January 2009

Expression Analysis and Stem Cell Engineering

Cara Rieger
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS
School of Engineering and Applied Science
Department of Biomedical Engineering

Dissertation Examination Committee:
David Gottlieb, co-chair
Robi Mitra, co-chair
Michael Brent
Donald Elbert
Jason Mills
Shelly Sakiyama-Elbert

EXPRESSION ANALYSIS AND STEM CELL ENGINEERING

by

Cara Rachel Rieger

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

December 2009
Saint Louis, Missouri

ABSTRACT OF THE DISSERTATION

Expression Analysis and Stem Cell Engineering

by

Cara Rachel Rieger

Doctor of Philosphy in Biomedical Engineering

Washington University in St. Louis, 2009

Research co-advisors: Professor David Gottlieb and Professor Robi Mitra

The overall goal of the thesis was to develop tools to advance investigations of stem cells in tissue engineering therapeutics for neurodegenerative disease and spinal cord injury. Two tools to characterize cell fate and a tool to separate a subset of neural cells were developed and evaluated. In the first study, a digital PCR technology, called polonies, was applied to measure mRNA from several key stem cell genes in small numbers of ES cells. Due to its properties, we hypothesized that polonies would be uniquely poised to profile stem cells. Polonies were counted for Oct3 in a sample of 10 ES cells and from three pluripotency genes: Oct3, Nanog, and Rex1 from a single blastocyst, containing 30 ICM cells. The polony method is sensitive, can be applied to most genes, and allows for a degree of multiplexing. Second, DNA methylation was explored as a tool to measure cell fate as ES cells are differentiated into neural cells. We tested the hypothesis that promoter DNA

methylation correlates to gene silencing. Promoter methylation of a pluripotency and neural fate determining genes in ES cells, ES derived neural cells, and non-neural tissues was measured by direct bisulfite sequencing. As expected Oct3, was methylated in differentiated cells and tissues. Unexpectedly, neural genes: Sox1, Olig1, and Olig2 were unmethylated in non-neural cells and tissues. The correlation between methylation and silencing was not universal; it was gene specific. In the third study, ES cells were genetically engineered to permit drug selection of subset of ES derived neural cells. We hypothesized that engineering ES cells with puromycin acetyltransferase gene (PAC) under the control of the Olig2 promoter would allow for selection of Olig2 expressing neural cells with puromycin. Two targeted ES cell lines were generated with PAC inserted into the Olig2 gene. Both lines have the expected functional properties and enable purification of Olig2 expressing cells from ES derived neural cells by puromycin selection. Overall, the tools developed in this thesis are a small step toward generating well-defined cell populations from stem cells needed to advance tissue engineering therapeutics for neurodegenerative disease and injury.

Acknowledgments

I would like to thank my advisor, David Gottlieb. Thank you for your patience and guidance through the completion of this work. The lessons that I have learned in becoming an experimentalist will stay with me: to have a solid endpoint, to trust the data and to be critical of it at the same time, and that sometimes nature intervenes. From you, I have learned the value of basic science research and to appreciate how much biology is still unknown. Thank you for giving me the opportunity to be a "cell engineer."

I would like to thank my co-advisor Rob Mitra for his contribution to my engineering training. Your expertise in nucleic acid technology development and genomics contributed greatly to the polony and methylation projects.

I would like to thank my thesis committee members: Michael Brent, Donald Elbert, Jason Mills, and Shelly Sakiyama-Elbert for taking the time to advise me through this process and review this work. I would especially like to thank Shelly Sakiyama-Elbert for her collaboration in the cell engineering project and guidance as a fellow tissue engineer. Throughout my life I have been fortunate to have excellent teachers, coaches, and role models that have helped me to succeed. I count all of you among them, and I hope I can follow your lead and mentor others.

I would like to thank my colleagues and co-workers who provided support and camaraderie which made research productive and enjoyable. I am thankful for everyone on the ninth floor of McDonnell Sciences who welcomed me, even as a BME student. I am grateful for the friendships that I have made with members of the Gottlieb Lab past and present: Renea Poppino, Xiaodong Zhang, Scott Horrell, Deany Delaney, Julia Kuhn, David Lorberbaum, and Beryl Ojwang. I valued the opinions of Mitra Lab members: K.T. Varley, Lee Tessler, and Francesco Vallarina. Members of the Sakiyama-Elbert Lab: Nicole Moore, Nithya Jesuraj, Dylan McCreedy, and (now) Hao Xu always provide a good reason to travel to the other campus.

Throughout my time in St. Louis, my teammates kept me sane. Thank you, Brentwood volleyball, Clayton women's basketball, BME kickball, basketball, and broomball teammates.

By making me laugh and making sure that I eat, Megan Kaneda and Christina Ambrosi, helped me to enjoy graduate school and time in St. Louis. As fellow BME graduate students, running partners, and knitting buddies, I would like to thank them for their friendship.

Most importantly, I would like to thank my family: my parents, Terry and Nancy Rieger, and my sister, Sharon Rieger. They have made sacrifices and provided the support and love throughout my life that have made this possible. I met Matt Silverman, the love of my life, just before starting graduate school. Matt and his family have become a part of my family and a source of humor and encouragement over the last five years.

I never would have finished the marathons and this five and a half year marathon without all of you.

Cara Rachel Rieger

Washington University in St. Louis
December 2009

For Matt & our future together

Contents

List of Tables

List of Figures

List of Abbreviations

DNA - Deoxyribonucleic acid
cDNA - Complementary DNA
RNA -Ribonucleic acid
mRNA -Messenger RNA
RT - Reverse Transcriptase Reaction
PCR – Polymerase Chain Reaction
RT-PCR - Reverse Transcription followed by PCR
ES cell - Embryonic Stem Cell
iPS cell - Induced Pluripotent Cell
ICM -Inner Cell Mass
TE -Trophectoderm
C -Cytosine
T -Thymine
G -Guanine
A -Adenine
Poly(A+) – Polyadenylated, chain of many adenines
Oligo(dT) – chain of many thymines
CpG - Cytosine joined by a phosphodiester bond to a Guanine
GFP - Green fluorescent protein
PAC – Puromycin N-acetyltransferase
FACS - Fluorescence activated cell sorting

Chapter 1

Introduction

Harnessing the promise of stem cells holds the potential to revolutionize modern medicine. Many devastating diseases are a result of cellular deficiencies. Stem cells, as the originators of all the tissues of the body, are a viable source of therapeutic replacement cells and a potential treatment for these diseases. Neurodegenerative disease and spinal cord injury are of particular interest because cell loss results in devastating loss of function for which no effective treatments currently exist.

Development of cellular replacement therapeutics for neurodegeneration and spinal cord injury is a focus of tissue engineering research. Tissue engineering combines therapeutic cells with instructive growth factors and biomaterial scaffolds to provide an organized method to deliver therapeutic cells. The ability of stem cells to self-renew and differentiate into all cell types gives great hope for their use as the cellular component in tissue engineering based therapeutics. One of the challenges with this process is to define and limit the heterogeneity of stem cells and differentiated cells to the desired cell type. To achieve this, tools to investigate and limit stem cell heterogeneity must be developed.

This thesis focuses on the development of tools to further the use of stem cells as part of tissue engineering approaches for neurodegenerative disease and spinal cord injury. The first study explores a new method to characterize stem cell heterogeneity. The second study investigates an assay for monitoring differentiation from ES cell to neural

cell fate. The final study builds a tool that will aid in limiting cellular heterogeneity enabling isolation and delivery of a defined set of ES derived neural cells. These tools were developed with emphasis on the use of stem cells in tissue engineering approaches for treatment of neurodegenerative disease and injury, but will have general implications for the use of stem cells in other applications.

The introduction will cover several topics encompassing the broad scope of the thesis. First, a primer on the biology of embryonic stem cells and benefits of their use for tissue engineering will be provided. Next, research in tissue engineering approaches for SCI using ES derived neural cells, instructive growth factors, and biomaterial scaffolds will be reviewed. Challenges encountered due to cellular heterogeneity will be described which provide the rationale for experimental work. Finally, tools for discovering and limiting cellular heterogeneity will be described.

1.1 Embryonic Stem Cells

Embryonic stem cells are ideal cells for tissue engineering applications because they can be grown on a large scale, differentiated into the needed lineage, and be genetically engineered to select for therapeutic cell type(s). The promise of embryonic stem (ES) cells as a renewable source of therapeutic cells is based on their fundamental ability to self-renew and differentiate into all cell types of the body. This is important for tissue engineering research and applications which require "therapeutic cells" at a clinically relevant scale [$50x10^6$-$5x10^9$] (Palsson and Bhatia 2004).

1.1.1 Mouse Embryonic Stem Cells

Mouse embryonic stem cells were discovered nearly thirty years ago (Evans and Kaufman 1981; Martin 1981). ES cells are isolated from the inner cell mass of the embryonic day 3.5 blastocyst. Signaling molecules, including leukemia inhibitory factor (LIF,(Williams, Hilton et al. 1988)), are added in vitro to instruct daughter cells to remain undifferentiated and divide. ES cells are pluripotent. The differentiation potential of ES cells in vitro directly reflects their role in the embryo as the founder of all three germ layers: ectoderm, mesoderm, and endoderm which later form all somatic tissues in the adult animal (Rossant 2007). Years of research on mammalian embryonic development using mice and mouse ES cells as a model system, has given strong evidence that ES cells are a promising candidate for therapeutic use. For clincal applications, it is clear human ES cells will be needed.

1.1.2 Human Embryonic Stem Cells

In 1998, human embryonic stem (hES) cells were derived from the inner cell mass of human embryos produced by *in vitro* fertilization (Thomson, Itskovitz-Eldor et al. 1998). Like mouse ES cells, hES cells can be expanded in an undifferentiated state *in vitro*, and are capable of differentiation into cells of all three germ layers, making them capable of producing all somatic cell types. In contrast to mouse ES cells, LIF does not promote hES cell self renewal (Thomson, Itskovitz-Eldor et al. 1998). Instead, the combination of activin and FGF2 is required to maintain hES cells in an undifferentiated state in the absence of feeder-cell layers, conditioned medium or serum replacement (Vallier, Alexander et al. 2005). Differentiation of hES cells to neural lineages uses many of the

same instructive factors as mouse ES cells (Wichterle, Lieberam et al. 2002; Hu, Du et al. 2009). The discovery of hES cells enabled the use of human cells to study human development and model disease. Because of their properties, hES cells will be a critical component in the development and application of cellular based therapeutics.

1.1.3 Induced Pluripotent Stem Cells

Recent discovery of the ability to reprogram differentiated cells into a pluripotent state has the potential to advance the use of stem cells for therapeutic applications. Induced pluripotent stem (iPS) cells are pluripotent cells artificially derived from differentiated cells by transfection of stem-cell associated transcription factors. The capability to generate iPS cells was first discovered by Yamanaka et al (Takahashi and Yamanaka 2006) using adult mouse fibroblasts and has been repeated using human fibroblasts (Takahashi, Tanabe et al. 2007; Wernig, Meissner et al. 2007) and other differentiated cell types (Reviewed in (Hochedlinger and Plath 2009). Growth profiles, gene expression profiles, and DNA methylation patterns closely resemble ES cells (Takahashi, Tanabe et al. 2007; Wernig, Meissner et al. 2007). Initial work with iPS cells used retroviruses to deliver reprogramming factors, raising concerns for clinical application of these cells. Other approaches including soluble signaling molecules are promising alternatives for clinical use (Shi, Desponts et al. 2008; Woltjen, Michael et al. 2009; Yu, Hu et al. 2009). The discovery of iPS may supplant needs for hES cells as the source of pluripotent cells, alleviating ethical concerns regarding the production of hES cells. The ability to reprogram adult cells has potential benefits for tissue engineering

applications because for the first time pluripotent cells can be obtained from a patient that will not be immunologically rejected.

1.1.4 Engineering ES by Gene Targeting

Gene targeting allows for precise manipulation of both mouse and human ES cell genomes (Smithies, Gregg et al. 1985; Thomas and Capecchi 1987; Capecchi 1989; Zwaka and Thomson 2003). This technology is widely used by biologists to disrupt gene function (a "knock-out") to study its function. Most often "knock-out" ES cells are injected into early-stage mouse embryos which are implanted in the uterus of a female mouse and allowed to develop into a mouse in order to study the effects of gene disruption on physiology. Gene targeting technology is also of great interest to the tissue engineer. ES cells can be engineered to facilitate monitoring and selection of specific cell fates. Introducing reporter and selector genes under the control of fate specific promoters, (a "knock-in" line), has shown great utility (Fehling, Lacaud et al. 2003; Xian, McNichols et al. 2003; Ying, Stavridis et al. 2003; Bibel, Richter et al. 2004; Xue, Wu et al. 2009). Knock-in lines will be instrumental for success in monitoring differentiation strategies and selecting for desired cell types.

In summary, the properties of ES cells make them promising candidates for therapeutic use. ES can be grown on a large scale, can differentiate into cells of many lineages, and furthermore can be genetically engineered to select for therapeutic cell type(s). Discoveries of hES and iPS have brought the dream of therapeutic cells even

closer. Nevertheless, harnessing stem cells for therapeutic applications is going to be a challenge. A tissue engineering approach combining engineered stem cells, instructive growth factors, and biomaterial scaffolds is desirable.

1.2 Tissue Engineering with ES Derived Neural Cells

ES cells offer a new opportunity for cell based therapy for both neurodegenerative disease and injuries. Neurodegenerative diseases are characterized by loss of function due to death of specific neural cell types. Traumatic injuries sever critical neural connections and result in cell death. In both cases there is hope that delivering healthy cells will be able to replace missing cells, repair connections, and restore function. ES cells can be differentiated to generate neural cells. Growth factors and culture conditions have been defined to differentiate both mouse and human ES cells into various neural lineages useful for therapeutics including: dopaminergic neurons (Kawasaki, Mizuseki et al. 2000; Wernig, Zhao et al. 2008) for Parkinson's disease, oligodendrocytes for myleinating deficiencies (Brustle, Jones et al. 1999; Nistor, Totoiu et al. 2005; Kang, Cho et al. 2007; Hu, Du et al. 2009), spinal motor neurons (Wichterle, Lieberam et al. 2002; Silani, Cova et al. 2004; Hu and Zhang 2009) for Lou Gerhig's Disease and spinal cord injury, and retinal cells for age related macular degeneration (Lamba, Karl et al. 2006; Zhao, Liu et al. 2006).

For the purposes of this dissertation, we will focus on the use of ES cells alone and as part of tissue engineering approaches, combined with growth factors and scaffolds, to aid in regeneration of cells following spinal cord injury (SCI) (Reviewed by (Coutts and Keirstead 2008; Willerth and Sakiyama-Elbert 2008). This is particularly relevant given the clinical trial using hES cells in human SCI by Geron.

1.2.1 Spinal Cord Injury

Traumatic injury to the spinal cord initiates a complex biological cascade characterized by inflammation, swelling, cell death, and demyelination. In the early stages of spinal cord injury, severed connections result in degradation and retraction of axons. Both neurons and oligodendrocytes are lost due to SCI. Resident astrocytes respond to the injury immediately and over time create a glial scar. This scar will inhibit axonal growth and myelination making the body unable to repair the severed connections (Coutts and Keirstead 2008). Depending on the location of the injury and its extent, the patient could have minor to catastrophic sensory and motor impairment. One strategy with the potential to regenerate the severed connections is to deliver healthy motor neurons and oligodendrocytes to the injury site to return some motor control.

1.2.2 ES Cells as Therapeutic Cells for Spinal Cord Injury

For safety reasons, undifferentiated ES cells cannot be used directly as a therapeutic because they will form teratomas or other cancers (Martin 1981). In addition, normal or injured adult tissues do not send proper inducing signals to direct stem cells to

desired fates (Nussbaum, Minami et al. 2007). Therefore, ES cells must be differentiated towards the therapeutic cell type *in vitro* before delivery.

1.2.3 From ES Cells to ES Derived Neural Cells Using Growth Factors

While many combinations of growth factors have been investigated to direct ES cells to a neural lineage, retinoic acid and sonic hedgehog have been the primary focus of research to direct ES cells to therapeutic cells for SCI. Initially, ES cells were directed to a general neural lineage by induction with retinoic acid (RA) (Bain, Kitchens et al. 1995). The ES cells are first cultured without LIF and RA for four days as embryoid bodies, spherical clusters of 100-1000 ES cells. After four days without RA, cells are then cultured in the presence of RA for four days to direct ES cells to a general neural lineage. These cultures are heterogeneous, containing astrocytes, oligodendrocytes and neurons. When these neural cells were implanted in a rat spinal cord injury model, cells survived and differentiated into oligodendrocytes, astrocytes, and neurons. Rats that received cells exhibited a small, but measurable functional improvement in locomotor function compared to animals that did not receive cells (McDonald, Liu et al. 1999). This early transplantation study showed delivery of ES derived neural cells contribute to improved outcomes and these cells warrant further investigation as a therapeutic for SCI.

The combination of retinoic acid and sonic hedgehog (Shh) further specifies differentiation of ES cells to cell types from the ventral portion of the central nervous system (Wichterle, Lieberam et al. 2002). ES cells are cultured as embryoid bodies for two days and then in the presence of RA and Shh for up to six days. A Shh agonist or purmophamine, which also acts on the hedgehog signaling pathway can be used to

substitute for Shh (Wichterle, Lieberam et al. 2002; Sinha and Chen 2006; Xue, Wu et al. 2009). These cultures are enriched in motor neurons and oligodendrocyte precursor cells, but still contain many cell types. ES derived ventral neural cells have been transplanted into rat spinal cord injury models and show surviving transplanted cells are enriched in motor neurons (Deshpande, Kim et al. 2006) where as ES cells treated with RA alone did not form motor neurons. Transplantation of ES derived ventral neural cells also resulted in improved motor function over controls with no therapeutic cells and general neural cells. Optimized protocols for further differentiation of hES derived ventral neural cells to motor neurons (Hu and Zhang 2009) and fully functional oligodendrocytes (Hu, Du et al. 2009) were recently described and may become useful for the next phase of research in spinal cord injury models.

1.2.4 Biomaterial Scaffolds for SCI

For SCI, scaffolds are important to provide a structural conduit to aid cells in bridging severed connections and enable delivery of additional differentiation signals or therapeutic agents. Tissue engineers have incorporated ES derived neural cells into a variety of biomaterial scaffolds to explore their utility for treatment of SCI. Types of scaffolds that are under investigation include: fibrin (Willerth, Arendas et al. 2006; Johnson, Parker et al. 2009), polymeric including poly(lactic-co-glycolic) acid (PLGA) poly L-lactic acid (PLLA) (Levenberg, Burdick et al. 2005), and poly(ethylene glycol) (PEG) (Mahoney and Anseth 2007), and electrospun fibers (Xie, Willerth et al. 2009). Along with cells many of these biomaterials can also encapsulate additional growth factors and therapeutic agents that can be released in a controlled fashion (Johnson,

Tatara et al. 2009) to aid in differentiation and cell migration, inhibit cell death, and facilitate axon extension. Transplantation of ES derived neural cells with biomaterial scaffolds into SCI models have also led to small, but significant, improvements in motor function (Johnson 2009). Yet, there is still much research to be done in this area. These studies are excellent models for future research on the types of cells, growth factors, and materials that will produce the best outcomes in SCI.

1.2.5 Heterogeneity is a Challenge for Cell Therapies

The central nervous system contains hundreds of different cell types. Each class of cells performs a specialized task which contributes to the overall function of the central nervous system. Disease or traumatic injury preferentially attacks certain cell types. Once we know which cells are affected, we can begin to investigate delivery of replacement cells as part of a therapeutic strategy. To be successful we need to experimentally determine which types of cells are most beneficial.

Many cell types are lost in SCI, most notably motor neurons and oligodendrocytes, but there are also other cell types that may play a critical supporting role. As a starting point, two different types of ES derived therapeutic cells for SCI will be considered: relatively undifferentiated early neural stem cells and differentiated young motor neurons. Delivery of early neural stem cells offers several advantages. Because early neural stem cells are able to proliferate they can help repopulate the cord, aiding in bridging the severed connections. Early neural stem cells will respond to endogenous or provided signals to differentiate and be directed to the needed therapeutic cell type(s). More than one type of therapeutic cell type can be delivered based on the differentiation

potential of the early neural stem cell. These advantages introduce the risk of several disadvantages. Early neural stem cells may continue to proliferate beyond what is needed, may not differentiate into the therapeutic cell type(s), or differentiate into undesirable cell types (Roy, Cleren et al. 2006), (Johnson 2009).

As a potential alternative, young motor neurons have the advantage of being both a more homogeneous population of cells, and will not proliferate or generate undesirable cell types. However, there are also potential problems with delivery of differentiated cells. Because young motor neurons will not proliferate, enough cells need to be provided that will survive and become incorporated as functional units following the transplantation procedure. This is a concern due to the fragile nature of motor neurons and their delivery to an inhospitable injury site (Harper, Krishnan et al. 2004). In addition, there may be other cell types that are needed for survival and therapeutic gain which are not being supplied. In summary, both cell types described here offer potential benefits and pitfalls. We cannot determine a priori which will be best. The optimal cell types for therapeutic use must be determined by experimentation.

To be able to experimentally determine the most beneficial cell type for delivery, we need to be able to both distinguish between cell types and have a means of isolating specific cell types. Sections 1.3.3 and 1.3.4 describe approaches to perform these tasks. Improvements in the tools that are used to characterize and separate cells will address the challenge that cellular heterogeneity poses to the development of tissue engineering strategies.

1.3 Cell Fate Characterization

Cell fate characterization is an important part of any tissue engineering strategy that incorporates ES cells. The properties of the ES cells should be known, and methods for measuring the transition of pluripotent ES cells to the desired therapeutic cell fate must to be developed. For use of ES cells in therapeutic applications, we must be able to define measurable characteristics to monitor the pluripotent state and differentiation. Candidate characteristics include: epigenetic marks, mRNA expression, protein expression, and metabolite profiles. Recent advances in genomics, PCR, and sequencing technologies make nucleic acid based analyses particularly attractive.

1.3.1 mRNA Profiling

Transcription of RNA marks one of the key steps in determining cellular fate. The mRNA encodes the information necessary for synthesis of proteins that are necessary for the cell to generate and maintain phenotype. The presence or absence, and abundance of particular mRNAs will determine cellular phenotype. For example, the mRNA profile of an ES cell is very different from a neuron (Abranches, Silva et al. 2009).

Biological tissues contain multiple cell types arranged in highly organized structures. For tissue engineering mRNA profiling can be used to first identify the characteristics of the needed therapeutic cell type(s). Once the desired mRNA profile is known it can then be used as a quality control standard for the production of therapeutic cells from ES cells. An ideal technology for mRNA profiling in tissue engineering would

have three characteristics. It must be able to measure mRNA for many genes in a sample containing single cells or small numbers of cells and also must be quantitative. Standard mRNA expression profiling techniques such as microarrays, northern analysis, or RT-PCR do not meet these criteria. New digital methods (section 1.3.3, Chapter 2) are likely to lead to advances in mRNA profiling technologies. Advances in mRNA profiling technology will be important to advance the use of ES cells in tissue engineering.

1.3.2 DNA Methylation Analysis

DNA methylation analysis is a complementary and alternative approach to mRNA expression analysis for characterizing cell fate (Baron, Turbachova et al. 2006). Mammalian DNA contains cytosine residues that are marked by covalent addition of a methyl group on the 5-carbon. At the time this thesis was initiated, DNA methylation was thought to function in silencing gene expression (Wolf, Jolly et al. 1984; Watt and Molloy 1988; Boyes and Bird 1991; Li, Beard et al. 1993; Panning and Jaenisch 1996; Walsh, Chaillet et al. 1998). Therefore, genes silenced during differentiation would become methylated in tissues where they are not needed. Consequently, changes in methylation patterns could be used to monitor differentiation of ES cells to a desired cell type.

Perceptions of the functional role of DNA methylation are changing (see chapter 3, Reviewed by (Suzuki and Bird 2008) as increased methylation data has become available through high throughput sequencing platforms (section 1.3.3). DNA methylation is no longer thought to be a definitive silencer. However, DNA methylation has several attributes that make it an attractive method to characterize cells. Foremost,

DNA methylation can be measured with base-pair precision using bisulfite treatment and PCR (Frommer, McDonald et al. 1992; Clark, Harrison et al. 1994). This technology has facilitated discovery of cell and tissue specific DNA methylation profiles for individual neural genes (Takizawa, Nakashima et al. 2001) and on a genome wide level (Eckhardt, Lewin et al. 2006; Illingworth, Kerr et al. 2008). For example, the promoter of Oct3/4, a critical marker of ES cell pluripotency, becomes methylated when ES cells are differentiated (Fouse, Shen et al. 2008); (Gidekel and Bergman 2002); (Hattori, Nishino et al. 2004); (Li, Pu et al. 2007) and unmethylated when fibroblasts are reprogrammed to a pluripotent cell (Takahashi and Yamanaka 2006; Takahashi, Tanabe et al. 2007; Wernig, Meissner et al. 2007). While the function of methylation is not well understood, correlation of methylation status and phenotype may be useful in a diagnostic context.

DNA methylation profiles could provide a diagnostic marker that would open the way for a high-resolution analysis of cell fate. After cell type and tissue specific methylation profiles are identified, they could be used as criteria to monitor the quality of therapeutic cells by their differentiation status. Therefore, discovery of cell type specific methylation profiles and development of DNA methylation assays have the potential to further the use of ES cells in tissue engineering research.

1.3.3 Digital Methods for mRNA Profiling and Methylation Analysis

Digital PCR relies on two principles: segregation of DNA molecules and clonal amplification by PCR to quantitatively detect single molecules. Messenger RNA can be treated in a similar fashion once RNA is reverse transcribed into complementary DNA (cDNA). The first embodiment of this method used dilution of DNA into a well plate to

separate molecules (Vogelstein and Kinzler 1999). Subsequently, gel matrices (Mitra and Church 1999; Chetverina, Samatov et al. 2002) and bead based emulsions (Dressman, Yan et al. 2003) have been used to segregate DNA molecules. Capitalizing on the extraordinary amplification power of PCR, colonies of DNA with identical sequence are obtained from each molecule. The sequence of these colonies can then be interrogated in numerous ways (Ronaghi, Uhlen et al. 1998; Vogelstein and Kinzler 1999; Mitra, Shendure et al. 2003) including high-throughput sequencing. Due to their sensitivity, digital methods are uniquely positioned for profiling stem cells ((Rieger, Poppino et al. 2007), See chapter 2). At the time this thesis was initiated digital PCR strategies were just beginning to be investigated for mRNA expression (Warren, Bryder et al. 2006; Kim, Porreca et al. 2007) and methylation analysis (Taylor, Kramer et al. 2007; Korshunova, Maloney et al. 2008). Currently, digital PCR joined to low-cost, high-throughput sequencing technologies sold by companies such as Illumina, Roche (454), and Applied Biosystems (SOLiD) have set the new standard for mRNA profiling for research and industry. These new methods are already providing powerful tools to gain insights into cell fate and will become critically important in the development of tissue engineering strategies.

1.4 Cell Separation

Despite considerable progress in directing stem cells to specific lineage by the application of growth factors, the population of differentiated cells is typically still heterogeneous. For therapeutic applications of ES cells, it will be important to enrich for the desired cell types and remove pluripotent and other rapidly dividing cells which may

have adverse clinical effects. Cell separation technologies are a necessary part of the future of the tissue engineering field (Palsson and Bhatia 2004).

Cell separation technologies that are able to provide a high level of purity and yield are desired for therapeutic applications. Separation based on physical properties such as microdissection or density centrifugation give moderate to good purity but do not scale. Biological separation using surface markers with immuno-affinity allows for high yield, but often low purity (Schweitzer, van der Schoot et al. 1995; Corti, Nizzardo et al. 2009). Individual cells can be separated by surface markers and genetically engineered cells containing reporter genes can be selected with fluorescence activated cell sorting (FACS). Using the FACS allows for purification of cells with multiple characteristics, but operates at relatively low throughput (Palsson and Bhatia 2004; Murry and Keller 2008). While not widely used, selection using antibiotics and cells engineered with corresponding resistance gene have shown the ability to yield clinically relevant numbers of cells in bioreactors at high purity (Zandstra, Bauwens et al. 2003).

1.5 Concluding Remarks

This thesis is motivated by the potential of stem cells as a source of replacement cells in cellular and tissue engineering therapeutics for treatment of neurodegenerative disease and injury. Because the nervous system is comprised of hundreds of cell types which are differentially affected by disease and injury, it will be necessary to experimentally determine which cell types are most beneficial as part of a therapeutic

strategy. For this purpose tools are needed that can both define heterogeneity and purify the desired cell types.

This thesis will detail the development of three tools to advance the use of stem cells as part of tissue engineering approaches for neurodegenerative disease and spinal cord injury. The first two studies will describe the application of two nucleic acid-based approaches to define cell fate. In chapter two, a digital PCR technology, called polonies, is applied to quantitatively measure mRNA levels from several key stem cell genes in small numbers of ES cells. To explore DNA methylation as a way to characterize differentiation from ES to neural cells, chapter three will measure DNA methylation surrounding genes that are instructive to pluripotency and neural cell fate in ES cells, ES derived neural cells, and non-neural tissues. Chapter four develops an approach to purify subsets of neural cells. ES cells are genetically engineered to permit drug selection of a subset of ES derived neural cells. Overall, this thesis contributes tools to define and limit cellular heterogeneity which will be useful in the development of cellular and tissue engineering therapeutics for neurodegenerative disease and injury.

Chapter 2

Polony Gene Expression Analysis of Stem Cells

2.1 Abstract

Expression profiling of stem cells is challenging due to their small numbers and heterogeneity. The PCR colony (polony) approach has theoretical advantages as an assay for stem cells but has not been applied to small numbers of cells. An assay has been developed that is sensitive enough to detect mRNAs from small numbers of ES cells and from fractions of a single mouse blastocyst. Genes assayed include Oct3, Rex1, Nanog, Cdx2 and GLUT-1. The assay is highly sensitive so that multiple mRNAs from a single blastocyst were easily detected in the same assay. In its present version, the assay is an attractive alternative to conventional RT- PCR for profiling a small population of stem cells. The assay is also amenable to improvements that will increase its sensitivity and ability to analyze many cDNAs simultaneously.

Chapter 2 has been adapted from the following published manuscript:
Rieger, C., Poppino, R., Sheridan, R., Moley, K., Mitra R., and D. Gottlieb. (2007). "Polony analysis of gene expression in ES cells and blastocysts." Nucleic Acids Res **35**(22): e151.

2.2 Introduction

Stem cells are currently the focus of intense interest because of their importance in normal development and adult physiology as well as their potential application in clinical medicine. Expression profiling of stem cells poses a special challenge and lack of appropriate methods constrains progress in many branches of stem cell research. The challenge arises because stem cells occur as small populations surrounded by other cell types and because even the stem cell populations themselves are heterogeneous and encompass multiple cell populations. An ideal profiling method would have three capabilities. The first is the sensitivity to assay mRNAs in small populations and single cells and thus deal with heterogeneity. Second, because cell fate is determined by sets of genes rather than any single gene, the method must allow parallel analysis of multiple genes. Finally the method must be quantitative, since levels of expression rather than mere presence or absence of transcripts determines phenotype. While multiple expression analyses of stem cells based on PCR have been published, no method fulfils all these criteria (Chiang and Melton 2003; Tietjen, Rihel et al. 2003; Bengtsson, Stahlberg et al. 2005; Hartmann and Klein 2006; Tang, Hajkova et al. 2006). The method of PCR colony (polony) analysis differs in important ways from conventional PCR and has potential to be very useful for profiling stem cells.

In polony (also called molecular colony (Chetverina, Samatov et al. 2002)) analysis, individual DNA molecules are amplified clonally in a polyacrylamide gel matrix (Mitra and Church 1999; Zhu, Shendure et al. 2003). Analysis is very efficient, with 80% of the input DNA molecules forming polonies, so the method is inherently very sensitive

(Mitra and Church 1999). All polonies signify one starting template DNA molecule so variations of amplification efficiency do not influence the final count of input templates. Cross- interference of different amplicons is largely avoided, since the reactions are effectively isolated from one another by the gel matrix. The DNA sequence of individual polonies can be ascertained by either sequence-specific fluorescent hybridization probes or an in situ sequencing procedure, opening the way for parallel multigene analysis (Mitra, Shendure et al. 2003). Because of these features, the polony method is an excellent candidate approach for profiling stem cells. However, previous expression studies with polonies have used relatively large starting samples of cells (Kim, Porreca et al. 2007) so it is not known if the technique can be applied to small numbers of cells and be useful for stem cell profiling.

In this report we demonstrate that the polony method can be used on small numbers of stem cells including ES cells and blastocysts. A method for isolating RNA and synthesizing cDNA from small samples was coupled with polony analysis and the sensitivity of the overall approach and the ability to do parallel analyses of multiple genes was evaluated. Our results represent significant progress towards the ideal profiling method described above and will encourage further technical developments of the polony approach.

2.3 Materials and Methods

ES cell culture

All ES cell experiments were done with the RW4 line of ES cells derived from Sv129 mice. Undifferentiated (ES) cells were grown on gelatin-coated tissue culture

plastic in the presence of leukemia inhibitory factor (LIF) using standard methods (Bain, Kitchens et al. 1995; Bain 1998). In preparation for polony experiments, undifferentiated ES cells were trypsinized with 0.25% trypsin-EDTA (GIBCO) to detach cells from the surface, and counted using a hemocytometer. For the 1,000 ES cell isolation, cells were spun down and diluted to obtain a concentration of 500,000 cells/mL.

Embryo Recovery and Culture

Embryos were recovered as previously described (Moley, Chi et al. 1998). In brief, three-week-old female mice (B6 x SJL F1, Jackson Laboratories; Bar Harbor, ME) were given free access to food and water and were maintained on a 12-hour light/dark cycle. Female mice were superovulated with an intraperitoneal injection of 10 International Units (IU)/animal pregnant mare serum gonadotropin (PMSG, Sigma; St. Louis, MO) followed 48 hours later by 10 IU/animal of human chorionic gonadotropin (hCG, Sigma; St. Louis, MO). Female mice were mated with males of proven fertility overnight following the hCG injection. Mating was confirmed by identification of a vaginal plug. Mice were sacrificed 96 hours post hCG injection to recover embryos at the blastocyst stage (3.5 d.p.c). Embryos were recovered by flushing dissected uterine horns and ostia with human tubal fluid medium (HTF, Irvine Scientific; Santa Ana, CA) containing 0.25% BSA (Bovine serum albumin fraction V, Sigma; St. Louis, MO).

RNA Isolation

Cells (either ES cells or blastocysts) were delivered to PCR tubes containing either 10μg for blastocysts or 100μg for ES cells of Dynabeads Oligo $(dT)_{25}$ (Invitrogen) in 20μL or 100μL lysis-binding buffer (100 mM Tris-HCl, pH 7.5, 500 mM LiCl, 10 mM EDTA, pH 8.0, 1% LiDS, 5 mM dithiothreitol (DTT)). Cells were lysed by pipetting up

and down five times in the lysis-binding solution. Tubes were rotated for 10 minutes at room temperature to promote hybridization of the polyA$^+$ mRNA with the oligo(dT) tails of the Dynabeads. After hybridization of mRNA with Dynabeads, a series of washes was performed to prepare the mRNA for reverse transcription. Two washes were performed in wash solution A (10 mM Tris-HCl, pH 7.5, 0.15 M LiCl, 1 mM EDTA, 0.1% LiDS). Next, Dynabeads were incubated in 100 μL wash solution B (10 mM Tris-HCl, pH 7.5, 0.15 M LiCl, 1 mM EDTA) with 0.01% Tween-20 for 5 minutes to allow the beads to equilibrate. This was followed by a second wash in wash solution B without Tween and a final wash in 10mM ice cold Tris-HCl, pH 7.5. In some experiments mRNA attached to the beads was used directly in an RT reaction. In others the mRNA was eluted in 10μL Tris-HCl by heating at 90°C for 2 minutes.

cDNA Synthesis

Reverse-transcription reactions were performed using the RETROscript kit (Ambion, Austin, TX). Final concentrations of components were as follows: 1x RT buffer (50mM Tris-HCl, pH 8.3, 75mM KCl, 3mM $MgCl_2$, 5mM DTT), 5μM oligo(dT_{18}), 500μM each dNTP, 0.5U/μL RNase Inhibitor, 5U/μL MMLV reverse transcriptase, .05 μg/μl BSA was added as a carrier. For cDNA synthesis reactions performed on mRNA hybridized to Dynabeads the oligodT primer was omitted. cDNA synthesis reactions were carried out at 42°C on a roller for 1 hour. An RT-minus reaction was always prepared in parallel by substituting water for MMLV RT-enzyme.

Primers

The primers used are listed in Appendix A, Tables A1 – A3; all primers are from Integrated DNA technologies (IDT, Coralville, IA). Primers were selected using Primer3

with the restriction of being within 800bp of the 3' end. All polony reverse primers include an acrydite group (Ac) on the 5' ends (Mitra and Church 1999). The 5' end of the hybridization primers are covalently linked to a fluorescent dye (Cy5).

Polony Reactions

Polony reactions were prepared according to Mitra and Church (Mitra and Church 1999). Template cDNA was added to a liquid phase acrylamide gel mix containing PCR components. Templates were amplified using PCR within the gel. cDNA template or RT-minus suspension was added to a liquid-phase PCR mix (polony mastermix) [10mM Tris-HCl pH 8.3, 50mM KCl, 0.01% gelatin, 1.5mM $MgCl_2$, 200μM dNTPs, 1 μM Forward Primer, 1μM Primer Reverse_Ac, 3.3U or 3.8U Jumpstart Taq (Sigma, St. Louis, MO), 9% Acrylamide, 0.05% Bisacrylamide (Sigma)]. Then, 0.667 μL of degassed 5% ammonium persulfate (Sigma) and 0.667 μl 5% temed (Sigma) were added to the polony mix to a total volume of 28μL or 40μL. Nineteen microliters of this solution was pipetted underneath a clean No. 2 coverslip (18 x 30 mm Fisher) on a bind-silane (Sigma) treated Teflon coated oval well slide (Erie Scientific, Portsmouth NH).

Slides were cycled using a PTC-200 thermal cycler (Bio-Rad, Hercules, CA) adapted for glass slides (16/16 twin tower block). The following program was used: denaturation (2 min at 94°C) followed by 43 cycles of denaturation, primer annealing, and extension (30 s at 94 °C, 30 s at 56 °C, 30 s at 72 °C). After cycling the Secure-Seal™ chamber was removed and slides were washed in hexane (Sigma) for 5 minutes to remove mineral oil and remaining adhesive. Coverslips were removed and slides were washed twice in solution 1E (100mM Tris pH 7.5, 500mM KCl, 2mM EDTA, 0.01% Triton-X 100) for 4 minutes with gentle shaking.

Hybridization for Polony Detection

Slides were incubated in 70% formamide (Sigma) in 1X SSC at 70°C on a roller for 2 minutes to denature double stranded DNA. Formamide was removed by washing with water for 3 minutes followed by washing with solution 1E. A blue Frame Seal chamber base (Bio-Rad) was applied to each slide and annealing mix was added (5.6μM hybridization probe in 125μL of 6x SSPE buffer with 1% Triton X-100). Slides were heated (2 minutes at 94 °C, 7 minutes at 56°C). Frame seal chambers were removed quickly and slides were placed in wash1E to dilute away excess primer to limit non-specific binding. Slides were washed and stored in wash1E.

Visualizing and Scoring Polonies

Polony slides were coverslipped, and imaged using a GenePix 4000B (Axon Instruments, Union City, CA) microarray scanner and GenePix software. Optimal signal intensity for the Cy5 fluorophore was obtained for laser PMT gain of 700 (635 laser) and 82 (532 laser). Images were saved as TIF and JPEG files. Polonies were counted manually using ImageJ software and cell counter applet.

Competitive PCR

DNA competitors with a 50bp deletion of the corresponding native amplicons were synthesized by standard methods. The competitors have the same terminal sequences as the native amplicons to ensure equal amplification. Forward primers, reverse primers and deletion primers are described in Table S4. For polony and competitive PCR analysis, RNA was extracted in a series of reactions containing 2,000 ES cells and 100μg of Dynabeads in 100μL of lysis-binding buffer as previously described. RNA was eluted from Dynabeads in 20μL DEPC H_20 and 4μL oligo$(dT)_{18}$ and reverse transcription

performed as previously described in a total volume of 42μL. Competitive PCR reactions were carried out with a fixed amount of sample and varying amounts of competitor to determine the equivalence point.

Model RNA

To analyze the efficiency of RT mimic mRNAs were constructed. The mimic consists of the BNI5 yeast gene fused to a polyA+ tail taken from a region of the Xenopus elongation factor-1α gene (pTRI-XEF, Megascript kit, Ambion). The mimic was synthesized by knitting PCR followed by cloning into pBluescriptSKII(+) vector (Stratagene). RNA was transcribed from this plasmid by standard methods using T7 RNA polymerase. Model RNAs were purified by standard methods and quantified by OD260 absorption.

2.4 Results

Polony technology has been used extensively to analyze genomic DNA and in a few instances cDNAs from large numbers of yeast or mammalian cells (Zhu, Shendure et al. 2003; Mikkilineni, Mitra et al. 2004; Kim, Porreca et al. 2007), but has not been used to profile gene expression from small numbers of cells. Our first question was whether polony analysis could be applied to small numbers of mammalian stem cells. Mouse ES cells were chosen because they offer a pure population of stem cells where the gene expression pattern is clearly related to cell fate choice (Boyer, Plath et al. 2006; Lee, Jenner et al. 2006). We also analyzed blastocysts, a stage of mammalian development comprised mainly of stem cells including a subset which corresponds to ES cells.

In the first experiment 1000 ES cells were used as the starting sample for isolating mRNA. Several methods of RNA extraction were investigated and it was found that hybridization capture of mRNA on oligo $(dT)_5$ Dynabeads was particularly efficient. An illustration of the overall experimental methods is provided in Figure 2.1.

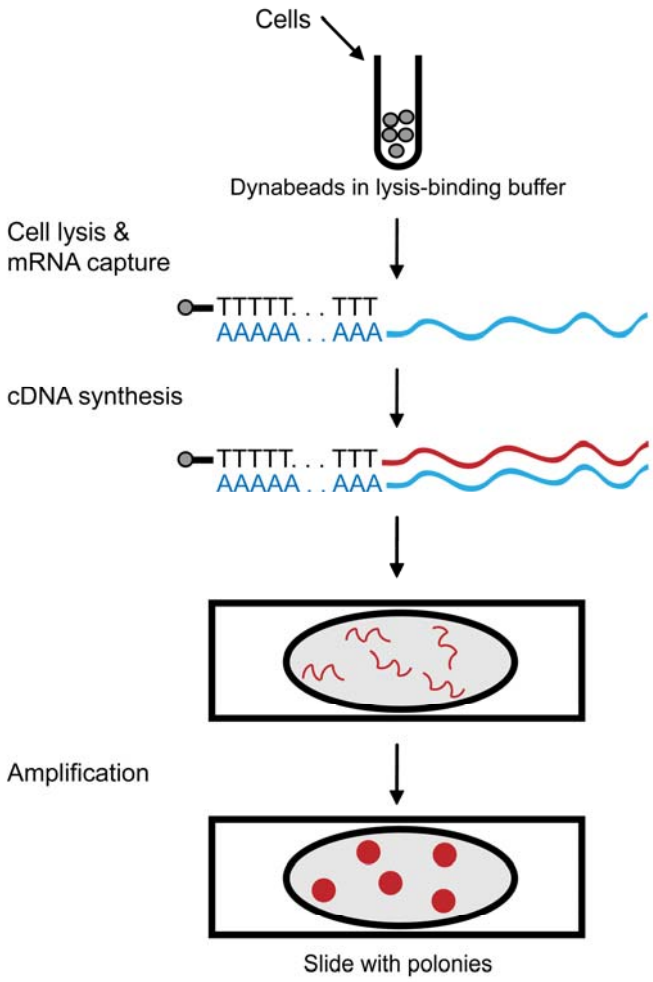


Figure 2.1: Flow Chart of Typical Polony Gene Expression Experiment. Cells (either 1000 ES cells, single blastocysts or multiple blastocysts) were delivered to a lysis-binding solution containing oligo$(dT)_{25}$ Dynabeads. After cell lysis, mRNA was captured by hybridization of poly-A tails on the beads and mRNA was reverse transcribed into cDNA. cDNA was added to non-polymerized polyacrylamide gel mix containing PCR components and deposited in an oval well on a microscope slide. After polymerization of the gel, slides were thermocycled so that cDNA templates gave rise to polonies. Polonies were visualized by hybridization with a labeled gene-specific probe.

The mRNA from 1000 ES cells was captured on Dynabeads and added to a reverse transcriptase (RT) reaction with the Oligo-dT of the beads serving as primer. After cDNA synthesis, a small fraction of the beads was delivered to a polony slide with primers designed to amplify Oct3, a transcription factor involved in maintaining the pluripotency of ES cells (Pesce and Scholer 2001). Each polony slide received the equivalent of 10 cells worth of cDNA or an equal volume of a control reaction lacking RT. Slides were thermocycled and then stained with a labeled hybridization primer for Oct3. In this and all subsequent experiments hybridization probes are internal to the amplifying primers and are labeled with Cy5 coupled to the 5' terminus. It is crucial that the assays be highly specific for the intended transcript and not show false positives. As with any PCR method, there is the potential of primer dimers and other unintended amplified sequences. Our results are very likely to be free of this sort of error for two reasons. All experiments include RT control samples and these do not produce polonies. Second, scoring polonies by hybridization of an internal primer which does not share sequence with the amplifying primers prevents signals from primer dimmers and other unintended amplicons. Polonies were visualized on an Axon microarray scanner and were abundant, evenly distributed and clearly distinguishable from background on the slides with cDNA. Importantly, polonies were absent from the RT control slide demonstrating that cDNA rather than genomic DNA is detected (Figure 2.2).

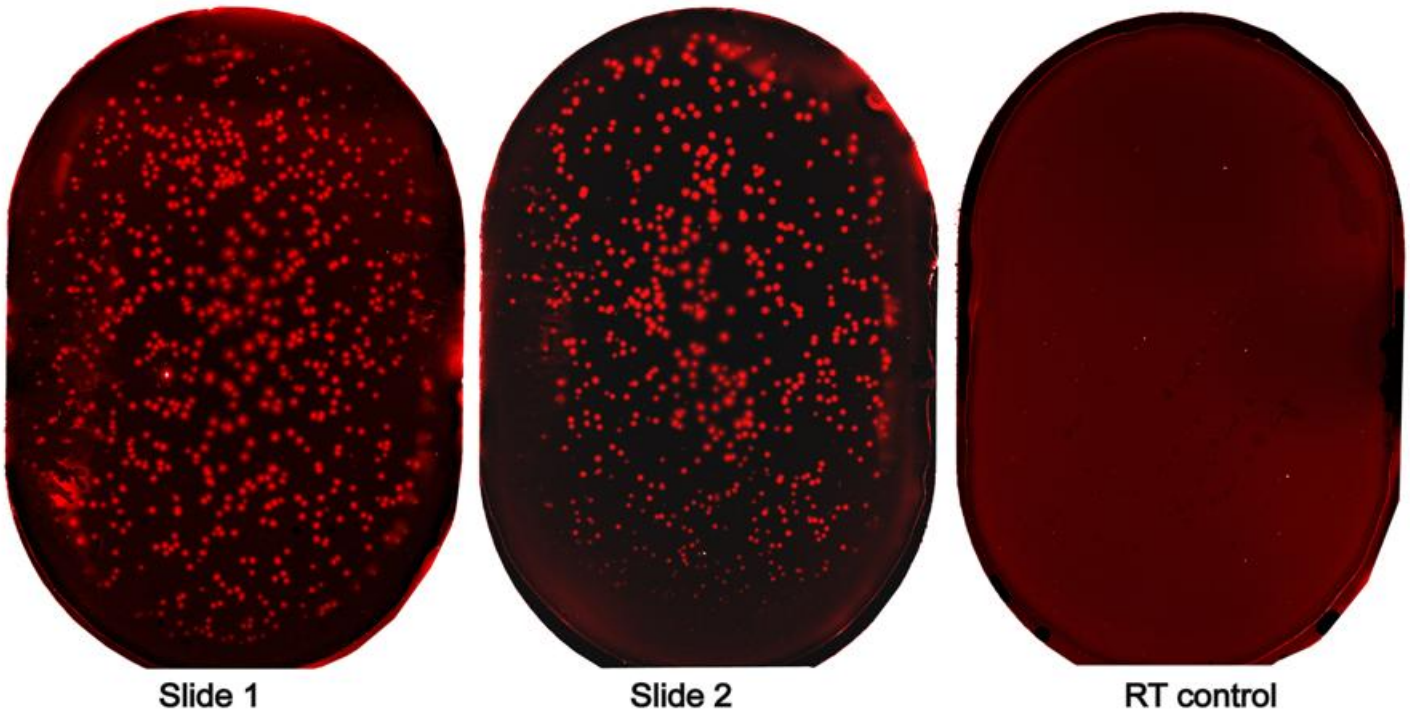


Figure 2.2: Oct3 polonies from ES cells. Slides 1 and 2 each received cDNA equivalent to 10 ES cells that was amplified to create Oct3 polonies. Polonies were visualized with a Cy5 gene-specific probe. The RT control slide is from a reaction without reverse transcriptase and has no polonies.

To investigate reproducibility, an experiment with two independent RT reactions was performed. Each RT reaction was assayed on 4 slides and the number of Oct 3 polonies on each slide were counted (Table 2.1).

Table 2.1: Oct3 Polony Counts from Slides Containing the Equivalent of 10 ES Cells

Slide number	Sample	Oct3 polonies per slide	Oct3 polonies per cell
1	10 ES cell equivalents	656	66
2	10 ES cell equivalents	542	54
3	10 ES cell equivalents	618	62
4	10 ES cell equivalents	513	51
5	10 ES cell equivalents	403	40
6	10 ES cell equivalents	292	29
7	10 ES cell equivalents	509	51
8	10 ES cell equivalents	460	46
9	RT control	0	0
Average		499	50
SD		116	
SE			37
Slides 1–4 originated from the same RT reaction			
Slides 5–8 originated from the same RT reaction			

mRNA was isolated from 1,000 ES cells and reverse transcribed to cDNA in 2 separate reactions. 10 ES cell equivalents of cDNA from each synthesis was delivered to 4 polony slides. After amplification, polonies were visualized by an Oct3 gene-specific hybridization probe. Scans of slides 1 and 2 appear in Figure 2. Oct3 polony counts ranged from 292 to 655 polonies per slide with an average of 499 per slide, equivalent to an average of 50 polonies per ES cell. Slide 9 is an RT control and does not contain polonies.

The mean of all 8 slides was 499 polonies with a standard deviation of 116; this is equivalent to a mean of 50 Oct3 polonies per cell. As discussed below this is a minimum estimate of the number of mRNAs per cell as it does not take into account the efficiency of mRNA isolation and conversion to cDNA. We conclude that the polony approach allows the assay of expression from small numbers of ES cells.

Having demonstrated that polonies can detect mRNA from small numbers of ES cells, we wanted to see if they could be used to detect mRNAs in a normal biological structure that contains stem cells and is made up of a small number of cells. We chose the mouse blastocyst since it is an intensely studied stage of mammalian development, is easily obtainable, and is comprised of only 75-100 cells (Chisholm, Johnson et al. 1985; Johnson and McConnell 2004). About 40% of the cells are in the inner cell mass (ICM) and phenotypically resemble ES cells. The transcription factor Oct3 is exclusively expressed in the ICM (Carayannopoulos, Schlein et al. 2004). In a range-finding experiment, 10 mouse blastocysts were pooled, their mRNA isolated, and cDNA synthesized. Polony assays for Oct3 were conducted on two slides each containing cDNA equivalent to one half of a blastocyst (Table 2.2).

Table 2.2: Oct3 Polony Counts from Pooled and Individual Blastocysts

Slide number	Sample (number of blastocysts in starting sample)	Oct3 polonies per slide	Oct3 polonies per blastocyst
1	1/2 Blastocyst equivalent (10)	967	1868
2	1/2 Blastocyst equivalent (10)	901	
3	1/2 Blastocyst equivalent (1)	857	
4	1/2 Blastocyst equivalent (1)	871	1728
5	RT control (5)	0	
6	RT control (5)	0	0
Average		899	1798
SD		48.9	

mRNA was isolated from a pool of 10 blastocysts and diluted so that the mRNA equivalent to one half of a blastocyst was delivered to slides 1 and 2. Next, mRNA isolated from a single blastocyst was divided between slides 3 and 4. Two RT control slides contained mRNA from the equivalent of 5 blastocysts. Oct3 polony counts from the pooled blastocyst sample are in good agreement with the individual blastocyst sample. Data demonstrates that analysis from one half of a blastocyst is feasible.

There were 967 and 901 polonies on the two slides for a total of 1868 polonies per blastocyst. Next, mRNA from a single blastocyst was isolated, reverse transcribed and 2 slides prepared. The average of these slides detected 1728 polonies per blastocyst. Two RT controls were done with the mRNA equivalent of five blastocysts; no polonies were present. We conclude that the polony method is sensitive to the level of a single blastocyst and that the entire analysis from mRNA preparation through polony analysis is scaleable in the range of one to ten blastocysts. The sensitivity of the polony assays compares very favorably with conventional RT-PCR analysis of expression in blastocysts, where multiple blastocysts are pooled to detect gene expression (Carayannopoulos, Schlein et al. 2004; Riley, Heeley et al. 2004). For some genes, however, expression of multiple genes can be measured from a single embryo (Zhang, Tam et al. 2006).

In order for the assay to be useful, it is essential to know the sample-to-sample variability due to cDNA synthesis and polony reactions. In this and all subsequent experiments we used two minor refinements of the previous protocol: mRNA was eluted from the beads prior to cDNA synthesis and the amount of Taq per slide was increased threefold. Taken together these two steps increased polony counts by about 30% (data not shown). To measure variability, mRNA was isolated from a pool of five blastocysts and split into three sub-pools, each the equivalent of a single blastocyst (Figure 2.3).

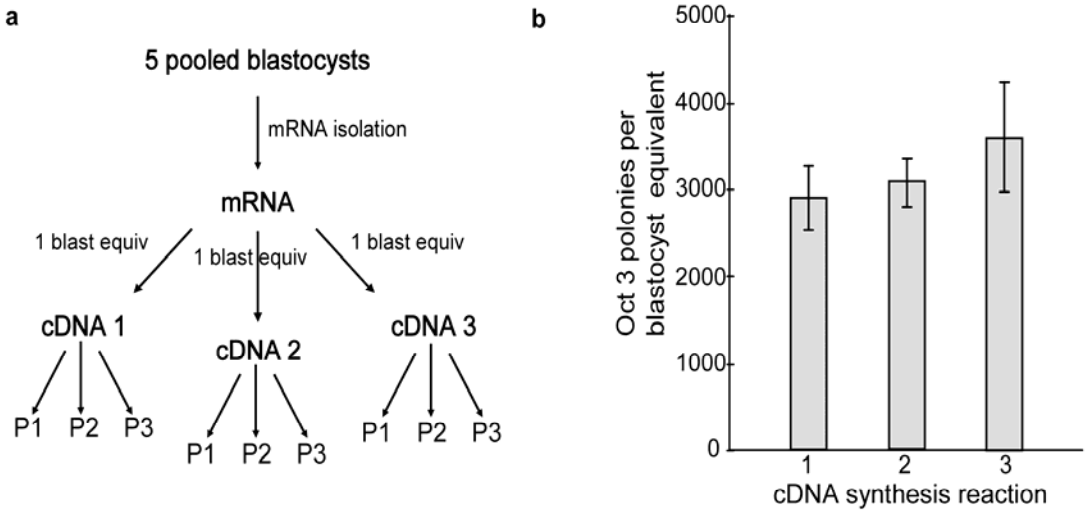


Figure 2.3: Analysis of Variation of RT and Polony Generation Steps. (a) Flow chart of the experiment. Five blastocysts are used in a single mRNA prep and one blastocyst equivalent is used in three separate RT reactions. Each RT reaction is analyzed on 3 separate polony slides (P1-3) for Oct3. (b) Bar graph where each bar is the average number of polonies for three slides from the same cDNA synthesis. The error bars are the standard deviation. ANOVA indicates that independent cDNA preparations are indistinguishable (P>.05)

These were reverse-transcribed in parallel, and cDNA analyzed for Oct3 transcripts in three polony reactions for each reverse transcriptase reaction. The variation between the polony numbers on replicate slides with the same reaction was acceptable, with the standard deviation being no more than 17.7 % of the mean. There was also good agreement between the means for the three different cDNA syntheses which differed by no more than 23%. An ANOVA analysis revealed that the different cDNA reactions were comparable to one another ($p>0.05$) with an overall average value of 3213+/-462 polonies per blastocyst. In conclusion, sample-to-sample variability is comparable to other widely used assays.

The ability to measure expression of multiple genes from a single sample is highly desirable, and we next investigated whether the polony assay could detect expression of multiple genes from a single blastocyst. We chose two other transcription factors expressed in ES cells and the blastocyst inner cell mass (ICM): Nanog and Rex1 (Rogers, Hosler et al. 1991; Mitsui, Tokuzawa et al. 2003). Gene-specific amplification and hybridization primers were designed for these mRNAs and validated with ES cells (data not shown). Next, individual blastocysts were assayed. RNA was extracted and cDNA synthesized by the same method as above, and the cDNA from each blastocyst split and delivered to three individual slides with primers for either Oct3, Nanog or Rex1, and the slides assayed with the appropriate gene-specific hybridization probe. As shown in Figure 2.4, all reactions yielded polonies; counts from this experiment are given in Table 2.3.

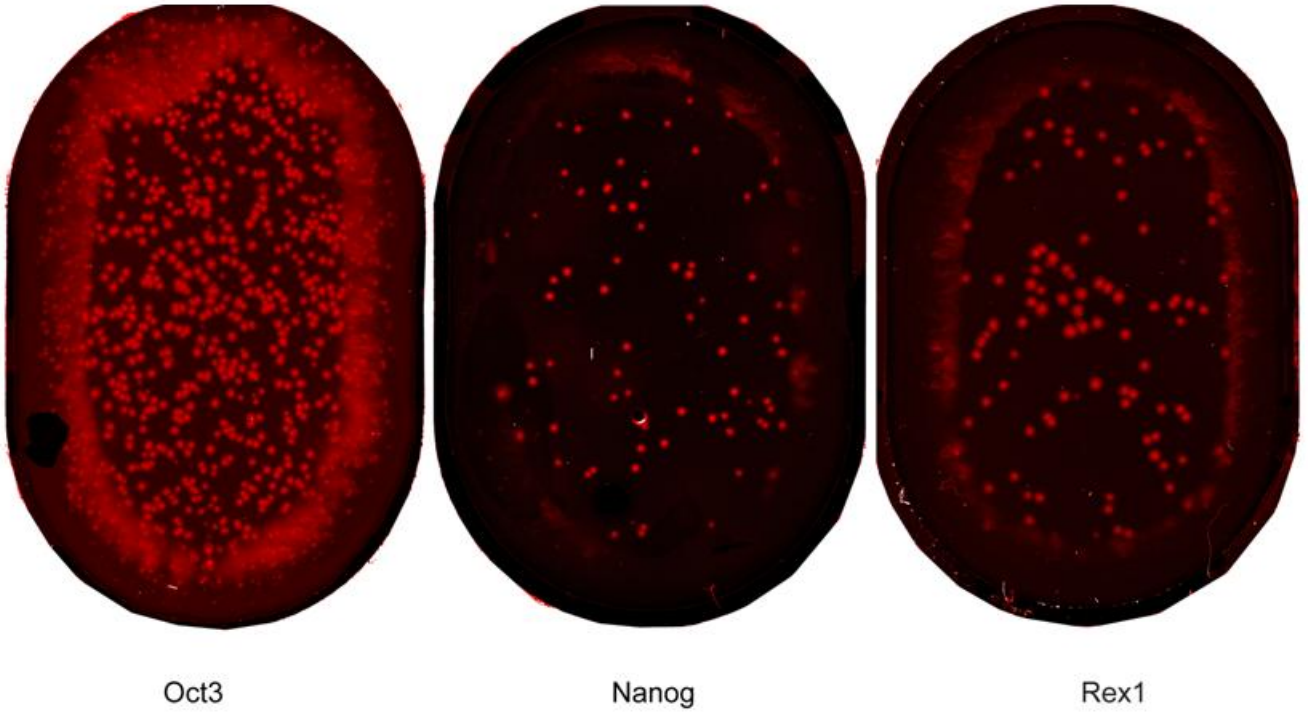


Figure 2.4: Detection of three genes from a single blastocyst. One-fifth of the cDNA was assayed for each of three genes: Oct3, Nanog, and Rex1. The polony method is sensitive enough to detect transcripts from only one-fifth of a blastocyst..

Table 2.3: Multigene Analysis of a Single Blastocyst

Gene	Polonies per blastocyst
Oct3	3242 ± 189
Nanog	361 ± 168
Rex1	664 ± 210

Three individual blastocysts were analyzed for 3 genes: Oct3, Nanog and Rex1. Three separate polony reactions were performed using mRNA from each blastocyst. The numbers in the table represent average and standard deviations for each gene. Gene specific hybridization probes were used to detect and quantify polonies

Oct3 gave the highest number of polonies; the number of Oct3 polonies per blastocyst was consistent with those of previous experiments. Nanog had the lowest number (~10% of Oct3) and Rex1 about twice as many as Nanog. The lower number of polonies for Nanog and Rex1 might mean that there are fewer mRNAs per blastocyst than Oct3. Alternatively, it could be because their isolation is less efficient or that cDNA synthesis is less efficient. We conclude that expression of at least three genes from a single blastocyst can be readily detected. This is in contrast with many current experiments with standard RT-PCR that require pooling multiple blastocysts (Carayannopoulos, Schlein et al. 2004; Riley, Heeley et al. 2004).

Blastocysts contain two layers termed the inner cell mass (ICM) and the trophectoderm. Oct3, Nanog and Rex1 are all expressed in the blastocyst ICM. To test the generality of the polony method we assayed expression of Cdx2, a gene selectively expressed in the trophectoderm (Strumpf, Mao et al. 2005; Deb, Sivaguru et al. 2006). Four individual blastocysts were analyzed for Cdx2 and Oct3 (Table2. 4).

Table 2.4: Analysis of Trophectoderm Gene Cdx2

Sample	Cdx2 polonies per blastocyst	Oct3 polonies per blastocyst
Single blastocyst	809	2268
Single blastocyst	2105	4305
Single blastocyst	1682	3614
Single blastocyst	1986	4027
Average	1646	3554
SD	585	903

RNA was extracted from each of four blastocysts and reverse transcribed separately. Each cDNA was split and analyzed for Cdx2 (2 slides) and Oct3 (1 slide). Polonies per blastocyst and the average and standard deviation are indicated.

Cdx2 polonies were present in all four blastocysts and there was a large variation among the four blastocysts with a range from 809 to 2105 Cdx2 polonies. The range for Oct3 was 2268 to 4305 which is consistent with previous experiments. We conclude that the polony approach can detect expression of a gene that is specifically expressed in the trophectoderm lineage of the blastocyst. All of the genes assayed above are for transcription factors and it is desirable to show that polonies can detect another class of genes. We therefore assayed the expression of GLUT-1, a membrane protein that is one of the primary glucose transporters in blastocysts (Figure 2.5) (Heilig, Saunders et al. 2003).

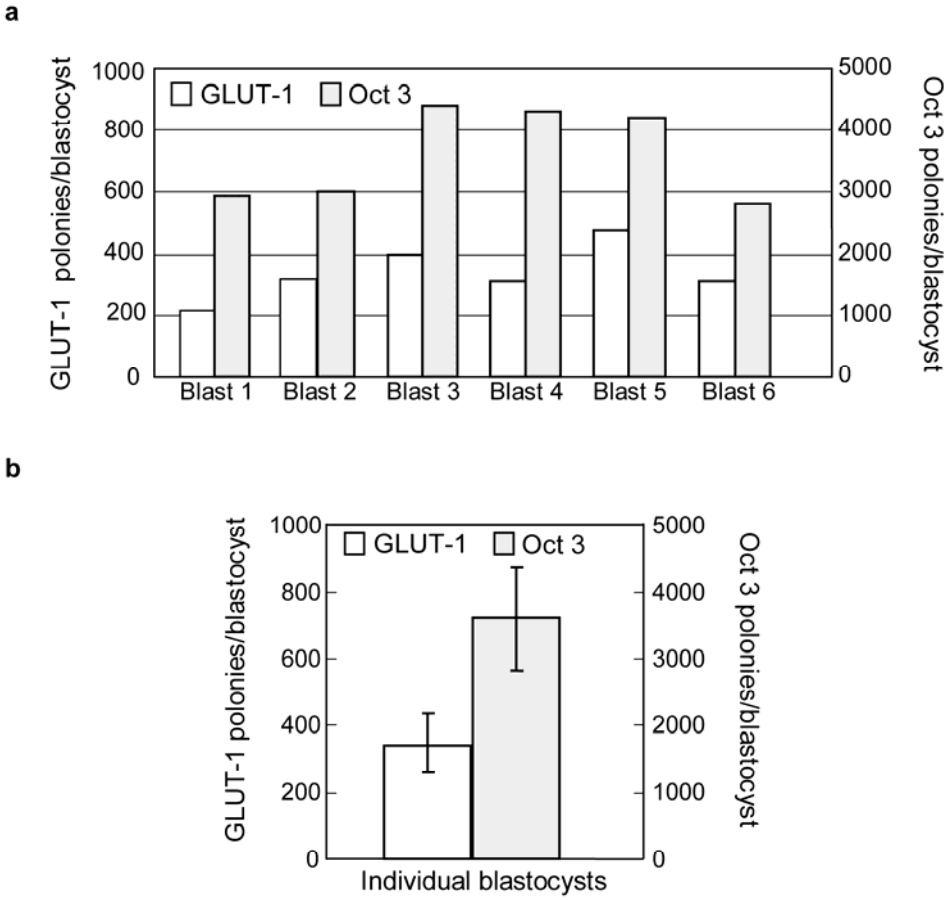


Figure 2.5: GLUT-1 Expression Assayed by Polonies. (a) GLUT-1 and Oct3 polonies per slide for six individual blastocysts. Each data point represents an average of two replicate slides for GLUT-1 and one slide for Oct3. All slides contain one-fifth of the cDNA from a single blastocyst. (b) Comparison of GLUT-1 and Oct3 polonies per blastocysts for averaged individual samples.

GLUT-1 assays were done on six individual blastocysts and Oct3 was measured as a control. GLUT-1 polonies are present in each blastocyst with an average of 348 ± 84 polonies/blastocyst. The blastocysts had 3340 ± 674 polonies for Oct3, in accordance with previous experiments. We conclude that all the blastocysts tested express GLUT-1 and that the polony method is suited for analysis of this gene.

In order to further validate the use of polonies for small numbers of cells a direct comparison with an established PCR method was performed. Competitive PCR was chosen as the standard method because of its sensitivity and rigorous quantitative design (Zentilin and Giacca 2007). Assays were done on ES cells for Oct3, Nanog and Rex1 expression by polonies and competitive PCR and the results compared (Figure 2.6).

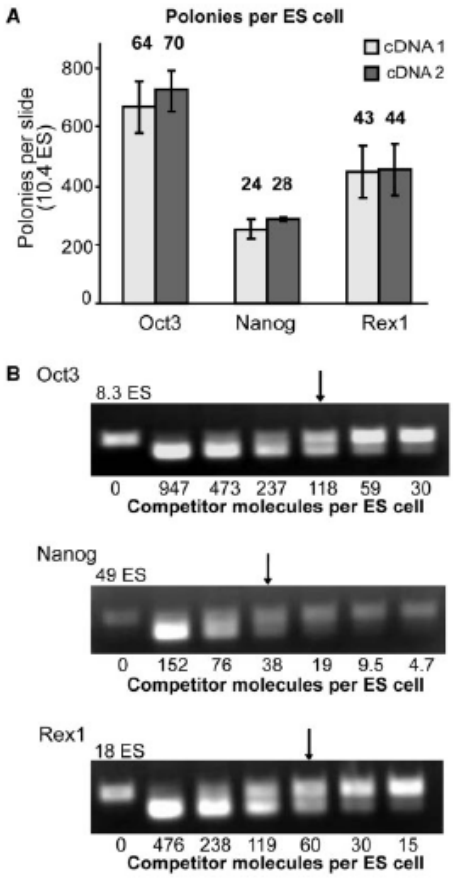


Figure 2.6: Polonies and competitive PCR for three genes. (A) Polonies per slide and per ES cell equivalent for Oct3, Nanog, and Rex1. The average number of polonies and standard deviation of three replicate slides containing 10.4 ES equivalents are shown. Calculated number of polonies per cell for each set of slides is indicated. (B) Competitive PCR for each of three genes. DNA competitors with 50-bp deletions were generated for Oct3, Nanog, and Rex1. Competitive PCR reactions with the indicated number of ES cell cDNA equivalents and varying amount of competitor is shown.

Polonies were counted on slides containing cDNA from 10.4 ES cell equivalents for the three genes. Average and standard deviation of polony counts for three replicate slides and calculated polonies per ES cell equivalent are shown in Figure 2.6. The polony method shows an average of 67 Oct3 cDNAs per cell, 26 Nanog cDNAs per cell, and 43 Rex1 cDNAs per cell. Competitive PCR gels for each of the three genes are also shown in Figure 2.6. Note that the number of ES cell equivalents used to obtain an equivalence point using PCR differed for each of the three genes. Using competitive PCR we obtain an estimate of 118 Oct3 cDNAs per cell, 38 Nanog cDNAs per cell and 60 Rex1 cDNAs per cell. The number of polonies per ES cell is thus similar to the number of cDNAs measured by competitive PCR for each of the three genes. RT controls for each gene using competitive PCR and polonies showed no background. In summary, polony assays and competitive PCR assays give comparable results

The numbers of polonies per cell is less than the actual number of mRNAs per cell due to inefficiencies in extracting mRNA and reverse transcription of mRNA to cDNA. Determining the efficiency from RNA to cDNA (reverse transcription) is a step towards extrapolating polony counts to actual number of mRNAs present in a cell. To this end a model RNA was constructed, a known amount reverse transcribed and the efficiency of the reaction determined with polonies (Figure 2.7).

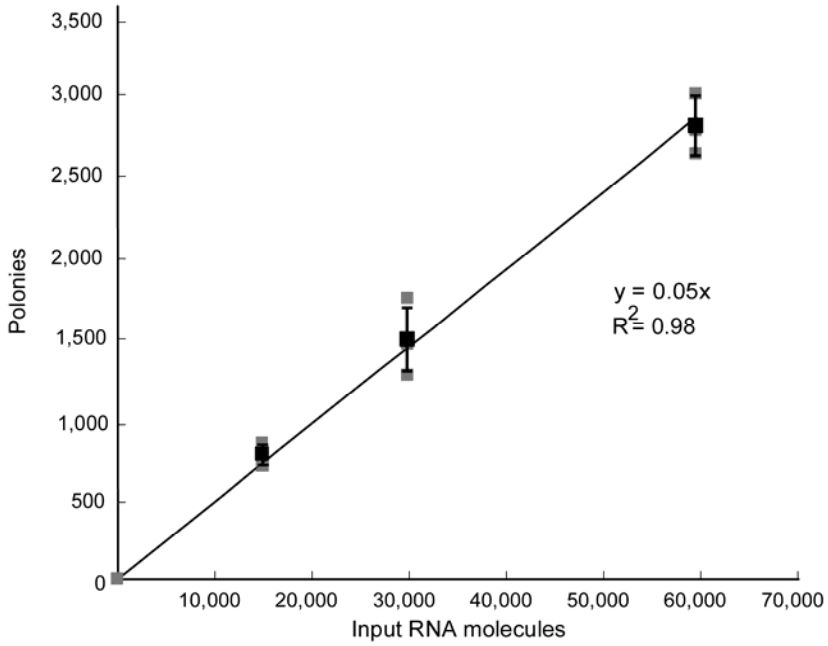


Figure 2.7: Efficiency from RNA to Polony. The number of starting RNA molecules is plotted against polony output for three levels of RNA input. The number of polonies increased linearly with the number of RNA molecules added to the polony reaction. Polony counts from each slide are shown by a gray box. Mean values and error bars (standard deviation) for each set of three slides at a particular dilution are shown in black.

A plasmid for generating model RNA was constructed by joining a yeast gene (BnI5) to the poly(A+) rich region from the Xenopus elongation factor-1α gene. The model RNA (1.6kb containing A_{70}) was synthesized by T7 polymerase. For three dilutions of model RNA, the number of polonies increased linearly with increasing amount of template (Figure 2.7). RNA template conversion to polony ranged from 4.8%-6.1% on individual slides and averaged 5%. Control polony slides without RNA did not produce polonies proving that the polony reaction is specific to the model RNA; RT controls were also negative. To explore the generality of this finding, polonies for other regions of this model RNA were tested. Efficiency from RNA to polony for these other amplicons was similar (data not shown). These data are in good agreement with measurements of RT

efficiency in the literature (Dufva, Svenningsson et al. 1995). Recently, differences in the efficiency of reverse transcription among templates have been shown (Warren, Bryder et al. 2006), although the reasons for the variability of the RT step have not been discovered.

2.5 Discussion

The polony method of analysis was adapted for use with small numbers of stem cells. The method is sensitive, can be applied to most genes, and allows a degree of multiplexing; it gives comparable results to competitive PCR, an established method for quantifying cDNAs (Zentilin and Giacca 2007). The approach is also amenable to future refinements that will extend its powers.

The method is sensitive enough to detect mRNAs in fractions of a single mouse blastocyst which is comprised of only 75-100 cells. Specifically we have detected mRNAs in as little as 1/5 of a single blastocyst. In the case of Oct3 expression is confined to the ICM, which is comprised of about 35 cells, demonstrating the method is sensitive to 7 cells (1/5 of 35 cells) for this particular RNA. It is significant that the number of Oct3 polonies per ES cell (~50) predicts that there would be about 1750 polonies per blastocyst (50/cellX35 ICM cells/blastocyst) a number close to what is measured. The generality of the method was demonstrated by performing assays on five separate genes representing two classes: transcription factors and a membrane transporter. They also include genes exclusive to the ICM (Oct3 and Nanog), a mRNA expressed in both ICM and TE (Rex1) and a mRNA expressed specifically in the TE (Cdx2) (Deb, Sivaguru et al. 2006). Taken together, these results suggest the method will be applicable to most genes of interest. The number of mRNAs present per cell is likely to be greater than the number of polonies due

to losses of mRNA in extraction and inefficiency in conversion of mRNA to cDNA by reverse transcriptase. Future developments of the method are needed to discover the efficiencies of the steps leading up to polonies.

In this study we measured the mRNA from three genes from individual blastocysts by performing parallel assays on fractions of the cDNA from a single blastocyst. Polonies for multiple templates can be analyzed on the same slide by including multiple primer pairs (Mitra, Butty et al. 2003) so it is likely that as many as ten genes can be amplified by a simple extension of the method we used. Much greater increases in the number of genes that can be assayed might be achieved by using universal amplifying primers and applying fluorescence in situ sequencing of the polonies (Mitra, Shendure et al. 2003). Thus future enhancements of our method could easily assay dozens of genes per blastocyst.

In summary, the results of these studies show that the polony approach may be applied to the problem of stem cell expression profiling and should encourage efforts to further develop this system for the special needs of stem cell biology.

Chapter 3

From ES to Neural Cells: DNA Methylation as a Marker of Neural Cell Fate

3.1 Abstract

The aim of this work was to test the hypothesis that DNA methylation represses transcription in genes involved in neural cell fate specification. We investigated the correlation between CpG methylation and silencing of gene expression by mapping methylation in key developmental and neural fate determining genes in expressing and non-expressing cells and tissues. The direct bisulfite sequencing method employed mapped methylation loci more quickly and cheaply than conventional clonal bisulfite sequencing, appropriate for our semi-quantitative focused analysis of a few genes along a pathway. We found that Oct3/4 and Pgk1 promoter methylation was correlated with silencing, but methylation status of neural fate determining genes Sox1, Olig1, and Olig2 was not indicative of expression. This analysis showed that DNA methylation of neural fate determining genes was not associated with transcriptional repression, refuting the hypothesis. The leading hypothesis for the function of DNA methylation is under question and tools to uncover its role have not been developed.

Deany Delaney assisted with experiments performed in Chapter 3. She performed differentiations and prepared DNA. Methylation analysis was done autonomously.

3.2 Introduction

DNA methylation is both precisely distributed and dynamic in the mammalian genome. In mammals DNA methylation is found exclusively on the 5-carbon of cytosine residues and is confined to CpG dinucleotides, a cytosine joined by a phosphodiester bond to a guanine. Methylation at CpG sites is symmetric; CpGs are either methylated or unmethylated on both strands (Bird 1978). Overall, CpG dinucleotides are underrepresented, present at one-fourth of expected frequency based on overall base compositions (Bird 1980; Jabbari and Bernardi 2004), and are highly methylated with 60-90% of CpG sites methylated (Ehrlich, Gama-Sosa et al. 1982; Ehrlich 2003). Genome-wide methylcytosine levels vary significantly from 0.7 – 1% in different cell types with somatic tissues having higher overall methylation levels than both germ cells and placental tissues (Ehrlich, Gama-Sosa et al. 1982; Ehrlich 2003). Not only is DNA methylation essential for proper embryonic development (Frommer, McDonald et al. 1992; Li, Beard et al. 1993; Okano, Bell et al. 1999), periods of demethylation and remethylation occur during specific stages of early embryonic development (reviewed by (Ehrlich 2003)). Because DNA methylation is necessary for proper development, and actively changing during development, as well as exhibiting tissue specific differences, it was hypothesized that methylation was involved in the regulation of cell fate.

A prevailing hypothesis in the field is that DNA methylation functions to silence transcription. Many studies have shown that DNA methylation correlates with decreased transcription and gene expression. Artificially methylated promoters show decreased transcription factor binding in cell extracts (Watt and Molloy 1988; Boyes and Bird 1991)

and decreased expression when transfected into cells (Boyes and Bird 1991). Embryonic cells methylate retroviral elements rapidly silencing active retroviruses (Jahner, Stuhlmann et al. 1982). In imprinting, the silent allele is preferentially methylated while the unmethylated allele is actively transcribed (Li, Beard et al. 1993). Expression of Xist, the X-inactivation RNA, is also silenced by methylation (Panning and Jaenisch 1996), and CpG Islands are methylated on the silent X chromosome (Wolf, Jolly et al. 1984). Expression is restored from inactive retroelements (Walsh, Chaillet et al. 1998), imprinted alleles (Li, Beard et al. 1993), X-linked and developmental genes (Fouse, Shen et al. 2008) when methylation is inhibited, when cells are treated with *5*-deoxyazacytidine or have knocked-out DNA methyltransferases, providing strong evidence that methylation is correlated to gene silencing.

Therefore, there is great interest in mapping DNA methylation genome-wide and in localized regions in various cell types. Initial studies digested of genomic DNA with methylation sensitive restriction enzymes revealing CpG sites, are unevenly distributed into two distinct populations in the mammalian genome: CpG Islands, defined as unmethylated CpG-rich DNA (Bird, Taggart et al. 1985) and heavily methylated rest of the genome. With the availability of DNA sequence data, CpG Islands were redefined *in silico*, as CpG-rich DNA independent of methylation status, based on sequence GC content (>50%) and observed CpG frequency (obs /expect >0.6) over a length window of 200 bases or greater ((Gardiner-Garden and Frommer 1987), now >500bp). CpG islands were found to be associated with the promoter regions of housekeeping genes and 56% of human genes (Gardiner-Garden and Frommer 1987; Antequera and Bird 1993); these regions were typically unmethylated and associated genes were transcribed. Methylation mapping of

specific genes within the context of developmental pathways also provided support for the role of DNA methylation in silencing. Mapping Glial Acidic Fibrillary protein (GFAP), revealed an unmethylated CpG site in its promoter in astrocytes, which expressed the gene, and the same site was methylated in neurons and other non-expressing tissues (Takizawa, Nakashima et al. 2001), providing additional support for the hypothesis that methylation is involved in silencing.

Mapping DNA methylation was revolutionized by the bisulfite conversion method and PCR (Frommer, McDonald et al. 1992; Clark, Harrison et al. 1994) enabling analysis of localized regions at single base resolution. With the completed human and mouse genomes, genome-wide methylation mapping is now being revisited using direct bisulfite sequencing (Eckhardt, Lewin et al. 2006). Representative regions of the genome have been mapped by first performing enrichment by affinity (Illingworth, Kerr et al. 2008) or digestion with methylation-sensitive restriction enzymes followed with bisulfite treatment and PCR. Methylation status containing sequence information is then obtained by application of PCR products to arrays (Shen, Kondo et al. 2007) or to short-read sequencing reactions (Meissner, Mikkelsen et al. 2008). This recent flood of genome-wide methylation data is revealing exceptions to once held generalizations in the field (Reviewed in (Suzuki and Bird 2008)).

The coverage and scope of genome-wide methylation maps are rapidly growing, yet these maps are incomplete. Focused methylation mapping still needs to be performed to elucidate the relationship between gene and tissue specific methylation patterns and gene regulation within the context of specific pathways. Traditionally, bisulfite treatment and gene specific PCR are coupled with cloning of PCR fragments and a small number of

clones are sequenced to get a quantitative estimate of methylation in a gene of interest. Direct bisulfite sequencing, where PCR products obtained following bisulfite treatment are sequenced directly, is a quick and cheap alternative for assaying methylation in a small number of loci in carefully chosen tissue types and provides a semi-quantitative picture of methylation for the pathway of interest.

The aim of this work is to test the hypothesis that DNA methylation is associated with repressed transcription in genes involved in neural cell fate. We investigated the correlation between CpG methylation and silencing of gene expression by mapping methylation in several regions of key developmental and neural fate determining genes in expressing and non-expressing cells and tissues by direct bisulfite sequencing. Two genes studied here support the hypothesis. The promoter of Pgk-1, a constitutively expressed housekeeping gene, is unmethylated in all cells and tissues; and the promoter of Oct3/4, a pluripotency gene, is unmethylated in ES cells where it is expressed, and becomes increasingly methylated in differentiated cells and tissues where Oct3/4 is silenced. The three neural fate determining genes analyzed: Sox1, Olig1, and Olig2 do not support the hypothesis as the regions mapped were largely unmethylated in cells and tissues that do not express these genes. The current study adds to the growing epigenome map and refutes the leading hypothesis that DNA methylation marks repressed transcription. Currently, there are no strong alternative hypotheses for the function of DNA methylation.

3.3 Materials and Methods

ES cell culture

All ES cell experiments were done with the RW4 line of ES cells derived from Sv129 mice. A set of experiments with cell line G-Olig2, where GFP has been knocked into the Olig2 locus of RW4 ES cells (Xian, McNichols et al. 2003), was also performed. Undifferentiated (ES) were cells were grown in DMEM supplemented with 10% Fetal Bovine Serum (FBS, Sigma), 10% Newborn Bovine Serum (NCS, Sigma) and supplemental nucleosides on gelatin-coated tissue culture plastic in the presence of leukemia inhibitory factor (LIF 1000U/ml, Invitrogen) and beta-mecaptoethanol (0.11mM, Sigma) (Bain, Kitchens et al. 1995; Bain 1998).

Neural differentiation

To generate neural cultures, undifferentiated ES cells were washed twice with DMEM/F12 medium and scraped off the bottom of flasks in the presence of 6mL DFK5. Cell clumps were triturated and then cultured for 2 days in 10ml of M-DFK5 medium in non-adhesive agarose coated tissue culture treated dishes. M-DFK5 medium consists of DMEM/F12 supplemented with 1X non-essential amino acids (NEAA, Invitrogen), 1x nucleosides, 0.1mM beta-mercaptoethanol, 50ug/mL transferin (Sigma), 5ug/mL insulin (Invitrogen), 30nM Na-selenite (Sigma) and 5% Knockout serum replacement (Invitrogen) as previously described (Bain, Kitchens et al. 1995; Bain 1998; Zhang, Horrell et al. 2008). EBs were transferred to 6 well gelatin coated dishes in M-DFK5 in the presence of retinoic acid (RA 2uM) and sonic hedgehog agonist (Shh 30nM, Curis Inc) for 4 days to generate neural cultures.

Hematopoetic differentiation

Hematopoietic precursors were generated by switching undifferentiated ES cells from DMEM to ES-IMDM media two days prior to differentiation. ES-IMDM media consists of IMDM supplemented with 15% FBS, 1% NEAA, 1% L-Glutamine (Invitrogen), Monthioglyceriol (MTG .15mM, (Sigma)) and LIF (1000U/ml). To begin differentiation, ES cells were washed in HEPES and then trypsinized (0.25% Trypsin/EDTA(Sigma)) for 1-2 minutes until cells just started to detach from the surface. FBS (1mL) and IMDM (5ml) were added to stop the action of trypsin. Cells were then washed with IMDM, resuspended in 5ml IMDM with 10% FBS, and 20,000 cells were added to an agarose coated dish containing 10ml of B-differentiation media and grown for 6 days. B-differentiation media consists of IMDM supplemented with 15% FBS, 1% Ascorbic acid (Sigma), 2mM L-glutamine, 0.039uL/mL MTG, Bone morphogenic protein (BMP-4 (10ng/mL) (R&D Systems)), and Vascular Endothelial Growth Factor (VEGF (5ng/mL) (R&D Systems)) (Lugus, Chung et al. 2007).

Fluorescence activated cell sorting of G-Olig2 cells

G-Olig2 ES cells were differentiated into neural precursor cells for either 4 or 6 days as described and then sorted on the MoFlo or Turbo sorter (Flow cytometry core, Washington University School of Medicine). The highest expressing cells (20%) were sorted from the lowest (20%) expressing cells and DNA was harvested.

Isolation of DNA from Cells and Tissues

Cells were harvested by trypsinization at appropriate differentiation stage. DNA was isolated from 1-2 million cells per reaction. Approximately, 70-150mg of heart, liver, and lung tissue were harvested from an anesthetized mouse. For all samples DNA was

purified using the ArchivePure DNA purification kit (5 Prime) according to manufacturer's instructions.

Construction of *in vitro* methylated template

Plasmid pBluescriptIISK+ was used to construct mock unmethylated and methylated templates to assess the efficiency of bisulfite treatment. An initial PCR reaction was performed to obtain a linear unmethylated template. This linear template was gel purified to remove primers using a Qiaquick gel purification kit (Qiagen) and then methylated at CpG sites *in vitro* at 37°C for two hours using m.SSSI methyltransferase (New England Biolabs). The reaction was conducted in 10 mM Tris-HCl, 50 mM NaCl, 1 mM Dithiothreitol pH 7.9 @ 25°C in the presence of s-adenosylmethionine (SAM 160uM). The methylated template was then cut with BstUI which is sensitive to methylated CpG, and full-length methylated template was gel purified (Qiagen). Controls verifying the methylation specificity of m.SSSI and cofactor SAM were performed in parallel. Primers for construction and validation are given in Appendix C, Table C.1.

Treatment with Sodium Bisulfite

DNA was treated with sodium bisulfite using EZ DNA methylation kit (Zymo Research). Modified conversion conditions were applied. Approximately 500ng of DNA were added to 7.5ul of M-Dilution buffer in a total volume of 50ul and incubated at 42ºC for 30 minutes to denature the DNA. Following denaturation, 97.5ul CT conversion reagent (resuspended in 750uL $H_2 0$ and 185uL M-dilution Buffer) were added and bisulfite treatment was performed overnight in a thermocycler for 16 cycles of 95 ºC for 30 seconds and 50 ºC for 1 hour. DNA was stored at 4 ºC until washing and desulphonation were

completed the following morning according to manufacturer's instructions. Bisulfite treated DNA was eluted from spin columns in 20ul of 10mM Tris 1mM EDTA.

PCR Amplification and Purification

PCR was carried out on 7-9uL (~140-180ng) of bisulfite treated DNA in a total reaction volume of 120ul using primers indicated in Appendix C, Tables C.1 and C.2. Gene specific PCR primers were designed using MethPrimer (Li and Dahiya 2002), a program based on Primer3 to design bisulfite specific PCR primers (http://www.urogene.org/methprimer/), and perl scripts were used to append universal sequencing tails. PCR reactions were performed in 1X Jumpstart Taq buffer with $MgCl_2$ at a final concentration of 1.5mM, Jumpstart Taq polymerase, 0.5uM each primer, 200uM dNTP mix, and Jumpstart Taq to 0.05U/ul supplemented with BSA (1X) and Betaine (1.6M). Thermocycling parameters were as follows: 94 ºC for 2 minutes, 94 ºC for 30 seconds, 50 ºC for 30 seconds, 72 ºC for 1 minute for 43 cycles with a final extension of 72 ºC for 5 minutes. PCR products were concentrated and purified from primers using a 1.5% agarose gel and Qiaquick gel purification kit (Qiagen) according to manufacturer's instructions, or gel purified by pipette from a 1.5% agarose gel with a 0.2% low melting temperature agarose window (Ma and Difazio 2008).

Sequencing and Analysis

Gel-purified PCR products (~8ng) were sent for sequencing at PNACL facility at Washington University School of Medicine. DNA was sequenced using ABI V1.1 from both ends. Trace files and called bases were downloaded from the PNACL server. Several commercially available and custom software programs were used to analyze data. First, traces were viewed in Gap4 databases assembled using Pregap4 from the Staden Package

((Staden, Beal et al. 2000) http://staden.sourceforge.net/). Following visual inspection, called sequences were analyzed using a combination of Perl scripts and ClustalX ((Thompson, Higgins et al. 1994), http://bips.u-strasbg.fr/fr/Documentation/ClustalX/#R) to align called bases with a known reference sequence and quantify bisulfite conversion. Further sequence analyses were performed using CpG Viewer ((Carr, Valleley et al. 2007), http://dna.leeds.ac.uk/cpgviewer/), which facilitated alignment and visual inspection of traces, called bases, and construction of lollipop diagrams. CpG dinucleotides were called methylated when the highest peak and called base at a particular position was cytosine followed by guanine in the sense direction and unmethylated when the highest peak and called base was thymine followed by guanine in the sense direction. A site was considered heterogeneous when reads from sense and antisense directions disagreed; most often this was a result of dual peaks (C/T) of near equal height.

3.4 Results

We investigated promoter CpG methylation status of three genes to test the hypothesis that promoter methylation is correlated to gene silencing in neural lineage specification. Methylation was mapped in loci of a housekeeping gene, Pgk-1 and several developmental regulatory genes, Oct3/4, Sox1, Olig1, and Olig2 in expressing and non-expressing tissues; we expected all cells and tissues that did not express a particular gene to be methylated. We began by first evaluating the direct bisulfite sequencing method, so that it could be applied to methylation analysis of biological samples.

3.4.1 Direct bisulfite sequencing is sensitive and specific, but does not quantify heterogeneity

To examine the sensitivity and specificity of the direct bisulfite sequencing method, a model linear 2.5kb template with known methylation status was analyzed. Sequencing was performed on 5 regions of unmethylated, unmethylated bisulfite treated, and in vitro CpG methylated bisulfite treated template using 5 amplicons of approximately 300bp, the same size as were used in the gene specific studies described. The efficiency of bisulfite conversion [#C's converted to T's / total number of C's sequenced] and extent of protection by methylation [#CpG protected / total # of CpG's sequenced] was determined using called sequences and shown in Table 3.1.

Table 3.1 In vitro methylation experiment

	% Bisulfite Conversion	% Methylation
Unmethylated	17 / 768 = 0.022	185 / 188 = 0.984
Unmethylated and bisulfite treated	779 / 782 = 0.996	0 / 196 = 0
CpG methylated and bisulfite treated	587 / 786 = 0.747	190 / 195 = 0.974
	[#C's converted to T's / total number of C's sequenced]	[#CpG protected / total # of CpG's sequenced]

The unmethylated, untreated template shows the rate of sequencing errors (false positives) that falsely predict conversion (0.022) and CpGs that should have been measured (1-0.984= 0.016, false negatives). Bisulfite conversion is very efficient, converting 0.996 of cytosines to called thymines. The method was also very sensitive in detecting completely *in vitro* methylated template, as the percent of CpGs measured is close to that of the

unmethylated, untreated template (185/188 unbisulfite treated, 190/195 methylated and bisulfite treated, $p>0.1$ binomial test). Methylation specific to CpGs will protect against bisulfite treatment 97% of the time while 99% of unmethylated CpGs will be converted in the presence of methylated CpGs (587/591 cytosines that are not CpGs are converted in M. SSSI methylated template). Errors generally occurred within the low-quality, first 50bp of sequences and must be taken into account when analyzing direct bisulfite sequencing data from biological samples. These errors can be further minimized by sequencing overlapping reads from both directions. Use of universal sequencing tails facilitated easy set up of sequencing reactions and improved amplification efficiency and sequence quality. This study shows that the direct bisulfite technique employed is very effective at detecting a completely methylated template. A thorough quantitative analysis based on mixtures of methylated and unmethylated templates, mimicking actual biological conditions was not performed. However, this analysis gave a strong indication that the direct bisulfite method was sensitive and specific enough to perform analysis on biological samples.

3.4.2 Promoter methylation correlates to silenced Oct3/4, but not Olig2

In the first study, direct bisulfite sequencing was used to measure methylation in promoter regions of Pgk1, Oct3/4, and Olig2 genes in ES, ES derived neural (neural), and ES derived hematopoietic (blood) cells, and adult mouse heart, lung, and liver tissue. Isolated DNA was treated with sodium bisulfite and amplified by PCR. PCR products were gel purified and sequenced (Figure 3.1). PCR primers for the promoter analysis are given in Appendix C, Table C.2.a.

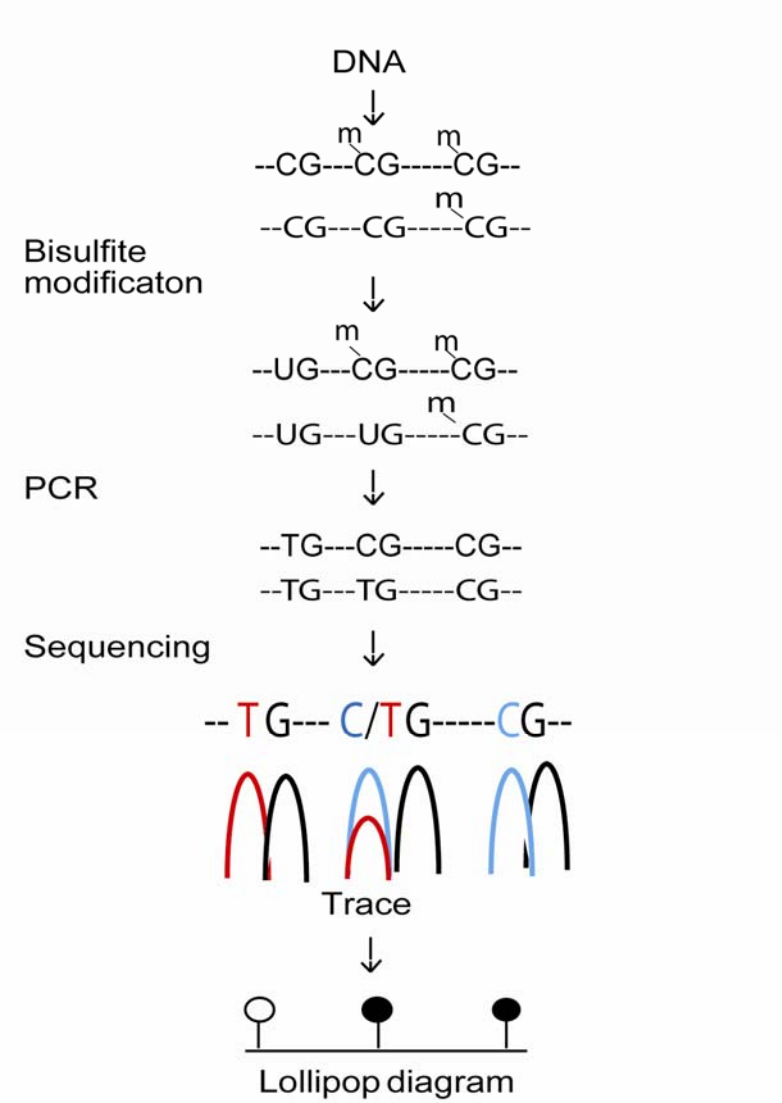


Figure 3.1: Methods for direct bisulfite sequencing. DNA is isolated from cells and tissues treated with sodium bisulfite. Methylated cytosines are protected whereas unmethylated cytosines are converted to uracils. Regions of interest are then PCR amplified with gene specific primers. Purified PCR products were Sanger sequenced. Traces show unmethylated CpGs as TpGs and methylated CpGs as CpGs. At heterogeneous sites, methylation status is determined by the tallest peak at each CpG dinucleotide. In lollipop diagrams methylated CpGs are filled circles and unmethylated CpGs are open circles.

The three genes studied have unique structures (Figure 3.2) and different expression profiles.

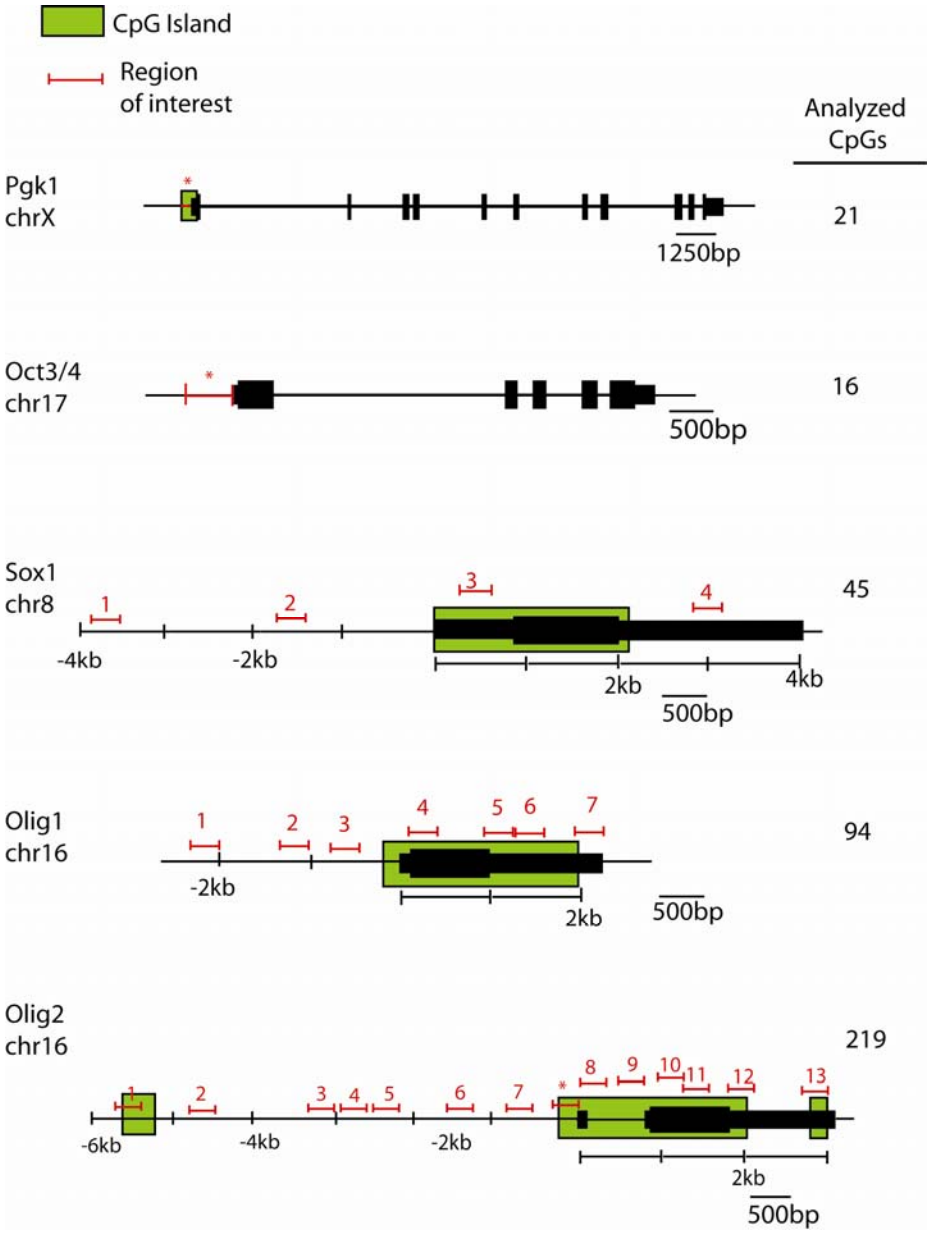


Figure 3.2: Structures of all genes studied: Pgk-1, Oct3/4, Sox1, Olig1, and Olig2. Exons are indicated by black rectangles with 5' and 3' untranslated regions by slightly smaller rectangles shown. Introns are shown by lines connecting exons. Called CpG Islands (UCSC browser) are shown in green. Regions of interest are given by red bars. Proximal promoter regions are denoted by an asterisk (*). Numbers over regions of interest correspond to regions described in the text. The total CpG sites assayed are indicated.

The Pgk-1 gene encodes the enzyme 3-phosphoglycerate kinase, a house-keeping gene located on the X chromosome where it is constitutively transcribed in all tissues from the active X (McBurney, Sutherland et al. 1991). Oct3/4, a transcription factor and pluripotency regulator, is expressed in the inner cell mass of the blastocyst and in ES cells. Oct3/4 plays an important role in development turning off as cells acquire somatic phenotypes (Pesce and Scholer 2001). On the other hand, Olig2 is a transcription factor expressed in the developing nervous system in oligodendrocyte precursor cells and in oligodendrocytes of the mature nervous system ((Lu, Yuk et al. 2000); Reviewed in (Ligon, Fancy et al. 2006)); Olig2 is not expressed in pluripotent cells. Expression of Olig2 is highly specific to subsets of cells at each stage of nervous system development (Takebayashi, Nabeshima et al. 2002).

Promoter CpG methylation was investigated in three genes: Pgk-1, Oct3/4 and Olig2 by direct bisulfite sequencing. Bisulfite sequencing traces revealed complete conversion of non-CpG cytosine residues in all three genes, indicating effective bisulfite treatment. Pgk1 and Olig2 promoters were unmethylated at all CpG sites examined in all cells and tissues. Only the Oct3/4 gene showed differential methylation across the different cell types (Figure 3.3).

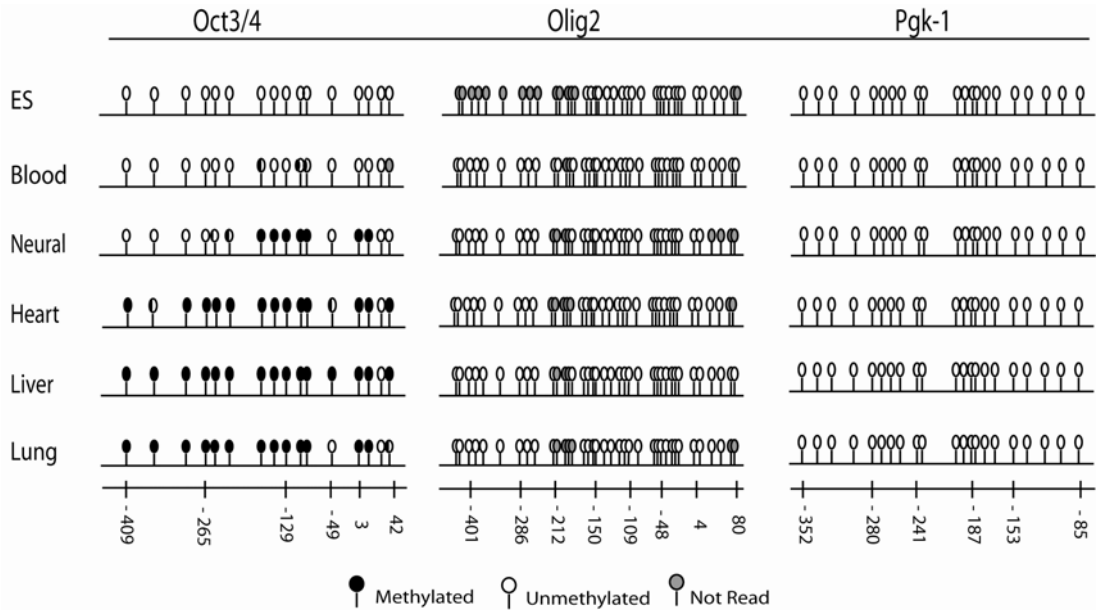


Figure 3.3: Promoter region methylation of Oct3/4, Olig2, and Pgk-1 in cultured cells and mouse tissues. Lollipop diagram of methylation in ES, ES derived neural, ES derived hematopoietic cells, heart, liver, and lung tissues by direct bisulfite sequencing. Partially filled circles indicate replicate reads with different calls. Oct3/4 promoter shows increased methylation in differentiated cells and near complete methylation in tissues whereas Olig2 and Pgk-1 promoter regions remain unmethylated in all cells and tissues.

Of the 16 Oct3/4 CpG sites assayed, all were unmethylated in ES cells. Blood progenitors showed methylation heterogeneity at three Oct3/4 sites (-167, -105, -101) proximal to the transcription start site. Heterogeneous bases, indicated by half filled circles are where dual C/T peaks were present and called bases differed in forward and reverse sequences (Figure 3). Neural cells had seven Oct3/4 methylated sites (-167 –
-107, 3, 18), two heterogeneous sites (-253,-229), and four called unmethylated sites that showed some methylation (-265, -49, +37, +42). The remaining three distal downstream sites (-409,-360,-298) have no evidence of methylation. The Oct3/4 promoter became increasingly methylated in adult tissues where Oct3/4 is silenced. For heart, liver, and lung

tissue 13/16, 15/16, and 13/16 CpGs showed methylated sites, with methylation spanning the entire promoter region. Heart tissue had many called methylation sites showing a mixture of methylated and unmethylated CpGs. Liver had the strongest methylation signal with 15/16 sites producing methylated CpGs. Even the only called unmethylated site (+37) showed a moderate dual C/T peak. The Oct4 promoter was also highly methylated in lung tissue with methylation detected at 13/16 sites, one heterogeneous CpG site (+42), and the remaining two unmethylated CpGs (-49, +37) showing dual C/T peaks, with the thymine peak slightly taller than the cytosine peak. In conclusion, our Oct3/4 results support the established hypothesis that DNA methylation is associated with transcriptional repression; Oct3/4 becomes methylated when the gene is no longer expressed, as pluripotent cells differentiate into the somatic lineages of the body.

3.4.3 The Olig2 promoter is not methylated in Olig2 expressing and non-expressing neural cells

Since methylation was not observed in the Olig2 promoter region in heart, liver, and lung tissues that have silenced Olig2, an additional study was conducted with G-Olig2 ES cell line to definitively determine if methylation is correlated to Olig2 expression in ES derived neural cells. Olig2 expression is easily monitored by GFP expression in the G-Olig2 cell line so that methylation could be measured in neural cells that expressed and did not express Olig2. G-Olig2 ES cells underwent neural differentiation in the presence of retinoic acid and sonic hedgehog agonist for 4 or 6 days as previously described. A sub-population of the neural cells, approximately 50%, were bright green under the fluorescence microscope indicating Olig2 was being expressed in agreement with

previously published results (Xian, McNichols et al. 2003). FACS sorting was used to separate Olig2 expressing from non-expressing cells and Olig2 promoter methylation was assayed. Like the cells and tissues studied in Figure 3.3, both Olig2 expressing and non-expressing neural cell populations showed no methylation in the Olig2 promoter region, refuting the hypothesis that methylation is associated with silenced Olig2 transcription and showing that methylation is not associated with Olig2 expression.

3.4.4 The Olig2 locus is hypomethylated in normal cells and tissues

Since Olig2 promoter methylation was not found in any cells or tissues assayed, we hypothesized that alternative regions may be responsible for methylation correlating with Olig2 regulation. Like Olig2, Sox2 is transcription factor expressed in the developing central nervous system involved in neural differentiation; its promoter does not have methylation correlated to silencing. Rather, methylation of an upstream epigenetic enhancer 4kb upstream is correlated to cell type specific Sox2 expression (Sikorska, Sandhu et al. 2008). Gene body methylation and methylation of 3'UTR regions have also been shown to occur and correlate to silencing in some genes (Suzuki and Bird 2008).

We proceeded to map methylation across the Olig2 locus in ES derived neural cells and in heart, liver, and lung tissues by direct bisulfite sequencing. Results of the analysis are shown in Table 3.2. Primers for analysis are given in Appendix C, Table C.2.b.

Table 3.2: Olig2 methylation in ES derived neural cells and non-expressing tisssues

ROI	1	2	3	4	5	6	7	*	8	9	10	11	12	13
Location [start end]	-5254 -5516	-4851 -4551	-3376 -3025	-2982 -2768	-2697 -2399	-1528 -1221	-881 -598	-450 100	234 609	526 779	950 1161	1135 1354	1841 2054	2679 2914
Neural Cells	0/16	1/9	------	0/13	0/7	1/10	0/10	0/28	0/19	0/10	0/19	0/17	0/4	0/8
Heart	0/16	0/9	0/18	0/13	0/13	0/9	0/10	0/30	0/19	0/10	0/19	0/20	0/11	0/8
Liver	0/15	1/9	0/19	0/13	0/11	0/10	0/10	0/34	0/19	0/10	0/19	0/20	0/7	0/8
Lung	0/15	0/9	0/15	0/5	0/8	0/10	0/10	0/33	0/19	0/10	0/10	0/16	0/11	0/8
# sites	18	9	19	13	13	10	10	37	19	10	19	21	11	10

Each measurement corresponds to the number of methylated CpGs/ total number of CpGs read.
*promoter region (see Figures 3.2, 3.3)

As shown in Table 3.2, bisulfite sequencing revealed that the Olig2 locus is largely unmethylated, as an additional 182 CpG sites were surveyed and only 2 methylated CpG sites were detected. Liver tissue and neural cells both had a methylated CpG at (-4795). The neural cells also had methylation heterogeneity at (-1426) whereas all cell types were unmethylated at this CpG site. In conclusion, there was no correlation between Olig2 silencing and methylation status, refuting the hypothesis that methylation is correlated with silencing.

3.4.5 Methylation of neural transcription factors Sox1 and Olig1 does not correlate to silencing

Although methylation was minimal across the Olig2 locus, there is evidence that DNA methylation of neural genes, such as brain-derived neurotrophic factor (BDNF) (Martinowich, Hattori et al. 2003) and in the astrocyte lineage GFAP (Takizawa, Nakashima et al. 2001), plays a role in cell fate specification. The next set of experiments focused on determining if methylation was at all correlated to gene silencing in non-neuronal cells and tissues in Sox1, a transcription factor instrumental in establishing early neural fate, and Olig1, a transcription factor closely related to Olig2 and oligodendrocyte precursor marker.

Sox1 is a transcription factor expressed in dividing neural cells. These neural progenitors are cells that are competent to become neurons; upon differentiation Sox1 is down-regulated (Pevny, Sockanathan et al. 1998; Kiefer 2007). Sox1 only has one exon which also coincides with a called CpG Island. Methylation was mapped in four regions

spanning the Sox1 locus (Figure 3.2). Primers are given in Supplemental Table C.2.c. A summary of all CpGs mapped is provided in Table 3.3.

Table 3.3: Sox1 methylation in cells and tissues

ROI	1	2	3	4
Location [start end]	-3725 -3412	-1609 -1312	316 635	2866 3183
ES cells	--------	7/7	0/23	5/5
Neural Cells	2/7	8/8	0/23	0/5
Heart	0/9	0/7	0/23	0/5
Liver	0/9	1/6	0/19	0/5
Lung	0/9	0/7	0/23	0/5
# sites	9	8	23	5

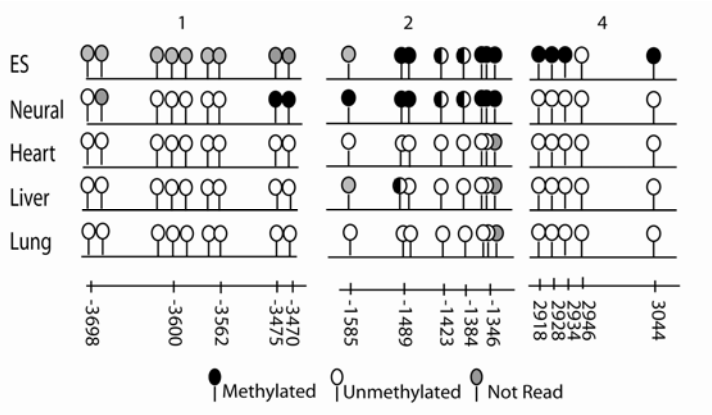
Fractions given are the number of methylated CpGs/ total number of CpGs read

As shown in Table 3.3, Sox1 upstream regions contained methylated sites in ES and neural cells, whereas the amplicon in the 5'UTR and called CpG island region was highly unmethylated in all tissues. The 3'UTR region, downstream of the CpG island, only showed methylation in ES cells. Detailed lollipop diagrams of differentially methylated sites are shown in Figure 3.4.

Neural cells show methylation at heterogeneous sites -3740 and -3745, while all other tissues are unmethylated; ES cells were not analyzed. The second upstream region contains the majority of differentially methylated sites with ES and neural cells having a high degree of methylation. In particular a pair of CpG sites, at -1489 and -1486 bases upstream, are highly methylated in ES and neural cells while -1489 is heterogeneous in liver; both sites (-1489, -1486) largely unmethylated in other tissues. Heterogeneity is seen at -1423 and -1384 with methylation being dominant in ES and neural cells and

unmethylation dominant in the tissues. The upstream triplet of CpG sites surrounding -1346 also showed methylation in ES and neural cells. Finally, unlike any of the other cell types and tissues analyzed ES cells had heterogeneously methylated sites covering the entire region in the 3'UTR, as five of the five sites were called methylated in separate reads (Figure 3.4). To summarize, methylation of the Sox1 locus did not occur in tissues that had silenced the Sox1 gene, refuting the hypothesis that methylation is correlated with transcriptional repression.

a.



b.

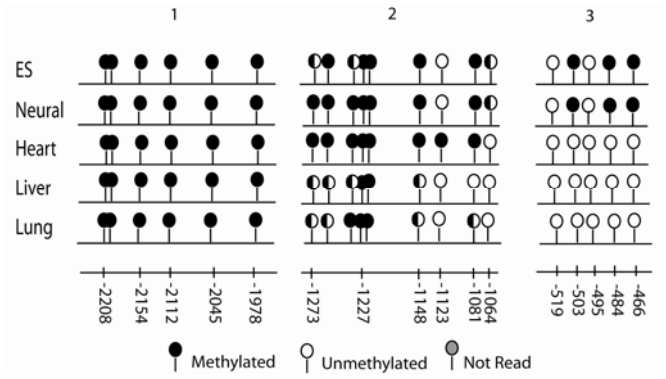


Figure 3.4: Methylation mapping of neural genes Sox1 and Olig1. a) Lollipop diagram for Sox1 differentially methylated regions. b) Lollipop diagram for Olig1 differentially methylated regions. Detailed methylation data is provided in Tables 3.3 and 3.4. Structural maps of both genes are provided in Figure 3.2.

Olig1 is a transcription factor involved in neural development. The gene is located 40 kb downstream of Olig2 in mice and in humans and is thought to have arisen out of a gene duplication event. Olig1 differs from Olig2 structurally in that it only has one exon (Figure 3.2). By itself, Olig1 promotes the formation of oligodendrocyte progenitor cells and mature oligodendrocytes in culture and in developing mice(Lu, Sun et al. 2002), but also functions in concert with Olig2 in establishing motor neuron and oligodentrocyte lineages in the primitive neural tube (Lu, Yuk et al. 2000; Ligon, Fancy et al. 2006).

Direct bisulfite sequencing was used to measure methylation of seven regions that span the Olig1 locus (Figure 3.2). Primers are given in Appendix C, Supplemental Table C.2.c. Based on the hypothesis, we expected to find that tissues that did not express Olig1, i.e. lung, liver, and heart would be highly methylated. We found that this was not true. An overview of all the sites analyzed is provided in Table 3.4.

Table 3.4: Olig1 methylation in cells and tissues

ROI	1	2	3	4	5	6	7
Location [start end]	-2235 -1955	-1324 -1009	-647 -399	202 450	883 1140	1149 1454	1756 2106
ES cells	6/6	8/9	3/5	------	0/8	0/16	1/13
Neural cells	5/5	8/9	3/5	0/14	0/9	0/15	0/20
Heart	5/5	8/9	0/5	0/20	0/9	0/15	0/20
Liver	6/6	6/9	0/5	0/15	0/9	0/15	0/24
Lung	6/6	7/9	0/5	0/20	0/12	0/16	0/24
# sites	6	9	6	20	13	16	24

Fractions are number of methylated CpGs/ total number of CpGs read.

Methylation was found in three regions upstream of the Olig1 transcription start site (Figure 4b) while all regions within the gene were highly unmethylated. The first region of interest analyzed (-2235 - -1955) showed methylation at all CpG dinucleotides analyzed.

The first three and last CpG dinucleotides were highly methylated (-2208,-2196,-2154, -1978). Dual C/T peaks were apparent at -2112, -2045 and in ES, neural heart, and lung samples, however methylation was called. Differential methylation was found in regions 2 and 3. A more detailed look at the status of particular sites is provided in Figure 4b. The second region (-1324 - -1009) showed a majority of methylated sites in all of the tissues, ranging from 6/9 in liver tissue to 9/9 sites called methylated in ES and neural cells and heart tissue. Heart tissue has the only called methylated CpG at -1123. In all of the tissues that site presents as a heterogeneous peak, with the rest of the tissues being more unmethylated at that site. The most highly methylated sites in all tissues were -1227 and -1225 which form a CGCG site. The final CpG site analyzed in region 2 at position -1064 displays methylation heterogeneity in ES and neural cells, but is distinctly unmethylated in heart, liver, and lung tissues. The upstream region proximal to the Olig1 promoter (-647 - 399) contained three sites that differed in methylation status (-503, -484,-466), being methylated in ES and neural cells and unmethylated in heart, liver, and lung tissue. The first and last of these three are heterogeneous and more highly methylated (-503, -466), while the other (-484) shows a strong clear methylation signal. One read from ES cells showed one methylated site (1954) in region 7 in a very CpG rich area in the 3' UTR of Olig1. Heterogeneity was apparent in five CpG sites from (1948-1958). This may be due to true methylation at all of these loci or protection from bisulfite treatment at unmethylated CpGs by neighboring methyl groups (Warnecke, Stirzaker et al. 2002). A repeat of this read would provide more information as only 13 of the 24 CpGs could be analyzed. There is no simple correlation between methylation of the Olig1 promoter region and silencing of the

Olig1 gene. Methylation measurements of the Olig1 gene refute the hypothesis that methylation is associated with transcriptional silencing of Olig1.

3.5 Discussion

The precise spatial and temporal regulation of gene expression is essential for development of a functional organism. DNA methylation is also necessary for proper development and exhibits tissue specific differences. These two observations led to investigations which have shown a strong correlation between DNA methylation and transcriptional silencing (Boyes and Bird 1991), and the emergence of the hypothesis that DNA methylation was involved in the regulation of cell fate by functioning to silence genes during development once fate decisions have been made and gene expression is no longer needed.

We tested the hypothesis that DNA methylation is correlated with repressed transcription in genes involved in neurogenesis by mapping methylation in loci of a housekeeping gene and several developmental regulatory genes in expressing and non-expressing tissues, expecting all cells and tissues that did not express a particular gene to be methylated. We found two genes that fit this hypothesis, a housekeeping gene Pgk-1 (Hansen and Gartler 1990) and pluripotency regulator Oct3/4 [(Pesce and Scholer 2001; Gidekel and Bergman 2002; Hattori, Nishino et al. 2004; Li, Pu et al. 2007; Yeo, Jeong et al. 2007; Zhang, Siu et al. 2008)], in agreement with previous studies. On the other hand, methylation maps of neural lineage genes in non-neural tissues obtained in this study contradicted the hypothesis.

Sox1 is the earliest transcription factor expressed specifically in neural cells. Sox1 expressing cells are dividing neural progenitors that are competent to become neurons upon differentiation and Sox1 down-regulation (Pevny, Sockanathan et al. 1998; Kiefer 2007). All Sox1 non-expressing tissues, heart, liver, or lung, assayed were largely unmethylated in the upstream regions and throughout the gene contrary to the hypothesis. Methylation was found ~1.4kb upstream of the Sox1 transcription start site in ES cells and in neural cells. Two methylated CpG sites were also found in only neural cells (~3.5kb upstream); ES cells were not analyzed for this region. Methylation was only observed in ES cells in the 3'UTR of Sox1 which also was an unanticipated finding. To summarize, methylation is not correlated to silencing of Sox1 expression.

Olig1 and Olig2 are transcription factors expressed in the developing nervous system in oligodendrocyte precursor cells and in oligodendrocytes of the mature nervous system (Lu, Yuk et al. 2000; Ligon, Fancy et al. 2006). Of all of the cells assayed in this study, Olig1 and Olig2 expression is limited to a subset of ES derived neural cells, therefore if methylation represses transcription of Olig1 or Olig2 according to the hypothesis, we would expect some methylation of the Olig1 and Olig2 promoters in neural cells, and complete methylation in all other cell types. Olig2 was largely unmethylated in all cells and tissues while Olig1 methylation profile was more complex.

Methylation analysis of the Olig1 locus showed regions that were completely methylated in all cells and tissues (~2kb upstream), had some methylated sites in all cells and tissues (~1.2 kb upstream), were methylated in ES and neural cells and not in adult tissues (~500bp upstream), or were completely unmethylated in all cells and tissues (body of the gene). The two upstream regions that had methylation in all cells and tissues (~1.2kb

and ~2kb upstream) support the hypothesis, as cells and tissues that do not express Olig1 are methylated. While all sites in the ~2kb upstream region were methylated in all tissues, the degree of methylation differs between tissue types in the ~1.2kb upstream region, with some sites being methylated, heterogeneously methylated, and unmethylated in different tissues. The ES derived neural cells containing a subset of Olig1 expressing cells are also methylated in these two regions, refuting the hypothesis. It is possible that the direct bisulfite PCR method employed here is not sensitive enough to pick up the heterogeneity of the ES derived neural cells, therefore analysis of a pure Olig1 expressing population of cells or use of a more quantitative method might resolve Olig1 expressing cells as unmethylated in this region and support the hypothesis. Analysis of the region immediately upstream of the Olig1 transcription start site, ~500bp , showed that methylation in this region was not correlated to silencing of Olig1 as heart, liver, and lung tissue were unmethylated in this region. In summary, our analysis shows that methylation is not associated with Olig1 silencing.

Like Olig1, Olig2 expression is limited to a subset of ES derived neural cells. Therefore if methylation represses transcription of Olig2 according to the hypothesis, we would expect some methylation of the Olig2 promoter in neural cells, and complete methylation in ES cells, and heart, liver, and lung tissues. There was no DNA methylation in the Olig2 promoter region, 500bp upstream to 100bp downstream of transcription start site, in cells and tissues that express and do not express Olig2. Recent work by other groups confirms our results that the Olig2 promoter is not methylated in tissues that do not express Olig2 (Meissner, Mikkelsen et al. 2008). In addition, Olig2 promoter methylation was observed in acute myeloid leukemia blood cells while normal blood cells

were unmethylated (Kroeger, Jelinek et al. 2008), indicating that methylation of the Olig2 promoter may be a sign of disease. We conclude that Olig2 promoter methylation is not correlated to Olig2 silencing and does not appear to have any involvement in regulating Olig2 transcription in normal cells and tissues.

Surprisingly, the entire Olig2 locus, from 6kb upstream through the body of the gene, was hypomethylated in the neural cells and in all tissues. Only two CpG sites upstream of Olig2 were methylated in neural cells and liver tissue and unmethylated in all other tissues. Recent restriction based illumina mapping also showed low levels of methylation directly upstream (~5kb) of Olig2 in ES and neural progenitors (Meissner, Mikkelsen et al. 2008). In addition, high levels of methylation ~14kb upstream in ES and neural cells and differential methylation with ES cells being methylated and neural progenitors largely unmethylated were found ~90 kb upstream of Olig2, far outside the scope of this study. In summary, methylation of the Olig2 locus is not correlated with Olig2 silencing and neural cell fate specification.

There are several alternative hypotheses for the role of methylation in gene regulation including: silencing of retroviral insertion elements, marking proliferative status in a cell line, or activating transcription. Silencing of retroelements by methylation is the most widely accepted of these. Retroviral insertion elements are preferentially methylated in order to protect the organism ((Jahner, Stuhlmann et al. 1982; Walsh, Chaillet et al. 1998). ;Reviewed in (Hoelzer, Shackelton et al. 2008)). The analyzed region of the Olig1 locus contains several retroviral insertion elements, one LINE, three SINES, and one LTR, and a large (6.8kb), nearly complete retroviral LTR element between Olig2 and Olig1 on chromosome 16, 21kb upstream of Olig1. Two of the three regions with methylation

mapped in this study correspond to a repetitive element. Olig1 region 2 (~1.2kb upstream) partially maps to a SINE element (-1446 – -1244), and the first 2 CpGs, corresponding to the SINE element, were methylated or heterozygously methylated in all cells and tissues. Olig1 region 3 (~500bp upstream) encompasses a LINE element (-593- -497) where there was methylation in ES and neural cells, but not in any of the adult tissues. An Alu sine element (-438—351), known for methylation (Reviewed in (Hoelzer, Shackelton et al. 2008)), was nearby to mapped region 3, but the CpG site in it (-365) was not mapped. In the limited scope of this study, the data support the association of retroviral insertion elements with methylation, however methylation of the Olig1 locus was not limited to retroviral elements.

Genome-wide methylation has been observed to increase in propagated cell lines over freshly isolated primary cultures, possibly as a result of in vitro culture conditions (Antequera, Boyes et al. 1990; Meissner, Mikkelsen et al. 2008). Our analysis shows some evidence for this in the methylation patterns in the Sox1 and Olig1 loci. Methylation was observed ~1.4kb upstream of the Sox1 transcription start site in ES cells and in neural cells. Similarly, in the Olig1 promoter region ES and ES derived neural cells were methylated at the same 3 sites, while all tissues were unmethylated. The two methylated CpGs only found in neural cells ~3.5kb upstream directly contradict the correlation between methylation and repression. Upon further analysis, if ES cells are not methylated in these two sites and methylation is specific to neural progenitors, methylation may be playing an alternative role as a transcriptional activator (Chahrour, Jung et al. 2008). Our study provides limited evidence to support the methylation as proliferative mark or transcriptional activator hypotheses.

There are structural and functional clues as to why the Oct3/4, Sox1, and Olig1 genes are able to recruit methyltransferases and Olig2, especially the promoter, does not. Unlike Oct3/4, which has a high GC content (59%), but low CpG density (observed/expected=0.34), the Olig2 promoter is associated with a called CpG Island and the analyzed region has a GC content of 61% and a CpG observed/expected of 0.76 similar to Pgk-1 (62%, 0.71) which is also associated with a called CpG Island and unmethylated on the active X in all tissues. All methylated regions in neural genes Sox1 and Olig1 found in this study did not fit the CpG Island criteria. Other groups have also found the majority of CpG Island promoters (87.9%) remain unmethylated in all tissue types, whereas 50% of the non-CpG Island 5'UTRs were methylated (Eckhardt, Lewin et al. 2006). This study also provides support for the prevailing idea that areas of high CpG density, CpG islands, are unmethylated, which may have the function of providing protection from mutagenesis to these critical regions (Bird 1980). The Olig2 promoter region analyzed also contains 2 Sp1 sites. A transcription factor involved in early development, binding sequences (GGCGG) and lack of methylation at these sites allows for binding and transcription. Functional assays of the Olig2 promoter region have shown non-specific transcription activity from a region 1.1kb upstream of the transcriptional start site (Zhang, Horrell et al. 2008). Lack of methylation at the Olig2 promoter would allow for non-specific transcriptional activity.

The overall conclusion of this study is that there is no sensible hypothesis that unifies all of the data. For two genes, DNA methylation is associated with repression whereas for three genes it is not. Larger datasets, including genome-wide studies are also coming to the conclusion that the prevailing hypothesis is not correct and the function of

DNA methylation remains elusive (Meissner, Mikkelsen et al. 2008; Suzuki and Bird 2008). Determining what makes a particular locus amenable to methylation and measuring the effects of methylation on a particular gene depends on being able to selectively perturb methylation within its genomic context. Current technologies allow for genome-wide methylation disruption by methyltransferase knockout or chemical inhibitor, but tools to perform targeted methylation or methylation inhibition are still under development.

Chapter 4

Engineering ES Cells for Drug Selection of a Subset of Neural Cells

4.1 Abstract

ES cells genetically engineered with drug resistance expressed under the control of a neural cell type specific promoter offers the potential to provide access to large numbers of well-defined neural cells for research and therapeutic applications. Two ES cell lines were generated where a puromycin acetyltransferase (PAC) gene was inserted into the Olig2 gene. In both lines, PAC expression recapitulated the expression of the native Olig2 gene to permit purification of a subset of ES derived neural cells by selection with puromycin. Selected neural cells were arranged in rosettes and expressed Olig2. In addition, continued culture and selection of purified Olig2 expressing cells was possible. This system enables convenient and cost-effective purification of large numbers of Olig2 expressing cells for biochemical analysis, transplantation research, and toxicity testing. A similar selection strategy may be applied to other neural subsets.

David Lorberbaum and Julia Kuhn assisted with experiments performed in Chapter 4. They built the expression vector, performed targeting detection PCRs, and handled clones. The rest of the work was done autonomously.

4.2 Introduction

The two fundamental properties of ES cells, the ability to self-renew and potential to differentiate, have raised exciting new possibilities for their use in biomedical research and applications in regenerative medicine. Methods have been developed to differentiate mouse ES cells into various neural lineages including: general neural(Bain, Kitchens et al. 1995), ventral neural (Wichterle, Lieberam et al. 2002), retina(Zhao, Liu et al. 2006), cerebellum(Salero and Hatten 2007), and cerebral cortex (Gaspard, Bouschet et al. 2008). These methods are being extended to human ES cells (Thomson, Itskovitz-Eldor et al. 1998), enabling human cells to become the focus of study of human biology and to recently discovered iPS cells (Takahashi, Tanabe et al. 2007; Wernig, Meissner et al. 2007), making it possible to develop patient specific models of disease. Stem cell derived neural cells have shown promise as a tool to study the development of the nervous system (Abranches, Silva et al. 2009), to model neurodegenerative disease (Di Giorgio, Carrasco et al. 2007; Hochedlinger and Plath 2009), for development of cellular replacement therapies (Reviewed by Murray and Keller, 2008; (Nayak, Kim et al. 2006), and as a testing ground for pharmaceuticals on neural cells ({Broom, 2006 #181; Murry and Keller 2008) Reviewed by (Rubin 2008)).

The ability to generate large numbers of well-defined ES cell derived neural cells is critical for all of the applications described. ES derived neural cells recapitulate much of the complex biology and heterogeneity of the mammalian nervous system, both in terms of genetics and physiology (Bain, Kitchens et al. 1995; Wichterle, Lieberam et al. 2002). Given their origins as proliferating stem cells, generating large numbers of ES cell derived

neural cells is not a problem. However, limiting the differentiation of stem cells to the desired population of neural cells is a challenge. Two approaches to direct stem cell derived neural cells to specific lineages are the use of instructive growth factors and manipulating stem cell genetics by promoter-reporter constructs.

ES cells have the potential to differentiate into all cell types of the body and without signaling factors will generate a very heterogeneous population that includes a small percentage of neural cells (Martin 1981). Instructive factors have been identified to direct ES cells to the neural lineage. Retinoic acid (RA) enriches for differentiation of ES cells along a general neural pathway: generating cultures of neurons, astrocytes, and oligodendrocytes (Bain, Kitchens et al. 1995). The combination of RA and sonic hedgehog (Shh) further limits the phenotypes of ES derived neural cells to the ventral nervous system, responsible for motor output (Wichterle, Lieberam et al. 2002), Figure 4.1A). Yet, even ventral neural cultures contain many cell types. Further definition of ES derived ventral neural cells can be based on the expression of transcription factors or other cell type specific marker genes.

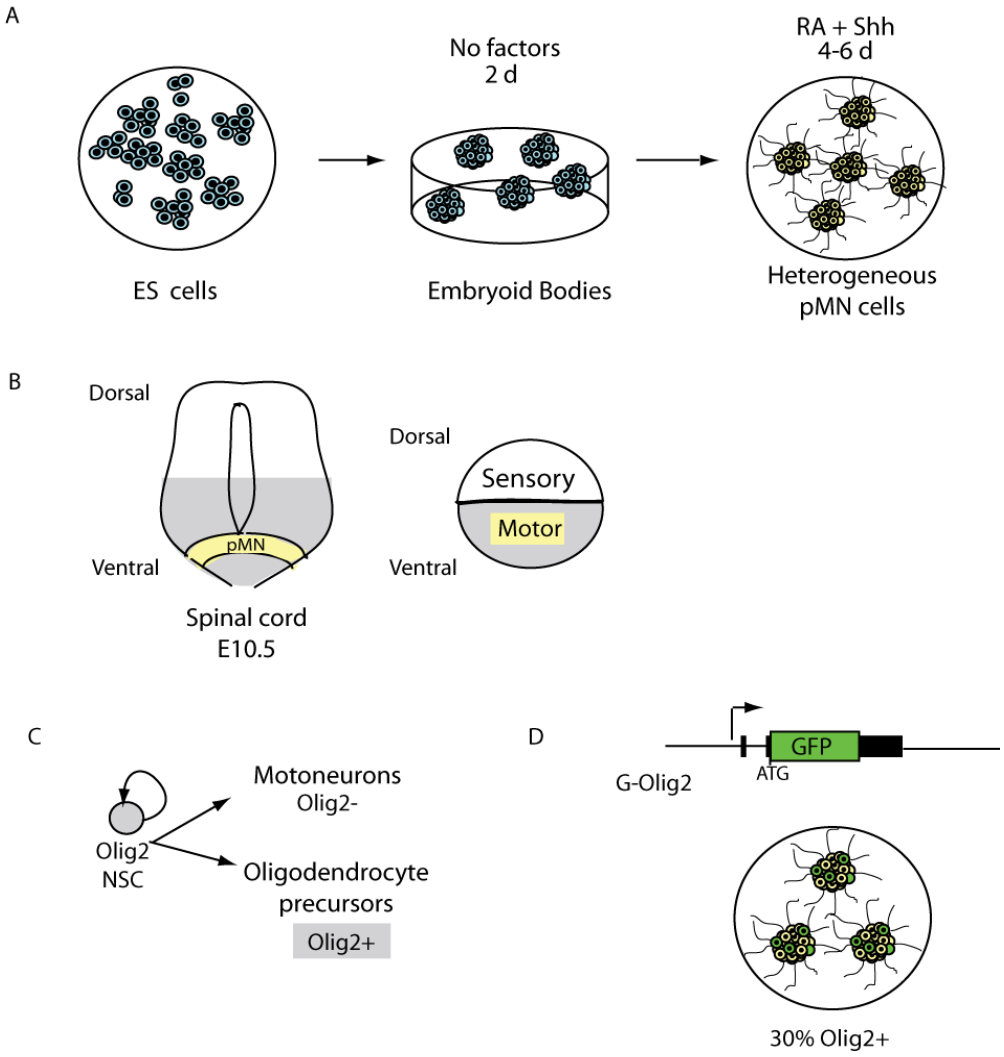


Figure 4.1: Taking Clues from Embryonic Development to Direct Stem Cells into Ventral Neural Cells. (A) ES cells can be differentiated to produce heterogeneous motor neuron progenitor (pMN) cultures by applying inducing factors retinoic acid and Shh (Wichterle 2002, Xian and Gottlieb 2005) (B) Diagram of the pMN domain, a region in the developing ventral spinal cord responsible for generating motoneurons and oligrodendrocyte precursor cells. (C) Olig2 is transcription factor expressed in the developing ventral spinal cord (Takebayashi, Yoshida et al. 2000). Olig2 expressing cells are multipotent and can differentate into motoneurons and oligodendrocyte precursors (based on (Rowitch, Lu et al. 2002)). (D) G-Olig2, an ES cell line with green fluorescent protein gene targeted to Olig2 in ES cells (Xian, McNichols et al. 2003; Xue, Wu et al. 2009), allows for easy visualization of Olig2 expression in pMN cells.

Olig2 is a basic helix-loop-helix transcription factor expressed in the ventral portion of the developing brain and spinal cord (Takebayashi, Yoshida et al. 2000; Ligon, Fancy et al. 2006), See Figure 4.1B). It is an important fate regulator in the central nervous system (Zhou and Anderson 2002). Olig2 expression defines a population of neural stem cells that differentiate into motoneurons and oligodendrocyte precursor cells (OPC). As depicted in Figure 4.1C, motoneurons shut off Olig2, but Olig2 expression persists in the OPC lineage (Lu, Sun et al. 2002; Wu, Wu et al. 2006).

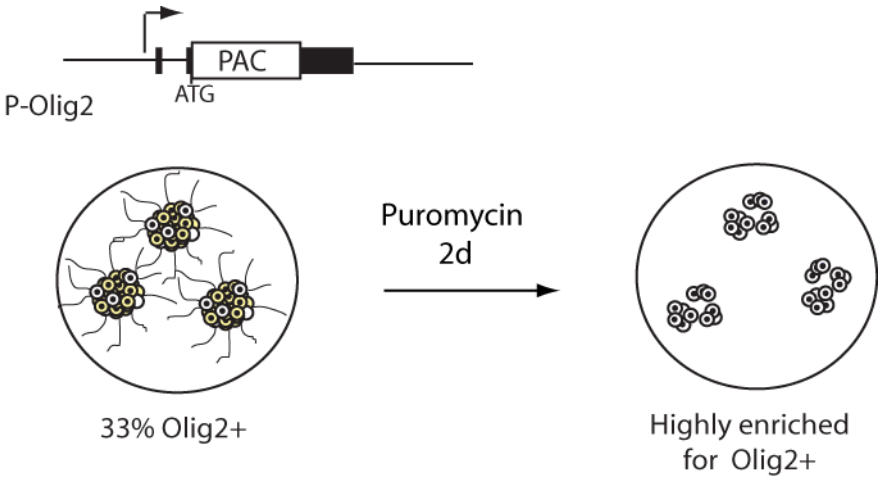
GFP knock-ins to the Olig2 gene have been particularly useful for observing and selecting for Olig2 expressing cells in heterogeneous ES derived neural cultures. In 2003, Xian et al (Xian, McNichols et al. 2003) generated a mouse Olig2 GFP knock-in ES cell line (G-Olig2). In G-Olig2 ES cells a promoterless green fluorescent protein gene (GFP) is inserted into the protein coding region of Olig2 so that the Olig2 promoter drives GFP expression. Cells expressing Olig2 in the G-Olig2 line express GFP, and appear green using fluorescence microscopy. With the knock-in line, Olig2 expression became easily observable in ES derived neural cultures. In addition, Olig2 expressing cells could be purified from heterogeneous neural cultures by fluorescence activated cell sorting (FACS, (Xian, McNichols et al. 2003; Xian and Gottlieb 2004; Xian, Werth et al. 2005)). By monitoring GFP expression, chemical signals that induce Olig2 expression could be easily identified by simple observation, as depicted in Figure 4.1D. Also, expression profiles of homogeneous Olig2 expressing cells can be obtained (Xian, Werth et al. 2005; Shin, Xue et al. 2007). Because of the utility of the mouse ES GFP knock-in, an Olig2-GFP knock-in was made recently in human ES cells (R-Olig2, (Xue, Wu et al. 2009)). One of the main purposes of this work was to purify Olig2 expressing cells, so the properties of these cells

could be explored in biochemical experiments. To that end, global gene expression analysis was performed on selected R-Olig2 at two time points during differentiation, revealing distinct expression profiles. In addition, R-Olig2 selected cells were transplanted into a rat spinal cord and differentiated into oligodendrocytes *in vivo*.

The Olig2 GFP knock-ins have provided a useful tool for observing Olig2 expression in the ES derived neural cells and a glimpse into the value of a genetically selected cell population for biochemical analysis and transplantation research. As the focus of research turns to relatively homogeneous cells, using GFP as a selector will limit the scope and scale of research.

ES cells can be targeted with a drug resistance cassette to select for cells expressing a targeted gene (Klug, Soonpaa et al. 1996; Li, Pevny et al. 1998; Billon, Jolicoeur et al. 2002; Ying, Stavridis et al. 2003; Zandstra, Bauwens et al. 2003). Drug selection works by utilizing a drug toxic to all cells and a gene encoding an enzyme that renders the drug non-toxic. An ES cell knock-in is made where the inserted gene encodes an enzyme that renders the drug inactive. In the knock-in line, cells not expressing the targeted gene do not express enzyme and are killed by the drug. In contrast, cells expressing the targeted gene make an enzyme which inactivates the drug, allowing only cells expressing the target gene to survive in the presence of drug.

A



B

	Drug selection	Cell Sorter
Where	Your lab	Facility
When	At your convenience	Schedule with facility
# selected cells	100 x 10^6	10^6
# of selections (wells)	Many parallel samples	Constrained
Project cost	Low	High
Continuous selection	Very simple	Multiple serial sorts

Figure 4.2: Advantages of Drug Selection. (A) Selection of Olig2 expressing neural cells is simple using the P-Olig2 cell lines. Puromycin is applied to P-Olig2 neural cultures, killing cells that are not expressing Olig2 and puromycin acetyltransferase (PAC). PAC expression protects Olig2 expressing cells, theoretically enriching cultures from 33% to 100% of cells expressing Olig2. (B) Using a drug based selection allows for the researcher to obtain up to 100X more selected cells with the convenience of doing the selection in their own lab on their schedule. Project costs for selection are much lower using drug selection. Experiments using continuous selection are now simple instead of requiring multiple serial sorts.

For studies on selected cells, drug selection provides many advantages over traditional GFP selection (Figure 4.2). FACS selection with GFP lines limits selection to a specific time point whereas drug selection can be applied in a continuous manner. Drug selection increases convenience with the benefit of decreasing cost. Selection can be performed in any laboratory at any time, thereby providing an easily accessible source of the selected cell population. In addition, parallel selection of multiple independent samples is challenging using FACS but facile using drug selection. In conclusion, using drug selection will enable research that is simply not feasible with GFP knock-ins.

Not only will drug selection work on smaller research scales, but it will be favored over FACS selection to obtain large numbers of selected cells for screening assays in research and industry (Zandstra, Bauwens et al. 2003). Drug selection dramatically increases the number of cells that can undergo selection, and decreases cost. Drug selection on 10^9 cells costs less than $100 and can be performed directly in culture vessels. Sorting the same 10^9 cells using FACS requires a dedicated FACS machine directly attached to a bioreactor operating around the clock. Sorting 10^9 cells with a FACS would occupy the machine for two to eleven days depending on flow-rate [25,000/sec – 4,000/sec], and require a substantial project investment.

Here we describe an approach using gene targeting to allow for drug selection of a subset of ES derived neural cells. The drug for selection was puromycin. Puromycin is an aminonucleoside antibiotic that is toxic to mammalian cells and will kill up to 99% of cells within two days (Watanabe, Kai et al. 1995). ES cell lines (P-Olig2-1 and P-Olig2-2) were generated where the Olig2 gene was targeted with the gene encoding the enzyme

puromycin acetyltransferase (PAC, (Gomez Lahoz, Lopez de Haro et al. 1991)), which modifies puromycin making it non-toxic. P-Olig2 cells expressing Olig2 will express PAC and survive puromycin selection. The P-Olig2 ES cell lines allow for convenient, efficient, cost-effective purification of Olig2 expressing cells. The principle of drug selection described here will be applicable to other neural genes.

4.3 Materials and Methods

4.3.1 Construction of P-Olig2 Knock-In Vectors

The Olig2 BAC 227 clone previously used by Xian et al. (Xian, McNichols et al. 2003) was verified by PCR of 10 short amplicons spanning the Olig2 locus (Appendix D.2). All were purchased from IDT technologies (Coralville, IA). All primers used for the entire construction process are listed in the tables in Appendix D.1.

Targeting vectors were constructed by a combination of recombineering and restriction enzyme based cloning as previously described (Wu, Ying et al. ; Xue, Wu et al. 2009). Because targeting vectors ranged in size from 12-20kb, to reduce the possibility of plasmid instability all bacterial strains were grown at 31ºC for the duration of the project. First, the Olig2 region was subcloned from its BAC into a kanamycin resistant plasmid (pStartK, (Wu, Ying et al. 2008)) using red recombination (Figure 4.3A). The rescue plasmid pStartK was amplified by PCR using primers tailed with 50bp of homology to the Olig2 region. The homology region was carefully screened to eliminate regions of repetitive sequence. Linear product was electroporated into recombinogenic Olig2 BAC 227 cells and recombinants were selected with kanamyacin. The resulting clones were screened by

looking for single band following mini-prep, and verified by restriction digests (Appendix D.2) and sequencing into the homology region. We generated 8 non-identical plasmids named pOlig2_(1-8) spanning the Olig2 locus, showing the robustness of the method .

Second, the open reading frame of Olig2 +1 to +973 was surrounded by AscI restriction sites using red recombination (Figure 4.3B). The homology containing plasmids, including pOlig2_1 and pOlig2_8, were co-electroporated with the PCR amplified chloramphenicol cassette flanked by AscI sites and 50bp of homology immediately outside of the Olig2 open reading frame into recombinogenic DH5α cells. Recombinants were identified by chloramphenicol resistance. Retransformation was used to isolate concatemers or to remove unrecombined plasmids that appeared during screening. Restriction sites were added to 7/8 homology plasmids as shown by digest and partial sequencing. These plasmids were named pOlig2_(1-8)Asc.

The third step was to bring the puromycin resistance cassette into AscI sites by cloning (Figure 4.3C). The resistance cassette was prepared by a combination of PCR and cloning. The resistance cassette consists of the following elements from 5' to 3': AscI site, Kozak sequence (Kozak 1986; Thomas and Capecchi 1987), puromycin cassette with bgh polyA signal (Stratagene, PKO-Select Puro), LoxP, phosphoglycerate kinase I promoter driving the neomycin phosphotransferase gene (PGK-neo) with bgh polyA signal, Lox P (NCI Biological Resources Branch pL452), and AscI site. This resistance cassette was cloned into an ES cell expression vector (Appendix D.3) to confirm functional puromycin and neomyocin genes. It was then cloned from the expression vector into AscI sites in the targeting vectors pOlig2_1Asc and pOlig2_8Asc. These two vectors were chosen because one had a longer region of homology upstream of the Olig2 gene and the other one had a

longer region of homology downstream of the Olig2 gene. Clones were screened by chloramphenicol sensitivity and orientation of the insert was verified by PCR of both junctions (Appendix D.2). Resulting plasmids were named pOlig2_1PN and pOlig2_8PN.

The final step adds a herpes simplex virus thymidine kinase gene for negative selection and moves the targeting constructs into a high copy number origin (Figure 4.3D). The targeting vectors pOlig2_1PN and pOlig2_8PN were incubated with a multisite gateway plasmid (pWS-TK3) containing attR1 and attR2 site and TK gene. The P-Olig2 fragment which contained attL1 and attR1 sites recombined with the gateway plasmid in the presence of clonase enzyme and exchanged by LR recombination. These final constructs were selected using ampicillin and verified by restriction digest (Appendix D.2). The final targeting vectors were named pOlig2_1PN-TK3 and pOlig2_8PN-TK3

4.3.2 Generating P-Olig2 ES Cell Lines from RW4

Normal Culture of ES Cells

All ES cell experiments were done on the RW4 mouse embryonic stem cell line derived from Sv129 mouse. ES cells were maintained on gelatin-coated tissue culture plastic in the presence of leukemia inhibitory factor (LIF) according to standard methods (Bain, Kitchens et al. 1995; Rieger, Poppino et al. 2007; Zhang, Horrell et al. 2008).

Electroporation

Targeting in ES cells was performed with both pOlig2_1PN-TK3 and pOlig2_8PN-TK3 constructs. To generate the P-Olig2 knock-in lines, pOlig2-1 and

pOlig2-2, $1x10^7$ RW4 ES cells were dissociated using 0.25% Trypsin-EDTA (Sigma, St. Louis, MO) and resuspended in electroporation buffer (20mM HEPEPS, pH 7.5, 137mM NaCl, 5mM KCl, 0.7mM Na_2HPO_4, 6mM dextrose), with 10-15μg of ScaI-linearized targeting vector. Electroporation was conducted at 0.23kV and 960μF in a 0.4cm cuvette (Bio-Rad, Hercules, CA). Electroporated cells were plated on gelatin and dosed with G418 (200μg/mL, Invitrogen) and 1-(2-Deoxy-2-fluoro-ß-D-arabinofuranosyl)-5-iodouracil (FIAU 100nM, Movarek Biochemicals, Brea, CA) 24 hours after electroporation. Resistant clones were picked 8 days following electroporation, trypsinized, and half of each clone was expanded on STO monolayers in 96 well plates. The other half was frozen in a 96 well PCR plate at -70 ºC and retrieved following identification of targeted clones.

Targeting Detection by PCR

For each ES clone, genomic DNA (2uL) was assayed by PCR for a novel junction generated by targeting of the short arm for each construct. Genomic DNA was prepared from confluent 96 well plates (Sambrook and Russel 2001). PCR was conducted using KTLA polymerase [1X KLA buffer 50mM Tris Base, 16mM Ammonium Sulfate, 0.1% Tween-20, 3.5mM MgCl2], 100 μM Betaine, 200 μM each dNTP, 200 nM each primer, 4μg/mL BSA, and 0.01U KTLA per 20μL reaction. Thermocyling parameters were: 94 ºC degrees for 10 minutes, followed by 35 cycles of 94 ºC for 30 seconds, 58 ºC for 30 seconds, and 72 ºC for 4 minutes. Primers for PCR detection are given in Appendix D.1.

Southern Hybridization

Genomic DNA was prepared from $12x10^6$ ES cells (RW4, B9, D9) using the ArchivePure DNA Cell/Tissue and Tissue Kit (5 Prime, Gaithersburg, MD**)** according to the manufacturer's instructions**.** Proteinase K (300μg) was added to 3mL of Cell Lysis Solution to digest proteins bound to DNA and remove nucleases. For targeting confirmation genomic DNA (1-10μg) was digested with either HindIII (200U) or SpeI (50U) overnight at 37 ºC. Additional enzyme was added and digestion was continued for one hour. DNA was ethanol precipitated and electrophoresed on a 0.8% agarose gel in 0.5X Tris-borate-EDTA (45 mM Tris-borate and 1 mM EDTA) buffer at 1V/cm for 19hrs. DNA was transferred to Hybond-XL (GE Healthcare Biosciences, Piscataway, NJ) membrane under neutral conditions for 20hrs. DNA was crosslinked to the membrane using UV Stratalinker 2400 on auto-crosslink setting. Probes were prepared using [^{32}P] dCTP (Easy Tide, Perkin-Elmer, Waltham, Massachusetts) and Rediprime kit (GE Healthcare Biosciences), and purified with illustra ProbeQuantG-50 columns (GE Healthcare Biosciences). Blots were hybridized in Rapid hybe for two hours at 65 ºC (GE Healthcare Biosciences), washed, and autoradiographed by standard methods.

Cre Excision and Subcloning

To excise the floxed PGK-neo cassette contained in the targeted lines, $2x10^6$ ES cells from each line were trypsinized and electroporated using an Amaxa nucleofector II (Amaxa Biosystems) with 5μg of Cre recombinase expressing plasmid (p1411, gift of Tim Ley, Washington University). Transfected cells were diluted and plated on STO feeder

monolayers. Single clones were picked and genomic DNA was assayed by PCR (Primers in Table 6) to demonstrate the removal of the PGK-neo cassette. Cre-d lines (B9-1, D9-1) were used for the duration of the study. Following expansion of cre-excised ES cell lines, lines were further subcloned by dilution on STO monolayers. Subclones were expanded and re-validated by PCR and neomyocin sensitivity to ensure complete removal of PGK-neo cassette.

Neural Differentiation of ES Cell Lines

Ventral neural differentiation of ES cells was performed by an established method using retinoic acid (RA) and sonic hedgehog (Shh) (Wichterle, Lieberam et al. 2002; Xian, Werth et al. 2005; Zhang, Horrell et al. 2008). The Shh agonist HhAg1.4 from Curis, Inc. (Cambridge, MA) was used throughout. Undifferentiated ES cells were scraped from flasks and cultured for 2 days as embryoid bodies (EBs) in DFK5 medium without inducing factors. EBs were then transferred to adhesive gelatin wells in DFK5, and RA (2μM) and Shh agonist (30 nM) were added to induce ventral neural phenotype. Culture was continued as indicated in the text.

mRNA Analysis

RNA was isolated from $5x10^6$ ES cells, one 6 well of EBs, or one 6 well of neural cultures in 1mL of TRI Reagent (Life Technologies, Carlsbad, CA) according to manufacuter's instructions. RNA was resuspended in 10mM Tris pH 8.0. and electrophesed on a 1.5% formaldehyde denaturing agarose gel alongside RiboRuler High Range RNA ladder (Fermentas, Glen Burnie, MD) for quantification. Total RNA was

quantified assuming 18S and 28S ribosomal RNA bands comprised 80% of total RNA. Reverse transcription reactions were conducted with 500ng of total RNA using the Retroscript kit according to manufacturer's instructions (Life Technologies) using random decamers as primers. Two step reverse transcription was conducted by first heating RNA and primers to 80 ºC for 3 minutes then adding the remainder of components. The reaction was incubated at 44 ºC for 1 hour, and enzyme was heat inactivated at 92 ºC for 10 minutes. Parallel reactions were prepared omitting reverse transcriptase enzyme. PCR reactions were performed with 1/20 of total cDNA. PCR was performed for the native and transgene Olig2 amplicon using KTLA polymerase for 35 cycles with an annealing temperature of 59 ºC and extension time of 45 seconds. Primers are given in Appendix D.1.

Puromycin Treatment of Cell Cultures

Puromycin dihydrochloride (Sigma Aldrich, St. Louis, MO) was dissolved in dH_20 at a concentration of 10mg/mL. Cells were dosed with 4μg/ml puromycin delivered in the appropriate culture media at stages described in the text.

Immunocytochemistry

Nuclear staining was performed by incubating cultures with 4′,6-diamidino-2-phenylindole dihydrochloride (Molecular Probes Inc., Eugene, OR) for 1 hour at 37°C before fixation. Then the wells were rinsed with cytoskeletal buffer (CB) (1.95 mg/ml 2-N-morpholino ethanesulfonic acid, 8.76 mg/ml NaCl, 5 mM EGTA, 5 mM $MgCl_2$, and 0.9 mg/ml glucose; pH 6.1) twice, fixed in 3% paraformaldehyde in CB for 15 min, treated

with 0.2% Triton X-100 (Sigma Aldrich) for 12 minutes, and then incubated for 1 hour in 3% bovine serum album in Tris-buffered saline (TBS) at pH 7.4. Primary antibodies to Olig2 (Millipore 1:500, Abcam 1:200) were applied overnight (4 °C). After washing in TBS for 15 minutes three times, Texas red-conjugated (Life Technologies) goat anti-rabbit IgG secondary antibody was applied (diluted 1:200). Experiments were performed in triplicate. Specificity of staining was confirmed by similar staining by two different Olig2 antibodies and by lack of staining on ES cell cultures, and with nonimmune serum controls performed with each experiment.

Live Dead Assay

Living cells were visualized by staining with calcein and dead cells were visualized by staining with EthD-1 (Life Technologies). Wells were washed by removing serum containing medium with detached cells to a tube and adding with Dulbecco's PBS (D-PBS). 3X Live/Dead solution was added to resuspend detached cells and returned to stain the attached cultures. Cultures were incubated in 1X Live/Dead reagent (1.3μM calcein, 2.7 μM EthD-1) at room temperature for 30 minutes and visualized.

Microscopy

Cells were viewed with a Nikon TE2000S fluorescence microscope (Nikon Instruments Inc., Melville, NY). Images were acquired on the MetaVue image analysis software (Molecular Devices, Sunnyvale, CA) and analyzed using Adobe Photoshop. A smart sharpening filter with a radius of 10.4 pixels was applied to remove lens blur from fluorescence images.

4.4 Results

The goal of this project was to generate an ES cell line that would enable us to purify Olig2 expressing cells from ES derived neural cultures. Gene targeting was used to engineer two ES cell lines with a drug resistance cassette expressed only in Olig2 expressing neural cells. The ES cell lines, P-Olig2-1 and POlig2-2, were generated by replacing the protein coding region of Olig2 with a gene encoding puromycin acetyltransferase (PAC). P-Olig2 ES cells are differentiated into neural cells where approximately 33% (Xian, Werth et al. 2005; Zhang, Horrell et al. 2008) will express Olig2. Olig2 expressing cells are then easily purified from heterogeneous P-Olig2 neural cultures by selection with puromycin.

First, targeting vectors were constructed to contain 10-20kb of the Olig2 locus, where the coding region was replaced by a gene conferring resistance to puromycin, PAC (Gomez Lahoz, Lopez de Haro et al. 1991). Next, ES cells were electroporated with a targeting vector to generate targeted events by recombination. Targeted ES cell clones were identified by PCR and confirmed by Southern analysis for two independent vectors. In targeted P-Olig2 ES cells, puromycin resistance is encoded in one Olig2 allele and therefore is expected to be expressed in cells that are expressing Olig2. Functional studies of two independently generated P-Olig2 ES cell lines show that puromycin resistance is conferred to a subset of neural cells that express Olig2. We conclude that the P-Olig2 knock-in lines work as designed and will facilitate selection of Olig2 expressing cells from heterogeneous neural cultures with puromycin.

4.4.1 Pharmacology of Puromycin Selection on Neural Cells

Before building the targeting vectors, puromycin was first investigated as a selector drug in neural cultures. Puromycin N-acetyltransferase (PAC) confers effective resistance to puromycin in ES cells and is often used as a selector for genetic engineering ES cells (Watanabe, Kai et al. 1995). However, the use of puromycin selection on neural cells has not been characterized. It is conceivable that neural cells cannot be protected from puromycin by PAC. For example, neural cells have extensive cellular processes and these might not receive PAC even if enzyme is made in the cell body. Unable to receive enzyme at its processes, a neural cell with the resistance gene could be killed by puromycin. To see if PAC protects neural cells we transformed RW4 ES cells with a linearized ubiquitously expressed PAC cassette. Random integrant ES clones were selected by puromycin and pooled together to generate a stable ES polyclonal culture where a PAC cassette was randomly integrated at multiple sites in the ES cell genome. This was then differentiated in parallel with nontransgenic RW4 ES cells and selected with puromycin drug for 24 hours. RW4 ES derived neural cells were sensitive to puromycin, exhibiting extensive cell death whereas some ES derived neural cells containing the PAC cassette remained alive. This experiment showed that the puromycin selection works in neural cells and so it was worthwhile to create the P-Olig2 knock-in.

4.4.2 Design and Construction of Targeting Vectors

As a first step in construction of targeting vectors, a 227kb Sv129 BAC clone containing the Olig2 region was obtained (Xian, McNichols et al. 2003). A Sv129 BAC was chosen because the RW4 ES cells used to generate the line are also derived from the Sv129

mouse strain. There is evidence that exact sequence homology, as would be derived from an isogenic strain, improves targeting efficiency (te Riele, Maandag et al. 1992). The frequency of targeting in ES cells is also proportional to the length of homology to the target sequence (Thomas and Capecchi 1987). Although BACs have been used for targeting (Valenzuela, Murphy et al. 2003), BAC DNA was not used directly because handling BAC DNA and detecting a BAC targeting event in ES cells are challenging. Instead, an approach described by Wu et al (Wu, Ying et al. 2008) was followed, which uses recombineering to subclone BAC DNA. A linear vector is tailed with 50bp of homology to the ends of the region of interest in the BAC by PCR, and recombination is catalyzed by lambda red recombinase (Figure 4.3A). This method for constructing targeting vectors allows for generation of long regions of homology, minimizes mutations from PCR amplification of homology arms, and reduces time and expense of vector construction compared to traditional cloning. In addition, keeping one homology arm short, between 1kb and 2kb, makes PCR detection of targeting events feasible. By using different primers for amplification of the rescue plasmid, eight independent plasmids between 10kb and 20kb containing the Olig2 region were generated [pOlig2_1-8]; two with short 5' arms and six with short 3' arms. The structures of these plasmids were verified by restriction digest (8/8) and sequencing (7/8) from the vector backbone into the homology region.

Next, recombineering was used to introduce restriction sites with base-pair precision at the boundaries of the open reading frame (Figure 4.3B). The restriction sites were brought into the targeting plasmids by recombination of a chloramphenicol cassette tailed with AscI sites and 50bp of homology just outside the Olig2 open reading frame, with the homology containing plasmid. Recombinants were selected by chloramphenicol

resistance and verified by restriction digest (7/8) and sequencing (5/5). Seven out of eight plasmids [pOlig2-1-3Asc, pOlig2_5-8Asc] were successfully modified with restriction sites.

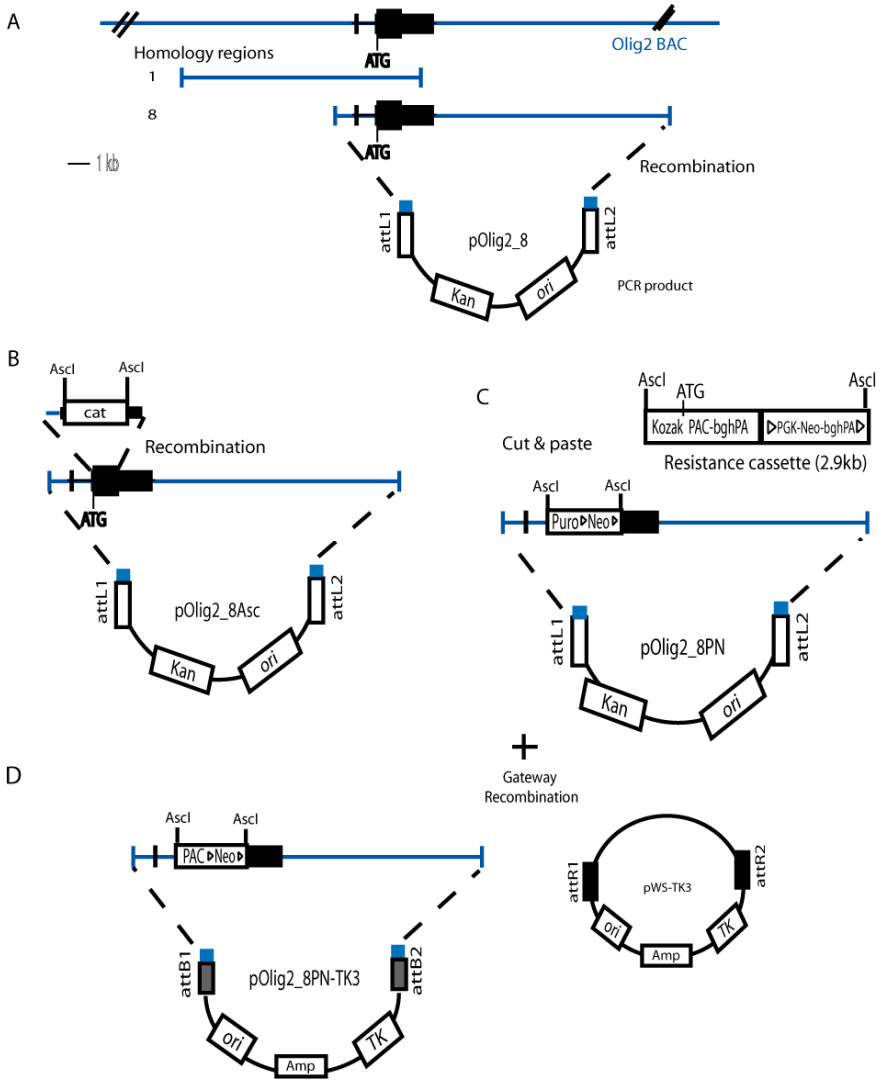


Figure 4.3: Engineering P-Olig2 Targeting Vectors by Recombineering and Cloning (based on (Wu, Ying et al. 2008)). (A) Two independent Olig2 regions of (12-15kb) were subcloned into plasmid vectors by recombination using 50bp of homology from a 227kb BAC (Xian, McNichols et al. 2003) containing the Olig2 region. (B) AscI restriction sites were introduced at the boundaries of the Olig2 open reading frame by recombination of a linear chloramphenicol cassette flanked with AscI and 50bp of Olig2 homology. (C) The AscI flanked resistance cassette, containing the promoterless puromycin cassette and floxed PGK promoter neomyocin cassette, was then transferred the targeting vector by conventional cloning. (D) Gateway clonase enzyme was used to catalyze site specific recombination between attB and attR sites and introduce a high-copy replication origin and thymidine-kinase negative selection cassette to generate final targeting vectors.

The next step was cloning the expression cassette into the targeting vectors. Two vectors were chosen, pOlig2-1Asc, with a short 3' homology region, and pOlig2-8Asc, with a short 5' homology region (Figure 4.3). Constructing more than one targeting vector was desired because it is unknown how specific sequences contribute to targeting efficiency in ES cells. Having two targeting vectors with unique ends should increase the likelihood of obtaining a targeting event. The second reason these two vectors, with short ends on opposite sides of the expression cassette, were chosen was to enable pilot testing of the targeting detection PCR described in the following section. The expression cassette was then cloned into both pOlig2-1Asc and pOlig2-8Asc by conventional cloning (Figure 4.3C). The expression cassette consists of a promoterless puromycin cassette which will be expressed under the control of the Olig2 promoter and ubiquitously active phosphoglycerate kinase (PGK) promoter driving a neomycin resistance cassette to select for insertion in ES cells. The cassette is flanked with AscI sites and consists of the necessary components for effective expression of puromycin resistance under the control of Olig2 promoter including: Kozak translational initiation sequence (Kozak 1986), start codon, and promoterless puromycin acetyltransferase and a bGH polyA+ signal. The neomycin resistance cassette is flanked by loxP sites and contains: the PGK promoter to control expression of the neomycin cassette with a bGH polyA+ signal. The PGK promoter is facing in the same direction as the Olig2 promoter. The flanking loxP sites (Sauer and Henderson 1988) allow for removal of the neomycin cassette by cre-recombinase once the targeted line has been obtained.

The expression cassette was generated by a combination of PCR and cloning. Due to the possibility of mutagenesis during PCR of the puromycin, the amplified puromycin

cassette with left most AscI site was cloned into a custom built expression vector (Appendix D.3) before being inserted in the targeting vectors. In the expression vector an EF-1alpha promoter (Chung, Andersson et al. 2002), a strong promoter in ES cells, was used to drive expression of the amplified puromycin acetyltransferase gene. Puromycin resistant ES cells were generated with the expression vector confirming the functionality of the amplified puromycin cassette. The expression vector also contained the PGK-neo cassette and right-most AscI site, so that the entire expression cassette was flanked by AscI sites. The expression cassette was cloned conventionally into the targeting vectors using AscI sites. A PCR assay for both junctions was used to detect correct insertion of the expression cassette in the targeting vectors. The two plasmids were named pOlig2_1PN and pOlig2_8PN.

All of the steps described above had been performed on low copy number vectors to reduce possibility of plasmid instability (Wu, Ying et al. 2008). As a last step, the targeting vectors were transferred to a high-copy number plasmid for production of DNA for targeting ES cells and to add a negative selection cassette containing herpes simplex virus thymidine kinase (TK) gene for enrichment of targeting events by selection with 1-(2-Deoxy-2-fluoro-ß-D-arabinofuranosyl)-5-iodouracil (FIAU, Movarek Biochemical, Brea, California (Capecchi 1989; Wu, Ying et al. 2008). Random insertion events in the ES cell genome will contain the TK gene and exhibit sensitivity to FIAU, whereas homologous recombination events will exclude the cassette from the ES cell genome. This backbone transfer was performed by a site specific recombination reaction between L1/R1 and L2/R2 catalyzed by clonase enzyme (Invitrogen, Carlsbad CA) (Figure 3D). Final vectors pOlig2_1PN-TK3 and pOlig2_8PN-TK3 were validated by restriction digest with three

different enzymes. The final vectors pOlig2_1PN-TK3 and pOlig2_8PN-TK3 have been though rigorous structural analysis throughout the construction process and have been validated by multiple methods including: restriction enzyme digests, junction PCR assays, appropriate antibiotic sensitivity or resistance, and sequencing.

4.4.3 Generation of the P-Olig2 Knock-In ES Cell Lines

Targeting in ES cells is typically detected by either short arm junction PCR or Southern analysis. Short arm junction PCR has the advantage of being an amplifying method and being easily scaled to analyzing many clones. Detection of a targeting event relies on amplification across a novel junction between the expression cassette and the neighboring genome. The novel junction needs to be detected in a genomic DNA, so the PCR assay needs to be very robust. Since the junction is novel in ES cells, typically there is no positive control for the PCR assay. By having two complementary targeting vectors with opposite short arms, the structure of the targeting event in ES cells is present the in long arm of the complementary targeting vector, providing a positive control template for optimization of the PCR assay.

To generate the targeted P-Olig2 ES cell lines, RW4 ES cells were transfected by electroporation of two independently generated targeting vectors, pOlig2_1PN-TK3 and pOlig2_8PN-TK3, in separate experiments. Successful recombination (Figures 4.4.1A, 4.4.2B) resulted in replacement of the Olig2 open reading frame (+1 to +973) with a promoterless puromycin cassette followed by a PGK-neo cassette for both vectors. Four independent electroporations were performed for each construct. Following positive selection with neomycin and negative selection with FIAU, clones were screened by junction PCR across the short arm (Figure 4.4.1A, 4.4.2A) from inside the targeting

construct into neighboring genomic DNA. Targeting efficiency was 5.2% (15/288) for pOlig2 _1PN-TK3 and 3.7% (14/376) for pOlig2 _8PN-TK3.

4.4.4 Structural and Functional Validation

Southern analysis was used to independently verify correct integration of the puromycin cassette in one clone derived from each targeting event. By using a probe for the Southern analysis external to the targeting vector in the Olig2 genomic region the unaltered allele could be compared to the engineered allele by examining shifts in restriction fragment size due to correct integration of the cassette at the Olig2 locus. Southern analysis on a P-Olig2 targeted clone from each construct and RW4 unengineered line (Figure 4.4.1C, 4.4.2C) produced the predicted structures. Targeting of pOlig2_1PN-TK3 to the Olig2 locus introduces two new SpeI sites reducing the detected band size from 7.8kb to 3.5kb for the targeted allele in clone D9 (Figure 4.4.1C). Clone B9 was produced by targeting pOlig2_8PN-TK3 to the Olig2 locus lengthening the HindIII fragment from 4.8kb to 6.7kb in the targeted allele (Figure 4.4.2C). In summary, southern analysis confirms correct chromosomal integration of the puromycin cassette into one allele of Olig2 in the targeted lines.

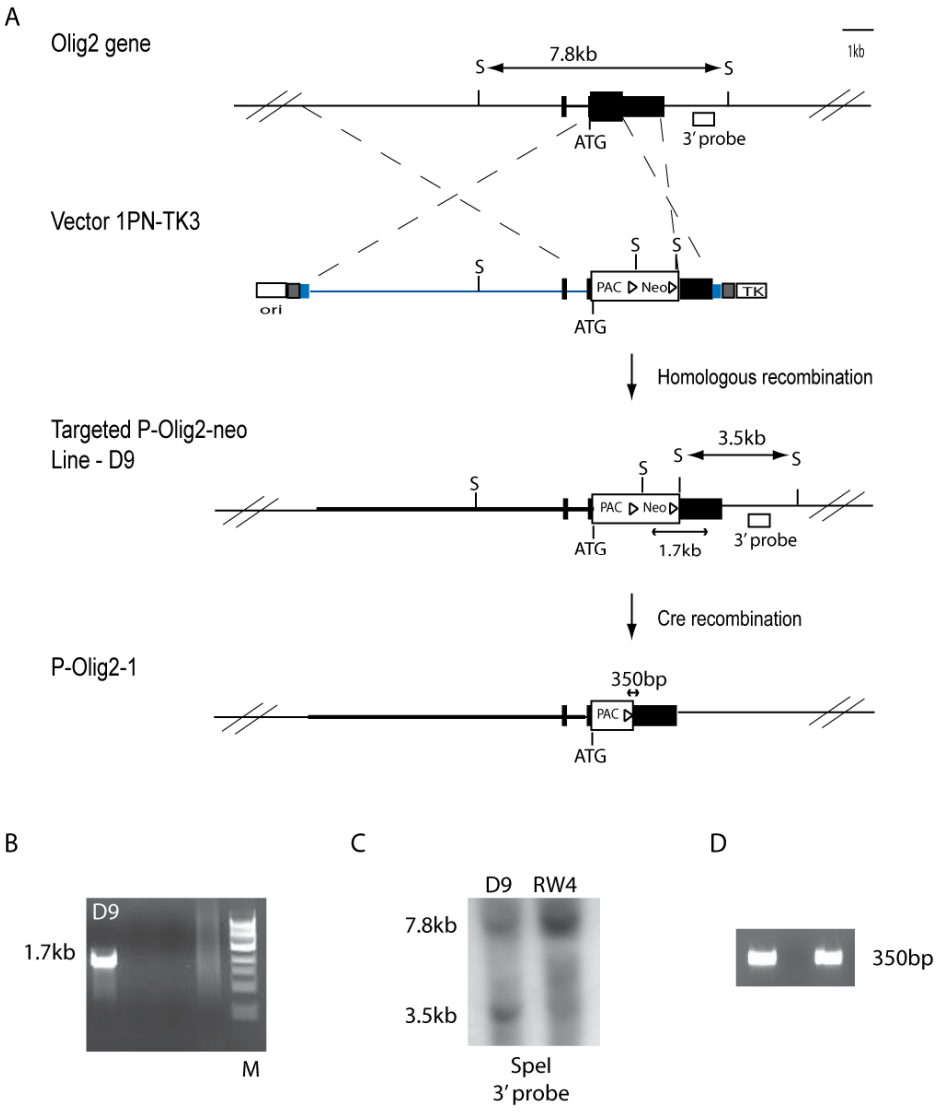


Figure 4.4.1: Successful Targeting of Puromycin Cassette to Olig2 Locus by Two Independent Targeting Events. (4.4.1) Targeting of pOlig2_1PN-TK3 to the Olig2 Locus in ES Cells. (A) The blue line represents the homology arms and black boxes are the two exons of Olig2. The targeting vector contained a promoterless puromycin cassette followed by a floxed neo cassette which was used for positive selection. A TK cassette was used for negative selection in ES cells. (B) Targeting events were identified by junction PCR. Junction PCR identified clone D9, a targeted clone by appearance of 1.7kb band. (C) Genomic DNA from D9 was digested with SpeI and Southern blot confirmed targeting of the cassette to the Olig2 locus by a new 3.5kb band detected by the 3' probe. (D) Cre-excision of the PGK-neo cassette was confirmed by the presence of a 350bp band and sensitivity to neomycin (not shown).

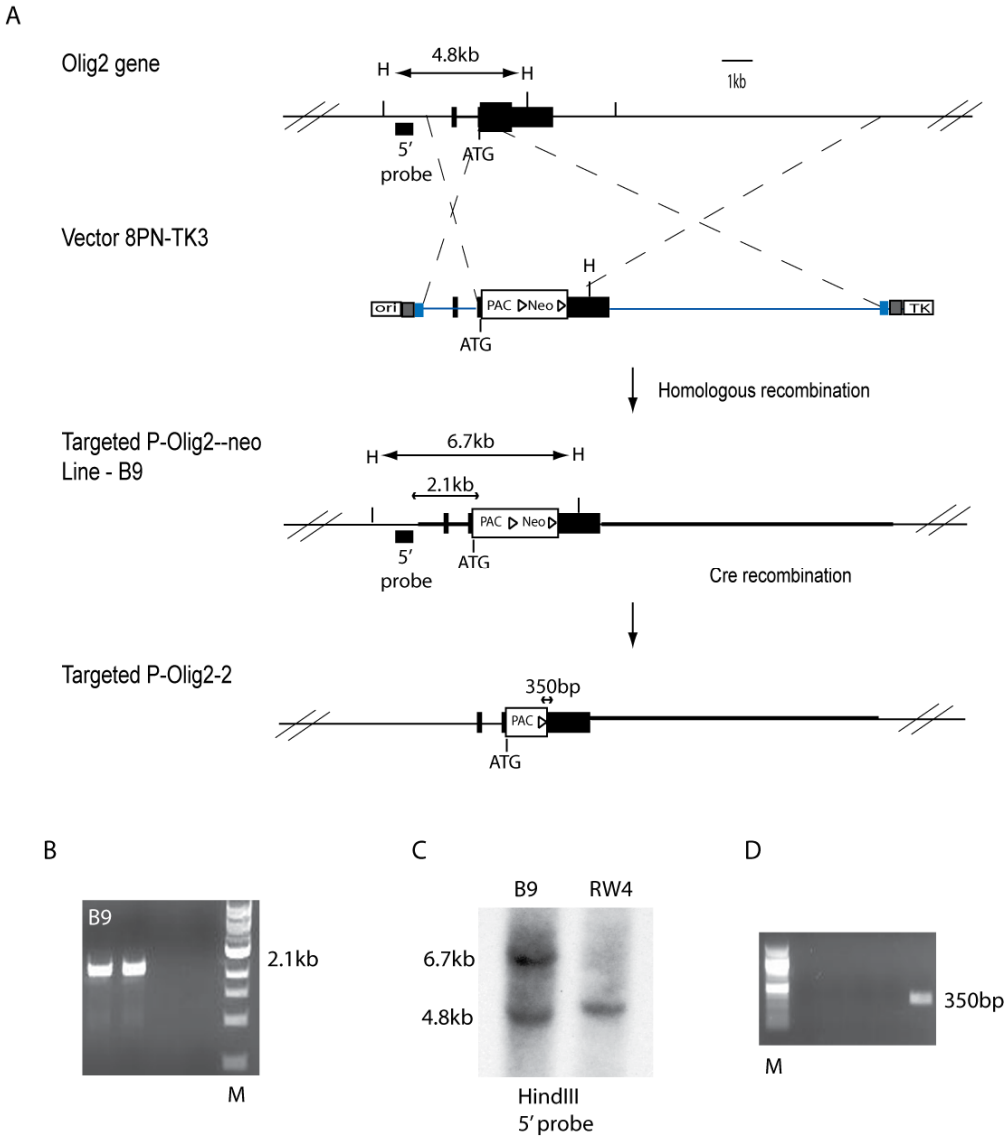


Figure 4.4.2: Successful Targeting of Puromycin Cassette to Olig2 Locus by Two Independent Targeting Events. (4.4.2) Targeting of pOlig2_8PN-TK3 to the Olig2 locus in ES Cells. (A) The blue line represents the homology arms and black boxes are the two exons of Olig2. The targeting vector contained a promoterless puromycin cassette followed by a floxed neo cassette which was used for positive selection. A TK cassette was used for negative selection in ES cells. (B) Targeting events were identified by junction PCR. Junction PCR identified clone B9, a targeted clone by appearance of 2.1kb band. (C) Genomic DNA from B9 was digested with HindIII and Southern blot confirmed targeting of the cassette to the Olig2 locus by a new 6.7kb band detected by the 5' probe. (D) Cre-excision of the PGK-neo cassette was confirmed by PCR by the presence of a 350bp band and sensitivity to neomycin (not shown).

Cre-excision of the PGK-Neomyocin Cassette is Required for Correct mRNA Regulation

Once the structure of the targeted P-Olig2neo lines, B9 and D9, had been validated, we next investigated the expression profile of the puromycin resistance transgene in the targeted lines. Since the lines were puromycin acetyltransferase knock-ins into Olig2, we expected expression of the P-Olig2 PAC transgene to follow the same temporal and cellular specificity of the native Olig2 gene. To test this we isolated RNA from ES, EB and 2-/4+ ventral neural cells from targeted lines and RW4 as an unengineered control. Olig2 and the P-Olig2 PAC mRNAs were measured by RT-PCR. Based on knock-in design principles, we expected to find that the P-Olig2 PAC mRNA would be expressed only in neural cells with the same regulation as the native Olig2 allele. As expected, Olig2 mRNA was detected in only 2-/4+ ventral neural cells in the targeted lines B9 & D9, and RW4. Unexpectedly, we found that the P-Olig2 PAC mRNA was expressed at all stages (the ES, EB and 2-/4+stages) and is not restricted to neural cells like the native mRNA (Figure 4.5).

To explore this unexpected finding further, we treated P-Olig2-neo ES cells with 4μg/ml puromycin and found that they were sensitive to the drug. This result directly contradicts the presence of the PAC mRNA in ES cells. It suggests that the PAC mRNA we measured was not being translated into a functional PAC enzyme at the ES cell stage. Furthermore, puromycin treatment of P-Olig2-neo ventral neural cells resulted in a subset of cells that was resistant to the drug. This contradictory result suggests that in neural cells PAC mRNA is translated into functional PAC enzyme. Together, these results led us to suspect that the misregulation of the transgene was only occurring at the mRNA level.

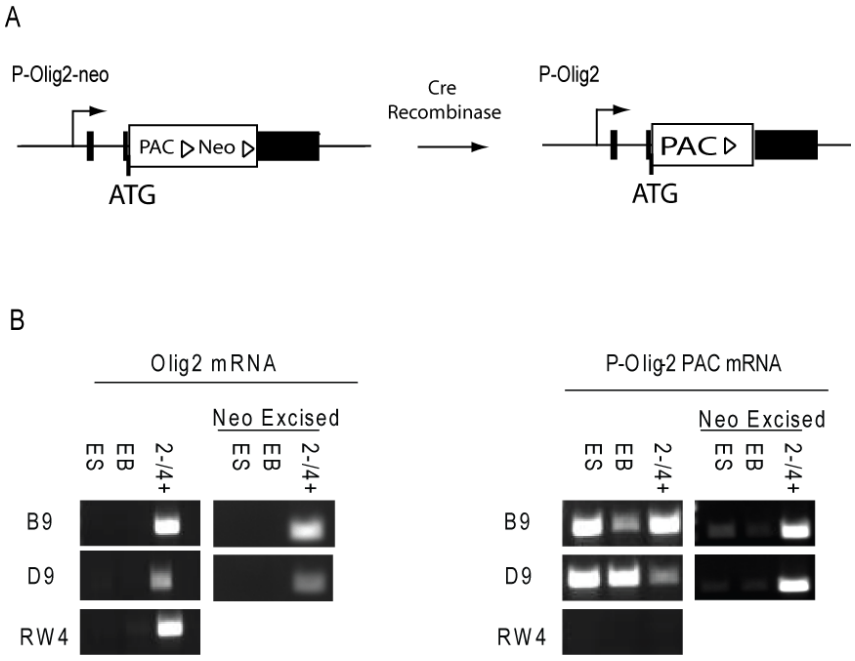


Figure 4.5: Excision of the PGK Promoter Is Required for Correct Regulation of P-Olig2 PAC mRNA. (A) Diagram of engineered allele before and after cre-excision. (B) Targeting the Olig2 locus with a PGK-neo cassette disrupts mRNA regulation of the transgenic allele. P-Olig2 PAC mRNA is expressed in all cell types unlike Olig2 which is restricted to neural cells. Despite unregulated mRNA expression, ES, EB and a subset of neural (2-/4+) cells show sensitivity to puromycin.

Since the mRNA expression of the P-Olig2 PAC mRNA appeared to follow the ubiquitous expression pattern of the PGK promoter, instead of Olig2 we reasoned that the very strong PGK promoter immediately adjacent to the puromycin cassette was interfering with proper regulation of the P-Olig2 message, despite the fact that the PGK promoter is arranged to direct transcription of the neomycin cassette downstream. There have been published accounts of nonspecific transcriptional activity and incorrect regulation of cassettes due to the presence of the PGK promoter (Pham, MacIvor et al. 1996; Scacheri, Crabtree et al. 2001). To attempt to eliminate the inappropriate expression of the knock-in, we used Cre recombinase to excise the floxed PGK-neo cassette. Resulting clones where PGK-neo had been excised were identified by junction PCR (Figure 4.4.1D, Figure 4.4.2D)

and neomyocin sensitivity. Excision of the PGK promoter restored proper mRNA regulation in both lines, restricting P-Olig2 PAC mRNA expression to induced ventral neural cells (2-/4+). From this we conclude that the PGK-neo cassette was responsible for the unexpected P-Olig2 PAC mRNA expression. This carries a general lesson for gene targeting that introducing sequences, especially strong promoters such as PGK, can have unintended effects. All subsequent data is from Cre-excised lines.

P-Olig2 Undifferentiated ES cells are Sensitive to Puromycin

Because Olig2 is not expressed in ES cells, it is expected that PAC will not be expressed in ES cells and P-Olig2 ES cells will be sensitive to puromycin. To learn about the functional regulation of the PAC, RW4 and P-Olig2 undifferentiated ES cells were plated and treated with puromycin over three days. A Live/Dead assay was performed to observe the action of the drug after 24, 48, and 72 hours. Cells are treated with two compounds, calcein and EthD-1, for Live/Dead staining. Living cells are visualized by calcein, a membrane permeant esterase substrate dye that is cleaved by intracellular esterases in living cells causing the cytoplasm of living cells to fluoresce bright green. Dead cells fluoresce red because EthD-1 is excluded by the membrane of living cells and can permeate the membrane of dead cells and bind to nucleic acids. As shown in Figure 4.6, untreated ES cells show bright green calcein stained living cells and very few EthD-1 stained dead cells. Within 24 hours of selection, 50-75% of RW4 and P-Olig2 ES cells are dead as shown by increased EthD-1 staining and a small percentage of detached rounded cells stained with calcein. Viability is less than ~4% for all three cell types after 48 hours. There are no calcein-stained surviving cells after 72 hours of puromycin for all three cell types. P-Olig2 ES cell lines are as sensitive to puromycin treatment as non-transgenic RW4

control ES cells. The knock-in is regulated as expected; the P-Olig2 transgene is not conferring functional resistance to puromycin in ES cells.

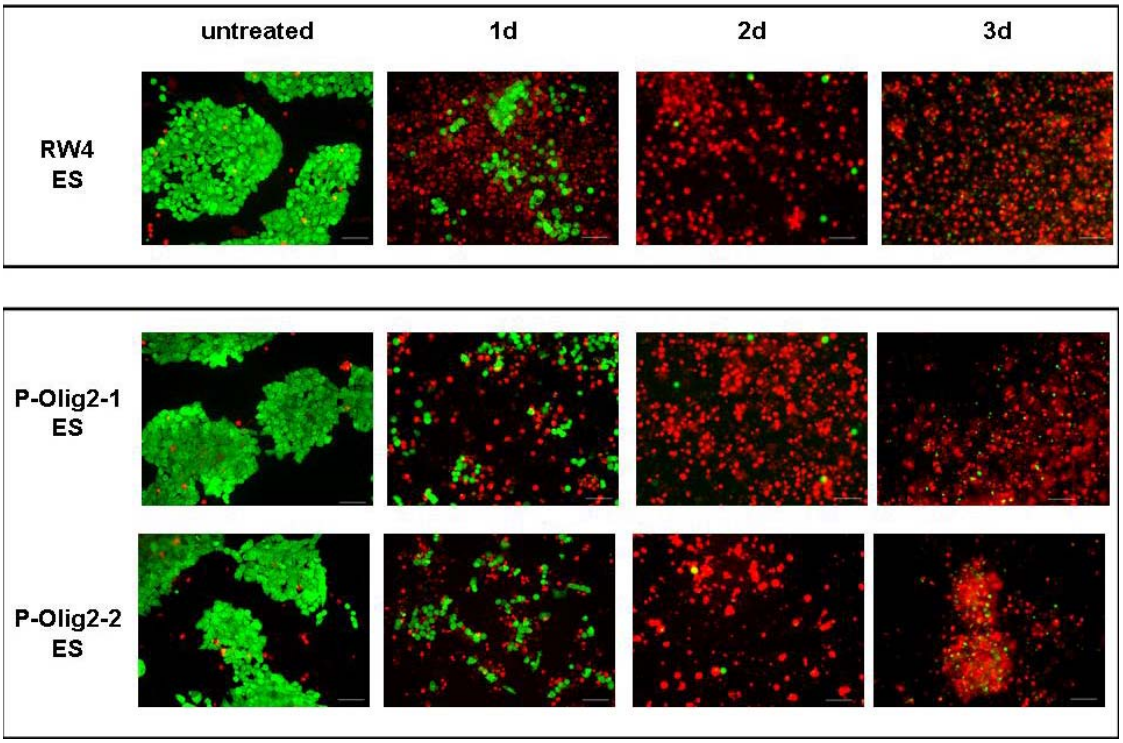


Figure 4.6: P-Olig2 ES Cells are Sensitive to Puromycin. Native RW4 ES cells as well as targeted P-Olig2 ES cells (POlig2-1, P-Olig2-2) are sensitive to puromycin. ES were seeded in 24 well plates and dosed as with 4μg/mL puromycin for the number of days shown. Live/Dead staining of untreated cells shows abundant living cells (green) and few dead cells (red). RW4 and transgenic ES cells are susceptible to puromycin as 24 hours and near complete cell death ~<4% green within 48 hours and no remaining viable cells after 72hrs. Scale bars are 50μm.

A Subset of P-Olig2 Ventral Neural Cells Survive Puromycin Treatment

If the PAC transgene is regulated similarly to the native Olig2 gene, at 2-/5+ and 2-/6+ 30% of P-Olig2 expressing ventral neural cells (Xian, Werth et al. 2005; Zhang, Horrell et al. 2008) should be protected from puromycin by expression of PAC. After treatment with puromycin for two days (2-/4+ to 2-/6+), RW4 and P-Olig2 neural cells (2-/6+) were observed by phase contrast microscopy and living cells were visualized

calcein staining as shown in Figure 4.7. Under phase contrast, RW4 and P-Olig2 neural cultures both display clear signs of cell death resulting from puromycin treatment. Many of the RW4 EBs detached from the surface of the dish during the course of treatment, but some remain attached. In contrast, calcein staining reveals living cells in the selected P-Olig2 neural cultures that are not seen in RW4 controls. The simplest interpretation is that expression of the transgene is protecting P-Olig2 neural cells from puromycin.

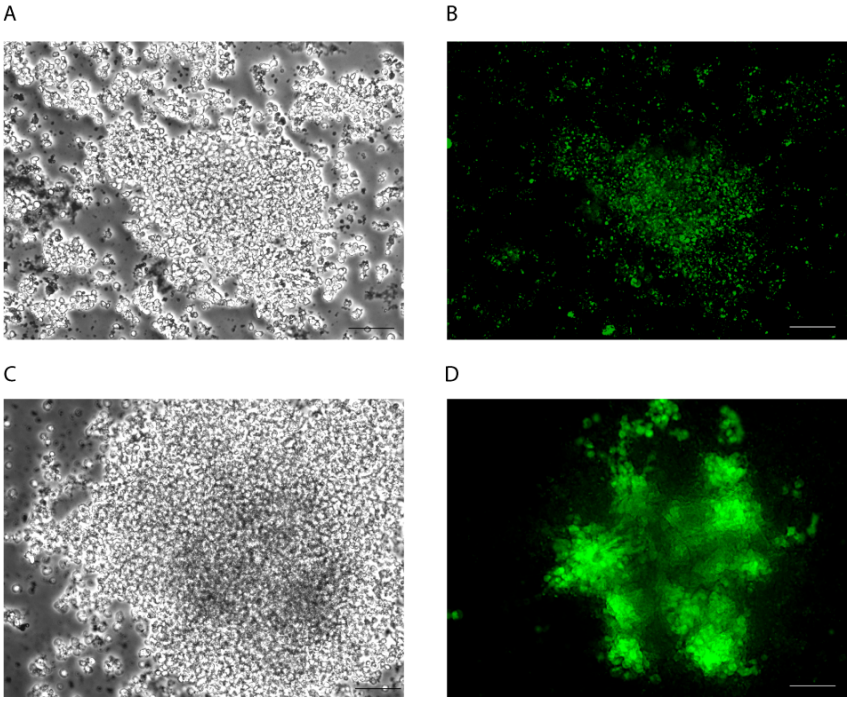


Figure 4.7: Expression of the Puromycin Acetyltransferase Gene (PAC) Protects P-Olig2 Neural Cells from Puromycin. Native RW4 and P-Olig2 ES cells were differentiated into neural cells for six days (2-/4+) and treated with 4μg/mL puromycin for 2 days (2-/4+ to 2-/6+). (A) Phase contrast image of treated puromycin treated RW4 neural cells shows dead cells and many EBs that have detached from the dish. (B) Calcein staining confirms that puromycin has killed all RW4 cells. (C) Phase contrast shows an attached P-Olig2 EB that looks very similar to RW4. (D) Calcein staining of P-Olig2 neural cell reveals abundant living cells beneath the debris providing evidence for a subset of neural cells protected from puromycin. Scale bars are 50μm.

The cellular phenotypes resulting from selection were visualized by calcein staining in treated and untreated cultures of P-Olig2 cells. After two days of puromycin treatment P-Olig2 cultures contain surviving cells, arranged in rosettes found both within and on the outskirts of most EBs (white arrows). Neural extensions are also visible (white arrows). Flattened neural cells that surround EBs in untreated cultures have been selected against (Figure 4.8).

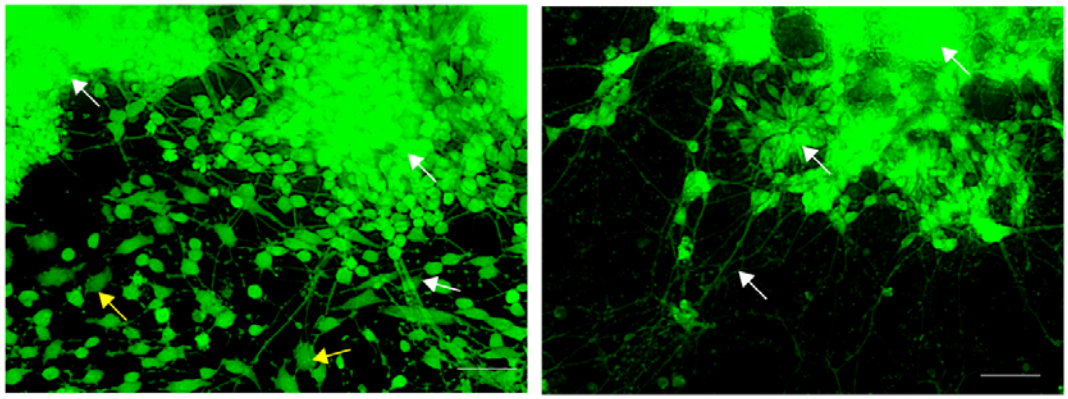


Figure 4.8: Treatment with Puromycin Selects a Subset of P-Olig2 Neural Cells. (Left) Calcein staining of Olig2-puro neural cells (2-/6+) shows rosettes and neurites (white arrows) but also other flattened neural cell types that surround embryoid bodies and decorate the surface of the dish (yellow arrows). (Right) Calcein staining of Olig2-puro neural cells (2-/6+) treated for two days with puromycin (from 2-/4+ to 2-/6+). Cells surviving treatment are arranged in rosettes both outside and inside of EBs and show some weak neural extensions (white arrows). Flattened neural cells that surround EBs present in untreated cultures have been selected against. Scale bars are 50μm.

In a separate experiment, selection was extended for one additional day. Rosettes and neural extensions become even more pronounced after three days of selection, and neural cells that surround EBs on the surface of the dish have not returned. In addition, there are surviving cells radiating from a few rosettes that have a radial glial morphology (Figure 4.9).

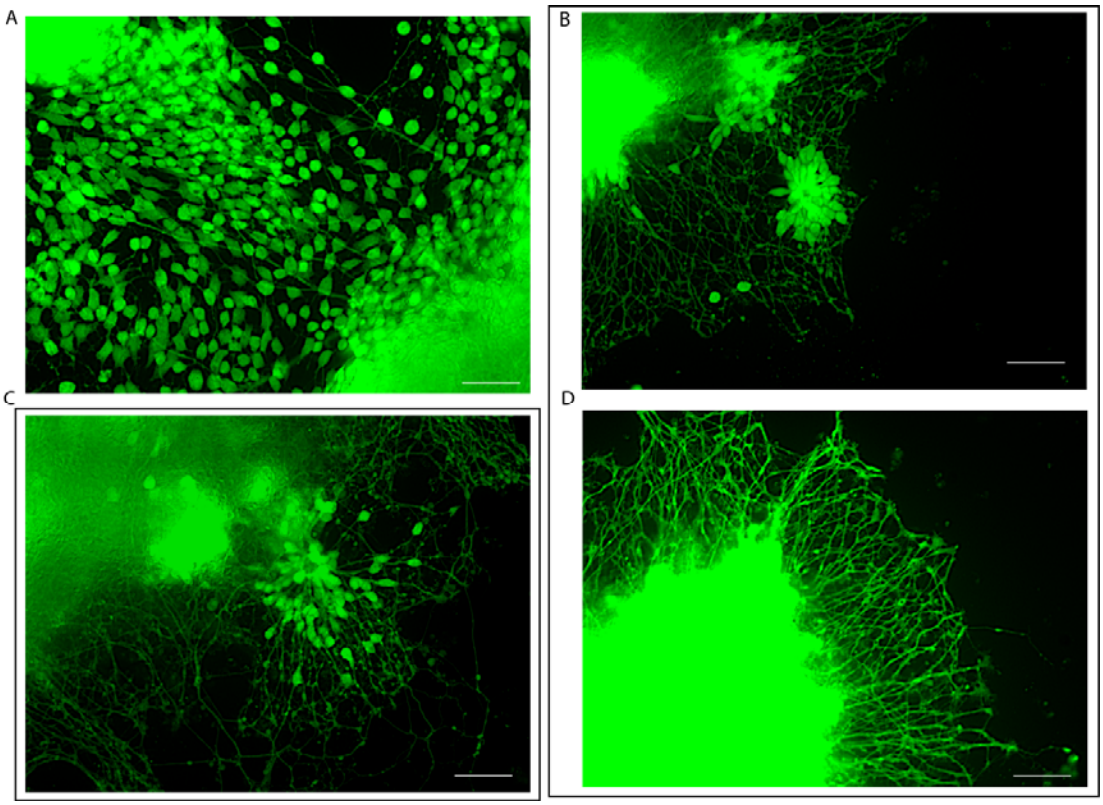


Figure 4.9: After Three Days of Selection, Resistant P-Olig2 Cells Thrive. Calcein staining of P-Olig2 neural cells (2-/7+) and P-Olig2 neural cells (2-/7+) selected for three days by puromycin (from 2-/4+ to 2-/7+). (A) Untreated P-Olig2 (2-/7+) have flattened neural cell types that surround embryoid bodies and decorate the surface of the dish which are absent in selected cultures. (B) After three days of selection P-Olig2 neural cells have cells surviving treatment arranged in rosettes similar to what is seen after two days of selection. (C) Selected P-Olig2 cells show a more pronounced radial glial morphology after three days. (D) Extensive neural processes are visible in P-Olig2 neural cultures after 3 days of puromycin selection.

Drug based selection allows for continuous selection of Olig2 expressing cells

Unlike selection using GFP and a FACS machine, drug based selection allows for continuous selection in culture. Selection of Olig2 expressing cells can be performed beyond the two days necessary for the action of the drug. Induction with retinoic acid and sonic hedgehog was extended to thirteen days, and drug selection was applied for nine days beginning at day (2-/4+). To insure that active puromycin was present, cells were redosed with puromycin at 2-/9+ and 2-/11+. Diverse cellular phenotypes were seen in selected P-Olig2 neural cultures stained with calcein including: rosettes, neural extensions, radial glial-like cells, large flat cells, and small rounded cell bodies with and without extensions (Figure 10). Continuous selection of Olig2 expressing cells is possible using drug selection.

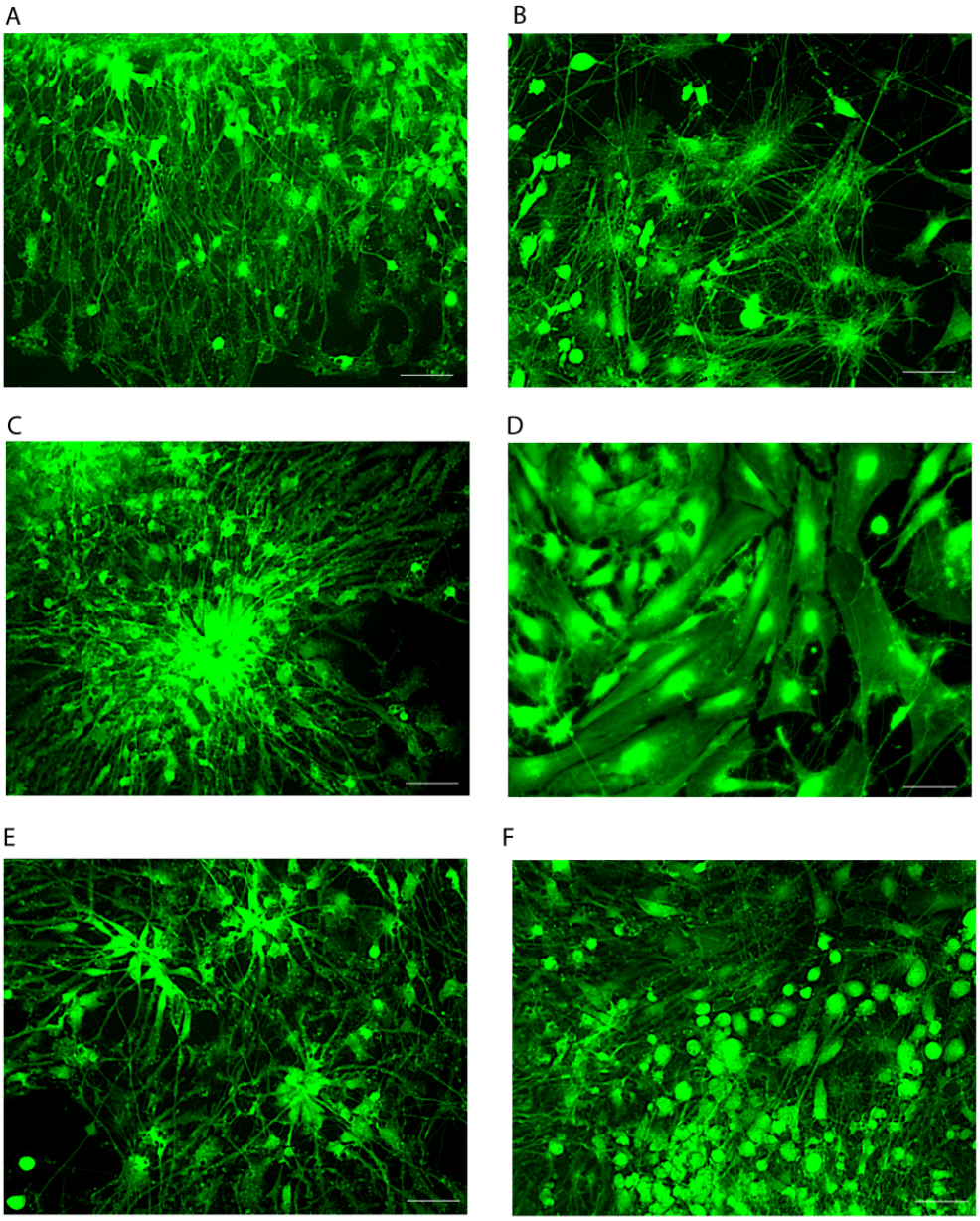


Figure 4.10: Continuous Selection Reveals Diversity of Olig2-Expressing Cells (A-F). Calcein staining of P-Olig2 neural cells (2-/13+) selected with 4μg/mL puromycin for 9 days (2-/4+ - 2-/13+). Cells were given a half feeding and redosed at 2-/9+ and 2-/11+. (A) Radial glial-like cells (seen in Figure 4.9) have continued to extend long processes through the course of selection. (B) Cells with flattened cell bodies and spindly processes appear. (C) Large rosettes are still visible which radiate processes. (D) Very large flat cells absent at 3 days of selection have emerged. (E) Smaller rosettes appear to be losing their shape as cells bodies radiate outward (F) Numerous small round cell bodies are visible through abundant processes.

Surviving Cells Express Olig2

Specificity of resistance to puromycin selection was investigated by staining selected cultures with Olig2 antibody. P-Olig2 neural cells were differentiated for 7 days (2-/5+) with selection applied to one set of cultures for 24 hours and stained with Olig2 antibody (Figure 4.11).

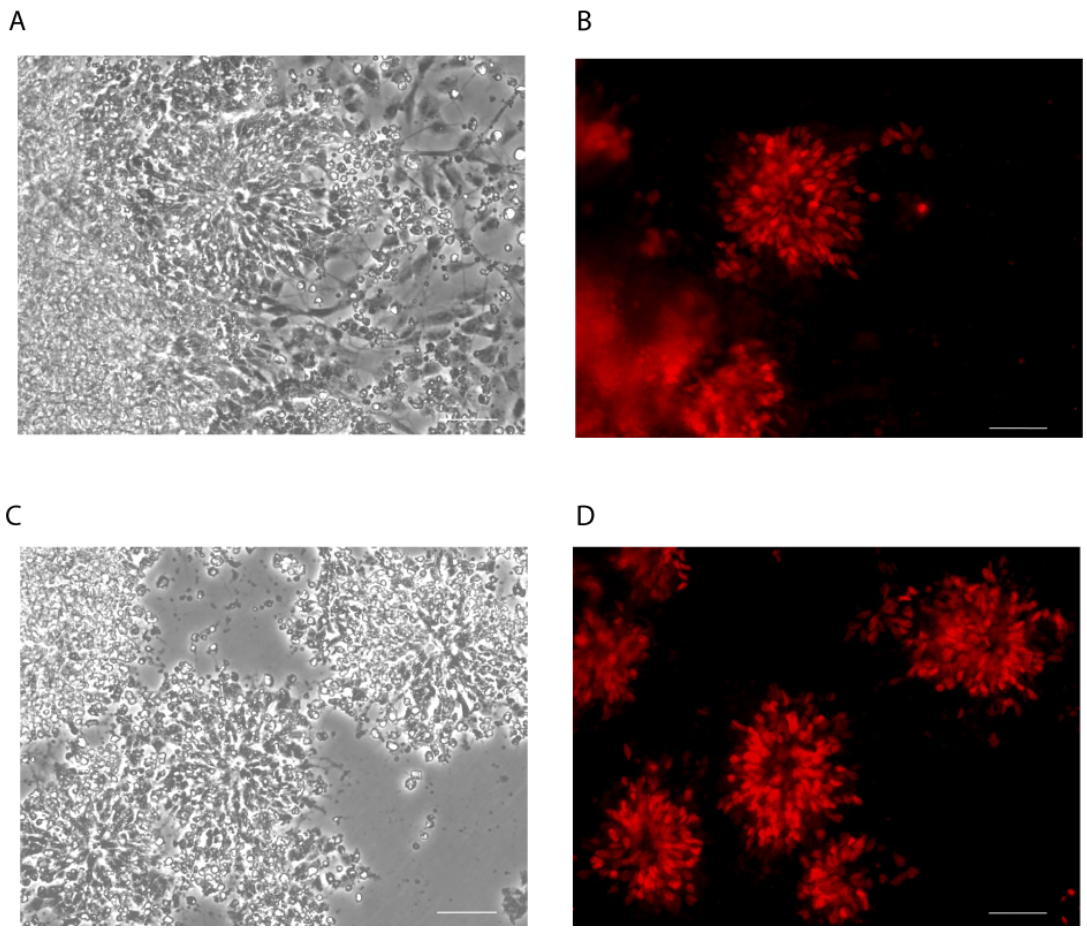


Figure 4.11: Surviving Cells Express Olig2. (A) P-Olig2 (2-/5+) neural cells contain many cell types including: rosettes, neurites, and flattened supporting cells that are visible using phase contrast. (B) Only rosettes stain with Olig2 antibody. (C) Selection of P-Olig2 neural cells for 24 hours with 4μg/mL puromycin (from 2-/4+ to 2-/5+) kills supporting cells and neurites, while rosettes survive. (D) The surviving rosettes express Olig2. Scale bar for all images is 50μm.

Under phase contrast unselected P-Olig2 cultures have diverse phenotypes. These include cells arranged in rosettes, neural extensions, and flatted supporting cells that surround embryoid bodies. However, only cells localized in rosettes stain with Olig2 antibody. After 24 hours of selection P-Olig2 neural cultures, only cells localized to rosettes are present. Neural extensions and supporting cells that were visible in unselected cultures are missing. Antibody staining of selected P-Olig2 cells shows that the selected cells stain with Olig2 antibody. The loss of several cell types from the selected cultures and strong Olig2 antibody staining of remaining selected cells demonstrates that selection of P-Olig2 cells with puromycin has indeed enriched for Olig2 expressing neural cells; the ES cell lines are working as designed.

4.5 Discussion

The ability to generate large numbers of well-defined ES cell derived neural cells is essential for stem cells to fulfill their potential in biomedical research and applications in regenerative medicine. A strategy employing a drug and cells genetically engineered with corresponding resistance gene has the ability to improve access, yield, and purity of selected cell populations (Klug, Soonpaa et al. 1996; Li, Pevny et al. 1998; Billon, Jolicoeur et al. 2002; Ying, Stavridis et al. 2003; Zandstra, Bauwens et al. 2003; Meissner, Wernig et al. 2007). Therefore, we investigated drug selection to isolate a subset of ES derived neural cells. Our goal was to generate an ES cell line where puromycin could be used to select for Olig2 expressing cells in ES derived neural cultures.

We successfully generated two knock-in ES cell lines termed P-Olig2-1 and P-Olig2-2. Both are generated by inserting a promoterless puromycin acetyltransferase gene

(PAC) into one allele of Olig2. In both knock-in lines PAC expression recapitulates the expression of the native Olig2 gene. Therefore, when P-Olig2 ES cells are differentiated into ventral neural cells, Olig2 expressing cells express PAC and are resistant to puromycin. The P-Olig2 ES lines are a new tool for research. By simply adding puromycin, Olig2 expressing cells can be selected from heterogeneous ES derived neural cells in a convenient, effective, and inexpensive manner.

The puromycin resistance profile of both P-Olig2 ES cell lines is highly similar to Olig2 expression. As expected, P-Olig2 undifferentiated ES cells were sensitive to puromycin. After two days of selection, approximately 95% of P-Olig2 ES cells were killed. After three days there were no surviving cells. On the other hand, a subset of differentiated P-Olig2 ES derived neural cells survived two and even three days of puromycin selection. Calcein staining allowed us to observe the surviving cells, which were arranged in rosettes found most often within embryoid bodies. Survival of rosettes gives strong evidence that the cell lines are indeed selecting for Olig2 expressing neural cells. Previous work with the G-Olig2 cell line has shown that rosettes in ES derived neural cultures contain Olig2 expressing cells (Xian and Gottlieb 2004). Olig2 antibody staining of P-Olig2 ES derived neural cultures gives further evidence for Olig2 expression in rosettes and has been shown previously (Zhang, Horrell et al. 2008). We also saw that P-Olig2 ES derived neural cultures become enriched in Olig2 containing rosettes following selection. Based on this evidence, we believe that puromycin selection in this system is purifying Olig2 expressing cells.

Drug selection allows for application of continuous selection and permits observation of highly organized and interconnected cellular networks of selected cells.

Following three days of puromycin selection of P-Olig2 ES derived neural cells, we observed surviving Olig2 expressing cells by calcein staining. We observed extensive processes radiating from cell bodies located in rosettes both within and outside of embryoid bodies. In some cases, cell bodies appeared to migrate away from the center of rosettes in a radial glial morphology. This agrees with observations of Olig2 expression in radial glial cells in the E11 spinal cord (Masahira, Takebayashi et al. 2006). Continued culture with retinoic acid and sonic hedgehog and selection over 9 days reveals multiple Olig2-expressing cellular phenotypes. These cells could belong to oligodendrocyte precursor lineage, which emerges from Olig2 expressing cells (Masahira, Takebayashi et al. 2006; Mukouyama, Deneen et al. 2006; Wu, Wu et al. 2006). Alternatively, these cells may represent phenotypes that do not exist in nature. In the embryo Olig2 expressing cells are subject to signals from neighboring Olig2 negative cells. These signals may restrict lineage progression of Olig2 expressing cells to a small number of allowable cell fates. Continued selection *in vitro* removes these signals, alleviating the restrictions. It is conceivable that Olig2 expressing cells can then differentiate into cells that do not occur *in vivo*. As a next step, gene expression profiles and antibody staining can be used to further characterize selected cells.

Genetically engineering ES cells introduces new sequences into the genome. These sequences, especially strong promoters such as PGK (Pham, MacIvor et al. 1996; Scacheri, Crabtree et al. 2001), can have unintended effects. In the initial targeted lines, the Olig2 puromycin acetyltransferase transgene was located directly upstream of a PGK promoter, an active promoter in ES cells. In these lines the Olig2-PAC mRNA was expressed in ES cells unlike the native Olig2 gene. Yet, despite the presence of Olig2-PAC mRNA, both ES

cell lines were sensitive to puromycin. We hypothesize that the proximity of active PGK promoter to the PAC transgene may have resulted in chromatin modifications in the region that permitted transcription. However, this transcript was not translated into functional enzyme as it did not confer resistance to puromycin. Removing the PGK-neo cassette, restored proper regulation to the PAC mRNA. Notably, excision of a PGK-neo cassette was also necessary for detectable GFP expression in the human ES R-Olig2 line (Xue, Wu et al. 2009). Techniques that facilitate removal of a positive selection cassette should be considered in design of future targeting vectors.

Several aspects of this work will be able to be used to develop selection methods of other subsets of neural cells. First, we have shown that puromycin selection can work in neural cells. Puromycin has very favorable kinetics for selection, allowing for selection in two days (Watanabe, Kai et al. 1995). We have also shown puromycin selection can be continuously applied for up to nine days in neural cells. It is very likely that puromycin selection will work with other neural promoters, and has been recently shown to function with endothelial specific promoters (Kim and von Recum 2009). Second, advances in recombineering technology have improved vector construction methods, increased targeting efficiency (Wu, Ying et al. 2008), and are also applicable to human ES (Xue, Wu et al. 2009). Using these methods we were able to construct two vectors and detected targeting in ES by PCR with an efficiency ranging from 3.7-5.2%. These approaches should be applicable to other neural genes.

Differentiation and puromycin selection of the P-Olig2 ES cell lines provide an easily accessible source of large numbers of Olig2 expressing cells. These selected cells will

become a key research tool in three areas: basic science investigations, transplantation research, and as a platform for drug toxicity and discovery.

For basic science studies of gene expression and epigenetic regulation, having a reliable source of selected cells will facilitate biochemical analysis, leading to new knowledge. Gene expression profiles of sorted Olig2 GFP cells have been obtained (Xian, Werth et al. 2005; Shin, Xue et al. 2007; Xue, Wu et al. 2009) at one or two points during the course of differentiation and interestingly show different coordinately expressed genes depending on the extent of differentiation. By making selection convenient and efficient, gene expression profiles of Olig2 expressing cells can be systematically investigated. Heterogeneity of ES derived neural cultures has also confounded investigations into epigenetic mechanisms that participate in regulation of Olig2 expression. An easily accessible population of Olig2 expressing cells will help decipher whether chromatin marks and non-coding RNAs (Lorberbaum D.S., Rieger, C.R, and Gottlieb, D.I., in preparation) observed in heterogeneous ventral neural cultures occur in cells that are expressing Olig2 or are from cells with silenced Olig2 expression. Several models have been proposed for fate determination of Olig2 expressing cells (Takebayashi, Yoshida et al. 2000; Rowitch, Lu et al. 2002; Wu, Wu et al. 2006). These cell lines provide a new tool to investigate these models. Selection can be applied at any time, in a pulsed or continuous manner, and the resulting cell phenotypes observed.

Several groups are currently investigating ES derived ventral neural cells as therapeutics for neurodegenerative disease and spinal cord injury (McDonald, Liu et al. 1999; Deshpande, Kim et al. 2006; Nayak, Kim et al. 2006; Coutts and Keirstead 2008; Willerth and Sakiyama-Elbert 2008; Johnson 2009). One great challenge for transplantation

researchers is to determine which subset of neural cells is most effective for restoring lost structures and function. For safety purposes, it is also critical for the therapeutic cells to be well-defined and not cause secondary tumors. Obtaining selected cells with a GFP line and FACS sorting is suitable for a proof of principle transplantation study. However, GFP selection becomes prohibitive for the necessary series of transplantation experiments.

Several practical concerns limit the application of GFP selected cells for transplantation research, namely scheduling, cost, and convenience. Transplantation studies require coordination of an animal surgery with the availability of replacement cells. Many laboratories that study transplantation have only limited access to a FACS. This is further compounded by the need to coordinate an animal surgery with cell selection at the FACS, with the additional problem that cell culture and surgical facilities are unlikely to be located in proximity to the FACS. Unforeseen delays in either surgery or cell sorting could easily introduce new variables, including contamination of sorted cells, which have the potential for derailing an expensive transplantation study. Using GFP selection with neural cells raises unique concerns because neural cells are delicate and must be dissociated from highly interconnected structures before selection. A drug selection based approach avoids many of the concerns described and will create opportunities for transplant experiments that otherwise would not be performed.

One drawback of the P-Olig2 lines for transplantation research is that the transplanted P-Olig2 cells will not be easily identified in an animal model. Transplantation researchers favor GFP lines because they allow for easy tracking of transplanted cells in the animal host. To this end, we are currently adding a ubiquitously expressed GFP to P-Olig2 line to aid in identification of transplanted cells.

Even before selected cells are transplanted into an animal model, drug selected cells can be used to test tissue engineering approaches. With a ready supply of cells, biomaterial scaffolds and combinations of growth factors (Reviewed by (Willerth and Sakiyama-Elbert 2008), (Levenberg, Burdick et al. 2005; Mahoney and Anseth 2007; Johnson, Parker et al. 2009)) can be optimized *in vitro* to obtain a desired differentiated cell population for transplantation.

An area of increased interest in research and industry is using subsets of differentiated ES cells as a testing platform to discover new growth factors, evaluate new drugs, and test compounds for toxicity (Reviewed by (Rubin 2008)). Screening assays require large numbers of biologically relevant cells. Once the desired population of cells is identified, methods to obtain sufficient yield and purity must be optimized. Recently, using a differentiation paradigm similar to that described here, optimum conditions to differentiate human ES to spinal motor neurons and oligodendocyte precursors were identified. Both of these important cell populations are derived from an Olig2 expressing intermediate (Hu, Du et al. 2009; Hu and Zhang 2009). Adding a drug selection strategy similar to P-Olig2 to these differentiation schemes could lead to increased yield and purity. Drug selection has the potential to facilitate production of a selected population of cells at a scale pivotal for investigations by the pharmaceutical industry.

Chapter 5

Conclusion

5.1 Summary of Findings

The overall goal of the thesis was to develop tools to advance investigations of stem cells in tissue engineering based therapeutics for neurodegenerative disease and spinal cord injury. Stem cells offer a renewable source of cells that can be differentiated into neural cells and have the potential to replace lost cells. Due to the complexity and heterogeneity of the nervous system, the optimal cell types for therapeutic use must be determined experimentally. The ability to both characterize and purify subpopulations of stem cells and neural cells is essential to advance these experiments. To accomplish the goal of this thesis, two tools to measure cell fate and a tool to purify a subset of cells were developed and evaluated.

In chapter 2, a digital method to measure mRNA as a tool for defining cell fate was developed and evaluated. Because stem cells occur as small populations and are heterogeneous, an mRNA profiling method that is sensitive, quantitative and allows for analysis of many genes simultaneously is desirable. We adapted polony technology (Mitra and Church 1999), a digital PCR based method, to measure mRNA expression from several key stem cell genes in small numbers of ES cells. Polonies measured mRNA from samples as small as 10 ES cells and one-fifth of a blastocyst, which contains approximately 35 ES cells (Chisholm, Johnson et al. 1985; Johnson and McConnell 2004). Polony counts

for three pluripotency genes were obtained from a single blastocyst. Furthermore, polony counts were comparable to the number of cDNAs measured by competitive PCR, a standard highly quantitative method (Zentilin and Giacca 2007). In summary, the results of chapter two show that the polony method is sensitive, can be applied to most genes, and allows for a degree of multiplexing.

In chapter 3, DNA methylation assays were developed and evaluated as a tool to characterize differentiation from ES to neural cells. We tested the hypothesis that methylation is associated with silencing of genes involved in neural cell fate specification. If true, methylation has the potential to be useful as a marker of cell fate. First, we developed and validated direct bisulfite sequencing to measure methylation. Then, methylation was measured for five genes in ES cells, ES derived neural cells, and non-neural tissues. In accordance with previous studies (Gidekel and Bergman 2002; Hattori, Nishino et al. 2004; Li, Pu et al. 2007; Fouse, Shen et al. 2008) pluripotency gene Oct3 was unmethylated in ES cells where it is expressed, and methylated in differentiated cells and tissues where Oct3 is silenced. Phosphoglycerokinase-1 (Pgk-1) a constitutively expressed housekeeping gene was unmethylated in all cells and tissues. Methylation status of three neural fate determining genes Sox1, Olig1, and Olig2 was not indicative of expression. These neural genes were largely unmethylated in non-neural cells and tissues where they are silenced. In summary, for some genes, such as Oct3, methylation status can serve as a diagnostic marker of cell fate. In contrast, methylation status of neural fate determining genes does not provide any information about cell fate.

In chapter 4, genetically engineered drug selection was developed and evaluated as a tool to purify subsets of neural cells. Genetically engineered drug selection offers a convenient, effective, and inexpensive method to obtain large numbers of purified cells, thereby facilitating transplantation research with the purified cells. We hypothesized that engineering ES cells with puromycin acetyltransferase gene (PAC) under the control of the Olig2 promoter would allow for selection of Olig2 expressing neural cells with puromycin. Before proceeding to the genetic engineering, we showed that PAC was able to confer resistance to puromycin in neural cells. Then, targeting vectors were designed and built with PAC inserted into the Olig2 gene. Correct targeting resulted in two ES cell lines, P-Olig2-1 and P-Olig2-2; one from each targeting vector was detected using PCR. Southern analysis confirmed the expected genomic structure for both ES cell lines. In both P-Olig2 lines PAC expression recapitulated the expression of the native Olig2 gene. As expected, P-Olig2 undifferentiated ES cells were sensitive to puromycin. As intended, a subset of P-Olig2 ventral neural cells was resistant to puromycin. Selected P-Olig2 neural cells were arranged in rosettes, typical of Olig2 expressing cells (Xian and Gottlieb 2004; Zhang, Horrell et al. 2008). Olig2 antibody staining confirmed that surviving cells express Olig2. A subset of P-Olig2 neural cells survived continuous selection for up to nine days. In summary, puromycin selection of P-Olig2 ES derived neural cells enables purification of Olig2 expressing cells.

Overall, this thesis achieves its goal by contributing tools to measure cell fate and purify subsets of neural cells. The tools help define and limit cellular heterogeneity, which

will be useful in the development of cellular and tissue engineering therapeutics for neurodegenerative disease and spinal cord injury.

5.2 Future Directions

5.2.1 Characterizing Cell Fate

The landscape of nucleic acid-based technologies for cell characterization has vastly improved since this thesis was initiated over 4 years ago. At that time digital PCR (Vogelstein and Kinzler 1999), polonies (Mitra and Church 1999; Mikkilineni, Mitra et al. 2004), and molecular colonies (Chetverina, Samatov et al. 2002) were at the forefront of the field. Advancements in high-throughput DNA sequencing technology have also advanced mRNA analysis. Newly developed mRNA profiling methods including: Illumina mRNAseq, ABI SOLiD SAGE, and Roche454 GS FLX Titanium Series Transcriptome Sequencing, are now standard tools in research and industry.

To use these technologies to measure mRNA, mRNA must be reverse transcribed into cDNA. However, reverse transcription introduces several biases and artifacts into the measurement of mRNA. First, conversion of mRNA to cDNA is inefficient. Similar to results found by others (Dufva, Svenningsson et al. 1995), in chapter 2 we measured a conversion efficiency of 5% using a model mRNA transcript. Second, reverse transcription is not equally efficient for all genes; so, the quantitative relationships measured between particular cDNAs may not reflect the abundance of their mRNAs. Reverse transcription efficiency is influenced by the structural properties of the particular mRNA, choice of primer to initiate reverse transcription, processivity of the particular enzyme, and location

of DNA sequence readout relative to priming site. Finally, cDNAs measured may not reflect natural mRNAs. Reverse transcriptases have demonstrated template switching artifacts (Cocquet, Chong et al. 2006), synthesis of spurious second-strand cDNA (Gubler 1987), and primer-independent reverse transcription. In summary, the ability to measure mRNA directly would alleviate biases and artifacts due to reverse transcription.

One approach that is poised to revolutionize mRNA profiling is the development of direct RNA sequencing technology (Ozsolak 2009). Measuring mRNA directly has the potential to increase sensitivity and make mRNA analysis more quantitative. In direct sequencing technology, polyadenylated mRNAs are captured on a surface coated with oligodT_{50} nucleotides. Using the oligo dT_{50} as a primer, a modified polymerase is used to synthesize a complementary strand of each tethered mRNA. Incorporated nucleotides are labeled by a fluorescent terminator. After each nucleotide is incorporated, the surface is imaged to capture sequence information. Then, the fluorescent terminator is cleaved to prepare for addition of the next nucleotide. The sequences obtained correspond to the particular mRNAs, and the numbers of molecules with a particular sequence reflect abundance. Thus, mRNA analysis can be performed without using reverse transcriptase. Direct RNA sequencing technology is sensitive, quantitative, and highly multiplex. It has all of the desirable attributes to be an improvement over current methods for stem cell mRNA profiling.

5.2.2 Cell Separation

Genetically engineered drug selectable ES cell lines that can be used to obtain relatively homogeneous populations of neural cells are not widely available ((Billon,

Jolicoeur et al. 2002; Ying, Stavridis et al. 2003), chapter 4). Most researchers still rely on FACS sorting of GFP knock-in and promoter reporter lines to obtain the purified cells that they need. Our ability to translate an established successful GFP reporter line (G-Olig2) into new successful drug selectable lines (P-Olig2) has strong implications for the development of new drug selectable lines for neural genes where valuable GFP reporter lines already exist.

An example of a neural gene where reporter lines have provided great insight is Homeobox gene Hb9. Hb9 is a transcription factor expressed specifically in spinal motor neurons (Arber, Han et al. 1999; Thaler, Harrison et al. 1999). Hb9 plays an essential role in specifying motoneuron identity during development and its expression persists in adult motoneurons. In 2002, Wicheterle et al (Wichterle, Lieberam et al. 2002) generated an ES cell line where GFP is expressed under the control of the Hb9 promoter. GFP expression recapitulates native Hb9 gene expression *in vivo* and is an effective tool for both visualizing and isolating motoneurons from ventral neural cultures. Development of a drug selectable line based on the expression of Hb9 could be very powerful. With new optimized growth factors regimens to differentiate hES cells to motoneurons (Hu and Zhang 2009) and the ability to purify these motoneurons using drug selection, we would be one stem closer to clinically relevant numbers of well-defined motoneurons for transplantation research and therapy.

Appendix A
A Gene Targeting Cis-Trans Test Using Polonies

A.1 Abstract

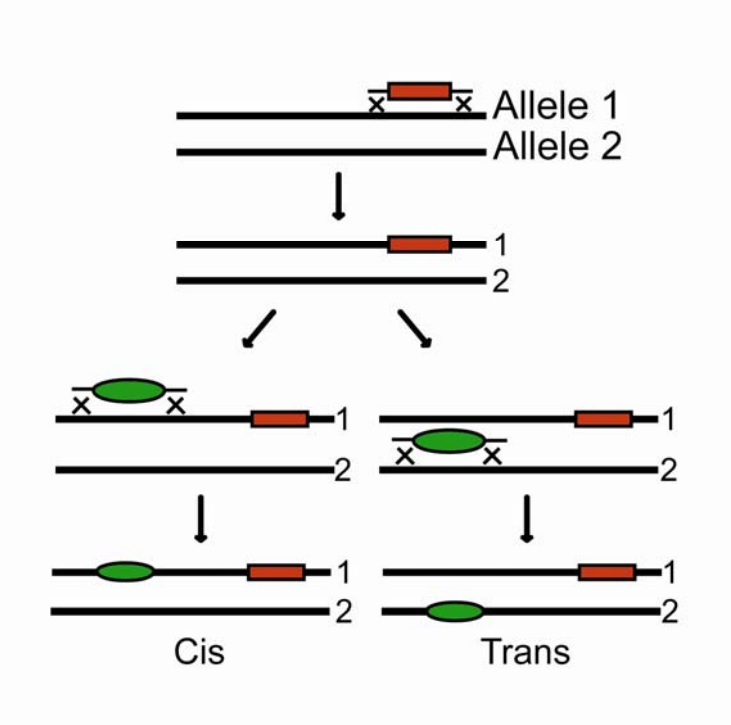
Repeated cycles of gene targeting to the same allele are desired in chromosomal engineering. After the first targeting event, it is necessary to determine if the secondary targeting event has occurred on the same allele (in cis) or has altered the unmodified allele (in trans). Typical cis-trans tests including Southerns and PCR lose the ability to detect linkage when the distance between targeting events is greater than 10kb. Here we describe a polony cis-trans test which easily detects linkage over 14kb and can be extended for targets spanning megabase distances.

A.2 Introduction

Chromosomal engineering is an emerging discipline for studying gene function and regulation. Genomic DNA is targeted to generate insertions, deletions (knock-outs), and replacements (knock-ins) through homologous recombination. To map regulatory elements of a particular locus, repeated cycles of gene targeting to the same allele are desired. For example, a primary targeting event introduces a reporter gene such as GFP into the open reading frame of a gene of interest. Gene regulation by neighboring cis-DNA elements could then be studied by secondary targeting events to predicted promoter, enhancer, or

repressor regions. Once a subsequent new targeting event is obtained it is necessary to determine whether the new targeting event occurred in cis or trans to an existing targeting event (Figure A.1).

Figure A.1: Two Outcomes of Sequential Targeting Events



Two outcomes result from sequential targeting events. Events can occur on the same allele (cis) or on different alleles (trans) of the same chromosome. A cis-trans test is required to identify desired cis-events which will facilitate the study of gene regulation.

Traditional gene targeting cis-trans tests are performed using Southern Analysis (Leung, Malkova et al. 1997), PCR(Langston and Symington 2004), and Cre-Lox excision(Zheng, Sage et al. 2000). As the genomic distance between the two targeting events increases, the limitations of all three of these methods become apparent. Southern detection of cis-targeting events requires generation of a restriction fragment that spans the two targeting events. As fragment sizes approach 15kb, few restriction sites are available and bands become weak. Similarly, the practicality of using liquid phase PCR for cis-trans

detection decreases as the distance between the two targeting events increases beyond 10kb. Another common, but limited approach for determining cis-targeting events relies on LoxP site insertion flanking each targeting event. The addition of cre-recombinase is used to selectively loop out the region between the LoxP sites in the cis-linked events, producing changes in expression of a selectable marker or a new junction that can be assayed by PCR.

In principle, the polony method is better suited for performing cis-trans analysis on longer range targeting events than existing techniques. In a polony reaction amplification products from two independent targets in nearby genomic loci occupy the same micro-region in the gel matrix. Cis elements are clearly identified by overlapping polonies, while targets in trans conformation are distinguished by unique polony signatures. The polony cis-trans test can be performed in a relatively short period of time. More importantly, the polony method is not constrained by the distance between the targeting events. Cis-linkage has been observed with polonies in the form of genotyping and haplotyping on targets 45kb apart (Mitra, Butty et al. 2003), and with targets separated by 50-100Mb with modifications to the DNA preparation method (Zhang, Zhu et al. 2006). Here we demonstrate the feasibility of a polony gene targeting cis-trans test.

A.3 Materials and Methods

DNA was extracted from ES cells (maintained by DD) by 5 Prime prep (5 Prime Gaithersburg, MD, done by XZ or SH) and diluted to an appropriate concentration for polonies in 10mM Tris-EDTA, pH 8.0 (B5 1:50, TG25 1:20). Polony slides were prepared according to previous work (Mitra and Church 1999). DNA was added to a liquid phase acrylamide gel mix containing PCR components and both the Neo gene and genomic target were amplified within the gel as described in Chapter 2.

Each amplicon was detected by its internal Cy-5 linked hybridization probe as previously described. Images for each probe pair were overlayed using the screen function in Adobe Photoshop. Polonies were counted manually using ImageJ software and cell counter applet. Olig2+14kb hybridization staining was pseudo colored red and Neo cassette hybridization staining was pseudo colored green.

A.4 Results

To establish the polony method as an alternative cis-trans test, we used polonies to determine cis linkage in DNA from two ES cell lines containing a Neomycin cassette (Neo) and two copies of the genomic target, an arbitrary region 14kb upstream of the Olig2 gene (UCSC browser build mm9; Chr. 16: 91,100,054 – 91,100,400). The cis-linked cell line (TG25) contains Neo in the open reading frame of one allele of the Olig2 gene (Chr. 16: 91,114,409- 91,117,534; (Xian, McNichols et al. 2003)). The unlinked cell line (B5) contains Neo on an entirely different chromosome (UCSC browser build mm9; Chr. 12: 32,013,051) than the genomic target and thereby serves as a model for a trans relationship.

Polonies were prepared from a cis-linked (TG25) cell line and an unlinked cell line (B5). Slides were prepared with DNA from a single cell line and included primers for both the genomic target and a region of the Neomycin gene. Primer sequences are given in Table A.1.

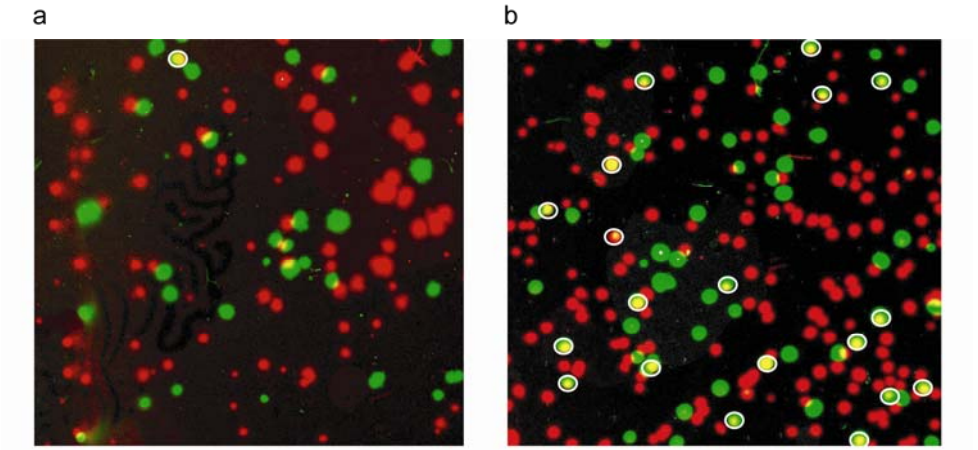
Table A.1: Primers for Cis-Trans Polony Assay

Primer Name	Sequence (5'-3')
NeoF4	ctttcgacctgcagccaatatggga
NeoR1_Ac	Ac-gcaggagcaaggtgagatgacagga
Hybridization 1	Cy5-tgaatgaactgcaggacgag
Olig2_+14kbF3	taggagggagggaaacctgggatgtaataa
Olig2_+14kbR3_Ac	Ac-agtgtggagcagggtagctacacaaatggt
Olig2_+14kbhybe3	Cy5- tcaatcaatcaatcaatcagca

Reverse primers have an Acrydite functional group on the 5' end to crosslink one strand of the polony amplicon into the gel matrix. After amplification, the free complementary strand is denatured, facilitating identification of each amplicon with an internal gene specific Cy5 labeled hybridization probe. For the second gene assayed, the first probe was removed before the second hybridization sequence was annealed. Amplicons for both genes were ~350bp.

Polonies for each gene were sequentially identified by unique hybridization probes, images were pseudo colored for each gene and images of the same slide overlaid. Sample regions from polony slides for the B5 and TG25 cell lines are shown in Figure B.2.

Figure A.2: Polony Cis-Trans Test on DNA from Two Transgenic ES Cell Lines



Genomic target polonies 14kb upstream of the Olig2 gene are red while Neo polonies are green. Overlapping polonies, which are yellow and have been circled in white, indicate presence of both genomic target and Neo amplicons in close proximity and predict a cis relationship. a) Region of a polony slide containing DNA from unlinked (B5) cell line. b) Region of a polony slide from DNA from cis-linked (TG25) cell line.

In Figure A.2 genomic target polonies 14kb upstream of the Olig2 gene are red while Neo polonies are depicted in green. Yellow polonies, a result of overlapping red and green polonies, depict the presence of both genomic target and Neo amplicon in close proximity and predict a cis relationship. By visual inspection of the slides, many yellow polonies were observed on slides prepared with TG25 DNA while only a few yellow polonies were seen on slides prepared with B5 DNA, demonstrating cis-linkage of Neo with the genomic target in the TG25 cell line.

To confirm observations and obtain quantitative estimates of the efficiency of the assay, polonies were counted on 8 slides, four prepared with B5 DNA and four with TG25 DNA. Polony counts are given in Table A.2.

Table A.2: Polony Counts from Cis-Trans Test on Two Transgenic ES Cell Lines

Cell Line	Total Genomic Target polonies (Red)	Total Neo Polonies (Green)	Cis-events (Yellow)
B5 (B)	299	115	3
B5 (B)	167	112	4
B5 (D)	377	96	1
B5 (D)	343	103	2
TG25 (B)	286	119	24
TG25 (B)	724	244	47
TG25 (D)	222	87	16
TG25 (D)	356	87	15

Cis-elements can be distinguished from un-linked elements by polony counts. Four slides were prepared with DNA from the cis-linked cell line (TG25) and four slides were prepared with DNA from the un-linked cell line (B5). (B) stands for 1M Betaine incorporated into the reaction. (D) stands for 2% DMSO incorporated into the polony reaction. Betaine and DMSO were added to the polony mix to attempt to minimize differences in amplification efficiency due to amplicon GC content.

For a yellow polony to be counted, more than 75% of the area of both polonies needed to be overlapping based on visual observation. Based on the numbers of overlapping polonies, a cell line with two elements in cis orientation is clearly distinguished from a cell line with unlinked targets. The number of overlapping polonies for cis-linked TG25 DNA ranged from 15-47, while a background of 1-4 of overlapping polonies were present on polony slides prepared with unlinked B5 DNA. The tenfold increase in polony overlap apparent in TG25 over B5 is enough to set linked apart from unlinked.

Efficiency estimates can also serve as a metric independent of the actual number of overlapping polonies on a particular slide to distinguish cis-targeting events. This is particularly relevant when sample DNA concentrations differ. The percentage of cis-linked polonies of the total number of polonies, also termed the co-amplification efficiency

(Mitra, Butty et al. 2003), for a particular target can be calculated to aid in determining a cis targeting event from an unlinked or trans targeting event and also estimate efficiency of the polony reaction. For this experiment the percentage of cis-linked polonies of the total number of Neo polonies averaged 18.8% ± 1.3% for TG25 and 2.3% ± 1.1% for B5 and for cis-linked polonies of the total number of genomic target polonies averaged 6.6% ± 1.7% for TG25 and 1.1% ± .9% for B5.

A.5 Discussion

The field of chromosomal engineering and study of gene regulation will benefit from improved methods for detecting cis-targeted elements. We demonstrate that the polony cis-trans test successfully detects cis-linked elements 14kb apart from unlinked elements in DNA from two transgenic ES cell lines.

The polony cis-trans test uses two sets of primers, one for each event, to generate polonies. Polonies visualized for a particular amplicon are specific to that amplicon and corresponding internal hybridization probe. Polony experiments on a non-transgenic ES cell line (RW4) confirmed that genomic targets were present, and the cell line did not contain a Neo cassette (data not shown). There also were no polonies on slides prepared without DNA. Polonies that are amplified from independent molecules remain spatially distinct, although a small number of overlapping polonies do appear, likely as a result of co-localization of independent templates in the gel. When the two regions of interest reside on the same molecule, amplification of each target generates a polony molecule that occupies the same small region of the gel matrix, indicating cis-linkage of the two amplicons.

In this particular assay, if polonies were generated from all cis-linked targets with 100% efficiency, we would expect that 100% of Neo (green) polonies would overlap with genomic target Olig2+14kb (red) polonies and 50% of genomic target Olig2+14kb (red) polonies would overlap with Neo (green) polonies. The observed co-amplification efficiency in the linked cell line for this study for Neo was 18.8% and 6.6% for the genomic target. Previously reported co-amplification efficiencies from polony haplotyping studies varied from 4-34% and were influenced by the DNA quality (Mitra 2003). DNA degradation or fragmentation may be responsible for lower than anticipated co-amplification efficiency by disrupting linked elements. As long as co-amplification efficiency is greater than the background noise, increasing the number of polonies present on a slide will ease cis-trans discrimination.

Several factors will enhance the polony cis-trans test beyond what is demonstrated here. For the model system described, we expect the total number of genomic target polonies to the number of Neo polonies in both the cis-linked and unlinked cell lines to be in a ratio of 2:1, unlike a two targeting event recognition or haplotyping, where the two transgene targets would be expected in a 1:1 ratio. Thus, cis-events should be even easier to spot in a true two transgene targeting study due to decreased background. Further modification of the polony protocol by decreasing amplicon size from 350bp used here to 100bp and the use of an argon chamber to decrease the amount of unpolymerized acrylamide in the gel would likely improve coamplification efficiency (Mitra and Church 1999); (Mitra, Butty et al. 2003). The use of a bead-emulsion digital PCR amplification and flow cell detection (Kim, Porreca et al. 2007) instead of a gel based system would further increase the sensitivity and processivity of the cis-trans polony assay. With stringent

genomic DNA extraction techniques (Zhang, Zhu et al. 2006), one could imagine extending the polony cis-trans test for targets spanning megabase distances.

Appendix B
Supplemental information for Chapter 2: Polony profiling stem cells and blastocysts

Table B.1: Polony primers for ES and blastocyst expression analysis

Gene	Accession #	Primers	Sequence 5'-3'
Oct 3/4	NM_013633	Forward	CGCCAATCAGCTTGGGCTAGAGAAG
		Reverse	Ac-TGCCTCAGTTTGAATGCATGGGAG
		Hybridization	Cy5-GGGGCTGTATCCTTTCCTCT
Nanog	NM_028016	Forward	GTGCTTGAACACCCTTACCCACGCC
		Reverse	Ac-GGACTCCAAGGACAAGCAAGCACC
		Hybridization	Cy5-CACCCACCCATGCTAGTCTT
Rex 1	NM_009556	Forward	GCA AAGGCAGGGAAGAAATGCTGA
		Reverse	Ac-TCTGCCGTATGCAAAAGTCCCCAT
		Hybridization	Cy5-GATTGTCCTCAGGCTGGGTA
Cdx 2	NM_007673	Forward	ATC AGCCTCTTTTGCCCCAGCTCTT
		Reverse	Ac-CCACCCCTTCCTGATTTGTGGAGA
		Hybridization	Cy5'-AACTACAGGAGCCAGAGGCA
GLUT-1	NM_011400	Forward	GCCCTGTCCAGACACTTGCCTTCT T
		Reverse	Ac-CCAGTGCTTCCAACTGGTCTCAGG
		Hybridization	Cy5-CCTATGGCCAAGGACACACT

Table B.2: Primers for BNI5polyA RNA RT efficiency studies

Primer	Sequence 5'- 3'
Construction primers	
YeastF	TCGACTCTAGA**GGATCC**CCGGGTAATGCAT
Forward Linker	tggtctcaaatttggtgacagattttggcaTTTAGTTCCAATCCAAAAT TGGCCATTTTC
Reverse Linker	AAATGGCCAATTTTGGATTGGAACTAAAtgccaaaatctgt caccaaatttgagacca
pTRI-Xef	ccgggc**gagctc**gaattcgcgaaattccg
Polony Primers	
Forward 1	AATTGTCCCAATCGCTCTGGAGGAA
Reverse1_AC	AC GCCCAGCTCAGACGCTGATGTGATA
Hybridization 1	Cy5-GCCCTACAAATCCGTTCAGA
Forward 4	CGGTGAGCAGCAATTCCAATTCAAG
Reverse 4_AC	AC ACTCCGACCTCATATTTCTCCTTCG
Forward 5	CATCAGCGTCTGAGCTGGGCAGTAT
Reverse 5_AC	AC CGCCACGATCAATTTGCTTTCTCAG
Hybridization 3	Cy5-GGGCAGTATTGCCAAGCTGGA

All sequences from BNI5 plasmid are in CAPS. All sequences from pTRI-Xef (Megascript kit, Ambion) which contained a 70bp polyA+ region that was appended to the BNI5 gene are in lower case. BamHI and SacI restriction sites are in bold. Dilution experiment was performed using Forward 5, Reverse 5_AC, and Hybridization 3

Table B.3: Primers for competitive PCR analysis

Gene, Accession #, & Primer	Sequence 5'-3'	Amplicon length (bp)
Oct3/4 (296bp) NM_013633		
Forward primer	CGCCAATCAGCTTGGGCTAGAGAAG	296
Reverse primer	GTGCCTCAGTTTGAATGCATGGGAG	
Competitor with 50bp deletion	gtgcctcagtttgaatgcatgggagagccctgagtagagtgtggtgaagtgggggcttcc	246
Nanog (315bp) NM_028016		
Forward primer	GTGCTTGAACACCCTTACCCACGCC	315
Reverse primer	GGACTCCAAGGACAAGCAAGCACC	
Competitor with 50 bp deletion	gtgcttgaacacccttacccacgcccaccctcattttgaggggtgaggtttaaaagtata	265
Rex1 (276bp) NM_009556		
Forward primer	GCAAAGGCAGGGAAGAAATGCTGA	276
Reverse primer	TTCTGCCGTATGCAAAAGTCCCCAT	
Competitor with 50bp deletion	ttctgccgtatgcaaaagtccccatcccctggaactatgccatggaataccaaaagaggc	226

Appendix C
Supplemental Information for Chapter 3: From ES to Neural Cells: DNA Methylation as a Marker of Neural Cell Fate

This appendix contains tables with the primers used for methylation mapping experiments and source code for computational methods used for experimental design and data analysis.

C.1 Primers

Table C.1. Primers for construction and testing of mock methylated template

Primer	Sequence 5'-3'
SK F3	CACAATTCCACACAACATACGAGCCGGAAG
SK R4	TACGCGCAGCGTGACCGCTACACTTGCCAG
PSK_F1	gtaaaacgacggccagtgAAAAGGCCAGCAAAAGGCCAGGAAC
PSK_R1/2	GgaaacagctatgaccatgTGCACACAGCCCAGCTTGGAGC
PSK_BF1	GtaaaacgacggccagtgAAAAGGTTAGTAAAAGGTTAGGAAT
PSK_BR1/2	GgaaacagctatgaccatgTACACACAACCCAACTTAAAC
PSK_F2	GtaaaacgacggccagtgATGTGAGCAAAAGGCCAGCAAAAGGCCAG
PSK_BF2	GtaaaacgacggccagtgATGTGAGTAAAAAGGTTAGTAAAAGGTTAG
PSK_F3	GtaaaacgacggccagtgGCAGAAGTGGTCCTGCAACTTTATC
PSK_R3	ggaaacagctatgaccatgGAATTATGCAGTGCTGCCATAACCATGAGT
PSK_BF3	gtaaaacgacggccagtgGTAGAAGTGGTTTTGTAATTTTATT
PSK_BR3	ggaaacagctatgaccatgAAATTATACAATACTACCATAACCATAAAT
PSK_F4	gtaaaacgacggccagtgGTTTGGTATGGCTTCATTCAGCTC
PSK_R4	ggaaacagctatgaccatgCAATGATGAGCACTTTTAAAGTTCTGCTAT
PSK_BF4	gtaaaacgacggccagtgGTTTGGTATGGTTTTATTTAGTTT
PSK_BR4	ggaaacagctatgaccatgCAATAATAAACACTTTTAAAATTCTACTAT
PSK_F5	gtaaaacgacggccagtgCCAAGTCATTCTGAGAATAGTGTATG
PSK_R5	gtaaaacgacggccagtgAGAGTATGAGTATTCAACATTTCC
PSK_BF5	gtaaaacgacggccagtgTTAAGTTATTTTGAGAATAGTGTATG
PSK_BR5	ggaaacagctatgaccatgAAAATATAAATATTCAACATTTCC

Table C. 2. Gene primers for methylation analysis by bisulfite sequencing

a. Primers for promoter analysis. All primers were located within 500bp surrounding transcription start site (UCSC genome browser July 2007 build), GC content for overall region analyzed (see Figure 3.2) and CpG Observed/ Expected are also given. Sequencing primers are located at the 5' end of each primer and the last two rows of the table.

Primer	Sequence 5'-3'	GC %	Obs/ Exp
Pgk1_F6	gtaaaacgacggccagtgTCCCAAGGCAGTCTGGAGCATG	62%	0.71
Pgk1_R6	ggaaacagctatgaccatgTTCCCAGCCTCTGAGCCCAG		
Oct4_F6	gtaaaacgacggccagtgGCTGTCTTGTCCTGGCCTTGGACAT	59%	0.34
Oct4_R6	ggaaacagctatgaccatgGGTCCTCTCACCCCTGCCTTGGGT		
Oct4_F7	gtaaaacgacggccagtgTGCAGTGCCAACAGGCTTTGTGGTG		
Oct4_R7	ggaaacagctatgaccatgACCCTCTAGCCTTGACCTCTGGCC		
Oct4_F8	gtaaaacgacggccagtgGTGACCCAAGGCAGGGGTGAGAG		
Oct4_R8	ggaaacagctatgaccatgTGAAGCCAGGTGTCCAGCCATGGGG		
Olig2_F6	gtaaaacgacggccagtgAAACTCCAGGTGTGGCAAGCAACCT	61%	0.76
Olig2_R6	ggaaacagctatgaccatgCCTCTCAGCACACAGCCAATGGGC		
Olig2_F7	gtaaaacgacggccagtgCATTGGCTGTGTGCTGAGAGGAG		
Olig2_R7	ggaaacagctatgaccatgAATTAGATTTGAGGTGCTC		
Olig2_F8	gtaaaacgacggccagtgCCATTGGCTGTGTGCTGAGAGGA		
Olig2_R8	ggaaacagctatgaccatgCTGTAATAAGCATCCACACCTTTC		
m13 -20F	Gtaaaacgacggccagtg		
m13 -27R	Ggaaacagctatgaccatg		

b. Primers for Olig2 mapping. Sequence for each primer, location of each primer relative to the transcription start site (TSS) of each gene (UCSC genome browser July 2007 build), GC content, and CpG Observed/Expected are given for each amplicon. Sequencing primers are located at the 5' end of each primer and the last two rows of the table.

Primer_name	Sequence 5'-3'	Location rel to TSS	GC %	Obs/ Exp
Olig2_6_BF1	TGTGTTGGAGAGTAATGGATTAATAGA	-5516		
Olig2_6_BR1	AAATTCTAACCACCTAAACCAC	-5254	61%	0.82
Olig2_6_BF2	TTTGGGAATTAGGTTAATTAGATGG	-4851		
Olig2_6_BR2	CACAAACTCCTCCTATCCCTTTATA	-4551	58%	0.41
Olig2_6_BF3	TTGTATAATTGTTTGAGAAGGTTT	-3376		
Olig2_6_BR3	TCCTTCCTTCCTACTAACATTCAAA	-3025	60%	0.63
Olig2_6_BF4	AGGATTAAATAGGGAATTAGGGAAT	-2982		
Olig2_6_BR4	TCCTAAACTATTAACCTCCTCCTTCA	-2768	63%	0.68
Olig2_6_BF5	TAGGGAAGGTATATTGGTTTAGTTG	-2697		
Olig2_6_BR5	ATCTCCTTCTCCTCTCAACTACCTT	-2399	58%	0.57
Olig2_6_BF7	ATTGTTTGTTTGTTTGTTTGTTAAA	-1528		
Olig2_6_BR7	AACCCATACAAATAACCCTACCTAC	-1221	47%	0.67
Olig2_6_BF8	TAAGATGTTGGAAGTTTAGTGGTTG	-881		
Olig2_6_BR8	CTCCCTTAACTCCTTTCTACTTAATAACA	-598	51%	0.59
Olig2_g_BF1	TGATTGTTTGGGTGTTTATATTTAT	234		
Olig2_g_BR1	ACATACAAATTATCCTAATCTCTCTCCTC	609	52%	0.79
Olig2_g_BF2	TTTATTGAGGGTTATTTAGAAGTTT	526		
Olig2_g_BR2	AACAAATACACAAACACAAACACAC	779	56%	0.56
Olig2_g_BF3	TTTATAGGAGGGATTGTGTTT	950		
Olig2_g_BR3	AACTCAATCATCTACTTCTTATCTTTCTTA	1161	65%	0.90
Olig2_g_BF4	GAAAGATAAGAAGTAGATGATTGAG	1135		
Olig2_g_BR4	TCCTCCAACRAATTAATAAAC	1354	62%	1.11
Olig2_g_BF7	GTTTTGGAGAGTTAAGGGTTG	1841		
Olig2_g_BR7	CCCTTAAAATATTCAACCAAAA	2054	60%	0.69
Olig2_g_BF9	ATAATATTAAAAAGTTGTTGTGAATATAGT	2679		
Olig2_g_BR9	CATAAAAAAATTTACCCCTC	2914	45%	1.0
m13 -20F	gtaaaacgacggccagtg			
m13 -27R	ggaaacagctatgaccatg			

c. Primers for Olig1 and Sox1 analysis. Sequence for each primer, location of each primer relative to the transcription start site (TSS) of each gene (UCSC genome browser July 2007 build), GC content, and CpG Observed/Expected are given for each amplicon. Sequencing primers are located at the 5' end of each primer and the last two rows of the table.

Primer_name	Sequence 5'-3'	Location rel to TSS	GC%	Obs/ Exp
Olig1_5_BF1	TGAAGTTAGTAAGTGTGTTTATAGTAAA	-2235		
Olig1_5_BR1	ACTAACCAATAAATAACATCAAACC	-1952	46%	0.43
Olig1_5_BF2	GGGAGGTAGAGATAAGGAGGATTT	-1324		
Olig1_5_BR2	CAAATATATCCAATTAACAAACAACCA	-1009	45%	0.64
Olig1_5_BF3	AGTGAGAATTAAGATAAGTAGAGGG	-647		
Olig1_5_BR3	AATATAATCAACCAATAATCCCATC	-339	43%	0.46
Olig1_g_BF1	TTGTATGAGTTGGTGGGTTATAGGTA	202		
Olig1_g_BR1	CATAACCAAATTCAAATCCTACATAC	450	65%	0.83
Olig1_g_BF2	TTTAAGTGAGGGTTGGTTGG	883		
Olig1_g_BR2	AAACTTCTAACTCTAAACAAATAAA	1140	60%	0.62
Olig1_g_BF3	TTATTATAATTTATTTATTGGTAGGG	1149		
Olig1_g_BR3	TAAACTTTCTAAACCTCCTAAATCC	1454	63%	0.55
Olig1_g_BF4	GGATTTATTAAAATAGGTAGTAAGGTA	1756		
Olig1_g_BR4	AATCAAATAAAATACCAATTAAATTC	2106	50%	1.2
Sox1_5k_BF2	GGGTTTATATTGATAGGTTTAGGG	-3725		
Sox1_5k_BR2	ACAAACTCTAATATACCCTCACAAA	-3412	55%	0.45
Sox1_5k_BF3	GAATGGTTTGAAATGAAAGTTTAAT	-1609		
Sox1_5k_BR3	AACAAACCTAATCAAATACCTACAAAC	-1312	45%	0.60
Sox1_g_BF1	GGTATTTGGGATTAGTATATGTTTAG	316		
Sox1_g_BR1	AAACCACAACAACAACAACAAC	685	60%	0.72
Sox1_g_BF4	TTGTTGGAGGAGAAATTATTTGTAG	2866		
Sox1_g_BR4	TTAACCTTATCCCACCACTAAACAC	3183	44%	0.38
m13 -20F	Gtaaaacgacggccagtg			
m13 -27R	Ggaaacagctatgaccatg			

C.2 Source code for primer design, sequencing data analysis, and amplicon analysis

C.2.1 cg_search.pl - Mapping CpG sites in a region of interest

```
#!/usr/bin/perl
use strict;
############################################################
#Cara Rieger
#March 27, 2008
#cg_search.pl - use for defining a region of interest in a
#sequence and mapping CpG sites
#Also can be used for printing subsequence for primer design
############################################################
#declare variables
my $infile;
my $sequence="";
my @sites;
my $region_of_interest=700;

#Check inputs
if (@ARGV!=1){
die "\n Usage perl cg_search.pl <sequencefile>\n\n";}

#Get filename and open file handle
$infile=shift;
open IN, $infile or die "Couldn't open $infile\n";

####Read in contents of file line by line
while(my $line =<IN>){
    if ($line=~/^\w/){ #if line begins with a word char.keep it and
append seq
      chomp($line);
    $sequence.=uc($line);
    }
}
#print $sequence;
close IN;

######Split the sequence into array on each character
##Dual checking##
my $substring= substr ($sequence,0,$region_of_interest);
print "$substring\n\n"; #use for printing sequence and designing
primers
my $cg_count= $substring =~ s/CG/CG/g;
print "Number of CGs $cg_count \n\n";
my $cpg=0;
@sites=split(//,$substring); #split sequence on each character
my $location=-500; #starting location relative to TS startsite
my $i=0;
```

```
for($i=0; $i<@sites; $i++){
    if($sites[$i] eq "C" && $sites[$i+1] eq "G"){
      print "$location\n";
      $cpg++;}
    $location++;}

print "\nCpGs in breakdown $cpg\n";
print "number in sequence $i\n";
#!/usr/bin/perl
```

C.2.2 primer.pl - append universal tails to primers; identify unbisulfite treated cognate primers

```
use strict;
############################################################
#Cara Rieger
#March 27, 2008
#
#Before running this program
#Obtain sequences of interest and use MethPrimer to generate primers
#for methylation specific PCR
#Store primer results as text file with > header for each entry
#Generate fasta file of sequences of interest
###primer1.pl
#Input:Output from MethPrimer and sequence files
#Can also be used to add tails
#Output: Tailed Bisulfite primers | Tailed Regular Primers
#############
# To run: perl primer1.pl primerfile.txt sequences.fa
############################################################

#Take in sequences for analysis
##
#Check the contents of the file
if (scalar (@ARGV) !=2){
    die "Usage perl primer1.pl fastafile.fa primerfile.txt\n";
}

#Open the fasta file containing the sequences and store then
my $fasta_file=shift(@ARGV); #takes first argument for cmd line
open IN, $fasta_file; #IN is file handler to read file

#Read through the file one line at a time
#and store the sequences and headers as elements in an array
my @sequence;
my $sequence_counter=0;
my $sequence_counter=0;
my @header;

while (my $line=<IN>){
#seaching for the >= will break each record
    if ($line=~/^>/){
```

```
      chomp($line);
      $header[$sequence_counter]=substr($line,1);
      $sequence_counter++;
    }elsif($line=~/^\w/){
      chomp($line);
      $sequence[$sequence_counter-1]=$sequence[$sequence_counter-
1].$line;
    }
}
close IN;

########################################
#Open the primer file which has each gene annotated with a
#>Name
#Followed by copied output from MethPrimer
#####################################

#Variables:
my $line_counter=0;
my $lp_tail="gtaaaacgacggccagtg"; #M13 Forward -20
my $rp_tail="ggaaacagctatgaccatg"; #M13 Reverse -27
my $primer_num; #starting number for all primers
my $primer_name;
my $primer_counter=0; #total number of primers;
my @outputB;#Bisulfite output
my @outputR;#Regular output

#Open the primer file
my $primer_file= shift @ARGV; #takes first argument from cmd line
(file name)
open INPUT, $primer_file; # INPUT is file handler to read $primer_file

#Read through the file one line at a time
#Parse data keeping important stuff

while (my $line = <INPUT>){ #line is equal to next line of file
#Seaching for the >= will break each record
    if ($line=~/^>/){

#Capture primer name
      chomp($line); #get rid of new lines
       $primer_name = substr($line,1,7); #gets the sequence query name
      $primer_num=6; #starting number for all primers
      $line_counter=0;
    }
    elsif($line=~/^(\s)*$/){
      $line_counter++;
    }
################################################
#Divide the line into a data array with slots
#[num, left,primer, LP, LP_length, TM, GC%, C's, LSequence]
##[right, primer, RP, RP_length, TM, CG%, C's, RSequence]
##[Product, size:, PS, ProductTM, CpGs in Product, prodCpGs]
############################################
```

```
    elsif($line_counter>=1 && $line!~/^(\s)*$/){
      chomp($line);

      if ($line=~/Left/){
      #put primer name and primer F/R and number on array
        my $primer_nameL=$primer_name.'_'.'BF'.$primer_num; #generates
a unique name for each primer
        my @dataL = split (/\s{2,}/ , $line) ; #splits the string of
data into an array as defined by 2 or more spaces
      for(my $i=0; $i<3; $i++){
          shift(@dataL); #get rid of number, left, primer
      }
      my $lp_start= shift(@dataL);
      my $lp_length=shift(@dataL);
      my $lp_seq=$lp_tail.pop(@dataL);
        my $lp_start_len=join(",",$lp_start,$lp_length);
      push(@outputB, $primer_nameL,$lp_seq,$lp_start_len);
        $primer_counter++;

      }elsif($line=~/Right/){
      my @dataR = split (/\s{2,}/ , $line) ; #splits at 2 or more
spaces
      my $primer_nameR=$primer_name.'_'.'BR'.$primer_num;
    for(my $i=0; $i<2; $i++){
      shift(@dataR);#dump the text
    }
        my $rp_start= shift(@dataR);
        my $rp_length=shift(@dataR);
      my $rp_start_len=join(",",$rp_start,$rp_length);
        my $rp_seq=$rp_tail.pop(@dataR);
      push(@outputB, $primer_nameR, $rp_seq, $rp_start_len);
      $primer_num++; #advance the primer_numbers for both primers
      $primer_counter++; #increment the counter for overall primers
    }
    }
}

my @final_outputB=@outputB;

close $primer_file;

###################
#Generating regular primers from methprimer bisulfite primers
#Shift off primer name and exchange for regular primer name
#Match the gene name from primer name to header name in sequence file
#Get corresponding sequences using start and length of sequence
#Make reverse complement for reverse sequence
###############

for(my $i=0;$i<$primer_counter;$i++){
    my $primer_name= shift(@outputB);
    my @name= split(/_B/,$primer_name);
    $primer_name=join("_",shift(@name),pop(@name));
```

```
    push(@outputR, $primer_name);
#j corresponds for the number of sequences from fasta file
    for(my $j=0; $j<scalar(@header);$j++){
      if($primer_name=~/^$header[$j]/){
      shift(@outputB);#get rid of bisulfite sequence in outputB
          if($primer_name=~/_F/){
            my $lp_s_l=shift(@outputB);
            my @lp_start_length=split(/,/,$lp_s_l);
            my $lp_seq=substr($sequence[$j],($lp_start_length[0]-
1),$lp_start_length[1]);
            $lp_seq=$lp_tail.$lp_seq;
            push(@outputR,$lp_seq,$lp_s_l);
          }
      if($primer_name=~/_R/){
          my $rp_s_l=shift(@outputB);
          my @rp_start_length=split(/,/,$rp_s_l);
          my $rp_seq=substr($sequence[$j],($rp_start_length[0]-
$rp_start_length[1]),$rp_start_length[1]);
          $rp_seq=reverse($rp_seq);
#need to complement reverse sequence;
          my @rc_seq;
          my @seq= split(//,$rp_seq);
          for (my $i=0; $i<scalar(@seq); $i++){
            if ($seq[$i] eq "A"){
                $rc_seq[$i]="T";
            }elsif($seq[$i] eq "T"){
                $rc_seq[$i]="A";
            }elsif($seq[$i] eq "C"){
                $rc_seq[$i]="G";
            }else{
                $rc_seq[$i]="C";}
          }
          $rp_seq=join(//,@rc_seq);
          $rp_seq=$rp_tail.$rp_seq;
          push(@outputR, $rp_seq,$rp_s_l);
      }
    }
    }
}

##########################################
###Print out the bisulfite and regular primers in Tab delimited format
############################
print "Primer_name\tSeq\tStart,Length\n"; #header line
for (my $i=0; $i<$primer_counter/2; $i++){
    my $primer_nameL=shift(@final_outputB);
    my $lp_seq=shift(@final_outputB);
    my $lp_start_len=shift(@final_outputB);
    my $primer_nameR=shift(@final_outputB);
    my $rp_seq=shift(@final_outputB);
    my $rp_start_len=shift(@final_outputB);
print "$primer_nameL\t$lp_seq\t$lp_start_len\n";
print "$primer_nameR\t$rp_seq\t$rp_start_len\n";
}
```

```
print "\nPrimer_name\tSeq\tStart,Length\n"; #header line
for (my $i=0; $i<$primer_counter/2; $i++){
    my $primer_nameL=shift(@outputR);
    my $lp_seq=shift(@outputR);
    my $lp_start_len=shift(@outputR);
    my $primer_nameR=shift(@outputR);
    my $rp_seq=shift(@outputR);
    my $rp_start_len=shift(@outputR);
    print "$primer_nameL\t$lp_seq\t$lp_start_len\n";
    print "$primer_nameR\t$rp_seq\t$rp_start_len\n";}
```

C.2.3 seqsort.pl - take in sequencing data and parse out important data for alignment and calculation

```
#!/usr/bin/perl
use strict;

############################################################
#Cara Rieger
#April 1, 2008
#seq_sort.pl - use for taking a file of all sequencing reads and
parsing out important data for alignment and calculation
#derived from running
#%>strings *.seq.bin>sequences.seq in directory with all sequence data
#################################################################
#declare variables
my $infile; #file containing sequences of interest
my @header;
my @sequence;
my $record=0;

#Check inputs
if (@ARGV!=1){
die "\n Usage perl seq_sort.pl <sequences.seq>\n\n";}

#Get filename and open file handle
$infile=shift;
open IN, $infile or die "Couldn't open $infile\n";

####Read in contents of file line by line
while(my $line =<IN>){
    if ($line=~/seq/){
      chomp($line);
      $record++;
      my @name=split('_', $line);# get rid of well numbers
      $header[$record-1]=shift(@name);
      $header[$record-1]=substr($header[$record-1],2); #get rid of
initials
    }elsif($line!~/mBIN/ && $line!~/seq/){
#if line is not the .seqline or a mBIN line keep it and append seq
      chomp($line);
    $sequence[$record-1].=uc($line); #store each sequence as element
in array
```

```
    }
}
close IN;


#Want to reverse complement the reverse sequences
for (my $i=0; $i<scalar(@sequence);$i++){
    if ($header[$i]=~/R$/){
      $header[$i]=$header[$i]."_RC";
    my $r_seq=reverse($sequence[$i]);
#need to complement reverse sequence;
      my @rc_seq;
      my @seq= split(//,$r_seq);
      for (my $i=0; $i<scalar(@seq); $i++){
      if ($seq[$i] eq "A"){
          $rc_seq[$i]="T";
      }elsif($seq[$i] eq "T"){
          $rc_seq[$i]="A";
      }elsif($seq[$i] eq "C"){
          $rc_seq[$i]="G";
      }else{
          $rc_seq[$i]="C";}
    }
    $sequence[$i]=join(//,@rc_seq);
    }
}

#print "$header[0]\t$header[1]\n";
my $name=substr($header[0],0,3);
my $file="$name.fa";
open OUT, ">$file";


##Print the matching gene sequences to a new fasta file for algnment
purposes
for (my $i=0;$i<$record; $i++){
    print OUT  ">$header[$i]\n";
    print OUT "$sequence[$i]\n";
 if (substr($header[$i],0,3) ne substr($header[$i+1],0,3) &&
$i!=$record-1){
     close OUT;
     $name=substr($header[$i+1],0,3);
     $file="$name.fa";
     open OUT, ">$file";
 }
}close OUT;
```

C.2.4 count8.pl - take in sequencing data and generate a bisulfite conversion matrix and overall statistics

```
#!/usr/bin/perl -w
use strict;

#Cara Rieger
#April 23, 2008
#count8.pl

##################################
#Need a reference sequence name_ref in fasta sequence file
#Can be used with Fasta output of aligned ClusalW from bisulfite
sequencing
#To generate matrix of C->C and C->T conversions
####################################


# Check to make sure that one file was given on the command line, if
not
# print a correct usage statement.
#
if (@ARGV != 1) {
    die "\nUsage: perl count8.pl <fasta file>\n\n";
}

# Get the file name and open an instream to it.
#
my $file = shift;
open IN, $file or die "Couldn't open file: $file\n";

#Initialize two arrays for fasta headers and sequence information#
my $contig_no=-1; # keep track of number of contigs
my @all_sequences;
my @all_headers;

# Read in the contents of the file one line at a time and store it in
$line
#
while (my $line = <IN>) {

    # if it's a line that begins with a '>' or non-word char if
($line=~ /^\W/)
    # store the line in @all_headers - giving you the fasta info
    # chomp off the newline (chomp($line))
    # if it is sequence it will be stored in @all_sequences
    # uppercase it for easier counting (uc($line)
    # and append it to the end of $all_sequences, the string that is
the growing
    # within the array ($all_sequences[$contig_no]  .= uc($line)) .
    #
    #
    if ($line =~ /^>/) {
      chomp($line);
```

```
      $contig_no=$contig_no +1;
      $all_headers[$contig_no]=$line;
      #print "header: $all_headers[$contig_no]  \n";
       }
    else{
       chomp($line);
      $all_sequences[$contig_no].= uc($line);
      #print "$all_sequences[$contig_no] \n";
    }
}

#We're done with the file, so close it
#
close IN;

my @bases; #2D array of bases for all sequences
my @reference_seq;# array containing reference sequence separated by
bases
my $ref_length;#length of reference sequence
my $ref; #index of reference sequence

########################
#Find the reference sequence
###################################

for (my $i=0; $i<$contig_no+1; $i++){
    if($all_headers[$i]=~/ref/){
      $ref=$i;
    }
}

my $k=0; #counter for ref sequence location
my $j_start=0; #start of ref sequence relative to alignment
my $j_end; #end of ref sequence relative to alignment
my @C_sites; #array containing the location of all C sites in
reference
my @CpG_sites; #array containing the location of all CpG sites in
reference

########################
#Get the start and stop of reference sequence and indices of C's and
CpGs
########################
my @temp=split(//, $all_sequences[$ref]);
for (my $j=0; $j<scalar(@temp);$j++){
    if($temp[$j]=~/\w/ || $temp[$j+1]=~/\w/){ #takes care of one bp
gap in ref
      $reference_seq[$k]=$temp[$j];
      if($reference_seq[$k]=~/C/){
          push(@C_sites,$k);
      }if($reference_seq[$k]=~/C/ && $temp[$j+1]=~/G/){
          push(@CpG_sites,$k);
      }
      $j_end=$j;
```

```
      $k++;

    }elsif($k==0){
    $j_start++;
}
}

#####################################################
#Make 2D array of bases, cropping to same size as reference, by
getting rid of leading and trailing sequence
#####################################################
for (my $i=0; $i<$contig_no+1; $i++){
    @temp=split(//, $all_sequences[$i]);
    $k=0;
    for (my $j=$j_start; $j<$j_end; $j++){
      $bases[$i][$k]=$temp[$j];
      $k++;
    }
}

my @C_conversion_matrix; #contains 0 for C->T and 1 for C->C for each
location
my @C_counts;# maintains a count of total number of C's in sequence
my @CG_conversion_matrix; #contains 0 for C->T and 1 for C->for each
location
my @CG_counts; # maintains a count of total number of CG's in sequence
my @C_interest; #maintains a list of locations where C->C
my @CpG_interest;# maintains a list of locations where CG->CG
my @T_counts; # count of C->T converters
my @TpG_counts; #count of CG->TG converters
#################################################
#Have 2D array of bases, and indices of C's and CpGs
#Now put a 1 for every C match and 0 for every C->T, -1 for C-
>something else
#Keep array of sites of interest (those that remain C's)
###################################

for(my $i=0; $i<$contig_no+1; $i++){
    $C_counts[$i]=0;
    $T_counts[$i]=0;
    $C_interest[$i]="";
    for(my $j=0; $j<scalar(@C_sites); $j++){
      my $position=$C_sites[$j];
      if($bases[$i][$position]=~/C/){
          $C_conversion_matrix[$i][$j]=1;
          $C_counts[$i]++;
          $C_interest[$i].="$position,";
      }elsif($bases[$i][$position]=~/T/){
          $C_conversion_matrix[$i][$j]=0;
          $T_counts[$i]++;
      }else{
          $C_conversion_matrix[$i][$j]=-1;
      }
    }
```

```
}

#######Now for the CpG sites

for(my $i=0; $i<$contig_no+1; $i++){
    $CG_counts[$i]=0;
    $TpG_counts[$i]=0;
    $CpG_interest[$i]="";
    for(my $j=0; $j<scalar(@CpG_sites); $j++){
        my $position=$CpG_sites[$j];
        if($bases[$i][$position]=~/C/){
            $CG_conversion_matrix[$i][$j]=1;
          $CG_counts[$i]++;
          $CpG_interest[$i].="$position,";
        }elsif($bases[$i][$position]=~/T/){
            $CG_conversion_matrix[$i][$j]=0;
          $TpG_counts[$i]++;
        }else{
            $CG_conversion_matrix[$i][$j]=-1;
        }
    }
}
#################################################################
###Print out statistics about bisulfite conversion
#####################################
my $totalC= scalar(@C_sites);
my $totalCG= scalar(@CpG_sites);

###Print Names and C and CG counts
print "Name\tC->C\tCG->CG\tC->T\tCG->TG\n";
for (my $i=0; $i<$contig_no+1; $i++){
    print "$all_headers[$i]\t";
    print "$C_counts[$i]/$totalC\t";
    print "$CG_counts[$i]/$totalCG\t";
    print "$T_counts[$i]/$totalC\t";
    print "$TpG_counts[$i]/$totalCG\n";
}

for (my $i=0; $i<$contig_no+1; $i++){
    print "\n$all_headers[$i]\n";
    print "C's of interest: $C_interest[$i]\n";
    print "CG's of interest: $CpG_interest[$i]\n";
}

############################
###Print C_conversion matrix and CpG_Conversion_matrix
################################################
#print "\n----------------------------------------\n";
#print "Csites\t@C_sites\n";
#for (my $i=0;$i<$contig_no+1;$i++){
#    print "$all_headers[$i]\t";
#    for (my $j=0; $j<scalar(@C_sites);$j++){
#        print "$C_conversion_matrix[$i][$j] ";
#    }print "\n";
```

```
# }

#print "CGsites\t@CpG_sites\n";
#for (my $i=0;$i<$contig_no+1;$i++){
#    print "$all_headers[$i]\t";
#    for (my $j=0; $j<scalar(@CpG_sites);$j++){
#        print "$CG_conversion_matrix[$i][$j] ";
#    }print "\n";
#}
############################
```

C.2.5 cg_count.pl – used to calculate GC% and observed over expected CpGs for any sequence of fasta file of sequences

```
#!/usr/bin/perl
use strict;

##########################################
#Cara Rieger
#cg_count.pl - use for calculating observed/expected CpG over a fasta
file of sequences particular
#Adapted from Assignment 1 solution from Bio5488 course
#July 28, 2008
##########################################

#declare variables
my $infile;
my $sequence="";

#Check inputs
if (@ARGV!=1){
die "\n Usage perl cg_search.pl <sequencefile>\n\n";}

#Get filename and open file handle
$infile=shift;
open IN, $infile or die "Couldn't open $infile\n";

####Read in contents of file line by line
print "Gene_amplicon\tGC%\tLength\tObsCpG\tExpCpG\tObs_Expect\n";
while(my $line =<IN>){

    if ($line=~/^\w/){ #if line begins with a word char.keep it and
append seq
      chomp($line);
      $sequence.=uc($line);
    }

    elsif ($line=~/^\>/){ #if line begins with carat keep the text as
title for line and increment counter
      chomp($line);
      if($sequence=~/\w+/){ # if sequence is not empty we need to
count it
          ##Do counts on the sequence##
          my $c_count= $sequence =~s/C/C/g;
```

```
          my $g_count= $sequence =~s/G/G/g;
          my $t_count= $sequence =~s/T/T/g;
          my $a_count =$sequence =~s/A/A/g;

      my $total_bases= $c_count + $g_count + $t_count + $a_count;
      my $expect =
int((($c_count/$total_bases)*($g_count/$total_bases))*$total_bases-1);
      my $gc_percent= sprintf "%.2f", ($c_count +
$g_count)/$total_bases;
      my $cpg= $sequence =~s/CG/CG/g;
      my $o_e= sprintf "%.2f", $cpg/$expect;

      print "$gc_percent\t";
      print "$total_bases\t";
      print "$cpg\t";
      print "$expect\t";
      print "$o_e \n";

      }
      print "$line\t"; # print the line if sequence is empty
      $sequence=""; # reset the sequence
    }

}

close IN;

##Do counts on the last sequence##
my $c_count= $sequence =~s/C/C/g;
my $g_count= $sequence =~s/G/G/g;
my $t_count= $sequence =~s/T/T/g;
my $a_count =$sequence =~s/A/A/g;

my $total_bases= $c_count + $g_count + $t_count + $a_count;
my $expect =
int((($c_count/$total_bases)*($g_count/$total_bases))*$total_bases-1);
my $gc_percent= sprintf "%.2f", ($c_count + $g_count)/$total_bases;
my $cpg= $sequence =~s/CG/CG/g;
my $o_e= sprintf "%.2f", $cpg/$expect;
print "$gc_percent\t";
print "$total_bases\t";
print "$cpg\t";
print "$expect\t"; print "$o_e \n";
```

Appendix D
Supplemental Information for Chapter 4: Engineering ES Cells for Drug Selection of a Subset of Neural Cells

D.1 Primer Tables

All locations are relative to Olig2 ATG Chr. 16 - 91,226,645
(UCSC browser, mm9 July 2007 build)

Table D.1: Olig2 BAC Mapping Primers

Name	Sequence 5'-3'	Location	Amplicon size
Olig2_Bac_MapF1	CCACAGGGGTGGGTTAAAGAGATAGCC	-14,794	
Olig2_Bac_MapR1	GAAATCACTGCAGGAGCAGGAGGAATC	-14,514	306
Olig2_Bac_MapF2	AGTCATAGAGTTGGTGACAACGAGGAC	-12,277	
Olig2_Bac_MapR2	CCTCTTTATTTCCAACGTAGGATCAGG	-11,907	396
Olig2_Bac_MapF3	ACGTGGAAGGCTTATGTTCTTGTGTTA	-7,626	
Olig2_Bac_MapR3	CGGTACATTCCATGAGTACACAGACAG	-7,349	303
Olig2_Bac_MapF4	TACTCCTGAGAATGGGTCCGTGGAAAG	-5,370	
Olig2_Bac_MapR4	CAGTGTGTGCGATGTGGAGGTTTAGGT	-4,932	464
Olig2_Bac_MapF5	AGCTGGAAATGTCCGGATGTGAGAAAC	-2,192	
Olig2_Bac_MapR5	CTCCGCTGTGGATGGGAGTTGATACTT	-1,813	405
Olig2_Bac_MapF6	CGAAAGGTGTGGATGCTTATTACAGAC	-838	
Olig2_Bac_MapR6	TGGACCGGAGATCTGAATAGAGAAGTA	-410	454
Olig2_Bac_MapF7	GTGTTTGATGAGGATTCACCAGTCTCT	2,783	
Olig2_Bac_MapR7	AGTGATCACCCAACATGTCATCTGTTA	3,144	387
Olig2_Bac_MapF8	GGCAAAATCAGCTTAGAACTCTGAACA	4,967	
Olig2_Bac_MapR8	ATGAAGGAGTTACCTGGTTTGGTCAT	5,335	393
Olig2_Bac_MapF9	AGGCCACAAGCAATTTCCCTCTCTTTC	7,014	
Olig2_Bac_MapR9	GTTCTTCAGGATGATGCTGCACTGGTC	7,292	304
Olig2_Bac_MapF10	TCACACTCAAATTCTAGGAAGCCAGAG	12,290	
Olig2_Bac_MapR10	GGCGTTAGCTTTCTATGAACACTCAGA	12,613	349

Table D.2: BAC Subcloning Primers

Name	5' Upstream & Antisense Homology + cgactgaattggttcctttaaagc 3'
Ol2_rec2b	AGCTGGACCAGGATAACACAAGTGTGTCGCAGCA TTTTATTAGCCAAGGA
OL2_rec4b	AGCAGACAAACAGTAATTGGTCTTCTTGTAGCAAA GTCAGAACAGGCATG
OL2_rec5b	AGAGGCAACATTATTTGCAAAGAATTATCGGAGGCT CTCCGCTGTGGATG
	5' Downstream & Sense Homology + gccgcactcgagatatctagacccag 3'
Ol2_rec6b	ACTCTGAAGCGGTGCTACTGTCCCATGTTCAGGCCA GAGATCTATGTGAG
Ol2_rec7b	GAAATGATCAGTCAGGACCCAGGCATTTCCTAAAGC TAAAATGGACGATT
Ol2_rec8b	CATACAGCACCTACTCTGTTCATGGACAGACAGTGA CAGCAGTGACAAAG
OL2_rec9b	GTTAACACGAGGGGCAAACCCCTTCATGCGCGCTCC CTGGCGCGACCTGC
OL2_rec10b	TTCAGTTCCAGTCGCCCCTCTGTCACCAATTTCTGAA GTGTCAGCCTGAT
Upper case letters indicate Olig2 sequence. Lower case is appended on the 3' end of primer to amplify pStartK.	

Table D.3: Olig2 Homology Plasmids

Name	Left end	Right End	Total size of plasmid (bp) including pStartK region
pOlig2_1	-8894	2,036	13,676
pOlig2_2	-8894	3,076	14,716
pOlig2_3	-2415	5,600	10,761
pOlig2_4	-2415	8,905	14,066
pOlig2_5	-2415	12,962	18,123
pOlig2_6	-1800	5,600	10,146
pOlig2_7	-1800	8,905	13,451
pOlig2_8	-1800	12,962	17,508

Table D.4: Primers for Insertion of AscI Sites and Chloramphenicol Resistance

Name	5' Homology + GGCGCGCC + agcattacacgtcttgagcgattgt 3'	Relative to Olig2 ATG
Ol2_Asc_CatF	AAGGTTGAAAAAAGAAGGATCATTCGAGAGCTTAGATC ATCCCTGGGGCC	-1
	5' Homology + GGCGCGCC + cacttaacggctgacatgggaatta	
OL2_Asc_CatR	CGCGCTGCTCCGCAGCCCTTGGCTCTCCAGGACGCAC CCCCGCTGGCCGG	973

Table D.5: Primers for Junction PCR to Detect Cassette Cloning

Name	Sequence 5' to 3'
cr 1_29 casetteF	AAGGATCATTCGAGAGCTTAGATCA
Reverse Junction1	GCGCCAGGAGGCCTTCCATCTGTTGCT
neo detect forward1	GCCTTCTATCGCCTTCTTGACGAGTTCTT
cr 1_29 casetteR	AGTGTTCAGCCAAAGAGTCAACCAG

Table D.6: Primers for ES Targeting Detection

Name	Sequence 5' to 3'
	Detection of Targeting of 1PN-TK3
Neo detect F1	GCCTTCTATCGCCTTCTTGACGAGTTCTT
ES detect 3_3	ACACAAACTGAAAACACCTGCCTTGCTTTA
	Detection of Targeting of 8PN-TK3
ES detect 5_ 2	CCCAGATCCAGTTTCTCAGCTTTGTATGTG
Puro_R4	GTGAGGAAGAGTTCTTGCAGCTCGGTGAC

Table D.7: Primers for Southern Probes

Name	Sequence 5' to 3'
	Detection of Targeting of 1PN-TK3
S-3'F	ttaacagatgacatgttgggtgatcactgg
S-3'R	ttaaggacccacaaaagtcaatgtggtctg
	Detection of Targeting of 8PN-TK3
S-5'F	ccaagtgtacagaatgtcaggcagttctcc
S-5'R	catctcgtgccttctgaagcattgttttc

Table D.8: Primers for RT-PCR Detection

Name	Sequence 5' to 3'
PCR_N_F1	GGATGCTTATTACAGACCGAGCCAACACC
PCR_N_R2	ACGAGGACACAGTCCCTCCTGTGAAG
PCR_puroT_R3	GCGTGAGGAAGAGTTCTTGCAGCTC

Table D.9: Sequencing Primers

	Name	Sequence 5' to 3'
Gap Repair	WS275	TAAACTGCCAGGCATCAAACTAAGC
	WS276	AGTCAGCCCCATACGATATAAGTTG
Chloramphenicol Insertion	WS187	ATGCCGCTGGCGATTCAGGTTC
	WS188	GCCGATCAACGTCTCATTTTCG

D.2 Supporting Data for Targeting Vectors

Figure D.1: Mapping Olig2BAC 227 by Short Amplicon PCR. Primers and expected amplicon sized are given in Table D.1.

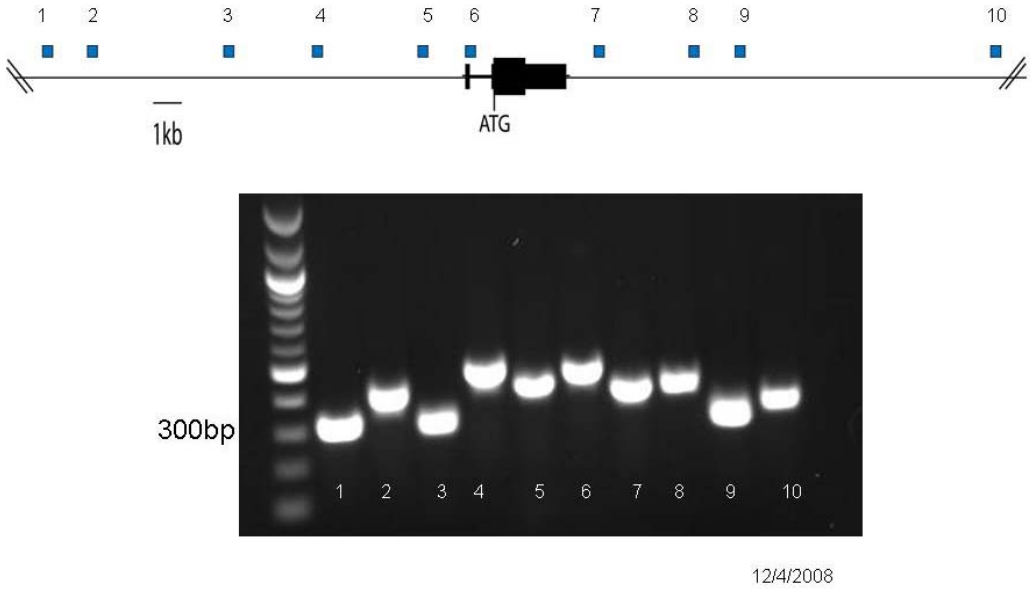
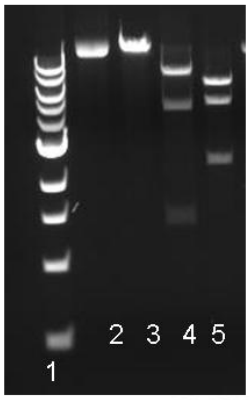
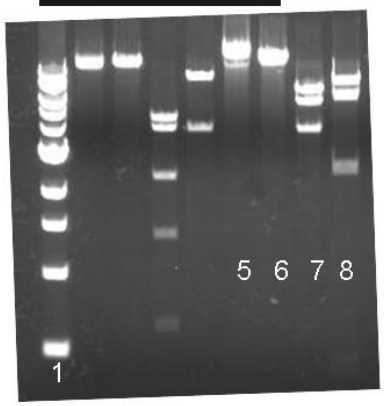


Figure D.2: Restriction Digests of Olig2 Subcloned Plasmids, pOlig2_1 and pOlig2_8

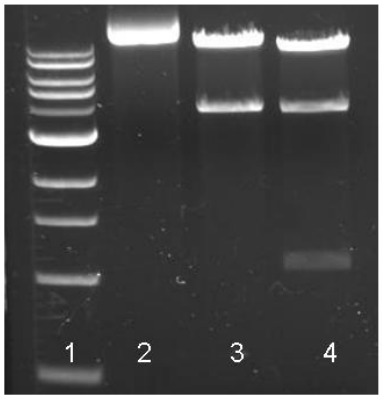


pOlig2_1 digest 12/18/08; 1/6/09
Lane 1: 1kb Ladder
Lane 2: 1-7 uncut – expect 13.6 kb
Lane 3: 1-7 BamHI – expect 13.6kb linearize
Lane 4: 1-7 XhoI – expect 7.7, 4.6,1.4kb
Lane 5: 1-7: HindIII- expect 6.8, 4.8, 2.4

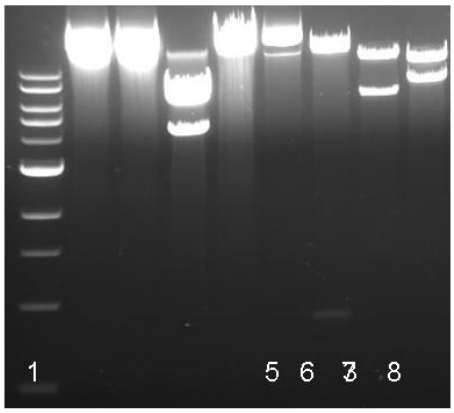


p_Olig2_8 digest 12/15/08; 1/6/09
Lane 1: 1kb Ladder
Lane 5: 8-2 uncut – expect 17.5kb
Lane 6: 8-2 BamHI – expect 17.5kb linearize
Lane 7: 8-2 SacI – expect 7.3, 6.0, 4.1kb
Lane 8: 8-2: XmaI- expect 8.2, 6.3, 2.5, 400bp

Figure D.3: Restriction Digests of AscI Modified Plasmids, pOlig2_1Asc and pOlig2_8Asc.



pOlig2_1Asc digest ; 03/05/09
Lane 1: 1kb Ladder
Lane 2: 1-1 uncut – expect 13.7 kb
Lane 3: 1-7 PvuI/ScaI– expect 9.9, 3.8 kb
Lane 4: 1-7 EcoRI – expect 8.9,3.7,1.1kb



pOlig2_8Asc digest; 01/27/2009
Lane 1: 1kb Ladder
Lane 5: 8-5 uncut – expect 17.5kb
Lane 6: 8-5 AscI – expect 16.5, 866bp
Lane 7: 8-5 SacI – expect 11.6, 6.0kb
Lane 8: 8-5: NdeI- expect 10.3, 7.2kb

Figure D.4: Junction PCR for Detection of PAC-neo Cassette Insertion. Primers are given in Table D.5.

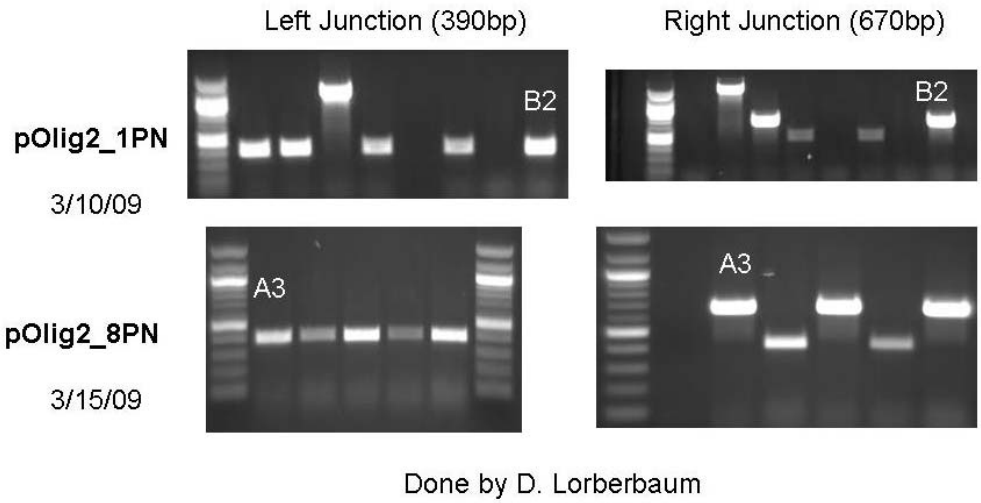
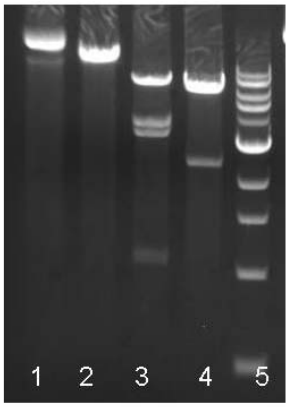
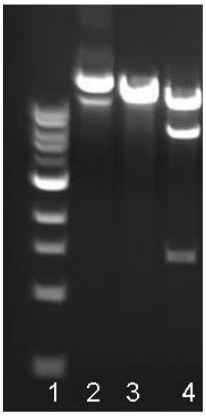


Figure D.5: Restriction Digests of PAC-neo Modified Plasmids, pOlig2_1PN and pOlig2_8PN

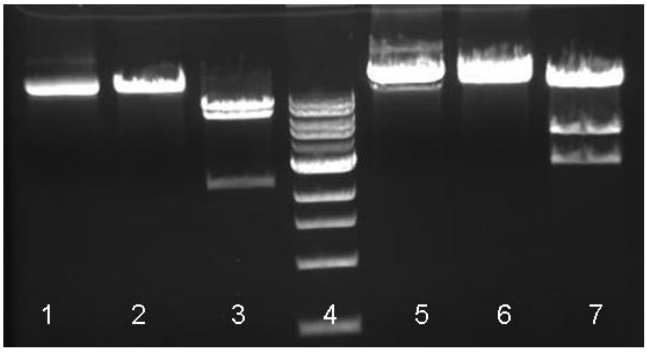


pOlig2_1PN digest; 3/17/2009
Lane 1: 1PN-B2 – uncut expect 15.6kb
Lane 2: 1PN-B2 –BamHI linear expect 15.6kb
Lane 3: 1PN-B2 – EcoRI expect 7.5,3.7,3.2,1
Lane 4: 1PN-B2 HindIII – expect 6.8,6.4, 2.4kb
Lane 5: 1kb Ladder



pOlig2_8PN digest; 3/11/2009 (DL)
Lane 1: 1kb ladder
Lane 2: 8PN-A3 – uncut expect 19.4kb
Lane 3: 8PN-A3 EcoRI –linear expect 19.4kb
Lane 4: 8PN-A3 – XhoI expect 12.1,5.9,1.4 kb

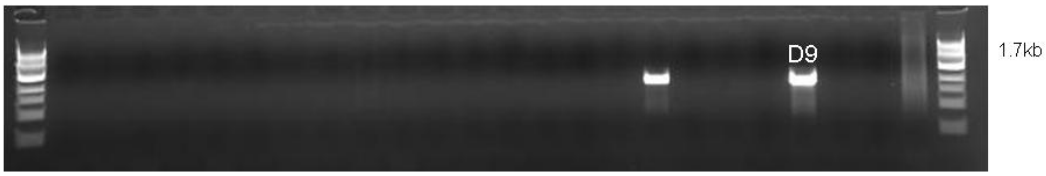
Figure D.6: Restriction Digests of Gateway Modified Plasmids, pOlig2_1PN-TK3 and pOlig2_8PN-TK3



pOlig2_1PN-TK3, pOlig2_8PN-TK3; Digests 4/1/2009
Lane 1: 1PN-TK3 – uncut expect 17.7kb
Lane 2: 1PN-TK3 –ScaI linear expect 17.7kb
Lane 3: 4: 1PN-TK3 HindIII – expect 8.5,6.9, 2.4kb
Lane 4: 1kb Ladder
Lane 5: 8PN-TK3- uncut expect 24.1kb
Lane 6: 8PN-TK3 – ScaI linear expect 24.1kb
Lane 7: 8PN-TK3 – EcoRI expect 14.2, 4.4, 2.8kb

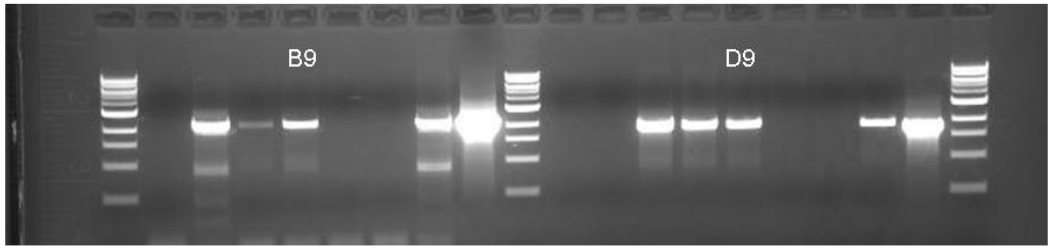
Figure D.7: Targeting Detection PCRs. Primers are given in Table D.6.

EP6: 1PN-TK3 C1-D12. PCR done: CR 4/15/2009



EP7: 8PN-TK3 A1-B12. PCR done: DL 4/16/2009

Validation after thaw. PCR done: DL 4/28/2009



D.3 ES Cell Expression Vector

This vector was designed in consultation with J. Kuhn and D. Lorberbaum. It was built and validated by J. Kuhn and D. Lorberbaum.

Rationale

In chapter 4, we engineered targeting vectors to insert a promoterless puromycin acetyltransferase gene (PAC) into the open reading frame of Olig2, the targeted gene. Once targeted to the Olig2 locus, PAC will be expressed by the Olig2 promoter. In order to incorporate the promoterless PAC into the targeting vector, it must be amplified using PCR to isolate the sequence from its source and incorporate restriction sites that are necessary for cloning into the targeting construct. Loss of function mutations of the PAC due to PCR are a concern. Because Olig2 is not expressed in ES cells, we would not know if PAC was functional or had incorporated a loss of function mutation until the ES cells are differentiated into Olig2 expressing neural cells. Therefore, it is best to avoid this scenario. Typically cDNA cassettes, like PAC, are extensively sequenced once they are cloned into a targeting vector to ensure that no mutations have occurred. This is tedious given that most targeting vectors range from 10-20kb.

It would be better to functionally test the PAC in ES cells before proceeding with the targeting experiment. To do this, we designed and build an ES cell expression vector, where expression of the amplified PAC is governed by a strong ES cell promoter (Chung 2002). This allows the activity of the PAC cDNA to be seen before targeted clones are

detected, expanded, and differentiated. Additionally, PAC can be sequenced once in the ~ 7kb expression vector for further validation. The functional PAC can then be transferred to the targeting vector by cut and paste cloning, eliminating mutagenesis of PAC.

Building the Vector

The cDNA expression vector is comprised of three parts. The first component is the promoter that will be used to drive expression of the inserted cDNA in ES cells. We chose to use cellular polypeptide chain elongation factor 1 alpha (EF-1a) which has been shown to robustly drive gene expression in ES cells (Chung, Andersson et al. 2002). The second part of the expression vector is the cDNA we want to express. We chose to use the puromycin acetyltransferase gene (PAC, (Gomez Lahoz, Lopez de Haro et al. 1991); (Watanabe, Kai et al. 1995)) as our cDNA because we are interested in generating targeted ES cells where Olig2 expressing neural cells can be purified with puromycin. The third component of the expression vector is the positive selection cassette which consists of a floxed PGK-neo. PGK-neo is a standard positive selection cassette for gene targeting in ES cells. It is flanked by loxP sites, "floxed", so that it can be excised from the ES cell genome once targeted clones have been obtained.

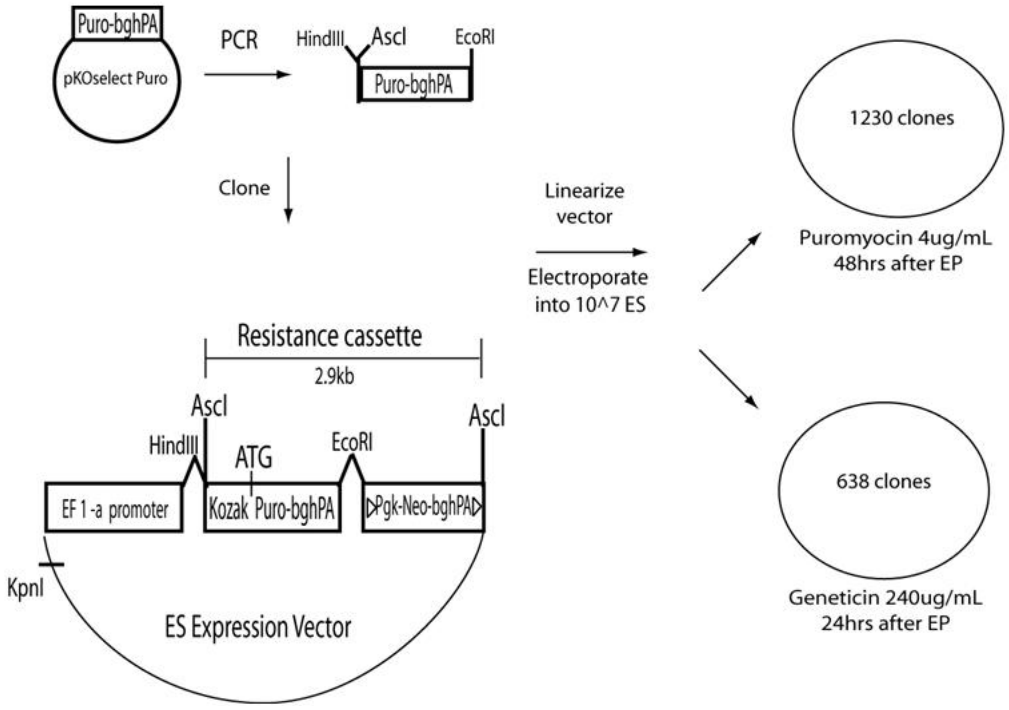
A map of the complete puromyocin-ES expression vector (pBSKANPe) is shown in Figure D.8. All primers used for construction are listed in Table D.10. Four steps were performed to assemble the expression vector. First, pBluescriptSKII(+) (Stratagene) was modified to contain an AscI site by inverse PCR with primers tailed with AscI sites. The linear PCR fragment was digested with AscI and ligated to form pBSAscI. Next, a floxed

PGK-neo cassette with bgh Poly(A+) was cloned from pL452 (NCI Recombineering Website) from EcoRI to NotI. Then, the puromyocin cDNA cassette containing bgh Poly(A) was amplified from pKO SelectPuro (Stratagene) by PCR and cloned into the expression vector from HindIII to EcoRI sites. The 5' puromyocin primer contained a HindIII and an AscI site which will be used for later cloning into the targeting vector. The first 6 bases of the Kozak sequence were also added in the 5' primer to facilitate translation upstream of the puromyocin ATG. The 3' primer contained an EcoRI site for cloning into the expression vector. Finally, the EF-1α promoter element was cloned into pBSAscI by 5' SalI and 3' HindIII restriction sites added during PCR amplification of EF-1α from pTracer – EF/BsdA (Invitrogen) completing the expression vector.

Table D.10: Primers for ES cDNA Expression Vector

Name	Sequence 5'-3'
Inverse AscI fit F	gatc**GGCGGCC**GGAGCTCCAGCTTTTGTTCCCTTTAGTGAGGGT
Inverse AscI fit R	gatc**GGCGCGCC**CCGCGGTGGCGGCCGCTCTAGAACTAGTGGAT
EF-1α F	ttat**GTCGAC**GTGAGGCTCCGGTGCCCGTCAGTGG
EF-1α R	cggc**AAGCTT**CAAGCTAATTCCTCACGACACCTGAAATGG
Puro F	tgcaa**AAGCTTGGCGCGCC**GCCAATATGACCGAGTACAAGCCCACGGTGC
Puro R	cgtaat**GAATTC**CCCAGCTGGTTCTTTCCGCCTCAGAAG
Bold or underlined sequences	SalI = **GTCGAC**; HindIII = **AAGCTT**; Kozak=GCCAAT; AscI = **GGCGCGCC**; EcoRI =**GAATTC**

Figure D.8: ES Cell Expression Vector Tests Amplified Puromycin Resistance Gene



Testing the Vector

As shown in Figure D.8, the completed PBSKANPe vector was linearized with KpnI electroporated into 10^7ES cells. The electroporated cells were divided evenly among two 100mm dishes. One dish was treated with 240μg/mL geneticin 24 hours after electroporation and the other was treated with 4μg/mL puromycin 48 hours following electroporation. Clones were counted seven days following selection. There were abundant puromycin resistant clones ~1200 indicating that the amplified puromycin resistance gene was indeed functional. In addition, approximately ~600 neomycin resistant were counted clones indicating that the PGK-neo cassette was also functional. The quantitative differences in the number of clones may be due uneven splitting of electroporated cells, or

it could reflect the relative activity of the EF-1a promoter compared to the PGK promoter in ES cells.

Conclusion

Using an ES cell expression vector, we show that a simple ES cell electroporation experiment is sufficient to ensure that the amplified promoterless puromycin acetyltransferase cDNA is functional. This functional PAC can be directly transferred to targeting vectors without the concern of mutagenesis. In addition, the vector contains a functional floxed positive selection PGK-neo cassette downstream of the promoterless cDNA which can be used for positive selection in ES cells during gene targeting. Once targeting events have been detected, the PGK-neo cassette then is able to be removed by Cre recombinase. The expression vector it is well suited to our targeting vector construction strategy (Wu, Ying et al. 2008) and will aid in future targeting experiments. In conclusion, we have designed and built an ES cell expression vector that allows us to test functionality of any amplified cDNA in ES cells before targeting it to our gene of interest.

References

Abranches, E., M. Silva, et al. (2009). "Neural differentiation of embryonic stem cells in vitro: a road map to neurogenesis in the embryo." PLoS One **4**(7): e6286.

Antequera, F. and A. Bird (1993). "Number of CpG islands and genes in human and mouse." Proc Natl Acad Sci U S A **90**(24): 11995-9.

Antequera, F., J. Boyes, et al. (1990). "High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines." Cell **62**(3): 503-14.

Arber, S., B. Han, et al. (1999). "Requirement for the homeobox gene Hb9 in the consolidation of motor neuron identity." Neuron **23**(4): 659-74.

Bain, G., D. Kitchens, et al. (1995). "Embryonic stem cells express neuronal properties in vitro." Dev Biol **168**(2): 342-57.

Bain, G., Yao,M., Huettner, J.E., Finley,M., & Gottlieb, D.I. (1998). Neuronlike cells derived in culture from P19 embryonal carcinoma and embryonic stem cells. Culturing Nerve Cells. Cambridge Mass, MIT Press.

Baron, U., I. Turbachova, et al. (2006). "DNA methylation analysis as a tool for cell typing." Epigenetics **1**(1): 55-60.

Bengtsson, M., A. Stahlberg, et al. (2005). "Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels." Genome Res **15**(10): 1388-92.

Bibel, M., J. Richter, et al. (2004). "Differentiation of mouse embryonic stem cells into a defined neuronal lineage." Nat Neurosci **7**(9): 1003-9.

Billon, N., C. Jolicoeur, et al. (2002). "Normal timing of oligodendrocyte development from genetically engineered, lineage-selectable mouse ES cells." J Cell Sci **115**(Pt 18): 3657-65.

Bird, A., M. Taggart, et al. (1985). "A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA." Cell **40**(1): 91-9.

Bird, A. P. (1978). "Use of restriction enzymes to study eukaryotic DNA methylation: II. The symmetry of methylated sites supports semi-conservative copying of the methylation pattern." J Mol Biol **118**(1): 49-60.

Bird, A. P. (1980). "DNA methylation and the frequency of CpG in animal DNA." Nucleic Acids Res **8**(7): 1499-504.

Boyer, L. A., K. Plath, et al. (2006). "Polycomb complexes repress developmental regulators in murine embryonic stem cells." Nature **441**(7091): 349-53.

Boyes, J. and A. Bird (1991). "DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein." Cell **64**(6): 1123-34.

Brustle, O., K. N. Jones, et al. (1999). "Embryonic stem cell-derived glial precursors: a source of myelinating transplants." Science **285**(5428): 754-6.

Capecchi, M. R. (1989). "Altering the genome by homologous recombination." Science **244**(4910): 1288-92.

Carayannopoulos, M. O., A. Schlein, et al. (2004). "GLUT9 is differentially expressed and targeted in the preimplantation embryo." Endocrinology **145**(3): 1435-43.
Carr, I. M., E. M. Valleley, et al. (2007). "Sequence analysis and editing for bisulphite genomic sequencing projects." Nucleic Acids Res **35**(10): e79.
Chahrour, M., S. Y. Jung, et al. (2008). "MeCP2, a key contributor to neurological disease, activates and represses transcription." Science **320**(5880): 1224-9.
Chetverina, H. V., T. R. Samatov, et al. (2002). "Molecular colony diagnostics: detection and quantitation of viral nucleic acids by in-gel PCR." Biotechniques **33**(1): 150-2, 154, 156.
Chiang, M. K. and D. A. Melton (2003). "Single-cell transcript analysis of pancreas development." Dev Cell **4**(3): 383-93.
Chisholm, J. C., M. H. Johnson, et al. (1985). "Developmental variability within and between mouse expanding blastocysts and their ICMs." J Embryol Exp Morphol **86**: 311-336.
Chung, S., T. Andersson, et al. (2002). "Analysis of different promoter systems for efficient transgene expression in mouse embryonic stem cell lines." Stem Cells **20**(2): 139-45.
Clark, S. J., J. Harrison, et al. (1994). "High sensitivity mapping of methylated cytosines." Nucleic Acids Res **22**(15): 2990-7.
Cocquet, J., A. Chong, et al. (2006). "Reverse transcriptase template switching and false alternative transcripts." Genomics **88**(1): 127-31.
Corti, S., M. Nizzardo, et al. (2009). "Motoneuron transplantation rescues the phenotype of SMARD1 (spinal muscular atrophy with respiratory distress type 1)." J Neurosci **29**(38): 11761-71.
Coutts, M. and H. S. Keirstead (2008). "Stem cells for the treatment of spinal cord injury." Exp Neurol **209**(2): 368-77.
Deb, K., M. Sivaguru, et al. (2006). "Cdx2 gene expression and trophectoderm lineage specification in mouse embryos." Science **311**(5763): 992-6.
Deshpande, D. M., Y. S. Kim, et al. (2006). "Recovery from paralysis in adult rats using embryonic stem cells." Ann Neurol **60**(1): 32-44.
Di Giorgio, F. P., M. A. Carrasco, et al. (2007). "Non-cell autonomous effect of glia on motor neurons in an embryonic stem cell-based ALS model." Nat Neurosci **10**(5): 608-14.
Dressman, D., H. Yan, et al. (2003). "Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations." Proc Natl Acad Sci U S A **100**(15): 8817-22.
Dufva, M., A. Svenningsson, et al. (1995). "Differential regulation of macrophage scavenger receptor isoforms: mRNA quantification using the polymerase chain reaction." J Lipid Res **36**(11): 2282-90.
Eckhardt, F., J. Lewin, et al. (2006). "DNA methylation profiling of human chromosomes 6, 20 and 22." Nat Genet **38**(12): 1378-85.
Ehrlich, M. (2003). "Expression of various genes is controlled by DNA methylation during mammalian development." J Cell Biochem **88**(5): 899-910.

Ehrlich, M., M. A. Gama-Sosa, et al. (1982). "Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells
DNA is digested into mononucleotide triphosphates by DNAse1 and P1 nuclease." Nucleic Acids Res **10**(8): 2709-21.
Evans, M. J. and M. H. Kaufman (1981). "Establishment in culture of pluripotential cells from mouse embryos." Nature **292**(5819): 154-6.
Fehling, H. J., G. Lacaud, et al. (2003). "Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation." Development **130**(17): 4217-27.
Fouse, S. D., Y. Shen, et al. (2008). "Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation." Cell Stem Cell **2**(2): 160-9.
Frommer, M., L. E. McDonald, et al. (1992). "A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands." Proc Natl Acad Sci U S A **89**(5): 1827-31.
Gardiner-Garden, M. and M. Frommer (1987). "CpG islands in vertebrate genomes." J Mol Biol **196**(2): 261-82.
Gaspard, N., T. Bouschet, et al. (2008). "An intrinsic mechanism of corticogenesis from embryonic stem cells." Nature **455**(7211): 351-7.
Gidekel, S. and Y. Bergman (2002). "A unique developmental pattern of Oct-3/4 DNA methylation is controlled by a cis-demodification element." J Biol Chem **277**(37): 34521-30.
Gomez Lahoz, E., M. S. Lopez de Haro, et al. (1991). "Use of puromycin N-acetyltransferase (PAC) as a new reporter gene in transgenic animals." Nucleic Acids Res **19**(12): 3465.
Gubler, U. (1987). "Second-strand cDNA synthesis: mRNA fragments as primers." Methods Enzymol **152**: 330-5.
Hansen, R. S. and S. M. Gartler (1990). "5-Azacytidine-induced reactivation of the human X chromosome-linked PGK1 gene is associated with a large region of cytosine demethylation in the 5' CpG island." Proc Natl Acad Sci U S A **87**(11): 4174-8.
Harper, J. M., C. Krishnan, et al. (2004). "Axonal growth of embryonic stem cell-derived motoneurons in vitro and in motoneuron-injured adult rats." Proc Natl Acad Sci U S A **101**(18): 7123-8.
Hartmann, C. H. and C. A. Klein (2006). "Gene expression profiling of single cells on large-scale oligonucleotide arrays." Nucleic Acids Res **34**(21): e143.
Hattori, N., K. Nishino, et al. (2004). "Epigenetic control of mouse Oct-4 gene expression in embryonic stem cells and trophoblast stem cells." J Biol Chem **279**(17): 17063-9.
Heilig, C. W., T. Saunders, et al. (2003). "Glucose transporter-1-deficient mice exhibit impaired development and deformities that are similar to diabetic embryopathy." Proc Natl Acad Sci U S A **100**(26): 15613-8.
Hochedlinger, K. and K. Plath (2009). "Epigenetic reprogramming and induced pluripotency." Development **136**(4): 509-23.

Hoelzer, K., L. A. Shackelton, et al. (2008). "Presence and role of cytosine methylation in DNA viruses of animals." Nucleic Acids Res **36**(9): 2825-37.
Hu, B. Y., Z. W. Du, et al. (2009). "Differentiation of human oligodendrocytes from pluripotent stem cells." Nat Protoc **4**(11): 1614-22.
Hu, B. Y. and S. C. Zhang (2009). "Differentiation of spinal motor neurons from pluripotent human stem cells." Nat Protoc **4**(9): 1295-304.
Illingworth, R., A. Kerr, et al. (2008). "A novel CpG island set identifies tissue-specific methylation at developmental gene loci." PLoS Biol **6**(1): e22.
Jabbari, K. and G. Bernardi (2004). "Cytosine methylation and CpG, TpG (CpA) and TpA frequencies." Gene **333**: 143-9.
Jahner, D., H. Stuhlmann, et al. (1982). "De novo methylation and expression of retroviral genomes during mouse embryogenesis." Nature **298**(5875): 623-8.
Johnson, M. H. and J. M. McConnell (2004). "Lineage allocation and cell polarity during mouse embryogenesis." Semin Cell Dev Biol **15**(5): 583-97.
Johnson, P. J. (2009). Embryonic Stem Cell Derived Neural Progenitor Cells Transplanted in Tissue Engineered Fibrin Scaffolds to Treat Spinal Cord Injury. Biomedical Engineering. Saint Louis, Washington University in St. Louis. **Ph.D.**
Johnson, P. J., S. R. Parker, et al. (2009). "Fibrin-based tissue engineering scaffolds enhance neural fiber sprouting and delay the accumulation of reactive astrocytes at the lesion in a subacute model of spinal cord injury." J Biomed Mater Res A **92**(1): 152-63.
Johnson, P. J., A. Tatara, et al. (2009). "Controlled release of neurotrophin-3 and platelet derived growth factor from fibrin scaffolds containing neural progenitor cells enhances survival and differentiation into neurons in a subacute model of SCI." Cell Transplant.
Kang, S. M., M. S. Cho, et al. (2007). "Efficient induction of oligodendrocytes from human embryonic stem cells." Stem Cells **25**(2): 419-24.
Kawasaki, H., K. Mizuseki, et al. (2000). "Induction of midbrain dopaminergic neurons from ES cells by stromal cell-derived inducing activity." Neuron **28**(1): 31-40.
Kiefer, J. C. (2007). "Back to basics: Sox genes." Dev Dyn **236**(8): 2356-66.
Kim, J. B., G. J. Porreca, et al. (2007). "Polony multiplex analysis of gene expression (PMAGE) in mouse hypertrophic cardiomyopathy." Science **316**(5830): 1481-4.
Kim, S. and H. A. von Recum (2009). "Endothelial Progenitor Populations in Differentiating Embryonic Stem Cells II: Drug Selection and Functional Characterization." Tissue Eng Part A.
Klug, M. G., M. H. Soonpaa, et al. (1996). "Genetically selected cardiomyocytes from differentiating embronic stem cells form stable intracardiac grafts." J Clin Invest **98**(1): 216-24.
Korshunova, Y., R. K. Maloney, et al. (2008). "Massively parallel bisulphite pyrosequencing reveals the molecular complexity of breast cancer-associated cytosine-methylation patterns obtained from tissue and serum DNA." Genome Res **18**(1): 19-29.

Kozak, M. (1986). "Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes." Cell **44**(2): 283-92.
Kroeger, H., J. Jelinek, et al. (2008). "Aberrant CpG island methylation in acute myeloid leukemia is accentuated at relapse." Blood.
Lamba, D. A., M. O. Karl, et al. (2006). "Efficient generation of retinal progenitor cells from human embryonic stem cells." Proc Natl Acad Sci U S A **103**(34): 12769-74.
Langston, L. D. and L. S. Symington (2004). "Gene targeting in yeast is initiated by two independent strand invasions." Proc Natl Acad Sci U S A **101**(43): 15392-7.
Lee, T. I., R. G. Jenner, et al. (2006). "Control of developmental regulators by Polycomb in human embryonic stem cells." Cell **125**(2): 301-13.
Leung, W., A. Malkova, et al. (1997). "Gene targeting by linear duplex DNA frequently occurs by assimilation of a single strand that is subject to preferential mismatch correction." Proc Natl Acad Sci U S A **94**(13): 6851-6.
Levenberg, S., J. A. Burdick, et al. (2005). "Neurotrophin-induced differentiation of human embryonic stem cells on three-dimensional polymeric scaffolds." Tissue Eng **11**(3-4): 506-12.
Li, E., C. Beard, et al. (1993). "Role for DNA methylation in genomic imprinting." Nature **366**(6453): 362-5.
Li, J. Y., M. T. Pu, et al. (2007). "Synergistic function of DNA methyltransferases Dnmt3a and Dnmt3b in the methylation of Oct4 and Nanog." Mol Cell Biol **27**(24): 8748-59.
Li, L. C. and R. Dahiya (2002). "MethPrimer: designing primers for methylation PCRs." Bioinformatics **18**(11): 1427-31.
Li, M., L. Pevny, et al. (1998). "Generation of purified neural precursors from embryonic stem cells by lineage selection." Curr Biol **8**(17): 971-4.
Ligon, K. L., S. P. Fancy, et al. (2006). "Olig gene function in CNS development and disease." Glia **54**(1): 1-10.
Lu, Q. R., T. Sun, et al. (2002). "Common developmental requirement for Olig function indicates a motor neuron/oligodendrocyte connection." Cell **109**(1): 75-86.
Lu, Q. R., D. Yuk, et al. (2000). "Sonic hedgehog--regulated oligodendrocyte lineage genes encoding bHLH proteins in the mammalian central nervous system." Neuron **25**(2): 317-29.
Lugus, J. J., Y. S. Chung, et al. (2007). "GATA2 functions at multiple steps in hemangioblast development and differentiation." Development **134**(2): 393-405.
Ma, H. and S. Difazio (2008). "An efficient method for purification of PCR products for sequencing." Biotechniques **44**(7): 921-3.
Mahoney, M. J. and K. S. Anseth (2007). "Contrasting effects of collagen and bFGF-2 on neural cell function in degradable synthetic PEG hydrogels." J Biomed Mater Res A **81**(2): 269-78.
Martin, G. R. (1981). "Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells." Proc Natl Acad Sci U S A **78**(12): 7634-8.

Martinowich, K., D. Hattori, et al. (2003). "DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation." Science **302**(5646): 890-3.

Masahira, N., H. Takebayashi, et al. (2006). "Olig2-positive progenitors in the embryonic spinal cord give rise not only to motoneurons and oligodendrocytes, but also to a subset of astrocytes and ependymal cells." Dev Biol **293**(2): 358-69.

McBurney, M. W., L. C. Sutherland, et al. (1991). "The mouse Pgk-1 gene promoter contains an upstream activator sequence." Nucleic Acids Res **19**(20): 5755-61.

McDonald, J. W., X. Z. Liu, et al. (1999). "Transplanted embryonic stem cells survive, differentiate and promote recovery in injured rat spinal cord." Nat Med **5**(12): 1410-2.

Meissner, A., T. S. Mikkelsen, et al. (2008). "Genome-scale DNA methylation maps of pluripotent and differentiated cells." Nature.

Meissner, A., M. Wernig, et al. (2007). "Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells." Nat Biotechnol **25**(10): 1177-81.

Mikkilineni, V., R. D. Mitra, et al. (2004). "Digital quantitative measurements of gene expression." Biotechnol Bioeng **86**(2): 117-24.

Mitra, R. D., V. L. Butty, et al. (2003). "Digital genotyping and haplotyping with polymerase colonies." Proc Natl Acad Sci U S A **100**(10): 5926-31.

Mitra, R. D. and G. M. Church (1999). "In situ localized amplification and contact replication of many individual DNA molecules." Nucleic Acids Res **27**(24): e34.

Mitra, R. D., J. Shendure, et al. (2003). "Fluorescent in situ sequencing on polymerase colonies." Anal Biochem **320**(1): 55-65.

Mitsui, K., Y. Tokuzawa, et al. (2003). "The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells." Cell **113**(5): 631-42.

Moley, K. H., M. M. Chi, et al. (1998). "Maternal hyperglycemia alters glucose transport and utilization in mouse preimplantation embryos." Am J Physiol **275**(1 Pt 1): E38-47.

Mukouyama, Y. S., B. Deneen, et al. (2006). "Olig2+ neuroepithelial motoneuron progenitors are not multipotent stem cells in vivo." Proc Natl Acad Sci U S A **103**(5): 1551-6.

Murry, C. E. and G. Keller (2008). "Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development." Cell **132**(4): 661-80.

Nayak, M. S., Y. S. Kim, et al. (2006). "Cellular therapies in motor neuron diseases." Biochim Biophys Acta **1762**(11-12): 1128-38.

Nistor, G. I., M. O. Totoiu, et al. (2005). "Human embryonic stem cells differentiate into oligodendrocytes in high purity and myelinate after spinal cord transplantation." Glia **49**(3): 385-96.

Nussbaum, J., E. Minami, et al. (2007). "Transplantation of undifferentiated murine embryonic stem cells in the heart: teratoma formation and immune response." FASEB J **21**(7): 1345-57.

Okano, M., D. W. Bell, et al. (1999). "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development." Cell **99**(3): 247-57.

Palsson, B. and S. Bhatia (2004). Tissue engineering. Upper Saddle River, N.J., Pearson Prentice Hall.

Panning, B. and R. Jaenisch (1996). "DNA hypomethylation can activate Xist expression and silence X-linked genes." Genes Dev **10**(16): 1991-2002.

Pesce, M. and H. R. Scholer (2001). "Oct-4: gatekeeper in the beginnings of mammalian development." Stem Cells **19**(4): 271-8.

Pevny, L. H., S. Sockanathan, et al. (1998). "A role for SOX1 in neural determination." Development **125**(10): 1967-78.

Pham, C. T., D. M. MacIvor, et al. (1996). "Long-range disruption of gene expression by a selectable marker cassette." Proc Natl Acad Sci U S A **93**(23): 13090-5.

Rieger, C., R. Poppino, et al. (2007). "Polony analysis of gene expression in ES cells and blastocysts." Nucleic Acids Res **35**(22): e151.

Riley, J. K., J. M. Heeley, et al. (2004). "TRAIL and KILLER are expressed and induce apoptosis in the murine preimplantation embryo." Biol Reprod **71**(3): 871-7.

Rogers, M. B., B. A. Hosler, et al. (1991). "Specific expression of a retinoic acid-regulated, zinc-finger gene, Rex-1, in preimplantation embryos, trophoblast and spermatocytes." Development **113**(3): 815-24.

Ronaghi, M., M. Uhlen, et al. (1998). "A sequencing method based on real-time pyrophosphate." Science **281**(5375): 363, 365.

Rossant, J. (2007). "Stem cells and lineage development in the mammalian blastocyst." Reprod Fertil Dev **19**(1): 111-8.

Rowitch, D. H., Q. R. Lu, et al. (2002). "An 'oligarchy' rules neural development." Trends Neurosci **25**(8): 417-22.

Roy, N. S., C. Cleren, et al. (2006). "Functional engraftment of human ES cell-derived dopaminergic neurons enriched by coculture with telomerase-immortalized midbrain astrocytes." Nat Med **12**(11): 1259-68.

Rubin, L. L. (2008). "Stem cells and drug discovery: the beginning of a new era?" Cell **132**(4): 549-52.

Salero, E. and M. E. Hatten (2007). "Differentiation of ES cells into cerebellar neurons." Proc Natl Acad Sci U S A **104**(8): 2997-3002.

Sambrook, J. and D. W. Russel (2001). Molecular Cloning: A Laboratory Manual. New York, Cold Spring Harbor Laboratory Press.

Sauer, B. and N. Henderson (1988). "Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1." Proc Natl Acad Sci U S A **85**(14): 5166-70.

Scacheri, P. C., J. S. Crabtree, et al. (2001). "Bidirectional transcriptional activity of PGK-neomycin and unexpected embryonic lethality in heterozygote chimeric knockout mice." Genesis **30**(4): 259-63.

Schweitzer, C. M., C. E. van der Schoot, et al. (1995). "Isolation and culture of human bone marrow endothelial cells." Exp Hematol **23**(1): 41-8.
Shen, L., Y. Kondo, et al. (2007). "Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters." PLoS Genet **3**(10): 2023-36.
Shi, Y., C. Desponts, et al. (2008). "Induction of pluripotent stem cells from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds." Cell Stem Cell **3**(5): 568-74.
Shin, S., H. Xue, et al. (2007). "Stage-dependent Olig2 expression in motor neurons and oligodendrocytes differentiated from embryonic stem cells." Stem Cells Dev **16**(1): 131-41.
Sikorska, M., J. K. Sandhu, et al. (2008). "Epigenetic modifications of SOX2 enhancers, SRR1 and SRR2, correlate with in vitro neural differentiation." J Neurosci Res **86**(8): 1680-93.
Silani, V., L. Cova, et al. (2004). "Stem-cell therapy for amyotrophic lateral sclerosis." Lancet **364**(9429): 200-2.
Sinha, S. and J. K. Chen (2006). "Purmorphamine activates the Hedgehog pathway by targeting Smoothened." Nat Chem Biol **2**(1): 29-30.
Smithies, O., R. G. Gregg, et al. (1985). "Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination." Nature **317**(6034): 230-4.
Staden, R., K. F. Beal, et al. (2000). "The Staden package, 1998." Methods Mol Biol **132**: 115-30.
Strumpf, D., C. A. Mao, et al. (2005). "Cdx2 required for trophectoderm. Oct4 restricted to ICM." Development **132**(9): 2093-102.
Suzuki, M. M. and A. Bird (2008). "DNA methylation landscapes: provocative insights from epigenomics." Nat Rev Genet **9**(6): 465-76.
Takahashi, K., K. Tanabe, et al. (2007). "Induction of pluripotent stem cells from adult human fibroblasts by defined factors." Cell **131**(5): 861-72.
Takahashi, K. and S. Yamanaka (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors." Cell **126**(4): 663-76.
Takebayashi, H., Y. Nabeshima, et al. (2002). "The basic helix-loop-helix factor olig2 is essential for the development of motoneuron and oligodendrocyte lineages." Curr Biol **12**(13): 1157-63.
Takebayashi, H., S. Yoshida, et al. (2000). "Dynamic expression of basic helix-loop-helix Olig family members: implication of Olig2 in neuron and oligodendrocyte differentiation and identification of a new member, Olig3." Mech Dev **99**(1-2): 143-8.
Takizawa, T., K. Nakashima, et al. (2001). "DNA methylation is a critical cell-intrinsic determinant of astrocyte differentiation in the fetal brain." Dev Cell **1**(6): 749-58.
Tang, F., P. Hajkova, et al. (2006). "MicroRNA expression profiling of single whole embryonic stem cells." Nucleic Acids Res **34**(2): e9.

Taylor, K. H., R. S. Kramer, et al. (2007). "Ultradeep bisulfite sequencing analysis of DNA methylation patterns in multiple gene promoters by 454 sequencing." Cancer Res **67**(18): 8511-8.

te Riele, H., E. R. Maandag, et al. (1992). "Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs." Proc Natl Acad Sci U S A **89**(11): 5128-32.

Thaler, J., K. Harrison, et al. (1999). "Active suppression of interneuron programs within developing motor neurons revealed by analysis of homeodomain factor HB9." Neuron **23**(4): 675-87.

Thomas, K. R. and M. R. Capecchi (1987). "Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells." Cell **51**(3): 503-12.

Thompson, J. D., D. G. Higgins, et al. (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Res **22**(22): 4673-80.

Thomson, J. A., J. Itskovitz-Eldor, et al. (1998). "Embryonic stem cell lines derived from human blastocysts." Science **282**(5391): 1145-7.

Tietjen, I., J. M. Rihel, et al. (2003). "Single-cell transcriptional analysis of neuronal progenitors." Neuron **38**(2): 161-75.

Valenzuela, D. M., A. J. Murphy, et al. (2003). "High-throughput engineering of the mouse genome coupled with high-resolution expression analysis." Nat Biotechnol **21**(6): 652-9.

Vallier, L., M. Alexander, et al. (2005). "Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells." J Cell Sci **118**(Pt 19): 4495-509.

Vogelstein, B. and K. W. Kinzler (1999). "Digital PCR." Proc Natl Acad Sci U S A **96**(16): 9236-41.

Walsh, C. P., J. R. Chaillet, et al. (1998). "Transcription of IAP endogenous retroviruses is constrained by cytosine methylation." Nat Genet **20**(2): 116-7.

Warnecke, P. M., C. Stirzaker, et al. (2002). "Identification and resolution of artifacts in bisulfite sequencing." Methods **27**(2): 101-7.

Warren, L., D. Bryder, et al. (2006). "Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR." Proc Natl Acad Sci U S A **103**(47): 17807-12.

Watanabe, S., N. Kai, et al. (1995). "Stable production of mutant mice from double gene converted ES cells with puromycin and neomycin." Biochem Biophys Res Commun **213**(1): 130-7.

Watt, F. and P. L. Molloy (1988). "Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter." Genes Dev **2**(9): 1136-43.

Wernig, M., A. Meissner, et al. (2007). "In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state." Nature **448**(7151): 318-24.

Wernig, M., J. P. Zhao, et al. (2008). "Neurons derived from reprogrammed fibroblasts functionally integrate into the fetal brain and improve symptoms of rats with Parkinson's disease." Proc Natl Acad Sci U S A **105**(15): 5856-61.

Wichterle, H., I. Lieberam, et al. (2002). "Directed differentiation of embryonic stem cells into motor neurons." Cell **110**(3): 385-97.

Willerth, S. M., K. J. Arendas, et al. (2006). "Optimization of fibrin scaffolds for differentiation of murine embryonic stem cells into neural lineage cells." Biomaterials **27**(36): 5990-6003.

Willerth, S. M. and S. E. Sakiyama-Elbert (2008). "Cell therapy for spinal cord regeneration." Adv Drug Deliv Rev **60**(2): 263-76.

Williams, R. L., D. J. Hilton, et al. (1988). "Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells." Nature **336**(6200): 684-7.

Wolf, S. F., D. J. Jolly, et al. (1984). "Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation." Proc Natl Acad Sci U S A **81**(9): 2806-10.

Woltjen, K., I. P. Michael, et al. (2009). "piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells." Nature **458**(7239): 766-70.

Wu, S., Y. Wu, et al. (2006). "Motoneurons and oligodendrocytes are sequentially generated from neural stem cells but do not appear to share common lineage-restricted progenitors in vivo." Development **133**(4): 581-90.

Wu, S., G. Ying, et al. (2008). "A protocol for constructing gene targeting vectors: generating knockout mice for the cadherin family and beyond." Nat Protoc **3**(6): 1056-76.

Xian, H. and D. I. Gottlieb (2004). "Dividing Olig2-expressing progenitor cells derived from ES cells." Glia **47**(1): 88-101.

Xian, H. Q., E. McNichols, et al. (2003). "A subset of ES-cell-derived neural cells marked by gene targeting." Stem Cells **21**(1): 41-9.

Xian, H. Q., K. Werth, et al. (2005). "Promoter analysis in ES cell-derived neural cells." Biochem Biophys Res Commun **327**(1): 155-62.

Xie, J., S. M. Willerth, et al. (2009). "The differentiation of embryonic stem cells seeded on electrospun nanofibers into neural lineages." Biomaterials **30**(3): 354-62.

Xue, H., S. Wu, et al. (2009). "A targeted neuroglial reporter line generated by homologous recombination in human embryonic stem cells." Stem Cells **27**(8): 1836-46.

Yeo, S., S. Jeong, et al. (2007). "Characterization of DNA methylation change in stem cell marker genes during differentiation of human embryonic stem cells." Biochem Biophys Res Commun **359**(3): 536-42.

Ying, Q. L., M. Stavridis, et al. (2003). "Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture." Nat Biotechnol **21**(2): 183-6.

Yu, J., K. Hu, et al. (2009). "Human induced pluripotent stem cells free of vector and transgene sequences." Science **324**(5928): 797-801.

Zandstra, P. W., C. Bauwens, et al. (2003). "Scalable production of embryonic stem cell-derived cardiomyocytes." Tissue Eng **9**(4): 767-78.
Zentilin, L. and M. Giacca (2007). "Competitive PCR for precise nucleic acid quantification." Nat Protoc **2**(9): 2092-104.
Zhang, H. J., M. K. Siu, et al. (2008). "Oct4 is epigenetically regulated by methylation in normal placenta and gestational trophoblastic disease." Placenta **29**(6): 549-54.
Zhang, J., W. L. Tam, et al. (2006). "Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1." Nat Cell Biol **8**(10): 1114-23.
Zhang, K., J. Zhu, et al. (2006). "Long-range polony haplotyping of individual human chromosome molecules." Nat Genet **38**(3): 382-7.
Zhang, X., S. Horrell, et al. (2008). "ES Cells as a Platform for Analyzing Neural Gene Transcription." Stem Cells.
Zhao, X., J. Liu, et al. (2006). "Differentiation of embryonic stem cells to retinal cells in vitro." Methods Mol Biol **330**: 401-16.
Zheng, B., M. Sage, et al. (2000). "Engineering mouse chromosomes with Cre-loxP: range, efficiency, and somatic applications." Mol Cell Biol **20**(2): 648-55.
Zhou, Q. and D. J. Anderson (2002). "The bHLH transcription factors OLIG2 and OLIG1 couple neuronal and glial subtype specification." Cell **109**(1): 61-73.
Zhu, J., J. Shendure, et al. (2003). "Single molecule profiling of alternative pre-mRNA splicing." Science **301**(5634): 836-8.
Zwaka, T. P. and J. A. Thomson (2003). "Homologous recombination in human embryonic stem cells." Nat Biotechnol **21**(3): 319-21.

Vita

Cara R. Rieger

765 Westwood Drive

Apartment 207

Clayton, MO 63105

Email: cara.rieger@gmail.com

Cell phone: (314) 302-8172

EDUCATION

2009 **Ph.D. Biomedical Engineering,** Washington University in St. Louis, St. Louis, MO

2007 **M.S. Biomedical Engineering,** Washington University in St. Louis, St. Louis, MO

2004 **B.S. Bioengineering,** Rice University, Houston, TX

RESEARCH EXPERIENCE

2004-2009 **Ph.D. Thesis. "Expression Analysis and Stem Cell Engineering."**

Advisors: David Gottlieb and Robi Mitra

Engineered ES cells by gene targeting for selection of a subset of neural cells

Measured DNA methylation in cell lines and tissues by direct bisulfite sequencing

Developed a colony PCR (polony) based assay for gene expression analysis

2003-2004 **Undergraduate research.** "**Cell Responsive Polymers for Tissue Engineering**."
Advisor: Antonios Mikos, Directed by: Joerg Tessmar
Modified a biodegradable synthetic peptide for cross-linking into OPF-hydrogels. Studied the effect of cross-linker size, concentration, and degrading enzyme on release.

2003 **Summer research intern**, Northwestern University. Advisor: Robert Linsenmeier. Developed a model of glucose levels in the retina using ordinary differential equations in MATLAB.

TEACHING AND MENTORING EXPERIENCE

2009 **Mentor**, Washington University in St. Louis. Designed and directed stem cell engineering rotation project for first year graduate student.

2008 **Supervisor**, Washington University in St. Louis. Designed and supervised expression vector construction project performed by research technicians.

2006, 2007 **Mentor**, Washington University in St. Louis. Directed and supervised research projects in PCR technologies for high school and undergraduate summer interns.

2005 **Teaching assistant**, Washington University in St. Louis. Obtained background materials and assisted undergraduate students during quantitative physiology laboratory course.

2004 **Summer research intern**, Northwestern University School of Education and Social Policy. Advisor: David Kanter. Provided support for Elementary School Science Teacher Workshop. Supervised an undergraduate intern.

PUBLICATIONS

Rieger C, Poppino R, Sheridan R, Moley K, Mitra R, Gottlieb D. Polony analysis of gene expression in ES cells and blastocysts. Nucleic Acids Res. 2007; 35(22):e151. Epub 2007 Dec 10.

Kanter D, Smith HD, McKenna, A, **Rieger C**, Linsenmeier RA. Inquiry-based Laboratory Instruction Throws Out the "Cookbook" and Improves Learning. Proceedings of the 2003 American Society for Engineering Education Annual Conference and Exposition.

POSTERS AND PRESENTATIONS

Rieger C, Delaney, D., Gottlieb D. Methylation Analysis for Neural Tissue Engineering. Presentation. Biomedical Engineering Department Student Seminar. September 2008.

Rieger C, Poppino R., Sheridan R., Moley K., Mitra R., Gottlieb D. Polony Analysis of Gene Expression in ES cells and Blastocysts. Presentation. Focus Group in Reproduction, Development and Cancer. December 2007.

Rieger CR, Poppino R, Riley JK, Moley KH, Mitra RD, and Gottlieb DI**.** Quantitative Expression Analysis of Stem Cells Using Polonies. Poster. Annual Fall Meeting of the Biomedical Engineering Society. October 2006.

Rieger C, Poppino R, Mitra R, Gottlieb D. Transcription profiling with polonies. Presentation. Biomedical Engineering Department Seminar. April 2006.

INVITED TALKS

Rieger C, Kuhn J, Lorberbaum D, Gottlieb D. Gene Targeting for Stem Cell Selection. Cellular Dynamics International. Madison, WI. June 2009.

Rieger C, Kuhn J, Lorberbaum D, Gottlieb D. Gene Targeting for Stem Cell Selection. Joslin Diabetes Center. Boston, MA. September 2009.