFs are exclusive to the UGa annotation and are not present in the CMR or RefSeq annotations, which contain 127 and 9 exclusive ORFs, respectively (see Fig. S2 in the supplemental material). Note that an exclusive ORF is defined as having start and stop codon that are not recognized as such in the other annotations, regardless of whether the ORF they define overlaps with any other ORFs in those annotations.

The number of exclusive ORFs underscores the differences between the various annotation programs. Nevertheless, one may conclude that there are potentially a maximum of 277 additional ORFs in the *P. furiosus* genome (Table 1), depending on the annotations that are used. For the purpose of this paper we have chosen to focus our validation efforts on a subset of the UGa ORFs. As will be demonstrated, virtually all of them that appear to be genuine ORFs are also annotated in the RefSeq and CMR versions of the genome.

Experimental evidence for expression of UGa ORFs in *P. furiosus*. DNA microarray analysis was used to investigate the validity of some of the 127 UGa ORFs, which were not recognized in the original GenBank annotation. A total of 61 of the UGa ORFs did not overlap with any of the GenBank ORFs and full-length PCR products were obtained for each of them. These were added to the DNA microarray containing the 2,065 GenBank ORFs (34), which were used to assess their expression in *P. furiosus* in cells grown under six different conditions, where each condition yielded between 8 and 12 data points per ORF. These conditions are important because they are known to cause major changes in the expression levels of a large number of genes, which may include the UGa ORFs. Note that the DNA arrays are used here to assess absolute gene expression rather than relative changes in gene expression.

Since the raw fluorescence intensities can be highly variable, a minimum likely signal intensity (MLSI) value was calculated (see Materials and Methods) where ORFs with values above 1,500 were considered to be expressed at a significant level. By this criterion, of the 61 UGa ORFs examined, 11 are expressed in one or more of the six growth conditions. In fact, as shown in Table 3, five of the 11 were expressed under more than one growth condition. Many of the 61 ORFs examined that are apparently not expressed are less than 150 bps in length (see Table S1 in the supplemental material), raising the possibility that they are expressed but that the corresponding cDNA is simply beyond detection under the conditions of the microarray experiment. Nevertheless, it does appear that 11 of the 61 UGa ORFs examined are expressed at sufficient levels to allow detection.

Properties of 11 ORFs expressed in *P. furiosus*. Some sequence properties of the 11 newly recognized ORFs are summarized in Table 4. They vary in size from 35 to 134 amino acids (4.0 to 15.4 kDa). Two of the 11 are predicted to have at least one transmembrane domain and a signal sequence, suggesting that they may be membrane associated or extracellular (data not shown). None of the 11 ORFs have significant se-

TABLE 2. Comparison of automated annotations for the genomes of two other organisms

Organism (reference)	Parameter	No. of ORFs	
		RefSeq (January 2005)	TIGR (April 2005)
Clostridium perfringens 13 (36)	Total ORFs	2,660	2,841
	Identical	2,259	2,259
	Similar	387	387
	TIGR not in RefSeq		195
	RefSeq not in TIGR	14	
Pyrobaculum aerophilum IM2 (14)		RefSeq (October 2004)	TIGR (April 2005)
	Total ORFs	2,605	2,995
	Identical	1,353	1,353
	Similar	1,209	1,209
	TIGR not in RefSeq		433
	RefSeq not in TIGR	43	

TABLE 3. Transcriptional analysis of 11 previously unannotated (UGa) ORFs

ORF	MLSI value[a]					
	Batch PS 95°C	Batch MS 95°C	Kinetic 0 h 72°C	Kinetic 1 h 72°C	Kinetic 2 h 72°C	Kinetic 5 h 72°C
PF0016.1	431	546	0	0	0	**2,048**
PF0031.1[b]	**1,597**	**1,832**	14	**2,413**	672	**1,956**
PF0437.1	0	0	423	1,117	47	**1,730**
PF0653.1	**1,781**	1,205	0	0	0	0
PF0670.1[b]	457	360	1,414	**1,764**	1,481	**1,508**
PF0736.1[b]	0	0	**10,350**	**6,915**	**7,520**	**6,372**
PF0897.1[b]	**1,575**	**3,569**	**2,188**	875	849	1,314
PF1112.1	397	493	**2,786**	173	0	799
PF1461.1	324	957	303	1,257	1,400	**1,712**
PF1735.2[b]	0	**9,613**	**1,765**	0	0	660
PF1755.1	**1,668**	1,002	279	227	0	935

[a] The growth conditions were as described in Materials and Methods. Abbreviations: PS, peptides plus sulfur; MS, maltose plus sulfur. The carbon source for the 72°C cultures was maltose (in the absence of sulfur).
[b] Indicates ORFs that were expressed under multiple microarray conditions. Boldfacing indicates expression 1.5-fold above the detection limit (see Materials and Methods). All MLSI values less than zero were set to zero for illustrative purposes.

quence similarity to ORFs in other genomes. All of them have a typical start codon (ATG or GTG), and six have a predicted bacterial-type ribosome binding site within 15 nucleotides upstream from the start codon (30). All but one of the 11 ORFs is also present in the CMR annotation (Table 4), eight of which are identical. The other three (PF0031.1, PF1112.1, and PF1735.2) differ in their start nucleotides. In RefSeq, five of the 11 ORFs have identical start and stop codons, and two have the same stop but a different start codon.

In *E. coli* it is known that consecutive ORFs on the same strand with small intergenic distances (~18 nucleotides or less) are likely to be part of an operon (33). Most of the 11 new ORFs are well separated from any neighboring ORFs by at least 50 nucleotides, but there are notable exceptions. One is PF0897.1 which is located only five nucleotides upstream of a possible ATP-binding cassette (ABC) transporter (11). As shown in Fig. 2, this comprises PF0895, the putative transporter, and two hypothetical proteins, PF0896 and PF0897, that are each predicted to contain ten transmembrane domains. PF0897.1 is predicted to have a signal sequence (data not shown). Interestingly, *P. furiosus* contains paralogues of all four of these genes in another possible ABC transporter (PF1090 to PF1093, see Fig. 2), where PF0897.1 shows high sequence similarity to PF1093.

Recombinant protein production using *P. furiosus* ORFs. In addition to expression in the native organism, it should be possible to obtain an indication of the validity of an ORF assignment by attempting to obtain a stable recombinant form of the protein that it encodes. It would seem unlikely that an incorrect ORF would generate an artifactual protein that would be resistant to proteolysis by the host organism. It seems even less likely that an artifactual protein would fold into a thermally stable protein of the type found in a hyperthermophilic organism, such as *P. furiosus*. Therefore, any putative ORF in the *P. furiosus* genome that encoded a heat-stable recombinant protein was assumed to be genuine. Thus, a total of 54 out of the 127 UGa ORFs were cloned, and attempts were made to express them in *E. coli* and characterize the recombinant proteins. These included all 11 of the UGa ORFs discussed above that are expressed in *P. furiosus*, as determined by the microarray analysis (see Table 3). They also included 27 of the 66 UGa ORFs that were not investigated by the microarray approach because they overlapped an original ORF.

Of the 54 UGa ORFs investigated, seven of them yielded detectable amounts of recombinant protein on an SDS-PAGE gel after the Ni affinity purification step (data not shown). The 47 UGa ORFs that did not yield protein products included 10

TABLE 4. Properties of the 11 previously unannotated ORFs expressed in *P. furiosus*

ORF	Start codon	Ribosome-binding site (nucleotides)	Protein length (amino acids)	Protein Mass (Da)	pI	CMR match	RefSeq match
PF0016.1	ATG	None	65	7,652	8.0	NT01PF0021	None
PF0031.1[a]	ATG	GGGAG(−14)	57	7,454	3.6	None	None
PF0437.1	ATG	TTGTG(−12)	83	9,442	5.7	NT01PF0489	PF0437.1n
PF0653.1	GTG	TGGAG(−13)	35	4,026	4.2	NT01PF0731	None
PF0670.1[a]	ATG	AGGTG(−10)	67	7,431	9.3	NT01PF0752	PF0670.1n
PF0736.1[a]	TTG	None	64	6,274	3.9	NT01PF0828	PF0736.1n
PF0897.1[a]	GTG	AGGAG(−15)	87	9,772	10.8	NT01PF1018	PF0897.1n
PF1112.1	ATG	None	47	5,568	8.2	NT01PF1254[b]	PF1112.1n[b]
PF1461.1	ATG	AGGGG(−13)	52	6,198	4.6	NT01PF1618	None
PF1735.2[a]	ATG	None	134	15,403	9.4	NT01PF1904[b]	PF1735.2n[b]
PF1755.1	ATG	None	61	6,872	8.3	NT01PF1925	PF1755.1n

[a] Indicates ORFs that were expressed under multiple growth conditions.
[b] Indicates the matching ORF is similar but not identical to the UGa ORF (with the same stop codon but a different start codon).

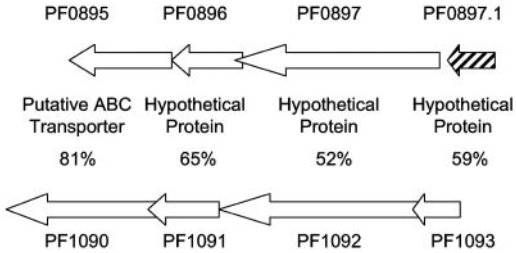


FIG. 2. ORF PF0897.1 is potentially part of an operon. A potential ABC transporter that is closely associated with a previously unannotated (UGa) ORF (striped) is aligned with part of a known *P. furiosus* ABC transporter. Both are within 16 nucleotides of the neighboring ORF. Overlapping ORFs are indicated by the overlapping nature of the arrows. PF numbers and functional annotations are given to those ORFs that have been previously annotated by GenBank. Percent similarity is indicated by the numbers between each gene pair.

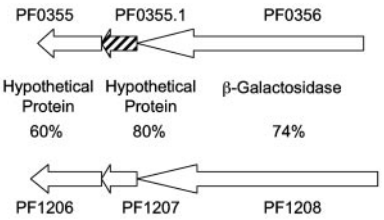


FIG. 3. ORF PF0355.1 is potentially part of an operon. A potential oligosaccharide-related operon with a previously unannotated (UGa) ORF (striped) is aligned with another conserved oligosaccharide-related operon in *P. furiosus*. See the legend to Fig. 2 for more details.

that were analyzed by the DNA microarray (none of which were expressed in *P. furiosus*). Of course, one cannot draw any conclusion about the validity of an ORF from the absence of recombinant protein, as this is frequently the case with ORFs that encode well-characterized proteins (1). Conversely, the production of heat-stable recombinant proteins would strongly suggest that the seven new (UGa) ORFs are genuine and that they are expressed by *P. furiosus*.

All seven recombinant proteins were readily purified from heat-treated *E. coli* cell extracts by multistep chromatography, although all but one (PF0706.1) yielded multiple bands after SDS-PAGE analysis (data not shown). Each major band was excised, digested with trypsin and analyzed by MALDI-TOF-MS, which confirmed the identity of each band. Moreover, analysis by electrospray ionization mass spectrometry confirmed that all but one of the recombinant proteins (PF0712.1) had the predicted mass (Table 5). PF0712.1 is smaller than the predicted mass of 10,078 Da by 920 Da, indicating that the protein may be proteolytically degraded in *E. coli*. Hence, all seven recombinant proteins are stable and can be purified from *E. coli* and in all but one case are of the expected size.

The CMR annotation contained all but one (PF0712.1) of the seven UGA ORFs that yielded stable recombinant protein. However, as shown in Table 5, only four of six ORFs that CMR does contain have identical start and stop sites, while the remaining two have a different start site. The RefSeq annotation contains five of the seven ORFs, four of which have identical start and stop sites. Two of the seven ORFs also have homologs in *P. abyssi* and *P. horikoshii*, further confirming the validity of the annotation. All seven are predicted to encode cytoplasmic proteins, and all but one of them are predicted to encode acidic proteins (Table 5). As shown in Fig. 3, one of the seven ORFs, PF0355.1, appears to be part of a three ORF operon because of its close proximity to two previously annotated ORFs. Interestingly, *P. furiosus* also contains a three-ORF paralog of this putative operon, including an ORF with a high similarity to PF0355.1 (Fig. 3).

Conclusions. Of the 127 previously unannotated ORFs in the *P. furiosus* genome, 61 were examined for expression in *P. furiosus* via DNA microarray analysis, and attempts were made to express 54 of them in *E. coli* (27 of the 54 were examined by both approaches). A total of 17 of the 88 ORFs examined (61 by array and 27 by expression) are likely to be functional in *P. furiosus*. Seven of them encode heat-stable proteins that could be obtained in a recombinant form, and 11 are expressed in *P. furiosus* under one or more growth conditions (which includes only one of the UGa ORFs that also generates a stable recombinant protein even though expression of all them was attempted). The remaining 39 of the 127 UGa ORFs were not studied due to overlapping sequences and lack of resources. Seventeen previously unannotated ORFs can therefore be added to the 2065 ORFs in original annotation of this genome.

Of considerable concern, however, is the fact that more than 25% of these 2,065 ORFs have ambiguous start sites according to the public databases. Primer extension and quantitative PCR experiments will be required to ascertain the true nature of these ORFs in order to provide a more accurate and complete picture of the *P. furiosus* proteome. Knowledge of correct start sites is essential for understanding gene overlap and regulation, and for practical concerns such as identification of natively purified proteins by N-terminal amino acid sequenc-

TABLE 5. Properties of the seven previously unannotated ORFs that yield stable recombinant proteins

ORF	Protein length (amino acids)	pI	Protein mass (Da)	CMR match	RefSeq match	*P. abyssi* homolog	*P. horikoshii* homolog
PF0355.1	77	8.0	9,113	NT01PF0397	PF0355.1n	None	None
PF0706.1	76	6.1	8,350	NT01PF0787[b]	PF0706.1n[b]	None	None
PF0712.1	93	8.4	10,078	None	None	None	None
PF1214.2	92	6.2	10,946	NT01PF1364	PF1214.2[b]	None	None
PF1461.1[a]	60	5.6	7,105	NT01PF1618	None	None	None
PF2047.1	79	6.3	9,397	NT01PF2240[b]	PF2047.1n	PAB0037.2n	PH0065.2n
PF2048.1	79	6.1	9,161	NT01PF2242	PF2048.1n	PAB0037.1n	PH0065.1n

[a] Indicates ORFs that were expressed in *P. furiosus*.
[b] Indicates the matching ORF is similar but not identical to the UGa ORF (with the same stop codon but a different start codon).

ing, as well as heterologous protein production. Addition of excess residues to the N terminus may well affect protein folding and cause proteolytic degradation by the host. What cannot be overemphasized is that this phenomenon is not unique to *P. furiosus*, as shown by cursory analyses of the genomes of both *Clostridium perfringens* strain 13 (36) and *Pyrobaculum aerophilum* strain IM2 (14), which indicate similarly large differences in start sites between annotations, as well as different numbers of genes (compare Tables 1 and 2).

There can be no doubt that all genome annotations need to be examined carefully, both for the accuracy of existing ORFs, and for the possibility of currently unannotated ORFs. Confirming any conclusions experimentally represents a major challenge particularly on a genome-wide scale, and particularly for ORFs exclusive to a given genome, many of which are likely to be currently unannotated. Adding to this burden is the constant discovery of novel ORFs, often exclusive to one genome, by the plethora of different automated procedures that are being developed. Moreover, currently unannotated ORFs are potentially of immediate biological significance. For example, we show here that the expression of one of the 17 novel *P. furiosus* ORFs is dramatically up-regulated by maltose, and its protein product presumably plays a key role in maltose metabolism, a process that would appear to be well understood. Thus, as more and more genomes become sequenced, it is imperative that ORFs such as this not be lost in the ever-expanding world of sequence space, particularly since these ORFs may well be the very essence of species individuality and represent novel biochemistry.

ACKNOWLEDGMENTS

We thank James F. Holden, Matthew S. Eckman, and William A. Lancaster for their assistance throughout the course of this study and Angeli Lal Menon, Frank J. Sugar, Jerry Eichler, and Dawei Lin for their many helpful discussions.

This work was supported by grants from the National Institutes of Health (GM62407), the Department of Energy (DE-FG02-05ER15710), and the National Science Foundation (MCB 0129841 and BES-0317911).

REFERENCES

1. **Adams, M. W. W., H. A. Dailey, L. J. Delucas, M. Luo, J. H. Prestegard, J. P. Rose, and B. C. Wang.** 2003. The Southeast Collaboratory for Structural Genomics: A high-throughput gene to structure factory. Accounts Chem. Res. **36:**191–198.
2. **Albers, S. V., S. M. Koning, W. N. Konings, and A. J. Driessen.** 2004. Insights into ABC transport in archaea. J. Bioenerg. Biomembr. **36:**5–15.
3. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
4. **Azad, R. K., and M. Borodovsky.** 2004. Probabilistic methods of identifying genes in prokaryotic genomes: connections to the HMM theory. Brief Bioinform. **5:**118–130.
5. **Bateman, A., E. Birney, L. Cerruti, R. Durbin, L. Etwiller, S. R. Eddy, S. Griffiths-Jones, K. L. Howe, M. Marshall, and E. L. Sonnhammer.** 2002. The Pfam protein families database. Nucleic Acids Res. **30:**276–280.
6. **Besemer, J., A. Lomsadze, and M. Borodovsky.** 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Research. **29:**2607–2618.
7. **Bocs, S., S. Cruveiller, D. Vallenet, G. Nuel, and C. Medigue.** 2003. AMIGene: annotation of microbial genes. Nucleic Acids Res. **31:**3723–3726.
8. **Brachat, S., F. S. Dietrich, S. Voegeli, Z. H. Zhang, L. Stuart, A. Lerch, K. Gates, T. Gaffney, and P. Philippsen.** 2003. Reinvestigation of the *Saccharomyces cerevisiae* genome annotation by comparison to the genome of a related fungus: *Ashbya gossypii*. Genome Biol. **4:**R45.
9. **Chen, L. L., and C. T. Zhang.** 2003. Gene recognition from questionable ORFs in bacterial and archaeal genomes. J. Biomol. Struct. Dyn. **21:**99–109.
10. **Cohen, G. N., V. Barbe, D. Flament, M. Galperin, R. Heilig, O. Lecompte, O. Poch, D. Prieur, J. Querellou, R. Ripp, J. C. Thierry, J. Van der Oost, J. Weissenbach, Y. Zivanovic, and P. Forterre.** 2003. An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*. Mol. Microbiol. **47:**1495–1512.
11. **Davidson, A. L., and J. Chen.** 2004. ATP-binding cassette transporters in bacteria. Annu. Rev. Biochem. **73:**241–268.
12. **Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg.** 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. **27:**4636–4641.
13. **Fiala, G., and K. O. Stetter.** 1986. *Pyrococcus furiosus* sp-nov represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100-degrees C. Arch. Microbiol. **145:**56–61.
14. **Fitz-Gibbon, S. T., H. Ladner, U. J. Kim, K. O. Stetter, M. I. Simon, and J. H. Miller.** 2002. Genome sequence of the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. Proc. Natl. Acad. Sci. USA **99:**984–989.
15. **Gill, S. C., and P. H. von Hippel.** 1989. Calculation of protein extinction coefficients from amino acid sequence data. Anal. Biochem. **182:**319–326.
16. **Guo, F. B., H. Y. Ou, and C. T. Zhang.** 2003. ZCURVE: a new system for recognizing protein-coding genes in bacterial and archaeal genomes. Nucleic Acids Res. **31:**1780–1789.
17. **Haft, D. H., B. J. Loftus, D. L. Richardson, F. Yang, J. A. Eisen, I. T. Paulsen, and O. White.** 2001. TIGRFAMs: a protein family resource for the functional identification of proteins. Nucleic Acids Res. **29:**41–43.
18. **Holden, J. F., F. L. Poole, S. L. Tollaksen, C. S. Giometti, H. Lim, J. R. Yates, and M. W. W. Adams.** 2001. Identification of membrane proteins in the hyperthermophilic archaeon *Pyrococcus furiosus* using proteomics and prediction programs. Comp. Functional Genomics **2:**275–288.
19. **Iliopoulos, I., S. Tsoka, M. Andrade, P. Janssen, B. Audit, A. Tramontano, A. Valencia, C. Leroy, C. Sander, and C. Ouzounis.** 2000. Genome sequences and great expectations. Genome Biol. **2:**interactions0001.1–interactions0001.3.
20. **Jenney, F. E., Jr., P. S. Brereton, M. Izumi, F. L. Poole, 2nd, C. Shah, F. J. Sugar, H. S. Lee, and M. W. Adams.** 2005. High-throughput production of *Pyrococcus furiosus* proteins: considerations for metalloproteins. J. Synchrotron Radiat. **12:**8–12.
21. **Jungblut, P. R., E. C. Muller, J. Mattow, and S. H. E. Kaufmann.** 2001. Proteomics reveals open reading frames in *Mycobacterium tuberculosis* H37Rv not predicted by genomics. Infect. Immun. **69:**5905–5907.
22. **Kawarabayasi, Y.** 2001. Genome of *Pyrococcus horikoshii* OT3. Methods Enzymol. **330:**124–134.
23. **Kawarabayasi, Y., M. Sawada, H. Horikawa, Y. Haikawa, Y. Hino, S. Yamamoto, M. Sekine, S. Baba, H. Kosugi, A. Hosoyama, Y. Nagai, M. Sakai, K. Ogura, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Ohfuku, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, and H. Kikuchi.** 1998. Complete sequence and gene organization of the genome of a hyperthermophilic archaebacterium, *Pyrococcus horikoshii* OT3. DNA Res. **5:**55–76.
24. **Kellis, M., N. Patterson, M. Endrizzi, B. Birren, and E. S. Lander.** 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature **423:**241–254.
25. **Larsen, T. S., and A. Krogh.** 2003. EasyGene–a prokaryotic gene finder that ranks ORFs by statistical significance. BMC Bioinformatics **4:**21.
26. **National Library of Medicine.** October 2002. The NCBI handbook. National Library of Medicine, Bethesda, Md.
27. **Ouyang, Z., H. Zhu, J. Wang, and Z. S. She.** 2004. Multivariate entropy distance method for prokaryotic gene identification. J. Bioinform. Comput. Biol. **2:**353–373.
28. **Peterson, J. D., L. A. Umayam, T. Dickinson, E. K. Hickey, and O. White.** 2001. The comprehensive microbial resource. Nucleic Acids Res. **29:**123–125.
29. **Pruitt, K. D., and D. R. Maglott.** 2001. RefSeq and LocusLink: NCBI gene-centered resources. Nucleic Acids Res. **29:**137–140.
30. **Ringquist, S., S. Shinedling, D. Barrick, L. Green, J. Binkley, G. D. Stormo, and L. Gold.** 1992. Translation initiation in *Escherichia-coli*-sequences within the ribosome-binding site. Mol. Microbiol. **6:**1219–1229.
31. **Robb, F. T., D. L. Maeder, J. R. Brown, J. DiRuggiero, M. D. Stump, R. K. Yeh, R. B. Weiss, and D. M. Dunn.** 2001. Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: implications for physiology and enzymology. Methods Enzymol. **330:**134–157.
32. **Rudd, K. E., I. Humphery-Smith, V. C. Wasinger, and A. Bairoch.** 1998. Low molecular weight proteins: A challenge for post-genomic research. Electrophoresis **19:**536–544.
33. **Salgado, H., G. Moreno-Hagelsieb, T. F. Smith, and J. Collado-Vides.** 2000. Operons in *Escherichia coli*: genomic analyses and predictions. Proc. Natl. Acad. Sci. USA **97:**6652–6657.
34. **Schut, G. J., S. D. Brehm, S. Datta, and M. W. Adams.** 2003. Whole-genome DNA microarray analysis of a hyperthermophile and an archaeon: *Pyrococcus furiosus* grown on carbohydrates or peptides. J. Bacteriol. **185:**3935–3947.
35. **Schut, G. J., J. Zhou, and M. W. Adams.** 2001. DNA microarray analysis of the hyperthermophilic archaeon *Pyrococcus furiosus*: evidence for a new type of sulfur-reducing enzyme complex. J. Bacteriol. **183:**7027–7036.
36. **Shimizu, T., K. Ohtani, H. Hirakawa, K. Ohshima, A. Yamashita, T. Shiba, N. Ogasawara, M. Hattori, S. Kuhara, and H. Hayashi.** 2002. Complete

genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. Proc. Natl. Acad. Sci. USA **99:**996–1001.
37. **Shockley, K. R., D. E. Ward, S. R. Chhabra, S. B. Conners, C. I. Montero, and R. M. Kelly.** 2003. Heat shock response by the hyperthermophilic archaeon *Pyrococcus furiosus*. Appl. Environ. Microbiol. **69:**2365–2371.
38. **Snyder, M., and M. Gerstein.** 2003. Genomics. Defining genes in the genomics era. Science **300:**258–260.
39. **Sugar, F. J., F. E. Jenney, Jr., F. L. Poole, 2nd, P. S. Brereton, M. Izumi, C. Shah, and M. W. Adams.** 2005. Comparison of small- and large-scale expression of selected *Pyrococcus furiosus* genes as an aid to high-throughput protein production. J. Struct. Funct. Genomics, in press.
40. **Weinberg, M. V., G. J. Schut, S. Brehm, S. Datta, and M. W. Adams.** 2005. Cold shock of a hyperthermophilic archaeon: *Pyrococcus furiosus* exhibits multiple responses to a suboptimal growth temperature with a key role for membrane-bound glycoproteins. J. Bacteriol. **187:**336–348.
41. **Wilkins, M. R., E. Gasteiger, A. Bairoch, J. C. Sanchez, K. L. Williams, R. D. Appel, and D. F. Hochstrasser.** 1999. Protein identification and analysis tools in the ExPASy server. Methods Mol. Biol. **112:**531–552.
42. **Zdobnov, E. M., and R. Apweiler.** 2001. InterProScan–an integration platform for the signature-recognition methods in InterPro. Bioinformatics **17:** 847–848.
43. **Zhu, H. Q., G. Q. Hu, Z. Q. Ouyang, J. Wang, and Z. S. She.** 2004. Accuracy improvement for identifying translation initiation sites in microbial genomes. Bioinformatics **20:**3308–3317.