UNIVERSITY OF CALGARY

Identification and Characterization of Murine Se70-2

by

Stephanie E. Minnema

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY
CALGARY, ALBERTA
OCTOBER 2006

UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "Identification and Characterization of Murine Se70-2" submitted by Stephanie E. Minnema in partial fulfilment of the requirements of the degree of Doctor of Philosophy.

Supervisor, Dr. Derrick Rancourt
Department of Biochemistry and Molecular Biology

Dr. Douglas Demetrick
Department of Biochemistry and Molecular Biology

Dr. Stephen Robbins
Department of Biochemistry and Molecular Biology

Dr. Christoph Sensen
Department of Biochemistry and Molecular Biology

Dr. Peter Bridge
Department of Medical Genetics

External Examiner, Dr. Angela Brooks-Wilson
Genome Sciences Centre, BC Cancer Research Centre

Date

Abstract

Differentiation of embryonic stem (ES) cells enables the study of cellular differentiation *in vitro*. With an interest in regulatory genes in neural differentiation, the Rancourt lab established methods for *in vitro* neural differentiation of ES cells, and pursued a screen identifying genes expressed during this process. The Rancourt EST sequence dataset resulted; from this dataset the project described in this thesis arose.

The work presented herein surrounds the identification and characterization of murine Se70-2 (mSe70-2), which is encoded by the gene *Rbm26*, represented in the Rancourt EST dataset as clone 3B6. Work presented includes the development and implementation of a data enrichment, management, and dissemination strategy for the Bain-Rancourt EST dataset: The Rancourt EST Database (RED). Analysis of the dataset by RED led to identification of 3B6, which adheres to criteria suggesting it may encode a nucleic acid binding protein involved in neural differentiation. The genomic structure of *Rbm26* and a number of transcripts from the gene are identified. The analysis of *Rbm26* mRNA expression in differentiating ES cell cultures, embryos, and adult tissues is shown. Expression of *Rbm26* is higher in neural precursor enriched cell cultures than in spontaneously- or non-differentiating ES cells. Expression in embryos is greatest in neural and mesenchymal tissues and the timing of expression corresponds to the peak period of neurogenesis in the developing brain.

Rbm26 encodes the murine homolog of human Se70-2, a putative nucleic acid binding protein. A paralog of mSe70-2 also exists, Psc1, which is a putative SR family splicing factor. Phylogenetic analysis indicates that mSe70-2 is conserved in eukaryotes, with gene duplication in vertebrate lineages. Characterization of mSe70-2 via antibody production and expression of the protein was pursued, but met with challenges. The possibility that mSe70-2 is regulated at the translational level is suggested. Knockdown of its *C. elegans* homolog indicates a possible role in positive regulation of the cell cycle, while the ability to bind RNA is shown using *in vitro* produced mSe70-2.

Collectively, the results of this project suggest a role for mSe70-2 in cell fate decisions associated with differentiation along a subset of possible cell lineages.

Acknowledgements

Derrick Rancourt, for accepting me into your lab and being my academic father. Thank you also for allowing me to pursue my interests, even though they can be all over the place, figuratively and literally! I've learned an immense amount over the past years under your direction, and I know whichever direction I take in my career, I will be stronger with the skills you've given me.

Doug Demetrick, for sitting on my committee. You have always been so approachable, and full of ideas and enthusiasm, even when it seemed the project was heading nowhere. I also want to thank you for the words of advice you gave me in 2005, when I thought my project was over: "fail now or fail later!" You gave me the perspective I needed to keep on going.

Steve Robbins, for sitting on my committee despite your reservations about the project. Thanks you for all of your input into my project, and for helping me make sense of results that just don't make sense.

Christoph Sensen, for sitting on my committee and introducing me to Canadian Bioinformatics Workshops. I am indebted to you for the opportunities you've given me, and for your concern with my progress. I've really appreciated being able to drop in and chat (when you're in town!) about science and life, and steal your never-ending supply of jelly beans. I'll always remember the geese too!

Torben Bech-Hansen, Angela Brooks-Wilson, Peter Bridge, and Jim McGhee, for taking time out of your busy schedules to contribute to my examinations or committee meetings during this process. Your time and interest are greatly appreciated… It also takes a village to raise a scientist!

Rancourt Lab members, past and present. You guys have been my family for the past seven years. I've learned so much with you, not just about science, but also about life and myself. I'll be taking with me so many memories when I go… the questionable lunch conversations, parties, window painting, joking around, Charles and Virginia the travelling mice, fighting over the music, trips to Lick's, old Lloyd the drooling cat and Maggie's weekend visits, creative pumpkin carving…

Mike Wride and Fiona Mansergh, for establishing much of the work my project was built on. I miss you two; the lab just hasn't been the same since you left!

Shi-Ying Liu, for your endless efforts to get the 3B6 *in situs* working. I am genuinely touched by the care you put into helping me; I won't forget your kindness.

Eileen Rattner, for teaching me the finer points of ES cell culture. It's an art form which you have mastered, and I feel lucky to have learned from the best.

Ken Ito, the Zen master of molecular biology, and fellow mountain lover. I've learned so many little things from you… is there anything you don't know?

Jackie Irwin, fellow neat freak and mountain lover, for all of your assistance, great conversation, and organizational skills. It was a dark day when you left our lab; you were the only person who could keep on top of things!

Knut Woltjen. We're the last of the "original students" and it's amazing we never tried to throw ourselves off the building. I could always count on you for an obscure conversation built on song lyrics, and great ideas from your molecular toolkit. Take care of Corkman, I'll miss you both.

Rebecca Everitt. Rebecca, I can't even begin to thank you for everything you have contributed to this project, I could not have done it without you. You have taught me so much, and tolerated so many of my stupid questions. I envy your ability to stay calm even at the most stressful of times! I miss sharing an office with you, and I'll miss the little visits with Alexandra and Trevor.

Beattie Lab members, past and present. Thank you for welcoming me into your space and for the many reagents I borrowed. Your lab is a place of calm between the whirlwinds of the Rancourt labs, and I'm so lucky you guys moved in. I especially want to thank Tara, Brant, Deirdre, Chris, and Nick, for all of their excellent advice and ideas, which helped the progression of my work immensely.

Jennifer Pelley, Sarah Attwell, Deirdre Lobb, and Maggie Renaud-Young, for critical reads of my thesis. You know you've got friends when they agree to read your thesis… I'm really thankful for your input! Ladies, you are fabulous!

The friends I've met throughout the years. A most brilliant bunch, you are too many to name, but you know who you are. Graduate school: it's the best of times; it's the worst of times. I wouldn't have made it through without you! You mean so much to me,

even though some of you have long moved on from Calgary, and some of you are relative newcomers. Remember the times we've had… Fernie, Gary Bucy's, Festivus, MSGSA kegs, martini parties, wine tasting, grad school complain-a-thons, Stampede BBQ, powder days, Robbie Burns, TGIF, Moose's, Survivor watching, shakin' it at the Embassy (or wherever), Halloween, science buddies, and on and on and on. And to all the observers of PL day, remember: we must observe this illustrious time each March until our minds and bodies fail us. And if you ever need anything, I'll be around for you.

Laura Lucier. We've come a long way since grade 5, and your intelligence and independent spirit still impress me. I'm really proud of you. And to think all of this science and engineering stuff started with the heat chamber…

The Schorn Family, for being wonderful friends to me and my family. You really helped ease my transition into living in Calgary. I'll never forget the absolute hilarity of all the stories around the dinner table I enjoyed at your house, the Panorama ski trips, and of course Brian and Rob's skating antics.

The Calgary Minnema Family; My uncle and aunt Peter and Pat, my cousins Chris and Pam, my cousins 'in law' Angie and Mike, and all of my 'little cousins' Caleb, Jacob, Brady, Annie, Ben, and Georgia. Thank you for all of the times we've had together, from playing hallway hockey with Caleb, to the Easter brunches, and welcoming the new additions to the family. You are my reminder that there's more to life than science. For the first time I've felt that I'm really part of an extended Minnema family, and for that I am grateful.

The Calgary Thomas Family; Martin, Nancy, Peter, Libby, and Ian. Thank you for inviting me into your home and life. I enjoy all of the time I spend with you, whether it's out in the mountains, around the table playing cards, chatting in the yard… anything at all! You've really made me feel incredibly welcome, and I feel like you are family to me.

Brad Thomas. I am the luckiest girl to have you; you are my partner and best friend. I can't find the right words to thank you… I am so thankful for all of your love and support in everything. You are my superstar, and knowing you makes my life so much better. XOXOXO.

My brother, Brian Minnema. My friend for life, I really miss having you around to chat with every day. We've had so many great memories together. I guess it's what I like the least about being in Calgary. You are so brilliant… I hope I can achieve a fraction of what you have the potential to do. I also hope life brings you everything you want but don't forget that you can always count on me for anything, no matter what.

My parents, Don and Ann Minnema. You have given me everything and I don't think I could ever start to repay you. You have no idea how much I love you, and how proud I am of you and your lives. Thank you for loving me and supporting me in all of the paths I've chosen, even if you didn't totally agree with some of them. I promise to do my best to make you proud.

When I was little, maybe eight, someone told me that I'd never finish anything. I'm not sorry to disappoint that person again.

Dedication

This thesis is dedicated to my family:

Don, Ann, and Brian Minnema

…for everything.

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Units of Measure

Abbreviations:

αMEM	Minimum essential medium, alpha medium
aa	Amino acids
Amp	Ampicillin
AP	Anterior-posterior
bHLH	Basic helix-loop-helix
BLAST	Basic local alignment search tool
BLASTp	Protein-protein BLAST
BLASTn	Nucleotide-nucleotide BLAST
BLASTx	Translated query BLAST
bl2seq	BLAST 2 sequences
BSA	Bovine serum albumen
Cam	Chloramphenicol
CCCH ZnFn	CCCH Zinc Finger
CNS	Central nervous system
CT antigen	Cancer-testes antigen
CTCL	Cutaneous T-cell lymphoma
ddH2O	Ultra-purified water
DEPC	Diethylpyrocarbonate
DIG	Digoxigenin
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DPBS	Dulbecco's phosphate buffered saline
dpc	Days post-coitus
dpi	Days post-induction
dsRNA	Double Stranded RNA

DTT	Dithiolthreitol
EB	Embryoid body
EC	Embryonal carcinoma
EDTA	Ethylenediaminetetraacetic acid
ES	Embryonic stem
EST	Expressed Sequence Tag
EtBr	Ethidium bromide
EtOH	Ethanol
FACS	Fluorescence activated cell sorting
FBS	Fetal Bovine Serum
gi	GeneInfo identifier
GO	Gene ontology
HMM	Hidden Markov model
HSC	Hemotopoietic stem cell
hTERT	Human telomerase reverse transcriptase
ICD	Intracellular domain
IMAGE	Integrated molecular analysis of genomes and their expression
IP	Immunoprecipitation
IPTG	Isopropyl-beta-D-thiogalactopyranoside
Kan	Kanamycin
KLH	Keyhole limpet hemocyanin
Ku80	Ku antigen, 80 kDa subunit
LB	Luria-Bertani medium
LBM	Luria-Bertani medium with Mg^{++}
LIF	Leukemia inhibitory factor
MEF	Mouse embryonic fibroblast
MGI	Mouse genome informatics
miRNA	Micro RNA
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	Messenger RNA
mRNP	Messenger ribonucleoprotein particle

MSC	Mesenchymal stem cell
mSe70-2	Murine cutaneous T-cell lymphoma tumor antigen Se70-2
NCBI	National Center for Biotechnology Information
ncRNA	Non-coding RNA
NGM	Nematode growth medium
nr	Non-redundant
OD_n	Optical density at wavelength of *n*
pBS	pBluescript II
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
Perl	Practical extraction and report language
pfu	Plaque forming units
Phi-BLAST	Pattern hit initiated-BLAST
PLB	Protein loading buffer
pri-miRNA	Primary micro RNA
Psc1	Peri-implantation stem cell 1
PVP	Polyvinylpyrrolidone
RA	Retinoic acid
RACE	Rapid amplification of cDNA ends
RAR	Retinoic acid receptor
RARE	Retinoic acid response element
Rbm27	RNA binding motif protein 27
Rbm	RNA binding motif
RBP	RNA binding protein
RPB	Random prime buffer
RdRP	RNA dependent RNA polymerase
RED	Rancourt EST Database
RefSeq	Reference sequence
REST	Relative expression software tool
RISC	RNA induced silencing complex
RNA	Ribonucleic acid

RNAi	RNA interference
rNTP	Ribonucleotide triphosphate
rpm	Rotations per minute
RRM	RNA recognition motif
RS domain	Arginine-Serine dipeptide repeat domain
RT	Reverse transcription
RXR	Retinoic X receptor
SDS	Sodium dodecyl sulphate
SDS-PAGE	SDS polyacrylamide electrophoresis
Se70-2	Cutaneous T-cell lymphoma tumor antigen Se70-2
shRNA	Short hairpin RNA
siRNA	Short interfering RNA
SMART	Simple modular architecture research tool
TBS	Tris buffered saline
TnT	*In vitro* transcription and translation
tRNA	Transfer RNA
UTR	Untranslated region

Units of Measure:

cm	Centimeter
cm^2	Square centimeter
g	Gram
mg	Milligram
µg	Microgram
L	Litre
mL	Millilitre
µL	Microlitre
M	Molar
mM	Millimolar
µM	Micrimolar

nM	Nanomolar
pM	Picomolar
Da	Dalton
kDa	Kilodalton
bp	Base pairs
kb	Kilobase pairs
nt	Nucleotide
x g	Times gravity
V	Volts
Ω	Ohm
µF	Micro Farads
Ci	Curie
h	Hours
min	Minutes
°C	Degrees Celsius

Epigraph

"We learn more by looking for the answer to a question and not finding it than we do from finding the answer itself"

-LloydAlexander

Chapter One: Introduction

Stem cells derived from embryonic sources are recognized as having a tremendous therapeutic potential. There is strong impetus for successful use of embryonic stem (ES) cells in regenerative medicine, however when implanted into adult hosts, ES cells commonly contribute the development of teratomas and teratocarcinomas[1]. Because of this oncogenic potential, a more thorough knowledge of directed differentiation, as well as regulatory factors and mechanisms governing differentiation is crucial to bringing ES cells into clinical use.

Treatment of neurological and neurodegenerative disorders has long been perceived as an area that would benefit from ES cell derived therapies. However, the nervous system is incredibly complex, and the existing understanding of mechanisms governing its development and maintenance is still likely to be at an elementary level. Over the past decade though, significant gains have been made in this field with the establishment of methods for differentiating ES cells into neural cell types *in vitro*. These *in vitro* neural differentiation methods have provided unparalleled opportunities for the study of molecular mechanisms underlying neural differentiation, as well as differentiation processes in general. A thorough understanding of neural differentiation *in vitro* will contribute a wealth of knowledge to the understanding of neural development, which in turn will assist in the advancement of regenerative therapies for treatment of neurological and neurodegenerative disorders.

1.1 Stem Cells

The emergence of the ability to isolate and manipulate stem cells is one of the most significant recent advances in the biological sciences. Stem cells, regardless of origin, are defined by several common properties. First, all stem cells are primitive in nature and do not have any specialized functions associated with any tissue or organ systems. Stem cells however, are multipotent. They have the ability to differentiate from their unspecified state along multiple lineages. Finally, stem sells are theoretically able

to limitlessly proliferate, maintaining their own population without any loss of their development potential. Because of this ability to self renew and their multipotency, the potential uses of stem cells in cell based regenerative therapies is seemingly limitless. In addition, stem cells present an outstanding model system for studying cellular renewal, differentiation, and toxicology. Therefore, they are an excellent system for conception and development for non-cell-based therapies such as pharmaceutical compounds.

There are two major generalized sources for stem cells: Embryonic tissues, and post-embryonic tissues. While cell from these sources possesses the defining features of stem cells they exhibit some differences, in particular with regards to their expected potency.

1.1.1 Embryonic stem cells

Embryonic stem (ES) cells are typically isolated from the inner cell mass (ICM) of the blastocyst or blastula [Figure 1, A, B], and can be maintained in culture indefinitely. If reintroduced into the inner cell mass of a developing blastocyst, mouse ES cells have the potential to give rise to any cell type of the embryo, including the germline, and are therefore referred to as being pluripotent. This pluripotent population of cells is thought only to persist until gastrulation, which gives rise to the three germ layers and restricts cells to developmental fates arising from their respective layer with few exceptions[Figure 1, C][2]. Maintenance of mouse ES cell pluripotency in culture requires the addition of leukemia inhibitory factor (LIF) to the culture medium, as in the absence of LIF they are prone to spontaneous differentiation[3,4].

Early work leading to the isolation of ES cell lines involved the dissociation and culturing of mouse teratocarcinomas. A unique subset of cells arising from culture of teratocarcinomas, called embryonal carcinoma (EC) cells, were found to proliferate indefinitely, and have morphological similarities to early mouse embryos when cultured as aggregates in the absence of adhesive substrate[5,6] EC cells were also found to contribute to many cell types of the embryo when introduced into developing blastocysts for chimera development[7]. These observations suggested that EC cells were of an embryonic nature, and so efforts were made to isolate cell lines directly from embryos.

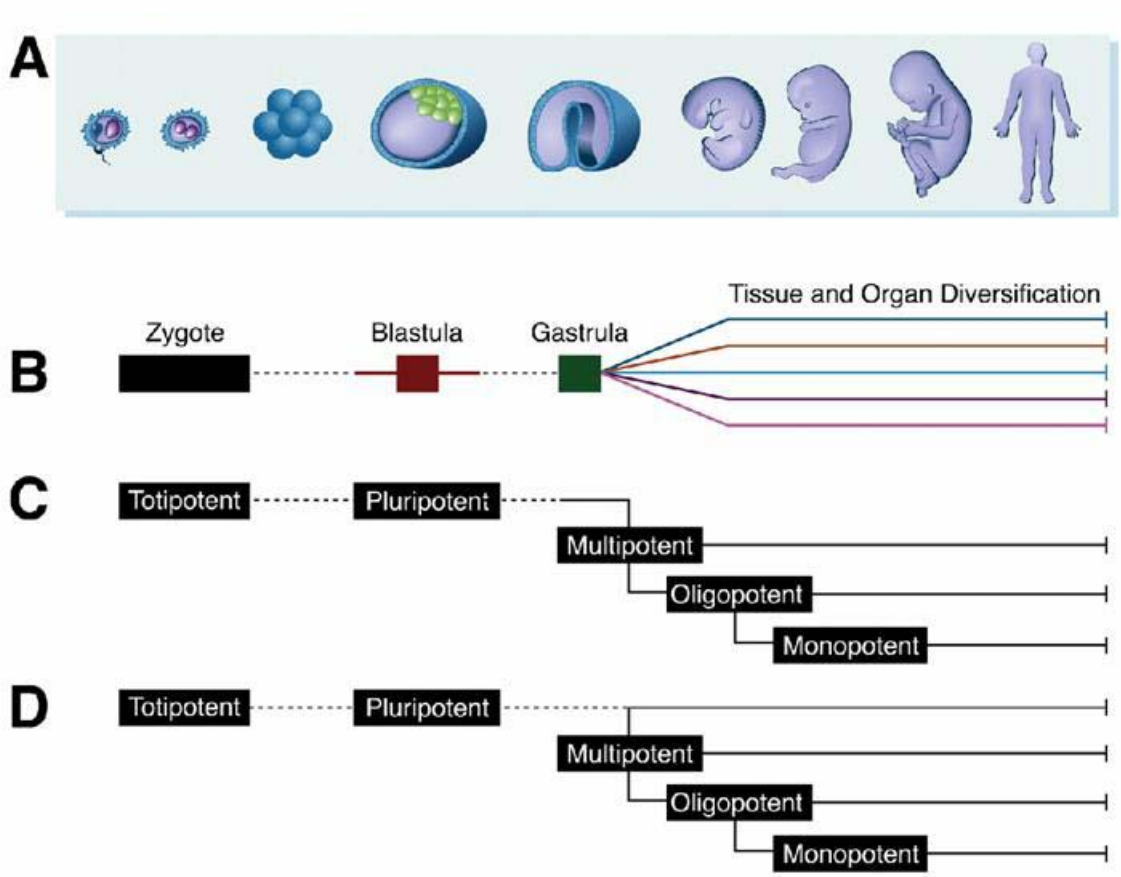


Figure 1: Origin and persistence of stem cells with differing potency during development. A,B: General scheme for development of an organism. C: Current paradigm for origins and potential of stem cells arising during development. Decreasing stem cell potential is associated with progression of development. Pluripotent stem cells are not expected to persist past gastrulation, while multipotent stem cells persist throughout the life of the organism. D: More recent findings indicate that pluripotent stem cells may also persist throughout the life of an organism. Adapted from Zipori, 2005[2].

ES cells were first isolated from embryonic mice by two independent groups in the early 1980's[8,9]. ES cell lines have since been reported from a number of species including humans, rats, primates, dogs, and several economically important farmed animals[10-16]. Despite these achievements, only mouse ES cells have been definitively proven to be able to differentiate into all embryonic cell types[17].

Since their initial identification, ES cells have become powerful tools for researching gene function, development, and differentiation. The ability of ES cells to contribute to the germline of chimeras enabled the development of transgenic mice; coupled with gene targeting by homologus recombination, ES cells have become the premiere system for studying gene function in mammals[17]. Great progress has also been made in the area of directed differentiation of ES cells *in vitro*, which will be further discussed.

1.1.2 Adult stem cells

Adult stem cells, also known as tissue resident stem cells, are multipotent cells that are found in the tissues or organs of developed organisms. Within tissues or organs, these cells are present in stem cell niches, which are anatomically and functionally defined microenvironments that support the existence and continued self renewal of stem cells[18]. Initially, stem cells were thought to only exist in the rapidly proliferating and differentiating tissues of the adult, including the hematopoietic system, epithelial linings of the digestive tract, and skin, all of which undergo continual renewal[19,20]. Adult stem cells have now been identified in a wide range of tissues including but not limited to the central nervous system, pancreas, heart, liver, and hair follicles, presenting the possibility that they may be isolated from most if not all adult tissues[21-29]. The best characterized adult stem cells are those originating from the hematopoietic system, originally isolated in the late 1980s[30]. Hematopoietic stem cells (HSCs) are relatively easy to harvest and purify, and have an array of markers available for their characterization[19]. Generally though, adult stem cells are rare and difficult to isolate, as they lack unique identifying markers and *in vitro* assays for functional testing[31]. Because of this lack of markers it is often difficult or impossible to differentiate between adult stem cells and their more

developmentally restricted progeny, and so putative adult stem cell populations are often referred to as stem/progenitor cells.

Arising from gastrulation onwards, adult stem cells are able to self renew and are thought to persist throughout the life of an organism [Figure 1, C]. In comparison to ES cells they have a more restricted developmental potential, as it is commonly accepted that they are biased towards differentiation along particular sets of lineages[2]. Specifically, adult stem cells can only give rise to terminally differentiated cell types corresponding to their tissue of origin, and do not cross tissue or germ layer boundaries to differentiate into cell types from different lineages[19] The distinction between pluripotent ES cells and adult stem cells is becoming blurred however, as there is increasing evidence for the ability of some adult derived stem cells to differentiate across the boundaries of the germ layers, thus suggesting persistence of some pluripotent cells throughout the life of an organism [Figure 1, D][2]. An example of an adult stem cell population that may exhibit pluripotency is mesenchymal stem cells (MSCs). MSCs are most often isolated from bone marrow, but they can also be found in adipose and muscle tissue, umbilical cord blood, and peripheral blood, amongst other locations[32-37]. They are well known to be able to differentiate into cells deriving from the mesodermal germ layer, such as adipocytes, osteoblasts, and chondrocytes, but there have been many reports of their ability to take on fates deriving from the ectodermal or endodermal germ layers as well.[38]. This apparent ability of MSCs to 'lose' their germ layer specific phenotype in order to adopt others, such as ectodermal and subsequently neural, is of particular interest due to the potential therapeutic applications, and promises to be a highly active area of research.

1.2 Early neural differentiation and development

1.2.1 A paradigm for cellular differentiation

Embryonic development begins with a single totipotent cell, possessing the potential to give rise to any embryonic or extraembryonic cell fate, which through finely orchestrated programs must give rise to all of the countless diverse cell types of an

organism. The cell types constituting an organism can be incredibly distinct from one another. These distinctions are based on many measures including metabolic activity, morphology, and gene expression, amongst others. Development and differentiation of varying cell types occurs via divisions of multipotent precursor cells with concurrent and progressive restriction of cell fate. One of the earliest events restricting the developmental potential of embryonic cells is establishment of the three germ layers during gastrulation. This event assigns cells to one of the mesodermal, ectodermal, or endodermal layers, hence restricting the fate of a cell to identities arising from the layer in which it exists, with few exceptions[19]. As cells progress along set developmental programs, they continue to lose their potential to differentiate into a broader variety of cell types, giving rise to tissue and organ systems.

Progressive lineage restriction is a common theme in development and clearly illustrates the hierarchical nature of cellular differentiation. Two classic examples of hierarchical cell fate determination are the fate map of *C. elegans*, and the population of the hematopoietic system by HSCs. *C. elegans* is a remarkable model organism particularly in that a complete fate map for every one of its 959 cells is established[39,40]. From its totipotent single cell origin, every cell division and associated lineage restriction is known, and can be drawn as a hierarchical tree. In the hematopoietic system, a single HSC has the potency to give rise to every cell type in the system, and does so by following a well characterized program of progressive lineage restriction[30,41,42]. HSCs have the capacity to either to self-renew, or give rise to progeny that are more restricted in their potential, such as the myeloid and lymphoid progenitor cells[43]. These progenitor cells, while still multipotent, are restricted to give rise to the myeloid and lymphoid cell lineages respectively. All cells of the embryo are believed to adhere to a similar process of lineage restriction, and the hierarchical structure of the process implies that cells generated earlier have a greater developmental potential[2]. This concept fits with commonly held views that ES cells, originating from the blastocyst, have the greatest developmental potential and stem/progenitor cells, originating later in development, are more restricted.

From a molecular perspective, cellular differentiation depends on coordination of cues both intrinsic and extrinsic to the cell. Despite the diversity of the cell types generated by developmental programs, common themes exist in their differentiation. One such theme is sequential, cascading expression of transcription factors containing the basic helix-loop-helix (bHLH) motif underlying cell determination and differentiation[44]. The ability of bHLH family proteins to modulate transcription depends of their ability to interact with DNA, for which either homo- or heterodimerization is requisite[45]. Dimers are formed via interactions between the helix-loop-helix regions of each protein forming a bundle of four helices[46,47]. Through their basic regions, dimers bind to DNA containing conserved E-box sequences, comprised of CANNTG hexamers[45]. Although the majority of bHLH proteins are known to act as transcriptional activators, several, such as the mammalian HES proteins and Olig2, have been demonstrated to function as transcriptional repressors[48,49] Activation of bHLH activity in stem/progenitor cells promotes cascading[48,49] expression of bHLH factors ultimately leading to terminal differentiation of varying cell types. bHLH factors have been shown to play a role in the fate determination of many cell types including, but not limited to, neural, muscular, hematopoietic, hepatic, mammary, and testicular, as well as epithelial cells of the digestive tract, urinary system, and respiratory system[50-56]. Most bHLH proteins are expressed in a cell type specific manner, allowing them to influence cell type specification by influencing transcription of target genes determining cell fate.

Myogenesis is an excellent example of cell fate determination and differentiation governed by cascading bHLH factors. Muscle cells arise from the mesoderm under the influence of myogenic bHLH proteins. Myogenesis is viewed as occurring in two general steps: determination of myogenic precursors, followed by terminal differentiation of these precursors[57]. In the determination step, uncommitted mesodermal cells are transformed into myogenic precursors, called myoblasts, via the activity of pro-myogenic bHLH proteins MyoD and Myf5[58]. If the activity of these pro-myogenic proteins becomes inhibited in mesodermal cells, then determination of myoblasts does not occur[57]. In determined myoblasts, MyoD and Myf5 activate expression of myogenic

differentiation genes such as *myogenin*, whose bHLH protein product promotes terminal differentiation of muscle cells[59].

The model of differentiation in which hierarchical lineage restriction occurs under the direction of cascading bHLH factors presents a challenge in development. It suggests that as lineage restriction progresses, there is a loss of multipotent progenitors. Clearly though, this is not the case as organisms are able to grow and develop from a single cell and renew their tissues, to some capacity, throughout their lives. Central to the model discussed herein is the ability of stem/precursor cells to extensively self renew, and regulation of cellular proliferation and differentiation signals. Such signals must be present in a fine balance; unfettered differentiation may lead to loss of stem/progenitor cell populations, while uncontrolled proliferation may lead to neoplasms and other disruptions in normal tissue architecture. A well characterized mechanism for maintaining balance between cellular differentiation and proliferation is lateral inhibition. Lateral inhibition describes the process in which single cells from amongst a group of equivalent stem/precursor cells become singled out for terminal differentiation. It occurs via a conserved mechanism involving signalling through the cellular receptor Notch.

The conserved core elements of the Notch pathway are the Notch receptor itself and its transmembrane ligands from the Delta and Serrate protein families[60]. Upon ligand interaction with Notch, the intracellular domain of Notch (Notch ICD) is proteolytically cleaved, allowing it to translocate to the nucleus where it acts as a transcriptional activator in conjunction with homologs of the Drosophila proteins suppressor of hairless, su(H), and mastermind, MAM, [Figure 2]. Activity of pro-determination bHLH factors in equivalent stem/precursor cells induces expression of Notch ligands. Determination and differentiation is inhibited in cells receiving relatively high levels of signalling through Notch, initiated by surrounding ligand expressing cells[61-63]. This inhibition is mediated by targets of the Notch pathway, including the inhibitory bHLH Hes proteins[64,65]. Hes proteins are known to antagonize the action of other bHLH factors, therefore repressing differentiation in cells receiving signalling through Notch[66]. Notch

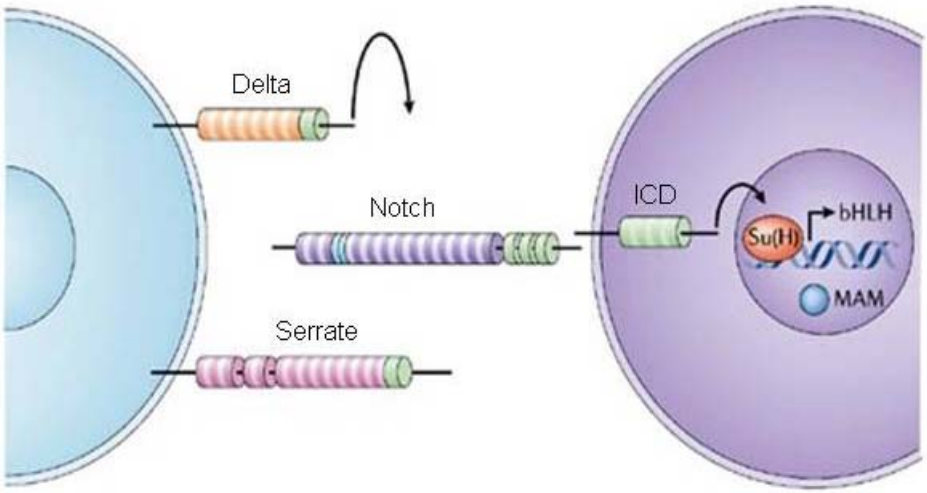


Figure 2: The core notch signalling pathway. Interaction of the Notch receptor with its ligands from the Delta or Serrate family of proteins leads to proteolytic release of the Notch intracellular domain (ICD), and its translocation to the nucleus. In the nucleus, the Notch ICD interacts with suppressor of hairless, Su(H), and mastermind (MAM), to regulate the transcription of genes encoding proteins such as the inhibitory bHLH factors of the Hes family. Adapted from Louvi and Artavanis-Tsakonas, 2006[60].

signalling is also known to promote cellular proliferation, though the exact mechanism promoting proliferation is not clear[67-69]

There is a broad dependence on Notch signalling in differentiation and maintenance of many adult tissues. Regulation of the balance between cellular proliferation and differentiation by lateral inhibition has been noted as a mechanism in differentiation of hematopoietic, neural, muscular, epithelial, and skeletal cell types, among others[60,69-72]. The end result of lateral inhibition through Notch signalling in many systems is a proliferative burst among uncommitted cells, thus ensuring maintenance of adequate stem/progenitor cell populations[70].

Overall, a paradigm for cellular differentiation exists in which stem/precursor cells with the potential to take on multiple fates undergo hierarchical fate restriction, eventually withdrawing from the cell cycle to become terminally differentiated. Fate restriction is governed by the interplay of multiple signals, including cascading expression of bHLH factors, and lateral inhibition via Notch activation. These signals ensure an appropriate balance between cellular differentiation and proliferation required for development and maintenance of an organism throughout its life. This paradigm applies to a wide range of systems, making the exploration of molecules and mechanisms involved in differentiation and development of a particular cell system relevant for insight to mechanisms at play in other systems.

1.2.2 Early neural development: an overview

Neural development involves highly intricate systems of numerous intrinsic and extrinsic signals leading to differentiation of diverse cell types. Development of the vertebrate nervous system begins very early in embryogenesis with the onset of gastrulation, which gives rise to the three germ layers of the embryo. The nervous system is derived from the ectoderm, which during the cellular movements of gastrulation comes to overlay mesodermal cells. The mesodermal cells induce the overlying ectoderm to become the neuroectoderm forming the neural plate, specified to take on neural fates[73]. This process is known as neural induction, and the nature of the inductive signals remained elusive for many years. Using molecular approaches in the

Xenopus model, several groups were able to identify the neural inducing activities of *noggin*, *follistatin*, *cerberus*, and *chordin*[74-78].

The proteins encoded by the identified neural induction genes were expected to be instructive cues, acting as switches to 'turn on' neural fate specification. In earlier experiments however, ectodermal explants were taken from *Xenopus* embryos, dissociated, and cultured in isolation, resulting in neuralization of the cells even without signalling from the organizer[79,80]. In addition, experiments in Drosophila had been completed showing inhibition of a bone morphogenic protein (BMP) homolog, known to induce epidermal fate in ectodermal cells, by a chordin homolog[81]. Taken together, these results led to the idea that neural induction was actually attained via an inhibitory mechanism. Hemmati-Brivanlou and colleagues showed that neural induction was a product of inhibition of BMP signalling, which normally acts between ectodermal cells to inhibit neural fate[74,82]. This concept of neural inducers inhibiting the neuro-inhibitory BMP signal forms the basis for the widely accepted default model of neural induction, in which ectodermal cells would all adopt neural fate in the absence of BMP signalling (reviewed by Muñoz-Sanjuán and Hemmati-Brivanlou, 2002[83]).

Following induction of the neuroectoderm forming the neural plate, the underlying mesoderm condenses to form the notochord. Simultaneously, the outer edges of the neural plate fold to meet each other, eventually fusing to separate from the surface ectoderm and form the neural tube[73]. This process is known as neurulation. Following neurulation, the neural tube is comprised of equivalent neuroectodermal cells, which constitute neural stem cells based on their ability to self-renew and differentiate into multiple neural cell types. During the course of neural development, these cells undergo numerous proliferative divisions in order to provide enough cells for establishment of the nervous system. Although these neuroectodermal cells will eventually give rise to diverse neural cell types, a generalized program of sequentially expressed bHLH transcriptional regulators is necessary for their determination and differentiation. bHLH factors in vertebrate neural development carry out an array of functions, including early neuronal determination, inhibition of neural cell fate in lateral inhibition, and

differentiation of specific neural cell subtypes[62,84-86]. Similar to the bHLH governed myogenesis program, neurogenesis also occurs in a two step process of determination and differentiation [Figure 3]. The first bHLH factors to be expressed in the neuroectodermal cells of the neural tube are pro-neural[87]. The role of these pro-neural bHLH factors, such as neurogenin and Mash1, is to promote determination of neural precursor cells, which are also known as neuroblasts[62].

All of the neuroectodermal cells initially expressing pro-neural factors are equivalent, but only selected cells will be singled out for determination, withdrawal from the cell cycle, and terminal differentiation. Similar to the process seen in other differentiation programs, the selection of cells that will differentiate from amongst a field of equivalent proliferating cells is attributable to the process of lateral inhibition, mediated by Notch signalling. Signalling through Notch inhibits the expression and activity of pro-neural genes, thereby facilitating continued progression through the cell cycle. Cells with low levels of Notch signalling, and therefore maintained activity of pro-neural genes become determined neuronal precursors which then express differentiation bHLH factors, such as NeuroD, promoting their terminal differentiation (reviewed by Bertrand *et. al.* 2002[88]). The exact cellular identity that the precursor cell will take on as it differentiates is dependent on the sequential complement of differentiation factors that it is induced to express[50].

1.2.3 Retinoic acid in early neural development

The retinoid family of molecules are metabolites of vitamin A, also known as retinol, which are implicated in a wide array of developmental processes. Vitamin A was discovered as an essential dietary compound prior to 1930, when adult farm animals fed vitamin A deficient diets displayed a range of dysfunctions including anaemia, poor immune function, blindness and motor neuron degeneration in the spinal cord (reviewed by Maden, 2002[89]). Vitamin A deprivation studies in pregnant pigs indicated that it was essential for normal embryonic development, as litters born to vitamin A deprived mothers lacked eyeballs. Later anatomical studies of vitamin A deprivation in avian and rodent model systems provided more detailed descriptions of abnormalities such as a lack

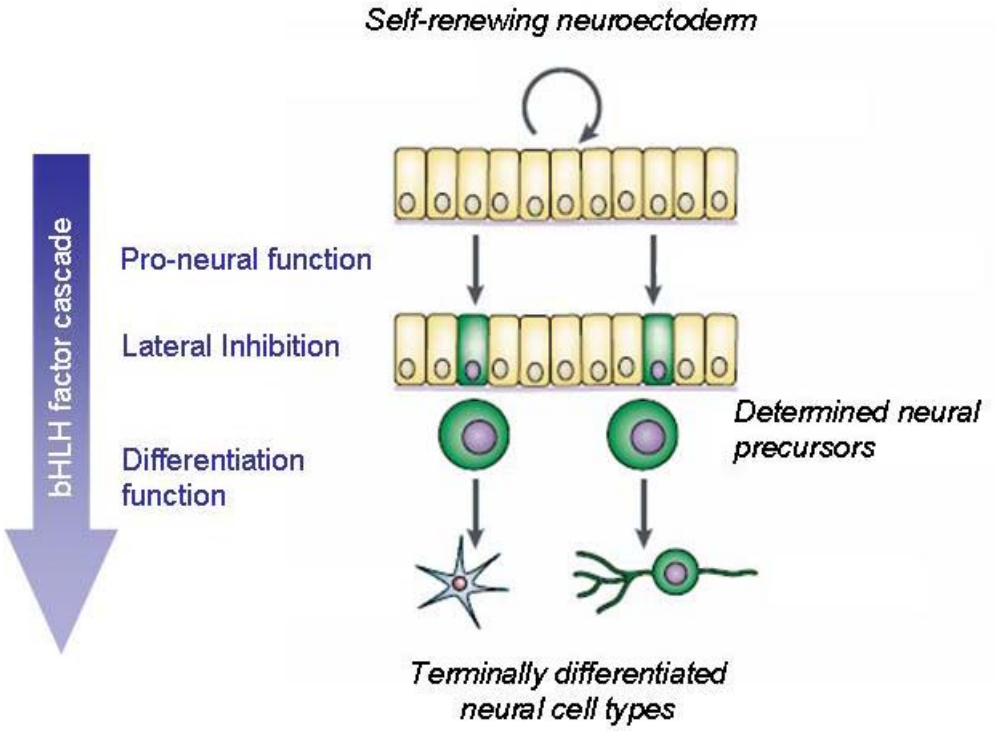


Figure 3: The general bHLH transcriptional factor cascade in neurogenesis. Selected self-renewing neuroectodermal cells differentiate into diverse neural cell types in a process mediated by a general cascade of bHLH factors. The bHLH factors involved exhibit a variety of functions including promotion of cellular determination, facilitation of lateral inhibition in conjunction with Notch signalling, and specified differentiation of determined cells into neural sub-types. Adapted from Bertrand *et.al.* 2002[88].

of peripheral neurons, neural crest cell death, incomplete or absent posterior hindbrains, as well as heart and circulatory defects[90]. Reciprocal experiments examining the effects of excess vitamin A were also carried out, and indicated its teratogenic properties affecting many systems including the CNS.

The vitamin A metabolic pathway was established in 1953, and retinoic acid (RA) was identified as the major biologically active retinoid[89]. RA is made from vitamin A in two major steps: Vitamin A is first converted into retinaldehyde by retinol- or alcohol-dehydrogenases, which is then converted to RA by retinaldehyde dehydrogenases. RA exerts its action on cells by binding its nuclear receptors, of which there are six in mammals: three retinoic acid receptors (RARs) and three retinoid X receptors (RXRs)[91,92]. These receptors act as heterodimeric transcriptional regulators when bound with ligand, and modulate the transcription of genes containing retinoic response elements (RAREs)[93].

Retinoic acid has been extensively described as a regulator of neuronal birth and differentiation. The requirement for RA in early neural development was first noted in *Xenopus* embryos. The earliest formed neurons in *Xenopus*, the primary neurons, withdraw from the cell cycle and differentiate nearing the end of gastrulation[94]. These neurons for an elementary nervous system that controls the reflexive movements required for the early embryo[95]. Present in defined numbers and spatial arrangement, primary neuron development is influenced by RA signalling. Studies by Sharpe and Goldstein, in which they injected mRNAs for nuclear RA receptors into early *Xenopus* embryos, resulted in increased numbers of primary neurons being formed[96,97]. They also showed through the use of dominant negative receptors and RA antagonists that in the absence of adequate RA signalling, the primary neurons fail to form in appropriate numbers or spatial arrangement[97,98]. It was also noted in these studies that the *Xenopus* orthologs of Delta and neurogenin responded to retinoid signalling, indicating the role of RA in early neural differentiation. Further studies of RA in early neurogenesis corroborate these findings, with observations that it falls upstream of several pro-neural genes[99,100].

Furthermore, hundreds of genes with a broad cross section of functions have been shown to be influenced by RA signalling during neural differentiation and outgrowth[101].

Despite its role in neuronal birth and differentiation, RA is best known for its role in patterning of the developing nervous system, which occurs concurrently with neuronal differentiation. As the neural tube forms in vertebrates it immediately becomes regionalized, in a manner related to the organization of the mature brain. The majority of the caudal (posterior) neural tube contributes to the spinal column, while the rostral (anterior) regions contribute to the three major divisions of the brain: forebrain (prosencephalon), midbrain (mesencephalon), and hindbrain (rhombencephalon). As embryonic development progresses, this anterior-posterior (AP) patterning becomes increasingly complex, with continued growth and segmentation of the three original divisions of the brain. All of the progressive segmentation of the neural tube may be characterized by the expression of well-defined marker genes.

As informed by the early vitamin A studies, RA plays a significant role in patterning of the AP axis of the neural tube. In particular, RA is central in orchestrating the development of the hindbrain. As development of the hindbrain progresses, segmentation gives rise to eight transient regions known as the rhombomeres[102]. Rhombomere 1 (r1) is the most anterior of the rhombomeres, existing adjacent to the midbrain-hindbrain junction, and each successive rhombomere is more posterior than the last with r8 lying at the junction of the hindbrain and the spinal column. Each rhombomere is defined based on morphological properties and differential expression of marker genes, including ordered patterns of *Hox* gene expression[103]. RA acts synergistically with fibroblast growth factor 8 (FGF-8) signalling to establish hindbrain segmentation [Figure 4, A]. They act in opposing and overlapping gradients, with the highest levels of FGF-8 signalling seen at the midbrain-hindbrain junction, and the highest levels of RA signalling at the r8-spinal column junction[104-106]. The role of RA in patterning the hindbrain is to establish the posterior hindbrain, as indicated by studies of hindbrain segmentation in the absence of RA. In the absence of RA, r4 is not correctly formed, resulting in an r4-like segment, r5-r8 fail to form entirely, and the anterior

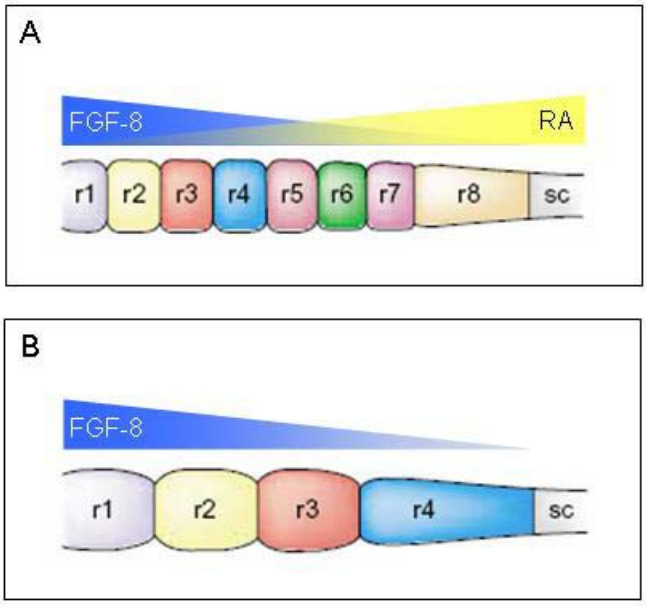


Figure 4: Retinoic acid is essential for patterning of the hindbrain. Panel A: RA and FGF-8 signalling act in opposing gradients to correctly establish hindbrain segmentation. Panel B: In the absence of RA signalling, r5-r8 are not formed, an R4-like structure is formed, and r1-r3 are expanded. Sc: spinal cord; r1-r8, rhombomeres 1-8. Drawn with components from Gavalas 2002[102].

rhombomeres r1-r3 are expanded [Figure 4, B][107,108]. This effect is expected to be a result of mis-regulation of genes required for posterior hindbrain patterning. Required *Hox* gene expression has been observed to be significantly reduced in the absence of RA[109]. Direct RA regulation of Hox genes involved in hindbrain development has been demonstrated, and other genes involved in patterning are also expected to be under control of RA[102,110].

1.3 Neural differentiation of ES cells.

Differentiation of ES cells *in vitro* has extended the variety of cell types that can be effectively studied in culture[111]. Protocols for differentiation of a wide variety of cell types have been established and extensively described. Cell types that can be generated from ES cells *in vitro* include neural, skeletal, myogenic, cardiac, and hematopoietic cells[112-116]. Given their pluripotency and advances in cell culture techniques, it may be possible to eventually differentiate ES cells into most desired cell types *in vitro*, and possible that these *in vitro* derived cells will be appropriate for therapeutic transplant[111].

Despite the elucidation of genes and mechanisms controlling neural differentiation and development, a complete picture of the entire complement of controlling elements is not available. The methods available for inducing neural differentiation of ES cells provide an excellent opportunity to examine the molecular mechanisms governing neural development. In addition, the mouse ES cell model system is an integrated platform for study of genes involved in the process of differentiation, given the array of methods available for their manipulation including gene targeting, transformation, and blastocyst injection for creation of transgenic mice[117].

1.3.1 Early in vitro neural differentiation methods

The first *in vitro* neural differentiation protocols were developed in P19 embryonal carcinoma (EC) cells. P19 cells, derived from mouse, are immortalized, known to be multipotent, and can contribute to many somatic cell lineages when injected into mouse blastocysts[118,119]. Earlier experiments had indicated that EC cells could differentiate into a variety of cell types when cultured as aggregates and exposed to

different drugs[120]. In particular, P19 cells were found to be easily differentiated into both neural and muscle cell types[121,122]. For induction of neuronal differentiation, P19 cells were cultured as aggregates in suspension for three days, followed by exposure to RA for an additional two days, after which they were dissociated and plated on adherent surfaces, and examined after three days[121].

The result of the RA induction protocol was a mixed cell population enriched with neuron-, glia-, and fibroblast-like cells[121-123] Differentiation into neural cell types in response to RA was found to be dependent on aggregate culture, and this effect was thought to be attributable to cell-cell interactions similar to those occurring in early development[121,124]. Further analysis of the cells derived from P19 neural differentiation demonstrated that the neuron-like cells, in addition to having the morphology of neurons, also expressed many genes characteristic of neurons including a comprehensive set of neurotransmitters and neuroreceptors[125-128]. These results showed that it was possible to generate a wide range of apparently functional neural sub-types in culture, indicating promise for continued research into differentiation mechanisms.

1.3.2 The Bain ES cell neural differentiation method

While completing postdoctoral work at Washington University, Dr. Gerard Bain pioneered methods for reproducible neural differentiation of mouse ES cells, which he later brought with him to the Rancourt Lab. His work was informed by early neural differentiation results in P19 cells, and also important similarities between P19 and ES cells; both can differentiate into a broad array of cell types both *in vivo* and *in vitro*. The method he developed, which will be referred to as the Bain ES cell neural differentiation method, involves an eight day induction period consisting of two major steps: Generation of ES cell aggregates, known as embryoid bodies (EBs), and exposure of these aggregates to RA [Figure 5][116]. Generation of EBs is achieved by culturing ES cells in LIF-free medium, in the absence of an adhesive substrate for four days. The cellular characteristics of EBs are not well known, but they have been shown to express markers of all three embryonal germ layers[129]. If allowed to spontaneously differentiate, they give rise to a variety of cell types, with upwards of 30% of the culture consisting of

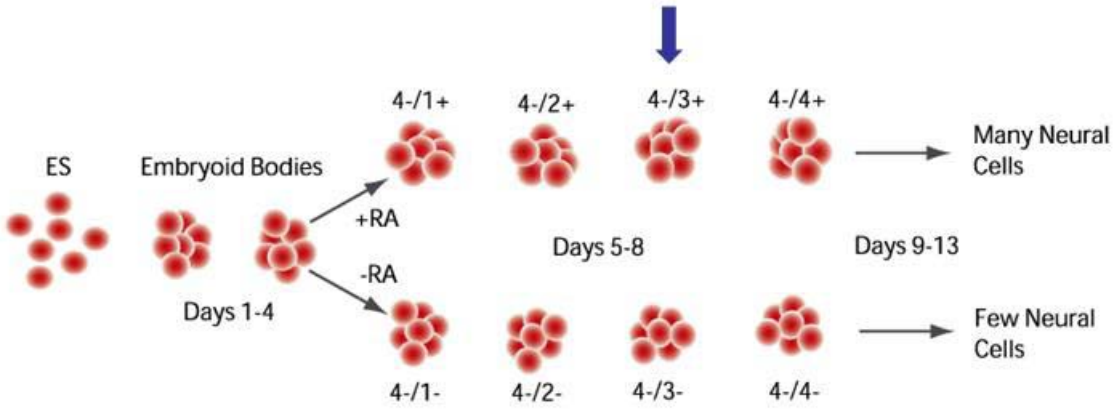


Figure 5: The Bain ES cell neural differentiation method. ES cells are grown in suspension as embryoid bodies for four days in the absence of LIF. Retinoic acid is then added to the culture medium at a concentration of 0.5 μM. After four days, the cells are dissociated and plated for outgrowth of differentiated cells. At the 4-/3+ stage of the induction, indicated by the blue arrow, the differentiating cultures are enriched for neural precursors expressing early markers of neural determination and differentiation.

cardiomyocytes[130]. The second step of the induction protocol is an additional four days of culture in which culture medium is supplemented with 0.5 μM RA[116]. Following this eight day induction period, often referred to as 4-/4+ (representing four days without RA / four days with RA) induction, the EBs are dissociated and transferred to tissue culture dishes coated with an adhesive substrate for cell attachment and outgrowth. Outgrowth of the dissociated aggregates is allowed to proceed for an additional five days, after which 40-50% of the cells in the resulting culture have glial or neuronal morphology[116].

Following development of the Bain ES cell neural differentiation protocol, other groups also used RA to induce neural differentiation of ES cells in culture, yielding similar results[1,112,131,132]. Surveys of known neural gene expression demonstrated that during *in vitro* neural differentiation, genes involved in neural development are expressed in an ordered manner similar to that seen during early neural development *in vivo*[112,131,133]. Thus, it is expected that ES cell neural differentiation recapitulates neural differentiation *in vivo*. At the 4-/3+ stage of the Bain protocol, at which time ES cells have been cultured as EBs for four days in the absence and three days in the presence of RA, the differentiating cultures were found to be enriched with neural precursor cells expressing early markers of neural differentiation[133]. Parallel control cultures that had not been exposed to RA (4-/3-) did not express any markers of neural differentiation; rather, they expressed mesodermal markers[133]. As seen in RA differentiated P19 cultures, the cells resulting from ES cell neural differentiation expressed an array of neurotransmitters and receptors, as well as various markers of neuronal sub-populations[112,131,134]. These results indicated that despite the known role of RA in hindbrain patterning, RA induced neural differentiation does not appear to be biased towards generation of hindbrain neurons[111,112,135]. The reasons for this observation are not clear, but the possibility exists that environmental cues are present *in vitro* influencing the differentiation of specific neural subtypes[111]. To further characterize the differentiated cells with neural morphology, electrophysiological studies were also pursued. Results of these studies showed that the neural-like cells in culture were likely functional neurons, as almost all of them possessed functional voltage gated channels[116].

Even with the efforts to analyse the neurons arising from ES cell neural differentiation, the characteristics of all the resulting cells are not well described. In particular, the nature of the cells not having neural or glial morphology has not been extensively examined, though it has been suggested that they may be multipotent neuroepithelial-like stem cells and lineage restricted neural cell precursors[132,136].

An ultimate goal for ES cell differentiation technologies is their application in regenerative cell therapies. Therefore, groups also pursued preliminary transplantation studies using ES cell derived neural cell types. Neural precursors derived from RA induction of ES cells have been introduced into the developing brain of embryonic rats, where they were shown to have the capacity to differentiate and migrate along with recipient cells, integrating themselves with the recipient nervous system[1]. Neural precursors transplanted into injured rat spinal cords were also able to differentiate and integrate into the recipient nervous system, resulting in improvement of motor function[137]. Even in light of these promising results, concerns regarding the safety of transplanting poorly characterized, mixed cell populations prevented clinical application of these cells. Clearly, there was a need for a more precise and complete understanding of neural differentiation and development both *in vitro* and *in vivo*.

1.4 The Bain-Rancourt EST sequencing project

Establishment of the Bain ES cell neural differentiation method in the Rancourt lab provided an excellent opportunity for the identification and study of genes involved in the neural differentiation process. Thus, a survey of genes expressed during the Bain ES cell neural differentiation protocol was undertaken.

Dr. Gerard Bain and colleagues in the Rancourt lab completed this survey using a subtractive hybridization approach to isolate genes which are highly expressed in ES cells that have been cultured as EBs in the presence of RA (4-/3+), in comparison to non-differentiating pluripotent ES cells. As noted earlier, 4-/3+ cell cultures have been shown to be enriched in neural precursors expressing early markers of neural differentiation. From this subtraction, a total of approximately 1200 ESTs were sequenced[135]. Each of

the sequences was compared to all of the other sequences in the dataset using a stand alone implementation of BLASTn, revealing a total of 604 non-redundant ESTs, which were subsequently submitted to Genbank (Accession numbers AW244216-AW244819). This submission made the Rancourt lab one of the biggest contributors to the Genbank EST database at that time.

Annotation of the Bain-Rancourt EST sequences was completed by comparing each to sequences in Genbank using BLASTn or BLASTx. Ninety-six of the ESTs showed no significant similarity to any existing sequences in Genbank and were therefore deemed novel. An additional 197 EST sequences showed similarities to other uncharacterized sequences, while the remaining ESTs could be attributed to known genes. The majority of known genes isolated in the screen, including *Hes1*, *Nova1*, *Wnt1*, *NFAT*, and *neurogenin*, have known roles in neural development and function. Notably, not all of the expected early markers of neural differentiation were isolated in the subtractive screen. Markers including Mash1, nestin, and Delta1, are not represented in the EST set, suggesting that the screen was not saturating and many other novel genes remained to be isolated from ES cells differentiating towards neural lineages.

1.5 Project objectives

The data obtained from the Bain-Rancourt EST sequencing project was analyzed with a large amount of human intervention and time investment. The resulting annotations and other pertinent data were contained in a single table, whose limitations were rapidly exceeded. Management and further enrichment of the data became a particular challenge, especially in light of the rapidly changing public data from which much of the information regarding the ESTs was gleaned. The requirement for a more effective data enrichment, management, and dissemination strategy was clear.

Also arising from the Bain-Rancourt EST sequencing project was an interest in focusing the lab on gene discovery and analysis of regulatory factors in neural differentiation. In particular, the lab was interested in pursuing characterization of genes in the dataset that met pre-determined criteria for putative regulators of other neural

specific genes. These criteria included indication of expression in neural tissues, and probability of encoding a nucleic acid binding protein.

In light of this research directive for the lab, the objectives for the project presented in this thesis were two fold:

A. To implement a new data enrichment, management, and dissemination strategy for the Bain-Rancourt EST set in order to facilitate candidate neuro-regulatory gene selection.

B. To identify and characterize a candidate gene from the dataset likely to encode a nucleic acid binding gene with probable roles in neural differentiation.

Chapter Two: Materials and Methods

2.1 General Considerations

The majority of solutions used during the course of this project were made according to recipes found in Molecular Cloning: a Laboratory Manual[138,139]. ddH_2O is ultra pure water from a MilliQ water filtration system (Millipore). Restriction endonucleases used in this work were purchased from Invitrogen or New England Biolabs. All synthetic oligonucleotides were ordered from University Core DNA Services (University of Calgary) or from Operon Technologies. For work involving RNA, standard RNAse free methods were employed. Nuclease free aerosol pipette tips, microfuge tubes, transfer pipettes, and disposable graduated pipettes were routinely used. All non-disposable glass and plastic ware used were treated with RNAzap RNAse decontamination solution (Ambion) according to manufacturer directions. DEPC treated water was exclusively used. All solutions were treated with DEPC or made using RNAse free chemical stocks and DEPC treated solutions. Treatment of water or solutions with DEPC was carried out by adding DEPC to a final concentration of 1% v/v, and mixing vigorously in a fume hood. After overnight incubation, the solutions were autoclaved for sterilization and hydrolysis of the DEPC.

2.2 Computational Methods

2.2.1 General bioinformatic methods

During the course of this work, a wide range of general bioinformatic methods were employed. The most widely used methods were those based on BLAST, including but not limited to BLASTp, BLAST 2 sequences (bl2seq), psiBLAST, and rpsBLAST[140-142]. Wherever possible and appropriate, sequence comparisons were conducted on the level of the amino acid sequences. All alignments were inspected by eye, with alignments with scores of less than 100 and expect-values of greater than 10^{-3} receiving particular attention. For sequence comparisons against genomic sequences, SSAHA against the Ensembl genome databases was generally used[143,144].

For manipulations and analysis of protein sequences the ExPASy proteomics server was typically used[145]. Tools most commonly used include ExPASy translate tool, and InterPro Scan[146], for conceptual translations and detection of conserved domains respectively.

2.2.2 Refinishing and annotation of RED dataset

Original sequence chromatogram files generated during sequencing of the RED ESTs were obtained from the archives of the University of Calgary Sequencing Core Facility. Each sequence was refinished using the Staden package[147,148]. This process included use of pregap4 for pre-processing of the sequence chromatogram files. Pre-processing steps included quality assessment and clipping, vector sequence clipping, detection of host sequence contamination, and sequence file conversions. Sequences were then subjected to assembly using gap4, so as to merge overlapping sequences, and allow visual inspection of low quality base calls.

Refinished EST sequences were re-annotated using MAGPIE[149,150]. MAGPIE configuration and initiation of the RED EST annotation project was completed by Mr. Paul Gordon, of the Sun Centre of Excellence in Visual Genomics. Once the relevant data transactions were completed, each new annotation assigned by MAGPIE was accepted or rejected based on personal inspection of the evidence overlays generated by MAGPIE. ESTs for which no annotation was assigned, or for which the annotation was rejected based on weak supporting evidence, retained the annotation "novel". The refinished and annotated sequences were eventually used to populate RED.

2.2.3 Design and implementation of RED

The RED database schema, and front end design were conceived in consultation with intended end-users. RED was implemented using an integrated multi-platform distribution of open source software, specifically NuSphere 1.13.10 (http://www.nusphere.com), which is portable to Windows, Linux, and Sun operating systems. The user interface of the database was created using PHP 4.0.5. Automated external database query tools were produced using PERL 5.6.1. The data produced during the initial EST sequencing project, in subsequent bench-based experiments, and

the results from database queries is stored locally in a MySQL 3.23.38 relational database. Apache 1.3.20 acts as the web server for this database.

2.2.4 EST mapping and gene structure prediction

The sequence of 3B6 clone 4.2, contained in construct pCR-4.2 (section 2.7.9) was compared to the Ensembl mouse genome assembly 3, version 7.3a.1 (release date May 27, 2002) using SSAHA. One genomic region containing a single Genscan *ab initio* gene prediction[151], 2 non-overlapping Ensembl predicted transcripts, and with a significant number of ESTs mapping to it was returned, and this was further examined. The sequence for each of the EST sequences mapping to the given genomic region were downloaded. Each of the ESTs was placed in order according to their position in the genomic region. EST sequences were collapsed onto one another by assembly of overlapping sequences using Gene Tool (BioTools Inc), after which a single assembled sequence resulted. This sequence was then again compared to the mouse genome as above, to determine approximate sizes for each exon. PCR primers in each predicted exon were designed (section 2.5.1) and confirmation of the predicted gene structure was carried out via PCR using these primers followed by Topo-TA cloning and sequencing of the resulting amplimers (sections 2.5.2, 2.7.9, 2.7.8)

2.2.5 Phylogenetic analysis

Phylogenetic analysis was completed for a selected set of mSe70-2 protein homologs, starting with identification of homologs in the public datasets. Sequences homologous to mSe70-2 were gathered using BLASTp searches against the NCBI RefSeq protein datasets limited to each organism of interest. For organisms of interest for which RefSeq datasets did not contain Se70-2 homologs, converging Phi-BLAST searches were completed against the NCBI non-redundant (nr) protein datasets limited to each organism. All homologs were confirmed using reciprocal BLASTp searches against the NCBI *M. musculus* RefSeq dataset. Each sequence used in subsequent analyses was found to be the reciprocal best hit to the *M. musculus* Se70-2 sequence represented in the NCBI RefSeq dataset.

Alignment of the mSe70-2 homologs was completed using ClustalX[152]. The scoring matrix, gap opening penalties, and gap extension penalties were BLOSUM62, 11, and 1 respectively. The resulting alignment was imported into MacClade 4 (http://macclade.org/macclade.html) for visual inspection and identification of homologous amino acid sites of the homologous proteins[153]. Only homologous blocks of sequence across all sequences analysed were used in evolutionary model determination and tree generation.

Determination of the most appropriate model of evolution to describe the given dataset was completed using Tree-puzzle 5.2[154], and tree topologies with node support values were determined using two methods: Bayesian and Maximum Likelihood analysis. Bayesian analysis was done using Mr. Bayes v.3.0[155] for 1000000 generations with a burnin value of 100, sampling 10000 trees, resulting in a consensus tree topology and posterior probability values for node support. Bootstrapped maximum likelihood analysis was completed with PhyML v.2.4.4[156] analyzing 100 pseudo-replicate datasets, resulting in an optimized maximum likelihood tree with bootstrap values for node support. Parameter values determined by Tree-puzzle were used in these analyses. Trees were visualized and rooted with *D. discoidium* using TreeView v.1.6.6[157], and the multiple sequence alignment was shaded for presentation using GeneDoc (http://www.psc.edu/biomed/genedoc).

2.3 Cell Culture

2.3.1 ES cell lines, media, and reagents

Mouse ES cell line D3 was used throughout the course of this work. Dulbecco's modified eagle medium (DMEM) with high glucose, l-glutamine, and sodium pyruvate (Gibco, Cat#11995-065) was supplemented as required to make all ES cell culture media. Only sterile disposable plasticware and strerile glassware dedicated to ES cell culture were used in handling the cells. Special care was taken in ensuring all items entering the flow hood were dedicated to use in tissue culture, and thoroughly cleaned prior to use according to standard, yet stringent, tissue culture techniques.

ES cell growth medium for routine culture was DMEM supplemented with 20% fetal bovine serum (FBS; Gibco), 0.1 mM non-essential amino acids (Gibco), 50 units/mL penicillin and 50μg/mL streptomycin (penicillin-streptomycin solution; Gibco), 0.1 mM β-mercaptoethanol, and 1000 units/mL leukemia inhibitory factor (LIF; Gibco). ES cell control differentiation medium was the same as the ES cell growth medium, but without the addition of LIF or β-mercaptoethanol. ES cell neural differentiation medium was composed of ES cell control differentiation medium with the addition of 0.5μM all-trans retinoic acid (RA; Sigma, Cat# 2625). Freeze medium consisted of 90% FBS and 10% dimethylsulfoxide (DMSO). All medium was stored at 4°C for immediate usage.

All-trans RA was prepared as a 1mM stock in 95% EtOH, aliquoted, and stored protected from light at -20°C. Magnesium and calcium free Dulbecco's phosphate buffered saline (DPBS) was prepared as a sterile 10x stock solution containing 2 g/L KCl, 2 g/L KH_2HPO_4, 80 g/L NaCl, and 21.6 g/L $Na_2HPO_4{\cdot}7H_2O$. 1x DPBS was prepared by diluting an aliquot of the 10x stock in sterile ddH_2O in a sterile bottle. Sterile 0.1% gelatin was prepared and stored at 4°C, for use in coating dishes for feeder free culture of ES cells. 0.25% Trypsin/EDTA solution (Gibco) was stored at 4°C for regular use in passaging cells.

2.3.2 Other cell lines, media, and reagents

Two cell lines other than D3 ES cells were commonly used during this project: P19 mouse embryonal carcinoma, and human embryonic kidney (HEK) 293T cells, gracious gifts from the labs of Dr. J. Cross and Dr. Tara Beattie (University of Calgary) respectively. These cells will be referred to as P19 and HEK cells throughout the remainder of this text.

All media for P19 culture was made from minimum essential medium alpha medium (αMEM) with L-glutamine, ribonucleosides, and deoxyribonucleosides (Gibco, Cat# 12571-063). P19 growth medium (P19GM) consisted of αMEM supplemented with 10% FBS and 1% antibiotic/antimycotic liquid (Gibco). P19 induction medium (P19IM) consisted of αMEM with 5% FBS. P19IM was supplemented with all-trans RA to a final concentration of 0.5μM for induction of neural differentiation.

Medium used for HEK 293T culture, HEK growth medium (HEKGM), was high glucose DMEM with L-glutamine (Gibco, Cat# 11965-092) supplemented with 10% FBS and 1% antibiotic/antimycotic liquid.

Freeze medium for both cell lines consisted of 90% FBS with 10% DMSO. DPBS for general cell culture was prepared as for ES cell culture, described in section 2.3.1.

2.3.3 Routine growth and passaging of ES cells

ES cells were routinely grown as feeder free monolayers in gelatin coated petri dishes or culture flasks at 37°C in a humidified 5% CO_2 incubator. Cells were observed daily for changes in morphology or growth properties, and passaged when 60-80% confluence was reached, typically every second day. For routine passaging, growth medium was removed from the cells which were then rinsed with DPBS to remove any residual medium. DPBS was aspirated off and enough trypsin to just cover the monolayer was added to the culture vessel. Trypsinization was allowed to proceed at 37°C for 1-5 mins as determined by periodically observing the cells for dissociation from the culture vessel and each other. A volume of ES cell growth medium at least equal to that of the trypsin was then added to the culture vessel to stop the trypsinization reaction. If needed, the cells were further dissociated from each other mechanically by pipetting up and down using a cotton plugged Pasteur pipette. The ES cells were then split at ratios ranging from 1:3 to 1:5 into new gelatinized culture vessels, or back into the vessel which they were previously growing in. ES cell growth medium pre-warmed to 37°C was the added to an appropriate volume depending on the vessel being used, and the cultures were returned to the incubator for growth.

2.3.4 Storage and defrosting of ES cells

Preparation for long term storage of ES cells in liquid nitrogen (lN_2) was carried out by first trypsinizing cultures as described above. After stopping the trypsinization reaction, cells were collected into a 15 mL conical tube (Falcon) using a cotton plugged Pasteur pipette. The cell suspension was spun in a cell culture dedicated table top centrifuge at 200 x g for 10 mins. The medium was then aspirated and the cell pellet

gently resuspended in 1 mL of ES cell freeze medium per 25cm^2 of cells harvested. Using a cotton plugged Pasteur pipette, cells were aliquoted into cryopreservation vials, which were appropriately labelled using an ethanol indelible marker. Vials were placed into a small styrofoam box stuffed with paper towels which was then placed at -80°C to facilitate slow freezing of the cells, so as to maintain maximum possible cell viability. After 24 h, the vials were transferred into a liquid N_2 tank for storage.

Revival of frozen ES cells was achieved by first rapidly defrosting a vial of cells in a 37°C water bath until just completely thawed. The vial of cells is then quickly but thoroughly cleaned with 70% ethanol and place in the flow hood for manipulation of the cells. Using a cotton plugged Pasteur pipette, cells were transferred into a 15 mL conical tube containing 5 mL of ES cell growth medium, and then centrifuged at 200 x g for 10 min to facilitate removal of the DMSO containing medium. The resulting cell pellet was then resuspended in pre-warmed growth medium and distributed drop-wise using a cotton plugged Pasteur pipette into a 60 mm tissue culture dish containing a mouse embryonic fibroblast (MEF) feeder layer and an appropriate volume of warmed growth medium. After 24 h of incubation at 37°C in a humidified 5% CO_2 incubator, the cells were examined and either passaged onto a fresh feeder layer containing dish, or fed by exchanging depleted medium with fresh pre-warmed growth medium.

For transfer of ES cells from feeder supported to feeder free culture, a pre-plating step must be performed to separate the ES cells from the contaminating MEF cells. Cultures were trypsinized as described in section 2.3.2, and then transferred into a non-gelatinized cell culture dish. The cells were incubated in this dish for 30-45 min at 37°C. The growth medium containing the ES cells was then collected and transferred into a freshly gelatinized tissue culture vessel, while the dish containing the MEFs was discarded.

2.3.5 Routine passaging, freezing, and revival of P19 and HEK 293T cells

For P19 cells, routine passaging was carried out as described for ES cells in section 2.3.3, except that P19GM was used and split ratios were typically 1:5 to 1:8 every 2-3 days.

Routine passaging of HEK cells was carried out as previously described in section 2.3.3, but with several modifications. First, gelatinized culture dishes were not required. The trypsinization step was carried out for a short period (1-2 min) at room temperature with gentle agitation of the culture dish followed by addition of at least one trypsin volume of HEKGM to stop the trypsinization reaction. The cell suspension was then further mechanically dissociated using a plugged 10 mL tissue culture pipette, which has a wider bore than a Pasteur pipette. HEK cells were typically split at ratios of 1:8 to 1:10 every 2-3 days.

Freezing of both cell lines was conducted as described in section 2.3.4 except that each cell line's standard growth medium was substituted for ES cell medium wherever used. Revival of each cell line was also done as previously described but without use of MEF feeder layers, again substituting the appropriate type of growth medium wherever necessary.

2.3.6 Neural differentiation using the Bain ES cell neural differentiation method

Neural differentiation of ES cells was carried out following the Bain ES cell neural differentation method as previously described[116,158]. In short, ES cells are cultured as aggregates, also known as embryoid bodies (EBs) for four days in the absence and four days in the presence of 0.5 μM all-trans retinoic acid (RA). This process is known as 4-/4+ induction.

ES cells at 60-80% confluence are first trypsinized and mechanically separated as described in section 2.3.2, until single cells and small (2-5 cell) aggregates in suspension are achieved. This cell suspension is seeded for EB formation in non-gelatinized bacteriological petri dishes containing ES cell control differentiation medium (RA-), and allowed to incubate in a 37°C humidified 5% CO_2 incubator for four days (4-), with addition of fresh medium as necessary. After four days in culture without RA, the medium was changed to ES cell neural differentiation medium (RA+), and incubation was continued for an additional four days (4+) again with addition of fresh medium as necessary. Spontaneously differentiating parallel control cultures continued to be maintained in control differentiation medium. Following this 4-/4+ induction, the

aggregates are dissociated using trypsinization and gentle mechanical shearing until single cells and small aggregates are again achieved. The resulting cell suspension is then distributed into gelatinized cell culture dishes containing control differentiation medium, for neural outgrowth over five days. Stages of this differentiation protocol that are of note are 4-/3+, at which time the induced cells are neural precursors, and 5 days post induction (dpi) at which time maximal neural outgrowth is observed.

2.3.7 Neural differentiation of P19 cells

Neural differentiation of P19 cells was carried out using the Bain method as previously described for ES cells in section 2.3.6, except that P19IM and P19IM supplemented with 0.5μM RA were substituted for the ES differentiation media whenever used[117,158]

2.4 RNA Based Methods

2.4.1 Isolation of RNA

Total RNA was prepared from different sources including cultured cells, mouse embryos, and adult mouse tissues using TRIzol reagent (Invitrogen) according to manufacturer protocols with several changes: homogenization of samples in TRIzol reagent was carried out in standard microcentrifuge tubes for total volumes of up to 1 mL, and in round bottom polypropylene tubes for larger volumes. After homogenization and 15 min incubation at room temperature, samples were stored at -80°C if immediate processing was not required. Phase separation with 20%v/v chloroform was performed according to protocol for embryo and cultured cell samples, but twice for adult tissue samples. Precipitation of the RNA from the resulting aqueous phase was achieved by addition of a volume of RNAse free isopropanol equal to that of the aqueous phase, gentle mixing, and incubation at -20°C for 30-60 min. Following centrifugation and washing of the RNA pellet in 1 mL of RNAse free 75% ethanol, the RNA was redissolved in DEPC treated ddH_2O with warming to 55°C for 10 min, and then chilled to 4°C overnight prior to further processing.

Concentration of prepared RNA was estimated using spectrophotometry. All RNA samples prepared were treated with RNAse free DNAse I (Invitrogen) to eliminate any lingering contamination with DNA. For treatment of each µg of RNA, 10 units of DNAse I in a total volume of 10 µL 1X DNAse I buffer (supplied as a 10X stock) was used. The reaction was allowed to proceed at room temperature for 15 min, after which the DNAse I was inactivated by the addition of RNAse free EDTA to a concentration of 2.5 mM and incubation at 65°C for 10 min. If increased RNA concentration was required after DNAse I treatment, samples were precipitated using RNAse free sodium acetate (NaOAc) and ethanol using standard lab procedures. Samples which needed to be accessed on a regular basis were stored as prepared at -80°C. Samples for long term storage were stored at -80°C as ethanol precipitates.

For preparation of poly A+ RNA, the Oligotex Direct mRNA midi kit system (Qiagen) was used according to the manufacturer protocol.

2.4.2 Analysis of RNA

RNA was routinely analyzed for concentration, purity, and quality using a variety of methods. For general analysis of concentration and purity, spectrophotometry ($A_{260/280}$) was carried out. RNA samples destined to be reverse transcribed and used in real-time PCR applications were then subject to more sensitive RNA quantitation using a RiboGreen RNA quantitation reagent kit (Molecular Probes). Samples stained with RiboGreen were compared to a standard curve using a fluorescent microplate reader.

Quality of prepared RNA was examined using either standard agarose gel electrophoreseis (described in section 2.7.3) under RNAse free conditions, or formaldehyde gel electrophoresis. For all electrophoresis of RNA, RNAzap RNAse decontamination solution (Ambion) or 3% hydrogen peroxide treated apparatus were exclusively used. Formaldehyde gels were typically prepared with 1% agarose. For each 100 mL of gel needed, 1 g of agarose was melted in 72 mL of ddH_2O. 10 mL 10X MOPS buffer (0.2 M MOPS, 50 mM NaOAc, 10mM EDTA, in DEPC ddH_2O) and 18 mL formaldehyde were then added. The gel was cast in a fume hood and allowed to solidify for several hours beneath a tent of aluminium foil. Set gels were submersed in a

buffer tank containing formaldehyde gel running buffer composed of 83 mL formaldehyde and 100mL 10X MOPS buffer per litre in DEPC treated ddH_20.

Sample preparation for formaldehyde gel electrophoresis was as follows: 32 µL of master mix containing 4 µL 10X MOPS buffer, 7 µL formaldehyde, 20 µL deionized formamide, and 1 µL 1 mg/mL EtBr per sample was added to 10µg RNA in a total volume of 9 µL DEPC treated ddH_20. RNA size markers were also prepared in this way. Samples were heated at 55°C for 15 min and then quickly chilled on ice. 5 µL of RNAse free loading dye containing xylene cyanol and bromophenol blue was added to each sample prior to loading onto the prepared gel. Gels were run at 100V with occasional buffer mixing until the leading bromophenol blue dye front had travelled two-thirds the total gel length. RNA was then imaged in the gel using a UV lightbox, Kodak DC90 Zoom camera, and Kodak 1D v3.5 digital imaging software.

2.4.3 Reverse transcription

Reverse transcription of total RNA to cDNA was achieved using SuperScript II reverse transcriptase (Invitrogen) according to standard lab protocols. Up to 5 µg of RNA in a total of 6 µL DEPC treated ddH_2O was combined with 5 µL 50 ng/µL random hexamer primers, and 1µL 10mM dNTP mix. This mixture was incubated at 65°C for 5 min and then chilled on ice before adding 4 µL of 5X first strand buffer, 2 µL 0.1 M DTT, and 1 µL RNAse out (Invitrogen). The reaction was incubated at 25°C for 10 min followed by 2 min in a 42°C water bath. 1 µL of SuperScript II reverse transcriptase was then added to each reaction, which was allowed to further incubate at 42°C for one h. 15 min incubation at 70°C was allowed to inactivate the reverse transcriptase for termination of the transcription reaction. Matched reactions to which no reverse transcriptase was added were always carried out in parallel as controls.

2.4.4 Northern blotting and detection

Northern blotting was done according to the standard alkaline method commonly used in our lab. Formaldehyde gels were run and imaged with a fluorescent ruler as described in section 2.4.2. Gels were then trimmed, keeping the wells but discarding the marker lane, measured and soaked in ddH_2O for 30 min, followed by 20 min

equilibration in 3mM NaOH. During this time, four sheets of 3MM Whatman paper, positively charged nylon membrane (Hybond N+), and a stack of paper towels were cut to the size of the gel. Longer pieces of 3MM paper were also prepared to act as wicks for the blotting solution. The transfer apparatus used consists of a reservoir of 3mM NaOH over which a glass plate is placed, with the 3MM paper wicks extending from the glass plate into the reservoir. Onto the wicks, two pieces of 3mM NaOH saturated 3MM paper placed, and the gel was laid well side down onto the paper, ensuring no bubbles were present between any layers. Remaining layers were added in the following order, again ensuring that no air bubbles were present: nylon membrane pre-wet in ddH_2O, remaining two pieces saturated 3MM paper, cut paper towel stack, a glass plate for application of even pressure, and a weight such as a full 500 mL solution bottle. Blotting was allowed to proceed in this way overnight. At the end of the transfer, the blot is disassembled and the positions of the gel wells marked on the membrane before neutralizing briefly with 2X SSC. Blots were sandwiched between 3MM paper and wrapped in aluminium foil for 30 min of baking at 80°C. Blots could remain wrapped until ready for use.

For detection of RNA species of interest, ^{32}P labelled probes were used. Plasmid templates used for probe making were pCR-4.2, pBS-4.2, or pBS-3B6 (described in section 2.7.9), or sub-fragments thereof. For RNA probes, plasmids were linearized via digestion with an appropriate restriction endonuclease, and purified using either a QIAquick PCR cleanup kit (Qiagen), or organic extraction followed by ethanol precipitation. RNA probes were produced by *in vitro* transcription in the presence of high specific activity (about 8000 Ci/mmol and 10 mCi/mL) [α-^{32}P]UTP (GE Healthcare) using a MAXIscript *in vitro* transcription kit (Ambion) according to manufacturer protocols. DNA probes were produced via random primed labelling of PCR generated DNA fragments in the presence of high specific activity (about 6000Ci/mmol and 10mCi/mL) [α-^{32}P]dCTP (GE Healthcare) according to standard lab protocol. For both probe types unincorporated radiolabel was removed by filtration through a ProbeQuant G-50 micro column (GE Healthcare), and radiolabel incorporation was estimated by

comparison of the purified probes to the label retained in the filter after centrifugation in similar proximity to a Geiger detector.

Hybridization and wash steps were carried out in roller bottles in a rotating hybridization oven. ULTRAhyb hybridization buffer (Ambion) was used according to manufacturer protocols Membranes prepared as described in section 2.4.3 were prehybridized in enough hybridization buffer to keep the membrane uniformly wet, at 68°C or 42°C for RNA or DNA probes respectively, for 2-4 h. Radiolabelled probes were then added to the hybridization buffer, and hybridization to the membrane was allowed to proceed overnight at 68°C for RNA probes or 42°C for DNA probes. DNA probes were denatured by boiling for 5 min followed by quick chilling on ice prior to use in hybridization. Washes were performed at 68°C (RNA) or 42°C (DNA) as follows: 2x15 min in 2X SSC with 0.1% SDS, followed by 2x15 min in 0.2X SSC with 0.1% SDS. Following the final wash, excess liquid was drained from the membrane and it was smoothly wrapped in plastic wrap for autoradiography. The membrane was exposed to Kodak BioMax MS film at -80°C for exposures ranging from 24-72 h.

2.4.5 In situ hybridization

In situ hybridizations were done using digoxigenin (DIG) labelled RNA probes against 12.5 dpc mouse embryo sagittal sections. Plasmid templates used for probe making were pCR-4.2, pBS-4.2, pBC-3B6, or pBS-3B6 (described in section 2.7.9), or sub-fragments thereof. Two different methods were used for sectioning embryos: paraffin sectioning, and cryostat sectioning. Paraffin sections were prepared according to standard protocols used in the lab. Cryostat sections were a generous gift from the lab of Dr. Carol Schuurmans (University of Calgary).

To generate DIG labelled probes, plasmids were first linearized via digestion with an appropriate restriction endonuclease, and purified using either a QIAquick PCR cleanup kit (Qiagen), or organic extraction followed by ethanol precipitation. *In vitro* transcription was carried out using a MAXIscript *in vitro* transcription kit (Ambion) according to manufacturer directions and in the presence of DIG-11-UTP (Roche). Both antisense and sense probes were generated, so as to provide controls. Probes were tested

for length via agarose gel electrophoresis under RNAse free conditions. Probes were also tested for DIG incorporation using a spot test prior to use. The DIG nucleic acid detection kit (Roche) was used according to manufacturer protocols.

Hybridization to paraffin embedded sections was carried out following standard lab protocols. In short, slides of paraffin embedded sections were dewaxed in xylene, rehydrated through a nuclease free ethanol series, and then bleached with a 6% hydrogen peroxide solution before being treated with proteinase K. Slides were then treated with a solution of 1.32% triethanolamine and 0.5% acetic acid, followed by prehybridization in a humidified chamber for 2 h at 60°C. Prehybridization and hybridization was done in a prewarmed solution consisting of 50% deionized formamide, 5X SSC, 1% SDS, 50 mg/mL heparin, and 50 mg/mL Torula RNA. Probe was added at 0.5 ng/μL for hybridization, which was allowed to proceed overnight at 60°C. The following washes were then completed in prewarmed solutions: 2x30 min in 50% deionized formamide, 2x SSC at 65°C; 3x5 min in 2x SSC at 37°C; 30 min in 2x SSC with 20 μg/mL RNAse A at 37°C; 2x30 min in 50% deionized formamide, 2x SSC at 65°C; 3x5 min in TBST at room temperature. Probe detection was then completed using the DIG nucleic acid detection kit (Roche) according to manufacturer protocols.

Hybridization to cryostat sections were completed by Ms. Natasha Klenin in the lab of Dr. Carol Schuurmans, according to their stansard lab protocols. Images of stained embryos were captured using a Zeiss Axioplan 2i microscope with a Hamamatsu OrcaER digital camera and Axiovision 4.2 software.

2.5 PCR Based Methods

2.5.1 Primers

Unless otherwise indicated, primers for all PCR based methods were designed using Primer3 (http://frodo.wi.mit.edu/)[159]. The name and sequence for each primer is provided in Table 1, which will be referred to throughout the text. Primers for use in real-time PCR applications and the amplimers they create were analyzed using MFold[160] to ensure that no interfering secondary structures would be formed under PCR conditions

Table 1: Primers used during this project

Primer Name	Sequence, 5′-3′
1A9F	cagcgtataggaggtgatgg
1A9R	gctatccaagcccagaacg
1B4F	aggaagaggaactgtgtgtgg
1B4R	agaggatcatgggtttgagg
1H7F	ccgctctagttatgtcagttcc
1H7R	aaaacgtcctccccatcc
2B5F	cggtcaggcggatattaagc
2B5R	atgcagaggtgatgcacagg
2F10F	tgagcgaaacacaaacatcc
2F10R	agggagcacttggaaaaagg
3A6F	ggacaaaggccactacttcg
3A6R	agtgctgcgttttcttagcc
7G2F	cgcccagattcctacaagc
7G2R	aggcaaagaagcctgtgtcg
8E1F	tctgcacggatgaccttagc
8E1R	tataccgagatggcctttgg
12B11F	gcactttctgttggggtagg
12B11R	ccccttctgctcatatttcc
12G1F	tctgctctccctgaaagtgg
12G1R	cagaggaaggtggaaactgg
3B6-1F	cgtcaggaagcctgaaggta
3B6-1R	tggatctgcatcacagatgg
3B6-4F	tccagaaggttgaatcacagt
3B6-3R	ttggtgctgaagaaagtctacc
3B6-6F	aaagggacctggtgaaacc
3B6-6R	accgtttctttggcattgg
3B5-7F	aaaaggggttctgtatgagagg
3B6-7R	ggtggtccttcaacaacagg
3B6-7Rb	cagactgggtggtaatggac
3B6-9R	ccatgtcacctgatgttagtcc
3B6-11R	aaacctgggctgtttgtcc
3B6-13F	gtccagcagcccattttacc
3B6-13R	tctgctggctctgttgtagc
3B6-16R	tcctgcgaagtttcagtgc
3B6-19R	aatgcaaaatctccagtgc
3B6-20F	ggggaaattgaggactgtca
3B6-21R	agcttcttcagtgtccacagc
3B6-22R	ccaagaacgagattcattgtcc

3B6-1FBgl	aattagatctgatggtttctaagatgatc
3B6-1FSal	gcgagtcgactgatggtttctaagatgatc
3B6-1FNot	aattgcggccgccaccatggtttctaagatgatc
3B6-1FBam	aattggatcctaatggtttctaagatgatc
3B6-7RNot	gcgcggccgctggtccttcaacaacaag
3B6-11FSal	gcggtcgactccagaagaaggttcagtttgg
3B6-11FBam	aattggatcctccagaagaaggttcagtttgg
3B6-13RNot	gcgcggccgcactgactgcttcacaacagg
3B6-16FXba	cgctctagacacattgaaacacagaaga
3B6-20RBam	cgcggatccaatctgacagtcctcaatt
3B6-22RNot	aattgcggccgcccaagaacgagattcattgtcc
3B6-22RBam	aattggatccccaagaacgagattcattgtcc
3B6-22RSal	gcggtcgacccaagaacgagattcattgtcc
3B6-P363-R	cgcgtcgacttaattaattaaggggtctaagattcacagg
3B6-G461-F	aattgcggccgccaccatgggactaacatcaggtgacatgg
3B6-W1005-R	cgcgtcgacttaattaattaccaagaacgagattcattgtcc
3B6-E652-R	cgcgtcgacttaattaattattctgctggctctgttgtagc
3B6-D736-R	cgcgtcgacttaattaattaatcctgctgaagtttcagtgc
3B6-H748-F	aattgcggccgccaccatgcacattgaaacacagaagatgc
3B6-In18F	tgtgttaggatagagaggaaagagg
3B6-In18R1	caagaacaaaaccagaaagagg
3B6-In18R2	taggccagcatttcacagc
Geo-1R	agtatcggcctcaggaagatcg
Geo-2R	attcaggctgcgcaactgttggg
λGT10-F2	cagcctggttaagtccaagc
λGT10-R2	ggtggcttatgagtatttcttcc
λGT10-F	acgaagttcagcctggttaag
λGT10-R	cttatgagtatttcttccagggta
λGT11-F	ggtggcgacgactcctggagcccg
λGT11-R	ttgacaccacaccaactggtaatg
pT1ATG	gccagggttttcccagtcacg
FLAG-R	tcgtcgtcatccttgtaatcc
GAPDH-F	gcacagtcaaggccgagaat
GAPDH-R	gccttctccatggtggtgaa
T3	cgcaattaagggtcactaaag
T7	taatacgactcactataccc
pGex5′	cgtttggtgctggcgac
pGex3′	cgggagctgcatgtgtcagag
L4440F	cagtgttctttcctgcgttatc
L4440R	ctgcaaggcgattaagttg
M13R	agcggataacaatttcacacagga
M13F	gtaaaacgacggcaagt

at the desired annealing temperatures. Primers were received as lyophilized samples and resuspended in sterile ddH_2O to a concentration of 100µM. For use in PCR based applications working stocks of 10µM were made.

2.5.2 General PCR Method

A general PCR method giving very consistent results was established using a home-made 2x PCR supermix and recombinant *Taq* polymerase (Invitrogen) for regular use. To make the home-made 2x PCR supermix 2mL 10x PCR Buffer (Invitrogen), 1.2mL 50mM $MgCl_2$, 40µL each 100mM dNTP (G,A,T,C; Invitrogen), 100µL Tween-20, and 1mL glycerol were combined with sterile ddH_2O in a total volume of 10mL. Supermix was stored at -20°C in 1mL aliquots. For each standard 25µL PCR reaction needed, 13.5µL of supermix was combined with 12.5µL sterile ddH2O, 0.5µL each 10µM primer, and 0.2µL *Taq* polymerase. 24µL of this master mix was then aliquoted into each reaction tube, and 1µL of the desired template was added. In reactions requiring more or less than 1µL of template the amount of water added to the master mix was adjust to compensate. For colony screening using PCR, colonies were picked using a sterile pipette tip, touched to a replicate plate, and then directly added to tubes containing the desired reaction mix by pipetting up and down once. If increased amounts of reaction product were desired, the entire reaction could be scaled up as needed. Thermocycling was carried out in an iCycler PCR machine (BioRad). A typical themocycling protocol is 5 min at 95°C then 30 repetitions of 30 seconds at 95°C, 30 seconds at 57°C, and 1 min at 72°C. This is followed by a 10 min incubation at 72°C and cooling to a hold temperature of 16°C. This typical protocol was adjusted for varying primer annealing temperatures and amplimer lengths depending on each particular reaction.

2.5.3 RT-PCR

RT-PCR was routinely carried out using the general PCR method described in the preceding section. cDNA prepared as described in section 2.4.3 was commonly diluted ten fold for use as template in these reactions. All reactions were carried out in matched pairs: one reaction using primers for detection of the gene of interest, and the other using primers GAPDH-F and GAPDH-R, for detection of GAPDH as an internal control for

normalization of cDNA levels. GAPDH has been previously shown to be an appropriate internal standard for use in studies of differentiating ES cells[161]. Thermocycling was carried out as described above, however an appropriate number of cycles to be completed was determined for each target amplimer in order to prevent the reaction from entering the plateau phase. At the plateau phase, reactions generally become saturated, preventing reliable densitometry from being conducted on the resulting agarose gels.

2.5.4 High-fidelity PCR

High fidelity PCR was regularly completed using ProofStart DNA polymerase (Qiagen) to generate fragments for cloning. A typical 50 µL high fidelity reaction contained 5 µL 10x ProofStart reaction buffer, 1.5 µL 10mM dNTP mix, 5 µL each 10µM primer, 32.5 µL sterile ddH_2O, 1 µL ProofStart, and 1 µL template. The ProofStart thermocycling protocol is as follows: 95°C for 5 min; 30 repeats of 30 seconds at 95°C, 30 seconds at 51°C, and 3.5 min at 72°C; 10 min at 72°C; 16°C hold. A small sample of the resulting product was examined by agarose gel electrophoresis, while the remainder of the product was purified either using a QIAquick PCR cleanup kit (Qiagen), or by gel purification (section 2.7.5) in preparation for cloning.

2.5.5 Real-time PCR

Real-time PCR was carried out to more accurately detect subtle differences in gene expression. It was carried out using an iCycler iQ thermocycler, along with iQ Sybr Green Supermix (both BioRad). A typical reaction was made up of 12.5 µL of Sybr Green Supermix, 10.5 µL ddH_2O, and 0.5 µL of each 10 µM primer. Each reaction was set up as a master mix in triplicate, with matched GAP-DH and ddH_2O controls, in non-fluorescing 96 well PCR plates (BioRad). Every plate was also set up in triplicate, resulting in nine data points for each reaction carried out. Thermocycling was done using a four-step protocol, in which the data acquisition step, following the extension step, was carried out at a temperature several degrees lower than that of the target amplimer melting temperature. This works to eliminate incidental fluorescence from any primer dimers or spurious products that may be present in the reaction. Upon the completion of

the thermocycling protocol, a melt curve analysis was always carried out in order to ensure production only of desired reaction products.

Each new primer set introduced for real-time PCR was extensively tested both in standard PCR reactions and real-time PCR reactions to ensure their specificity and to establish their ability to produce the desired amplimer in a linear fashion for a wide range of starting template concentrations. Once validated, standard curves a ten fold serial dilution series of starting template were produced for each primer set, allowing for calculation of reaction efficiency. Though the efficiencies for each reaction performed were similar, attempts to equalize reaction efficiencies for each primer set were not made, as the $\Delta\Delta C_T$ method[162] was not used for the final analysis of the data. Rather, the Relative Expression Software Tool (REST), based on a mathematical model devised by M.W. Pfaffl at the Technical University of Munich, was used in order to minimize assumptions made in data analysis[163,164]

2.6 Bacterial Methods

2.6.1 Escherichia coli strains

Several strains of *Escherichia coli* (*E. coli*) were used during the course of this work. Strain DH5α (Invitrogen) was used for general purpose cloning and plasmid maintenance. DH5α is recombination deficient and also competent for *LacZ* α-complementation with plasmids encoding the *LacZ* α fragment, allowing for blue/white selection of desired clones. [139].

For protein expression in the prokaryotic system (section 2.10.4), *E. coli* strains Rosetta(DE3), BL21(DE3), and HMS174(DE3)pLysS (all Novagen) were used. Each of these strains are intended for use with expression plasmids under the control of either the T7 promoter or the T7/*lac* hybrid (*tac*) promoter, and also encoding the $lacI^q$ repressor. Such plasmids include pET (Novagen) and pGEX (GE Healthcare) vectors. Each of these strains carries the λ lysogen DE3 which encodes T7 polymerase under control of the IPTG inducible *lac* promoter[165]. Both Rosetta(DE3) and BL21(DE3) are *lon* and *OmpT* protease deficient. HMS174(DE3)pLysS carries the pLysS episome which

encodes T7 lysozyme for increased control of T7 polymerase production[166]. It is also recombination deficient allowing for stable propagation of recombinant constructs. Rosetta(DE3) carries the episome pRARE which encodes tRNAs that are rarely used in E. coli, therefore enhancing expression of eukaryotic proteins containing such codons[167]

Bacteriophage λ based methods (section 2.8) were carried out in *E. coli* strain LG75 or LE392 (Stratagene). LE392 is a permissive host encoding the *supF* tRNA, enabling suppression of amber stop codons where needed. LG75 is *supF°*, so *supF* must be supplied *in trans* in order to facilitate translation through an amber stop codon.

2.6.2 Media and routine growth of E. coli

For routine growth of *E. coli* Luria-Bertani (LB) was generally used, consisting of 10 g/L peptone, 5 g/L yeast extract, and 10 g/L NaCl. For preparation of media for plates, 15 g/L of granulated bacterial agar was added prior to autoclaving. For growth of *E. coli* carrying recombinant plasmids with either the *lac* or *tac* promoter, LB was commonly supplemented with 2% sterile glucose in order to help suppress leaky transcription from those promoters[139]. For bacterial strains to be used in the propagation of bacteriophage λ, LB supplemented with 1 g/L $MgCl_2 \cdot 6H_2O$ added prior to autoclaving was used. This medium will be referred to as LBM throughout this text. SOC medium consisting of 20 g/L tryptone, 5 g/L yeast extract, and 0.5 g/L NaCl with 20 mL 1M sterile glucose added after autoclaving, was used for recovery of bacteria post-elecroporation.

Antibiotics stocks for use in plasmid selection included ampicillin (amp) and kanamycin (kan) at 25mg/mL in ddH_2O, and chloramphenicol (cam) at 34 mg/mL in ethanol. Amp and Kan stocks were filter sterilized, and all stocks were stored at -20°C. IPTG and X-gal stock solutions were prepared for use in blue/white selection. IPTG was dissolved at 20 mg/ml in ddH_2O, filter sterilized, and stored at -20°C. X-gal was dissolved in dimethylformamide at 20 mg/mL and stored protected from light, also at -20°C.

All manipulation of *E. coli* was done using standard aseptic laboratory methods. Liquid culture of *E. coli* was routinely done in 2 mL volumes in sterile glass culture

tubes, or in 10 mL volumes contained in sterile 50 mL conical tubes (Falcon). LB with appropriate supplementation was inoculated from a single colony picked off of a selective medium plate using a sterile pipette tip. For the purpose of screening many colonies during cloning, the same pipette tip was also used to make a small streak on a replicate plate for later use. When conical tubes were used as the culture vessels, the caps were not tightened, but rather were secured with tape to allow adequate aeration of the growing *E. coli*. Liquid cultures were allowed to grow at 37°C with vigorous shaking at 260 rpm for 14-16 h. To make bacterial freezer stocks for long term storage, 0.6 mL of overnight liquid culture was added to 0.3 mL sterile 66% glycerol, mixed well and frozen at -80°C. Revival of glycerol stocks was achieved by transferring a small scrape of the -80°C stock to an appropriate selective plate and spreading using a sterile platinum inoculation loop. The plate was then incubated at 37°C for 12-14 h to allow for growth of colonies.

2.6.3 Transformation of E. coli using electroporation

Bacterial strains for growth and selection of recombinant plasmids were prepared for electroporation according to previously described methods[139], and stored at -80°C as 40 µL aliquots. Prior to electroporation, the needed number of aliquots was allowed to thaw on ice. 2µL of ligation reactions diluted two fold with ddH_20, or 0.5 µL of intact dilute plasmid DNA resuspended in ddH_20 was added directly to the aliquot of electrocompetent bacteria and this mixture was transferred into 0.2 cm electroporation cuvettes (BioRad) and kept on ice. Electroporation was conducted using a GenePulser (BioRad) under the following conditions: 2.5 mV, 25 µF, and 400Ω. After electroporation, 500 µL SOC was immediately added into the cuvette. The mixture of bacteria and SOC was then transferred into a sterile glass culture tube. 30-40 min of incubation in a 37°C shaking incubator was allowed for bacterial recovery and expression of selectable marker genes prior to plating on pre-warmed selective medium. If blue/white selection was possible, 40 µL each of stock X-gal and IPTG were pre plated and allowed to absorb into the medium. Typically, a range of amounts of the electroporated bacteria were plated, ranging from 1% to 50% depending on the plasmid DNA used. Plated *E. coli* were allowed to grow for 14-16 h at 37°C.

2.7 DNA Based Methods

2.7.1 Isolation of plasmid DNA

Plasmid DNA was typically isolated from 2-5 mL of overnight bacterial cultures using a QIAprep Spin Miniprep Kit (Qiagen) according to instructions provided with the kit. Elution from the kit spin columns was always done with 10 mM Tris (buffer EB) unless the plasmid DNA was destined for use in sensitive applications such as transfection of mammalian cells, *in vitro* transcription, or coupled *in vitro* transcription and translation. In these cases, elution from the column was done using sterile nuclease free ddH_2O.

A modified method of plasmid DNA preparation using components of the QIAprep Spin Miniprep Kit was developed for preparation of plasmid for batch colony screening by restriction digestion during cloning. The method is analogous to the alkaline lysis method,[139] but does not require the use of organic extraction steps. The method also does not use the columns provided with the kit, taking advantage of the excess solutions provided and saving columns for more critical application. Bacterial cultures are pelleted and then resuspended in 250 µL Buffer P1. As in the normal kit protocol, 250 µL Buffer P2 is added to lyse the cells, and 350 µL Buffer N3 is added to neutralize and clear the bacterial lysates. After centrifugation at 13000x g for 10 min, the clear supernatant containing the plasmid DNA is transferred to a fresh microcentifuge tube. The DNA is isolated from this supernatant by ethanol precipitation and subsequently resuspended in 10mM Tris. Plasmid DNA prepared by this method has been found to consistently be of good quality as judged by agarose gel electrophoresis and spectrophotometric analysis ($A_{260/280}$), and cuts cleanly using commonly used restriction endonucleases. Use of plasmid DNA prepared by this method has not been attempted in PCR or sequencing reactions.

2.7.2 Isolation of genomic DNA

Genomic DNA was isolated from ES cells grown in feeder free culture to ~80% confluence. Growth medium was aspirated from the cells which were then rinsed twice with DPBS. Collection of the cells was done by adding 125 µL/cm^2 ES cell lysis buffer

directly to the culture dish and incubating at 37°C for 45-60 min. ES cell lysis buffer consists of 100 mM NaCl, 0.5% SDS, 20 mM Tris pH 7.5, and 50mM EDTA. Cell lysates were collected as 500 µL aliquots into microcentrifuge tubes and treated with 10 µL 10 mg/mL protinease K at 55°C overnight. Each sample was then extracted with an equal volume of equilibrated phenol, followed by extractions with phenol:chloroform until no visible interphase remained after centrifugation. A final extraction using chloroform only was done, and then the DNA was precipitated from the aqueous phase using 20 µL 5M NaCl and 2 volumes 100% ethanol. DNA was collected by centrifugation for two min, washed once in 70% ethanol, and allowed to briefly air dry. Pellets were resuspended in 40-80 µL of 10 mM Tris pH 8.0.

2.7.3 Agarose gel electrophoresis

Agarose gel electrophoresis was carried out according to standard lab practices. For general use, 1% agarose gels were prepared in TAE buffer with 0.5µg/mL of ethidium bromide was added after microwaving. For resolution of fragments under 300 bp, 2% gels were used, also made in TAE buffer with added ethidium bromide. All gels were run in TAE buffer. If DNA fragments were to be isolated from the gel after running, fresh TAE buffer was sued and the buffer tank was washed prior to filling. Loading dye containing xylene cyanol and bromophenol blue was most commonly used, but for resolution of small fragments, Orange G loading dye was used to circumvent obscuring of small bands by the bromophenol blue dye front. Gels were typically run at 100V until either the bromophenol blue dye front had migrated three quarters the length of the gel, or the Orange G dye front had run to the end of the gel. All gels images were captured using a UV lightbox, Kodak DC90 Zoom camera, and Kodak 1D v3.5 digital imaging software.

2.7.4 Restriction digestion

Restriction digestion was routinely carried out during the course of cloning protocols and for the purpose of screening recombinant plasmids. The most common restriction endonucleases used during the course of this project were *Sal I, Not I, Bam HI, Eco RI,* and *Xcm I.* Digests were completed in the reaction buffers supplied by the

nuclease manufacturers, with the addition of BSA when necessary. For double digests, reactions were carried out in buffer conducive to maximum activity of both the enzymes in question. If the enzymes required very different buffer conditions, digestion with each enzyme was carried out sequentially, using the lower salt buffer condition first and then increasing the volume and adjusting the buffer to the higher salt condition. Alternatively, reaction cleanup was done using a QIAquick PCR cleanup kit (Qiagen) between sequential digestion steps, to facilitate buffer changes. A typical restriction digestion reaction included 1 µg DNA in 1x reaction buffer with 0.5 µL of enzyme. Reactions were allowed to proceed at 37°C unless otherwise prescribed for a particular enzyme for 2-3 h.

2.7.5 Gel purification

Purification of DNA fragments from agarose gel was routinely completed during the course of cloning protocols. For general cloning purposes, fragments were separated via agarose gel electrophoresis as imaged as described above, before the desired fragment was cut from the gel using a new razor blade. DNA was then extracted from the gel using a QIAquick Spin Gel Extraction kit (Qiagen), eluting the column with 10mM Tris.

For difficult cloning procedures fragments were gel purified using a freeze and squeeze method. Fragments were separated via gel electrophoresis in a low-melt agarose gel, imaged, and excised from the gel using a new razorblade. The gel slice was placed into a microcentrifuge tube and weighed. An equal volume of TE was added to the gel, which was then melted at 65°C with occasional vortexing. The resulting solution was immediately transferred into a dry ice ethanol bath for flash freezing. Once frozen, tubes were removed from the bath and allowed to thaw at room temperature after which they were then centrifuged for 10 min at top speed. Supernatants were transferred to fresh tubes. To the old tubes still containing the agarose, 100 µL of TE was added, combined with the agarose, centrifuged, and the supernatants again collected. DNA was precipitated from the supernatants by adding $1/10^{th}$ volume of 3M NaOAc and 2 volumes 100% ethanol, with 10µg of glycogen added as a carrier. After precipitation at -20°C for 30 min, and centrifugation at top speed for 30 min, resulting DNA pellets were washed

with 70% ethanol and air dried briefly. Pellets were resuspended in an appropriate volume of 10mM Tris, and subjected to quantitation via spectrophotometry.

2.7.6 Modification of fragment ends

Ends of DNA fragments commonly needed to be modified for the purpose of cloning methods. Modifications frequently used include addition of adenosine (A) overhangs to blunt ended fragments, end polishing of fragments, and phosporylation or dephosphorylation of fragments.

The terminal transferase activity of *Taq* polymerase was taken advantage of to add A overhangs to blunt ended DNA fragments, such as those produced using high fidelity PCR (section 2.5.4). This method was of particular utility for TA cloning of PCR generated fragments. A typical reaction consisted of 10 µL 2x PCR supermix (described in section 2.5.2) and 500 ng DNA in 10 µL ddH_2O, with 0.2 µL of *Taq* polymerase. The reaction was incubated at 72°C for 15 min and then cleaned up using a QIAquick PCR cleanup kit (Qiagen).

End polishing of DNA fragments was typically completed using T4 DNA polymerase (Invitrogen). Reactions were carried out directly after restriction digestion, before fragment purification, as T4 polymerase retains activity in restriction buffers. To the digest in question, 2µL 10 mM dNTP mix and 1 µL T4 DNA polymerase were added. The reaction was allowed to proceed for no longer than 30 min in a 12°C water bath.

Dephosporylation of linearized vectors was routinely done via treatment with calf intestinal alkaline phophatase (CIP; Roche). CIP treatment was also carried out immediately following restriction digestion by addition of 1 unit of CIP and incubation at 37°C for 30 min. This was then repeated with the addition of another unit of CIP and 30 min more incubation at 37°C. The reaction was stopped by addition of $1/100^{th}$ volume 500 mM EDTA and incubation at 75°C for 15 min.

Phosphorylation of cloning inserts such as double stranded synthetic oligos was carried out using T4 polynucleotide kinase (PNK; Roche). A typical 50µL reaction consisted of 0.1µg/µL DNA in 43.5 µL, heated to 70°C for 5 min and then chilled on ice before adding 5 µL 10x kinase buffer, 0.5 µL 20 mM ATP, and 1 µL T4 PNK. The

reaction was allowed to proceed for 30 min at 37°C, and then was heat inactivated for 10 min at 70°C.

2.7.7 Ligation

Ligations of appropriately digested and modified vectors and inserts were completed using 50-100 ng total of DNA in as small a reaction volume as possible, generally 10-15 µL, to facilitate molecular interactions. Insert to vector molar rations of 3:1 or 5:1 were used, as well as insert free control reactions. A 10µL ligation reaction used 2 µL 5x ligase buffer and 1 µL T4 DNA ligase (Invitrogen). Ligations were incubated either at 16°C for 2-4 h or 4°C 14-16 h.

TOPO TA cloning using DNA Topoisomerase I based ligations were conducted using a TOPO TA cloning kit (Invitrogen), with pCR2.1-TOPO, according to manufacturer protocols.

2.7.8 Sequencing

DNA sequencing reactions were carried out using Applied Biosystems BigDye terminator sequencing kits and capillary electrophoresis on an ABI 3730S Automated sequencer by the University of Calgary sequencing core facility. Samples were submitted to the facility as 50-100 ng/kb of template plus 3.2 pmol of primer premixed in a total volume of 12 µL. Sequencing from λ phage DNA was carried out by submitting 1-2 µg DNA with λGT10 or λGT11 primers as needed. Primers commonly used for sequencing plasmid inserts included T3, T7, pGEX-3′, pGEX-5′, M13F and M13R. For sequencing of long inserts, custom primers internal to the insert were also used to walk the entire sequence length. Constructs used for expression of mSe70-2 were completely sequenced from promoter through coding sequence multiple times to ensure accuracy of the construct. For each sequence returned, the trace files were examined for ambiguous base calls and end clipped for quality using Gene Tool (BioTools Inc). Wherever possible, contiguous sequences were assembled, also using Gene Tool, and examined for sequence discrepancies. Alignment of the sequence with known or predicted sequences was done using the NCBI BLAST 2 sequences (bl2seq) utility with appropriate parameters.

2.7.9 Recombinant DNA constructs

Numerous constructs were produced during the course of this project. Not every construct will be described in detail here, but rather those that were of central utility to the project will be described. All constructs were sequence verified, and maintained in *E. coli* strain DH5α.

Inserts from λ phage clones isolated in hybridization screens were amplified from purified λ DNA using appropriate λ primers. They were then TA cloned into pCR2.1-TOPO. The insert 3B6 clone 4.2 in pCR2.1-TOPO (pCR-4.2) was liberated via restriction digestion with *Spe I* and *Not I* and ligated into similarly digested pBC-SK$^+$ (Stratagene). The resulting construct was pBC-4.2.

To confirm and explore the predicted gene structure of 3B6, numerous standard RT-PCR reactions were carried out using various combinations of 3B6 nested primers. Resulting amplimers from each reaction were cloned into pCR2.1-TOPO. Each resulting clone was numbered in the following fashion: pCR-clone*n* where *n* represents the assigned number.

To accurately clone the full length coding sequence of 3B6, including exon 14 (see Appendix A for sequence), a multi step strategy was employed. It is important to note that any successful cloning of the 3B6 full length coding sequence could only be achieved in pBS II KS$^+$ (Stratagene) or pBC SK$^+$ if the orientation of 3B6 was opposite to that of the *LacZ* reporter hence preventing any semblance of production of mSe70-2 in the host bacterial strain. As exploration of the gene structure was done using PCR with *Taq*, not all of the cloned 3B6 fragments were free of small errors in their sequence. No pCR clone containing the full length coding sequence without errors was generated, and generating this sequence from cDNA using high fildelity PCR proved troublesome. The construct pCR-clone1-7 contained the coding sequence starting with the ATG codon in exon 1 through exon 7, and was free of any errors. It also contained a *Sal I* restriction site immediately preceeding the ATG. Within exon 7 lies a naturally occurring *Sma I* site. This site was taken advantage of by generating a fragment of 3B6 overlapping exon 7 and extending to the end of the known coding sequence in exon 22a. This fragment

was generated from cDNA using high fidelity PCR with primers 3B6-7F and 3B6-22R. The resulting products were TA cloned into linearized pBSXcmI, a TA cloning vector developed by Mr. Bob Winkfein at the University of Calgary. This vector is a modified pBS II KS^+ with an *Eco RI* site closely followed by two adjacent *Xcm I* sites situated between the *Pst I* and *Eco RI* sites of the parental vector. Digestion of pBSXcmI with *XcmI* results in T overhangs flanked by *Eco RI* sites. Multiple resulting clones were sequenced until one in the desired orientation and containing the desired splice variant without any sequence errors was identified. This construct was designated pBSXcmI-7-22.

To complete the cloning of the full length coding sequence, pBSXcmI-7-22 and pCR-clone1-7 were sequentially digested with *Sma I* and *Not I*. This digestion liberated the insert from pCR-clone1-7 and linearized pBSXcmI-7-22, releasing a small portion of the vector. These two fragments were ligated resulting in the construct pBSXcmI-3B6.

For expression of mSe70-2 in the prokaryotic system, several constructs in the vector pGEX-5x-1 (GE Healthcare) were built. This vector expresses proteins of interest as C-terminal fusions with glutathione S-transferase (GST) under control of the *tac* promoter. Generation of these constructs proved to be highly challenging. pGEX-3B6 was constructed by excising 3B6 from pBSXcm1-3B6 with *Sal I* and *Pvu I* and ligating it into pGEX-5x-1 that had been similarly digested. Digestion of pBSXcmI-3B6 with *Pvu I* was necessary to cleave the pBSXcmI backbone as both it and the insert are approximately 3 kb and therefore cannot be separated from each other using gel electrophoresis. Cloning inserts consisting of exons 1-7, 1-13, and 11-22 were generated using high fidelity PCR using primers 3B6-1FSal and 3B6-7RNot, 3B6-1FSal and 3B6-13RNot, and 3B6-11FBam and 3B6-22RNot respectively. These inserts were subjected to digestion with the required endonucleases and ligated with appropriately digested pGEX-5x-1. The resulting constructs were pGEX-3B6-1-7, pGEX-3B6-1-13, and pGEX-3B6-11-22. To enable growth of bacterial strains containing these constructs, medium supplementation with 2% glucose was necessary.

Generation of constructs for CMV promoter driven expression of mSe70-2 in mammalian cells necessitated generation of the construct pBS-3B6 as an intermediate step. pBS-3B6 was made by producing a high fidelity PCR product made with primers 3B6-1FNot, which includes a Kozak consensus sequence, and 3B6-22RBam from pBSXcm-3B6. This product was digested to give cohesive ends and then ligated into similarly digested pBS II KS^+. A CMV driven expression construct was built using vector pCMV-Tag4a (Stratagene) which produces proteins of interest with a C-terminal FLAG tag. The 3B6 insert was released from pBS-3B6 via digestion with *Not I*, *Bam HI*, and *Pvu I*. pCMV-Tag4a was also digested with *Not I* and *Bam HI*. The inserts were ligated with the vector, and construct pCMV-3B6-FLAG was yielded.

A construct for producing mSe70-2 as a C-terminal fusion with enhanced green fluorescent protein (eGFP) was built using the vector pEGFP-C1 (Clontech). pBSXcmI-3B6 was digested using *Sal I* and *Pvu I*. The resulting 3B6 insert was end polished with T4 DNA polymerase and then digested with *Hind III*, giving one blunt and one cohesive end. The vector was linearized with *Bgl II*, polished with T4 DNA polymerase, then digested with *Hind III*, also resulting in one blunt and one cohesive end. The vector and insert were ligated, conveniently reconstructing a *Bgl II* site for screening. The resulting construct was pEGFP-3B6.

Constructs for T7 *in vitro* transcription and translation (TnT) were built using pBS II KS^+. T7-FLAG-3B6 was created with the assistance of the four component ligaton expertise of Dr. Ken Ito, University of Calgary. 3B6 insert was prepared by digestion of pBSXcmI-3B6 with *Sal I*, *Hind III*, and *Pvu I*. The FLAG tag component was made up of synthetic oligos which when annealed provided *Kpn I* and *Sal I* cohesive ends, a Kozak consensus sequence as well as the FLAG tag. The 3′ most component was also made up of synthetic oligos which when annealed provided *Hind III* and *Bgl II* cohesive ends in addition to a *c-myc* tag and stop codons in all 3 frames. The synthetic oligos used were as follows:

FLAG-S: caccatggattacaaggatgacgacgataaggg

FLAG-AS: tcgacccttatcgtcgtcatccttgtaatccatggtggtac

MYC-S: agcttagagcagaaactcatctctgaagaggatctgtagtaattaattaa

MYC-AS: gatcttaattaattactacagatcctcttcagagatgagtttctgctcta

The oligos were annealed as described above, while pBS II KS^+ was linearized with *Kpn I* and *Bam HI*. All four components were ligated simultaneously resulting in the completed T7-FLAG-3B6 construct.

T7-3B6 was made by first generating an amplimer by high fidelity PCR from pBS-3B6 using primers 3B6-IFNot, which includes a consensus Kozak sequence and 3B6-W1005-R which contains stop codons in all 3 frames followed by a *Sal I* site. This insert was digested with *Not I* and *Sal I* to generate cohesive ends, and ligated into similarly digested pBS II KS^+.

Vector based RNAi constructs were built in the pSUPER.gfp.neo vector (OligoEngine). Synthetic oligos were designed that would target the 3B6 transcript for silencing at two different sites. The oligos were designed such that when annealed, cohesive *Bgl II* and *Xho I* ends would result. The oligo sequences were as follows:

KD1a: gatccacacttccagacctatgtattcaagagatacataggtctggaagtgtttttttc

KD1b: tcgagaaaaaaacacttccagacctatgtatctcttgaatacataggtctggaagtgtg

KD2a: gatccaggcgttatgcattgatcattcaagagatgatcaatgcataacgccttttttttc

KD2b: tcgagaaaaaaaggcgttatgcattgatcatctcttgaatgatcaatgcataacgcctg

Lyophilized oligos were resuspended at a concentration of 3 mg/ml in 10mM Tris. They were annealed in 50 µL volumes consisting of 1 µL each of the oligo a and b, 43 µL ddH_20, and 5 µL 10x SuRE/Cut buffer H (Roche). Each was heated to 95°C in a heating block for 5 min, and then the heating block was turned off to allow slow cooling to room temperature over approximately 4 h. pSUPER.gfp.neo was linearized using *Bgl II* and *Xho I*, and the oligos were subsequently ligated into the vector. The resulting constructs were pSUPER-KD1, and pSUPER-KD2.

Remaining constructs used during the course of this work were gifts from several individuals at the University of Calgary: pCMV-eGFP, from Dr. Knut Woltjen; pCMV-FLAG-hTERT, from Dr. Nick Ting; T7-Ku80 and T7-FLAG-hTERT, form Dr. Tara Beattie.

A number of other constructs were also pursued during this project, but due to technical challenges or other constraints, efforts were abandoned. These constructs included but were not limited to the following: pET-3B6 for His tagged expression in the eukaryotic system; pEGFP-N1-3B6 (pEGFP-N1; Clontech) for expression of mSe70-2 as an N-terminal fusion with GFP in mammalian cells; pCMV-Tag5-3B6 (pCMV-Tag5a; Stratagene) for expression of mSe70-2 with a C-terminal *c-myc* tag in mammalian cells; pREP-NTAP-3B6[168] for expression of mSe70-2 in the *Schizosaccharomyces pombe* system, with expertise from the lab of Dr. Dallan Young; pSUPER-scram, a scrambled RNAi control; λD-3B6 for expression of mSe70-2 as a fusion with the D-capsid protein on the surface of bacteriophage λ.

2.8 Bacteriophage λ Methods

2.8.1 Preparation of plating cells

E. coli cell line LE392 or LG75 was grown in a 2 mL overnight culture in LBM at 37°C with vigorous shaking, as described in section 2.6.2. For preparation of plating cells, a 50 mL subculture in LBM was inoculated with 500 µL of the overnight culture, and allowed to grow at 37°C with shaking until an OD_{600} of 0.6-1.0 was reached, approximately 3-4 h. This culture was transferred to a 50 mL conical tube and centrifuged at 3000 rpm for 10 min. The resulting cell pellet was gently resuspended in 20 mL 10 mM $MgSO_4$ and stored at 4°C until use. At 4°C these cells were found to remain viable for more than two weeks.

2.8.2 Determination of λ phage library titer

A cDNA library prepared from 12.5 dpc mouse embryos in λGT10 was received as a scraping from a -80°C freezer stock in 1 mL SM buffer. SM buffer consists of 0.1 M NaCL, 5 mM $MgSO_4 \cdot 7H_2O$, 50mM Tris-HCl pH7.5 and 0.01% w/v gelatin, autoclaved in ddH_2O.

For initial determination of the phage titer, establishing the order of magnitude of the titer, spot titration was carried out according to standard lab protocols. 10 cm LBM

agar plates were allowed to pre-warm to 37°C and LBM top agar (LBM with 7.5 g/mL agar) was melted and equilibrated to 50°C. For each plate, 100 µL of LE392 plating cells was aliquoted into a sterile glass bacterial culture tube. 4 mL of the molten LBM top agar was added to the tube which was then mixed by rolling the tube between hands. The mixture was then immediately poured "slowly but quickly" onto the pre-warmed LBM plate and allowed to solidify at room temperature. 10 fold serial dilutions of the phage library were prepared in SM buffer. 10 µL of each dilution were spotted onto the pre-poured bacterial lawn and allowed to dry before inversion of the plate and incubation in a 37°C incubator for 14-16 h. Order of magnitude of the titer was determined by counting visible plaques and calculating the estimated number of plaque forming units (pfu) per mL for the undiluted stock in SM buffer.

More accurate determination of phage titers were carried out by titering on full 10 cm LBM plates. Phage dilutions were selected that were expected to yield 1-100 plaques per 100 µL. 100µL of each of these dilutions was combined with 100 µL of LE392 plating cells in sterile glass culture tubes, and allowed to incubate for 15 min at 37°C to facilitate pre-adsorption of the phage to the bacteria. 4 mL of molten LBM top agar was added, and the mixture was poured onto pre-warmed plates as described above. After solidifying at room temperature, plates were inverted and incubated at 37°C for 14-16 h. Resulting plaques on each plate were counted and pfu/mL calculated for the undiluted stock in SM buffer.

2.8.3 Solutions for hybridization screening

Solutions required for hybridization screening of phage libraries were denaturing solution, neutralization solution, hybridization buffer, and 5x random prime buffer. Denaturing solution contained 20 g NaOH and 87.6 g NaCl per litre. Neutralization solution contained 500 mL 1M Tris-HCl pH 7.5 and 87.6 g NaCl per liter. Sterilization of these solutions was not necessary.

Each 100 mL of hybridization buffer is made in our lab from 69 mL ddH_2O, 25 mL 20x SSPE (175.3 g/L NaCl, 27.6 g/L $NaH_2PO_4 \cdot H_2O$, 7.4 g/L EDTA; pH 7.5), 5 mL 10% SDS, 0.1 g Ficoll (Sigma), 0.1 g polyvinylpyrrolidone (PVP; Sigma), and 0.1 g BSA

(sigma). Stock 10 mg/mL salmon sperm DNA was boiled and quickly chilled on ice before adding 1 mL to the above solution. The completed solution is stored in 50 mL aliquots at -20°C.

Homemade 5x random prime buffer (RPB) is made by combining 2 volumes solution A with 5 volumes solution B, and 3 volumes solution C. Solution A consists of 1 mL 1.25 M Tris-HCl pH8 + 0.125 M $MgCl_2$, 18 mL β-mercaptoethanol, and 5 µL each of 100mM dATP, dGTP, and dTTP. Solution B is 2M Hepes pH adjusted to 6.6 using 5M NaOH. Solution C is random hexanucleotides at 90 OD_{260}/mL in TE. RPB was stored at -80°C in 10 µL aliquots.

2.8.4 Hybridization screening of phage libraries

For primary library screening the λGT10 12.5 dpc cDNA library was plated on 150 mm LBM agar plates, as described for full plate library titering, except that 200 µL of LE392 plating cells were used and combined with 9 mL of LBM top agar. To ensure full library coverage, 30 plates of $5x10^4$ pfu were plated. After overnight incubation at 37°C for plaque formation, the plates were chilled to 4°C for at least one hour.

Plaque lifts onto 132 mm Hybond N+ nylon membranes was carried out by first placing a membrane directly onto each plate, and marking its orientation on the plate using a needle heated by a flame. Filters were left on the plates for four min after which they were transferred plaque-side-up using forceps to a shallow dish containing denaturing solution for 2 min. Membranes were then transferred to a shallow dish containing neutralization solution for 2 min, and then to a large container containing at least 1 L of 2x SSC to wash with gentle mixing for up to 30 min. Filters were then air dried on paper towels.

Hybridization steps were completed in a tightly sealing cylindrical Tupperware-like container. 5 mL of pre-warmed hybridization buffer was put into the container and membranes were added one at a time, plaque-side-up, to ensure each became soaked in the buffer. The final membrane added, was put plaque-side-down. The container was sealed and placed in a 65°C water bath shaking at 75 oscillations per min for 3-6 h of prehybridization.

During prehybriziation radioactively labelled probes were synthesized. Probes for library screening were made from DNA fragments corresponding to exons 13-19 of 3B6. These were produced by PCR using pCR-clone24, which contains exons 13-22 of 3B6. The primers used for this reaction were 3B6-13F and 3B6-19R. A typical probe synthesis reaction consisted of 50 ng DNA in 10 µL of ddH_2O, boiled and then quick chilled on ice, 8 µL 5x RBP, 10 µL [α-^{32}P]dCTP (GE Healthcare), and 1 µL Klenow (Invitrogen). The labelling reaction was allowed to proceed at 37°C for two h. Unincorporated radiolabel was removed by filtration through a ProbeQuant G-50 micro column (GE Healthcare), and radiolabel incorporation was estimated by comparison of the purified probes to the label retained in the filter after centrifugation in similar proximity to a Geiger detector. Before adding to the hybridization container, the probe was boiled for 5 min and the quickly chilled on ice. Hybridization was allowed to proceed at 65°C overnight at 75 oscillations per min.

After hybridization, the probe containing buffer was discarded and the membranes were washed 3 x 15 min in 2x SSC + 0.1% SDS at room temperature followed by one wash at 65°C in 0.2x SSC + 0.1% SDS. Membranes were then smoothly wrapped in plastic wrap and exposed to Kodak BioMax MS film overnight.

Positive plaques were identified, picked up using a sterile Pasteur pipette and eluted into microcentrifuge tubes containing 1 mL SM buffer. 10 µL chloroform was added to each, and vigorous mixing using a vortex was completed, followed by incubation at room temperature for 1-2 h. The tubes were then centrifuged for 1 min at top speed, and the supernatants, containing the positive phage, were transferred to fresh tubes. For secondary screening, the phage were diluted 1:1000 in SM buffer and 1µL, 10 µL, and 100µL of this dilution was plated as earlier described on 10 mm LBM plates. After overnight incubation, plaque lifts were done as described on plates containing well-separated plaques (one plate per positive plaque from primary screening). Hybridization was continued as done for the primary screen. Resulting positive plaques were again plated and a tertiary screen was performed to ensure isolation of the positive plaques.

2.8.5 Small scale phage DNA preparation

DNA from λphage isolated in hybridization screening was prepared using the small scale liquid lysis method. To 500μL of LE392 plating cells was added 100 μL of phage in SM buffer isolated in the tertiary screen. Pre-adsorption of the phage to the cells was allowed to take place for 15 min at 37°C. 10 mL of LBM was added to the mixture and incubated 37°C with shaking overnight. 20 μL of chloroform was then added and incubation at 37°C was continued for an additional 10 min. Cellular debris was pelleted at 3000 rpm fro 20 min, and then the supernatant was transferred to 15 mL conical tubes. 15 μL 10mg/mL RNAse and 15μL 10mg/mL DNAse were added, digestion was carried out for 30 min at 37°C. Precipitation of the phage was completed by addition of 2.5 mL 2.5M NaCl + 40% polyethylene glycol (PEG; Sigma), and two+ h incubation at 4°C. The precipitated phage was pelleted by 10 min centrifugation at 3000 rpm, and then resuspended in 0.5 mL SM buffer and transferred to a fresh microcentrifuge tube. 5 μL 10% SDS and 5 μL of 0.5 M EDTA were added to disrupt the phage capsid with incubation at 68°C for 15 min. Organic extractions were carried out and the DNA was ethanol precipitated following standard procedure. Resulting DNA pellets were resuspended in 100μL 10mM Tris. This DNA was used to screen λ phage inserts via restriction digestion, PCR, and sequencing.

2.9 Transfection and treatment of mammalian cell lines

2.9.1 Transient transfection

Transient transfections of both P19 and HEK cells were carried out in order to express mSe70-2 in a mammalian system. Expression constructs used in these experiments were pCMV-eGFP as a transfection control for FACS analysis, pCMV-FLAG-hTERT, a generous gift from Dr. N. Ting in the lab of Dr. T. Beattie (University of Calgary), as a positive control for FLAG-tagged expression of a large protein, and pCMV-3B6-FLAG, designed for expression of full length mSe70-2 with a C-terminal FLAG tag. Transfection of pEGFP-3B6, designed to produce full length mSe70-2 as a C-

termial fusion with GFP, and pEGFP-C1 (Clontech) as an empty vector control were also carried out. Transfections were typically done in triplicate for each of three timepoints and duplicate for each construct. This resulted in two dishes of identically treated cells to be collected for FACS analysis and/or cell lysate or RNA preparation, at each of 24, 48, and 72 h post transfection. A general protocol for the transfection of cells using FuGENE6 transfection reagent (Roche) according to manufacturer protocol is presented here.

Cells intended for transfection were split the day prior to transfection at ratios resulting in 30-40% confluence after 24 h of growth. Immediately preceding transfection, medium was aspirated from all cells and replaced with one-half of the normal volume required of antibiotic free growth medium. A transfection mix containing 2 µg of DNA and 4 µL of FuGENE6 in a total volume of 100 µL Opti-MEM medium (Gibco, Cat# 31985-062) was prepared in a master mix for each 60 mm dish to be transfected, and allowed to incubate for 30 min at room temperature enabling transfection complex formation. Transfection mixes not containing any DNA were also prepared as controls for transfection reagent toxicity. 100 µL of transfection mix was then added dropwise to the appropriate dish of cells, and mixed gently by swirling. Dishes were then placed in a standard 37°C humidified 5% CO_2 incubator for four h, after which each dish was brought up to its normal volume of antibiotic free growth medium and placed back into the incubator.

At each timepoint, cells destined for cell lysate or RNA preparation were collected into 15 mL conical tubes by trypsinization and centrifugation at 200 x g for 10 min. Cells were washed twice in DPBS and then either stored at -20°C or processed immediately for subsequent applications. Cultures to be collected at timepoints later than 24 h were fed with fresh antibiotic free medium each day for maintenance.

2.9.2 Proteasomal inhibition

Transfection experiments were set up as described in the previous section. At 24 h after transfection, cells were fed with fresh antibiotic free growth medium. At 48 h after transfection, medium was again aspirated, but this time replaced with antibiotic free

growth medium containing the proteasomal inhibitor MG-132(Sigma) at a concentration of 10 μM, 15 μM, or 20 μM. Cells were collected for cell lysate preparation as described in section 2.10.2, at 2, 4, and 6 h after commencement of treatment with MG-132.

2.9.3 FACS analysis

FACS analysis of cells transiently transfected with GFP expressing constructs was carried out by the University of Calgary Flow Cytometry Core Facility. Transfected cells were collected for FACS analysis at each timepoint in parallel to collection of cells for other applications. Cells were trypsinized as previously described and then subjected to manual disagregation by gentle pipetting to achieve single cell suspension, which is essential for successful FACS analysis. Suspensions were then collected in round bottom 5 mL polystyrene tubes (Falcon), and centrifuged at 500 x g for 5 min. Medium was aspirated off and then cells were washed twice in DPBS with centrifugation at 500 x g for 5 min after each wash. Resulting cell pellets were gently resuspended in 500 μL DPBS, placed on ice, and submitted for immediate sorting.

2.9.4 Vector mediated RNAi

Vector mediated RNAi knockdown of 3B6 transcripts was carried out in P19 cells. Vectors encoding dsRNA hairpins directed against 3B6 transcripts, pSUPER-3B6-KD1 and pSUPER-3B6-KD2, as well as the empty vector pSUPER.gfp.neo (OligoEngine), were transiently transfected in duplicate sets into P19 cells at 40% confluence using FuGENE6 transfection reagent as described in section 2.9.1. After 24 h one of the duplicate sets of cells for each transfection was collected for FACS analysis as described in the previous section. The remaining duplicate set was collected as described in section 2.9.1 for preparation of RNA. Alternately, all transfected cells were collected for FACS, but after sorting on the basis of GFP expression cells were returned for RNA preparation.

2.10 Protein Based Methods

2.10.1 Peptide antibody production and purification

Peptides corresponding to amino acids 276-285 and 403-412 of mSe70-2 were synthesized as both KLH- and BSA- conjugates, and HPLC at the Alberta Peptide

Institute (API) at the University of Alberta. These peptides were chosen as they are both in regions of perfect conservation between the human and mouse Se70-2 homologs, and neither falls within a predicted conserved domain. They were identified as having high probability for existing on the surface of the protein by the API using SurfacePlot analysis[169,170]. These peptides, named 3B6-2 and 3B6-3 respectively, were received lyophilized and were redissolved according to API protocols.

Peptide antibodies were raised in 3 rabbits per peptide according to methods previously described[171]. Briefly, 300 µg of KLH-conjugated peptide brought to a total volume of 1 mL in Freund's Complete Adjuvant was injected into each rabbit by technicians in the University of Calgary Animal Care Facility. Prior to the first injection, a sample of pre-immune serum was taken from each rabbit. Booster injections were prepared in Freund's Incomplete Adjuvant, and administered every two weeks following the initial injections, for a six week period, at which time samples of serum from each rabbit were taken to test for immune reaction.

Testing of serum for immune reaction to the injected peptide was done via SDS-PAGE and western blotting of 5 µg of each BSA-conjugated peptide. Detection was completed (section 2.10.3) using dilutions of serum varying from 1:500 to 1:1000. Once confirmed that an immune reaction had been raised, the rabbits were anesthetised and exsanguinated via cardiac puncture. Resulting serum was aliquoted and stored at -20°C until further use.

Affinity purification of the peptide antibodies was completed by Ms. Jackie Irwin, currently of Dr. Christoph Sensen's lab at the University of Calgary. BSA-conjugated peptides were immobilized on columns containing beaded agarose supports using an AminoLink Plus Immobilization Kit (Pierce), according manufacturer protocols. Prepared columns were equilibrated with PBS, and then 2 mL of serum was applied to the column followed by 2 mL 1x PBS. This was allowed to then incubate for at least one h at room temperature, after which the column was washed with 14 mL 1x PBS. Antibodies were eluted from the columns in 0.1 mL glycine, pH 2.5, and then neutralized by the addition of 50 mL/mL 1 M Tris, pH 9.5. Eluted fractions containing protein were

pooled and concentrated using a Macrosep 30,000 MW cutoff centrifugal filter (Pall/Gelman). Antibodies were eluted from the filter in 10% glycerol and stored at -20°C.

2.10.2 Isolation of protein from cultured cells

Cultured cells were trypsinized as previously outlined and collected via gentle centrifugation. Resulting cell pellets were washed twice in PBS, and resuspended in two pellet volumes of 1x protein loading buffer (PLB; 50mM Tris pH6.8, 2% SDS, 5% β-mercaptoethanol, 10% glycerol, 0.05% bromophenol blue). Samples were then heated to 95°C for 5 min and sonicated to shear genomic DNA. Alternatively, medium was aspirated and cells rinsed once in with DPBS before addition of 1 mL/78 cm^2 lysis buffer (50 mM Tris-HCl pH7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100) with 10 µL protease inhibitor cocktail (P-8340, Sigma). Lysates were collected into a microcentrifuge tube and centrifuges at top speed for 10 min, after which the supernatant was transferred into a fresh tube. Where required, protein concentrations were determined using a Bio-Rad DC protein assay, according to manufacturer directions.

2.10.3 Polyacrylamide gel electrophoresis, western blotting and detection

SDS-PAGE and transfer to membranes were carried out using standard lab protocols. Protein samples prepared in PLB were again heated to 95°C for 5 min and centrifuged at 13000x g prior to loading on gel. Protein separation was done in 1x SDS running buffer (25mM Tris, 192mM glycine, 0.1% SDS) at 150V until the dye front was run off of the gel. PageRuler prestained protein markers (Fermentas) were included as size standards on each gel. Where indicated, gels were run in duplicate with one gel reserved for staining with Coomassie brilliant blue. Otherwise, proteins were transferred to Hybond ECL nitrocellulose membranes (Amersham Biosciences) in western transfer buffer (25mM Tris, 192mM glycine, 20% methanol) at 120V for 1.5 h, after which the membranes were blocked in a 5% skim milk solution made up in TBS with 0.1% Tween-20 (TBST) and 0.1% Nonidet-P40 (NP-40). For detection of ubiquitinated proteins, membranes were autoclaved while submerged in ddH2O and sandwiched between sheets

of filter paper in order to improve ubiquitin immunoreactivity, as previously described[172]. This method was carried out immediately preceding the blocking step.

Incubation and wash steps for detection of proteins of interest were carried out at room temperature with gentle mixing. Membranes were incubated for 1 h in 5% skim milk in TBST with 0.1% NP-40 to which primary antibody was added in an appropriate dilution. Membranes were then washed three times in TBST + 0.1% NP-40 for 10 min per wash. Secondary antibodies were diluted in 5% skim milk in TBST + 1.0% NP-40 and incubated with membranes for 1 h, after which membranes were washed 3 times for 10 min in TBST + 0.1% NP-40. Detected proteins were visualized using enhanced chemiluminescence substrate (ECL; Amersham Biosciences) following the manufacturer protocol, and images were captured on Kodak BioMax MS film. Images from film were converted to digital format using an Epson Perfection 4870 Photo scanner and SilverFastSE software (LaserSoft Imaging Inc.).

Primary antibodies and their respective dilutions used during the course of this project are as follows: anti-FLAG rabbit polyclonal (Sigma), used at an optimized dilution of 1:5000; anti-mSe70-2 rabbit polyclonal antibodies 3B6-2 and 3B6-3, tested at dilutions ranging from 1:200-1:1000; anti-GFP mouse monoclonal JL8 (Clonetech), used at a recommended dilution of 1:10000; anti-GST mouse monoclonal (Amersham Biosciences) used at a recommended dilution of 1:10000; anti-ubiquitin mouse monoclonal (Zymed), used at a recommended dilution of 1:1000. Anti-rabbit and anti-mouse horseradish peroxidase conjugated secondary antibodies (Amersham Biosciences) were used at a 1:10000 dilution.

2.10.4 Protein expression in prokaryotic cells

E.coli strain DH5α containing pGEX-based constructs (section 2.7.9) was grown in 10 mL overnight cultures in LB supplemented with 2% sterile glucose and 50 μL/mL ampicillin. All culture steps in this protocol were carried out in this medium. Constructs were isolated via plasmid mini preps and a sample was electroporated into appropriate expression bacterial cell lines and plated for overnight growth on LB agar with 2% sterile glucose and 50 μL/mL ampicillin. From these plates, 2 mL overnight cultures were

inoculated, and 500 µL of these cultures were used to inoculate 5 mL cultures grown until an OD_{600} of 0.6-0.8 was reached. 1.5 mL of each uninduced culture was set aside in a microcentrifuge tube on ice, and expression was induced in the remaining culture by addition of 0.1mM IPTG with further incubation at 37°C for 2-4 h.

1.5 mL of each induced culture was transferred to a microcentrifuge tube. 10 µL of each cell sample was set aside in a fresh microcentrifuge tube at each subsequent step. Both induced and uninduced cells were pelleted by gentle centrifugation, resuspended in 125 µL cold MJK buffer, and incubated on ice for 1 h with periodic vortexing. MJK buffer is a proprietary detergent lysis buffer developed by Dr. Young Ou, at the University of Calgary. Cellular debris was pelleted by centrifugation, and the supernatants transferred to fresh microcentrifuge tubes. These lysates were either used directly in SDS-PAGE, or affinity purified.

For affininty purification of GST fusion proteins 20 µL of 50% Glutathione Sepharose 4B (GE Healthcare) slurry was added to the supernatants which were then incubated at 4°C for 2 h with gentle mixing. Beads were pelleted by centrifugation, and then washed three times in 1x PBS. Bound proteins were eluted from the beads twice in GST elution buffer (10 mM reduced glutathione [Sigma] in 50 mM Tris-HCl, pH 8.0), and subjected to SDS-PAGE.

2.10.5 In vitro protein production

In vitro transcription and translation of mSe70-2 was done using the TnT T7 coupled reticulocyte system (Promega), with constructs T7-3B6, and T7-FLAG-3B6. Reactions were carried out in the presence of ^{35}S-methionine. A typical 10 µL reaction consisted of 5 µL rabbit reticulocyte lysate (RRL), 2.4 µL ddH2O, 0.4 µL reaction buffer, 0.4 µL methionine-free amino acid mix, 0.2 µL RNAse out (Invitrogen), 0.4 µL Redivue L-[^{35}S]methionine (GE Healthcare), 0.2 µL T7 RNA polymerase, and 1µL 0.1 ng/µL template construct. Reactions were allowed to proceed in a 30°C water bath for 2 h, and then frozen at -20°C until use. If the produced protein was to be directly applied to SDS-PAGE, unincorporated methionine was first cleaned up from the sample by filtration through a Microcon YM10 spin column filter (Millipore).

2.10.6 Immunoprecipitation

Immunoprecipitation (IP) was routinely carried out using ANTI-FLAG M2 affinity gel (Sigma) according to manufacturer directions with some changes. 1 mL of cell lysates or 5 µL of ^{35}S methionine labelled TnT reaction in 1 mL TBS were precleared with 50 µL protein-G sepharose beads at 4°C for 1 h. 40 µL of FLAG affinity beads was transferred into a microcentrifuge tube and washed twice in TBS. All centrifugation steps were carried out at 4°C at 5000 x g. The protein G sepharose pelleted by centrifugation and the lysates were transferred to the tubes containing the FLAG affinity beads. IPs were carried out at 4°C mixing end over end overnight. The FLAG affinity beads were then washed three times in 1 mL TBS. Captured proteins were eluted from the beads by addition of 20 µL PLB and heating to 95°C for 5 min, before centrifugation and SDS-PAGE.

2.10.7 RNA affinity chromatography

RNA affinity chromatography was carried out as previously described [173], but with several modifications. Lyophilized pre-conjugated poly-A, poly-C, and poly-U homoribopolymer agarose beads (Sigma) were reconstituted in 0.1M NaCl pH 7.5 according to manufacturer directions. Poly-G homoribopolymer agarose beads and unconjugated adipic acid dihydrazide agarose beads (Sigma) were received as 50% slurries. All beads were washed according to manufacturer directions and then equilibrated and stored as a 50% slurry in Buffer D containing 0.5% CHAPS.

Homoribopolymer binding reactions consisted of 50 µL of 50% RNA-agarose or agarose only slurry, 6µL of ^{35}S labeled TnT protein reaction, 100ng/µL BSA, 0.5nmol tRNA, and Buffer D with 0.5% CHAPS to a final volume of 200µL. Binding reactions were incubated at 30°C for 1 h mixing end over end and were then washed 4 times with 1mL Buffer D with 0.5% CHAPS and 4mM $MgCl_2$. Proteins bound to the homoribopolymers were eluted in 20µL 1x PLB with heating to 95°C for 5 min, and then separated on 6% SDS-PAGE. The resulting gel was dried and exposed to Kodak BioMax MR film. Images from film were converted to digital format using an Epson Perfection 4870 scanner and SilverFastSE software (LaserSoft Imaging Inc.).

2.11 RNAi in the *C. elegans* system

2.11.1 Strains, maintenance, and general reagents

Wild type *C. elegans* N2 Bristol strain was used during this project. *C. elegans* were maintained and manipulated according to widely used methods. [174,175]. Briefly, *C. elegans* were cultured on lawns of *E. coli* strain OP50 grown on nematode growth medium (NGM) agar (standard feeding plates). Incubations were at 20°C, and *C. elegans* were physically manipulated and passaged using platinum picks.

NGM agar consists of 3 g/L NaCl, 2.5 g/L peptone, and 17 g agar with 1 mM $CaCl_2$, 1 mL of 5 mg/mL cholesterol in ethanol, 1 mM $MgSO_4$, and 25mM KPO_4 added after autoclaving. M9 buffer is 6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 5 g/L NaCl, and 0.25 g/L $MgSO_4{\cdot}7H_20$. Phosphate-EDTA solution is 10mM phosphate buffer with 0.1 mM EDTA, and dream juice is 0.2% tricaine methanesulphonate.

All images were captured using a Zeiss Axioplan 2i microscope with a Hamamatsu OrcaER digital camera and Axiovision 4.2 software.

2.11.2 Preparation of dsRNA

Double stranded RNA (dsRNA) corresponding to fragments of the worm homolog of 3B6, B0336.3, were prepared essentially as previously described[176,177]. The appropriate clone JA:B0336.3 was available from an RNAi feeding library described by Kamath and Ahringer,[178] and was received as a generous gift from the lab of Dr. J. Gaudet (University of Calgary).

A single bacterial colony containing the relevant clone was selected for growth in a 2 mL 2xYT medium with 50 µg/mL ampicillin liquid culture for 6h. The insert with flanking T7 promotors was amplified using the standard PCR reaction described in section 2.5.2 with the primers L4440F and L4440R, and 2 µL of the bacterial culture as template. 14 identical reactions were prepared in order to produce enough of the inset for further processing. Thermocycling for this reaction was carried out as follows: 94°C for 30 seconds, 60°C for 30 seconds, 72°C for 2 min, x 35 cycles. Product from two of the reactions was cleaned up using a QIAquick PCR cleanup kit (Qiagen), and subject to restriction mapping as well as sequencing to ensure the correct clone had been obtained.

The remaining 12 reactions were pooled into four microcentrifuge tubes, and cleaned up via ethanol precipitation, resuspending the clean DNA in 46µL nuclease free ddH_2O.

In vitro transcription was carried out by adding the following to the tubes containing 46µL of amplified DNA insert: 20 µL 5x transcription buffer, 10 µL 100mM DTT, 1 µL RNAse OUT RNAse inhibitor (Invitrogen), 20 µL pre-mixed 2.5mM rNTPs, and 2 µL T7 RNA polymerase (Promega). This reaction was allowed to proceed at 37°C overnight. The reactions were subsequently phenol-chloroform and chloroform extracted, pooling the supernatants into a single microcentrifuge tube. The dsRNA was precipitated and resuspended in 20 µL of phosphate-EDTA solution, and a 1 µL aliquot run out on agarose gel for confirmation of appropriate size and quality.

2.11.3 Injection of dsRNA

The concentration of the dsRNA prepared as described in section 2.11.1 was measured using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies). The concentration of the dsRNA was adjusted to 1 µg/µL and it was subsequently separated into 5 µL aliquots for storage at -20°C prior to injection.

Injection of young adult wild type N2 *C. elegans* was carried out by Dr. J. McGhee according to his standard protocol[179]. Ten *C. elegans* per experiment were each injected with dsRNA directed against B0336.3. A second group of ten *C. elegans* were injected with dsRNA directed against GFP in exact parallel, as a control. These injections were done blind, so that scoring for phenotype would be completed without knowledge of which group received which particular dsRNA injection.

Post-injection, the *C. elegans* were transferred from the injection slide to a dish containing M9 buffer, and then to standard feeding plates for overnight recovery at 16°C. After recovery, they were transferred to fresh feeding plates and incubated at 20°C as usual, for feeding and egg laying.

2.11.4 Scoring C. elegans for growth defects

To score for growth defects, F0 generation injected *C. elegans* were transferred to fresh feeding plates every 6h. 3h after transfer to a fresh plate was deemed as age 0h for the F1 generation laid on the plate in question. After the F0 generation was transferred

off of a given plate, the plate was maintained at 20°C to allow for hatching and development of the F1 generation. F1 generation *C. elegans* at 100h old (adult) or 38 h old (L4; fourth larval stage) were collected in M9 buffer for scoring. After gentle washing in M9 buffer, they were anesthetised in dream juice and mounted on microscope slides for imaging. Measurements of each imaged worm were conducted using ImageJ v1.33 software[180]. Mean length and standard deviation was calculated for each treatment group, and one sided T-tests of means were conducted to establish if any noted differences were significant.

2.11.5 Scoring of C. elegans for egg laying defects

To score for egg laying defects, F0 generation injected *C. elegans* were each given their own feeding plate and transferred to a fresh feeding plate every 6h. Three h after transfer to a fresh plate was deemed as age zero h for the F1 generation laid on the plate in question. After the F0 generation *C. elegans* was transferred off of a given plate, the eggs laid on that plate were counted and then was maintained at 20°C to allow for hatching and development of the F1 generation. At 38h old (L4), five F1 generation *C. elegans* were each placed onto their own fresh feeding plate, and transferred to a new plate every 6h. Eggs laid by the F1 generation were counted as for the F0 generation. Egg counting continued for each worm in each injection group until egg laying had ceased to occur. Total number of eggs laid for each *C. elegans* was determined, and the mean number of eggs laid and standard deviation for each treatment group and generation was determined. One sided T-tests of means were conducted to establish if any significant differences in egg laying existed.

Chapter Three: The Rancourt EST Database

3.1 Summary

Following the Bain-Rancourt EST project, the Rancourt lab began to pursue an interest in nucleic acid binding proteins with roles in the neural differentiation process. This chapter describes the design and implementation of the Rancourt EST Database (RED), which was developed in response to the need for a data analysis, management, and dissemination system for use with the Bain-Rancourt EST project data. RED was crested using an open-source package of software for web database development, Nushpere 1.13.10. It routinely gathers relevant information regarding each Bain-Rancourt EST, also now known as the RED dataset, and delivers it to users in a convenient format. RED may be viewed at http://136.159.173.207/ESTs/public/RED.php In addition, 12 ESTs from the RED dataset were selected on the basis of criteria set forth for identification of possible nucleic acid binding proteins with roles in neural differentiation. An initial screen of their expression is also presented.

3.2 Introduction

3.2.1 Need for a data management system

With the completion of the Bain-Rancourt EST sequencing project[135], it appeared that the key directional focus of the Rancourt lab would become discovery and characterization of genes involved in neural fate decisions, particularly those encoding nucleic acid binding proteins. The 604 non-redundant ESTs isolated during the sequencing project had been annotated using batch BLAST submissions to the NCBI and hand parsing of the results. Each of the EST sequences were submitted to the dbEST division of GenBank and assigned accession numbers starting with AW244216. The dbEST library identification number is 2757.

Additional information regarding each of the clones had been gathered manually using web-interface tools. All resulting data, including but not limited to clone identifiers, sequences, annotations, and BLAST hits for each EST was hoarded into a single Excel (Microsoft) spreadsheet. The data produced quickly grew beyond the limitations of the spreadsheet. In addition, due to the rapidly changing availability, quality, and varied analyses of molecular data in the public realm, the snapshots of information collected pertinent to the Bain-Rancourt ESTs rapidly became outdated. Data for each EST was critical to continued work related to the project, however the constant gathering of information consumed large tracts of research time as each analysis was labor intensive. This was particularly true as analyses needed to be frequently repeated to ensure that results remained current, and of desired quality. It was also recognized that management and dissemination of results yielded in bench experiments further examining the individual ESTs would be valuable to the lab. Data yielded from experiments included images of gels, blots, films, and equivalents, with associated text. The Excel spreadsheet provided no satisfactory method for management of these datatypes, and thus a more sophisticated data management system was needed.

Candidate gene selection based on the data contained within the existing spreadsheet also proved to be a challenge. Each user wanting to identify genes for follow up was required to sift through the masses of information available in the spreadsheet.

This was a daunting task, as the massive spreadsheet was difficult to navigate, the data was not cross-referenced to its original sources, and it was often incomplete, out of date, or both. The need for development of a data analysis, management and dissemination strategy in the lab to expedite further work was readily recognized.

3.2.2 The NuSphere Package

NuSphere (http://www.nusphere.com) provides integrated distributions of highly popular web development components, portable to various operating systems including Windows, Linux, Sun, amongst many others. NuSphere 1.13.10 is an open source distribution including MySQL 3.23.38, Perl 5.6.1, PHP 4.0.5, and Apache 1.3.20. Perl (Practical Extraction and Report Language) is generally the language of choice for bioinformatic tasks as it excels at parsing and integrating data files in a manageable fashion[181]. PHP and MySQL are widely used together for web database applications; MySQL providing the database structure, and PHP providing the necessary applications, from display of webpages to verification of complex queries[182]. Apache is the *de facto* standard for web servers. In the July 2006 Netcraft Web Server Survey (http://news.netcraft.com/archives/web_server_survey.html) it was found that over 63% of the web sites on the internet were using Apache as their server. Apart from the powerful team of components included in the NuSphere 1.13.10 package, other advantages for its use include that open source software is free of charge, and is widely supported and in perpetual use by countless developers.

3.3 Results

3.3.1 RED: The Rancourt EST Database

The data management, analysis, and dissemination challenges presented by the early treatment of the Bain-Rancourt EST dataset were rectified by development of a web database application subsequently named the Rancourt EST Database (RED)[183].

To ensure the quality of the original EST data, and assigned annotations, chromatograms produced during the initial EST sequencing project were retrieved from the archives of the University of Calgary Core Sequencing Facility and refinished using

the Staden package. Refinished sequences for the 604 non-redundant ESTs were given new putative annotations using MAGPIE with human intervention, as a starting point for this project.

RED, the Rancourt EST Database, was designed in concept via discussions with potential end users. Incorporating desired functionality and information retrieval was paramount in its development. RED represents a flexible template DNA sequence database that could be easily manipulated to suit the needs of different projects. It was developed using NuSphere 1.13.10, and is normally accessible from http://www.ucalgary.ca/~rancourt but is currently only available at http://136.159.173.207/ESTs/public/RED.php. Underlying RED is a MySQL 3.23.38 relational database. Automated data retrieval and analysis tools were written in Perl 5.6.1. PHP 4.0.5 was used to create the user interface and mediates information transactions between users and the database, while Apache 1.3.20 acts as the web server.

The most critical function of RED is to retrieve, maintain, and update data pertinent to each EST in an automated fashion with minimal administrator or user intervention [Figure 6]. On a bi-weekly basis, current data is automatically downloaded and unpackaged from the NCBI ftp server. The automated data retrieval and analysis tools parse the given data, extracting relevant information to be housed in the database, or information required to facilitate retrieval of additional data. The original data generated during the EST sequencing project, and data output from the automated scripts is stored locally in the MySQL database. The database is automatically populated with updated information contained in the output files generated by the scripts, using blocks of MySQL commands executed by additional Perl scripts. Hypertext links appearing in the pages of the interface for RED are dynamically generated by PHP accessing the data contained in the MySQL database, in response to user input.

To further illustrate the functional data flow in RED, the following example may be considered: One of the pieces of data that is retrieved for each EST in RED is the identifier for the UniGene cluster of ESTs it falls into, if any. To fetch each of these IDs, the entire *Mus musculus* UniGene dataset (mm.data.gz) is downloaded from the NCBI ftp

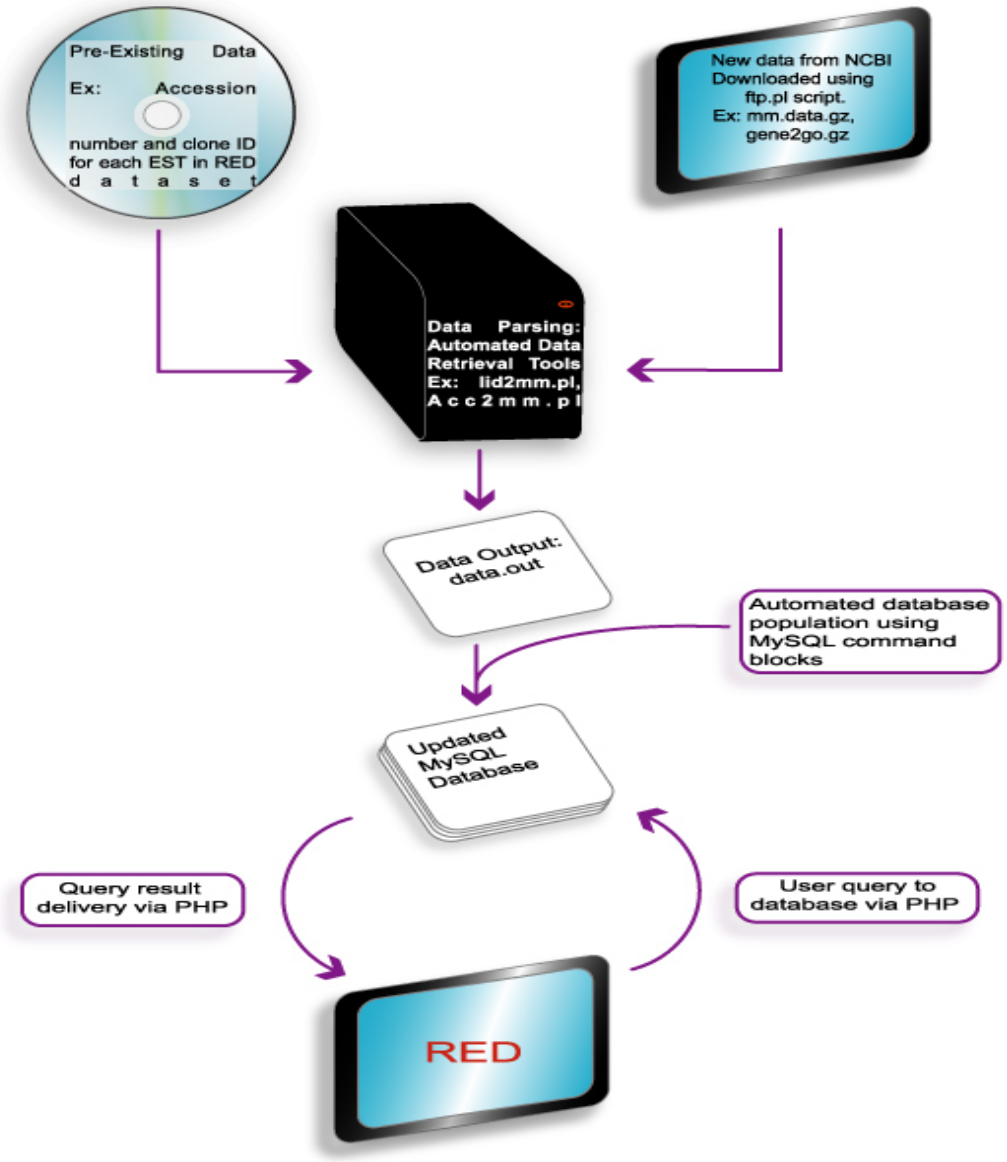


Figure 6: Schematic of RED function. The critical functions of RED are retrieving, and updating pertinent data. Updated data files from NCBI are downloaded on a bi-weekly basis using the ftp.pl script. Permanent, pre-existing data generated during the initial EST sequencing project and the new NCBI data are used by the PERL automated data retrieval tools, which parse out needed information both for database population and for usage by other scripts. Output data needed for population of database tables is automatically written to the necessary tables using MySQL command block. At any time the user may access the information contained in the database via the PHP generated user interface.

server and expanded using the script ftp.pl. From this dataset, the script lid2mm.pl extracts every UniGene cluster containing an EST from the Bain-Rancourt dbEST library (dbEST ID 2757). Each of these clusters is written to an output file lid2mm.out. This new file is used as input for the script acc2mm.pl which uses the accession number for each of the ESTs to find which UniGene cluster it falls into. This information is written to a new output file and stored in the database for access via the PHP interface.

3.3.2 The original RED user interface

A schematic illustrating the original RED user interface layout is shown in Figure 7. The RED interface is divided into public and private sections. The public portion of RED is freely accessible, whereas only authorized users may access the private portion. In the original version of the user interface, the home link and task-specific manual pages were available from both the public and private portions of RED. The public side of the interface included background and database pages, offering information on why RED was created and a description of the database. The graphs page displayed spontaneously generated statistics regarding the EST dataset, in graphical format. The ESTs page was the portal to the data, allowing the user to search RED with a gene ontology (GO) term or keyword(s), or to view all ESTs. Query results for each EST presented all available information pertaining to the sequence as hypertext links in a single tabular format. Following the links provided in each field of the table allowed access to locally stored data, as well as information gleaned from external resources. Fields of the results table included clone ID, putative annotation, gene ontology, GenBank record, and additional information. Links appearing in the gene ontology field connected the user to appropriate entries GO consortium's AmiGO browser. In the GenBank record field, the accession number for the desired EST was displayed, while its gi number (GenInfo identifier) was used covertly to link to the EST's GenBank record. The dynamically created links listed within the additional information field allowed the user to view any available EST summary, automated BLAST alignment results, and UniGene EST cluster data. The public portion of RED also included a BLAST page that enabled users to compare their sequences with the Bain-Rancourt ESTs using the stand-alone BLAST executables

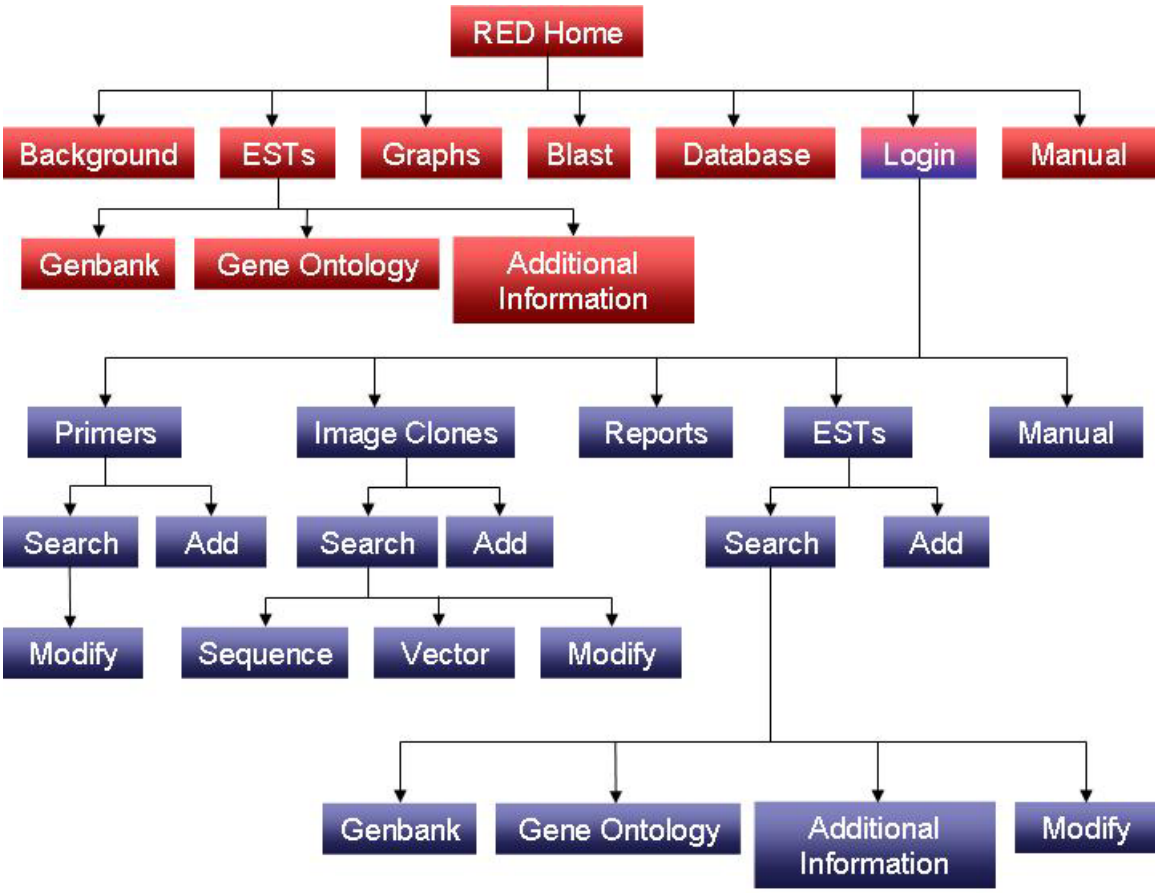


Figure 7: The original RED user interface layout. The user interface layout for RED was conceived in consultation with potential end-users so that desired functionality could be delivered. RED was divided into segments available to the public, and segments available only to authorized users. Red boxes represent pages included in the original publicly available graphical user interface. The login page is visible from the publicly accessible portion of the interface; it is the gateway to the private portion of the database which is shown in blue.

available from NCBI (http://www.ncbi.nlm.nih.gov/BLAST/download.shtml).

The login page was the authorized user's gateway into the private portion of the database. From the private ESTs page, the user could choose to search for existing EST sequences, or add new ESTs to the database. Searching the ESTs returned all of the information provided by the public search ESTs page, plus links to experimental results, available primers, and available corresponding clones from the IMAGE (Integrated molecular analysis of genomes and their expression) consortium, if any. Authorized users were also given the option of modifying EST records. IMAGE clones could be searched for from, or added to the database from the image clones page. IMAGE search results included the IMAGE identifier number, corresponding RED EST clone ID, vector, length, and sequences. The vector link accessed information about the IMAGE cloning vector from the IMAGE Consortium, while the sequences link allowed the user to view the sequences for each IMAGE clone. The option to modify IMAGE clone entries was also given. From the reports page authorized users could select from a series of data output reports such as FASTA formatted sequences, EST information summaries, or custom reports containing only desired information. The reports page also included a batch EST submission report, which is formatted for direct submission of ESTs that have not yet been submitted to GenBank. Information regarding available primers for each EST was available from the primers page. Searching the primer information returned the EST clone ID and size, 5' and 3' primer sequences, resulting amplimer size, and a link to modify the primer information.

3.3.3 The current RED user interface

Redesigning of the user interface of RED began after over a year of usage. It became clear at the time that several features of the original interface were severely underused, and others had become redundant in light of the progressive changes to public resources such as the NCBI and Ensembl. The current RED interface [Figure 8], is effectively a streamlined version of the original, facilitating more intuitive browsing while delivering access to an expanded volume of EST related information. RED remains divided into public and private portions, with the login page as the gateway to

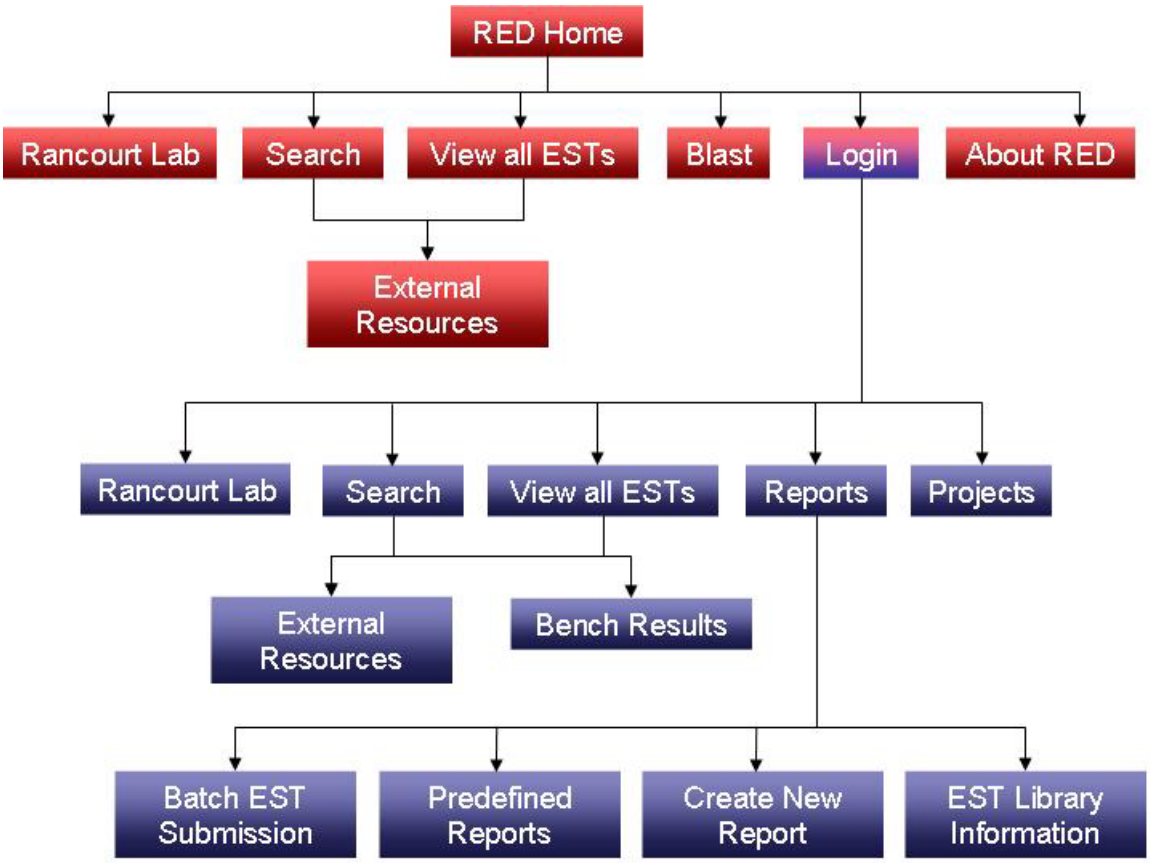


Figure 8: The current RED user interface layout. The current interface layout for RED is a streamlined version of the original layout. The layout was redesigned to facilitate more intuitive browsing while delivering an expanded amount of EST-centered information. Underused or redundant functions were also eliminated to reduce clutter, increasing ease of use. RED remains divided into segments available to the public, shown in red, and segments available only to authorized users, shown in blue.

the private portion. The most notable omissions from the current interface in comparison to the original are the IMAGE clones and primers pages, which were eliminated due to lack of use amongst authorized users. The graphs page was also eliminated due to lack of use and maintenance issues, and the new about RED page, providing literature about RED and surrounding projects, was substituted for the old background and database pages. The new projects page in the private portion of the database is the page authorized users are directed to when they initially logon. It provides notes and milestone news regarding particular EST follow-up projects ongoing in the lab, enhancing communication between individuals working on related projects. The private reports page is still available, but the report creation process has been made more user-friendly by incorporating separate pages for the desired report types.

The RED homepage is the entrance allowing user access to the RED data [Figure 9]. From the homepage a selection of intuitive links are available, still including the BLAST link allowing users to compare sequences against the entirety of RED. Now, from the homepage and every page, RED can be searched or the user can choose to view every EST. Searching may be done using a GO term, keyword(s), accession number, or EST ID. Search results pages [Figure 10] were newly redesigned to present many links to external resources in an intuitive fashion. The database entry returned for each EST includes general information about the EST including its RED ID, GenBank accession, as well as the gene description and product name for the gene to which it appears to be a segment of. This information is retrieved from the NCBI Entrez Gene data by the Perl automated data retrieval tools, and stored locally. All remaining information regarding each EST is delivered in the form of colour-coded hypertext links to external resources. Resources tapped include GenBank, UniGene, Entrez Gene, PubMed, OMIM, Homologene, Ensembl, AmiGO, and CD (conserved domain) search. In addition to these links, authorized database users are also provided with a link named bench, which provides access to any locally stored experimental results, if available.

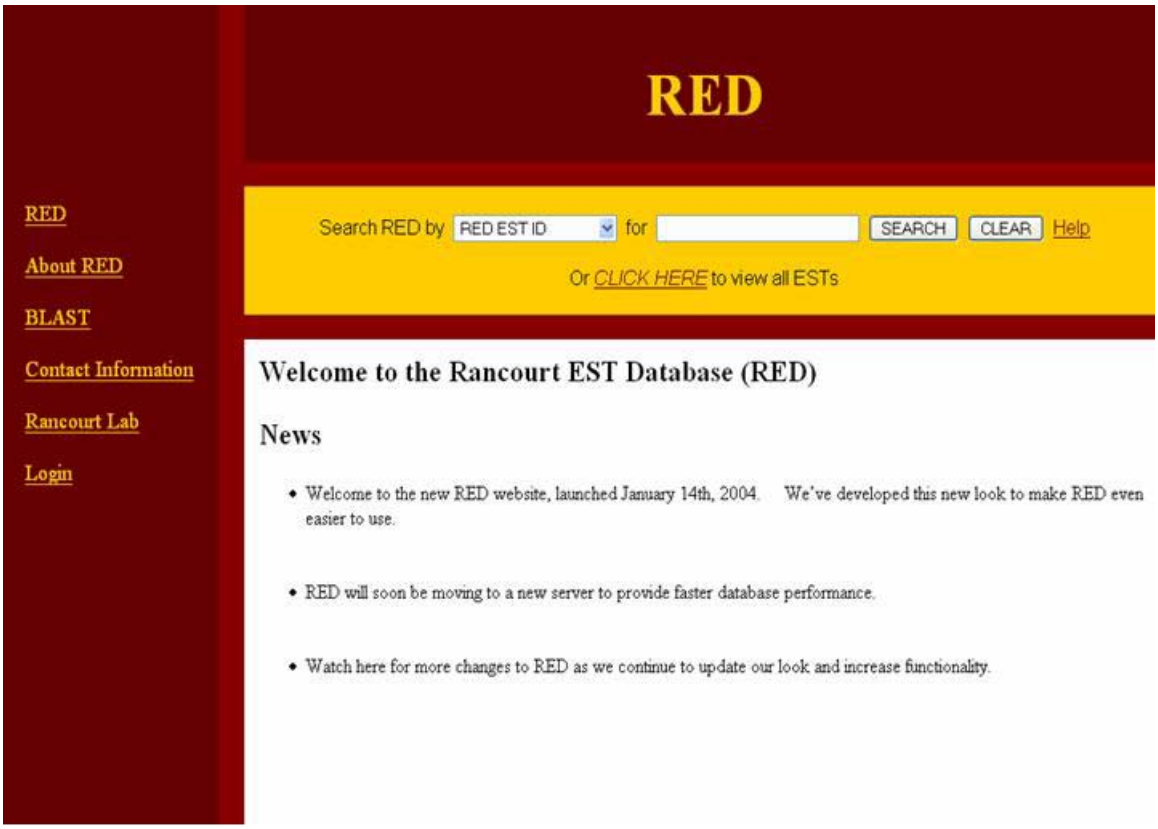


Figure 9: The RED homepage. The RED home page is the user's gateway to RED. From this and every page, RED can be searched on the basis of RED EST ID, Keyword(s), Accession number, or Gene Ontology term. All ESTs may also be viewed. The links on the left hand side of the page are as follows: RED, returns to RED homepage; About RED, literature regarding the Bain-Rancourt EST sequencing project and RED itself; BLAST, allows users to used a stand alone version of BLAST to compare query sequences with those contained in RED; Contact Information, information for contacting our lab; Rancourt Lab, lab home page; Login, portal to private portion of RED.

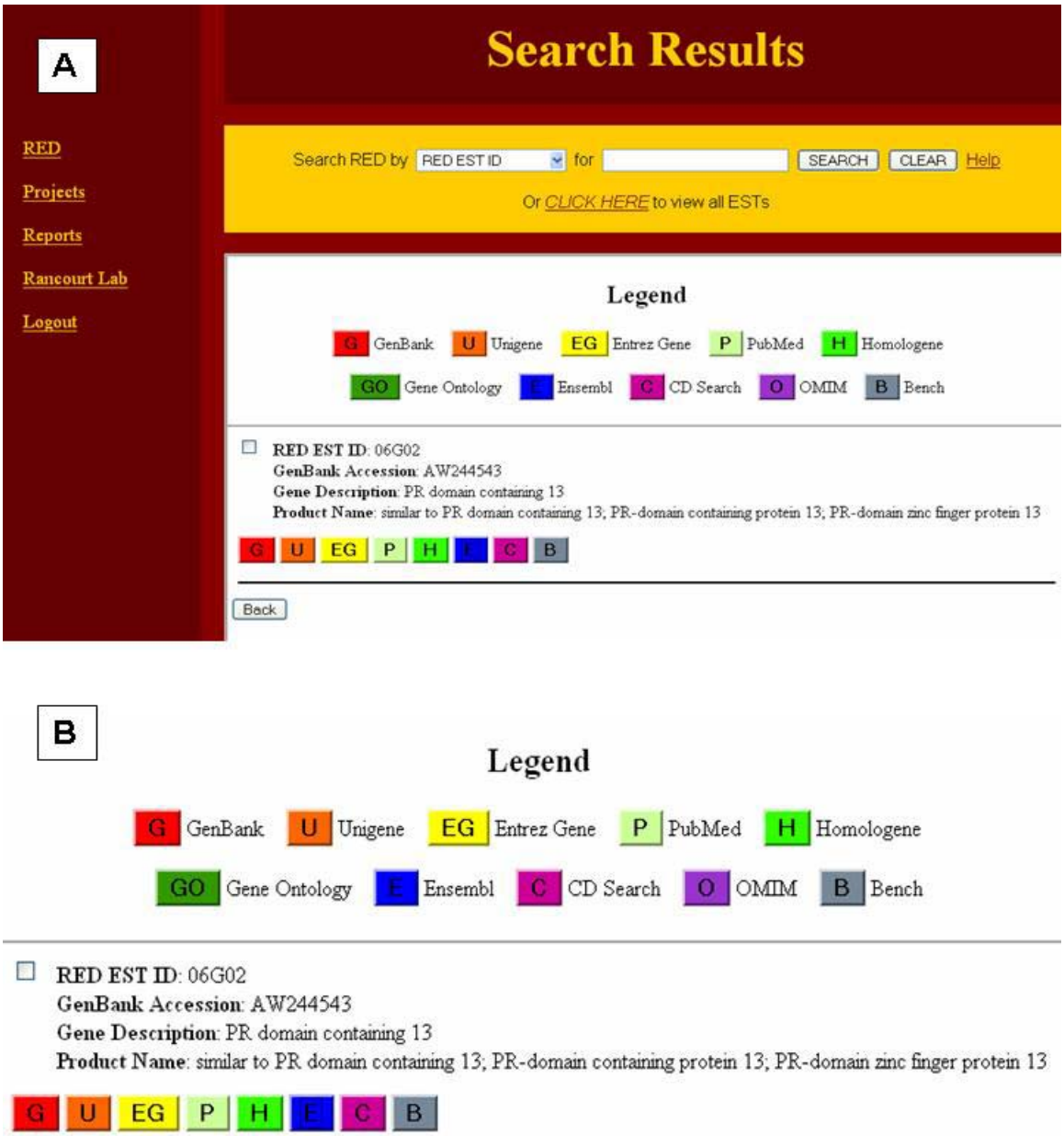


Figure 10: Representative RED search result page. A results page from a search originating in the private portion of RED is shown. Panel A shows the complete results page. Panel B shows the search result itself in isolation. Basic information regarding the identity of each RED clone is given, as well as a selection of links leading to external resources giving information pertinent to the selected clone. Links are dynamically created and represented as coloured coded buttons for ease of identification. Links when followed open their targets in new browser pages. Search results from the public portion of RED appear identical to that shown, except for the unavailability of the grey "bench" link.

3.3.4 Candidate gene selection using RED

Apart from data analysis, management, and dissemination, one of the goals of RED was to streamline the candidate gene selection process for further follow up within the wet-lab. The Rancourt lab was developing a directional focus surrounding the discovery and characterization of new genes encoding nucleic acid binding proteins, with potential roles in neural differentiation. Candidate ESTs selected for follow up were expected to meet the following criteria based solely on *in silico* analyses: evidence for expression in libraries originating from neural tissues, and novel or uncharacterized status, with related sequences originating from predictive methods or sequencing projects. GO terms implying nucleic acid binding functions were not included as criteria as at the time, GO terms were not assigned to uncharacterized genes and proteins.

Twelve candidate ESTs for follow up were selected using RED, and are listed in Table 2. Selection of candidates was completed simply by browsing the information available in RED, as well as the presented links to external resources. Each of the ESTs selected met the required criteria. Each was part of a UniGene EST cluster attributed to an uncharacterized gene. Most of these clusters were only recognized as transcribed loci defined by the Riken full-length enriched cDNA sequencing project[184], and others were hypothetical predictions based on the existence of a number of ESTs. The UniGene cluster containing one of the selected ESTs, 8E1, did show some similarities to a previously known but uncharacterized gene. All of the UniGene clusters that the selected ESTs fell into also contained ESTs sequenced from neural tissue-containing cDNA libraries, such as whole brain, head and neck, spinal cord, and brain stem, amongst other cDNA sources. The sequence for one of the ESTs itself, 3B6, showed evidence of encoding a nucleic acid binding functional motif: a CCCH type zinc finger detected using InterPro Scan against SMART HMMs (P=2.9e-4).

Primer sets for each of the ESTs selected for follow up were obtained, and RT-PCR was performed, examining their expression in ES cells, 4-/3- cells (spontaneously differentiating), 4-/3+ cells (Neurally differentiating; induced with RA), and several embryonic stages[Figure 11]. Of the 12 clones selected for follow up, six of them, 1B4,

Table 2: RED clones selected for follow up

Clone ID	Description	Unigene ID	Neural cDNA sources	Domains
01A09	Hypothetical	Mm.260925	Whole Brain, Prosencephalon	N/A
01B04	Hypothetical	Mm.383979	Head and neck,	N/A
01H07	Riken cDNA	Mm.152466	Whole brain, Brain stem, Eye	N/A
02B05	Riken cDNA	Mm.222224	Whole Brain, Spinal cord, Head and neck	N/A
02F10	Riken cDNA	Mm.177058	Whole Brain	N/A
03A06	Riken cDNA	Mm.126757	Head and neck	N/A
03B06	Riken cDNA	Mm.291542	Whole brain, Head and neck	CCCH Zinc Finger
06G02	Hypothetical	Mm.208541	Cerebellum	N/A
07G02	Riken cDNA	Mm.35470	Whole brain	N/A
08E01	Similar to known	Mm.247438	Whole brain, Brain stem	N/A
12B11	Riken cDNA	Mm.45203	Spinal cord, Whole brain	N/A
12G01	Riken cDNA	Mm.386888	Whole brain, Head and neck	N/A

ESTs from RED were selected for follow up on the basis of several features. Each EST fell into an uncharacterized UniGene cluster which contained ESTs from sources indicating neural expression. Each of these UniGene clusters was also attributed to uncharacterized genes. The description given for each clone originates from the time period at which the clones were selected, but the UniGene ID given reflects the current cluster membership, as clusters are commonly retired and/or merged. 03B06 was the only selected clone that showed evidence of encoding a functional domain.

Figure 11: Expression of RED clones selected for follow up. Expression analysis of selected RED clones was carried out using RT-PCR against RNA prepared from the indicated sources including ES cells, differentiating ES cells, and developing mouse embryos. Several clones, 1B4, 1H7, 2F10, 3A6, 3B6, and 8E1, were found to be expressed at a higher level in neurally differentiating ES cells (4-/3+ neural precursors) than in spontaneously differentiating ES cells (4-/3-).

1H7, 2F10, 3A6, 3B6, and 8E1, were found to be more highly expressed in 4-/3+ neural precursors than in spontaneously differentiating 4-/3- cells. Nine of the clones were found to be expressed in ES cells. None of the selected clones were found to be expressed in differentiating cells while not in ES cells. From the embryonic cDNA samples, all clones except 2B5 and 6G2 were found to be expressed in at least two of the three time points. In almost all clones expressed in the embryonic samples, expression at 10.5 dpc was lower than the expression that was observed at 12.5 dpc or 14.5 dpc.

3.4 Discussion

3.4.1 RED is a flexible template database

RED has assisted the Rancourt Lab in efforts to retrieve, collate, and share data obtained about the Bain-Rancourt ESTs, therefore addressing previous data management and enrichment challenges. Data gleaned from numerous sources is smoothly integrated and incorporated into RED, allowing further work on the characterization of the EST sequences to be expedited, both *in silico* and at the bench. RED has effectively eliminated the repetitive and time consuming tasks of manually maintaining the EST dataset. The automated data retrieval and analysis tools accompanying RED eliminate the need for research time to be invested in the continual bioinformatic re-analysis of the data set, and ensure that data contained within the database remains current and accurate.

The Rancourt lab has used information compiled in RED to more efficiently identify ESTs that may represent candidate regulatory genes in the neural differentiation process. RED streamlines the candidate selection process: rather than investing time combing through and verifying significant amounts of data of questionable quality contained within a spreadsheet, candidate selection is expedited as a broad range of current information pertaining to each EST is presented in a simple fashion. The need to search multiple external resources is eliminated, as each analysis is pre-completed and served up to the user in user-friendly hypertext links.

From a more general perspective, RED represents a flexible template database that could be manipulated to suit the needs of a variety of projects. RED is particularly well

suited to projects where information regarding many genes needs to be gathered and maintained. For example, it was recognized that RED is an effective system for management and enrichment of data yielded from analysis of microarray experiments. For use in other projects, modifications or additions to the automated data retrieval and analysis tools can be made to reflect the needs of the particular project, and the look of the user interface may also be changed as desired. Based on the extremely popular web database development team of MySQL and PHP, RED is completely scaleable. The only limitation to implementation of RED is the hardware platform it is run on. Another advantage of RED is its portability, achieved by utilizing open source tools for its construction. This also allows us to freely offer RED to anyone who wishes to use it as a framework for their own web database system. A version of RED was given to the Lab of Dr. Geoff Hicks at the University of Manitoba, who as members of the International Gene Trap Consortium used the RED framework as inspiration for their Embryonic Stem Cell database (ESdb; www.escells.ca).

The concept of creating custom in-house web database systems for management and analysis of projects is certainly no longer unusual. Concurrent with and since the time of the publication of RED, there has been emergence of numerous similar web databases for management of EST projects. Many of these, such as ESTree[185] db, the Diatom EST Database[186], GeneNotes[187], and ESTIMA[188], are very similar to RED in both concept and function. The proliferation of such projects illustrated the validity of the concepts and methods used while implementing RED. Indeed, it also speaks to the crucial role of data management and analysis in discovery based projects, as well as to the increasing computational savvy of labs undertaking these projects.

The past years have seen the directional focus of the Rancourt lab shift away from projects surrounding neural gene discovery, and hence RED has been seeing little use. It remains however, an excellent, flexible framework for a data management, analysis, and dissemination system which may be modified to meet the needs of a wide variety of projects.

3.4.2 Candidate EST expression

Although the candidate EST selection process was greatly simplified by RED, the results of the RT-PCR expression screen were not as anticipated. All of the ESTs in the dataset were isolated in a subtractive screen between cDNAs from ES cells and 4-/3+ neural precursors. It was initially anticipated that results would show expression of the ESTs in 4-/3+ cDNA samples, but not in ES cell cDNA samples, however, this was not the case. Almost all of the ESTs examined in this and other experiments conducted by the lab were found to be expressed in ES cells. Those that could not be detected at some level in ES cells often could not be detected at all in any cDNA samples (data not shown). Such results were consistently reproducible by multiple lab members, and therefore the unexpected expression patterns cannot be simply attributed to technical error arising from an individual. Although the ES cell cDNA used in the subtraction was biotinylated, and four rounds of subtraction with streptavidin were completed, it is apparent that the resulting library of sequences was not limited to those expressed exclusively in 4-/3+ cell cultures. It is also interesting to note that the paper originally describing the RED ESTs draws attention to the level of completeness of the subtraction, indicating that the subtraction was 'not saturating'[135].

The expression of the ESTs in both differentiating and non-differentiating cultures could suggest that these genes may play more generalized roles in cellular maintenance such as cell cycle or translational control. This idea though should not exclude the possibility that some of the genes represented in the RED dataset do have roles in the neural cell type fate decision, as it was noted that six of the screened ESTs were more highly expressed in neurally differentiating 4-/3+ cultures than in spontaneously differentiating 4-/3- control cultures. The Bain method for neural differentiation of ES cells does not yield a pure culture of neural cell types. Rather, mixed cell populations enriched with neural cells result. Similarly, spontaneously differentiating ES cell cultures also contain a proportion of neural cells. The contribution of neural precursors to the control 4-/3- cultures may explain the expression of what could be genes influencing neural differentiation, in spontaneously differentiating cell cultures.

Expression of many of the candidate ESTs was also seen in the selected embryonic stages. In the great majority of the ESTs screened, expression was at higher levels in 12.5 dpc and 14.5 dpc embryos than in 10.5 dpc embryos. This observation further supports the possibility that the genes in question may play roles in neural development, as the highest rates of neural proliferation in the mouse embryo begin at approximately 11.5 dpc[189-191].

To conclude, though the expression of the selected ESTs was not exactly as anticipated, it is still a valid possibility that some of the uncharacterized ESTs within the RED dataset represent genes with roles in the development of neural cell types.

Chapter Four: Rancourt EST Database clone 3B6: Gene Structure and Expression

4.1 Summary

Chapter 3 showed a preliminary expression screen of several RED clones selected as candidate nucleic acid binding proteins with roles in the neural differentiation process. In this chapter, one of the clones, 3B6, is selected for closer examination. 3B6 is mapped to mouse chromosome 14, and its gene structure is determined using predictive methods in concert with RT-PCR, cloning, and sequencing. It is found to have a total of 22 exons which are transcribed into several alternatively spliced transcripts. Conceptual translation of the transcripts shows that the 3B6 products contain two nucleic acid binding domains, which are not disrupted by the alternative splice events. Expression of 3B6 in differentiating ES cells, developing mouse embryos, and adult tissues is also examined. 3B6 is shown to be expressed in neural precursors, in embryos starting after 10.5 dpc lasting until after 14.5 dpc, and at lower levels in each adult tissue tested. The expression pattern in the 12.5 dpc embryo is found to be widespread. Overall, the results indicate that 3B6 may encode a nucleic acid binding protein with possible roles in cell fate decisions.

4.2 Introduction

4.2.1 Selection of clone 3B6 for further experimentation

Following the initial RT-PCR screen of ESTs selected for follow-up, described in section 3.3.4, RED clone 3B6 was chosen for further characterization. The lab was incubating an interest in nucleic acid binding genes involved in neural differentiation and development, and clone 3B6 fit the criteria for further follow-up. The EST sequence itself showed no similarity to any sequences in the public databases at that time, nor did the UniGene cluster into which it was grouped. An especially interesting finding was that a conceptual translation of the 3B6 EST sequence contained similarity to the SMART HMM representing CCCH zinc finger domains. In addition, the expression pattern seen for 3B6 in the initial RT-PCR screen was suggestive of involvement in neural differentiation and development. It was expressed at a higher level in 4-/3+ neural precursor-enriched cultures than in spontaneously differentiating 4-/3- cultures, and also its expression was increased in embryos after 10.5 dpc, concurrent with an increase in neural cell proliferation in the developing embryo. Also contributing to its selection for further follow-up was the fact that the 3B6 EST had previously been used by Dr. Hui Xu, formerly of the Rancourt lab, to screen a 12.5 dpc embryonic cDNA library in bacteriophage λGT10, resulting in the isolation of a single 787 bp clone, designated clone 4.2. Clone 4.2 contained the entire 443 bp sequence of the EST, plus 140 bp of additional 5' sequence and 204 bp of additional 3' sequence. All of these factors taken into account at the time this work was undertaken made 3B6 an attractive candidate for further experimentation. The remaining work of this project revolves around its characterization.

4.2.2 Gene structure prediction and Genscan

Among the central strategies for enhancing the value of eukaryotic genomic sequence is the goal of identifying all of the regions representing genes. This task has received considerable attention and numerous methods for gene prediction have been developed since the early 1990s, with the initial attempts to apply sequence recognition approaches to gene identification[192,193]. Generally, gene finding approaches may be

categorized according to three broad strategies: content-based, site-based, and comparative[194]. Content-based approaches rely on the overall properties of sequences contributing to genes such as codon usage and sequence complexity. Site-based methods use the presence of patterns and profiles describing sequence features. Finally, comparative methods use sequence similarity to known genes for prediction of new ones. Gene finding is a highly complex problem and none of the available methods show ideal performance. However, when multiple methods are combined, more robust predictions may result.

In a comparative evaluation of gene finding programs, several general trends regarding accuracy were noted[195]. First, the accuracy of exon prediction is dependent on the length of the exon, with the most accurate predictions occuring for exons ranging in size from 70-200 bp long. Second, internal exons are far more likely then terminal exons to be detected accurately, which is attributable to the poor ability to detect untranslated regions, and start and stop codons. More major pitfalls of gene structure prediction arise from inability to successfully predict events that commonly give rise to genomic complexity, such as non-canonical splice sites, overlapping genes, multi-exon untranslated regions, and alternative promoters and poly-A signals[196]. In addition, gene prediction methods are generally very poor at detecting alternatively spliced exons.

One of the most widely used methods for gene prediction is Genscan[151,197]. Genscan uses what its developers describe as a 'probabilistic model' of gene structures, effectively combining content- and site-based methods. In the comparative evaluation of gene predictive methods by Rogic and colleagues[195], Genscan was one of only two tools found to be reliable predictors of exons. When the gene structure for 3B6 was initially being explored, Ensembl contig view included Genscan predictions as one of the annotation tracks. At the time Genscan was the only external *ab initio* gene finding method being used to annotate the mouse genome. Although performance of Genscan had been shown to perform substantially better than other available gene finding methods, inspection of the Genscan predictions in the context of the Ensembl contig view with its tracks of evidence suggested that it clearly had limitations. The EST data track

was particularly illustrative of the shortcomings of Genscan. It showed, among other things, spliced ESTs both overlapping and flanking predictions, indicative of a failure to predict terminal exons.

Overall it has been recognized that integrated approaches combining all three gene finding strategies, content-based, site-based, and comparative, would provide added value to gene predictive methods. Indeed such methods have been under development over recent years. Previously however, it has been noted that use of gene annotation platforms with graphical interfaces, such as MAGPIE or Ensembl, is highly complementary to gene finding methods for one-at-a-time gene structure prediction[196]. During elucidation of the 3B6 gene structure presented in this thesis, the Ensembl genome browser containing Genscan predictions was utilized in this manner, along with other independent software.

4.2.3 Alternative pre-mRNA splicing

The transcriptional landscape of the mammalian genomes is incredibly complex. At the time of the writing of this thesis, the most recent assembly of the mouse genome annotated by Ensembl (assembly NCBIm36 version 36.39)[144] is annotated as having approximately 24 000 genes, while the mouse transcriptome is estimated to contain in excess of 181 000 independent transcripts[198]. This order of magnitude discrepancy illustrates the critical nature of alternative pre-mRNA splicing, which gives rise to much of the variation in the transcriptome.

An estimated 94% of genes are multi-exon, interspersed with intronic sequences[199]. Transcription produces pre-mRNA containing intronic sequences, which must be processed out of the transcript to form the mature mRNA. Excision of introns from the transcript and ligation of exons is directed by sequence motifs marking the boundaries of the intron/exon junctions. The 5' splice site of an intron is marked by a GU dinucleotide, while the 3' splice site is designated by an AG dinucleotide, also referred to as the splice donor and splice acceptor sites, respectively[200]. The location of the 3' splice site is also designated by the presence of a branch point sequence containing a conserved

A residue, followed by a poly-pyrimidine tract upstream of the AG dinucleotide. These canonical splice sites are present in over 98% of exons, and are recognized by the major U2-type spliceosome, a macromolecular complex which catalyses the reactions required for splicing[201,202].

Several types of alternative splicing exist, classified based on where the splice variation occurs [Figure 12] The most common type of alternative splicing is exon skipping, in which a cassette exon is excluded from a transcript, accounting for 38% of alternative splicing events[202]. The use of alternative splice donor and splice acceptor sites accounts for 18% and 8% of splicing events respectively. Alternative splicing leading to retention of an intron, or inclusion of mutually exclusive exons also occurs[202]. Splicing patterns are highly regulated and are often specific to different tissue types, temporal cues, or extracellular stimuli[200]. Alternative splicing is of particular prevalence in the nervous and immune systems, where it may play roles in precise control of cellular differentiation, activation, and functional specificity[203].

4.2.4 Real time PCR analysis

Real time PCR analysis combines RT-PCR with fluorescent detection methods to quantify and compare expression levels of genes of interest under varying conditions. It is an appropriate method for use when the amount of available sample RNA is small, and subtle differences in expression may be detected. It also eliminates the need for end-point analysis, which may be skewed by reactions reaching their saturation point. Relative quantification using SYBR green I, a double stranded DNA specific dye, has become a popular method for analysis of gene expression. In this method, a measurement of fluorescence is taken at the end of each elongation step. At this point, all reaction products are expected to be double stranded, and therefore fluoresce in the presence of SYBR green I. With each successive cycle, increasing fluorescence occurs and is monitored. Plotting each measurement results in generation of a sigmoid curve illustrating the progression of the PCR. During the linear phase of the reaction, the exponential amplification of the target molecule is described by:

$$N_C = N_O \times E^C$$

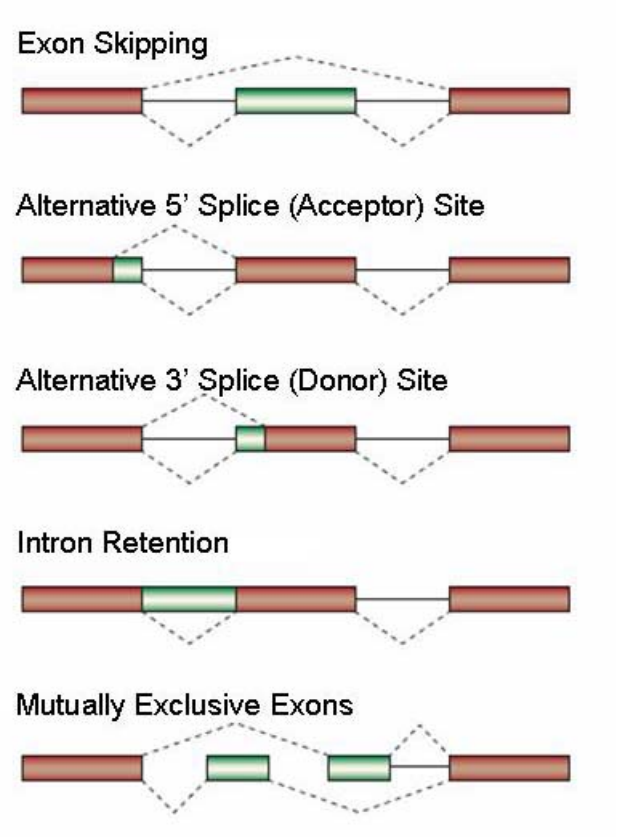


Figure 12: Types of alternative splicing. Adapted from Ast, 2004[202].

where N_C is the final number of target molecules, N_O is the initial number of target molecules, E is the efficiency of the reaction, and C is the number of cycles[204]. In an ideal PCR, E=2, indicating that each target molecule is exactly doubled during each cycle of the reaction. As it turns out, molecular biology is not ideal, and reactions in which E=2 seldom occur. Rather, reaction efficiencies vary with primer and amplimer sequence, amongst other variables.

Data analysis for real-time PCR revolves around determination of a threshold cycle, C_T, for each reaction. The C_T is reached during the linear phase of the reaction, and is the cycle number at which a set amount of target DNA has been produced, and therefore a set level of fluorescence has been met. The most commonly used method of analysis for relative quantification is the $\Delta\Delta C_T$ method, as described by Livak and Schmittgen[162]. The fold change in target gene expression between two conditions, control and treated, is expressed by $2^{-\Delta\Delta C_T}$ where:

$$\Delta\Delta C_T = \left(C_{T\,target} - C_{T\,reference}\right)_{treated} - \left(C_{T\,target} - C_{T\,reference}\right)_{control}$$

A critical assumption that must be applied to this analysis method refers to the reaction efficiency. For the above equation to be valid, the amplification efficiencies of the target and reference genes must be approximately equal. Reaction efficiencies can be established by completing replicates of the same reaction using a dilution series of cDNA as template. Plotting the C_T versus the starting template amount yields a standard curve, and the slope of this curve relates to the reaction efficiency:

$$E = 10^{\frac{-1}{slope}}$$

In most applications of real time PCR, calculated efficiencies are not applied, and are all too often ignored. It has previously been shown that the $\Delta\Delta C_T$ method is very sensitive to variation in PCR efficiency, and even as small a change in efficiency as 0.04 units can skew observed fold changes by 4-fold[204]. Therefore, to ensure validity of real time PCR results analyzed by the $\Delta\Delta C_T$ method, the reactions in question must be optimized so that near equal efficiencies are achieved. Alternatively, analysis methods which correct for variation in efficiency may be used.

Dr. Michael. W. Pfaffl proposed a new model for relative quantification using real-time PCR[163]. In this model, observed C_T values for each set of reactions are corrected for reaction efficiency. Fold change in target gene expression is expressed as the efficiency corrected difference in C_T between two conditions, control and treated, in comparison to a reference gene:

$$fold\ change = \frac{\left(E_{target}\right)^{\Delta C_T(control-treated)}}{\left(E_{reference}\right)^{\Delta C_T(control-treated)}}$$

A tool for analysis of real time PCR data, REST (relative expression software tool), has been developed based on the Pfaffl model; REST also incorporates a statistical test of significance[164]. Typically in $\Delta\Delta C_T$ analysis, tests of significance such as t-tests are used, though the appropriateness of these tests may be called into question. T-tests depend on the assumption that a given dataset is normally distributed, which for ratios comparing expression levels is doubtful. In contrast, REST includes a randomization test as its test of significance. If the fold change observed had arisen by chance, then randomizations of the data between control and treated groups would be expected to yield similar fold change results. Over many randomizations of the data, the proportion of combinations giving results strongly supporting a difference between the paired datasets is established and expressed as a P-value. If the resulting P-value is very small, then the hypothesis that the observed fold change difference did not occur by chance is supported. Since the C_T values used in the above equation are group means of C_T values arising from multiple, commonly nine, replicates of a reaction, this randomization method can be appropriately applied. The REST method for analysis of real time PCR data therefore effectively reduces the number of assumptions made in the analysis of such data. Additionally, REST has been wrapped in a user friendly interface.

4.3 Results

4.3.1 The genomic structure of 3B6 has 22 exons and produces several transcripts

In order to map 3B6 to the genome, the sequence of 3B6 clone 4.2 which was isolated from a screen of a bacteriophage λGT10 12.5 dpc cDNA library, was compared

to the Ensembl mouse genome assembly 3 version 7.3a.1 using SSAHA. A single hit in four segments indicative of multiple exons with near perfect identity was returned. The genomic region returned was annotated by Ensembl with two non-overlapping Ensembl predicted transcripts, one *ab initio* Genscan gene prediction, and a multitude of overlapping ESTs. Given that Genscan predictions have been shown to have limited reliability, and that ESTs annotated to the region did not remain within the bounds of the prediction, effort was made to amalgamate the given information in order to obtain a better predictive model for the genomic structure of the 3B6 gene.

Each of the ESTs annotated to the region bounded by the Genscan prediction was downloaded and ordered according to its position within the genomic region. EST sequences were collapsed into a single sequence by progressive contig assembly using GeneTool. Where discrepancies existed for particular bases, the base appearing in the majority of the overlapping sequences was accepted. If only two overlapping sequences were available and a discrepancy occurred, then the ESTs were aligned to the genome, and the base present at the corresponding genomic position was accepted. A single sequence representing a predicted cDNA for the 3B6 gene was obtained, and this sequence was compared to the genome again using SSAHA to facilitate prediction of exon boundaries, yielding a genomic structure model.

Primers for several exons were designed that could be paired in a combinatorial fashion to test the validity of the predicted gene structure via RT-PCR from 12.5 dpc mouse embryo cDNA. The majority of the PCR reactions performed spanned multiple predicted exons, therefore if the PCR product generated was of similar size to that predicted for the primer set, then it could be inferred that the gene structure prediction was approximately accurate. Indeed, for the RT-PCRs performed against 12.5 dpc cDNA with varying combinations of 3B6 primers the predicted amplimer sizes were produced, as shown in Figure 13.

While attempting to confirm the predicted 3B6 gene structure, it appeared that several reactions were generating more than one discrete product, as indicated by the appearance of diffuse bands on the agarose gels [Figure 13, B, C]. To resolve whether

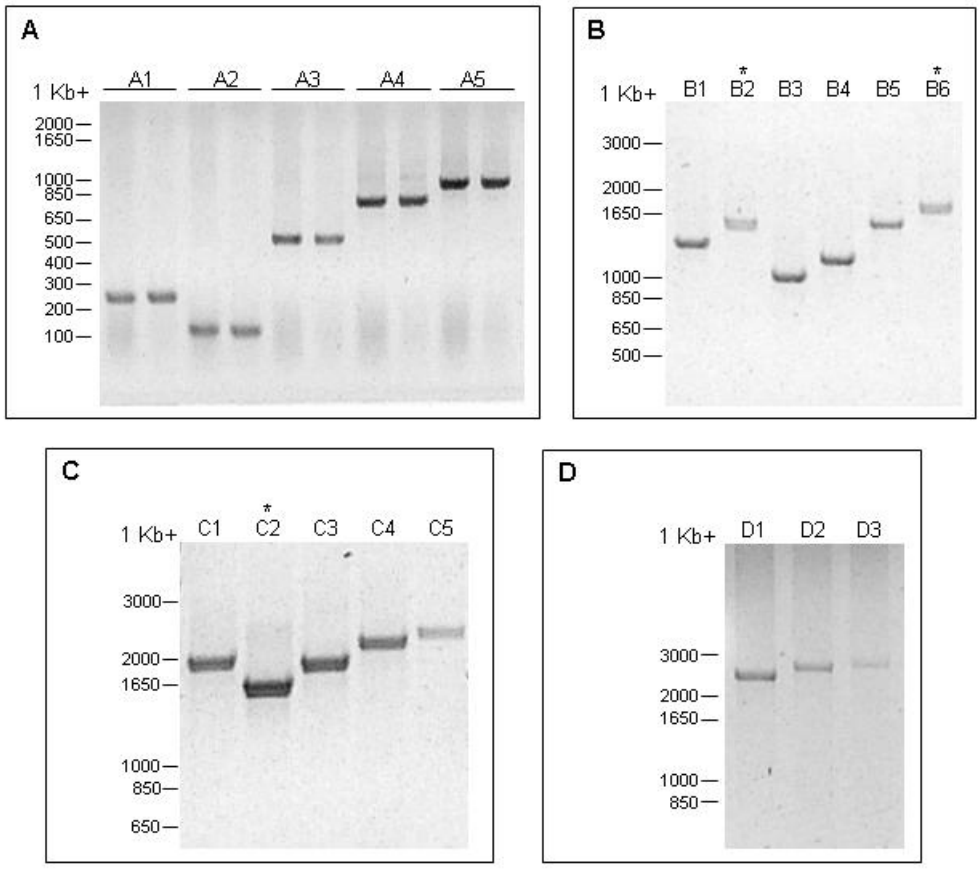


Figure 13: RT-PCR for validation of 3B6 predicted gene structure. The 3B6 gene structure prediction was validated using RT-PCR with 12.5 dpc embryo cDNA as template. All primers used were 3B6-nF or nR, where the n corresponds to the exon number that the primer falls within. The primers used and expected amplimer size for each lane is as follows:

Panel A: **A1**, 6F-6R, 246 bp; **A2**, 7F-7R, 130 bp; **A3**, 4F-6R, 520 bp; **A4**, 6F-9R, 768 bp; **A5**, 6F-11R, 928 bp.

Panel B: **B1**, 6F-13R, 1321 bp; **B2**, 6F-16R, 1573 bp; **B3**, 4F-9R, 1061 bp; **B4**, 4F-11R, 1221 bp; **B5**, 4F-13R, 1614 bp; **B6**, 4F-16R, 1866 bp.

Panel C: **C1**, 4F-16R, 1866 bp; **C2**, 6F-16R, 1573 bp; **C3**, 7F-19R, 1798 bp; **C4**, 7F-21R, 2021 bp; **C5**, 7F-22R, 2119 bp.

Panel D: **D1**, 1F-19R, 2693 bp; **D2**, 1F-21R, 2916 bp; **D3**, 1F-22R, 3005 bp.

Multiple products were generated by several reactions, appearing as bands with a diffuse doublet appearance. This is most evident in lanes B2, B6, and C2, indicated by *.

there were multiple products and to further confirm the gene structure, reaction products from every primer combination were cloned into pCR-Topo2.1, using TA cloning methods. A minimum of eight colonies representing each PCR product were screened by restriction digestion and sequenced. Note, the large PCR products seen in Figure 13, D were not successfully cloned. This issue will be revisited in later chapters.

The combination of PCR, cloning and sequencing revealed that the predicted structure for the 3B6 gene, which will herein be referred to simply as 3B6, was valid. 3B6 consists of 22 confirmed exons spanning a 62.1 Kb region of mouse chromosome 14E2.3 [Figure 14, top]. Sequencing further revealed that three of the exons are subject to alternative splicing. Exon 14, which is 72 bp long, was found to be a cassette exon. It was this alternative splicing of exon 14 that contributed to the diffuse doublets seen on the agarose gels. Exons 13 and 9 both have alternative splice acceptor sites. Use of the normal (most frequently sequenced) splice acceptor site of exon 13 results in usage of the predicted 132 bp exon in the resulting transcript. Use of the alternative splice acceptor results in loss of 9 bp from the transcript. The most frequently sequenced version of exon 9 is 156 bp long and use of its alternative splice acceptor results in addition of 15 bp to the transcript. The sequence of each of these exons is given in Appendix A.

With three exons subject to alternative splicing, eight permutations of these splice events are in theory possible. However, only five of the possible eight transcripts were identified experimentally [Figure 14, bottom]. By far the most common transcript sequenced was that which was initially predicted, containing exon 14 and not using either of the alternate splice sites. This transcript has been designated 1a. Transcript 2a lacks exon 14, while transcripts 1b and 2b are identical to 1a and 2a respectively, except that they both use the alternative splice acceptor site in exon 13. Transcript 1c is also identical to transcript 1a, except that the alternative splice acceptor site in exon 9 is used. It must be noted here that the ends of the gene have not been cloned and sequenced due to the placement of the primers used in this experiment, and therefore the predicted untranslated region (UTR) sequences for 3B6 have not been verified experimentally. Rapid amplification of cDNA ends (RACE) was attempted for confirmation of the UTR

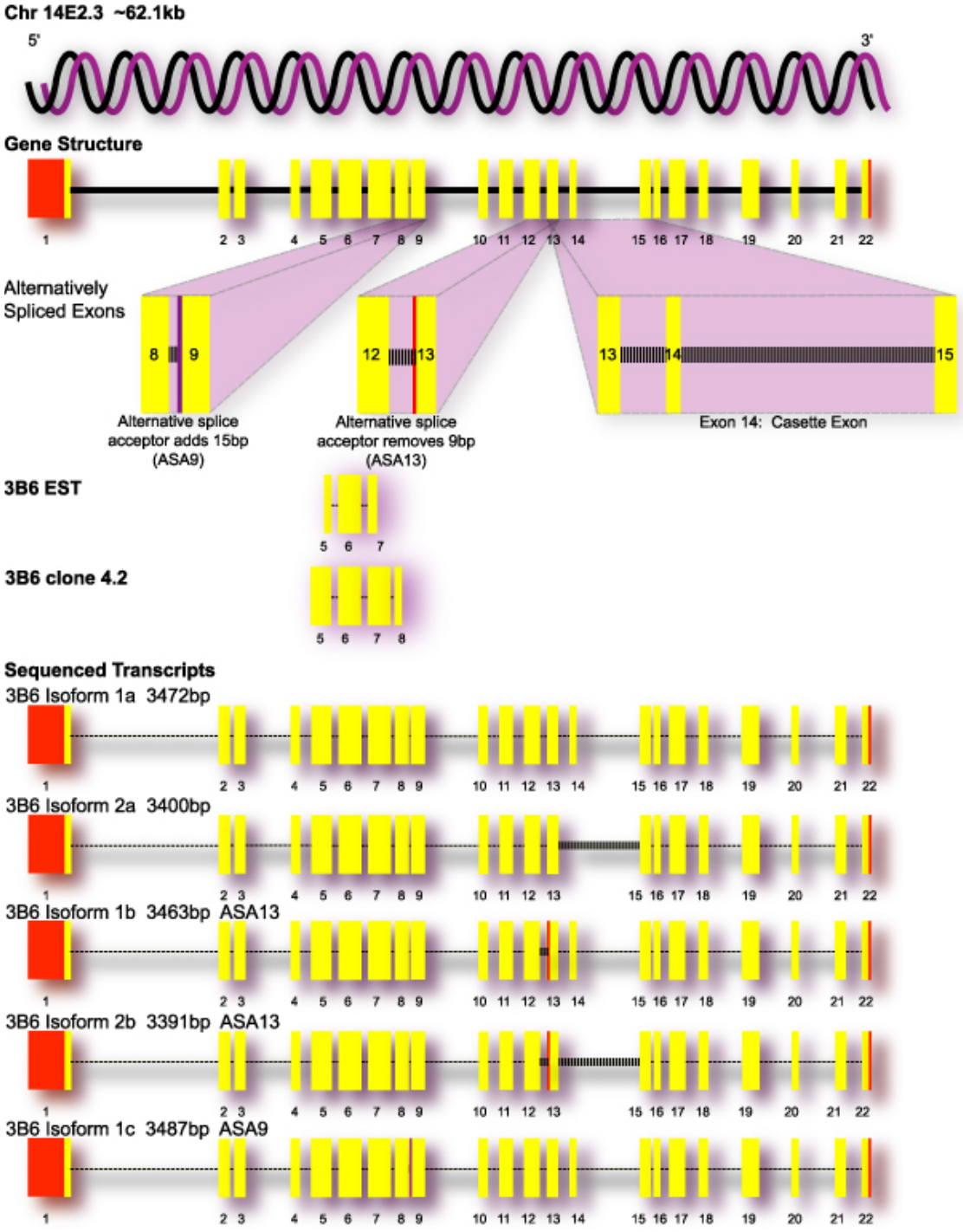


Figure 14: The 3B6 genomic structure and splice isoforms. Please see following page for legend.

Figure 14: The 3B6 genomic structure and splice isoforms. 3B6 consists of 22 exons, spanning a 62.1 Kb region of mouse chromosome 14E2.3. Sequencing further revealed that three of the exons are subject to alternative splicing. Exon 14, which is 72 bp long, was found to be a cassette exon. Exons 13 and 9 both have alternative splice acceptor sites. Use of the alternative splice acceptor in exon 13 (ASA13) results in loss of 9 bp from the transcript. Use of the alternative splice acceptor in Exon 9 (ASA9) results in addition of 15 bp to the transcript. The most common transcript sequenced, designated isoform 1a, contains exon 14 and does not using either of the alternate splice sites. Isoform 2a lacks exon 14. Isoforms 1b and 2b are identical to 1a and 2a respectively, except that they both use ASA13. Isoform 1c is also identical to isoform 1a, except that ASA9 is used. The predicted UTR sequences for 3B6 have not been verified experimentally, and are represented in red, flanking the yellow exon 1 and 22.

sequences, in collaboration with others in the lab; however these experiments were discontinued due to technical difficulties.

Each of the transcripts contain a complete open reading frame starting with an ATG within exon 1 and ending with a TGA stop codon within exon 22. Conceptual translation of each transcript indicated that none of the alternate splice events lead to a frameshift, since all of the numbers of base pairs added or lost are divisible by three. The alternate splice events also do not lead to any changes of single amino acids at the exon junctions. In addition, prediction of functional protein domains within the protein product from transcript 1a was completed using InterPro Scan. Similarities to the SMART HMMs for CCCH zinc finger domains and RNA recognition motifs (RRM; also known as RNA binding domain, RBD) were found with P-values of 2.9e-4 and 9.1e-3 respectively. These functional motifs are also predicted in each of the putative proteins, as they are not predicted to fall into regions affected by alternative splicing.

In attempts to confirm the number of 3B6 splice isoforms, northern blotting was performed. Over numerous attempts, 3B6 was not detected via northern blotting, and further attempts were abandoned. In addition, repeated screens of a 12.5 dpc embryo cDNA library in bacteriophage λGT10 were conducted in order to yield new sequence for 3B6, and possibly to detect additional isoforms. Neither new sequence, nor new isoforms were detected in these screens.

4.3.2 3B6 is expressed in cultures enriched for neural precursors, developing embryos, and adult tissues.

The initial RT-PCR expression screen discussed in section 3.3.4 [Figure 11] showed that 3B6 is expressed at a higher level in neural precursor enriched cultures than in spontaneously differentiating ES cell cultures, and also that is expressed more highly in 12.5 and 14.5 dpc than in 10.5 dpc embryos. To reconfirm these results, real time PCR was used. Though real-time PCR was not absolutely necessary for this experiment, it was used as the route for establishment of the technique in the Rancourt lab. In addition, by completing this experiment in real time, the assays were developed and ready for use in experiments which were anticipated as the project progressed.

The experiment was carried out using new RNA prepared from the same sources used in the earlier RT-PCR screen. 12.5 dpc and adult brain samples were also used in this experiment: 12.5 dpc brain in order to establish if any evidence for enrichment of 3B6 expression in the embryonic brain is present, and adult brain in order to examine whether 3B6 is expressed in adult mice. A pan-3B6 approach was taken in this and subsequent experiments, using primers that detect all possible transcript of 3B6. Standard curves were established for each reaction to be completed, to allow for calculation of reaction efficiency. Real time PCR reactions using 3B6 and GAP-DH primers were carried out in triplicate for both biological and technical replicates, resulting in nine data points for each reaction. Data was analyzed using the REST method and propagation of error [Appendix B]. Results were also analyzed using the $\Delta\Delta C_T$ for comparison, and are shown in Appendix C. Results were consistent with the previous RT-PCR experiment [Figure 15, A]. In comparison to the expression level in ES cells, 3B6 expression was 2.5-fold higher in neural precursor enriched 4-/3+ cultures, as well as 9.7- and 12.9-fold higher in 12.5 and 14.5 dpc embryos, respectively. 3B6 expression was also found to be 13.4-fold higher in 12.5 dpc brain and 2.1-fold higher in adult brain compared to ES cells. No significant difference in expression between ES cells and either spontaneously differentiating 4-/3- cells or 10.5 dpc embryos was found. The 3.7-fold difference in expression seen between 12.5 dpc embryo and 12.5 dpc brain was found to be significant by analysis with REST, as was the 11.4-fold difference observed between 12.5 dpc brain and adult brain.

The highest rate of neural proliferation in the developing mouse brain is seen starting at approximately 11.5dpc and extending through to approximately 15.5 dpc[191]. To characterize the expression of 3B6 during this period, embryos were collected at each time point from 10.5 dpc to 15.5 dpc for preparation of cDNA. RT-PCR was performed using primers for 3B6 and GAP-DH. Results indicate that 3B6 is expressed at a stable level from 10.5 dpc to 14.5 dpc, dropping off at 15.5 dpc [Figure 15, B]. This result is inconsistent with previous results indicating a lower level of expression at 10.5 dpc, and will be discussed later.

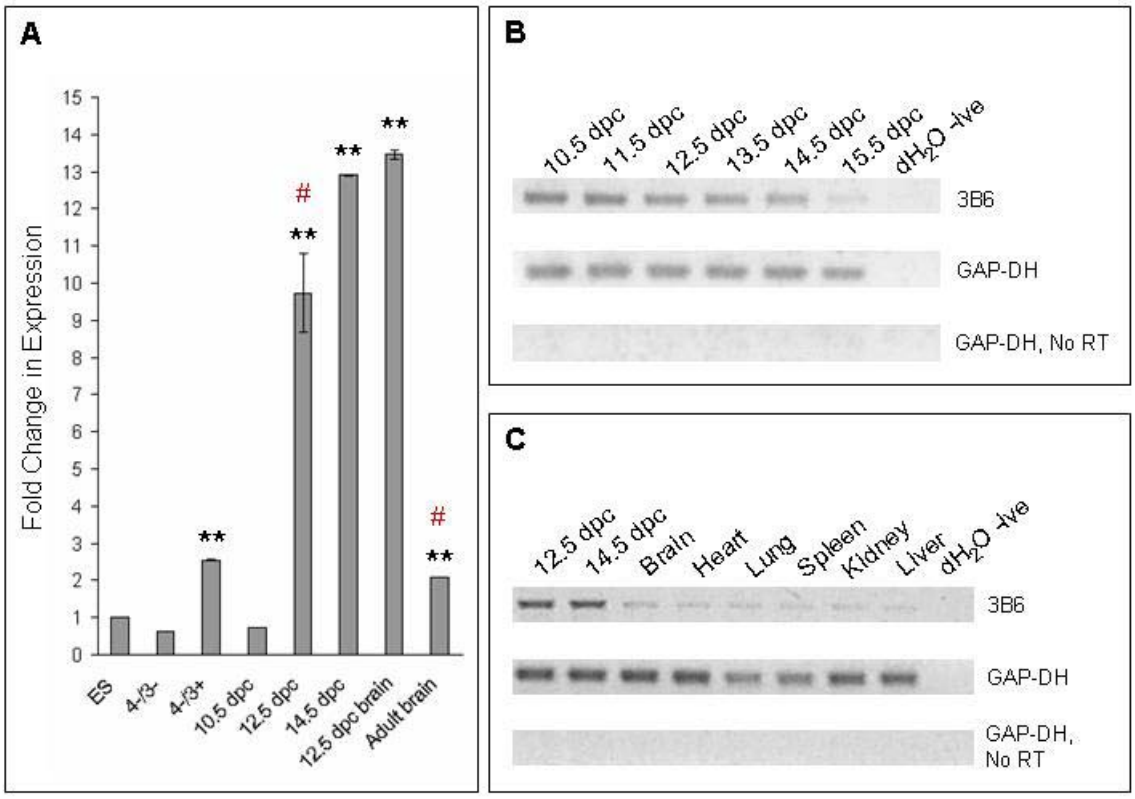


Figure 15: Expression of 3B6 developing embryos and adult tissues. Panel A: Real time PCR examining fold change in 3B6 expression in comparison to that seen in ES cells. Results were analyzed using REST with normalization to GAP-DH, and propagation of error. ** represents a significant fold change difference in expression compared to ES cells ($p<0.001$). # indicates a significant fold change in expression compared to 12.5 dpc brain ($p<0.001$). This relationship was only tested for 12.5 dpc embryos and adult brain. Panel B: RT-PCR analysis of 3B6 expression at timepoints spanning the maximal neural proliferation period in mouse embryos. Panel C: RT-PCR analysis of 3B6 expression in selected adult tissues in comparison to 12.5 dpc and 14.5 dpc embryos. GAP-DH not RT controls are replicates of the GAP-DH PCR reactions using templates to which no reverse transcriptase (RT), was added during the cDNA synthesis step.

Expression of 3B6 in the adult mouse was demonstrated in the real time PCR experiment. To examine whether this expression was restricted to the brain or was more widespread, cDNA from adult mouse tissues were used in an RT-PCR experiment. Results indicated that expression of 3B6 in adult mice was not restricted to the brain, but rather was consistently seen across all tissues tested [Figure 15, C]. The level of expression seen in adult tissues was significantly lower than that seen in 12.5 dpc and 14.5 dpc embryos, consistent with results from the real time experiment.

4.3.3 3B6 expression is not limited to the nervous system in the embryo

In light of results suggestive of a role for 3B6 in neural development, but also indicating generalized expression in adult mice, *in situ* hybridizations were pursued in order to establish whether 3B6 expression was spatially restricted to the nervous system in embryos. Paraffin embedded sagittal sections of 12.5 dpc mouse embryos were subjected to *in situ* hybridization using varying DIG labelled RNA probes directed at 3B6. A wide variety of probes were generated, including full length 3B6 RNA, and numerous sub-fragments yielding complete and redundant coverage of the 3B6 sequence. No staining was ever detected using paraffin embedded sections, despite numerous of repetitions of the protocol with variations for attempted optimization.

Standard protocols for *in situ* hybridization in the Genes and Development Research Group (GDRG) at the University of Calgary call for the use of cryopreserved embryos. Para-sagittal cryosections of 12.5 dpc mouse embryos were received as a generous gift from the lab of Dr. Carol Schuurmans in the GDRG. *In situ* hybridization using these sections were performed in collaboration with her lab, using DIG labelled RNA probes corresponding to the full length 3B6 1a transcript. Hybridization was detected using the anti-sense 3B6 probe [Figure 16], while the control slides hybridized with the control sense probe showed no staining [not shown]. 3B6 is expression is not restricted to the nervous system. 3B6 is expressed throughout the nervous system and mesenchymal tissues. Expression also appears to be excluded from the developing heart and liver.

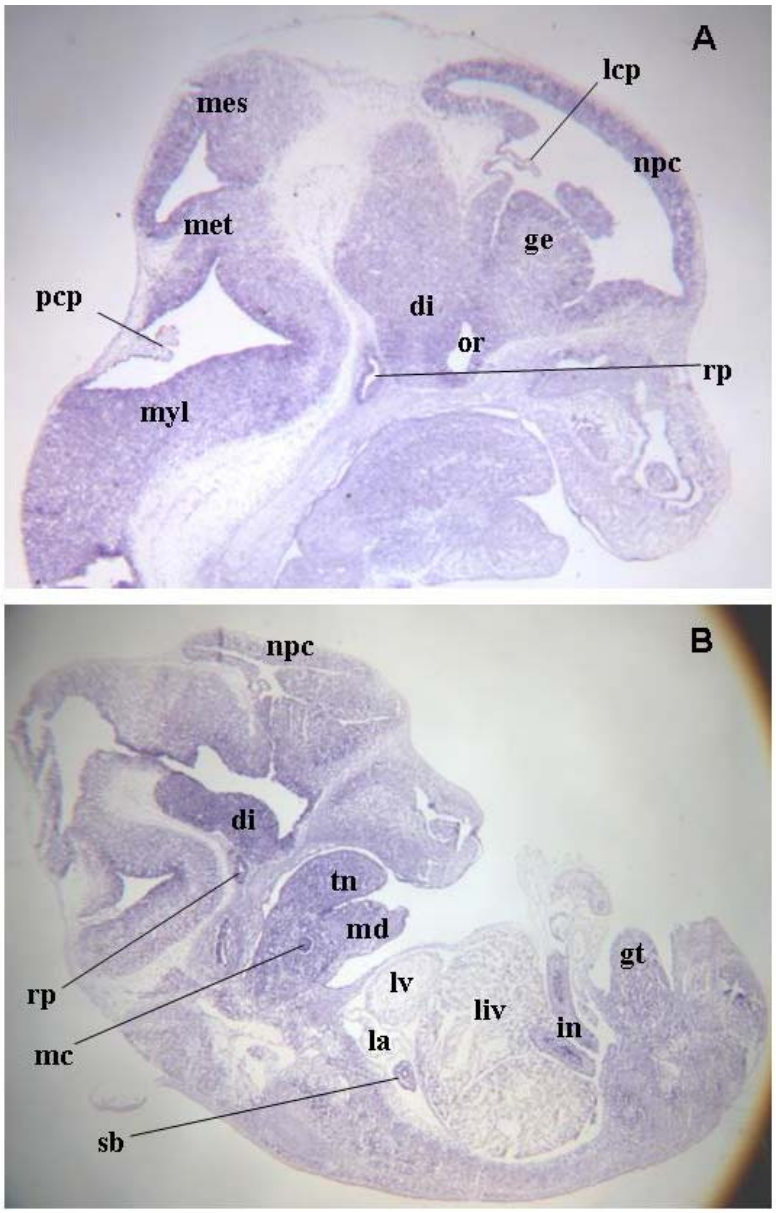


Figure 16: 3B6 is expressed throughout the nervous system and mesenchymal tissues of 12.5 dpc mouse embryos. Distribution of 3B6 transcripts was analyzed by RNA in situ hybridization with 3B6 probes on 12.5 dpc para-sagittal sections. Panel A shows distribution of 3B6 in the developing head, while panel B shows the entire embryo. Of interest is the apparent exclusion of staining from the developing heart and liver. Di, diencephalon; ge, ganglionic eminence; gt, genital tubercle; in, intestine; la, left atrium; lcp, lateral choroid plexus; liv, liver; lv, left ventricle; mc, Merckel's cartilage; md, mandible; mes, mesencephalon; met, metencephalon; myl, mylencephalon; npc, roof of neopallial cortex; or, optic recess of diencephalon; pcp, posterior choroids plexus; rp, Rathke's pocket; sb, segmental bronchus; tn, tongue.

4.4 Discussion

4.4.1 The 3B6 gene structure

Prediction of the gene structure for 3B6 started with mapping of 3B6 clone 4.2, isolated from a bacteriophage λ cDNA library, to the mouse genome using a sequence similarity search. It evolved into an intuitive method using the graphical display of Ensembl, including Genscan predictions and EST evidence, in combination with a simple sequence assembly method. The Genscan prediction given for the region 3B6 mapped to was fairly accurate, but use of the EST evidence allowed for refinement of the model in regions of apparent weakness: it had missed the first two exons of the gene, and the last as well as several internal exons were incomplete. Effectively, the method used during this work integrated all three approaches to gene structure prediction: content- and site-based, plus comparative methods were combined. This yielded a prediction for 3B6 which was subsequently validated by RT-PCR. Given the density of the EST data and the increasing numbers of *ab initio* gene predictions from different methods displayed in Ensembl, this method could easily be applied to other genes of interest.

One area where the gene structure prediction developed during this project was weak, was in its failure to predict that different splice isoforms may exist. In the RT-PCR experiments described herein, doublets were repeatedly detected from some reactions. This led to the need for cloning and sequencing all of the reaction products, which revealed that exons 9, 13, and 15 were subject to alternative splicing. In all, five splice variants of 3B6 were isolated and sequenced. The relevance of each of the transcripts is unknown, and elucidation of their individual expression patterns was determined to be beyond the scope of the project, therefore in subsequent expression studies, a pan-3B6 approach was taken.

A question that was approached but not successfully answered was that of how many transcripts exist in total and can be detected by northern blotting. It is possible that more transcripts exist, but were not isolated by sequencing. Northern blotting became a sizeable stumbling block of this project and attempts were eventually abandoned. 3B6 expression may be beneath the limit of detection for northern blotting.

Another piece of information missing from the 3B6 gene structure relates to the UTRs of the gene. Although UTRs were included in the gene prediction, primers designed for its validation only encompassed the predicted coding sequence. RACE was carried out for both the 5' and 3' ends of the gene but failed to provide any new sequence. This could be attributable to inaccuracy of the prediction in the UTRs, resulting in the amount of sequence between the known and real ends being larger than can effectively be isolated by RACE. Of course the possibility that RACE failed for technical reasons cannot be discounted.

In the version of Ensembl available at the time of writing (assembly NCBIm36 version 36.39) there were three transcripts predicted for 3B6. Only one of them, with Ensembl transcript identifier ENSMUST00000022715, shares the same 5' exon with those that have been sequenced in this project. It is exactly the same including the UTR, as the predicted 3B6 gene model. More discrepancy exists at the 3' end of the gene. One transcript, ENSMUST00000100327, contains the predicted 3' exon lacking any UTR. The other two transcripts predict alternate 3' exons. The 3' exon included in ENSMUST00000022715 is suspect, as it is separated from the preceding exon by a predicted intron of only two base pairs. The 3' exon proposed in the transcript ENSMUST00000077531 is entirely UTR, as are the preceding two exons due to a frameshift in the transcript. As a whole, not one of the transcripts predicted by Ensembl is exactly like those confirmed by this project. Though the internal predicted exons are generally accurate, the terminal exons are not clear, again illustrating the poor ability of predictive methods to detect terminal exons.

A putative full length cDNA clone for 3B6 was also available from the NIH Mammalian Gene Collection (MGC; http://mgc.nci.nih.gov/) at the time of writing. This clone, with IMAGE ID 6827086 and accession BC066051, includes the same coding sequence as predicted for 3B6 transcript 1a, with the exeption of exon 22. An alternative exon 22 sequence, encoding a different predicted C-terminal for the peptide, is presented. Also available in the MGC sequence is an apparent 5′ UTR sequence of 412 bp. The alternative exon 22 sequence also includes a 3′ UTR of 680 bp. Overall, this cDNA clone

supports the prediced transcript 1a presented here, although there is evidence for mutually exclusive 3′ terminal exons. It would be interesting to persue further experimental confirmation of the alternative 3′ exon, as well the UTR sequences, using the MGC sequence and methods similar to those used to describe the gene structure of 3B6.

Conceptual translation of the transcripts sequenced herein indicates that none of the alternative splicing events lead to frameshifts or changes in amino acid identites at exon junctions. This is not surprising in that it has previously been shown that alternative splicing is strongly biased towards preservation of the protein coding frame[203]. In addition, two conserved nucleic acid binding domains, and RRM and a CCCH zinc finger, were predicted to reside in the protein sequence. Neither domain is impacted by alternative splicing. The functions of the splice variants of 3B6 cannot be predicted at this time, but the possibility that they have differing temporal and/or spatial specificity, differing binding partners or affinity for nucleic acids, and a host of other possibilities cannot be discounted. Regardless, the evidence for presence of two functional domains piqued the interest in continued exploration of 3B6, and will be discussed in detail in later chapters.

4.4.2 A role for 3B6 in neural development?

Taken as a whole, the PCR-based expression studies outlined herein are suggestive of a role for 3B6 in neural development. The results of the original RT-PCR screen described in section 3.3.4 were corroborated by the real time PCR results. Expression of 3B6 in the neural precursor enriched 4-/3+ cultures is at least 2-fold higher than that seen in the spontaneously differentiating cultures. This difference in expression levels, though significant, is not as great as seen for embryos. It is possible that the expression level of 3B6 in the 4-/3+ cultures is artificially low due to variation in cell type as the Bain ES cell neural differentiation method is known to produce heterogeneous cultures[116,135]. Given the comparatively high level of expression in embryonic brain, it would be interesting to examine the expression of 3B6 in a pure population of neural precursors.

The real time PCR experiment further established that 3B6 expression is not limited to the embryo, as indicated by the observed expression in the adult brain. This expression level however, is significantly lower than that seen in the embryonic brain, again suggesting that 3B6 may play a role in embryonic brain development. Persistence of its expression in the adult brain could indicate a maintenance role for 3B6 in mature neural cells, or possibly a role in maintenance or differentiation of neural stem cells.

The significant difference in 3B6 expression noted between 12.5 dpc embryos and 12.5 dpc brain was thought to suggest that its expression may be restricted to, or at least enriched in the embryonic brain. The level of expression seen in the complete embryo could originate from expression in the brain. Expression levels seen in samples from isolated 12.5 dpc brain would not be diminished by other non-expressing tissues, and therefore would appear higher than in complete embryos, as observed. No conclusions regarding the distribution of 3B6 expression can be made based on this result, however. This result merely reconfirmed the need for *in situ* hybridizations to determine the expression patterns in the embryo.

Results of the time course experiment spanning embryonic development from 10.5 dpc to 15.5 dpc were not consistent with those previously seen in the real time PCR experiment for 10.5 dpc embryos, in that expression of 3B6 at 10.5 dpc was found to be approximately equal to that seen in 12.5 and 14.5 dpc embryos. As the greatest rate of neural proliferation during mouse development occurs from approximately 11.5 to 15.5 dpc[191], it was anticipated that results might show expression of 3B6 increased after 10.5 dpc, persistent until or though 15.5 dpc. The reason for the observed expression at 10.5 dpc is not known, but might be attributable to inherent, random error in measurement of the age of the embryos harvested. It is certainly not impossible that the embryos harvested for this series of experiments were hours older than desired, particularly if they were collected later in the day than intended.

An ideal follow-up to the time course experiment would be a real time PCR study examining the profile of expression of 3B6 starting with an earlier embryonic time point, such as 8.5 dpc, through every 12 hours until birth. Performing this experiment under the

rigidity of real time PCR would certainly provide a more comprehensive view of 3B6 expression during embryonic development.

The expression of 3B6 seen in the adult brain in the real time PCR experiment led to the question whether this adult expression was limited to the brain or was more widespread. RT-PCR results indicated that 3B6 is expressed at low levels in all tissues examined. This result led to many questions: could 3B6 still have expression restricted to the nervous system in embryos? Could it be playing a generalized role in the maintenance of tissues? The possibility that expression was more restricted in the embryo existed, and could not be ruled out without the results from *in situ* hybridizations.

4.4.3 3B6 expression is not limited to the nervous system in embryos

Although the PCR-based results were informative, they still did not indicate a desired piece of information: whether embryonic expression of 3B6 was neurally restricted. The need for *in situ* hybridization was recognized and large investments of time were made attempting the method using paraffin embedded sections, which yielded no results. It appeared that this method would also need to be abandoned, particularly in light if other projects pursuing *in situ* screens of neurally expressed sequences. In a high throughput whole-mount *in situ* hybridization screen of putative transcription factors to identify those showing restricted expression patterns in the developing mouse brain, Gray and colleagues assayed 3B6[205]. Their results are available in the Mouse Genome Informatics (MGI; http://www.informatics.jax.org) gene expression database, and although they suggest that 3B6 was detected, they then go on to say that the expression level was ambiguous, and expression levels could not be determined with any certainty. In a similar experiment examining putative RNA binding proteins using cryosectioned tissues, McKee and colleagues also assayed 3B6[206]. Their results are also available from MGI, as well as in supplementary materials accompanying their paper. Their results with respect to 3B6 are described as either absent, or 'background or ubiquitous' expression.

The lab of Dr. Carol Schuurmans in the Genes and Development Research Group routinely do *in situ* hybridization using cryosectioned embryos. Paraffin embedded sections generally show better morphology than cryomatrix embedded ones, but due to

the increase amount of handling required for wax embedding, RNA is often lost, resulting in reduced or lost signal detection[207]. Ambion Inc. technical notes suggest that cryostat sectioned samples should be the first choice for use in *in situ* hybridization, and that paraffin sections should be avoided for hybridizations where sensitivity may be an issue. Given that sensitivity was believed to be the overriding problem in the northern blotting experiments, paraffin embedded sections were not likely the appropriate choice for the *in situ* hybridizations, however at the time our lab was not aware of this issue and the paraffin method had been successfully used many times previous with different transcripts[208-210]. Regardless, after years of attempts, *in situ* hybridization results were finally obtained after several months of collaboration with Ms. Natasha Klenin from Dr. Schuurmans' lab. The switch to cryosectioned embryos immediately made the difference in detection of 3B6.

The results, though predictable based on the adult tissue data, were somewhat different than what had initially been expected [Figure 16]. 3B6 expression is seen throughout the nervous system, but also throughout the rest of the embryo primarily in mesenchymal tissues. Interestingly, although expression of 3B6 is present in the adult heart and liver tissues, it is effectively excluded from those tissues in the embryo. Though the heart and liver originate from different germ layers, cross talk between them during development is essential[211]. Cardiogenic mesoderm is essential for liver induction and development, as it delivers inductive signals to the presumptive liver endoderm[212,213]. Interestingly, 3B6 expression was not found to be up regulated in spontaneously differentiating control 4-/3- cultures during the Bain ES cell neural differentiation protocol, and these control cultures are known to consist of up to 30% cardiomyogenic precursors[130].

Two other genes encoding nucleic acid binding proteins, *Foxd3* and *Hes5*, have also been shown to be expressed in the nervous system and mesenchymal tissues of the mouse embryo, while being excluded from the heart and liver[205,214,215]. Foxd3 is a winged helix transcription factor that plays central roles in the differentiation and migration of neural crest cells[216]. Hes5, an effector of Notch signalling, is an inhibitory

bHLH transcription factor that is central to maintenance of neural precursors[66]. A common thread between these genes, apart from their ability to bind nucleic acids and their expression, is that they are both repressors of neural cell fates[214,217]. It is interesting to speculate that 3B6 may also be a negative regulator of neural fates, controlling rate of neural proliferation and promoting other cell fates, but again, too little is known about 3B6 to effectively propose a putative developmental role.

Chapter Five: Expression and analysis of the 3B6 Product, mSe70-2

5.1 Summary

Following up the findings from chapter 4, including the gene structure and expression patterns of 3B6, this chapter describes preliminary analysis of the 3B6 gene product, mSe70-2. mSe70-2 is the mouse homolog of human cutaneous T-cell lymphoma (CTCL) tumor antigen Se70-2. Polyclonal antibodies against mSe70-2 are raised, and expression of the protein in both bacterial and mammalian cell systems is pursued. Challenges encountered in this work are described and discussed. Results from this chapter may suggest that mSe70-2 production is highly regulated, and a mechanism for this regulation is proposed.

5.2 Introduction

5.2.1 Cutaneous T-cell lymphoma tumor antigen Se70-2

Cutaneous T-cell lymphomas (CTCLs) are heterogeneous skin neoplasms originating from T-lymphocytes[218]. Malignant T-lymphocytes localize to the epidermis where thay produce Cutaneous plaques, patches, and tumors[219]. CTCLs cannot currently be cured, and in advanced disese even the most aggressive therapies available are ineffective.

Immunological based methods of treatment have been successfully used in other skin neoplasms such as melanoma[220]. With the success of such treatments, therapeutic methods such as peptide vaccination have been an attractive concept for treatment of CTCLs. Before these methods can be developed though, therapeutic targets in the form of tumor antigens must be identified.

The SEREX method, serological identification of recombinantly expressed genes, has been proven to be a powerful method used to screen for new immunologically relevant tumor antigens. In this approach, a bacteriophage λ cDNA expression library is constructed from tumor tissues, or in a variation on the method, testis tissue [Figure 17][221]. Testes tissue may be used in order to identify cancer-testis (CT) antigens. In light of the immunoprivileged nature of the testes, many CT antigens may be considered immunologically specific to tumors[222]. The constructed library is used for lytic infection of *E. coli*, which expresses the recombinant proteins. These proteins are transferred to membranes and detected using patient serum. Clones that elicit an antibody response, and therefore are detected on the membranes, are isolated and sequenced in order to identify the antigen[221].

The SEREX method was used by Stefan Eichmüller and colleagues to identify CTCL associated antigens[218]. The library they used was constructed from testis tissue, but was not subtracted with other normal tissues, and therefore was not testis specific. In all they identified 15 CTCL associated antigens, one of which is Se70-2. Se70-2 is not testis specific, as demonstrated using RT-PCR, but rather is expressed broadly in adult tissues. It was also found to be expressed at the mRNA level in 100% of CTCL and

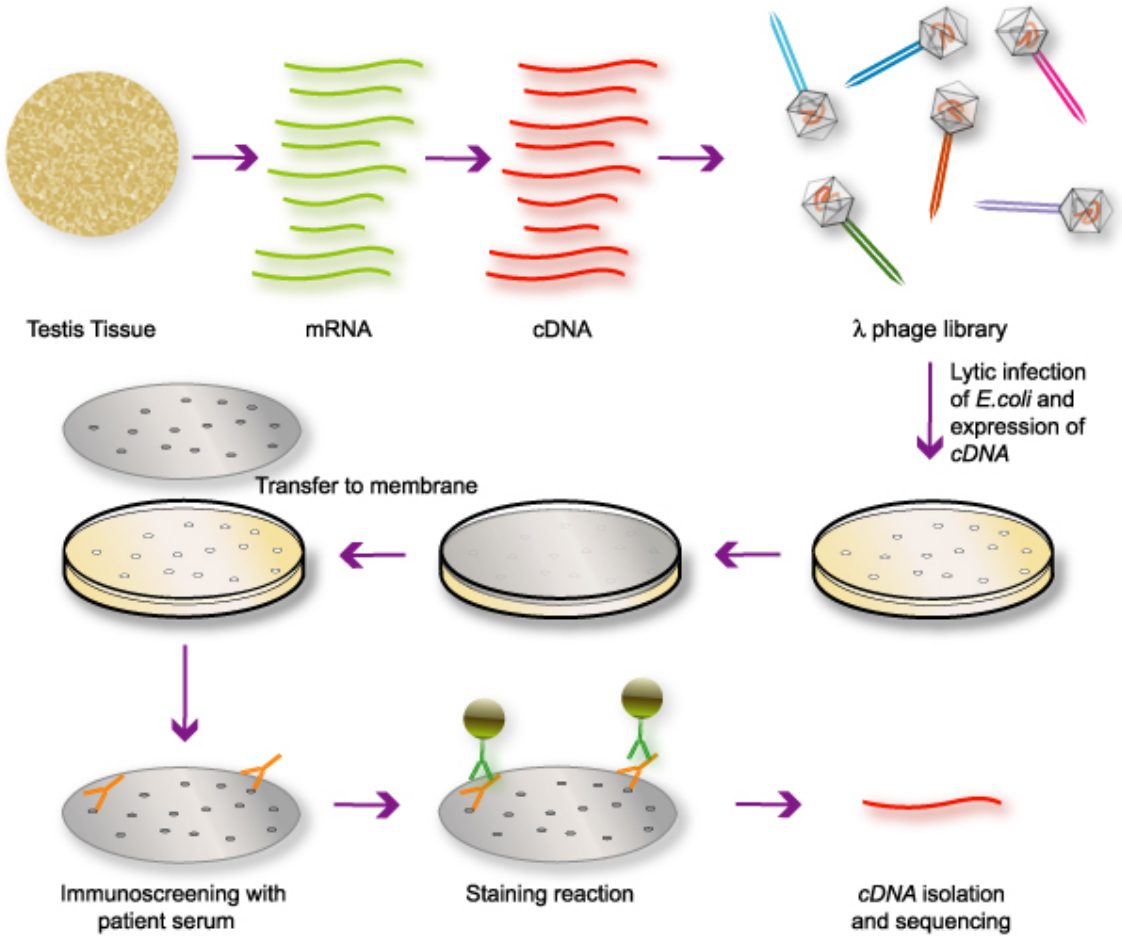


Figure 17: Identification of tumor antigens using the SEREX method.

leukemia cell lines tested. Analysis of the protein expression has not been carried out. Additionally, Se70-2 was not found to have any significant similarity to any GenBank published sequences.

As is the case with the majority of SEREX identified antigens, it is not known why Se-70-2 elicits an immunogenic response, particularly when its expression in a host of tissues is considered. Several possibilities exist. First, the expression of Se70-2 has been demonstrated only on the mRNA level, therefore the immune system may not have had previous exposure to the protein, and therefore does not recognize it as a 'self' protein. A variation on this point is that anomalous expression of the protein, be it due to mutation of the gene, amplification, expression of abnormal splice variants, or otherwise, leads to immunogenic response[223]. Finally, Se70-2 may be a cancer related autoantigen which is expressed at similar levels in neoplastic and normal tissues, but raise an immunogenic response in cancer patients for unknown reasons[221].

5.3 Results

5.3.1 3B6 encodes the mouse Se70-2

At the time of its selection for follow up, 3B6 did not show any significant similarities to any sequences in the public databases. With the continual update of information regarding 3B6 by RED however, similarities eventually became available. It was found using BLASTx that the translation of the cDNA sequence for Se70-2 submitted by S. Eichmüller and colleagues (accession: AF273052) shared 94% identity and 97% similarity with the 305 C-terminal amino acids of the predicted 3B6 protein product. The reason that the alignment covers only the C-terminal end of the protein is that the sequence submitted for Se70-2 was incomplete, corresponding to exons 15-22 of the predicted 3B6 transcripts. This partial sequence however was attributed to the human locus C13orf10 (current Entrez gene ID: 64062), and the product of this locus was annotated as Se70-2. Comparing the complete Se70-2 sequence (translation of the coding sequence of accession BC041655) with the predicted products of each 3B6 transcript using BLASTp, it was found that transcript 1a had the most remarkable

similarity to Se70-2: 94% identity and 96% similarity over the entire 1007 aa length of the 3B6 product. From this level of sequence similarity it could be confidently concluded that 3B6 encodes the mouse homolog of cutaneous T-cell lymphoma antigen Se70-2. Indeed, annotation of the mouse 3B6 gene locus, at that time referred to in the public data as the RIKEN cDNA 1700009P03 gene (current Entrez gene ID: 74216), soon also referred to the product as Se70-2. For the continued duration of this document, the product encoded by the 3B6 gene will be referred to as mSe70-2.

5.3.2 Generation of antibodies for mSe70-2

For continued characterization of 3B6 and its product mSe70-2, polyclonal antibodies were generated using two peptides judged to be on the surface of the protein. The peptides chosen for antibody production were identical between the human and mouse versions of Se70-2, so that the resulting antibodies could be used in both systems if desired. Also, both peptides avoided regions of the proteins that were expected to contain conserved functional domain in order to minimize the potential for cross-reactivity with other proteins containing the same domains. Peptide 3B6-2 corresponded to amino acids 722-731 of mSe70-2, with the sequence EAQKKKQEAL, and peptide 3B6-3 corresponded to amino acids 849-858 with the sequence GILSSGRGRG. Both peptides were received from the Alberta Peptide Institute as both KLH and BSA conjugates. Injections of the KLH conjugates in Freund's adjuvant were administered to rabbits for antibody production.

Sera from the rabbits who received repeated administrations of the peptides were tested for antibody response using western blotting. Samples of the peptides conjugated to BSA were subjected to SDS-PAGE, and detected on membranes using 1:1000 dilution of serum. Rabbits raised reactions to the peptides, as demonstrated by detection of the BSA conjugated peptides by serum from rabbits who received injections of the corresponding KLH conjugated peptide [Figure 18, A, B]. The antibodies were then affinity purified using the BSA conjugated peptides. The process of purification of the antibodies did not lead to loss of their ability to detect the peptides they were raised with [Figure 18, C].

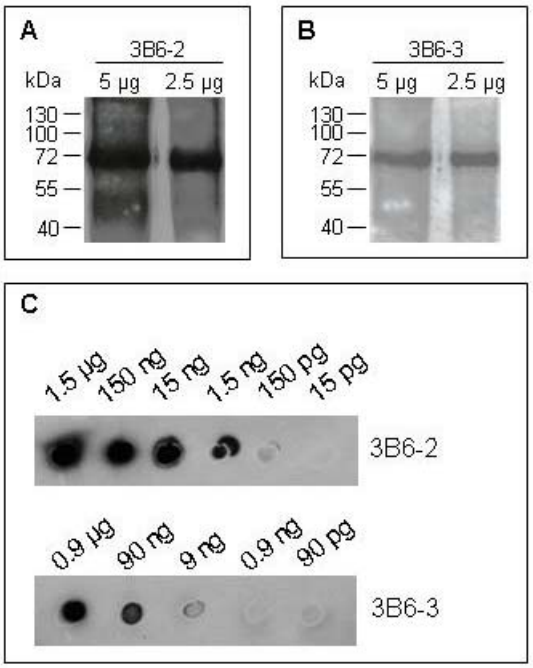


Figure 18: Polyclonal antibodies raised against mSe70-2 detect the peptides they were raised with. Panel A and B: 5 µg and 2.5 µg of each BSA conjugated peptide was subjected to SDS-PAGE and western blotting using a 1:1000 dilution of serum from rabbits injected with the corresponding KLH conjugated peptides. Panel C: Dilution series of the BSA conjugated peptides were spotted onto membranes for detection using purified a 1:1000 dilution of the purified antibodies.

As it was shown that the mSe70-2 antibodies do recognize the peptides they were raised with, experiments were then carried out to detect the native mSe70-2 protein in cell lysates. Three cell lines, ES, P19, and HEK, had previously been shown to express 3B6 transcripts via RT-PCR (data not shown). P19 and HEK cells had been introduced into the project for different reasons: P19 cells are more easily maintained and transfected than ES cells, but like ES cells can be differentiated into neural precursor enriched cultures using the Bain RA induced differentiation method[117,123,224]. They were intended to play roles in determining the function of mSe70-2 in neural differentiation studies. HEK cells were introduced as easily maintained and manipulated general use cell line for studies of mSe70-2. Cell lysates from each of these cell lines was prepared, and western blots were generated also including samples of the BSA conjugated peptides and cell lysates from the Cos-7 cell line, which were available in the lab. Blots were detected using purified mSe70-2 antibodies at dilutions ranging from 1:500 to 1:1000. This experiment was repeated several times using different amounts of protein for the blots. A typical result is shown in Figure 19. No detection of native mSe70-2, with an expected size of approximately 110-112 kDa, was ever confirmed, despite detection of the BSA conjugated peptides. A number of bands of varying sizes were often detected, but may be due to cross-reactivity with other epitopes present in the proteome. Similar results were obtained using either of the polyclonal antibodies. It could not be distinguished if mSe70-2 was not detected because it is not present, or if it was not detected because the antibodies are ineffective. Due to lack of a positive control for a complete mSe70-2 protein, further work using the antibodies was suspended until appropriate controls could be established.

5.3.3 Expression of mSe70-2 in the E. coli system

In continued effort to characterize the mSe70-2 antibodies, and also with the goal of characterization of the protein itself, efforts to express mSe70-2 in the *E. coli* system were undertaken. The coding sequence that was chosen for expression was that of 3B6 transcript 1a, which will be referred to as the full length 3B6 coding sequence [Appendix D]. This transcript appeared to be most predominant in mouse at 12.5 dpc, as it was by

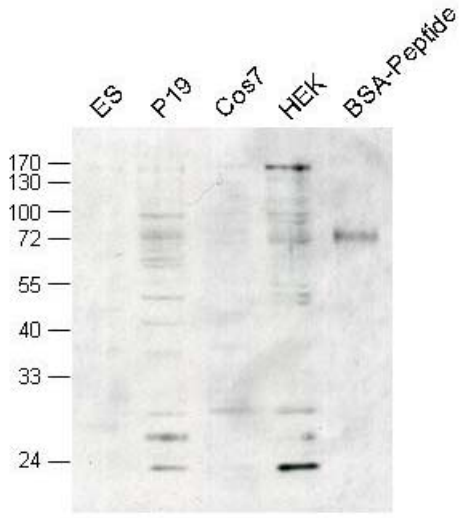


Figure 19: Polyclonal antibodies raised against mSe70-2 may not detect the native protein. 100 µg of total protein from various cell lysates plus 1 µg of the BSA conjugated 3B6-3 peptide was subjected to SDS-PAGE and western blotting with the 3B6-3 antibody at a dilution of 1:1000. Expected size of mSe70-2 is 112 kDa. Detection of mSe70-2 could not be confirmed given the number of alternatively sized bands present, and the lack of a positive control for the protein.

far the most commonly obtained sequence during the process of validation of the 3B6 gene structure, and also its predicted product was very similar to its human Se70-2 homolog.

At this point in the project, the complete coding sequence for 3B6 had not been cloned, as difficulties had been encountered with the TA cloning of the largest PCR products produced during the validation of the predicted gene structure (section 4.3.1). Cloning of the full length sequence was completed by assembly of two overlapping sequences into a pBS-based TA cloning vector. In the resulting construct, pBSXcm-3B6, the coding sequence for 3B6 was oriented opposite to the *LacZ′* gene in the vector. The vector pGEX-5x-1 was available in the lab and was hence chosen for expression of mSe70-2 as a fusion to the C-terminal end of GST, under control of the *tac* promoter. Cloning into this expression vector, and as it was rapidly determined, any other expression vector, was extremely difficult. In almost all cases no colonies containing desired constructs were returned, regardless of strategy, reagents, or bacterial strains used. Many months were spent believing that the cloning difficulties were due to technique or reagents, given the straight-forward nature of the development of pBSXcm-3B6.

Several observations were made during the cloning of several different constructs over this course of time: first, it was noted that when trying to clone 3B6 into the *Not I-Sal I* site of pBS, that the orientation of 3B6 with respect to *LacZ′* impacted the cloning results [Figure 20]. 3B6 was very simple to clone into pBS provided that it was in the opposite orientation to that of the *LacZ′* selectable marker. If attempts were made to clone 3B6 in the same orientation as *LacZ′*, then no clones were ever obtained. Second, during a particular attempt to clone 3B6 into pGEX-5x-1, many clones were yielded which appeared to be the desired construct when screened by restriction digestion. When sequenced though, it became apparent that 3B6 was not in frame with the GST coding sequence. These observations among others suggested that leakage of expression of the mSe70-2 protein has a detrimental effect on the survival of bacterial cells, and so steps were taken to prevent such leakage from occurring. Expression from pGEX-5x-1 is

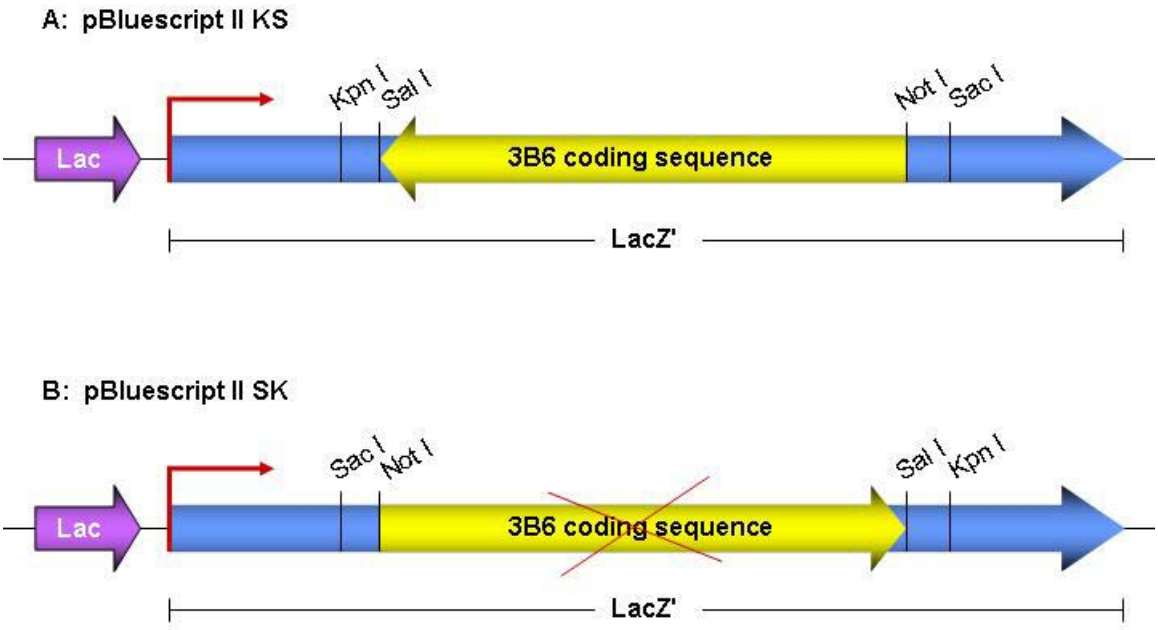


Figure 20: Orientation and frame of the 3B6 coding sequence influence its cloning. Challenges with cloning of the 3B6 coding sequence were first encountered when attempting to insert it into the Not I-Sal I site of pBluescript (pBS). *E. coli* colonies containing the recombinant construct were easily obtained when pBS II KS (A) was used, but were never obtained using the same cloning strategy with pBS II SK (B). The only difference between the two constructs was the orientation of the multiple cloning site, and hence the orientation of the 3B6 insert. The yellow and blue arrows indicate the orientation of the 3B6 and LacZ' reading frames. Red arrows indicate direction of transcription using the lac promoter, represented as a purple arrow. Similar effects were seen when cloning into pGEX-5x-1: if the coding sequence was out of frame with GST, the recombinant construct was easily propagated in *E. coli*.

under control of the *tac* promoter, a hybrid of the *trp* and *lac* promoters, and is known to be prone to leakage even in the presence of the *lac* repressor. The *lac* promoter is part of the *lac* operon, involved in lactose metabolism in *E. coli*. Lactose metabolism is suppressed in the presence of the preferred *E. coli* energy source, glucose.

In order to diminish any leaky expression of mSe70-2 media for recovery and growth of bacterial cells was supplemented with glucose. With continued efforts including new reagents, close monitoring of each step of the cloning process, and growth of bacteria at different temperatures, a single colony containing the desired pGEX-3B6 construct was eventually obtained. For the sake of confidence in the construct, it was completely sequenced six times and found to be correct. Efforts to build another construct for mSe70-2 expression in the prokaryotic system, pET-3B6, were also made. As was seen initially in the pGEX-3B6 construction process, colonies containing the desired construct were not obtained. Attempts to produce this construct were eventually discontinued.

The challenges presented by cloning 3B6 did not bode well for expression of mSe70-2. Sequence verified pGEX-3B6 plasmid DNA was electroporated into several expression bacterial strains: Rosetta (DE3), BL21 (DE3), and HMS174 (DE3) pLysS. It was hoped that one of these strains would appropriately support expression of mSe70-2. Cultures were induced to produce mSe70-2 by the addition of IPTG to the glucose supplemented medium. Affinity purification of GST fusion proteins was carried out, reserving samples from each step in the process for western blot analysis using anti-GST antibodies [Figure 21]. Though it was clear that IPTG induction of the cultures did lead to expression of GST fusion proteins, the desired GST-mSe70-2 fusion, with an expected size of approximately 140 kDa, was not present. It appeared that the products being produced were degraded by the bacteria. Use of protease inhibitors when lysing the cells or direct lysis of the cells in protein loading buffer (PLB) failed to provide any indication of a 140 kDa product produced at all. The observed protein bands produced by each bacterial strain were similar, except that in the HMS174 (DE3) pLysS strain some of the

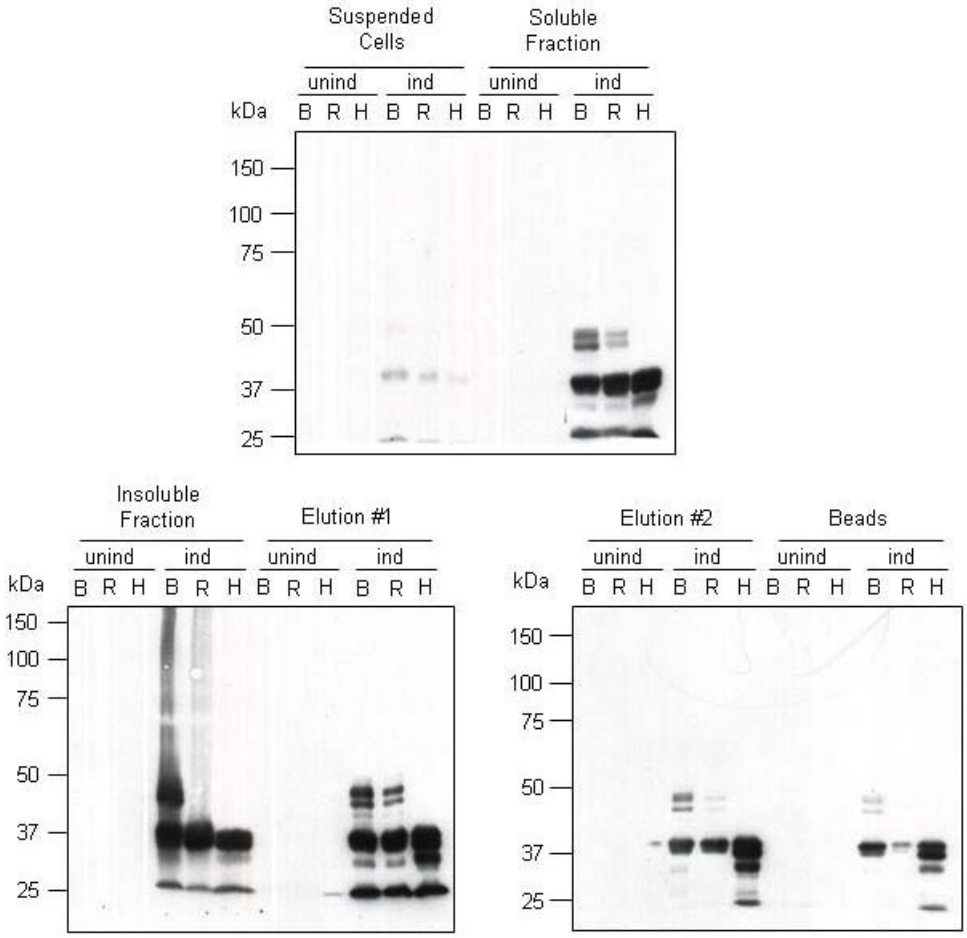


Figure 21: Induction of pGEX-3B6 in different *E. coli* strains does not improve mSe70-2 expression. Three *E. coli* strains, BL21 (DE3) [B], Rosetta (DE3) [R], and HMS 174 (DE3) pLysS [H], were transformed with pGEX-3B6 and grown in glucose supplemented medium. Cultures were induced to express mSe70-2 using 0.1 mM IPTG (ind), or left uninduced (unind) for 4 hours. Cells were collected, and lysed, and GST-fusion proteins were purified with Glutathione Sepharose 4b. Western blotting was carried out with anti-GST antibodies. Expected GST-mSe70-2 size is 140 kDa, and GST alone is 29 kDa.

upper bands seen in the Lon and OmpT protease deficient strains Rosetta (DE3) and BL21 (DE3) were not observed.

This observation suggested that GST-mSe70-2 may be degraded in a prescribed manner. To try and begin to understand if this was the case, several sub-fragments of the 3B6 coding sequence were cloned into pGEX-5x-1, with similar difficulties to those previously encountered. The cloned segments covered exons 1-7, 1-13, or 11-22. It was thought that expression of GST fusion proteins from these constructs may yield degradation products of varying sizes, allowing for mapping of areas of the peptide sequence, indicating probable sites of protein cleavage. Methods to prevent these cleavage events may then be better explored, or dual -tagged expression constructs could be produced to further characterize the cleavage events. Rosetta (DE3) cells were transformed with each of the pGEX constructs and expression of the GST fusion proteins was induced with IPTG at varying incubation temperatures. Cells were collected directly into PLB for western blotting with anti-GST antibodies. Perplexing results were yielded, with similar banding patterns being detected from each construct [Figure 22]. Although some similar bands could be anticipated from pGEX-3B6-1-7 and pGEX-3B6-1-13, the same could not be said for pGEX-3B6-11-22. This result was suggestive of problems beyond proteolytic processing or degradation, such as rearrangement of the plasmid constructs.

Because GST-mSe70-2 proteins were primarily intended for use as positive controls for characterization of the mSe70-2 antibodies, attempts to troubleshoot this system were discontinued in favour of expression in the mammalian system. The mammalian system was deemed as having more potential for the overall characterization of mSe70-2 in addition to the antibodies.

5.3.4 Expression of mSe70-2 in mammalian cell culture systems

With the goal of expressing mSe70-2 as a tagged protein in mammalian cells, development of a number of new expression constructs was undertaken. In total, attempts to build four constructs were made, but only two were eventually obtained: pCMV-3B6-FLAG, for expression of mSe70-2 with a C-terminal FLAG tag, and pEGFP-

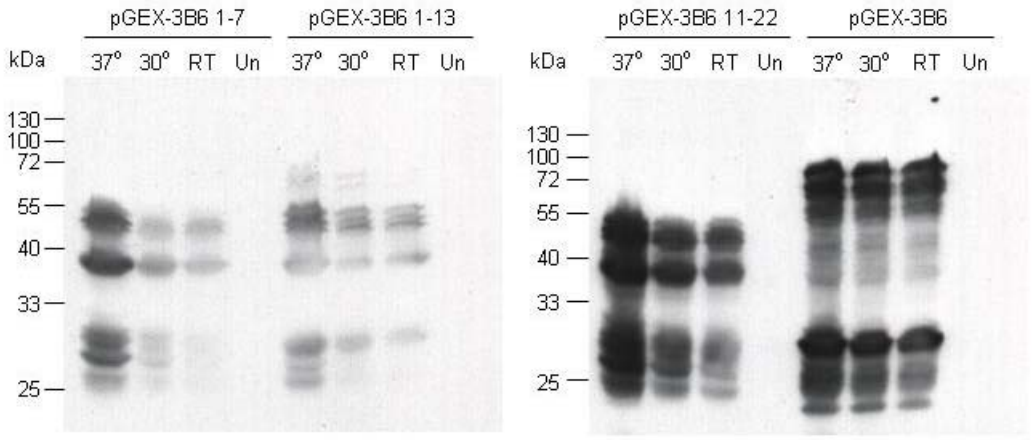


Figure 22: Induction of pGEX-3B6 constructs at different temperatures does not improve expression of mSe70-2. *E. coli* strain Rosetta (DE3) was transformed with pGEX-3B6 constructs containing exons 1-7, 1-13, 11-22, or the full coding sequence of 3B6. Liquid cultures for induction were grown in glucose supplemented medium, and induced with 0.1 mM IPTG at 37°C, 30°C, or room temperature (RT). Uninduced (Un) cultures were grown at 37°C. Cells were collected in PLB and subjected to western blotting with anti-GST antibody. Expected GST-mSe70-2 sizes: exons1-7, 67 kDa; exons 11-13, 99 kDa; exons 11-22, 83 kDa; full length 3B6, 140 kDa. GST alone is 29 kDa.

3B6, for expression of mSe70-2 as a fusion with the C-terminus of GFP. As with the pGEX-based constructs, each of these constructs were extensively sequenced to ensure accuracy. Cell lines chosen for the transfection studies were P19 and HEK, as described earlier, and transfection conditions were optimized using CMV promoter driven GFP expression constructs and FACS for estimation of transfection efficiencies.

For expression of GFP-mSe70-2, cells were transfected with pEGFP-3B6 or empty vector and grown for 24-72 hours. Cells were then collected for lysates and subjected to western blotting using anti-GFP antibodies. A typical result from transfection of HEK cells is shown in Figure 23. Although GFP could be detected, no evidence for expression of a 140 kDa GFP-mSe70-2 fusion protein was present. Similar results were also obtained using P19 cells (data not shown). This construct was quickly abandoned in lieu of work with the pCMV-3B6-FLAG construct.

For expression of FLAG -tagged mSe70-2, cells were transfected with pCMV-3B6-FLAG or empty pCMV-tag4a vector. pCMV-FLAG-hTERT, a gift from the lab of Dr. Tara Beattie was also transfected as a control to ensure expression of a large FLAG -tagged protein could be accomplished. Cells were harvested for cell lysates after 24-72 hours, and western blotting with anti-FLAG antibodies was carried out. In each experiment mSe70-2 was not detected [Figure 24, A]. FLAG-hTERT, was consistently detected indicating its successful production in transfected cells. These results raised the question as to whether the pCMV-3B6-FLAG construct was even being transcribed. To examine this possibility, HEK cells were transfected with pCMV-3B6-FLAG and collected for RNA after 24-72 hours. RT-PCR using a primer lying in exon 18, 3B6-18F, and one lying in the FLAG tag, FLAG-R, indicated that transcription was indeed occurring from the transfected construct [Figure 24, B].

Confirmation that the construct was transcribed, but yet there was no apparent production of the mSe70-2-FLAG protein called into consideration the possibility that Se70-2 was subject to very rapid degradation by the proteasome. To address this issue, HEK cells were transfected with pCMV-3B6-FLAG or left untransfected. After 48 hours, cells were treated with varying amounts of proteasomal inhibitor MG132 for up to

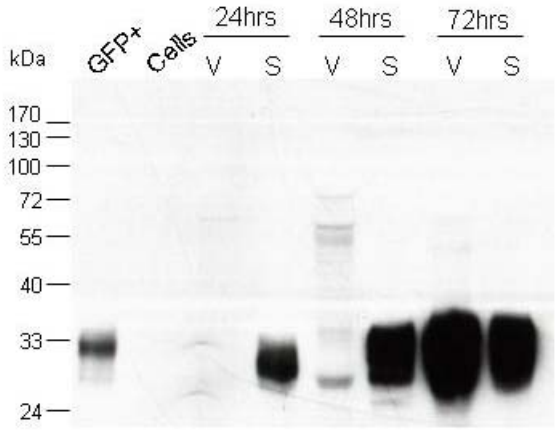


Figure 23: GFP-mSe70-2 is not expressed from pEGFP-3B6 in HEK cells. HEK cells were transfected with pEGFP-3B6 (S) or empty pEGFP-C1 vector (V). After 24, 48, or 72 hours, cells were collected in PLB for use in western blotting with anti-GFP antibodies. GFP+, purified GFP; Cells, untransfected HEK cells. The expected size of the GFP-mSe70-2 fusion protein is 140 kDa. GFP itself is 27 kDa.

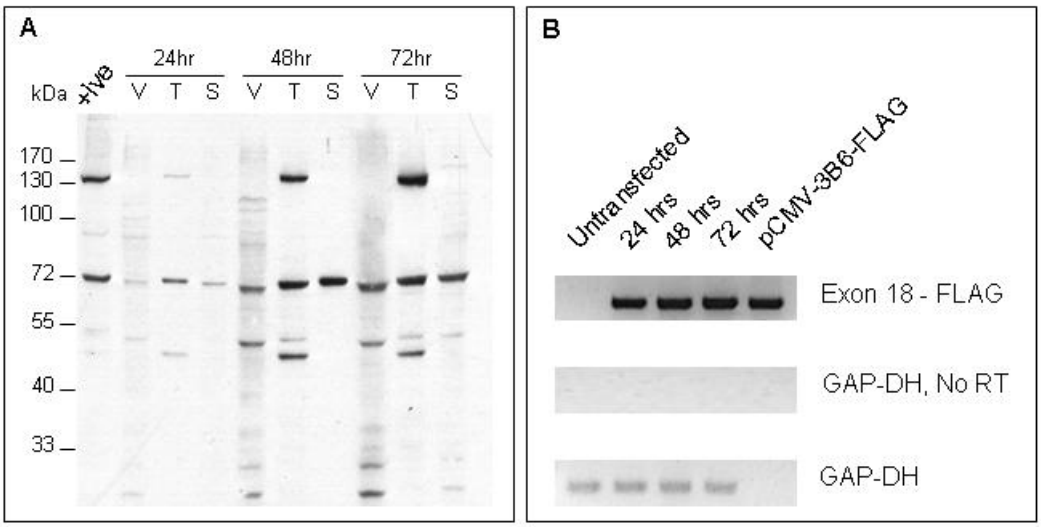


Figure 24: mSe70-2 is expressed on the mRNA level but not on the protein level from pCMV-3B6-FLAG. HEK cells were transfected with empty vector pCMV-Tag4a (V), pCMV-FLAG-hTERT (T), or pCMV-3B6-FLAG (S). At 24, 48, and 72 hours, cells were collected for lysates or cDNA preparation. Panel A: western blotting carried out using anti-FLAG antibody. Expected product sizes are 140 kDa and 112 kDa for FLAG-hTERT and mSe70-2 respectively. +ive, FLAG-hTERT positive lysate. Panel B: RT-PCR using primers 3B6-18F and FLAG-R. Untransfected, untransfected HEK cell control; pCMV-3B6-FLAG, vector PCR positive control.

6 hours, after which they were collected for cell lysates. Western blotting was carried out using anti-FLAG antibodies and anti-ubiquitin antibodies to ensure the effectiveness of the MG132 treatment. Results showed no apparent stabilization of mSe70-2 with proteasomal inhibition [Figure 25].

To examine whether mSe70-2 was being expressed but at low levels, immunoprecipitation of cell lysates from pCMV-3B6-FLAG transfected HEK cells with anti-FLAG antibodies was carried out. A large number of bands were immunoprecipitated, including a prominent one detected above 100 kDa [Figure 26]. This prominent band could not be confidently identified as mSe70-2-FLAG due to the apparent cross reaction of the FLAG antibodies with many proteins in the cell lysates. In an effort to clarify this result, a similar immunoprecipitation was carried out using the antibodies raised against mSe70-2 [Figure 27]. No protein of the size expected for mSe70-2-FLAG was immunoprecipitated by either antibody. Again it is not clear if this is attributable to a lack of presence of the protein or to ineffectiveness of the antibodies.

5.4 Discussion

5.4.1 Antibody development: are the mSe70-2 antibodies effective?

Antibodies are central to the characterization of any protein, as they facilitate a wide array of methods for elucidating expression and function. Two antibodies were raised using peptides from mSe70-2 with these methods in minds. The overriding problem with the antibodies produced is that it has so far been indeterminable whether or not they can effectively detect mSe70-2. The antibodies work in principle, as they both detect the peptides they were raised with. What has not been shown is that they can detect the entire mSe70-2 protein. In theory, if the antibodies can detect the peptides they were raised against, they should be able to detect the full protein especially if it is on a denaturing western blot. This was not observed with the generated antibodies, which may be attributable to several causes: first, it may be that the protein is normally expressed at low levels, or rapidly turned over in the cell, and therefore cannot be detected. Second, the protein may not be expressed at the time and place examined.

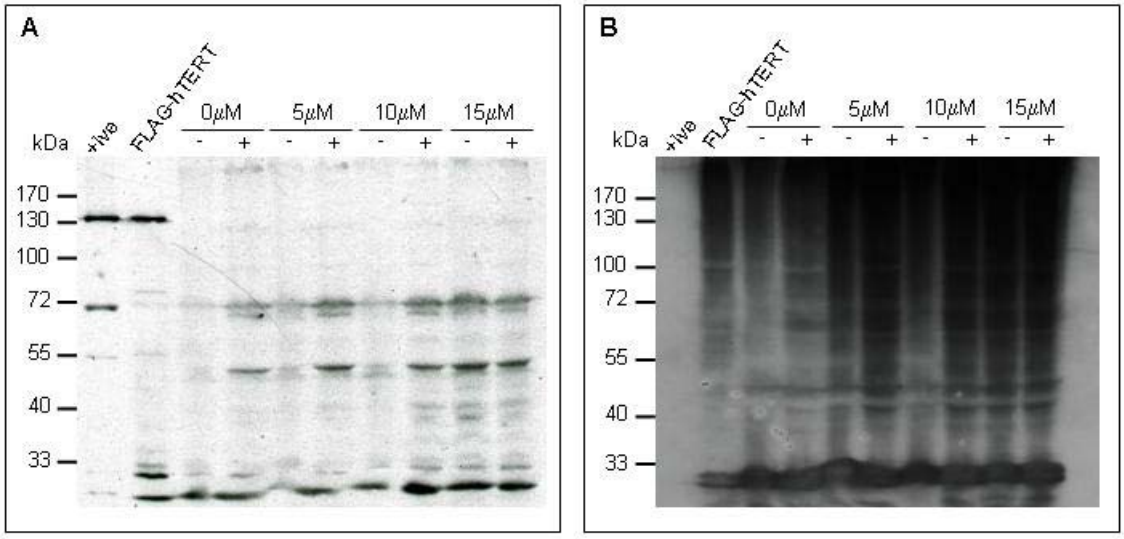


Figure 25: Proteasomal inhibition does not stabilize mSe70-2. HEK cells were transfected with pCMV-3B6-FLAG. 48 hours post-transfection, cells lest untreated or were treated with 0, 5, 10, or 15 μM proteasome inhibitor MG132 for 2, 4, and 6 hours, and then collected for lysates. Western blotting was performed using anti-FLAG (Panel A) and anti-ubiquitin (Panel B) antibodies. Blots shown are for the 6 hour time point. Expected sizes for FLAG-hTERT and mSe70-2-FLAG are 140 kDa and 112 kDa respectively. +ive, FLAG-hTERT positive lysate; pCMV-FLAG-hTERT, transfection control; -, untransfected cells; +, pCMV-3B6 transfected cells.

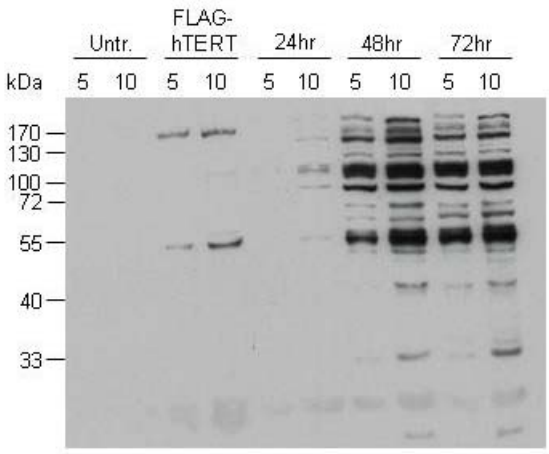


Figure 26: FLAG IP from pCMV-3B6-FLAG transfected HEK cells. HEK cells were transfected with pCMV-3B6-FLAG and collected for cell lysates at 24, 48, and 72 hours. Lysates were immunoprecipitated with anti-FLAG-M2 antibody conjugated beads. Proteins were eluted from beads by boiling in 20 µL PLB, 5 and 10 µL of which was loaded into neighbouring wells for PAGE. Western blotting was done with anti-FLAG polyclonal antibody. Expected sizes of FLAG-hTERT and mSe70-2-FLAG are 140 and 112 kDa respectively. Untr., untranscfected HEK cells; FLAG-hTERT, pCMV-FLAG-hTERT transfected cells, 24 hour time point.

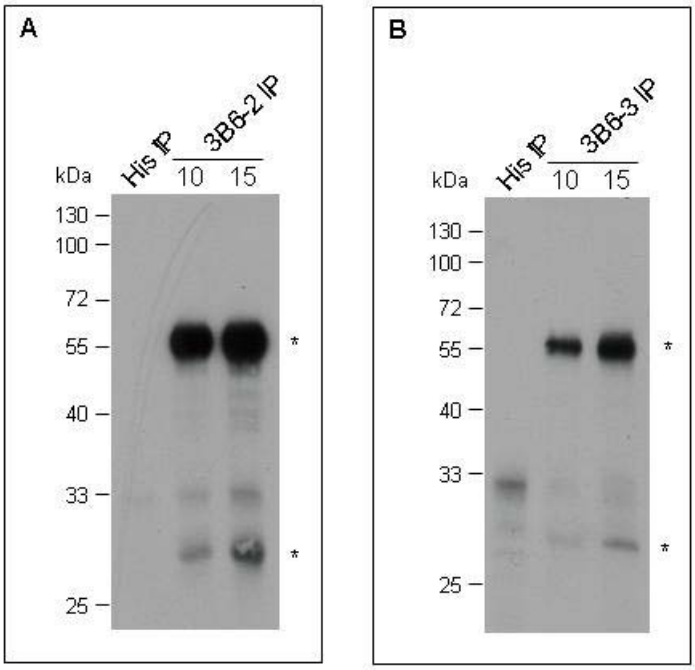


Figure 27: mSe70-2 IP from pCMV-3B6-FLAG transfected HEK cells. HEK cells were transfected with pCMV-3B6-FLAG and collected for cell lysates at 48 hours. Lysates were immunoprecipitated with antibodies raised against mSe70-2, 3B6-2 (Panel A) and 3B6-3 (Panel B), plus anti-His mouse monoclonal antibody as a control. Proteins were eluted from beads by boiling in 30 µL PLB, 10 and 15 µL of which was loaded into neighbouring wells for PAGE. Western blotting was done with the same mSe70-2 antibody as was used for the IP. Expected size of mSe70-2-FLAG is 112 kDa. * indicates IgG light and heavy chains.

Finally, the problem may be the antibodies themselves not being able to detect the proteins which are actually present.

To address the first issue, attempts were made to over-express mSe70-2 in a variety of systems, none of which were successful. Inhibition of the proteasome to try to stabilize protein production did not yield any answers. This is not particularly surprising as overexpression of proteins that are rapidly degraded often overcomes the degradation machinery, and hence they are readily detectable without proteasomal inhibition. In order to circumvent the possibility that mSe70-2 was being expressed at low levels immunoprecipitations were completed, again without demonstrating expression of mSe70-2. This result also is inconclusive though, as it may be that the antibodies are inappropriate for IPs, as not all antibodies are effective in every application.

The second issue, regarding the spatial and temporal expression of the protein could be the critical one. Only cell lines were examined for the presence of mSe70-2, but it may be that it is tightly regulated at the protein level, and therefore only produced under very specific circumstances. Given the isolation of the human Se70-2 as a tumor antigen, appropriate places to look for the protein may be the immunoprivileged brain or testes, the developing embryo, and CTCL cell lines. The final possibility is that the antibodies are not effective against mSe70-2. To conclusively prove this point, positive controls are needed, which currently are unavailable.

Overall, it cannot be concluded whether the mSe70-2 antibodies are effective or not. Without purified mSe70-2 protein as a positive control, it is not possible to conclusively differentiate between the case that the antibodies do not work, and the case that they do, but there is no protein present to be detected. This is a major problem in this body of work, which despite efforts has not been resolved.

5.4.2 Is mSe70-2 detrimental to E. coli?

Cloning of the 3B6 coding sequence was highly challenging, and there is evidence to suggest that mSe70-2 is detrimental to E. coli. Any cloning of the coding sequence that could not result in any leakage of mSe70-2, in any manner, was straightforward, but cloning the coding sequence in the correct orientation in the presence

of bacterial promoters was always difficult. The generation of the pGEX-3B6 construct was particularly challenging, as pGEX vectors contain the *tac* promoter which due to the contribution of the *lac* operator is known to be inherently leaky. Using glucose supplemented medium to suppress the *lac* operon appeared to be helpful, as the pGEX-3B6 construct was obtained, and can be propagated under these conditions. A contributing factor that resulted in finally obtaining this construct may have been the brute force used. It may have been the individual characteristics of a particular bacterium in combination with its local environment that allowed it to survive while propagating pGEX-3B6 (luck). Ultimately, pGEX-3B6 was established as a result of determined and persistent effort

Expression of mSe70-2 in the prokaryotic system was never achieved. There may be selective pressure on *E. coli*, to destroy mSe70-2 by proteolysis as it is made, and/or prevent its production because the protein has deleterious effects on the cell. The differences in product sizes seen between the *Lon* and *ompT* protease deficient bacterial strains Rosetta (DE3) and BL21 (DE3), and those seen in the non-protease deficient strain HMS174 (DE3) pLysS suggest that proteolysis of a product induced from pGEX-3B6 is occurring. However, even with use of protease inhibitors or direct preparation of cell lysates in PLB, no evidence for the full GST-mSe70-2 product was ever observed. Another possibility to be considered is that the expression bacterial strains were able to alter the construct after electroporation, and the cells carrying altered versions were selected for in culture. The likelihood of this occurrence is not great though, as although Rosetta (DE3) and BL21 (DE3) are both recombination competent, HMS174 (DE3) pLysS is not, and similar products were seen in each strain. Lastly, the expected molecular weight of mSe70-2 is approximately 112 kDa. With the addition of GST, it would be approximately 140 kDa, which is very large for a bacterial expression system to produce abundant product. Though the mechanism at use is not clear, what is evident is that mSe70-2 is not well tolerated by *E. coli*.

A possible solution to this problem would have been to obtain a construct based on a pET vector containing the 3B6 coding sequence. Construction of such a

recombinant plasmid was underway, but never achieved as prokaryotic expression efforts were abandoned. In retrospect, the cell line HMS174 (DE3) pLysS would have been an ideal host for both maintenance and expression of pET-3B6 as it allows tight regulation of proteins expressed from a T7 promoter, as in pET, and it is recombination deficient. The only drawback may have been its proteolytic competency. Overall though, a pET-based construct in an appropriate cell line is likely to have been the best chance for expression of mSe70-2 in prokaryotic cells.

5.4.3 Is mSe70-2 regulated at the translational level?

The broad expression of 3B6 transcripts combined with the inability to detect mSe70-2 even when expressed as a -tagged construct raises the possibility that the protein is not normally expressed by the majority of cells, but instead highly regulated at the translational level. Several lines of evidence are suggestive of this possibility: first, the pCMV-3B6-FLAG construct has been sequenced many times, and is known to be correct. It has been shown to be transcribed in HEK cells, and the ability to produce and detect a large FLAG -tagged protein (FLAG-hTERT) has been demonstrated. However, there is still no mSe70-2-FLAG detected. The possibility exists that FLAG gets cleaved from mSe70-2, although there is no evidence to support this claim, neither on the western blots nor with *in silico* analysis of the mSe70-2 sequence.

Another line of evidence suggesting that mSe70-2 may be regulated at the translational level is that it has been found to be expressed at the mRNA level in a wide array of normal tissues, but the immune system can raise reactions to it under certain circumstances, suggesting that it is not tolerized to the protein. This immune reaction could arise based on production of aberrant forms of the protein, or in an unknown autoimmune mechanism, but it also could arise on the basis that the immune system has never had exposure to the protein as it has not previously been expressed in non-immunoprivileged sites. Dr. Stefan Eichmüller of the group who identified human Se70-2 as a tumor antigen, has said that in his experience many tumor antigens are found expressed on the mRNA level in a variety of normal tissues but not on the protein level.

Also, his group has also had difficulties expressing recombinant Se70-2 (S. Eichmüller, pers. comm.).

The possibility that 3B6 encodes non-coding RNAs (ncRNAs) has been considered, but it is not believed to be the case. In a survey of mammalian transcriptomes, it was estimated that up to one third of transcripts produced are likely to be ncRNA[198]. ncRNAs are generally conserved and are found in sizes ranging from micro RNAs (miRNAs) of 20-25 nucleotides to large transcripts of over 100 kb[225]. Functionally, ncRNAs are expected to be meaningful, but the range of functions is currently not known. Almost all ncRNAs have been found to not contain any clear open reading frames[226]. 3B6 transcripts all have large open reading frames, and identification of se70-2 as an antigen demonstrates that is is at least under some conditions expressed, supporting the assessment that 3B6 does not encode ncRNAs.

Translation of mRNA has long been known to be under regulatory control, but it is not generally thought of as a reversible mechanism for control of protein production. Considerable amounts of research in this area are revealing that translational regulation is far more complex than ever previously thought. The life history of an mRNA is largely determined by a wide array of heterogeneous interacting factors constituting its messenger ribonucleoprotein particle (mRNP). mRNP components may be thought of as adaptors mediating mRNA related processes including localization, translation, and degradation[227]. Proteins such as YB-1 and TIAR are proteins which may contribute to the mRNP which negatively influence translation of mRNAs during stress conditions[228,229]. Upon silencing of translation during cellular stress, the inactive mRNPs congregate into transient stress granules, where they are held until stress conditions are alleviated[227]. It has recently been shown that during the period of stress, translation of some of the silenced mRNAs may be required for repair of stress induced damage, and that this can be achieved by reversal of the imposed repression[230]. Upon alleviation of the stress stimuli, inactive mRNPs are either returned to active translation, or delivered to GW bodies for degradation. GW bodies are cytoplasmic foci, defined by localization of a

marker protein GW182, where mRNA degradation machinery is concentrated, as they are the sites of normal mRNA decay[231,232].

In addition to proteins, mRNP components may include miRNAs. As previously mentioned, miRNAs are a subset of ncRNAs, and they are well known to facilitate regulation of mRNA via RNA interference (RNAi; reviewed in chapter 6). miRNA mediated translational silencing is expected to affect up to one third of transcribed genes, based on imperfect complimentarity of miRNA with its target mRNA. In short interfering RNA (siRNA) mediated post translational gene silencing mechanisms, perfect complimentarity between a siRNA and its target mRNA leads to direct cleavage of the mRNA by the RNA-induced silencing complex (RISC). In miRNA mediated silencing mechanisms, imperfect pairing of an miRNA and its mRNA target leads to RISC association and transcript silencing concurrent with localization of the complex to GW bodies[233]. In absence of localization to the GW bodies, miRNA translational silencing is impaired; it has been shown that the interaction of GW182 and the major RISC catalytic protein Argonaute is required for this process[234-236].

This general silencing mechanism is still not well understood, and is generally thought of as irreversible, with the repressed mRNA undergoing degradation in the GW bodies. A recent paper though demonstrates that miRNA mediated silencing can be reversed, with concurrent mRNA escape from GW bodies, under conditions of cellular stress[237]. Relief from miRNA transcriptional repression was shown for the cationic amino acid transporter CAT-1, whose activity is maintained at low levels in the liver, except during stresses requiring liver regeneration as the response. This work demonstrates that destruction of miRNA translationally repressed mRNAs is not immediate in the GW bodies, which may act as temporary storage sites for such mRNAs[231].

Where protein mediated translational repression is generally in response to cellular stress, miRNA mediated repression could occur constitutively and be relieved by stress. This suggests that a mechanism, analogous to stabilization of proteins normally targeted to the proteasome in response to stimuli, may exist on the mRNA level. A gene,

such as 3B6, could be widely transcribed but constitutively silenced at the translational level by miRNA, its mRNA still detectable as it is not undergoing immediate degradation. Translational repression could be relieved given appropriate cues, resulting in production of mSe70-2. What exactly the appropriate cues may be remains to be seen. Currently the only de-repression cue known is cellular stress, but others may exist such as developmental cues. This concept that mSe70-2 may be regulated at the transcriptional level is highly speculative, but none the less, it will be very interesting to see how the field surrounding mechanisms of transcriptional regulation will progress.

5.4.4 Concluding remarks

Few of the results appearing in this chapter were expected to be a part of this document. Many of them were intended as 'quick and dirty' results for the sole purpose of troubleshooting. Hence, several of the figures are not controlled in a complete manner and may be considered inconclusive. It was felt that it was important to include them though, to provide a more complete view and reflect the reality of this work. Despite time dedicated to troubleshooting, these problems have not been resolved.

It is recognized that there are many avenues of exploration yet to be exhausted with respect to protein expression and antibody characterization. However, it was agreed that, in light of the time invested in this line of research with negligible returns, continued effort in this area be suspended in favour of work which could provide more positive contributions to the project.

Chapter Six: Preliminary functional characterization of mSe70-2

6.1 Summary

Following the challenges presented in the preceding chapter, this chapter describes several projects undertaken to pursue a preliminary functional analysis of mSe70-2. RNAi mediated knockdown of 3B6 transcripts is carried out in both P19 cells and the *C. elegans* system, revealing a possible role for 3B6 in regulation of the cell cycle in *C. elegans*. mSe70-2 is produced *in vitro*, and RNA affinity chromatography shows that it is likely a functional RNA binding protein. In addition, a phylogenetic analysis of mSe70-2 and its homologs shows its deep conservation in eukaryotes, including the novel finding of conservation amongst fungi, and confirmation of a previously noted gene duplication in vertebrates. Results from this chapter indicate that mSe70-2 is a protein highly conserved in evolution, which is likely functioning as a splicing factor.

6.2 Introduction

6.2.1 RNA interference

RNA interference (RNAi) has become a widely used method for analysis of genes. Functional knockdowns can be completed, even in high throughput format, in a less time and resource intensive manner than targeted gene disruption. RNAi can most simply be described as the process by which dsRNA silences gene expression. The silencing may occur post-transcriptionally by two methods: Degradation or translational inhibition of mRNAs. RNAi mechanisms have also been shown to affect gene regulation at the transcriptional level.

The RNAi pathway may be initiated in similar but distinct ways by short RNA molecules originating exongenously or endogenously. RNA with an exogenous origin will here be referred to as short- interfering RNA (siRNA), and that with an endogenous origin will be referred to as micro RNA (miRNA), as touched upon in the last chapter. Our current understanding of post-transcriptional gene silencing via RNAi suggests that initiation of silencing by either siRNA or miRNA leads to activation of the same general pathway, although as yet unidentified distinct pathways may exist.

RNAi is an ancient pathway predating the evolutionary divergence of plants and *C. elegans*, which likely originated as a mechanism for silencing of viral sequences and rogue genetic elements[238]. Indeed, in plants and *C. elegans* with aberrant RNAi pathway function, it has been found that some mobile genetic elements cannot be suppressed[239,240]. Though it was not realized at the time, RNAi was first observed in petunias when efforts to enrich their purple colour by introducing a transgene resulted in co-suppression of both the transgene and the endogenous gene[241]. RNAi was again stumbled upon in the *C. elegans* system during injection of antisense RNA for inactivation of the gene *par-1*. It was observed that both the sense and the antisense injections resulted in similar silencing, but the mechanism for this phenomenon was unknown[242]. It was the sharp reasoning of Fire and Mello that identified dsRNA as the mediator of the observed silencing[243]. They suggested that the RNA preparations used in the *par-1* experiments may have been contaminated with dsRNA. To test their hypothesis, they injected sense, antisense and

dsRNA corresponding to the *unc-22* gene into the gut of *C. elegans*. They found that silencing by the dsRNA was orders of magnitude more effective than that seen with sense or antisense RNA. They also noted that the dsRNA effect was propagated into the progeny of the injected *C. elegans*, and that the gene inactivation was due to degradation of the target mRNA[243]. Shortly after the findings of Fire and Mello, and their coining of the term RNAi, siRNA mediated gene silencing was confirmed in plants and *Drosophila*[244,245].

Discovery of the first miRNA, *lin-4*, occurred in *C. elegans* in the early 1990s[246]. *Lin-4* has partial complimentarity to a segment of the 3' UTR of *lin-14*, which it negatively regulates. *Lin-4* was found to pair imperfectly with *lin-14*, but the mechanism for silencing was unknown. For seven years after the discovery of *lin-4* no other miRNAs were found, as its homologs evaded bioinformatic methods. *Let-7*, another miRNA regulator of *lin-14* was later identified, and evidence for its sharing the RNAi silencing mechanism was established[247]. Homologs of *let-7* were identified using similarity based methods, indicating that RNAi is a general conserved mechanism for gene silencing[248].

RNAi is a two phase process involving initiation and execution of the pathway [Figure 28]. During the initiation phase, biogenesis of siRNA and miRNA occurs, while in the execution phase the RNAs are joined by protein components of the pathway and silencing occurs. Biogenesis of miRNA begins with transcription of endogenous primary miRNAs (pri-miRNAs), which are long molecules containing approximately 70 hairpins with imperfect internal sequence complementarity[238]. They are processed into separate hairpins by the nuclear RNAse III protein Drosha, and its dsRNA-binding-protein partner DGCR8, and then exported from the nucleus[249,250]. These pre-miRNA hairpin loops are then cleaved into mature 21-25 nt miRNAs by the cytoplasmic RNAse III enzyme Dicer, leaving 2 nt 3' overhangs and 5' phosphate groups[251]. Biogenesis of 21-25 nt siRNA occurs in a similar manner, beginning with dsRNA introduced from an exogenous source which is cleaved by Dicer, again leaving 2 nt 3' overhangs and 5' phosphate groups. Amplification of siRNA catalyzed by a RNA dependent RNA polymerase (RdRP) can

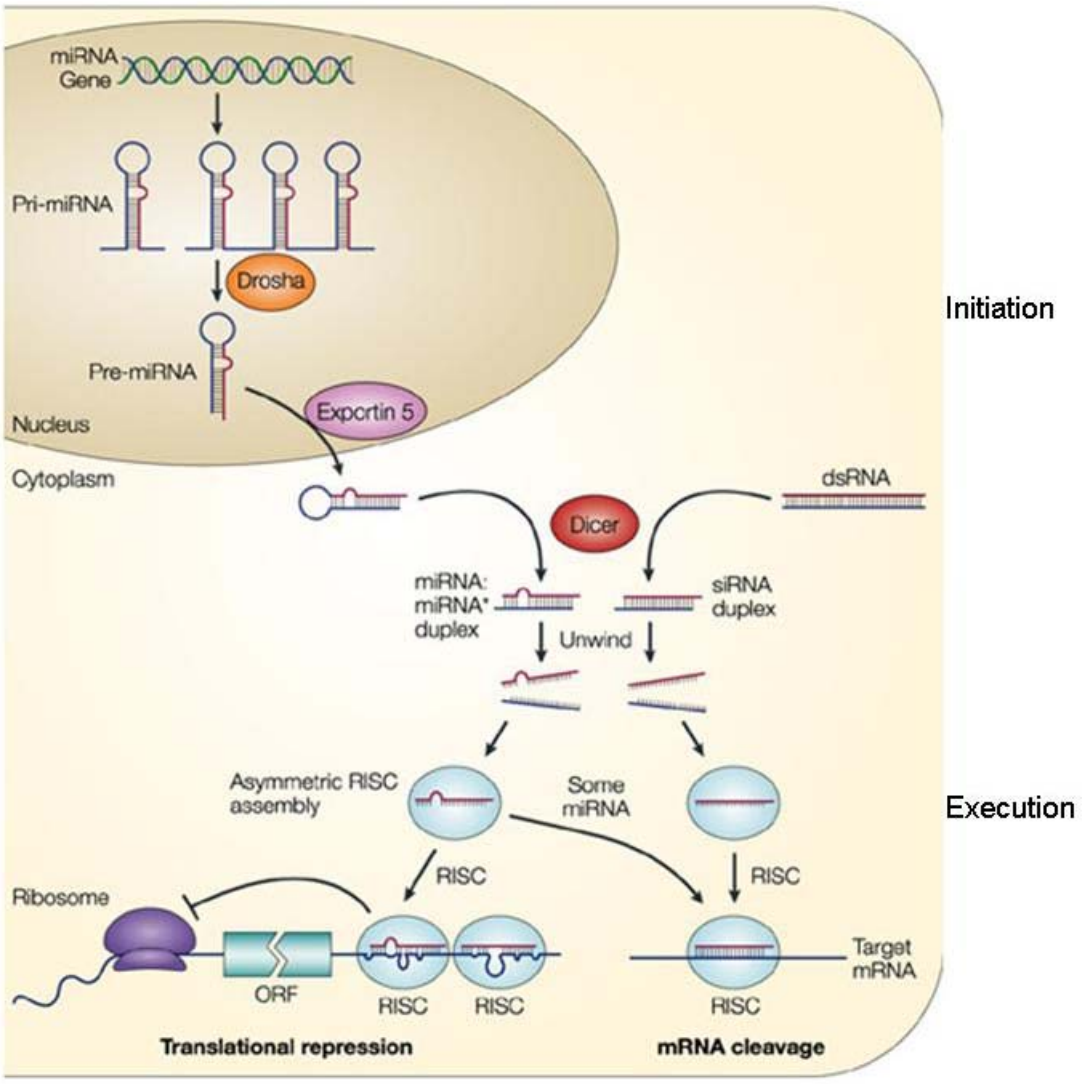


Figure 28: RISC facilitated mechanisms of RNAi mediated gene silencing. Description of the components and mechanisms is given in the text. Adapted from He and Hannon 2004[252].

occur in plants and *C. elegans*, likely by a siRNA acting as a primer for dsRNA elongation on the target mRNA[253]. This results in more dsRNA availability for siRNA production by Dicer, likely contributing to the efficiency of RNAi mediated silencing seen in plants and *C. elegans*. To date, RdRPs have not been found in *Drosophila* or mammals.

The execution phase of RNAi involves recognition of the target mRNA by miRNA or siRNA, and association of effector proteins which facilitate gene silencing. The best characterized RNA effector complex is the RNA induced silencing complex (RISC). Several activities are associated with RISC: Helicase, exonuclease and endonuclease (commonly referred to as slicer activity), and RNA binding. RISC is able to identify and preferentially incorporate the antisense strand from an unwinding si- or miRNA duplex[254]. RISC identification of miRNA and siRNA species is based on the presence of their 2 nt 3' overhangs and 5' phosphate groups, and recognition of the antisense strand occurs based on the 5' end stability of the duplexes. The strand with the least 5' end stability is selected for incorporation into RISC[255]. The Argonaute (Ago) protein, the major known component of RISC, confers this RNA binding activity, and also its slicer activity[234,256]. There are multiple proteins in the Ago family, suggesting that multiple RISCs may exist[257]. In addition, multiple Dicer proteins exist, and it has been suggested that different but analogous pathways exist to facilitate the different outcomes of RNAi[258].

The decision of whether translational inhibition or mRNA degradation will occur as the outcome of the RNAi pathway is based on the level of complementarity between the antisense RNA incorporated in RISC and its mRNA target. If the pairing of the miRNA and the target is imperfect, translational repression in association with GW bodies, as commented upon in chapter 5, occurs. In the case that perfect complimentarity exists between the siRNA or miRNA in RISC and the target mRNA, cleavage of the target mRNA by the slicer activity of RISC occurs. Another outcome of RNAi is transcriptional repression of the target gene via chromatin modification, involving a

different Ago-containing effector complex, the RNA initiated transcriptional silencing (RITS) complex[259-262].

RNA interference has been recognized as an inexpensive and straightforward way to perform loss of function studies. In plants, *C. elegans*, and *Drosophila*, introduction of dsRNA into the system is sufficient for induction of RNAi. RNAi in the *C. elegans* system is particularly convenient given that dsRNA can be delivered by both feeding of *E. coli* expressing dsRNA and direct soaking of *C. elegans* in dsRNA[263,264]. Microinjection is a more stringent method for dsRNA delivery into *C. elegans*, which yields optimal propagation of the RNAi effect into progeny.

With the ease of RNAi induction in *C. elegans* it was recognised that a new standard for systematic, genome-wide functional studies could be set. Major genome-wide RNAi screens have been carried out using dsRNA expressing E. coli feeding libraries. In what is the largest study to date, Ravi Kamath and colleagues generated a feeding library encoding dsRNA for nearly 17000 of the genes (~86.3%) in the *C. elegans* genome[265]. They used this library to systematically knock down expression of the genes in *C. elegans*, which were visually scored for a range of defects, and they found that approximately 10.3% of the genes screened gave obvious phenotypes. They noted that highly conserved genes were most likely to give aberrant phenotypes, while knock down of genes with highly specialized or redundant functions were least likely to show a detectable phenotype.

Methods for use of RNAi in mammalian systems have been more challenging to develop, as long dsRNA induces non-specific gene silencing via activation of dsRNA dependent protein kinase PKR, RNAseL activation, and interferon-α and –β production[266]. Important observations by Sayda Elbashir and colleagues revealed that induction of non-specific silencing could be avoided by directly introducing mature siRNAs or short hairpin RNA (shRNA) Dicer substrates rather than dsRNA into cells[267,268]. The potency of experimental RNAi in mammals is still not as effective as it is in *C. elegans*, as without RdRP there is no propagation or persistence of the effect. Vector driven methods for *in vivo* production of shRNA or siRNA can help to overcome

this issue, however, design of effective RNA sequences can still present challenges. To assist in siRNA design a variety of tools have been developed including, but not limited to, OptiRNA, Dharmacon siDESIGN tool (http://www.dharmacon.com/sidesign/default.aspx), and the Ambion siRNA target finder (http://www.ambion.com/techlib/misc/siRNA_finder.html)[269,270]. These tools and others incorporate research findings for rule-based siRNA design. Although siRNA design is still an inexact method, requiring experimental validation of several sequences to find one that is effective, many improvements have been made. Continued research in this area will surely reveal the overall importance of RNAi in development and maintenance of organisms, and will also likely yield methods for application of RNAi to therapeutics.

6.2.2 RS domain protein Peri-implantation stem cell-1

Peri-implantation stem cell-1 (Psc1) is a mouse RS domain protein encoded by *psc1*, which was isolated based on its differential expression in ES cells and early primitive ectoderm-like (EPL) cells. EPL cells are an *in vitro* derived pluripotent cell type equivalent to the primitive ectoderm of the early mouse embryo[271]. Psc1 contains a number of conserved protein domains including a CCCH zinc finger, a RRM, and an RS domain.

RS domains are defined by regions of arginine/serine dipeptides[272]. They are widely accepted as protein-protein interaction domains, but they also are capable of mediating protein-RNA interactions[273]. The most well known family of RS domain-containing proteins is the SR protein family, characterized by the presence of an RS domain along with one or two RRMs[272]. SR family proteins play multiple roles in pre-mRNA splicing, including roles in spliceosome assembly and splice site selection[274]. They are required for removal of constitutively spliced introns, and also function to regulate alternative splicing[273].

Aside from shared functional domains, SR family proteins are also characterized on the basis of their subcellular localization. They are commonly found localized to subnuclear structures referred to as nuclear speckles, which are sinks of pre-mRNA splicing factors defined by localization of an SR protein, SC35[275]. Nuclear speckles are

so enriched with splicing factors, that localization of a protein to them implies its involvement in pre-mRNA splicing[275]. Almost all splicing factors that localize to the nuclear speckles contain the RS domain, which has been shown to have roles in both the nuclear import and subnuclear localization of SR proteins[276,277]. Phosphorylation of the RS domain regulates SR protein localization, with hyperphosphorylated SR proteins found in the cytoplasm, and dephosphorylated ones found in the nuclear speckles[274]. In addition, partial phosphorylation of SR proteins increases their affinity for pre-mRNA and other proteins, thus promoting assembly of the spliceosome[274].

In a paper describing Psc1, Steven Kavanagh and colleagues examined its subcellular localization[278]. Psc, like other SR proteins, was found to localize to nuclear speckles, but also to smaller, non anti-SC35 staining nuclear regions generally found near the nuclear membrane. In addition, Psc1 was found in cytoplasmic foci which were determined to not be associated with GW bodies or stress granules. This localization profile is different than that for other SR proteins, they are typically not found in cytospeckles or non-anti-SC35 staining nuclear speckles. As with other SR proteins, the RS domain of Psc1 was shown to be required for its nuclear import and localization into nuclear speckles, and its RRM domain was also shown to be necessary for nuclear speckle localization. In addition the RRM was found to be necessary for cytospeckle formation, and functional in RNA binding.

The Kavanagh group also in their paper identified 13 homologs of Psc1, among which mSe70-2 was included as a paralog. Organisms represented in the list of homologs were mouse (*M. musculus*), human (*H. sapiens*), Frog (*X. laevis*), Fruitfly (*D. melanogaster*), mosquito (*A. gambiae*), worm (*C. elegans*), chicken (*G. gallus*), slime mold (*D. discoidium*), Puffer fish (*T. rubripes*), and rat (*R. norvegicus*). Psc1, mSe70-2, and Se70-2 share a total of eight regions of homology, including the CCCH zinc finger, RRM, RS domain, and several regions of unknown significance [Figure 29]. Eight of the homologs were presented as sharing a C-terminal acidic rich region, enriched for aspartate and glutamate residues. Based on this acidic rich region, the homologs were classified as a new family of acidic rich RS (ARRS) proteins.

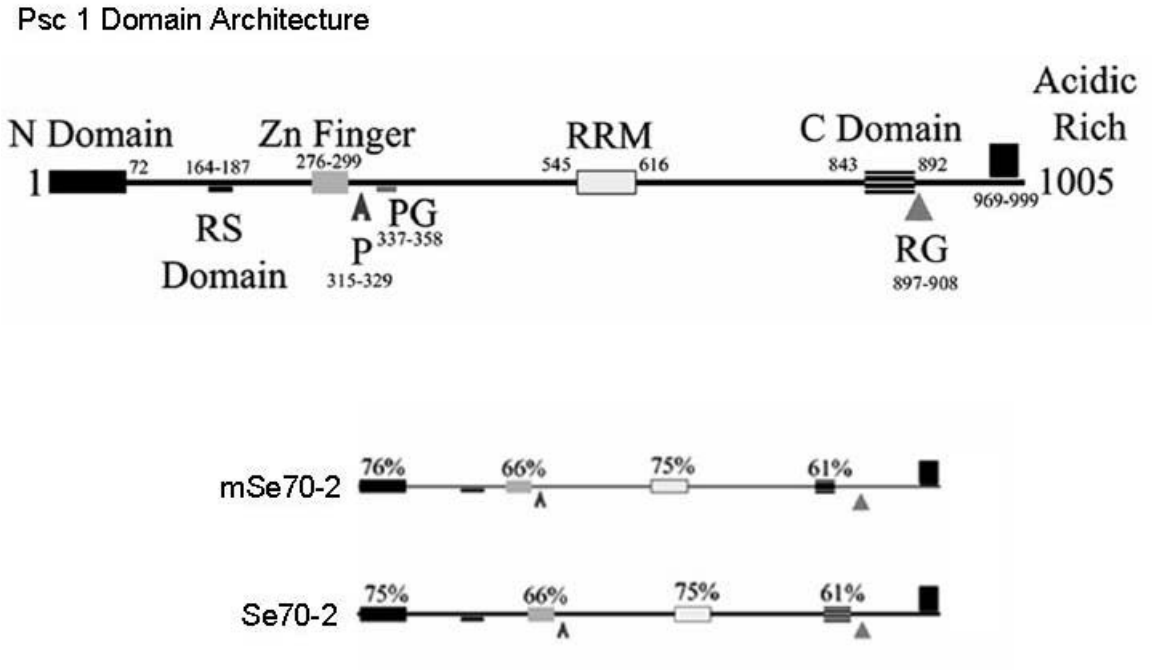


Figure 29: Conserved domain architecture of Psc 1, mSe70-2, and Se70-2. N domain, conserved N-terminal region of Psc 1 and its homologs; RS domain, arginine/serine repeats; Zn Finger, CCCH type zinc finger; P, proline-rich region; PG, proline/glycine repeats; RRM, RNA recognition motif; C domain, conserved C-terminal region of Psc 1 and its homologs; RG; arginine/glycine repeats; Acidic rich, aspartate/glutamate-rich region. Se70-2 and mSe70-2 do not contain the PG repeat region. Percentages associated with the conserved regions for Se70-2 and mSe70-2 represent the percent similarity throughout that region in comparison to Psc 1. Adapted from Kavanagh *et al.* 2005[278].

A basic phylogenetic analysis of the homologous proteins was carried out[278]. Although the method was not made completely clear, it appears that ClustalW alignment followed by neighbour joining or parsimony tree construction with bootstrap analysis was carried out. The most interesting result of this analysis is the identification of a gene duplication event in the vertebrate lineage. This analysis though, may not have been completely robust due to methodology. The amino acid substitution model was not determined, therefore assumption of homogeneous amino acid and site substitution rates was made, while it is known that this assumption does not hold true. It also seems that many non-homologous sites were included in the tree building, as there is no indication of their removal from the analysis. This is likely to lead to weak estimates of branch lengths, diminishing the ability to infer functional changes between diverging proteins. The tree presented may also be questionable in that although two homologs from chicken and one from rat are mentioned in the figure legend, they do not appear in the analysis.

Regardless of the strength of the phylogenetic analysis, it is of great interest that a paralog for mSe70-2 was identified. Based on localization studies and conserved domain detection, Psc1 is expected to function as a splicing factor, and it along with mSe70-2 and other homologs have been used to define the new ARRS family of proteins.

6.3 Results

6.3.1 RNAi mediated 3B6 knockdown in mammalian cells

With the lack of significant results yielded from attempts to express and detect mSe70-2, as described in chapter 5, it was clear that alternative methods for examination of mSe70-2 function were necessary. Continued updating of RED, which by this time itself had been redesigned, offered new information regarding 3B6. Among the new information were annotations from the International Gene Trap Consortium. One gene trapped ES cell line, F026D01, was annotated as disrupting the 3B6 locus. Over time, three more gene trapped ES cell lines, AP0798, AR0055, and AR0058, were also annotated to the 3B6 locus. All four of the cell lines were ordered form their respective divisions of the gene trap consortium with the intention of using them for differentiation

studies, blastocyst injections, and expression studies. None of the cell lines were found to be appropriate for use, and thus this line of investigation was discontinued. Details regarding the cell lines and their characterization can be found in Appendix E.

Following characterization of the gene trapped cell lines, it was decided that RNAi mediated knockdown of 3B6 would be pursued. A vector for targeted disruption of 3B6 was under development in the lab, but due to timing and resource constraints, RNAi was deemed the better approach. The cell line that was chosen for the knockdown study was P19, for their ease of culture, amenability to transfection, and ability to differentiate into neural cell types using the Bain RA induction protocol. If effective knockdown of 3B6 was achieved in P19 cells, then studies of the effect of reduced 3B6/mSe70-2 levels during *in vitro* neural differentiation could be carried out. The method for RNAi chosen was vector-based, using pSUPER.gfp.neo for expression of shRNA Dicer substrates. Design of the shRNA hairpins for 3B6 targeting was done using several web-based tools, and sequences identified as appropriate by more than one tool were selected for use. Sequences encoding two different shRNAs directed at 3B6 were cloned into pSUPER.gfp.neo resulting in the constructs pSUPER-KD1 and pSUPER-KD2. P19 cells were transfected with pSUPER-KD1 or pSUPER-KD2 to test for knockdown of 3B6. 24 hours post-transfection, cells were collected for RNA preparation, and RT-PCR using 3B6 primers was carried out to detect 3B6 transcript levels normalized to GAP-DH. No knockdown of 3B6 was seen in these initial tests (data not shown). The experiments were repeated, but 24 hours post-transfection the cells were sorted based on GFP expression from the vector. Sorted cells were collected and RNA was prepared from the GFP expressing cells as well as control cells. RT-PCR was again carried out using 3B6 primers, and densitometry was carried out on the resulting products after normalization to GAP-DH. Results indicated that the pSUPER-KD1 construct induced a 53% reduction in detectable 3B6 transcript levels in transfected P19 cells [Figure 30].

Despite the observed knockdown of 3B6, it was felt that the effect was not great enough, and this direction for the project was abandoned. The knockdown was estimated

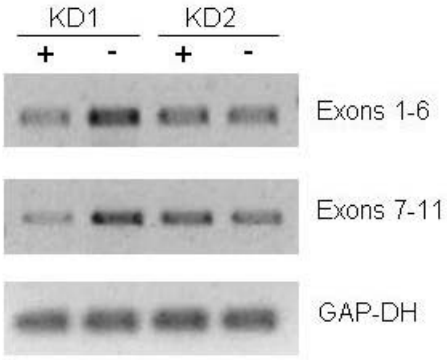


Figure 30: Vector-based RNAi knockdown of 3B6 in P19 cells. P19 cells were transfected with pSUPER-KD1 or pSUPER-KD2 (+), or left untransfected (-) to test for knock down of 3B6. 24 hours post transfection, cells were subjected to FACS on the basis of GFP expression and sorted cells were collected for RNA preparation. RT-PCR using primers directed at 3B6 exon 1-6 or exons 7-11 was completed. Densitometry was carried out using Bio-Rad Quantity 1 software after normalization to GAP-DH, indicating a ~53% reduction in 3B6 transcript levels is induced by pSUPER-KD1.

to be only in the 50% range, and it was suspected that this change might not be enough to cause a significantly measurable effect in RA induced neural differentiation, which already is a system susceptible to experimental noise. Other shRNA constructs could have been produced in order to find a more effective knock down, but again under time and resource constraints this was agreed to be imprudent.

6.3.2 RNAi mediated knockdown of the C. elegans 3B6 homolog

To gain indication of functional importance of 3B6/mSe70-2, the *C. elegans* (worm) model system was looked to. A homolog of 3B6 exists in *C. elegans*, B0336.3 (WormBase ID: WBGene0015143), whose expected product shares 26% identity and 37% similarity with mSe70-2, over its entire length. They are reciprocal best hits, and the B0336.3 protein is expected to have a domain architecture similar to that of mSe70-2.

In the genome wide RNAi screen conducted by Ravi Kamath and colleagues, B0336.3 was one of the genes for which a feeding library clone was successfully established[265]. The RNAi associated phenotype scored for B0336.3 was slow growth (gro), and this phenotype was supported by another high throughput RNAi screen as well[279]. However, there are also occurrences of the B0336.3 phenotype being scored as normal[280-282]. In order to resolve this discrepancy and gain insight into B0336.3 function, its RNAi mediated knockdown was undertaken. The feeding library clone used in the Kamath screen was obtained and dsRNA was produced by *in vitro* transcription from T7 promoters flanking the insert. dsRNA was introduced into *C. elegans* by microinjection in blind studies. 100 hour old F1 generation adult worms or 38 hour old F1 generation fourth larval stage (L4) worms were scored for size. At the adult stage worms injected with dsRNA targeting B0336.3 (si3B6) were observed to be on average 9.1% smaller than those injected with control dsRNA (siGFP) [Figure 31, B]. This result was found to be significant with $p<0.001$ using a one sided t-test of significance. The observed growth effect was more prominent at the L4 stage, as worms injected with si3B6 were on average 30% smaller than those injected with siGFP [Figure 31, C]. This effect was also significant with $p<0.001$. These results confirm the gro phenotype previously observed for B0336.3 knockdown worms. All of the worms in this study were found to be

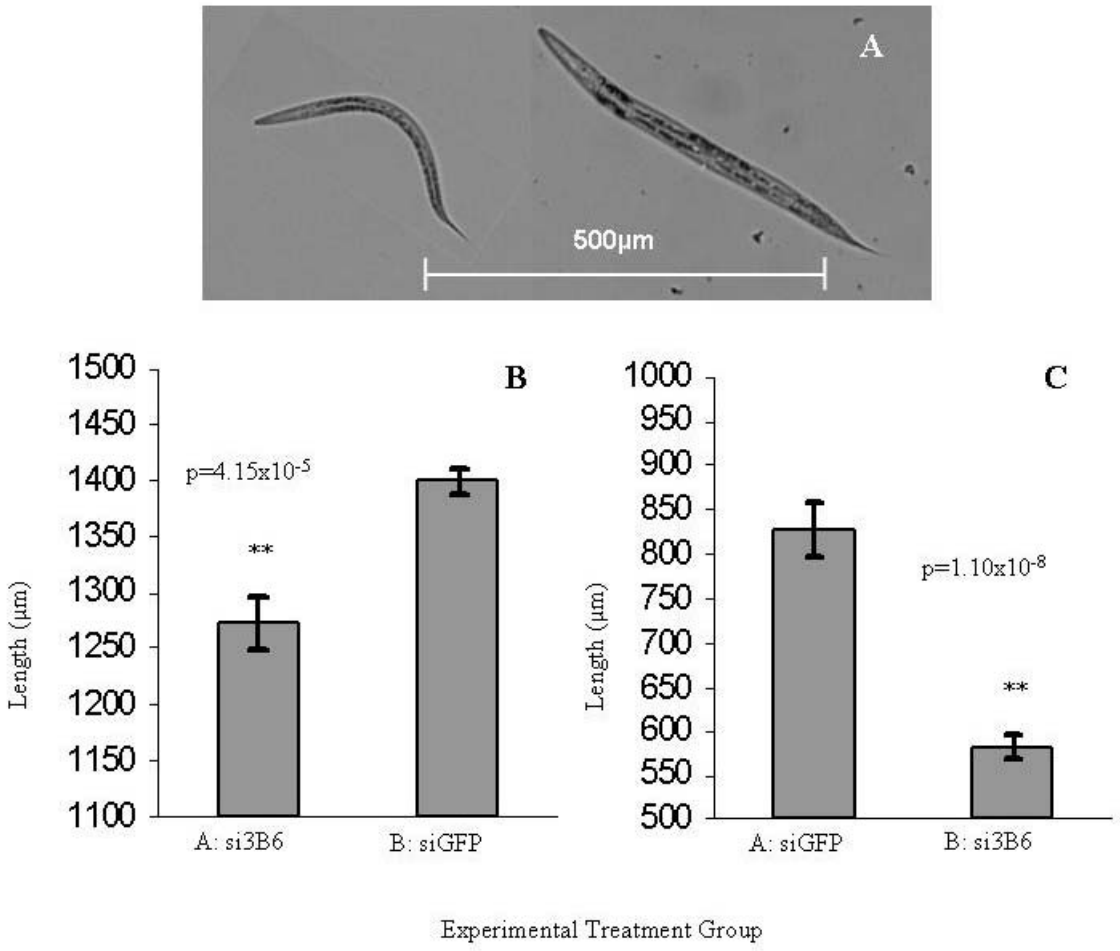


Figure 31: C. elegans with RNAi mediated knockdown of B0336.3 show growth defects. Wild type N2 F0 generation worms were injected with dsRNA targeting either B0336.3 (C. elegans 3B6 homolog; si3B6) or GFP (control; siGFP) and designated treatment group A or B in blind studies. F1 generation worms at 100 hours old (Adults; Panel B, experiment 1) or 38 hours old (L4 stage; Panel C, experiment 2) for each treatment group were pooled, anesthetized, and measured. Panel A shows representative worms from the second experiment. In the first experiment (panel B) treatment group A worms were found to be on average 9.1% smaller than those in treatment group B. The size difference between the two groups was found to be significant using a one-sided t-test ($p=4.15x10^{-5} < 0.001$), and it was confirmed that treatment group A was injected with si3B6. In the second experiment (panel C), group B worms were found to be an average of 30% smaller than those in group A, and this result was also found to be significant using a one-sided t-test ($p=1.10x10^{-8} < 0.001$). Group B was confirmed to have been injected with si3B6.

morphologically normal based on visual inspection (Dr. Jim McGhee, pers. comm.). During the first experiment scoring for worm size, it was observed that the si3B6 injected group may not have laid as many eggs as the control group. To follow this observation up, experiments scoring for abnormal egg laying phenotypes were carried out, however no difference in egg laying in either the F0 or F1 generation was noted between the si3B6 and control worms (data not shown). The initial observed difference in egg laying is likely attributable to the longer time taken for the si3B6 injected worms to reach egg laying maturity.

6.3.3 In vitro production of mSe70-2 and RNA affinity chromatography.

As mSe70-2 was not successfully produced in cells, its production in cell free systems was pursued. For this purpose, two new constructs behind T7 promoters in pBS were produced: T7-3B6 and T7-FLAG-3B6. Both of these constructs along with empty pBS vector and a T7-FLAG-hTERT construct were subjected to *in vitro* coupled transcription and translation (TnT) in the presence of ^{35}S methionine. To check for protein production, a sample of each reaction was separated by SDS-PAGE, and the resulting gel was dried and exposed to film. The expected 112 kDa mSe70-2 appeared to be produced both from the T7-3B6 and T7-FLAG-3B6 construct [Figure 32, A]. mSe70-2 was produced more efficiently from the T7-FLAG-3B6 construct. While production of extraneous products is a common problem encountered with TnT reactions, the upper 150 kDa band produced by both constructs was perplexing, given that both constructs contained stop codons in all reading frames following the 3B6 coding sequence. To determine if this product was the result of read-through past the stop codons, T7-FLAG-3B6 was linearized 3' of the stop codons, and again subjected to TnT in the presence of ^{35}S methionine. Results showed that mSe70-2 is more efficiently produced from the circular T7-FLAG-3B6 construct than the linear version, and that the 150 kDa product was not the result of read through past the stop codons, as it also appeared in the reactions using the linear template [Figure 32, B]. The origin of the 150 kDa product remained unknown.

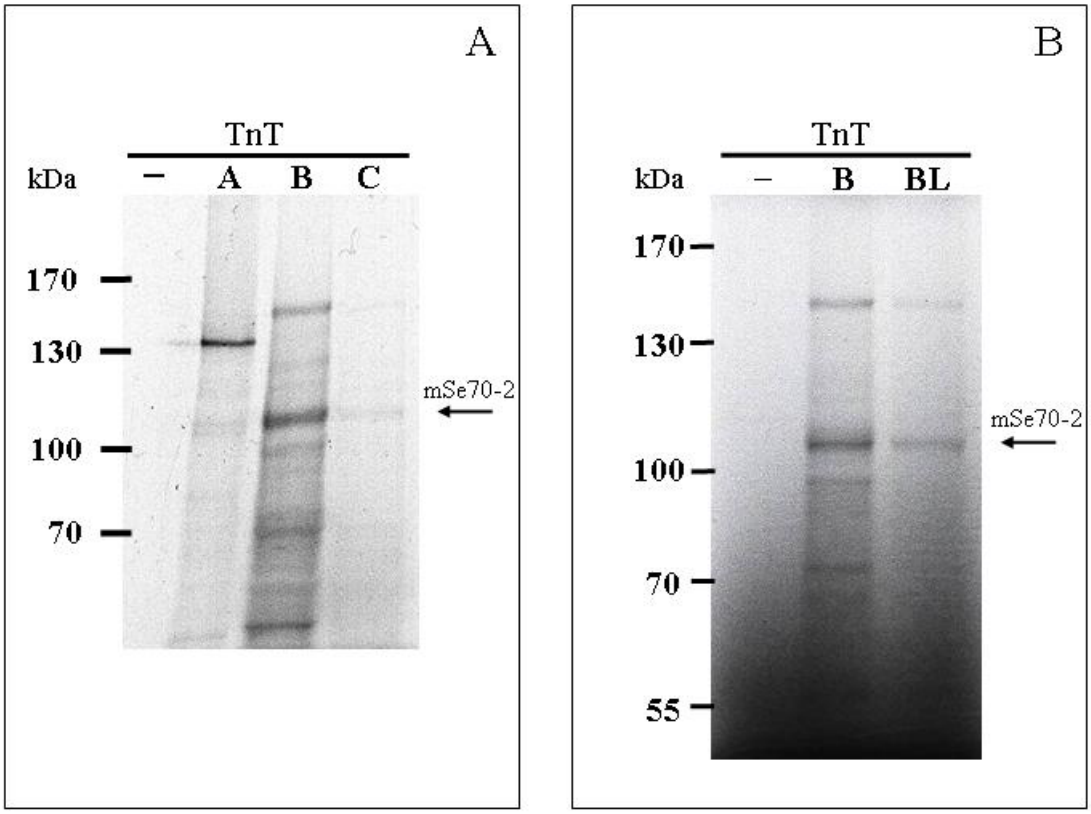


Figure 32: Production of mSe70-2 using TnT. Protein was produced using standard TnT reactions including ^{35}S-methionine and 100ng of each indicated template construct, and separated using 6% SDS-PAGE. Resulting gels were exposed to film. Expected sizes of the produced proteins were approximately 140 kDA from T7-FLAG-hTERT, and 112 kDa from T7-FLAG-3B6 and T7-3B6. The band expected to represent 3B6 is indicated in each panel. -, empty vector (negative control); A, T7-FLAG-hTERT (positive control); B, T7-FLAG-3B6; BL, linearized T7-FLAG-3B6; C, T7-3B6. Panel A: mSe70-2 is more effectively produced from the T7-FLAG-3B6 construct than the T7-3B6 construct. Panel B: mSe70-2 is more effectively produced using circular template (B) than linearized (BL) template. In addition, production of the larger product seen at approximately 150 kDa could not be eliminated by linearization of the template construct.

To provide more evidence that the ~112 kDa product generated from T7-FLAG-3B6 is FLAG-mSe70-2, TnT using T7-FLAG-3B6 followed by immunoprecipitation with bead conjugated anti-FLAG M2 antibodies was completed. Results showed that the 112 kDa product was immunoprecipitated using the anti-FLAG M2 antibodies, providing more evidence that FLAG-mSe70-2 was being produced [Figure 33]. Interestingly, the 150 kDa product of unknown nature was also immunoprecipitated using the anti-FLAG antibodies.

mSe70-2 likely has the ability to bind RNA, and it contains an RRM and a newly identified RS domain which may facilitate protein-RNA interactions. To examine if mSe70-2 could bind RNA, and if there was any preference for sequences rich in particular ribonucleotides, RNA affinity chromatography using homoribopolymer conjugated beads was carried out. Ku80 was used as a positive control as it is known to have affinity for RNA and had been used as a control for RNA binding studies in the lab of Dr. Tara Beattie, and hTERT was selected as a negative control as it was expected to bind only its RNA component, hTR, in a highly specific manner[283,284]. Results indicated that mSe70-2 could bind to poly-G homoribopolymers, indicating that it is a functional RNA binding protein which may have some preference for G-rich sequences [Figure 34]. Confidence in this result must be reduced however, as the negative control hTERT also was found to bind poly-G beads as well as poly-A beads, making the Se70-2 result inconclusive. In retrospect, hTERT was not an appropriate choice of negative control for these experiments as it has high affinity for the G-rich single stranded overhangs of telomeres, and may therefore also be able to bind G-rich RNA[285]. Although the results are inconclusive, they may indicate that mSe70-2 functions as an RNA binding protein.

6.3.4 Phylogenetic analysis of mSe70-2

A phylogenetic analysis of mSe70-2 was undertaken in order to establish the extent of its conservation in eukaryotes, and also to provide a more robust analysis with more accurate branch lengths than that presented in the Kavanagh paper describing Psc1[278]. First, homologs of mSe70-2 were identified from a broad Eukaryotic taxonomic sample, including mammals, amphibians, birds, boney fish, insects, nematodes, echinoderms, and

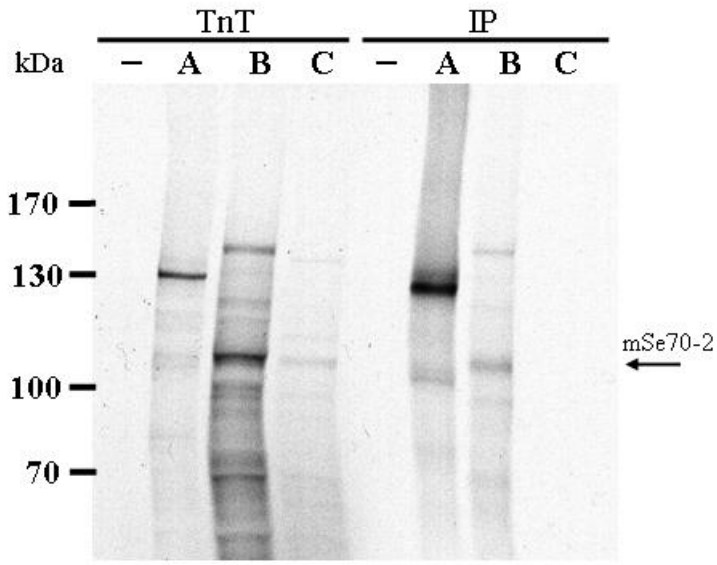


Figure 33: FLAG tagged mSe70-2 is produced using TnT, and can be immunoprecipitated using anti-FLAG antibodies. Protein was produced using standard TnT reactions including 35S-methionine and 100ng of each indicated template construct, and then immunoprecipitated using bead conjugated anti-FLAG M2 monoclonal antibody. -, empty vector (negative control); A, T7-FLAG-hTERT (IP positive control); B, T7-FLAG-3B6; C, T7-3B6 (IP negative control). FLAG tagged mSe70-2 is pulled down using anti-FLAG antibodies, as is the larger 150 kDa band, indicating the presence of FLAG.

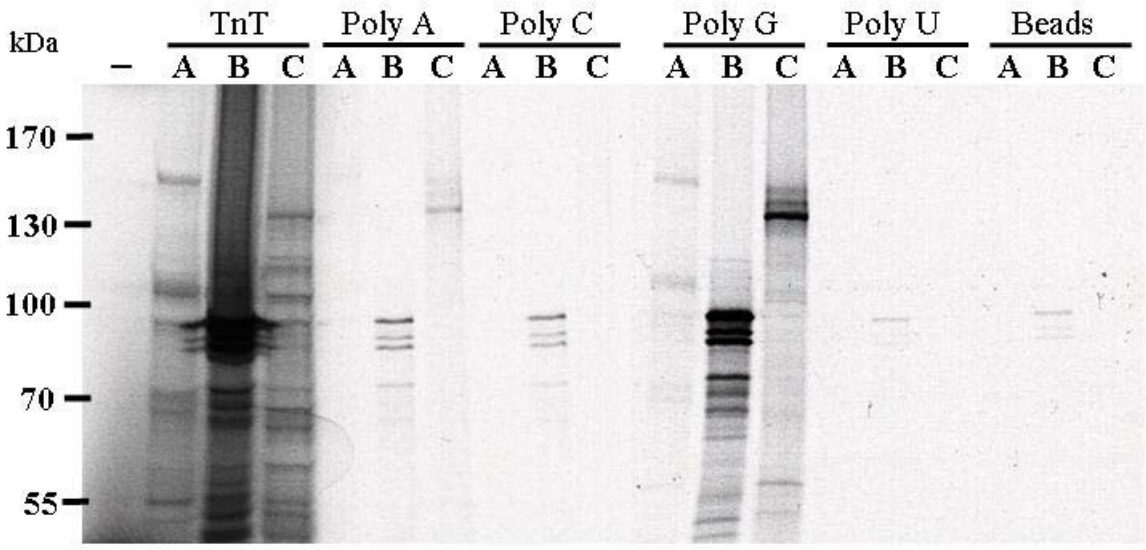


Figure 34: TnT produced mSe70-2 may associate with poly-G homoribopolymers. Protein was produced using standard TnT reactions including 35S-methionine and 100ng of each indicated template construct, and then subjected to RNA affinity chromatography using homoribopolymer conjugated agarose beads as indicated. -, empty vector; A, T7-FLAG-3B6; B, T7-Ku80 (positive control); C, T7-FLAG-hTERT (negative control). mSe70-2 and was seen to selectively associate with poly-G conjugated beads. The negative control hTERT was found to interact with poly-A and poly-G conjugated beads, therefore reducing confidence in the mSe70-2 interaction.

fungi, as well as the mycetozoan *Dictyostelium discoidium.* Sequences homologous to mSe70-2 were gathered using BLASTp searches against the NCBI RefSeq protein datasets limited to each organism of interest. For organisms of interest where RefSeq datasets did not contain Se70-2 homologs, converging Phi-BLAST searches were completed against the NCBI nr protein datasets limited to each organism. All homologs were confirmed using reciprocal BLASTp searches against the NCBI *M. musculus* RefSeq dataset. Each sequence was found to be the reciprocal best hit to the *M. musculus* Se70-2 sequence represented in the NCBI RefSeq dataset, and is listed in Table 3. Note that since the publication of Psc1, both the mouse protein and its human paralog have been renamed RNA binding motif protein 27 (Rbm27). In each vertebrate but *X. laevis*, two homologs were identified. It is expected that a second homolog does exist, but was not detected due to incomplete sequence coverage or assembly of the *X. laevis* genome at this time.

Multiple sequence alignment of the mSe70-2 homologs was completed using Clustal X with the default parameters of the BLOSUM62 scoring matrix, with gap opening and extension penalties of 11 and 1 respectively. The resulting alignment was imported into MacClade for identification and removal of the uninformative non-homologous sequence sites. The complete multiple alignment with indication of the sites included in further analyses is available in Appendix F. After determination of the most appropriate model of evolution to describe the given dataset, tree topologies were established using both Bayesian and bootstrapped Maximum Likelihood analysis of the data. The consensus Bayesian tree, which was corroborated by the Maximum Likelihood tree, is presented in Figure 35, with Maximum Likelihood bootstrap values and Bayesian posterior probability given as node support values. Results from the analysis indicate strong support for the vertebrate gene duplication with the vertebrate homologs forming two clades, one for Se70-2 sequences, and one for the paralogous Rbm27 sequences. Comparison of branch lengths given for the Se70-2 and Rbm27 clades suggests that the genes in the Rbm27 clade are more divergent in comparison to the evolutionary parental sequence than the genes in the Se70-2 clade are. It cannot be established which clade

Table 3: Sequences used in the phylogenetic analysis of mSe70-2.

Organism	Taxonomy	Gene Identifier	Description
Mus musculus	*Eu., met., ver., mam.*	N/A	mSe70-2
Mus musculus	*Eu., met., ver., mam.*	63723782	PREDICTED: RNA binding motif protein 27
Homo sapiens	*Eu., met., ver., mam.*	31652264	cutaneous T-cell lymphoma tumor antigen Se70-2
Homo sapiens	*Eu., met., ver., mam.*	37549972	PREDICTED: RNA binding motif protein 27
Gallus gallus	*Eu., met., ver., aves*	50730697	PREDICTED: similar to Cutaneous T-cell lymphoma tumor antigen se70-2
Gallus gallus	*Eu., met., ver., aves*	50755232	PREDICTED: similar to KIAA1311 protein
Tetraodon nigroviridis	*Eu., met., ver., acti*	47219258	unnamed protein product
Tetraodon nigroviridis	*Eu., met., ver., acti*	47230586	unnamed protein product
Xenopus laevis	*Eu., met., ver., amp.*	27696233	MGC52868 protein
Drosophila melanogaster	*Eu., met., art, ins.*	20129635	CG10084-PA, isoform A
Apis mellifera	*Eu., met., art, ins.*	66560588	PREDICTED: similar to cutaneous T-cell lymphoma tumor antigen se70-2
Caenorhabditis elegans	*Eu., met., nem.*	17551820	B0336.3
Strongylocentrotus purpuratus	*Eu., met., echinod.*	72007081	PREDICTED: similar to cutaneous T-cell lymphoma tumor antigen se70-2
Ustilago maydis 521	*Eu., fungi*	71013562	hypothetical protein UM02478.1
Schizosaccharomyces pombe	*Eu., fungi*	6522994	SPBC902.04
Neurospora crassa	*Eu., fungi*	28923887	hypothetical protein
Cryptococcus neoformans	*Eu., fungi*	57229727	conserved hypothetical protein
Dictyostelium discoideum	*Eu., mycetozoa*	66818002	hypothetical protein DDB0169078

All mSe70-2 homologs were confirmed using reciprocal BLASTp searches against the NCBI *M. musculus* RefSeq dataset. Each sequence was found to be the reciprocal best hit to the *M. musculus* Se70-2 sequence represented in the NCBI RefSeq dataset. Eu, eukaryota; met, metazoa; ver, vertebrata; mam, mammalia; acti, actinopterygii; amp, amphibia; art, arthropoda; ins, insecta; nem, nematoda; echinod, echinodermata.

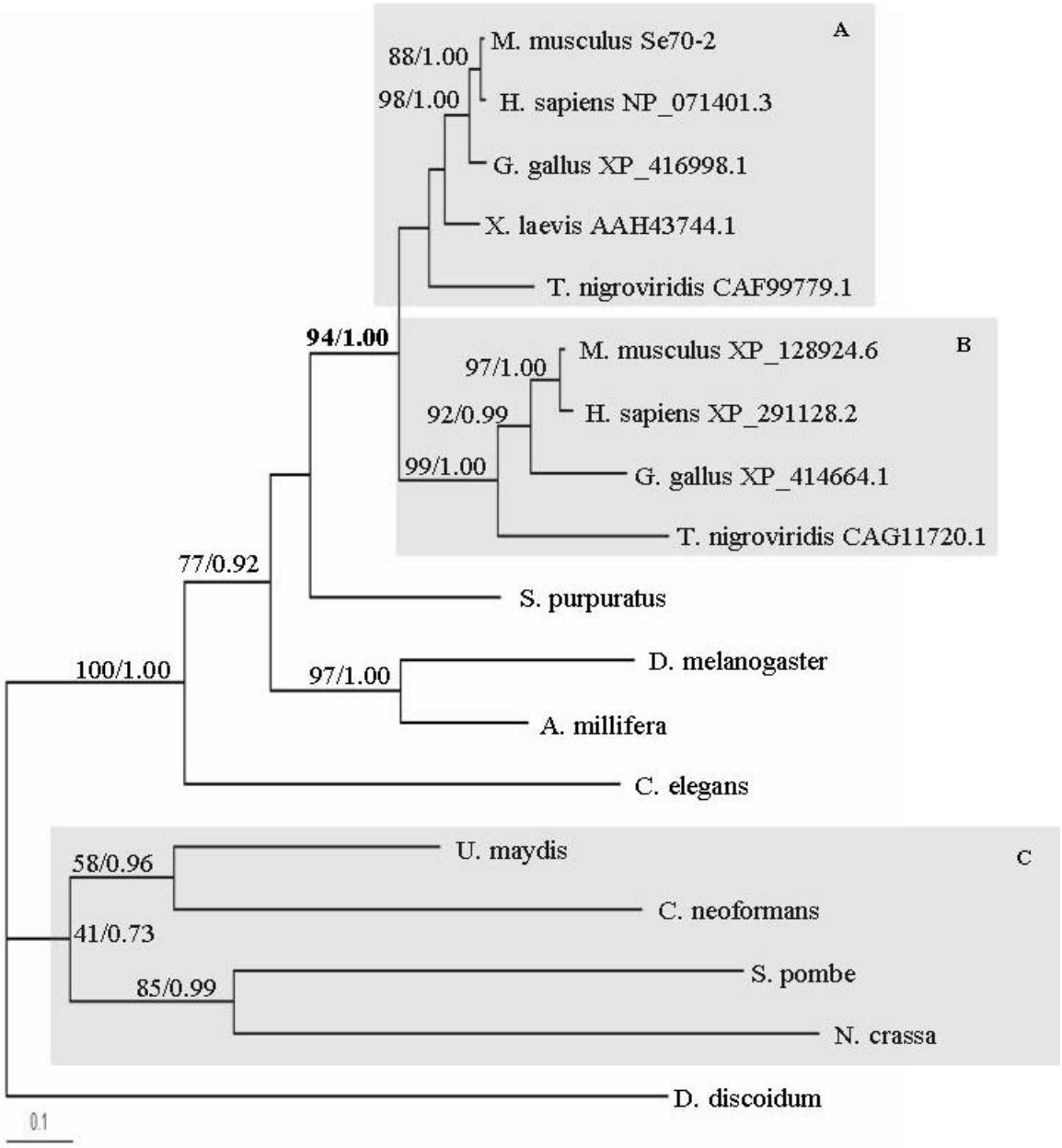


Figure 35: Phylogenetic tree illustrating the relationships between selected mSe70-2 homologs. The consensus Bayesian tree is shown, rooted with the outgroup sequence from *D. discoidum*. Node support values given are maximum likelihood bootstrap values and Bayesian posterior probability values respectively. There is strong support for a vertebrate gene duplication (in bold font), with shaded area A representing a Se70-2 clade, and shaded area B representing a Rbm27 clade. The fungal homologs also form a clade, albeit with weak support, as shown in shaded area C. Only nodes supported by better than 70% bootstrap support and 0.9 posterior probability are shown, with the exception of those for the fungal clade.

represents the evolutionary parental gene, and which gene represents the duplicated gene. Divergence or specification of protein function in the Rbm27 clade compared to the mSe70-2 clade may be inferred, though realistically cannot be confirmed in this analysis. These results also show the fungal homologs forming a single clade. Fungal homologs for Se70-2 or Rbm27 have not previously been identified, making this result novel and significant.

6.4 Discussion

6.4.1 The C. elegans homolog of 3B6 is involved in growth regulation

RNA interference is an effective method for investigating gene function when protein or antibodies are not available, as in the case of 3B6/mSe70-2. It was thought that RNAi mediated knockdowns in P19 cells would be most effective, allowing for examination of 3B6 function in differentiating cells. Indeed, had an effective knockdown of 3B6 been rapidly achieved in P19 cells, it is likely that an assay could have been developed to gage for alterations in neural differentiation, if any. However, this was not the case, and it was felt that the time needed for knockdown and assay development was not available for investment, hence precipitating the examination of 3B6 in the *C. elegans* system.

The *C. elegans* homolog of 3B6 is B0336.3, which is uncharacterized except for predictions and its appearance in a number of high throughput screens. Like mSe70-2, it contains a CCCH type zinc finger, a RRM, and a RS domain, with these domains arranged in an architecture similar to that of Psc1. It has been assigned the Gene Ontology terms 'Nucleic acid binding' and 'zinc ion binding'. In the Kavanagh paper describing Psc1, it is included in the ARRS family of proteins[278]. In a project working to achieve genome-wide gene expression analysis, the BC C.elegans Gene Expression Consortium have used a B0336.3::gfp transcriptional fusion to examine the expression pattern of B0336.3 *in vivo*[286]. GFP under control of the B0336.3 promoter was found to be widely expressed in both larval and adult *C. elegans*.

Expression of B0336.3 has been knocked down in several genome-scale RNAi screens[265,274,280-282]. In some of these screens the resulting phenotype was said to be slow growth (gro) while in others it was said to be normal. Reasons for this discrepancy are likely due to the methodological nature of these experiments. In each, a large number of knockdowns were being scored for phenotypes visually. This is likely to result in only obvious, unambiguous phenotypes being scored. In addition, high-throughput experiments are typically done by feeding which, though the fastest and least time consuming method, is less effective and reliable than administering dsRNA by microinjection[287].

The results from the experiments completed in this project confirm the gro phenotype that had previously been seen in other RNAi knockdown studies of B0336.3. This effect was more pronounced at the L4 stage than in adults, which may also have contributed to the earlier discrepancies in observed phenotype. Given the observation of the gro phenotype, it may be concluded that B0336.3 plays a role in positive regulation of cell cycle progression, though how it exerts this function is not known.

Other *C. elegans* genes that are expected to encode RNA binding proteins also showed the gro phenotype when subjected to RNAi in high throughput experiments. Among these genes are known and predicted splicing factors including *rbd-1*, *rnp-6*, *rsp-7* and *rsp-8*. The *rbd-1* product contains six RRMs and is necessary for ribosome biogenesis, as it has been shown to be essential for processing of pre-rRNAs[288,289]. The product encoded by *rnp-6* is homologous to PUF60 in mammals and Half-pint in Drosophila, both of which can bind the poly-pyrimidine tracts of introns and regulate constitutive or alternative splicing[290,291]. Neither *rsp-7* nor *rsp-8* are well characterized, but they are known to encode SR proteins containing the characteristic RRM and RS domain, with roles in spliceosome formation and regulation of pre-mRNA splicing[292]. It seems likely that these genes as well as 3B6 exert their ability to positively influence the cell cycle by regulating a variety of other genes on the mRNA level.

6.4.2 mSe70-2 is likely a splicing factor

A fair body of evidence implicates the 3B6-encoded protein mSe70-2 as a splicing factor. Probably most suggestive of this role is the domain architecture of mSe70-2, containing a CCCH zinc finger, a RRM, and a RS domain. The combination of the RS domain, discussed in section 6.2.2, and the RRM is characteristic of the SR family of splicing factors. The presence of the RS domain in mSe70-2 will likely lead to subcellular localization in nuclear speckles when it is actively being produced. This subcelleular localization is also diagnostic of SR proteins.

Even if the SR protein was non-existent in mSe70-2, the presence of the CCCH type zinc finger and the RRM alone is enough to suggest roles in pre-mRNA processing. Though zinc finger domains are typically known as DNA binding domains, the prototypical CCCH zinc finger from the tristetraprolin (TTP) protein is known to directly bind AU rich elements in the TNFα mRNA, causing its destabilization[293-295]. CCCH zinc finger domains in other proteins such as splicing factor U2AF have also been shown to be RNA binding, and in addition these zinc fingers are commonly found in proteins with involvement in cell cycle regulation[296]. U2AF also contains an RRM domain, which almost exclusively binds RNA, and is found in a variety of proteins implicated in RNA metabolism[297-299].

The mSe70-2 paralog Rbm27 (earlier referred to as Psc1), has already been shown to be an SR family protein, with its characteristic functional domains and subcellular localization. As shown, mSe70-2 possesses all of the sequenced-based characteristics of SR proteins, but due to the paralogous relationship between mSe70-2 and Rbm27 we cannot expect them to retain the exact same function. Results of the phylogenetic analysis indicate that the Rbm27 clade is diverging more rapidly than the mSe70-2 clade, hence mSe70-2 is more likely to retain the ancestral function. It is also apparent from the paper describing Psc1 that only the sequences of Rbm27 clade contain noted proline/glycine dipeptide repeats. Neither the mSe70-2 clade nor any of the non-vertebrate homologs contain this repeat region, further suggesting that mSe70-2 retains the ancestral gene function. This is not to propose that Rbm27 is a splicing factor, while

mSe70-2 is not; given their level of sequence identity and the conservation of functional domains it is most likely that both are splicing factors, but Rbm27 has diverged in order to function in a different niche than mSe70-2. In addition, the conservation of the mSe70-2 throughout eukaryotes, including fungi as shown in this analysis, implies the importance of its function in general cellular systems.

Also in support of mSe70-2's putative role as a splicing factor is the RNAi phenotype of its *C. elegans* homolog which is shared with a variety of other splicing factors, as discussed in section 6.4.1. In addition, it has been shown, though not definitively, via TnT and RNA affinity chromatography, that mSe70-2 does bind RNA. A question remaining from the TnT based experiments regards the identity of the 150 kDa product observed to interact both with anti-FLAG antibodies and poly-G homoribopolymer beads, just as mSe70-2 was observed to do. The TnT experiments were carried out using rabbit reticulocyte lysate (RRL), which contains all components necessary for protein synthesis. There is some evidence in the literature for direct interaction between proteins recruited to the spliceosome and the translation initiation complex[300,301]. It is possible that, as a spliceosomeal component, mSe70-2 could interact with protein components of the translation initiation complex which are present in the RRL. This interaction may be detected on SDS-PAGE if denaturation of the proteins was incomplete. Ideally, it would be of interest to pursue the identification of the 150 kDa product using mass spectrometry, to see if such an interaction is present.

Taken together, results point to the probability that mSe70-2 is a splicing factor. Experimental demonstration that mSe70-2 is a splicing factor would of course unequivocally confirm this assessment, however its mRNA substrates are completely unknown, as is its expression pattern at the protein level. Hence, its role in mouse development and cellular maintenance, if any, as well as the mechanism by which it is regulated at this point remains conjectured.

Chapter Seven: Synthesis and Conclusions

7.1 Recap of project

The project presented in this document surrounds the identification and characterization of mSe70-2, a protein encoded by *Rbm26*, formerly referred to as 3B6. Work presented includes the development and implementation of a data enrichment, management, and dissemination strategy for the Bain-Rancourt EST dataset: The Rancourt EST database (RED). This led to the identification of the candidate gene 3B6, which adhered to criteria suggesting that it may encode a nucleic acid binding protein involved in the process of neural differentiation. The genomic structure of *Rbm26* and a number of transcripts from the gene were identified. The analysis of *Rbm26* at the level of its mRNA expression was pursued, and its expression in differentiating ES cell cultures, embryos, and adult tissues was shown. Its mRNA expression was found to be at a higher level in differentiating ES cell cultures enriched for neural precursors in comparison to spontaneously differentiating or non-differentiating ES cell cultures. *Rbm26* mRNA expression in embryos was found to be widespread, particularly in neural and mesenchymal tissues. In addition, the timing of its expression corresponded to that of the peak period of neural cell birth in the developing brain. In all adult tissues tested, *Rbm26* was seen to be expressed at a lower level than that seen in embryonic tissues.

Rbm26 was found to encode the murine homolog of the human cutaneous T-cell lymphoma tumor antigen Se70-2. Work to characterize mSe70-2 was undertaken first by development of antibodies for its detection. The developed antibodies were found to recognize the peptides they were raised with, but the native protein was never detected. It remains unclear if this observation is due to a lack of translation of *Rbm26* mRNA or if the antibodies are not appropriate for detection of the full size mSe70-2. This difficulty was further compounded as efforts to produce mSe70-2 in a variety of systems neglected to yield any protein. It was however determined that mRNA was produced from tagged expression constructs, and that tagged mSe70-2 does not appear to be degraded by the proteasome. The possibility exists that mSe70-2 production is highly regulated by the

cell at the translational level. This may be supported by the discovery of Se70-2 as a tumor antigen, suggesting that the immune system does not recognize the protein as 'self'. The group who discovered the human Se70-2 also have not been able to express the protein (Dr. S. Eichmüeller, pers. comm.).

Other analyses of mSe70-2 independent of the antibodies or *in vivo* expressed protein were pursued. Early in the project *in silico* analysis of mSe70-2 indicated that it contained two conserved nucleic acid binding domains: a CCCH-type zinc finger, and an RNA recognition motif (RRM). The ability of *in vitro* generated mSe70-2 to bind RNA was established, and a binding preference for G rich sequences was seen. A loss of function RNAi study for the *C. elegans* homolog of mSe70-2, B0336.3, indicated that it may play a role in positive regulation of the cell cycle. Existence of *Rbm27*, which encodes Rbm27 (also known as Psc1) and is a paralog of *Rbm26*, was noted and phylogenetic analysis of mSe70-2 homologs in eukaryotes was carried out. This analysis demonstrated the deep conservation of mSe70-2, and gene duplication in the vertebrate lineage giving rise to a mSe70-2 and a Rbm27 clade. Rbm27 is an SR family splicing factor, showing all of the defining features of the family. With its high level of similarity to Rbm27, mSe70-2 also shows all of the defining sequence features of the SR family. These results establish with confidence that mSe70-2 is a developmentally regulated RNA binding protein (RBP), and it is highly probable that it functions as a splicing factor.

7.2 RNA binding proteins in cell fate determination

RNA binding proteins (RBPs) play important roles in cellular maintenance, differentiation, and tissue development. They may exert their functions using a variety of methods including regulation of pre-mRNA splicing and translational silencing. Alternative splicing is of particular prevalence in the nervous and immune systems, where it has roles in precise control of cellular differentiation and specified function[203], while translational silencing is becoming more recognized as a conserved mechanism for regulation of stem cell fate (reviewed by Wong *et al., 2005*[302]). The roles of several

RBPs, SRp38, Musashi1 (Msh1), and the Nova and Hu protein families, will be briefly surveyed here.

7.2.1 SRp38

Primary neuronal differentiation is mediated by a pro-neural cascade starting with *neurogenin* which then activates a series of bHLH transcription factors including *NeuroD*[62,87,303]. Along with early induction of this cascade, individual cells are specified to become neurons from a field of equivalent cells in the process of lateral inhibition, which is mediated by levels of Delta-Notch signalling[304]. Cells with significant levels of signalling through Notch remain undifferentiated contributing to the pool of self renewing stem or progenitor cells.

SRp38 is a SR family splicing factor which was identified as having a regulatory role in neurogenesis, and is induced by NeuroD[305]. SRp38 has also been identified in a screen for SR proteins in a neural specific cDNA library, and in yeast two-hybrid screens for interactions with other splicing regulators[306-308]. It is an atypical SR protein as it is unable to activate splicing; rather, SRp38 is a repressor necessary for both mitotic and stress related pre-mRNA splicing inhibition[307,309]. SRp38 has been shown to function in a negative feedback loop, as its ectopic expression leads to repression of *neurogenin*, *NeuroD*, and other genes in the pro-neural cascade, hence inhibiting neural differentiation[310]. Analogous effects on the differentiation of other cell types including neural crest, mesodermal, and endodermal were also noted in the same study. In each case, early markers of different tissue types were expressed, indicating that the cells retained competence to differentiate, but later markers of differentiation were not expressed; therefore, SRp38 can inhibit neural differentiation but does not effect neural induction[310]. This effect is expected to be carried out via induction of Notch signalling, as direct targets of the Notch pathway are induced in response to expression of SRp38, however, the immediate pre-mRNA targets of SRp38 remain unclear.

7.2.2 Musashi 1

Another conserved RBP with a role in cell fate determination is Musashi-1 (Msi1), which is strongly expressed in the nervous system and was initially identified in

Drosophila[311,312]. Mammalian Msi1 has been found to be expressed extensively in neural stem/progenitor cell populations, and was also initially shown to have preference for interaction with Poly-G RNA as demonstrated by homoribopolymer binding assays[313]. Expression of Msi1 is not seen in post-mitotic terminally differentiated cells, pointing to its role in stem/progenitor cell self renewal[314]. Its role is to promote proliferation of, undifferentiated cells, and it fulfills this role via enhancing activation of Notch signalling. Msi1 enhances Notch signalling by binding to and repressing translation of the Notch intracellular domain (ICD) antagonist Numb in a sequence specific manner[315]. The effect of Msi1 on positive regulation of the Notch pathway was confirmed, as transcription of the Notch ICD pro-proliferation target *Hes1* is dependent on the RNA binding abilities of Msi1[315].

A more recently identified target of Msi1 is $p21^{WAF-1}$, a cyclin dependent kinase inhibitor essential for the mitotic arrest required for terminal differentiation of cells[316]. Msi1 overexpression was found to enhance proliferation of neurally differentiating P19 cells, and conversely the ability of P19 cells to differentiate was diminished when Msi1 function was abrogated[317]. This effect was demonstrated to be mediated via the ability of Msi1 to bind $p21^{WAF-1}$ mRNA, repressing its translation. Notably, $p21^{WAF-1}$ is also regulated by Hes1, suggesting that Msi1 can influence the regulation of $p21^{WAF-1}$ on both the mRNA and protein level, thus ensuring positive regulation of cell cycle progression[317,318].

Msi1 expression is not limited to neural stem/progenitor cell populations. It has also been found to be expressed in germline, intestinal and stomach epithelial, mammary, and mesenchymal stem/precursor cells[319-322]. In mesenchymal stem cells, which can differentiate into neuronal or mesenchymal cell types such as bone, cartilage, and fat, Msi1 expression is increased when differentiation cues are present[38,319,323]. Similarly, Msi1 is found to be up regulated in response to injury of the stomach surface mucosal cells, leading to rapid proliferation of stem/precursor cells and their recruitment to injury sites for rapid regeneration of the mucosa[321]. On a more generalized scale, loss of Msi1 function leads to premature differentiation of stem cell populations[322]. Therefore Msi1 is

a cell intrinsic signal required for self renewal and stem cell identity during development and tissue repair.

7.2.3 Hu proteins

Two families of neural specific RBPs, the Nova family and the Hu family, were first discovered as paraneoplastic neuronal degeneration (PND) antigens. Paraneoplastic syndromes are dysfunctions of organs or tissues caused by cancer, but not via invasion of the primary tumor or metastases (Reviewed by Posner, 2003[324]). Paraneoplastic neuronal degeneration is an autoimmune response against normal neurons coupled with naturally occurring anti-tumor immunity, most commonly associated with small cell lung, breast, or ovarian cancer (reviewed by Albert *et al.,* 2004 and Roberts *et al.,* 2004[325,326]).

The Hu family of proteins show similarity to the *Drosophila* ELAV protein, which is a RBP required for neuronal development and maintenance[327]. Similarly, the neuronal Hu family proteins, including HuB, C, and D, are involved in neuronal cell fate decisions. The family members are expressed in a hierarchical fashion during neural differentiation, with HuB expressed in early post mitotic neurons, HuC expressed in mature neurons, and HuD expression overlapping the others and occurring primarily in young migrating neurons[328]. A number of putative targets of the RNA binding ability of the Hu proteins have been proposed, including $p21^{WAF-1}$ and GAP-43, a protein with critical roles in axonal growth and regeneration, although the mechanism by which the Hu proteins exert their influence on these mRNAs is not entirely clear[329,330]. It is expected that the Hu proteins interact with AU rich elements in the 3' UTRs of their targets. What is known is that Hu proteins induce neuronal differentiation in several cell lines when overexpressed[331,332]. The multiple functions of HuD in this process have more recently been elucidated. Based on loss of function studies in mice, HuD appears to negatively regulate neural stem cell proliferation and encourage neural progenitor cells to withdraw from the cell cycle, in addition to promoting differentiation of post-mitotic cells[333].

7.2.4 Nova proteins

The Nova (Neuro oncological ventral antigen) proteins, Nova1 and Nova2, are neuron specific splicing factors[334]. The better characterized of the Nova family, Nova1,

is known to bind pre-mRNAs in a sequence specific manner[335]. Known targets of Nova1 binding include the pre-mRNAs transcribed from the $GABA_A\gamma2$ and GlyRα2 receptor genes[336]. Both of these neurotransmitter receptors are expressed throughout the brain, but require regional alternative splicing[337,338]. The regional specific splicing of $GABA_A\gamma2$ and GlyRα2, which is necessary for the viability of the neurons in question, has been shown to be mediated by Nova1[336]. More recently it has been shown using custom microarrays with extensive exon junction probe sets that Nova2 targets numerous transcripts encoding components involved in synaptic functioning[339]. With Nova family proteins targeting gene products with functions at the synapse, it is expected that Nova-regulated alternative splicing plays a significant role in synaptic plasticity; that is, the ability of synapses to modulate their activity in response to neural activity (reviewed by Ule and Darnell, 2006[340]).

Nova1 has also been shown to regulate its own alternative splicing[341]. In the same study it was shown that Nova1 is bi-functional, having the ability to both enhance or repress inclusion of alternate exons in spliced transcripts. The activity of Nova1 is dependent on the position of its sequence specific binding. Binding within an exon leads to skipping of that exon, while binding within an intron leads to inclusion of upstream alternatively spliced exons[341]. Nova proteins do not only modify cassette exon splicing: In the microarray study it was noted that Nova may also function in specific splice site selection, likely acting in a combinatorial fashion with an array of other interacting proteins[339].

Overall, it is well documented that these and a range of other RBPs function in cell fate determination, cellular maintenance, and tissue development. As was postulated in surveys of the transcriptome, alternative splicing contributes immensely to the complexity of the products produced from the genome. This complexity however is not limited to the numbers and variety of products made, but extends to the functions of the resulting products, their regulation, and regulation of a range of cellular processes.

7.3 The function and regulation of mSe70-2: speculation and questions.

7.3.1 Proposed molecular function of mSe70-2

The results yielded by this project have led to the establishment that mSe70-2 is a developmentally regulated RNA binding protein (RBP), and it is highly probable that it functions as a splicing factor, although this has not been demonstrated experimentally. The gene encoding mSe70-2, *Rbm26*, was initially isolated in a subtractive screen from ES cells and differentiating ES cell cultures enriched with neural precursors. It is not surprising to find a putative splicing factor in the RED dataset, as alternative splicing is known to be highly prevalent in the brain and its development. Even if mSe70-2 does not function as a splicing factor, it would still be of great interest to explore what RNA sequences it preferentially interacts with. Like mSe70-2, Msi1 was initially identified as having binding preference for G-rich RNA sequences. To more accurately identify its target sequences, the SELEX method was used (described in Okano *et. al.* 2005[342]). Briefly, a random pool of RNAs flanked with known sequences is produced, and RNAs within that pool which can be bound by the protein of interest are isolated using affinity chromatography. After reverse transcription and sequencing, a consensus sequence for the preferred binding sites of the protein of interest is identified. Having the preferred RNA binding sequence for mSe70-2 would allow for screening of the genome sequence for its potential targets. Identification and follow-up of potential mSe70-2 targets would surely assist in the elucidation of both the molecular and biological functions of mSe70-2. This proposed work is dependent on establishment of tools such as effective protein expression methods and suitable antibodies. Pursuing such tools is critical to any further work on mSe70-2.

7.3.2 Proposed biological function of mSe70-2

The work presented in this thesis does not definitively indicate a biological role for mSe70-2, however, the results do allow for speculation along such lines. Given the prevalence of RBPs in cell fate determination, and the discovery of *Rbm26* in uncommitted cells, it appears that mSe70-2 may play a role in regulation of cell fate. Whether mSe70-2 is pro-differentiation or pro-proliferation is unclear, but its apparent

biological role in *C. elegans*, where it may be a positive regulator of the cell cycle, suggests that it may in fact be pro-proliferation. This pro-proliferation biological function may be of particular importance during cell differentiation towards a sub-set of possible cell fates such as neural. The expression pattern of *Rbm26* during ES cell neural differentiation may lend support to this hypothesis. A basal level of *Rbm26* expression was observed in pluripotent ES cells, but upon induction of a neural differentiation program *Rbm26* expression was seen to increase, as indicated by the expression level observed in 4-/3+ neural precursor enriched cultures. In the spontaneously differentiating 4-/3- control cultures, an increase in *Rbm26* expression was not seen. These control cultures, though mixed and likely containing a small proportion of neural precursors, tend to contain a significant proportion of cardiac myocytes, suggesting that *Rbm26* plays a role in neural, but not all, cell fate decisions.

The timing of *Rbm26* expression in developing embryos may also support the argument that it is important in cell proliferation coupled with differentiation. *Rbm26* expression was found to be up-regulated during the time period corresponding to a burst of neuronal birth and differentiation. Although *Rmb26* may not be a pro-neural gene *per se*, this timing may again suggest a role for *Rbm26* in the decision of an uncommitted cell to proliferate, hence maintaining the stem/progenitor cell population, or differentiate into a neural cell type. This does not necessarily indicate that the biological function of *Rbm26* is limited to cells with competence to differentiate into neural cell types. As indicated by the *in situ* hybridization results, expression of *Rbm26* is not limited to the nervous system of the developing embryo as expression was widely seen in mesenchymal cell types as well. It should be noted though that mesenchymal stem/progenitor cells do have the capacity to differentiate into neural cell types[38,323]. Interestingly expression of *Rbm26* appeared to be essentially excluded from the embryonic liver and heart. The reasons for this observation are unknown, but it may be that the cells contributing to these tissues at 12.5 dpc are restricted enough in their respective cell lineages that *Rbm26* expression is no longer required. Alternatively, it may be that the program of differentiation into hepatic or cardiac cell types does not require *Rbm26* function at all,

with another gene filling its putative role in maintaining the balance between proliferation and differentiation signals.

In what could appear as a paradoxical twist, *Rbm26* expresion was also observed at lower levels in all adult tissues tested. It is now widely accepted that many and possible most adult tissues house stem and/or uncommitted progenitor cell populations. It would be interesting to explore if the observed *Rbm26* expression could be attributed to these cell populations and not to terminally differentiated cells. In order to address whether *Rbm26* expression is limited to stem/progenitor cell populations, it would be interesting to examine if its expression persists at 5 dpi (days post induction) in the Bain ES cell neural induction protocol. At 5 dpi, the maximal number of terminally differentiated neural cells is expected, and *Rbm26* expression would be expected to be diminished or absent if its expression is limited to stem/progenitor cells.

To summarize, a biological role for *Rbm26* in cell fate decisions, analogous to the roles of SRp38 and Msi1, though speculative, may be proposed based on work presented in this thesis. *Rbm26* expression may be limited to stem cell and/or uncommitted progenitor cell populations where it may function in maintenance of the balance between cellular proliferation and differentiation. Like SRp38 and Msi1, *Rbm26* expression may be increased in response to specific developmental and/or damage cues, such as pro-neural gene expression, in order to ensure an adequate pool of self renewing stem or progenitor cells is maintained.

7.3.3 Proposed regulation of mSe70-2

The inability to express or detect mSe70-2 was a major stumbling block of this project, as antibodies and recombinant proteins are critical tools required for a vast array of analyses. Several observations during the course of the project call the regulation of mSe70-2 into question: First, the antibodies raised against mSe70-2 work in principle, as they recognise the peptides used in their production, but do not apparently detect the full length protein although *Rbm26* transcripts are easily detected; second, tagged mSe70-2 does not appear to be produced from expression constructs despite detection of transcripts from the given constructs; finally, the human Se70-2 was initially identified as a tumor

antigen, which gives rise to the possibility that the immune system had not been tolerized to the protein. These observations beg the question of whether mSe70-2 is normally produced or not, even though transcripts for its production are ever present. It may be proposed that *Rbm26* transcripts are normally translationally repressed, and only under certain conditions, such as bursts of differentiation during development or in response to regeneration cues in the adult, is mSe70-2 produced. It is notable that in regulation of stem/progenitor populations, translational repression appears to be a common, conserved mechanism (reviewed by Wong *et al., 2005*[302]). It is also possible that mSe70-2 is only produced at immune privileged sites in the adult, such as is the case with Hu and Nova family proteins, and translationally repressed in other tissues. A wider range of cell types and developmental stages, such as 12.5 dpc embryos, adult brain, and neural precursor enriched 4-/3+ cultures, must be explored for expression of mSe70-2. Development of positive controls for mSe70-2 expression, is crucial to these experiments

If translation of mSe70-2 is typically repressed, then the mechanism of silencing of the *Rbm26* transcripts must be considered, though this project has not yielded any results supporting such conjecture. Translational silencing may be mediated by proteins components of the mRNP. This silencing mechanism though is accepted as a response to stress conditions or mitotic progression and involves sequestration of the repressed mRNAs into transient stress granules (reviewed by Moore, 2005[227]). It appears however that mSe70-2 is constitutively silenced. Recently it has been shown that translational repression by miRNA can be constitutive and also reversible in response to appropriate cues[237]. In order to examine if *Rbm26* transcripts are targeted by miRNA, a computational approach may be taken examining similarities between known and predicted miRNAs and *Rbm26*. Such a method has previously been applied to the *C. elegans* genome, but no miRNA targeting the *Rbm26* homolog was identified[343]. This does not eliminate the possibility of *Rbm26* regulation by miRNA, as in silico methods for identification of miRNAs and their targets are not robust (reviewed by Brown and Sanseau, 2005[344]). Indeed, there is some evidence, albeit weak, for transcription from the opposite strand of the *Rbm26* genomic locus presented in Ensembl (data not shown).

Also, the possibility that *Rbm26* transcripts are subject to transcriptional repression mediated by proteins should not be dismissed. It would be of interest to explore protein interactions with the transcripts, as it may represent a new mechanism for constitutive gene regulation at the translational level.

7.4 Concluding remarks

> "Science is not the way to find the answers to all questions.
> Science is a way to find better questions." – C. Barrans

The project presented in this thesis has certainly resulted in the development of more questions than it has answered. The initial project objectives of implementing a new data enrichment, management, and dissemination strategy for the Bain-Rancourt EST set, as well as identification and characterization of a candidate gene likely to encode a nucleic acid binding protein with probable roles in neural differentiation have been met. The development and implementation of RED is significant in that it presents a model for scalable, user friendly systems for use in sequencing projects. Although systems of its type are now more ubiquitous, it was novel at the time of its publication, and received significant attention when presented at meetings. It was clear that RED met a need that was not being met, not only in the Rancourt lab, but in others. The characterization of mSe70-2 is significant in that it will further illustrate the role for RNA binding proteins in development and maintenance of stem cell populations, and will also reveal new insights into novel gene regulation mechanisms. Central to furthering knowledge of mSe70-2 will be development of reliable tools for its study, including antibodies and recombinant proteins. Without these tools, it is likely that mSe70-2 and its function will remain elusive. Many opportunities exist for the continuation of this project, and mSe70-2 may reveal itself to be a critical regulator of stem cell fate, which may eventually be leveraged to bring the field of regenerative medicine closer to more widespread clinical application.

References

1. Brustle, O. et al. In vitro-generated neural precursors participate in mammalian brain development. *Proc Natl Acad Sci U S A* **94**, 14809-14 (1997).
2. Zipori, D. The stem state: plasticity is essential, whereas self-renewal and hierarchy are optional. *Stem Cells* **23**, 719-26 (2005).
3. Smith, A. G. et al. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688-90 (1988).
4. Williams, R. L. et al. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**, 684-7 (1988).
5. Martin, G. R. & Evans, M. J. Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro. *Proc Natl Acad Sci U S A* **72**, 1441-5 (1975).
6. Martin, G. R. Teratocarcinomas as a model system for the study of embryogenesis and neoplasia. *Cell* **5**, 229-43 (1975).
7. Papaioannou, V. E., McBurney, M. W., Gardner, R. L. & Evans, M. J. Fate of teratocarcinoma cells injected into early mouse embryos. *Nature* **258**, 70-73 (1975).
8. Martin, G. R. 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**, 7634-8 (1981).
9. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-6 (1981).
10. Hatoya, S. et al. Isolation and characterization of embryonic stem-like cells from canine blastocysts. *Mol Reprod Dev* **73**, 298-305 (2006).
11. Saito, S. et al. Generation of cloned calves and transgenic chimeric embryos from bovine embryonic stem-like cells. *Biochem Biophys Res Commun* **309**, 104-13 (2003).
12. Saito, S. et al. Isolation of embryonic stem-like cells from equine blastocysts and their differentiation in vitro. *FEBS Lett* **531**, 389-96 (2002).
13. Li, M. et al. Isolation and culture of pluripotent cells from in vitro produced porcine embryos. *Zygote* **12**, 43-8 (2004).
14. Ouhibi, N. et al. Initial culture behaviour of rat blastocysts on selected feeder cell lines. *Mol Reprod Dev* **40**, 311-24 (1995).
15. Thomson, J. A. et al. Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145-7 (1998).
16. Thomson, J. A. et al. Isolation of a primate embryonic stem cell line. *Proc Natl Acad Sci U S A* **92**, 7844-8 (1995).
17. Solter, D. From teratocarcinomas to embryonic stem cells and beyond: a history of embryonic stem cell research. *Nat Rev Genet* **7**, 319-27 (2006).
18. Scadden, D. T. The stem-cell niche as an entity of action. *Nature* **441**, 1075-9 (2006).

19. Wagers, A. J. & Weissman, I. L. Plasticity of adult stem cells. *Cell* **116**, 639-48 (2004).
20. McCulloch, E. A. & Till, J. E. Perspectives on the properties of stem cells. *Nat Med* **11**, 1026-8 (2005).
21. Blanpain, C., Lowry, W. E., Geoghegan, A., Polak, L. & Fuchs, E. Self-renewal, multipotency, and the existence of two cell populations within an epithelial stem cell niche. *Cell* **118**, 635-48 (2004).
22. Stemple, D. L. & Anderson, D. J. Isolation of a stem cell for neurons and glia from the mammalian neural crest. *Cell* **71**, 973-85 (1992).
23. Rietze, R. L. et al. Purification of a pluripotent neural stem cell from the adult mouse brain. *Nature* **412**, 736-9 (2001).
24. Seaberg, R. M. et al. Clonal identification of multipotent precursors from adult mouse pancreas that generate neural and pancreatic lineages. *Nat Biotechnol* **22**, 1115-24 (2004).
25. Reynolds, B. A. & Weiss, S. Clonal and population analyses demonstrate that an EGF-responsive mammalian embryonic CNS precursor is a stem cell. *Dev Biol* **175**, 1-13 (1996).
26. Beltrami, A. P. et al. Adult cardiac stem cells are multipotent and support myocardial regeneration. *Cell* **114**, 763-76 (2003).
27. Oh, S. H., Hatch, H. M. & Petersen, B. E. Hepatic oval 'stem' cell in liver regeneration. *Semin Cell Dev Biol* **13**, 405-9 (2002).
28. Toma, J. G. et al. Isolation of multipotent adult stem cells from the dermis of mammalian skin. *Nat Cell Biol* **3**, 778-84 (2001).
29. Alonso, L. & Fuchs, E. Stem cells of the skin epithelium. *Proc Natl Acad Sci U S A* **100 Suppl 1**, 11830-5 (2003).
30. Spangrude, G. J., Heimfeld, S. & Weissman, I. L. Purification and characterization of mouse hematopoietic stem cells. *Science* **241**, 58-62 (1988).
31. Ohlstein, B., Kai, T., Decotto, E. & Spradling, A. The stem cell niche: theme and variations. *Curr Opin Cell Biol* **16**, 693-9 (2004).
32. Pittenger, M. F. et al. Multilineage potential of adult human mesenchymal stem cells. *Science* **284**, 143-7 (1999).
33. Gronthos, S. et al. Surface protein characterization of human adipose tissue-derived stromal cells. *J Cell Physiol* **189**, 54-63 (2001).
34. Jankowski, R. J., Deasy, B. M. & Huard, J. Muscle-derived stem cells. *Gene Ther* **9**, 642-7 (2002).
35. Miura, M. et al. SHED: stem cells from human exfoliated deciduous teeth. *Proc Natl Acad Sci U S A* **100**, 5807-12 (2003).
36. Kuznetsov, S. A. et al. Circulating skeletal stem cells. *J Cell Biol* **153**, 1133-40 (2001).
37. Rosada, C., Justesen, J., Melsvik, D., Ebbesen, P. & Kassem, M. The human umbilical cord blood: a potential source for osteoblast progenitor cells. *Calcif Tissue Int* **72**, 135-42 (2003).
38. Krabbe, C., Zimmer, J. & Meyer, M. Neural transdifferentiation of mesenchymal stem cells--a critical review. *Apmis* **113**, 831-44 (2005).

39. Sulston, J. E. Caenorhabditis elegans: the cell lineage and beyond (Nobel lecture). *Chembiochem* **4**, 688-96 (2003).
40. Hoffenberg, R. Brenner, the worm and the prize. *Clin Med* **3**, 285-6 (2003).
41. Spangrude, G. J., Brooks, D. M. & Tumas, D. B. Long-term repopulation of irradiated mice with limiting numbers of purified hematopoietic stem cells: in vivo expansion of stem cell phenotype but not function. *Blood* **85**, 1006-16 (1995).
42. Osawa, M., Hanada, K., Hamada, H. & Nakauchi, H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science* **273**, 242-5 (1996).
43. Goldsby, R. A. *Immunology* (W.H. Freeman, New York, 2003).
44. Garrell, J. & Campuzano, S. The helix-loop-helix domain: a common motif for bristles, muscles and sex. *Bioessays* **13**, 493-8 (1991).
45. Murre, C. et al. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **58**, 537-44 (1989).
46. Ellenberger, T., Fass, D., Arnaud, M. & Harrison, S. C. Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes Dev* **8**, 970-80 (1994).
47. Ma, P. C., Rould, M. A., Weintraub, H. & Pabo, C. O. Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell* **77**, 451-9 (1994).
48. Sasai, Y., Kageyama, R., Tagawa, Y., Shigemoto, R. & Nakanishi, S. Two mammalian helix-loop-helix factors structurally related to Drosophila hairy and Enhancer of split. *Genes Dev* **6**, 2620-34 (1992).
49. Novitch, B. G., Chen, A. I. & Jessell, T. M. Coordinate regulation of motor neuron subtype identity and pan-neuronal properties by the bHLH repressor Olig2. *Neuron* **31**, 773-89 (2001).
50. Ross, S. E., Greenberg, M. E. & Stiles, C. D. Basic helix-loop-helix factors in cortical development. *Neuron* **39**, 13-25 (2003).
51. Green, A. R. & Begley, C. G. SCL and related hemopoietic helix-loop-helix transcription factors. *Int J Cell Cloning* **10**, 269-76 (1992).
52. Murre, C. Helix-loop-helix proteins and lymphocyte development. *Nat Immunol* **6**, 1079-86 (2005).
53. Desprez, P. Y., Sumida, T. & Coppe, J. P. Helix-loop-helix proteins in mammary gland development and breast cancer. *J Mammary Gland Biol Neoplasia* **8**, 225-39 (2003).
54. Coppe, J. P., Smith, A. P. & Desprez, P. Y. Id proteins in epithelial cells. *Exp Cell Res* **285**, 131-45 (2003).
55. Dias, P., Dilling, M. & Houghton, P. The molecular basis of skeletal muscle differentiation. *Semin Diagn Pathol* **11**, 3-14 (1994).
56. Kazhdan, I., Rickard, D. & Leboy, P. S. HLH transcription factor activity in osteogenic cells. *J Cell Biochem* **65**, 1-10 (1997).

57. Weintraub, H. The MyoD family and myogenesis: redundancy, networks, and thresholds. *Cell* **75**, 1241-4 (1993).
58. Rudnicki, M. A. et al. MyoD or Myf-5 is required for the formation of skeletal muscle. *Cell* **75**, 1351-9 (1993).
59. Hasty, P. et al. Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene. *Nature* **364**, 501-6 (1993).
60. Louvi, A. & Artavanis-Tsakonas, S. Notch signalling in vertebrate neural development. *Nat Rev Neurosci* **7**, 93-102 (2006).
61. Chitnis, A. & Kintner, C. Sensitivity of proneural genes to lateral inhibition affects the pattern of primary neurons in Xenopus embryos. *Development* **122**, 2295-301 (1996).
62. Ma, Q., Kintner, C. & Anderson, D. J. Identification of neurogenin, a vertebrate neuronal determination gene. *Cell* **87**, 43-52 (1996).
63. Lewis, J. Notch signalling and the control of cell fate choices in vertebrates. *Semin Cell Dev Biol* **9**, 583-9 (1998).
64. Jarriault, S. et al. Delta-1 activation of notch-1 signaling results in HES-1 transactivation. *Mol Cell Biol* **18**, 7423-31 (1998).
65. Ohtsuka, T. et al. Hes1 and Hes5 as notch effectors in mammalian neuronal differentiation. *Embo J* **18**, 2196-207 (1999).
66. Kageyama, R. & Ohtsuka, T. The Notch-Hes pathway in mammalian neural development. *Cell Res* **9**, 179-88 (1999).
67. Henrique, D. et al. Maintenance of neuroepithelial progenitor cells by Delta-Notch signalling in the embryonic chick retina. *Curr Biol* **7**, 661-70 (1997).
68. Hitoshi, S. et al. Notch pathway molecules are essential for the maintenance, but not the generation, of mammalian neural stem cells. *Genes Dev* **16**, 846-58 (2002).
69. Wilson, A. & Radtke, F. Multiple functions of Notch signaling in self-renewing organs and cancer. *FEBS Lett* **580**, 2860-8 (2006).
70. Luo, D., Renault, V. M. & Rando, T. A. The regulation of Notch signaling in muscle stem cell activation and postnatal myogenesis. *Semin Cell Dev Biol* **16**, 612-22 (2005).
71. Nuttall, M. E. & Gimble, J. M. Controlling the balance between osteoblastogenesis and adipogenesis and the consequent therapeutic implications. *Curr Opin Pharmacol* **4**, 290-4 (2004).
72. Ahuja, V., Dieckgraefe, B. K. & Anant, S. Molecular biology of the small intestine. *Curr Opin Gastroenterol* **22**, 90-4 (2006).
73. Sanes, D. H., Reh, T. A. & Harris, W. A. *Development of the nervous system* (Academic, San Diego, Calif. ; London, 2000).
74. Hemmati-Brivanlou, A., Kelly, O. G. & Melton, D. A. Follistatin, an antagonist of activin, is expressed in the Spemann organizer and displays direct neuralizing activity. *Cell* **77**, 283-95 (1994).
75. Lamb, T. M. et al. Neural induction by the secreted polypeptide noggin. *Science* **262**, 713-8 (1993).

76. Sasai, Y., Lu, B., Steinbeisser, H. & De Robertis, E. M. Regulation of neural induction by the Chd and Bmp-4 antagonistic patterning signals in Xenopus. *Nature* **376**, 333-6 (1995).
77. Smith, W. C. & Harland, R. M. Expression cloning of noggin, a new dorsalizing factor localized to the Spemann organizer in Xenopus embryos. *Cell* **70**, 829-40 (1992).
78. Bouwmeester, T., Kim, S., Sasai, Y., Lu, B. & De Robertis, E. M. Cerberus is a head-inducing secreted factor expressed in the anterior endoderm of Spemann's organizer. *Nature* **382**, 595-601 (1996).
79. Grunz, H. & Tacke, L. Neural differentiation of Xenopus laevis ectoderm takes place after disaggregation and delayed reaggregation without inducer. *Cell Differ Dev* **28**, 211-7 (1989).
80. Sato, S. M. & Sargent, T. D. Development of neural inducing capacity in dissociated Xenopus embryos. *Dev Biol* **134**, 263-6 (1989).
81. Wharton, K. A., Ray, R. P. & Gelbart, W. M. An activity gradient of decapentaplegic is necessary for the specification of dorsal pattern elements in the Drosophila embryo. *Development* **117**, 807-22 (1993).
82. Wilson, P. A. & Hemmati-Brivanlou, A. Induction of epidermis and inhibition of neural fate by Bmp-4. *Nature* **376**, 331-3 (1995).
83. Munoz-Sanjuan, I. & Brivanlou, A. H. Neural induction, the default model and embryonic stem cells. *Nat Rev Neurosci* **3**, 271-80 (2002).
84. Lee, J. E. Basic helix-loop-helix genes in neural development. *Curr Opin Neurobiol* **7**, 13-20 (1997).
85. Ishibashi, M. et al. Targeted disruption of mammalian hairy and Enhancer of split homolog-1 (HES-1) leads to up-regulation of neural helix-loop-helix factors, premature neurogenesis, and severe neural tube defects. *Genes Dev* **9**, 3136-48 (1995).
86. Gowan, K. et al. Crossinhibitory activities of Ngn1 and Math1 allow specification of distinct dorsal interneurons. *Neuron* **31**, 219-32 (2001).
87. Kintner, C. Neurogenesis in embryos and in adult neural stem cells. *J Neurosci* **22**, 639-43 (2002).
88. Bertrand, N., Castro, D. S. & Guillemot, F. Proneural genes and the specification of neural cell types. *Nat Rev Neurosci* **3**, 517-30 (2002).
89. Maden, M. Retinoid signalling in the development of the central nervous system. *Nat Rev Neurosci* **3**, 843-53 (2002).
90. Maden, M. Heads or tails? Retinoic acid will decide. *Bioessays* **21**, 809-12 (1999).
91. Kastner, P., Chambon, P. & Leid, M. in *Vitamin A in health and disease* (ed. Blomhoff, R.) xviii, 677 p. (M. Dekker, New York, 1994).
92. Kliewer, S. A., Umesono, K., Evans, R. M. & Mangelsdorf, D. J. in *Vitamin A in health and disease* (ed. Blomhoff, R.) xviii, 677 p. (M. Dekker, New York, 1994).
93. Hoffmann, B. et al. A retinoic acid receptor-specific element controls the retinoic acid receptor-beta promoter. *Mol Endocrinol* **4**, 1727-36 (1990).

94. Jacobson, M. & Huang, S. Neurite outgrowth traced by means of horseradish peroxidase inherited from neuronal ancestral cells in frog embryos. *Dev Biol* **110**, 102-13 (1985).
95. Hartenstein, V. Early neurogenesis in Xenopus: the spatio-temporal pattern of proliferation and cell lineages in the embryonic spinal cord. *Neuron* **3**, 399-411 (1989).
96. Sharpe, C. & Goldstone, K. The control of Xenopus embryonic primary neurogenesis is mediated by retinoid signalling in the neurectoderm. *Mech Dev* **91**, 69-80 (2000).
97. Sharpe, C. R. & Goldstone, K. Retinoid receptors promote primary neurogenesis in Xenopus. *Development* **124**, 515-23 (1997).
98. Sharpe, C. & Goldstone, K. Retinoid signalling acts during the gastrula stages to promote primary neurogenesis. *Int J Dev Biol* **44**, 463-70 (2000).
99. Franco, P. G., Paganelli, A. R., Lopez, S. L. & Carrasco, A. E. Functional association of retinoic acid and hedgehog signaling in Xenopus primary neurogenesis. *Development* **126**, 4257-65 (1999).
100. Arima, K. et al. Global analysis of RAR-responsive genes in the Xenopus neurula using cDNA microarrays. *Dev Dyn* **232**, 414-31 (2005).
101. Maden, M. Role and distribution of retinoic acid during CNS development. *Int Rev Cytol* **209**, 1-77 (2001).
102. Gavalas, A. ArRAnging the hindbrain. *Trends Neurosci* **25**, 61-4 (2002).
103. Gavalas, A. & Krumlauf, R. Retinoid signalling and hindbrain patterning. *Curr Opin Genet Dev* **10**, 380-6 (2000).
104. Irving, C. & Mason, I. Signalling by FGF8 from the isthmus patterns anterior hindbrain and establishes the anterior limit of Hox gene expression. *Development* **127**, 177-86 (2000).
105. White, J. C., Highland, M., Kaiser, M. & Clagett-Dame, M. Vitamin A deficiency results in the dose-dependent acquisition of anterior character and shortening of the caudal hindbrain of the rat embryo. *Dev Biol* **220**, 263-84 (2000).
106. Berggren, K., McCaffery, P., Drager, U. & Forehand, C. J. Differential distribution of retinoic acid synthesis in the chicken embryo as determined by immunolocalization of the retinoic acid synthetic enzyme, RALDH-2. *Dev Biol* **210**, 288-304 (1999).
107. Gale, E., Zile, M. & Maden, M. Hindbrain respecification in the retinoid-deficient quail. *Mech Dev* **89**, 43-54 (1999).
108. Wendling, O., Ghyselinck, N. B., Chambon, P. & Mark, M. Roles of retinoic acid receptors in early embryonic morphogenesis and hindbrain patterning. *Development* **128**, 2031-8 (2001).
109. Dupe, V. & Lumsden, A. Hindbrain patterning involves graded responses to retinoic acid signalling. *Development* **128**, 2199-208 (2001).
110. Dupe, V. et al. In vivo functional analysis of the Hoxa-1 3' retinoic acid response element (3'RARE). *Development* **124**, 399-410 (1997).

111. Mansergh, F. C., Wride, M. A. & Rancourt, D. E. Neurons from stem cells: implications for understanding nervous system development and repair. *Biochem Cell Biol* **78**, 613-28 (2000).
112. Dinsmore, J. et al. Embryonic stem cells differentiated in vitro as a novel source of cells for transplantation. *Cell Transplant* **5**, 131-43 (1996).
113. Risau, W. et al. Vasculogenesis and angiogenesis in embryonic-stem-cell-derived embryoid bodies. *Development* **102**, 471-8 (1988).
114. Keller, G., Kennedy, M., Papayannopoulou, T. & Wiles, M. V. Hematopoietic commitment during embryonic stem cell differentiation in culture. *Mol Cell Biol* **13**, 473-86 (1993).
115. zur Nieden, N. I., Kempka, G., Rancourt, D. E. & Ahr, H. J. Induction of chondro-, osteo- and adipogenesis in embryonic stem cells by bone morphogenetic protein-2: effect of cofactors on differentiating lineages. *BMC Dev Biol* **5**, 1 (2005).
116. Bain, G., Kitchens, D., Yao, M., Huettner, J. E. & Gottlieb, D. I. Embryonic stem cells express neuronal properties in vitro. *Dev Biol* **168**, 342-57 (1995).
117. Bain, G. & Gottlieb, D. I. Neural cells derived by in vitro differentiation of P19 and embryonic stem cells. *Perspect Dev Neurobiol* **5**, 175-8 (1998).
118. Schmidt-Kastner, P. K., Jardine, K., Cormier, M. & McBurney, M. W. Absence of p53-dependent cell cycle regulation in pluripotent mouse cell lines. *Oncogene* **16**, 3003-11 (1998).
119. McBurney, M. W. P19 embryonal carcinoma cells. *Int J Dev Biol* **37**, 135-40 (1993).
120. Jones-Villeneuve, E. M., Rudnicki, M. A., Harris, J. F. & McBurney, M. W. Retinoic acid-induced neural differentiation of embryonal carcinoma cells. *Mol Cell Biol* **3**, 2271-9 (1983).
121. Jones-Villeneuve, E. M., McBurney, M. W., Rogers, K. A. & Kalnins, V. I. Retinoic acid induces embryonal carcinoma cells to differentiate into neurons and glial cells. *J Cell Biol* **94**, 253-62 (1982).
122. McBurney, M. W., Jones-Villeneuve, E. M., Edwards, M. K. & Anderson, P. J. Control of muscle and neuronal differentiation in a cultured embryonal carcinoma cell line. *Nature* **299**, 165-7 (1982).
123. Bain, G., Ray, W. J., Yao, M. & Gottlieb, D. I. From embryonal carcinoma cells to neurons: the P19 pathway. *Bioessays* **16**, 343-8 (1994).
124. Herbert, M. C. & Graham, C. F. Cell determination and biochemical differentiation of the early mammalian embryo. *Curr Top Dev Biol* **8**, 151-78 (1974).
125. Levine, J. M. & Flynn, P. Cell surface changes accompanying the neural differentiation of an embryonal carcinoma cell line. *J Neurosci* **6**, 3374-84 (1986).
126. McBurney, M. W. et al. Differentiation and maturation of embryonal carcinoma-derived neurons in cell culture. *J Neurosci* **8**, 1063-73 (1988).
127. MacPherson, P. A., Jones, S., Pawson, P. A., Marshall, K. C. & McBurney, M. W. P19 cells differentiate into glutamatergic and glutamate-responsive neurons in vitro. *Neuroscience* **80**, 487-99 (1997).

128. Staines, W. A., Morassutti, D. J., Reuhl, K. R., Ally, A. I. & McBurney, M. W. Neurons derived from P19 embryonal carcinoma cells have varied morphologies and neurotransmitters. *Neuroscience* **58**, 735-51 (1994).
129. Choi, D. et al. In vitro differentiation of mouse embryonic stem cells: enrichment of endodermal cells in the embryoid body. *Stem Cells* **23**, 817-27 (2005).
130. Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W. & Kemler, R. The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *J Embryol Exp Morphol* **87**, 27-45 (1985).
131. Fraichard, A. et al. In vitro differentiation of embryonic stem cells into glial cells and functional neurons. *J Cell Sci* **108 (Pt 10)**, 3181-8 (1995).
132. Li, M., Pevny, L., Lovell-Badge, R. & Smith, A. Generation of purified neural precursors from embryonic stem cells by lineage selection. *Curr Biol* **8**, 971-4 (1998).
133. Bain, G., Ray, W. J., Yao, M. & Gottlieb, D. I. Retinoic acid promotes neural and represses mesodermal gene expression in mouse embryonic stem cells in culture. *Biochem Biophys Res Commun* **223**, 691-4 (1996).
134. Renoncourt, Y., Carroll, P., Filippi, P., Arce, V. & Alonso, S. Neurons derived in vitro from ES cells express homeoproteins characteristic of motoneurons and interneurons. *Mech Dev* **79**, 185-97 (1998).
135. Bain, G. et al. ES cell neural differentiation reveals a substantial number of novel ESTs. *Funct Integr Genomics* **1**, 127-39 (2000).
136. Mujtaba, T. et al. Lineage-restricted neural precursors can be isolated from both the mouse neural tube and cultured ES cells. *Dev Biol* **214**, 113-27 (1999).
137. McDonald, J. W. et al. Transplanted embryonic stem cells survive, differentiate and promote recovery in injured rat spinal cord. *Nat Med* **5**, 1410-2 (1999).
138. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular cloning : a laboratory manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1989).
139. Sambrook, J. & Russell, D. W. *Molecular cloning : a laboratory manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2001).
140. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403-10 (1990).
141. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-402 (1997).
142. McGinnis, S. & Madden, T. L. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res* **32**, W20-5 (2004).
143. Ning, Z., Cox, A. J. & Mullikin, J. C. SSAHA: a fast search method for large DNA databases. *Genome Res* **11**, 1725-9 (2001).
144. Birney, E. et al. Ensembl 2006. *Nucleic Acids Res* **34**, D556-61 (2006).
145. Gasteiger, E. et al. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* **31**, 3784-8 (2003).
146. Quevillon, E. et al. InterProScan: protein domains identifier. *Nucleic Acids Res* **33**, W116-20 (2005).

147. Staden, R. The Staden sequence analysis package. *Mol Biotechnol* **5**, 233-41 (1996).
148. Staden, R., Beal, K. F. & Bonfield, J. K. The Staden package, 1998. *Methods Mol Biol* **132**, 115-30 (2000).
149. Gaasterland, T. & Sensen, C. W. MAGPIE: automated genome interpretation. *Trends Genet* **12**, 76-8 (1996).
150. Gaasterland, T. & Sensen, C. W. Fully automated genome analysis that reflects user needs and preferences. A detailed introduction to the MAGPIE system architecture. *Biochimie* **78**, 302-10 (1996).
151. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J Mol Biol* **268**, 78-94 (1997).
152. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876-82 (1997).
153. Madison, D. & Madison, M. (Sinauer and Associates, Sunderland, Connecticut, 2000).
154. Strimmer, K. & von Haeseler, A. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc Natl Acad Sci U S A* **94**, 6815-9 (1997).
155. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-4 (2003).
156. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696-704 (2003).
157. Page, R. D. TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**, 357-8 (1996).
158. Bain, G., Yao, M., Huettner, J. E., Finley, M. F. A. & Gottlieb, D. I. in *Culturing nerve cells* (eds. Banker, G. & Goslin, K.) 666 p. (MIT Press, Cambridge, Mass., 1998).
159. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**, 365-86 (2000).
160. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-15 (2003).
161. Murphy, C. L. & Polak, J. M. Differentiating embryonic stem cells: GAPDH, but neither HPRT nor beta-tubulin is suitable as an internal standard for measuring RNA levels. *Tissue Eng* **8**, 551-9 (2002).
162. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-8 (2001).
163. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**, e45 (2001).
164. Pfaffl, M. W., Horgan, G. W. & Dempfle, L. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res* **30**, e36 (2002).

165. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**, 113-30 (1986).
166. Moffatt, B. A. & Studier, F. W. T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell* **49**, 221-7 (1987).
167. Novy, R., Drott, D., Yaeger, K. & Mierendorf, R. Overcoming the codon bias of E. coli for enhanced protein expression. *InNovations* **12**, 1-3 (2001).
168. Rigaut, G. et al. A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* **17**, 1030-2 (1999).
169. Parker, J. M., Guo, D. & Hodges, R. S. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* **25**, 5425-32 (1986).
170. Parker, J. M. & Hodges, R. S. Prediction of surface and interior regions in proteins--Part I: Linear tripeptide sequences identify structural boundaries in proteins. *Pept Res* **4**, 347-54 (1991).
171. Harlow, E. & Lane, D. *Antibodies : a laboratory manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1988).
172. Swerdlow, P. S., Finley, D. & Varshavsky, A. Enhancement of immunoblot sensitivity by heating of hydrated filters. *Anal Biochem* **156**, 147-53 (1986).
173. Caputi, M., Mayeda, A., Krainer, A. R. & Zahler, A. M. hnRNP A/B proteins are required for inhibition of HIV-1 pre-mRNA splicing. *Embo J* **18**, 4060-7 (1999).
174. Brenner, S. The genetics of Caenorhabditis elegans. *Genetics* **77**, 71-94 (1974).
175. Wood, W. B. *The Nematode Caenorhabditis elegans* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1988).
176. Fukushige, T., Goszczynski, B., Tian, H. & McGhee, J. D. The evolutionary duplication and probable demise of an endodermal GATA factor in Caenorhabditis elegans. *Genetics* **165**, 575-88 (2003).
177. Fukushige, T., Goszczynski, B., Yan, J. & McGhee, J. D. Transcriptional control and patterning of the pho-1 gene, an essential acid phosphatase expressed in the C. elegans intestine. *Dev Biol* **279**, 446-61 (2005).
178. Kamath, R. S. & Ahringer, J. Genome-wide RNAi screening in Caenorhabditis elegans. *Methods* **30**, 313-21 (2003).
179. Goszczynski, B. & McGhee, J. D. Reevaluation of the role of the med-1 and med-2 genes in specifying the Caenorhabditis elegans endoderm. *Genetics* **171**, 545-55 (2005).
180. Abramoff, W. S., Magelhaes, P. J. & Ram, S. J. Image Processing with ImageJ. *Biophotonics International* **11**, 36-42 (2004).
181. Stein, L. D. in *Bioinformatics: a practical guide to the analysis of genes and proteins* (eds. Baxevanis, A. D. & Ouellette, B. F. F.) (Wiley-Interscience, New York, 2001).
182. Valade, J. (Wiley, Hoboken, N.J., 2004).
183. Everitt, R. et al. RED: the analysis, management and dissemination of expressed sequence tags. *Bioinformatics* **18**, 1692-3 (2002).

184. Carninci, P. et al. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res* **10**, 1617-30 (2000).
185. Lazzari, B. et al. ESTree db: a tool for peach functional genomics. *BMC Bioinformatics* **6 Suppl 4**, S16 (2005).
186. Maheswari, U. et al. The Diatom EST Database. *Nucleic Acids Res* **33**, D344-7 (2005).
187. Hong, P. & Wong, W. H. GeneNotes--a novel information management software for biologists. *BMC Bioinformatics* **6**, 20 (2005).
188. Kumar, C. G. et al. ESTIMA, a tool for EST management in a multi-project environment. *BMC Bioinformatics* **5**, 176 (2004).
189. Gardette, R., Courtois, M. & Bisconte, J. C. Prenatal development of mouse central nervous structures: time of neuron origin and gradients of neuronal production. A radioautographic study. *J Hirnforsch* **23**, 415-31 (1982).
190. Chenn, A., Braisted, J. E., McConnell, S. K. & O'Leary, D. D. M. in *Molecular and cellular approaches to neural development* (eds. Cowan, W. M., Jessell, T. M. & Zipursky, S. L.) x, 563 p. (Oxford University Press, New York, 1997).
191. Lawson, S. N. & Biscoe, T. J. Development of mouse dorsal root ganglia: an autoradiographic and quantitative study. *J Neurocytol* **8**, 265-74 (1979).
192. Mural, R. J., Einstein, J. R., Guan, X., Mann, R. C. & Uberbacher, E. C. An artificial intelligence approach to DNA sequence feature recognition. *Trends Biotechnol* **10**, 66-9 (1992).
193. Uberbacher, E. C. & Mural, R. J. Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc Natl Acad Sci U S A* **88**, 11261-5 (1991).
194. Baxevanis, A. D. in *Bioinformatics: a practical guide to the analysis of genes and proteins* (eds. Baxevanis, A. D. & Ouellette, B. F. F.) xviii, 470 , [16] of plates (Wiley-Interscience, New York, 2001).
195. Rogic, S., Mackworth, A. K. & Ouellette, F. B. Evaluation of gene-finding programs on mammalian sequences. *Genome Res* **11**, 817-32 (2001).
196. Mathe, C., Sagot, M. F., Schiex, T. & Rouze, P. Current methods of gene prediction, their strengths and weaknesses. *Nucleic Acids Res* **30**, 4103-17 (2002).
197. Burge, C. B. & Karlin, S. Finding the genes in genomic DNA. *Curr Opin Struct Biol* **8**, 346-54 (1998).
198. Carninci, P. et al. The transcriptional landscape of the mammalian genome. *Science* **309**, 1559-63 (2005).
199. Waterston, R. H. et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-62 (2002).
200. Black, D. L. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* **72**, 291-336 (2003).
201. Burset, M., Seledtsov, I. A. & Solovyev, V. V. Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Res* **28**, 4364-75 (2000).
202. Ast, G. How did alternative splicing evolve? *Nat Rev Genet* **5**, 773-82 (2004).

203. Modrek, B., Resch, A., Grasso, C. & Lee, C. Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Res* **29**, 2850-9 (2001).
204. Ramakers, C., Ruijter, J. M., Deprez, R. H. & Moorman, A. F. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett* **339**, 62-6 (2003).
205. Gray, P. A. et al. Mouse brain organization revealed through direct genome-scale TF expression analysis. *Science* **306**, 2255-7 (2004).
206. McKee, A. E. et al. A genome-wide in situ hybridization map of RNA-binding proteins reveals anatomically restricted expression in the developing mouse brain. *BMC Dev Biol* **5**, 14 (2005).
207. Carlson, C. & Zhongting, H. in *Ambion Technical Bulletin #507*.
208. O'Sullivan, C. M. et al. Uterine secretion of ISP1 & 2 tryptases is regulated by progesterone and estrogen during pregnancy and the endometrial cycle. *Mol Reprod Dev* **69**, 252-9 (2004).
209. Naylor, M. J., Rancourt, D. E. & Bech-Hansen, N. T. Isolation and characterization of a calcium channel gene, Cacna1f, the murine orthologue of the gene for incomplete X-linked congenital stationary night blindness. *Genomics* **66**, 324-7 (2000).
210. Rancourt, S. L. & Rancourt, D. E. Murine subtilisin-like proteinase SPC6 is expressed during embryonic implantation, somitogenesis, and skeletal formation. *Dev Genet* **21**, 75-81 (1997).
211. Gilbert, S. F. & Singer, S. R. *Developmental biology* (Sinauer Associates, Sunderland, Mass., 2000).
212. Zaret, K. S. Regulatory phases of early liver development: paradigms of organogenesis. *Nat Rev Genet* **3**, 499-512 (2002).
213. Jung, J., Zheng, M., Goldfarb, M. & Zaret, K. S. Initiation of mammalian liver development from endoderm by fibroblast growth factors. *Science* **284**, 1998-2003 (1999).
214. Dottori, M., Gross, M. K., Labosky, P. & Goulding, M. The winged-helix transcription factor Foxd3 suppresses interneuron differentiation and promotes neural crest cell fate. *Development* **128**, 4127-38 (2001).
215. Sumazaki, R. et al. Conversion of biliary system to pancreatic tissue in Hes1-deficient mice. *Nat Genet* **36**, 83-7 (2004).
216. Cheung, M. et al. The transcriptional control of trunk neural crest induction, survival, and delamination. *Dev Cell* **8**, 179-92 (2005).
217. Kageyama, R., Ohtsuka, T., Hatakeyama, J. & Ohsawa, R. Roles of bHLH genes in neural stem cell differentiation. *Exp Cell Res* **306**, 343-8 (2005).
218. Eichmuller, S. et al. Serological detection of cutaneous T-cell lymphoma-associated antigens. *Proc Natl Acad Sci U S A* **98**, 629-34 (2001).
219. Gemmill, R. Cutaneous T-cell lymphoma. *Semin Oncol Nurs* **22**, 90-6 (2006).
220. Nestle, F. O. et al. Vaccination of melanoma patients with peptide- or tumor lysate-pulsed dendritic cells. *Nat Med* **4**, 328-32 (1998).

221. Tureci, O., Sahin, U. & Pfreundschuh, M. Serological analysis of human tumor antigens: molecular definition and implications. *Mol Med Today* **3**, 342-9 (1997).
222. Tureci, O. et al. Identification of a meiosis-specific protein as a member of the class of cancer/testis antigens. *Proc Natl Acad Sci U S A* **95**, 5211-6 (1998).
223. Chen, Y. T. Identification of human tumor antigens by serological expression cloning: an online review of SEREX. *Cancer Immunity* (2004).
224. Yao, M., Bain, G. & Gottlieb, D. I. Neuronal differentiation of P19 embryonal carcinoma cells in defined media. *J Neurosci Res* **41**, 792-804 (1995).
225. Mattick, J. S. The functional genomics of noncoding RNA. *Science* **309**, 1527-8 (2005).
226. Costa, F. F. Non-coding RNAs: new players in eukaryotic biology. *Gene* **357**, 83-94 (2005).
227. Moore, M. J. From birth to death: the complex lives of eukaryotic mRNAs. *Science* **309**, 1514-8 (2005).
228. Mazan-Mamczarz, K., Lal, A., Martindale, J. L., Kawai, T. & Gorospe, M. Translational repression by RNA-binding protein TIAR. *Mol Cell Biol* **26**, 2716-27 (2006).
229. Nekrasov, M. P. et al. The mRNA-binding protein YB-1 (p50) prevents association of the eukaryotic initiation factor eIF4G with mRNA and inhibits protein synthesis at the initiation stage. *J Biol Chem* **278**, 13936-43 (2003).
230. Evdokimova, V. et al. Akt-mediated YB-1 phosphorylation activates translation of silent mRNA species. *Mol Cell Biol* **26**, 277-92 (2006).
231. Bruno, I. & Wilkinson, M. F. P-bodies react to stress and nonsense. *Cell* **125**, 1036-8 (2006).
232. Lian, S. et al. GW bodies, microRNAs and the cell cycle. *Cell Cycle* **5**, 242-5 (2006).
233. Pillai, R. S. MicroRNA function: multiple mechanisms for a tiny RNA? *Rna* **11**, 1753-61 (2005).
234. Liu, J. et al. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437-41 (2004).
235. Liu, J., Valencia-Sanchez, M. A., Hannon, G. J. & Parker, R. MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. *Nat Cell Biol* **7**, 719-23 (2005).
236. Jakymiw, A. et al. Disruption of GW bodies impairs mammalian RNA interference. *Nat Cell Biol* **7**, 1267-74 (2005).
237. Bhattacharyya, S. N., Habermacher, R., Martine, U., Closs, E. I. & Filipowicz, W. Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell* **125**, 1111-24 (2006).
238. Novina, C. D. & Sharp, P. A. The RNAi revolution. *Nature* **430**, 161-4 (2004).
239. Tabara, H. et al. The rde-1 gene, RNA interference, and transposon silencing in C. elegans. *Cell* **99**, 123-32 (1999).
240. Xie, Z. et al. Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol* **2**, E104 (2004).

241. Napoli, C., Lemieux, C. & Jorgensen, R. Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *Plant Cell* **2**, 279-289 (1990).
242. Guo, S. & Kemphues, K. J. par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* **81**, 611-20 (1995).
243. Fire, A. et al. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806-11 (1998).
244. Elbashir, S. M., Martinez, J., Patkaniowska, A., Lendeckel, W. & Tuschl, T. Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. *Embo J* **20**, 6877-88 (2001).
245. Hamilton, A. J. & Baulcombe, D. C. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**, 950-2 (1999).
246. Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843-54 (1993).
247. Reinhart, B. J. et al. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**, 901-6 (2000).
248. Pasquinelli, A. E. et al. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408**, 86-9 (2000).
249. Lee, Y. et al. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415-9 (2003).
250. Han, J. et al. The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* **18**, 3016-27 (2004).
251. Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363-6 (2001).
252. He, L. & Hannon, G. J. MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* **5**, 522-31 (2004).
253. Lipardi, C., Wei, Q. & Paterson, B. M. RNAi as random degradative PCR: siRNA primers convert mRNA into dsRNAs that are degraded to generate new siRNAs. *Cell* **107**, 297-307 (2001).
254. Khvorova, A., Reynolds, A. & Jayasena, S. D. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209-16 (2003).
255. Schwarz, D. S. et al. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199-208 (2003).
256. Ekwall, K. The RITS complex-A direct link between small RNA and heterochromatin. *Mol Cell* **13**, 304-5 (2004).
257. Tijsterman, M. & Plasterk, R. H. Dicers at RISC; the mechanism of RNAi. *Cell* **117**, 1-3 (2004).
258. Lee, Y. S. et al. Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**, 69-81 (2004).
259. Volpe, T. A. et al. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833-7 (2002).

260. Mochizuki, K., Fine, N. A., Fujisawa, T. & Gorovsky, M. A. Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. *Cell* **110**, 689-99 (2002).
261. Noma, K. et al. RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing. *Nat Genet* **36**, 1174-80 (2004).
262. Verdel, A. et al. RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**, 672-6 (2004).
263. Tabara, H., Grishok, A. & Mello, C. C. RNAi in C. elegans: soaking in the genome sequence. *Science* **282**, 430-1 (1998).
264. Timmons, L. & Fire, A. Specific interference by ingested dsRNA. *Nature* **395**, 854 (1998).
265. Kamath, R. S. et al. Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature* **421**, 231-7 (2003).
266. McManus, M. T. & Sharp, P. A. Gene silencing in mammals by small interfering RNAs. *Nat Rev Genet* **3**, 737-47 (2002).
267. Elbashir, S. M., Harborth, J., Weber, K. & Tuschl, T. Analysis of gene function in somatic mammalian cells using small interfering RNAs. *Methods* **26**, 199-213 (2002).
268. Elbashir, S. M. et al. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494-8 (2001).
269. Cui, W., Ning, J., Naik, U. P. & Duncan, M. K. OptiRNAi, an RNAi design tool. *Comput Methods Programs Biomed* **75**, 67-73 (2004).
270. Reynolds, A. et al. Rational siRNA design for RNA interference. *Nat Biotechnol* **22**, 326-30 (2004).
271. Pelton, T. A., Sharma, S., Schulz, T. C., Rathjen, J. & Rathjen, P. D. Transient pluripotent cell populations during primitive ectoderm formation: correlation of in vivo and in vitro pluripotent cell development. *J Cell Sci* **115**, 329-39 (2002).
272. Hedley, M. L., Amrein, H. & Maniatis, T. An amino acid sequence motif sufficient for subnuclear localization of an arginine/serine-rich splicing factor. *Proc Natl Acad Sci U S A* **92**, 11524-8 (1995).
273. Graveley, B. R. Sorting out the complexity of SR protein functions. *Rna* **6**, 1197-211 (2000).
274. Sanford, J. R. & Bruzik, J. P. Regulation of SR protein localization during development. *Proc Natl Acad Sci U S A* **98**, 10184-9 (2001).
275. Lamond, A. I. & Spector, D. L. Nuclear speckles: a model for nuclear organelles. *Nat Rev Mol Cell Biol* **4**, 605-12 (2003).
276. Kataoka, N., Bachorik, J. L. & Dreyfuss, G. Transportin-SR, a nuclear import receptor for SR proteins. *J Cell Biol* **145**, 1145-52 (1999).
277. Li, H. & Bingham, P. M. Arginine/serine-rich domains of the su(wa) and tra RNA processing regulators target proteins to a subnuclear compartment implicated in splicing. *Cell* **67**, 335-42 (1991).
278. Kavanagh, S. J. et al. A family of RS domain proteins with novel subcellular localization and trafficking. *Nucleic Acids Res* **33**, 1309-22 (2005).

279. Simmer, F. et al. Genome-wide RNAi of C. elegans using the hypersensitive rrf-3 strain reveals novel gene functions. *PLoS Biol* **1**, E12 (2003).
280. Gonczy, P. et al. Functional genomic analysis of cell division in C. elegans using RNAi of genes on chromosome III. *Nature* **408**, 331-6 (2000).
281. Sonnichsen, B. et al. Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. *Nature* **434**, 462-9 (2005).
282. Rual, J. F. et al. Toward improving Caenorhabditis elegans phenome mapping with an ORFeome-based RNAi library. *Genome Res* **14**, 2162-8 (2004).
283. Tuteja, R. & Tuteja, N. Ku autoantigen: a multifunctional DNA-binding protein. *Crit Rev Biochem Mol Biol* **35**, 1-33 (2000).
284. Feng, J. et al. The RNA component of human telomerase. *Science* **269**, 1236-41 (1995).
285. Kelland, L. R. Overcoming the immortality of tumour cells by telomere and telomerase based cancer therapeutics--current status and future prospects. *Eur J Cancer* **41**, 971-9 (2005).
286. McKay, S. J. et al. Gene expression profiling of cells, tissues, and developmental stages of the nematode C. elegans. *Cold Spring Harb Symp Quant Biol* **68**, 159-69 (2003).
287. Ahringer, J. in *Worm Book* (The C. elegans research community, April 6, 2006).
288. Bjork, P. et al. A novel conserved RNA-binding domain protein, RBD-1, is essential for ribosome biogenesis. *Mol Biol Cell* **13**, 3683-95 (2002).
289. Saijou, E., Fujiwara, T., Suzaki, T., Inoue, K. & Sakamoto, H. RBD-1, a nucleolar RNA-binding protein, is essential for Caenorhabditis elegans early development through 18S ribosomal RNA processing. *Nucleic Acids Res* **32**, 1028-36 (2004).
290. Van Buskirk, C. & Schupbach, T. Half pint regulates alternative splice site selection in Drosophila. *Dev Cell* **2**, 343-53 (2002).
291. Page-McCaw, P. S., Amonlirdviman, K. & Sharp, P. A. PUF60: a novel U2AF65-related splicing activity. *Rna* **5**, 1548-60 (1999).
292. Longman, D., Johnstone, I. L. & Caceres, J. F. Functional characterization of SR and SR-related genes in Caenorhabditis elegans. *Embo J* **19**, 1625-37 (2000).
293. Carballo, E., Lai, W. S. & Blackshear, P. J. Feedback inhibition of macrophage tumor necrosis factor-alpha production by tristetraprolin. *Science* **281**, 1001-5 (1998).
294. Lai, W. S. et al. Evidence that tristetraprolin binds to AU-rich elements and promotes the deadenylation and destabilization of tumor necrosis factor alpha mRNA. *Mol Cell Biol* **19**, 4311-23 (1999).
295. Lai, W. S., Carballo, E., Thorn, J. M., Kennington, E. A. & Blackshear, P. J. Interactions of CCCH zinc finger proteins with mRNA. Binding of tristetraprolin-related zinc finger proteins to Au-rich elements and destabilization of mRNA. *J Biol Chem* **275**, 17827-37 (2000).
296. Zorio, D. A. & Blumenthal, T. Both subunits of U2AF recognize the 3' splice site in Caenorhabditis elegans. *Nature* **402**, 835-8 (1999).

297. Dreyfuss, G., Swanson, M. S. & Pinol-Roma, S. Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem Sci* **13**, 86-91 (1988).
298. Kielkopf, C. L., Lucke, S. & Green, M. R. U2AF homology motifs: protein recognition in the RRM world. *Genes Dev* **18**, 1513-26 (2004).
299. Birney, E., Kumar, S. & Krainer, A. R. Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res* **21**, 5803-16 (1993).
300. Ferraiuolo, M. A. et al. A nuclear translation-like factor eIF4AIII is recruited to the mRNA during splicing and functions in nonsense-mediated decay. *Proc Natl Acad Sci U S A* **101**, 4118-23 (2004).
301. Shibuya, T., Tange, T. O., Sonenberg, N. & Moore, M. J. eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay. *Nat Struct Mol Biol* **11**, 346-51 (2004).
302. Wong, M. D., Jin, Z. & Xie, T. Molecular mechanisms of germline stem cell regulation. *Annu Rev Genet* **39**, 173-95 (2005).
303. Lee, J. E. et al. Conversion of Xenopus ectoderm into neurons by NeuroD, a basic helix-loop-helix protein. *Science* **268**, 836-44 (1995).
304. Chitnis, A., Henrique, D., Lewis, J., Ish-Horowicz, D. & Kintner, C. Primary neurogenesis in Xenopus embryos regulated by a homologue of the Drosophila neurogenic gene Delta. *Nature* **375**, 761-6 (1995).
305. Grammer, T. C., Liu, K. J., Mariani, F. V. & Harland, R. M. Use of large-scale expression cloning screens in the Xenopus laevis tadpole to identify gene function. *Dev Biol* **228**, 197-210 (2000).
306. Komatsu, M., Kominami, E., Arahata, K. & Tsukahara, T. Cloning and characterization of two neural-salient serine/arginine-rich (NSSR) proteins involved in the regulation of alternative splicing in neurones. *Genes Cells* **4**, 593-606 (1999).
307. Shin, C. & Manley, J. L. The SR protein SRp38 represses splicing in M phase cells. *Cell* **111**, 407-17 (2002).
308. Yang, L., Embree, L. J., Tsai, S. & Hickstein, D. D. Oncoprotein TLS interacts with serine-arginine proteins involved in RNA splicing. *J Biol Chem* **273**, 27761-4 (1998).
309. Shin, C., Feng, Y. & Manley, J. L. Dephosphorylated SRp38 acts as a splicing repressor in response to heat shock. *Nature* **427**, 553-8 (2004).
310. Liu, K. J. & Harland, R. M. Inhibition of neurogenesis by SRp38, a neuroD-regulated RNA-binding protein. *Development* **132**, 1511-23 (2005).
311. Nakamura, M., Okano, H., Blendy, J. A. & Montell, C. Musashi, a neural RNA-binding protein required for Drosophila adult external sensory organ development. *Neuron* **13**, 67-81 (1994).
312. Okano, H., Imai, T. & Okabe, M. Musashi: a translational regulator of cell fate. *J Cell Sci* **115**, 1355-9 (2002).
313. Sakakibara, S. et al. Mouse-Musashi-1, a neural RNA-binding protein highly enriched in the mammalian CNS stem cell. *Dev Biol* **176**, 230-42 (1996).

314. Sakakibara, S. & Okano, H. Expression of neural RNA-binding proteins in the postnatal CNS: implications of their roles in neuronal and glial cell development. *J Neurosci* **17**, 8300-12 (1997).
315. Imai, T. et al. The neural RNA-binding protein Musashi1 translationally regulates mammalian numb gene expression by interacting with its mRNA. *Mol Cell Biol* **21**, 3888-900 (2001).
316. Liu, M., Iavarone, A. & Freedman, L. P. Transcriptional activation of the human p21(WAF1/CIP1) gene by retinoic acid receptor. Correlation with retinoid induction of U937 cell differentiation. *J Biol Chem* **271**, 31723-8 (1996).
317. Battelli, C., Nikopoulos, G. N., Mitchell, J. G. & Verdi, J. M. The RNA-binding protein Musashi-1 regulates neural development through the translational repression of p21WAF-1. *Mol Cell Neurosci* **31**, 85-96 (2006).
318. Kabos, P., Kabosova, A. & Neuman, T. Blocking HES1 expression initiates GABAergic differentiation and induces the expression of p21(CIP1/WAF1) in human neural stem cells. *J Biol Chem* **277**, 8763-6 (2002).
319. Chu, M. S. et al. Signalling pathway in the induction of neurite outgrowth in human mesenchymal stem cells. *Cell Signal* **18**, 519-30 (2006).
320. Kayahara, T. et al. Candidate markers for stem and early progenitor cells, Musashi-1 and Hes1, are expressed in crypt base columnar cells of mouse small intestine. *FEBS Lett* **535**, 131-5 (2003).
321. Nagata, H., Akiba, Y., Suzuki, H., Okano, H. & Hibi, T. Expression of Musashi-1 in the rat stomach and changes during mucosal injury and restitution. *FEBS Lett* **580**, 27-33 (2006).
322. Siddall, N. A., McLaughlin, E. A., Marriner, N. L. & Hime, G. R. The RNA-binding protein Musashi is required intrinsically to maintain stem cell identity. *Proc Natl Acad Sci U S A* **103**, 8402-7 (2006).
323. Hung, S. C. et al. Immortalization without neoplastic transformation of human mesenchymal stem cells by transduction with HPV16 E6/E7 genes. *Int J Cancer* **110**, 313-9 (2004).
324. Posner, J. B. Immunology of paraneoplastic syndromes: overview. *Ann N Y Acad Sci* **998**, 178-86 (2003).
325. Roberts, W. K. & Darnell, R. B. Neuroimmunology of the paraneoplastic neurological degenerations. *Curr Opin Immunol* **16**, 616-22 (2004).
326. Albert, M. L. & Darnell, R. B. Paraneoplastic neurological degenerations: keys to tumour immunity. *Nat Rev Cancer* **4**, 36-44 (2004).
327. Robinow, S., Campos, A. R., Yao, K. M. & White, K. The elav gene product of Drosophila, required in neurons, has three RNP consensus motifs. *Science* **242**, 1570-2 (1988).
328. Okano, H. J. & Darnell, R. B. A hierarchy of Hu RNA binding proteins in developing and adult neurons. *J Neurosci* **17**, 3024-37 (1997).
329. Joseph, B., Orlian, M. & Furneaux, H. p21(waf1) mRNA contains a conserved element in its 3'-untranslated region that is bound by the Elav-like mRNA-stabilizing proteins. *J Biol Chem* **273**, 20511-6 (1998).

330. Chung, S., Eckrich, M., Perrone-Bizzozero, N., Kohn, D. T. & Furneaux, H. The Elav-like proteins bind to a conserved regulatory element in the 3'-untranslated region of GAP-43 mRNA. *J Biol Chem* **272**, 6593-8 (1997).
331. Akamatsu, W. et al. Mammalian ELAV-like neuronal RNA-binding proteins HuB and HuC promote neuronal development in both the central and the peripheral nervous systems. *Proc Natl Acad Sci U S A* **96**, 9885-90 (1999).
332. Kasashima, K., Terashima, K., Yamamoto, K., Sakashita, E. & Sakamoto, H. Cytoplasmic localization is required for the mammalian ELAV-like protein HuD to induce neuronal differentiation. *Genes Cells* **4**, 667-83 (1999).
333. Akamatsu, W. et al. The RNA-binding protein HuD regulates neuronal cell identity and maturation. *Proc Natl Acad Sci U S A* **102**, 4625-30 (2005).
334. Buckanovich, R. J., Posner, J. B. & Darnell, R. B. Nova, the paraneoplastic Ri antigen, is homologous to an RNA-binding protein and is specifically expressed in the developing motor system. *Neuron* **11**, 657-72 (1993).
335. Buckanovich, R. J. & Darnell, R. B. The neuronal RNA binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo. *Mol Cell Biol* **17**, 3194-201 (1997).
336. Jensen, K. B. et al. Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron* **25**, 359-71 (2000).
337. Macdonald, R. L. Ethanol, gamma-aminobutyrate type A receptors, and protein kinase C phosphorylation. *Proc Natl Acad Sci U S A* **92**, 3633-5 (1995).
338. Sommer, B. et al. Flip and flop: a cell-specific functional switch in glutamate-operated channels of the CNS. *Science* **249**, 1580-5 (1990).
339. Ule, J. et al. Nova regulates brain-specific splicing to shape the synapse. *Nat Genet* **37**, 844-52 (2005).
340. Ule, J. & Darnell, R. B. RNA binding proteins and the regulation of neuronal synaptic plasticity. *Curr Opin Neurobiol* **16**, 102-10 (2006).
341. Dredge, B. K., Stefani, G., Engelhard, C. C. & Darnell, R. B. Nova autoregulation reveals dual functions in neuronal splicing. *Embo J* **24**, 1608-20 (2005).
342. Okano, H. et al. Function of RNA-binding protein Musashi-1 in stem cells. *Exp Cell Res* **306**, 349-56 (2005).
343. Lall, S. et al. A genome-wide map of conserved microRNA targets in C. elegans. *Curr Biol* **16**, 460-71 (2006).
344. Brown, J. R. & Sanseau, P. A computational view of microRNAs and their targets. *Drug Discov Today* **10**, 595-601 (2005).
345. Stanford, W. L., Cohn, J. B. & Cordes, S. P. Gene-trap mutagenesis: past, present and beyond. *Nat Rev Genet* **2**, 756-68 (2001).
346. Skarnes, W. C. et al. A public gene trap resource for mouse functional genomics. *Nat Genet* **36**, 543-4 (2004).
347. Hansen, J. et al. A large-scale, gene-driven mutagenesis approach for the functional analysis of the mouse genome. *Proc Natl Acad Sci U S A* **100**, 9918-22 (2003).
348. Chen, W. V. & Soriano, P. Gene trap mutagenesis in embryonic stem cells. *Methods Enzymol* **365**, 367-86 (2003).

APPENDIX A: EXON SEQUENCES SUBJECT TO ALTERNATIVE SPLICING

Several isoforms of the 3B6 transcript were cloned and sequenced during the course of this project. The sequences of the exons subject to alternative splicing plus 25 bp of flanking sequence are given below. Intron sequences are shown in blue lower case font, while the exon sequences are in black upper case font. None of the alternative splices lead to shifting of the translational reading frame, nor have any evident impact on the putative CCCH Zinc Finger or RRM domains. The presentation of these sequences was made with the assistance of the Ensembl mouse genome browser exon view (www.ensembl.org/Mus_musculus/index.html).

A.1. Exon 9

Exon 9 contains alternative splice acceptor sites. The 3B6 transcript 1a which is most similar to its human homolog arises with the use of the splice acceptor site shown in italics. Use of the alternate splice acceptor site, shown underlined, results in the addition of 15 bp to the transcript, adding 5 aa to the translated sequence.

...ttttataattgttgtttaattacag<u>g</u>ttttttgtgttacc***ag***ATACATACGATACAGATGGCTACAATCCTG
AAGCCCCAAGCATAACAAACACTTCCAGACCTATGTATAGACACAGAGTCCA
TGCACAAAGGCCCAACTTGATAGGACTAACATCAGGTGACATGGATTTACCA
CCCAGAGgtattaacaatcttttcttattttc....

A.2. Exon 13

Exon 13 contains alternative splice acceptor sites. The 3B6 transcript 1a which is most similar to its human homolog arises when the use of the splice acceptor site shown in italics is used. Use of the alternate splice acceptor site, shown underlined, results in the loss of 9 bp from the transcript, leading to a loss of 3 aa from the translated sequence.

....ctttaacttttccttttgttttg***ag***GTAATACAGCCTTTAGTCCAGCAGCCCATTTTACCTGTT
GTGAAGCAGTCAGTCAAAGAGCGGCTGGGTCCTGTACCATCAGCTACAACAG
AGCCAGCAGAAGCCCAGAGTGCTACTTCAGAACTTCCCCAGgtaaacagcagctttgatc
agcagc...

A.3. Exon 14

Exon 14 has been found to be a cassette exon. Loss of the 72 bp that make up exon 14 results in a loss of 24 aa from the translated sequence. This is referred to as transcript 2a. Transcript 2b arises when exon 14 is lost, and the alternate splice acceptor in exon 13 is used, resulting in a total loss of 81 bp from the transcript and 27 aa from the translated sequence.

....aatatgtccattttaatttttttagAATGTAACTAAGTTATCTGTGAAGGACAGGTTGGGTTT
TGTATCAAAGCCATCTGTTTCAGCAACTGAAAAGgtaagaatagcatgatgtgtttgtc...

APPENDIX B: PROPOGATION OF ERROR

Propagation of error, also known as the Δ method, was used to determine the standard error associated with the fold-changes in expression resulting from REST analysis of real-time PCR data. This method was suggested as an appropriate method for this experiment by Dr. Alexandre Bureau, formerly of University of Lethbridge, now at Université Laval. Error propagation can be used to combine multiple independent sources of error contributing to the same measurement. Sources of error in analysis of real-time PCR include measurement of the mean C_T for each of the experimental and internal reference genes, and measurement of the efficiencies for each reaction. The propagation of error method as used for this project is as follows, where var = variance, obs = observations, SD = standard deviation, E = reaction efficiency, ES = ES cell samples, exp = experimental samples, ratio = fold change in expression and SE = standard error:

$$Var[mean_{ES}] = \frac{SD^2}{n\ obs} \qquad Var[mean_{\exp}] = \frac{SD^2}{n\ obs}$$

$$Var[\Delta C_{T\ \ 3B6}] = Var[mean_{ES}] + Var[mean_{\exp}]$$

$$Var[\Delta C_{T\ \ GAPDH}] = Var[mean_{ES}] + Var[mean_{\exp}]$$

$$Var[\log ratio] = \left(Var[\Delta C_{T\ 3B6}](\log[E_{3B6}])\right)^2 + \left(Var[\Delta C_{T\ GAPDH}](\log[E_{GAPDH}])\right)^2$$

$$SE(ratio) = (ratio)(Var[\log\ ratio])$$

APPENDIX C: ANALYSIS OF REAL-TIME PCR DATA: $\Delta\Delta C_T$ METHOD

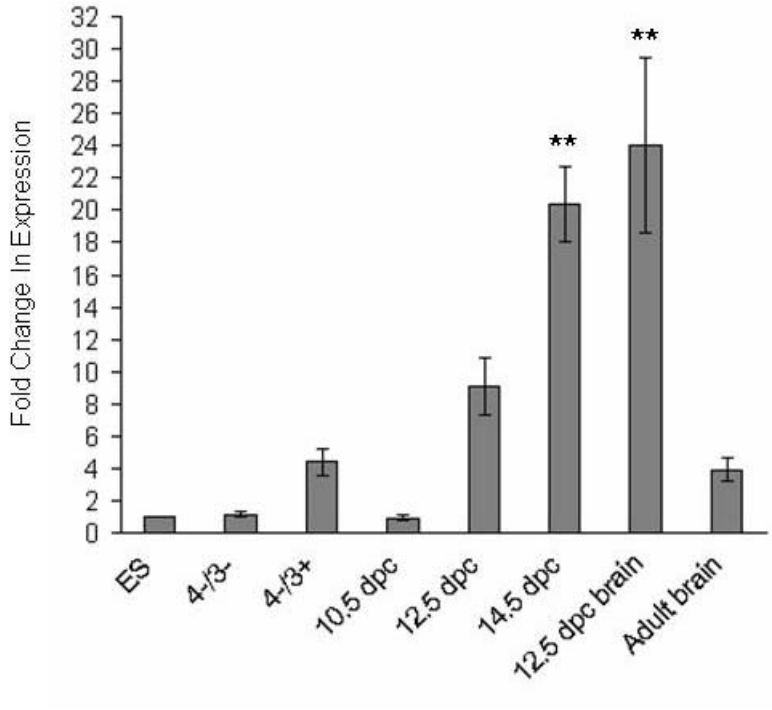


Figure 36: Real-time PCR results analyzed using the ΔΔCT method. Real time PCR data (Section 4.3.2) was analysed using the $\Delta\Delta C_T$ method as described by Livak and Schmittgen[162]. Fold changes in expression over that observed in ES cells is plotted. ** indicates statistically significant changes in expression level ($p<0.05$). In comparison to the results derived from use of the REST method, trends in expression are similar; however fold changes in expression are altered. This may be attributable to difference in efficiencies between reactions, which is assumed to be equal by the $\Delta\Delta C_T$ method. In addition, using a t-test of significance suggests that even a fold-change difference of greater than 9, as seen in 12.5 dpc embryos, is not significant.

APPENDIX D: FULL LENGTH CODING SEQUENCE OF 3B6

For expression of mSe70-2 in all utilized systems, the full length 3B6 coding sequence shown below was cloned. This sequence is most similar to its human homolog, and includes exon 14 and exon 22a. This coding sequence is 3024 bp long, translating to 1007 aa. Exons are illustrated in alternating black and bold blue print. The putative CCCH zinc finger is indicated by ***, and the putative RRM is indicated by ###.

```
1    ATGGTTTCTAAGATGATCATCGAAAACTTCGAGGCGCTCAAGTCCTGGCTCAGCAAGACC
1    -M--V--S--K--M--I--I--E--N--F--E--A--L--K--S--W--L--S--K--T-

61   CTCGAGCCCAT**CTGTGATGCAGATCCATCAGCCCTAGCAAAATATGTTTTGGCTCTGGTA**
21   -L--E--P--I--C--D--A--D--P--S--A--L--A--K--Y--V--L--A--L--V-

121  **AAGAAAGAAAAAAGTGAAAAAGAATTAAAGGCGTTATGCATTGATCAGCTTGATGTATTT**
41   -K--K--D--K--S--E--K--E--L--K--A--L--C--I--D--Q--L--D--V--F-

181  **CTTCAGAAAG**AGACACAAATATTTGTTGAAAAACTTTTTGATGCTGTGAATACAAAGAGT
61   -L--Q--K--E--T--Q--I--F--V--E--K--L--F--D--A--V--N--T--K--S-

241  TATCTACCTCCTCCAGAGCAGCCATCGTCAGGAAGCCTGAAGGTAGACTTTCTTCAGCAC
81   -Y--L--P--P--P--E--Q--P--S--S--G--S--L--K--V--D--F--L--Q--H-

301  CAAGAAAAAGATATAAAAAAAGAAGAG**CTCACAAAGGAGGAAGAGCGAGAGAAGAAGTTT**
101  -Q--E--K--D--I--K--K--E--E--L--T--K--E--E--E--R--E--K--K--F-

361  **TCCAGAAGGTTGAATCACAGCCCTCCCCAGTCAAGCTCCCGATACAGAGATAATAG**AAGC
121  -S--R--R--L--N--H--S--P--P--Q--S--S--S--R--Y--R--D--N--R--S-

421  CGTGATGAGAGGAAAAAAGATGATCGTTCTCGAAAAAGAGATTATGATCGAAACCCTCCT
141  -R--D--E--R--K--K--D--D--R--S--R--K--R--D--Y--D--R--N--P--P-

481  CGAAGAGATTCATATAGAGACCGGTATAACAGAAGGCGAGGGAGAAGTCGCAGTTACAGC
161  -R--R--D--S--Y--R--D--R--Y--N--R--R--R--G--R--S--R--S--Y--S-

541  AGGAGTCGAAGTCGAAGCTGGAGTAAAGAGAGGCTTCGTGACAGGGATAGAGATAGGAGC
181  -R--S--R--S--R--S--W--S--K--E--R--L--R--D--R--D--R--D--R--S-

601  AGAACGAGGAGCAGAAGCAGGACGCGGAGCAGGG**AAAGGGACCTGGTGAAACCTAAATAT**
201  -R--T--R--S--R--S--R--T--R--S--R--E--R--D--L--V--K--P--K--Y-

661  **GACCTGGATAGAACAGATCCGTTAGAAAATAATTATACACCAGTTTCTTCAGTATCTAAT**
221  -D--L--D--R--T--D--P--L--E--N--N--Y--T--P--V--S--S--V--S--N-


721  **ATTTCATCTGGCCACTATCCTGTACCTACTTTGAGCAGCACCATTACCGTCATTGCTCCT**
241  -I--S--S--G--H--Y--P--V--P--T--L--S--S--T--I--T--V--I--A--P-
```

```
781  ACTCATCACGGTAATAACACTACCGAAAGTTGGTCTGAATTTCATGAAGACCAGGTGGAT
261  -T--H--H--G--N--N--T--T--E--S--W--S--E--F--H--E--D--Q--V--D-

841  CACAACTCTTATGTAAGGCCACCAATGCCAAAGAAACGGTGTAGAGATTATGATGAAAAG
281  -H--N--S--Y--V--R--P--P--M--P--K--K--R--C--R--D--Y--D--E--K-
                          ************************************

901  GGGTTCTGTATGAGAGGAGACATGTGTCCTTTTGATCATGGAAGCGACCCAGTAGTTGTA
301  -G--F--C--M--R--G--D--M--C--P--F--D--H--G--S--D--P--V--V--V-
     **************************************************

961  GAAGATGTGAATCTGCCAGGCATGCTGCCTTTCCCAGCTCAGCCACCTGTTGTTGAAGGA
321  -E--D--V--N--L--P--G--M--L--P--F--P--A--Q--P--P--V--V--E--G-

1021 CCACCTCCTCCCGGGCTGCCTCCACCTCCTCCGATCCTTACACCGCCACCTGTGAATCTT
341  -P--P--P--P--G--L--P--P--P--P--P--I--L--T--P--P--P--V--N--L-

1081 AGACCCCCTGTGCCACCTCCCGGTCCATTACCACCCAGTCTGCCACCTGTCACAGGACCT
361  -R--P--P--V--P--P--P--G--P--L--P--P--S--L--P--P--V--T--G--P-

1141 CCTCCTCCACTTCCTCCTTTGCAGCCGTCTGGCATGGATGCGCCACCTAACTCTGCGACC
381  -P--P--P--L--P--P--L--Q--P--S--G--M--D--A--P--P--N--S--A--T-

1201 AGCTCTGTCCCTACTGTAGTAACAACTGGCATTCATCACCAGCCTCCTCCTGCTCCACCC
401  -S--S--V--P--T--V--V--T--T--G--I--H--H--Q--P--P--P--A--P--P-

1261 TCTCTCTTTACTGCAGATACATACGATACAGATGGCTACAATCCTGAAGCCCCAAGCATA
421  -S--L--F--T--A--D--T--Y--D--T--D--G--Y--N--P--E--A--P--S--I-

1321 ACAAACACTTCCAGACCTATGTATAGACACAGAGTCCATGCACAAAGGCCCAACTTGATA
441  -T--N--T--S--R--P--M--Y--R--H--R--V--H--A--Q--R--P--N--L--I-

1381 GGACTAACATCAGGTGACATGGATTTACCACCCAGAGAAAAGCCTCCTAATAAAAGCAGT
461  -G--L--T--S--G--D--M--D--L--P--P--R--E--K--P--P--N--K--S--S-

1441 ATGAGGATAGTAGTGGATTCAGAATCCAGAAAAAGAACTATTGGGTCTGGGGAACCTGGA
481  -M--R--I--V--V--D--S--E--S--R--K--R--T--I--G--S--G--E--P--G-

1501 GTTTCTACAAAGAAGACATGGTTTGATAAACCAAATTTTAATAGGACAAACAGCCCAGGT
501  -V--S--T--K--K--T--W--F--D--K--P--N--F--N--R--T--N--S--P--G-
                                              ##################

1561 TTCCAGAAGAAGGTTCAGTTTGGAAATGAAAATACTAAACTTGAACTTCGGAAGGTTCCT
521  -F--Q--K--K--V--Q--F--G--N--E--N--T--K--L--E--L--R--K--V--P-
     ############################################################

1621 CCAGAATTAAATAACATCAGCAAACTCAATGAACATTTTAGTCGATTTGGAACATTGGTT
541  -P--E--L--N--N--I--S--K--L--N--E--H--F--S--R--F--G--T--L--V-
     ############################################################

1681 AACTTACAGGTTGCCTATAATGGTGATCCTGAAGGTGCCCTTATTCAATTCGCAACTTAT
561  -N--L--Q--V--A--Y--N--G--D--P--E--G--A--L--I--Q--F--A--T--Y-
     ############################################################
```

```
1741 GAAGAAGCAAAGAAAGCAATCTCAAGTACAGAAGCAGTGTTAAATAATCGTTTTATTAAG
581  -E--E--A--K--K--A--I--S--S--T--E--A--V--L--N--N--R--F--I--K-
     ############################################################

1801 GTTTATTGGCACAGAGAAGGAACTACTCAGCAGTTACAAACTACTTCACCAAAGGTAATA
601  -V--Y--W--H--R--E--G--T--T--Q--Q--L--Q--T--T--S--P--K--V--I-
     ################################

1861 CAGCCTTTAGTCCAGCAGCCCATTTTACCTGTTGTGAAGCAGTCAGTCAAAGAGCGGCTG
621  -Q--P--L--V--Q--Q--P--I--L--P--V--V--K--Q--S--V--K--E--R--L-

1921 GGTCCTGTACCATCAGCTACAACAGAGCCAGCAGAAGCCCAGAGTGCTACTTCAGAACTT
641  -G--P--V--P--S--A--T--T--E--P--A--E--A--Q--S--A--T--S--E--L-

1981 CCCCAGAATGTAACTAAGTTATCTGTGAAGGACAGGTTGGGTTTTGTATCAAAGCCATCT
661  -P--Q--N--V--T--K--L--S--V--K--D--R--L--G--F--V--S--K--P--S-

2041 GTTTCAGCAACTGAAAAGGTGTTGTCTACATCTACTGGTCTCACAAAAACGGTGTATAAT
681  -V--S--A--T--E--K--V--L--S--T--S--T--G--L--T--K--T--V--Y--N-

2101CCAGCTGCTTTGAAGGCTGCACAGAAAACCTTATCAGTTTCTACCCCTGCAGTTGATAAT
701  -P--A--A--L--K--A--A--Q--K--T--L--S--V--S--T--P--A--V--D--N-

2161 AATGAAGCACAGAAAAAGAAACAGGAGGCACTGAAACTTCAGCAGGATGTAAGAAAAAGG
721  -N--E--A--Q--K--K--K--Q--E--A--L--K--L--Q--Q--D--V--R--K--R-

2221 AAGCAAGAAATTTTAGAAAAGCACATTGAAACACAGAAGATGCTAATTTCAAAACTAGAG
741  -K--Q--E--I--L--E--K--H--I--E--T--Q--K--M--L--I--S--K--L--E-

2281 AAAAACAAAACCATGAAGTCTGAAGATAAAGCAGAAATAATGAAAACATTGGAGATTTTG
761  -K--N--K--T--M--K--S--E--D--K--A--E--I--M--K--T--L--E--I--L-

2341 ACAAAAAATATTACCAAGTTAAAAGATGAGGTTAAATCTACATCTCCTGGACGCTGTCTT
781  -T--K--N--I--T--K--L--K--D--E--V--K--S--T--S--P--G--R--C--L-

2401 CCCAAAAGTATAAAAACAAAGACTCAGATGCAGAAAGAATTGCTTGACACAGAATTGGAT
801  -P--K--S--I--K--T--K--T--Q--M--Q--K--E--L--L--D--T--E--L--D-

2461 TTATATAAGAAGATGCAGGCTGGAGAAGAAGTCACTGAACTTAGGAGAAAATATACAGAA
821  -L--Y--K--K--M--Q--A--G--E--E--V--T--E--L--R--R--K--Y--T--E-

2521 TTACAGCTGGAAGCTGCTAAAAGAGGAATTCTTTCATCTGGCCGAGGTAGAGGAATTCAT
841  -L--Q--L--E--A--A--K--R--G--I--L--S--S--G--R--G--R--G--I--H-

2581 ACAAGAGGTCGAGGTACAGCTCACGGCCGGGGCAGGGGTAGAGGCCGAGGCCGAGGTGTG
861  -T--R--G--R--G--T--A--H--G--R--G--R--G--R--G--R--G--R--G--V-

2621 CCTGGGCATGCTGTAGTGGATCACCGACCCCGAGCACTGGAGATTTCTGCATTTACAGAG
881  -P--G--H--A--V--V--D--H--R--P--R--A--L--E--I--S--A--F--T--E-


2681 AGTGATAGAGAAGATCTTCTTCCTCATTTTGCGCAATATGGGGAAATTGAGGACTGTCAG
901  -S--D--R--E--D--L--L--P--H--F--A--Q--Y--G--E--I--E--D--C--Q-
```

```
2741 ATTGATGACGCTTCACTCCATGCAATAATCACATTCAAGACAAGAGCAGAAGCAGAAGCA
921  -I--D--D--A--S--L--H--A--I--I--T--F--K--T--R--A--E--A--E--A-

2801 GCTGCAATTCACGGCGCTCGCTTCAAAGGGCAGGATTTGAAGCTGGCCTGGAACAAACCA
941  -A--A--I--H--G--A--R--F--K--G--Q--D--L--K--L--E--E--E--F--Q-

2861 ATAGCTAACATGTCAGCTGTGGACACTGAAGAAGCTGAACCTGATGAAGAGGAATTTCAG
961  -I--A--N--M--S--A--V--D--T--E--E--A--E--P--D--E--E--E--F--Q-

2921 GAAGAGTCTTTGGTGGATGACTCATTACTTCAAGATGATGACGAAGAAGAAGAGGACAAT
981  -E--E--S--L--V--D--D--S--L--L--Q--D--D--D--E--E--E--E--D--N-

3001 GAATCTCGTTCTTGGAGAAGATGA
1001 -E--S--R--S--W--R--R--*-
```

APPENDIX E: CHARACTERIZATION OF GENE TRAP CELL LINES

Gene trap mutagenesis is a high throughput technique that can be used to generate random insertional mutations across a given genome[345]. Gene trapping in ES cells provides many experimental opportunities such as generation of null mice and mice expressing reporters such as β-gal from a gene's endogenous promoter. The international gene trap consortium (IGTC) is a worldwide collaboration with the goal of trapping every gene in the mouse genome in ES cell lines for large scale functional analysis[346]. As of June 1, 2006, 57 000 (www.genetrap.org) gene trapped cell lines had been annotated, and the IGTC freely provides these ES cell lines to researchers on a cost recovery basis.

During the course of this project four cell lines annotated as having 3B6 traps became available, and all were procured for use in functional analysis of 3B6/mSe70-2. The first cell line was F026D01, generated by the German Gene Trap Consortium (GGTC), a collaboration of several German universities. The remaining three cell lines came available simultaneously from the Sanger Institute Gene Trap Resource (SIGTR; Cambridge, UK). These cell lines were AP0798, AR0055, and AR0058.

The F026D01 cell line was generated using the pT1ATGβgeo vector.[347] The other three cell lines were generated using the vector pGT01xr (Unpublished; www.sanger.ac.uk/PostGenomics/genetrap/). Both pT1ATGβgeo and pGT01xr are typical gene trap vectors containing the promoterless reporter/selectable marker fusion of β-galactosidase and neomycin transferase known as β-geo, flanked by a splice acceptor and a polyadenylation sequence[348].

In the F026D01 cell line, the vector insertion was annotated as occurring between exon 6 and 7, in the sixth intron. Though this trap did not give a complete knockout of 3B6, it was very likely to give a functional knockout as truncation of the mSe70-2 protein would occur prior to both of the putative functional domains. With this in mind, the intent was to inject this cell line into blastocysts for the production of transgenic mice.

Cell lines AP0798, AR0055, and AR0058 were annotated as having the same gene trap insertion within intron 18 of 3B6. This insertion was not expected to lead to a

functional knockout of 3B6, as both functional domains would be retained in mSe70-2. However, it was expected that one or more of these cell lines would prove experimentally useful, in cellular localization studies or studies of 3B6 distribution in the developing mouse embryo.

All gene trapped cell lined were received frozen on dry ice, and were revived as described in section 2.3.3. After several feeder supported passages, they were transferred to feeder free culture and maintained as described in section 2.3.2, with the exception that an active concentration of 150 μg/mL G418 was added to the ES cell growth medium. G418 selection was used to ensure the presence of the neomycin resistance conferring gene trap in each cell line. All cell lines were viable under G418 selection with the exception of AP0798.

As the GGTC was known to not karyotype their gene trapped cell lines, a sample of the F026D01 cell line was given to Ms. Eileen Rattner of the University of Calgary Embryonic Stem Cell Targeted Mutagenesis facility for karyotyping. The karyotype of the F026D01 cell line was found to be grossly abnormal (data not shown). For transgenic mouse production, karyotypic normality is essential to yield viable embryos and germline chimeras. Further work with this cell line was immediately discontinued.

Prior to karyotyping of the AR0055 and AR0058 cell lines, confirmation of the gene trap position within 3B6 was carried out. RNA was prepared from each of the cell lines (section 2.4.1), reverse transcribed (section 2.4.3), and subjected to RT-PCR (section 2.5.3) using primer 3B6-13F paired with primers Geo-1R or Geo-2R for detection of a fusion transcript consisting of 3B6 and β-geo. RT-PCR using the primer pair 3B6-13F and 3B6-19R was included as a control. All primer sequences are available in Table 1. If the insertion of the gene trap vector existed within intron 18 as annotated, then the control reaction would not produce any product as the fusion of 3B6 with β-geo would occur after exon 18. Conversely, if the gene trap was not present, the reactions using the Geo-1R and -2R primers were expected to not produce any product. Results unfortunately suggested absence of the gene trap (data not shown). The control reactions produced the product expected if no gene trap was present, and the Geo-1R reactions

produced no product indicating absence of the gene trap at the annotated locus. The reactions using primer Geo-2R also did not produce any specific product. Examination of the genomic locus was undertaken in order to reconfirm this result. Genomic DNA was prepared from the AR0055, and AR0058 cell lines as described in section 2.7.2. This DNA was used as template for standard PCR reactions (section 2.5.2) using the Geo-1R and -2R primers along with primers designed to fall within intron 18. Primers 3B6-In18F and 3B6-In18R1 both lie 5′ of the annotated gene trap insertion site, and primer 3B6-In18R2 lay across the site disrupted by insertion. Although some non-specific amplification occurred in several reactions, results confirmed that the gene trap insertion was not present at the 3B6 genomic locus [Figure 37]. Although cell lines AR0055, and AR0058 were likely to contain a gene trap as evidenced by their resistance to G418, it was clear that they had been mis-annotated by the SIGTR gene trap analysis pipeline. Further work with these cell lines was therefore discontinued.

In conclusion, none of the four cell lines annotated as 3B6 gene traps were suitable for use. In the context of this project, the acquisition and characterization of gene trap cell lines turned out to be an unfortunate use of research time and funds.

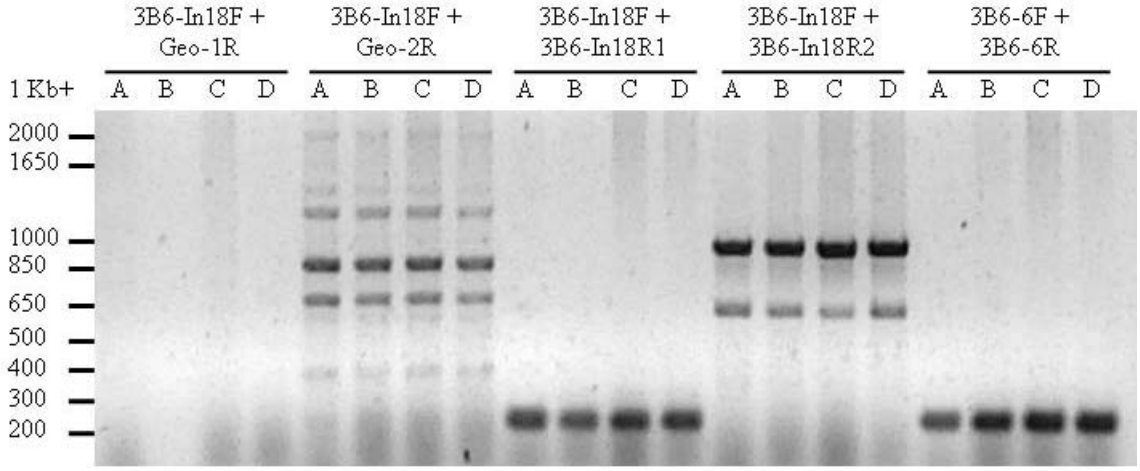
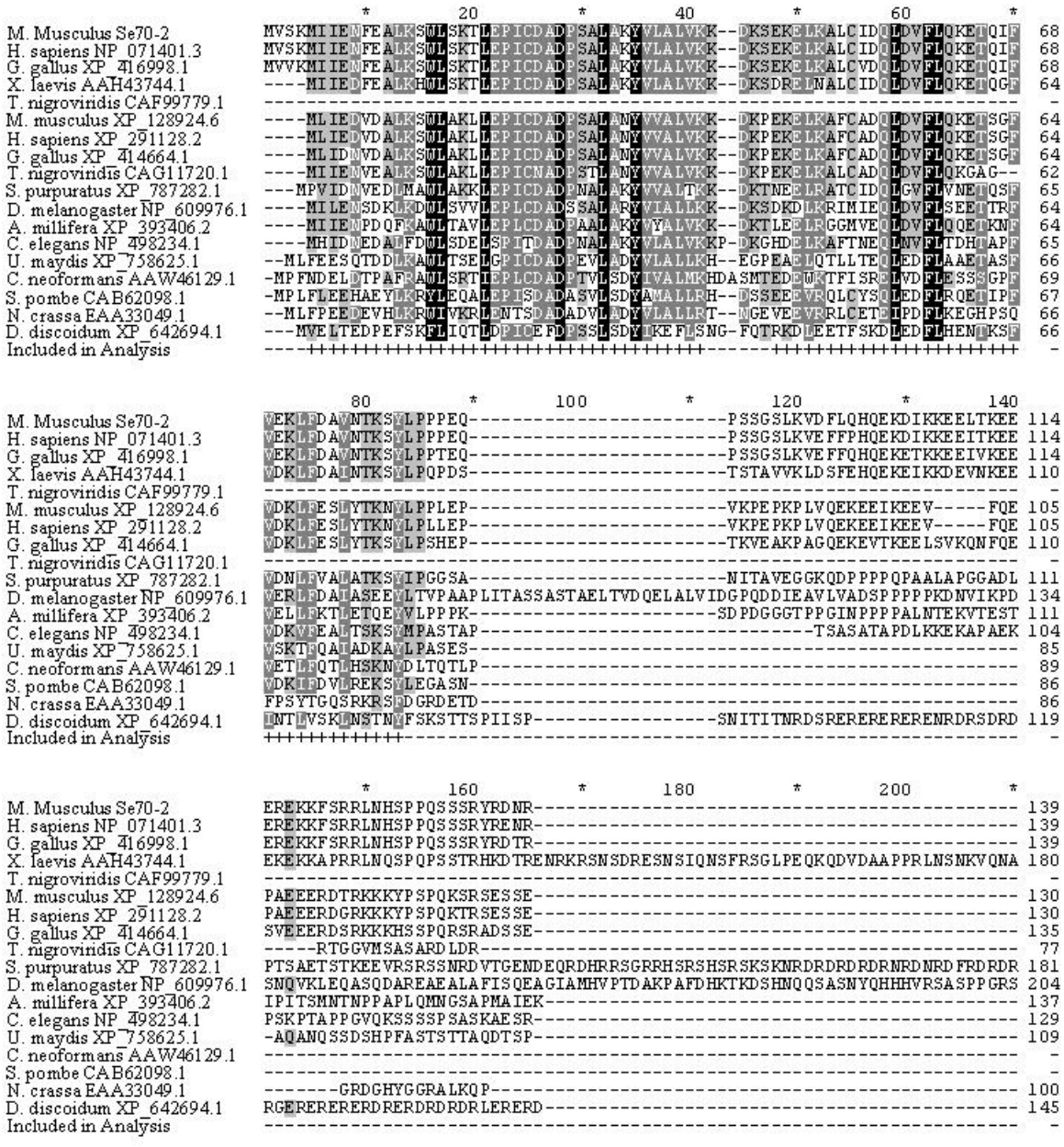
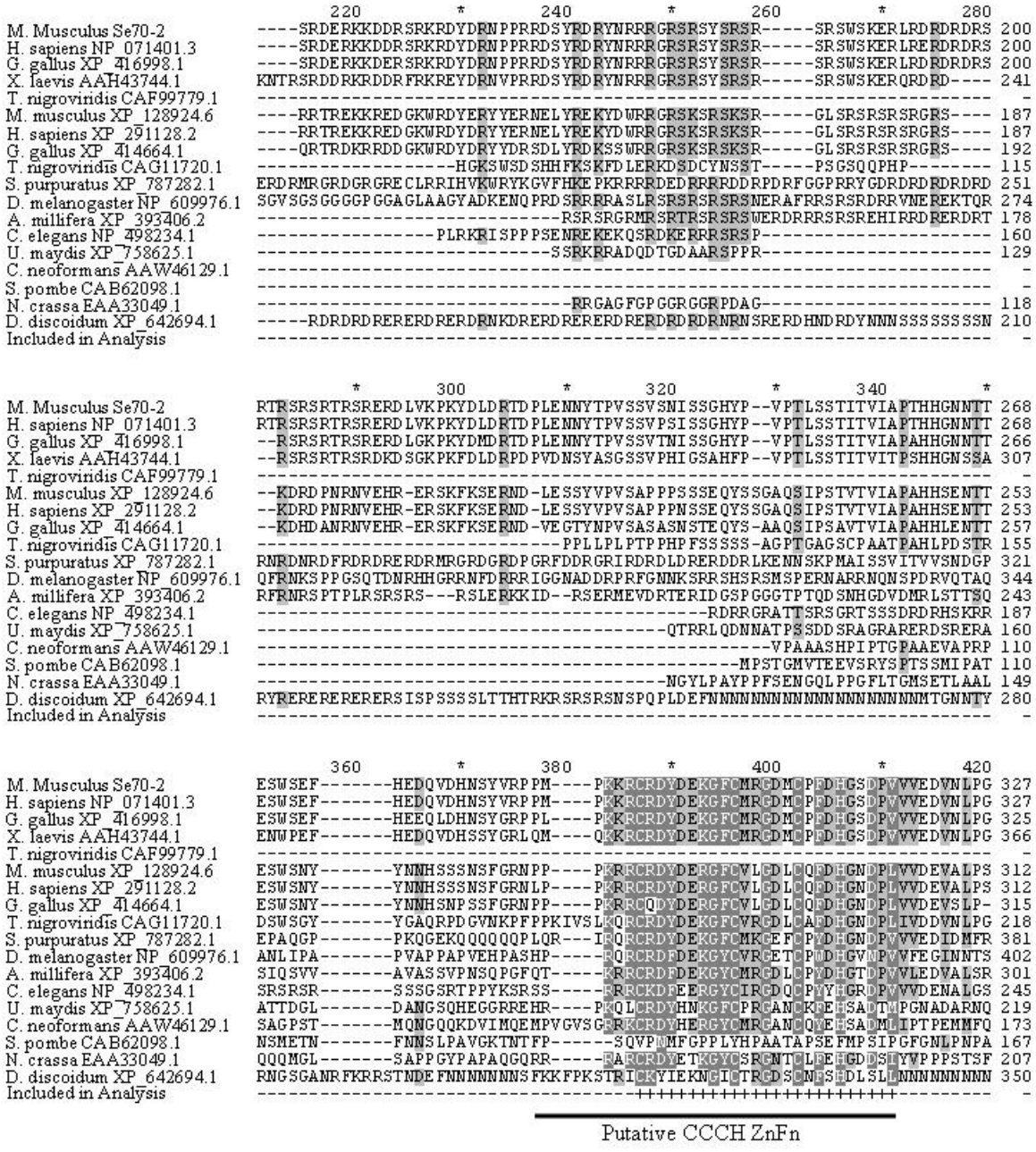


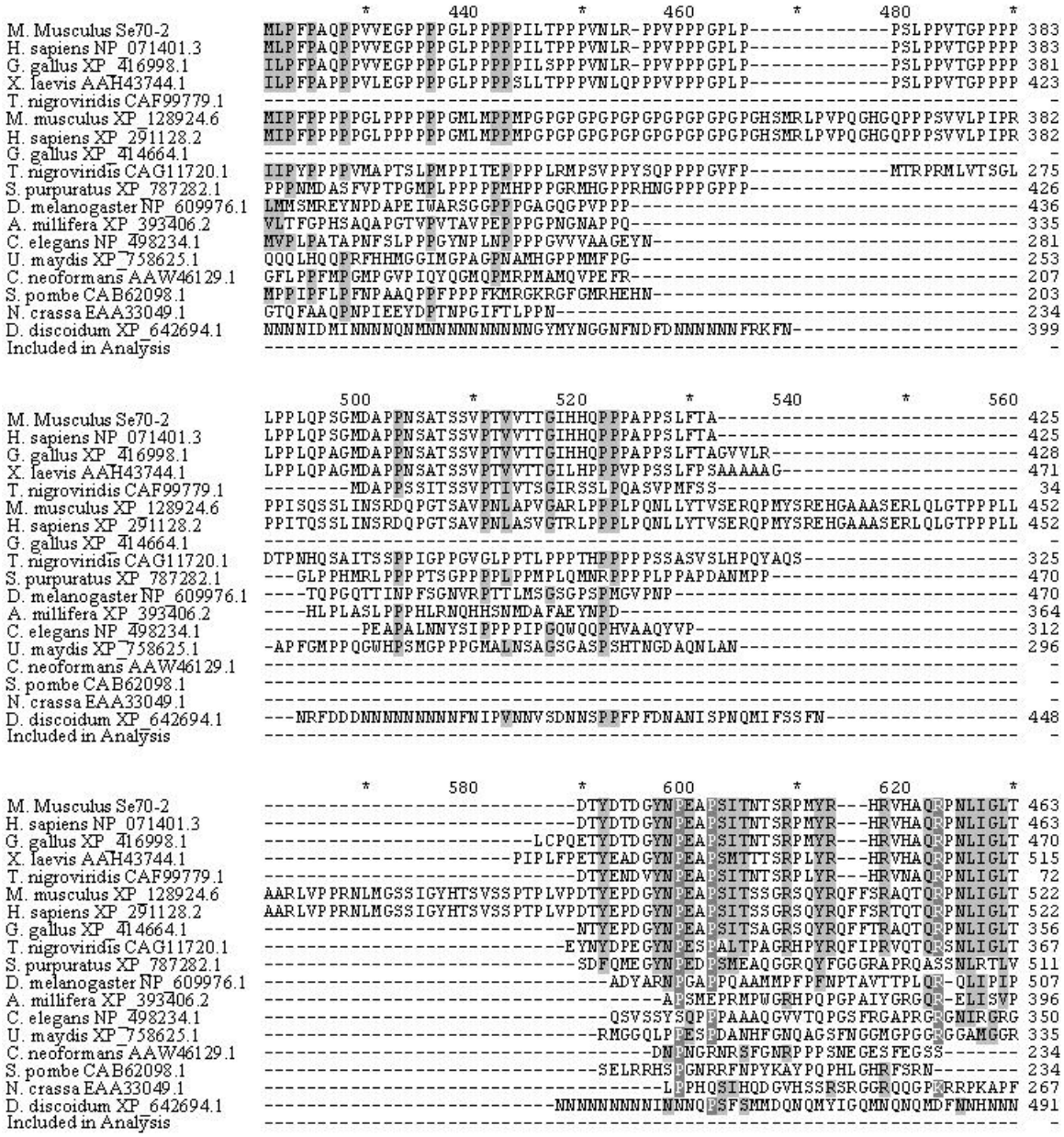
Figure 37: Cell lines AR0055 and AR0058 do not contain a gene trap insertion in the 3B6 locus. Genomic DNA was prepared from cell lines AR0055 and AR0058 and subjected to PCR with the indicated primers, for confirmation of a gene trap insertion within intron 18 of 3B6. Expected bands were as follows: 3B6-In18F + Geo1R, 1106 bp if insertion is present; 3B6-In18F + Geo-2R, 1025 bp if insertion is present; 3B6-In18F + 3B6-In18R1, 239 bp (control); 3B6 –In18F + 3B6-In18R2, 956 bp if insertion is not present; 3B6-6F + 3B6-6R, 240 bp (control). Lane markers: A, wild type R1 ES cells; B, AR0055; C, AR0058 preperation 1; D, AR0058 preparation 2. Although some non-specific products were produced, results clearly indicate the absence of the gene trap insertion at the 3B6 locus.

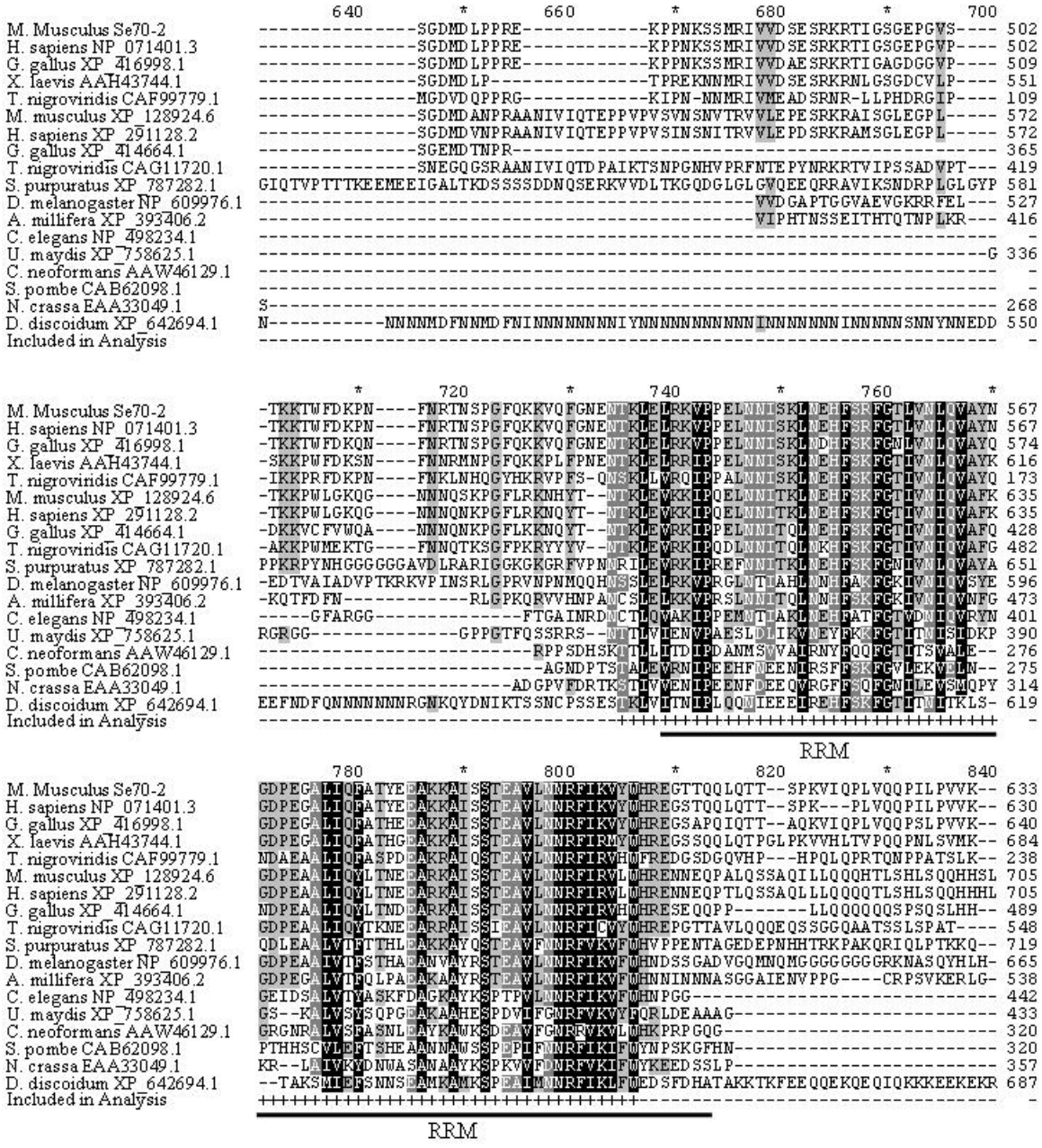
APPENDIX F: MULTIPLE SEQUENCE ALIGNMENT OF SE70-2 HOMOLOGS

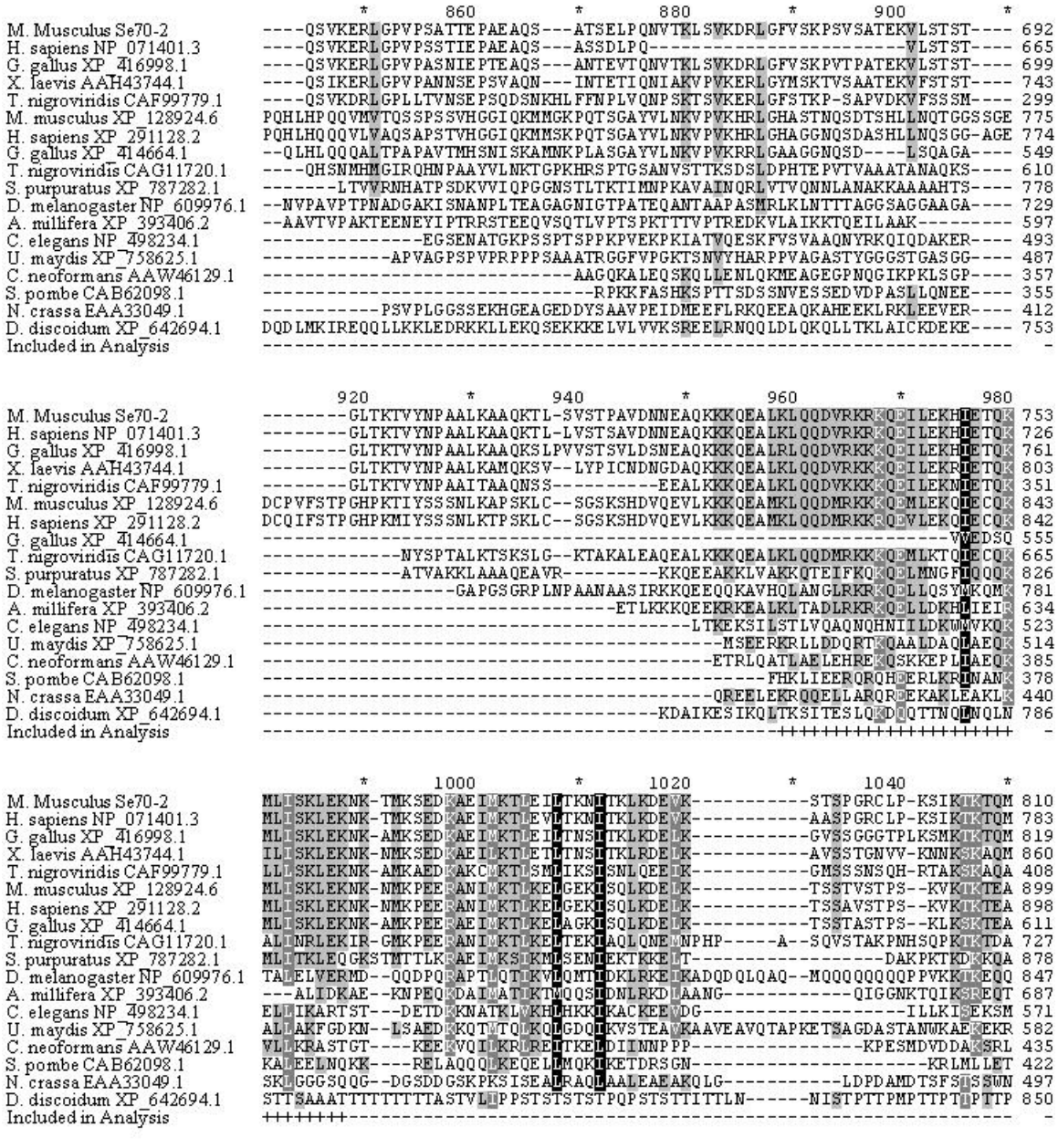
Figure 38: Multiple sequence alignment generated for mSe70-2 Homologs. Alignment of the mSe70-2 homologs was completed using ClustalX. The scoring matrix, gap opening penalties, and gap extension penalties were BLOSUM62, 11, and 1 respectively. The alignment was shaded according to conservation levels using GeneDoc, with black shading representing 95%, dark grey representing 80%, and light grey representing 60% conservation across sequences. Regions of the alignment marked with + in the "Included in analysis" track were used in the phylogenetic analysis. Areas under regions of the alignment marked with a heavy line represent predicted functional motifs.

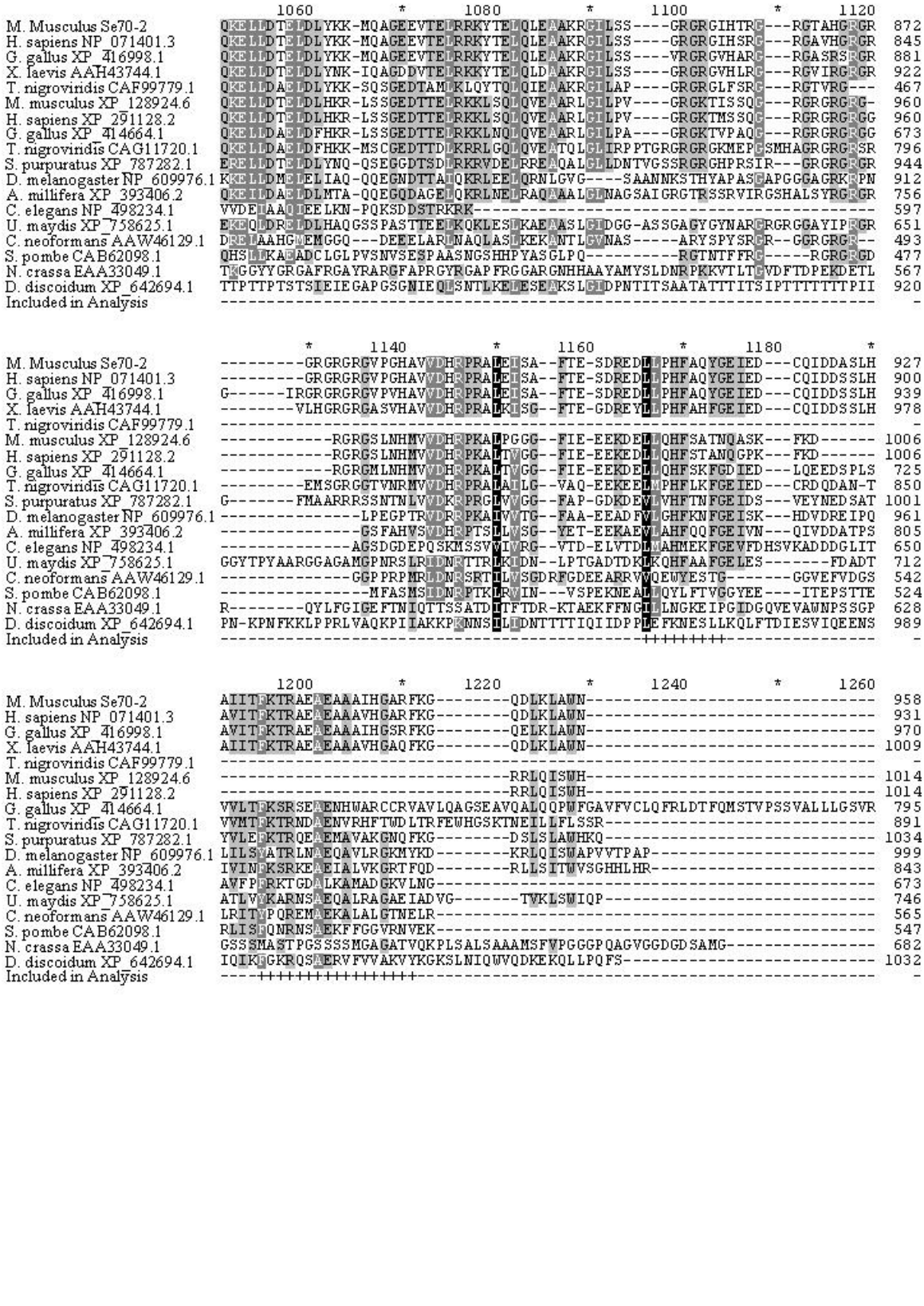

M. Musculus Se70-2
H. sapiens NP_071401.3
G. gallus XP_416998.1
X. laevis AAH43744.1
T. nigroviridis CAF99779.1
M. musculus XP_128924.6
H. sapiens XP_291128.2
G. gallus XP_414664.1
T. nigroviridis CAG11720.1
S. purpuratus XP_787282.1
D. melanogaster NP_609976.1
A. millifera XP_393406.2
C. elegans NP_498234.1
U. maydis XP_758625.1
C. neoformans AAW46129.1
S. pombe CAB62098.1
N. crassa EAA33049.1
D. discoidum XP_642694.1
Included in Analysis

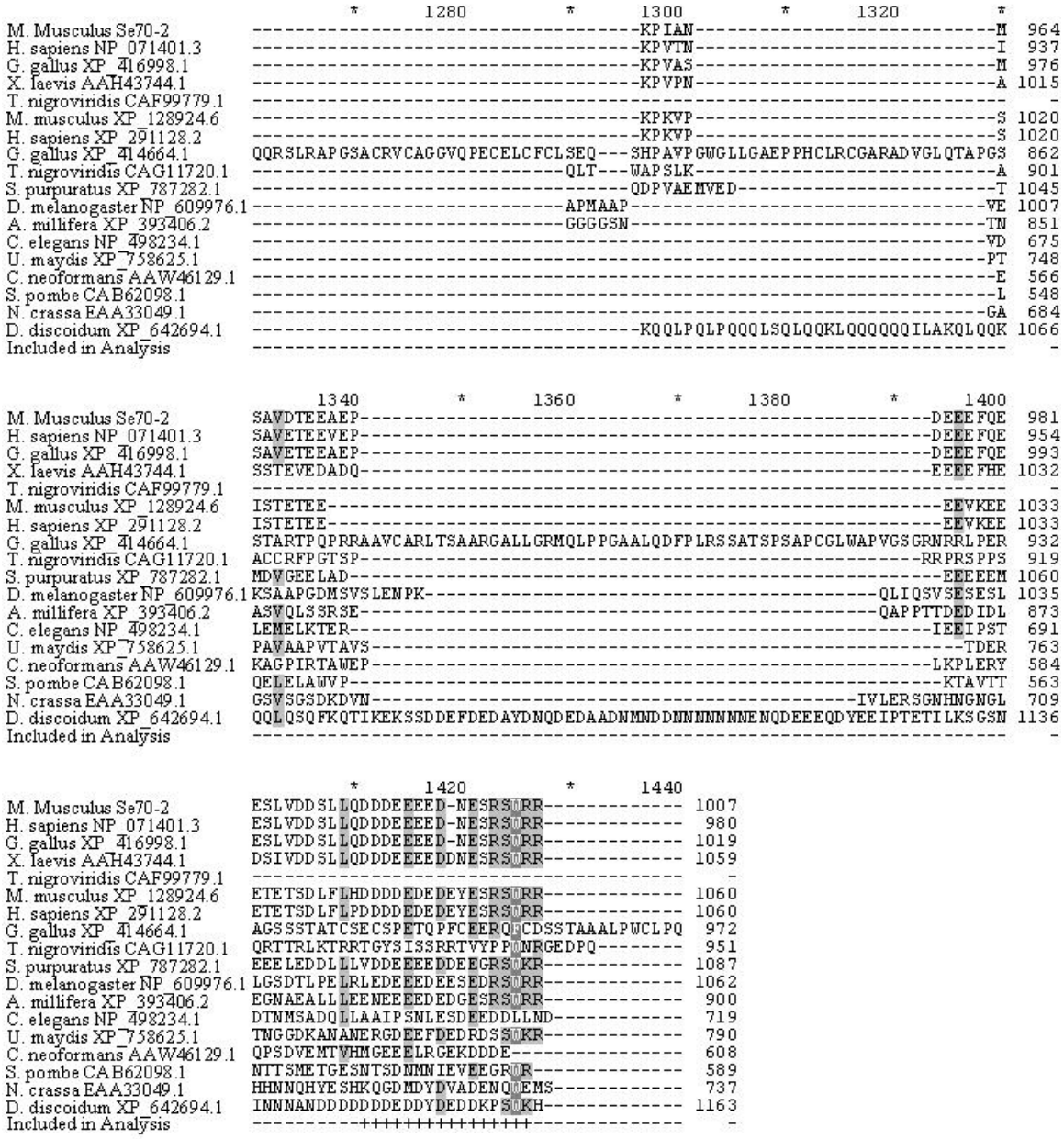

220 * 240 * 260 * 280
M. Musculus Se70-2 ----SRDERKKDDRSRKRDYDRNPPRRDSYRDRYNRRRGRSRSYSRSR-----SRSWSKERLRDRDRDRS 200
H. sapiens NP_071401.3 ----SRDERKKDDRSRKRDYDRNPPRRDSYRDRYNRRRGRSRSYSRSR-----SRSWSKERLRERDRDRS 200
G. gallus XP_416998.1 ----SRDERKKDERSRKRDYDRNPPRRDSYRDRYNRRRGRSRSYSRSR-----SRSWSKERLRDRDRDRS 200
X. laevis AAH43744.1 KNTRSRDDRKRDDRFRKREYDRNVPRRDSYRDRYNRRRGRSRSYSRSR-----SRSWSKERQRDRD---- 241
T. nigroviridis CAF99779.1 ---------- -
M. musculus XP_128924.6 ----RRTREKKRED GKWRDYERYYERNELYREKYDWRRGRSKSRSKSR-----GLSRSRSRSRGRS---- 187
H. sapiens XP_291128.2 ----RRTREKKRED GKWRDYDRYYERNELYREKYDWRRGRSKSRSKSR-----GLSRSRSRSRGRS---- 187
G. gallus XP_414664.1 ----QRTRDKRRDD GKWRDYDRYYDRSDLYRDKSSWRRGRSKSRSKSR-----GLSRSRSRSRGRS---- 192
T. nigroviridis CAG11720.1 ----------HGKSWSDSHHFKSKFDLERKDSDCYNSST-----PSGSQQPHP-------- 115
S. purpuratus XP_787282.1 ERDRMRGRDGRGRECLRRIHVKWRYKGVFHKEPKRRRRDEDRRRDDRPDRFGGPRRYGDRDRDRDRDRD 251
D. melanogaster NP_609976.1 SGVSGSGGGGPGGAGLAAGYADKENQPRDSRRRRASLRSRSRSRSRSNERAFRRSRSRDRRVNEREKTQR 274
A. millifera XP_393406.2 ----------RSRSRGRMRSRTRSRSRSWERDRRRSRSREHIRRDRERDRT 178
C. elegans NP_498234.1 ----------PLRKRISPPPSENREKEKQSRDKERRRSRSP---------- 160
U. maydis XP_758625.1 ----------SSRKRRADQDTGDAARSPPR---------- 129
C. neoformans AAW46129.1 ---------- -
S. pombe CAB62098.1 ---------- -
N. crassa EAA33049.1 ----------RRGAGFGPGGRGGRPDAG---------- 118
D. discoidum XP_642694.1 -----RDRDRDRERERDRERDRNKDRERDRERERDRERDRDRDRNRNSRERDHNDRDYNNNSSSSSSSSN 210
Included in Analysis ---------- -
* 300 * 320 * 340 *
M. Musculus Se70-2 RTRSRSRTRSRERDLVKPKYDLDRTDPLENNYTPVSSVSNISSGHYP--VPTLSSTITVIAPTHHGNNTT 268
H. sapiens NP_071401.3 RTRSRSRTRSRERDLVKPKYDLDRTDPLENNYTPVSSVPSISSGHYP--VPTLSSTITVIAPTHHGNNTT 268
G. gallus XP_416998.1 --RSRSRTRSRERDLGKPKYDMDRTDPLENNYTPVSSVTNISSGHYP--VPTLSSTITVIAPAHHGNNTT 266
X. laevis AAH43744.1 --RSRSRTRSRDKDSGKPKFDLDRPDPVDNSYASGSSVPHIGSAHFP--VPTLSSTITVITPSHHGNSSA 307
T. nigroviridis CAF99779.1 ---------- -
M. musculus XP_128924.6 --KDRDPNRNVEHR-ERSKFKSERND-LESSYVPVSAPPPSSSEQYSSGAQSIPSTVTVIAPAHHSENTT 253
H. sapiens XP_291128.2 --KDRDPNRNVEHR-ERSKFKSERND-LESSYVPVSAPPPNSSEQYSSGAQSIPSTVTVIAPAHHSENTT 253
G. gallus XP_414664.1 --KDHDANRNVEHR-ERSKFKSERND-VEGTYNPVSASASNSTEQYS-AAQSIPSAVTVIAPAHHLENTT 257
T. nigroviridis CAG11720.1 ----------PPLLPLPTPPHPFSSSSS-AGPTGAGSCPAATPAHLPDSTR 155
S. purpuratus XP_787282.1 RNRDNRDFRDRDRERDRMRGRDGRDPGRFDDRGRIRDRDLDRERDDRLKENNSKPMAISSVITVVSNDGP 321
D. melanogaster NP_609976.1 QFRNKSPPGSQTDNRHHGRRNFDRRRIGGNADDRPRFGNNKSRRSHSRSMSPERNARRNQNSPDRVQTAQ 344
A. millifera XP_393406.2 RFRNRSPTPLRSRSRS---RSLERKKID--RSERMEVDRTERIDGSPGGGTPTQDSNHGDVDMRLSTTSQ 243
C. elegans NP_498234.1 ----------RDRRGRATTSRSGRTSSSDRDRHSKRR 187
U. maydis XP_758625.1 ----------QTRRLQDNNATPSSDDSRAGRARERDSRERA 160
C. neoformans AAW46129.1 ----------VPAAASHPIPTGPAAEVAPRP 110
S. pombe CAB62098.1 ----------MPSTGMVTEEVSRYSPTSSMIPAT 110
N. crassa EAA33049.1 ----------NGYLPAYPPFSENGQLPPGFLTGMSETLAAL 149
D. discoidum XP_642694.1 RYRERERERERSISPSSSSLTTHTRKRSRSRSNSPQPLDEFNNNNNNNNNNNNNNNNNNNNNMTGNNTY 280
Included in Analysis ---------- -
360 * 380 * 400 * 420
M. Musculus Se70-2 ESWSEF-------HEDQVDHNSYVRPPM----PKKRCRDYDEKGFCMRGDMCPGDHGSDPVVVEDVNLPG 327
H. sapiens NP_071401.3 ESWSEF-------HEDQVDHNSYVRPPM----PKKRCRDYDEKGFCMRGDMCPGDHGSDPVVVEDVNLPG 327
G. gallus XP_416998.1 ESWSEF-------HEEQLDHNSYGRPPL----PKKRCRDYDEKGFCMRGDMCPGDHGSDPVVVEDVNLPG 325
X. laevis AAH43744.1 ENWPEF-------HEDQVDHSSYGRLQM----QKKRCRDYDEKGFCMRGDMCPGDHGSDPVVVEDVNLPG 366
T. nigroviridis CAF99779.1 ---------- -
M. musculus XP_128924.6 ESWSNY-------YNNHSSSNSFGRNPP----PKRRCRDYDERGFCVLGDLCQGDHGNDPLVVDEVALPS 312
H. sapiens XP_291128.2 ESWSNY-------YNNHSSSNSFGRNLP----PKRRCRDYDERGFCVLGDLCQGDHGNDPLVVDEVALPS 312
G. gallus XP_414664.1 ESWSNY-------YNNHSNPSSFGRNPP----PKRRCQDYDERGFCVLGDLCQGDHGNDPLVVDEVSLP- 315
T. nigroviridis CAG11720.1 DSWSGY-------YGAQRPDGVNKPFPPKIVSLCQRCRDYDEKGFCVRGDLCAGDHGNDPLIVDDVNLPG 218
S. purpuratus XP_787282.1 EPAQGP-------PKQGEKQQQQQQPLQR---IPQRCRDYDEKGFCMKGEFCPYDHGNDPVVVEDIDMFR 381
D. melanogaster NP_609976.1 ANLIPA-------PVAPPAPVEHPASHP-----PQRCRDFDEKGYCVRGETCPYDHGVDPVVFEGINNTS 402
A. millifera XP_393406.2 SIQSVV-------AVASSVPNSQPGFQT-----KRRCRDFDEKGYCMRGDLCPYDHGTDPVVLEDVALSR 301
C. elegans NP_498234.1 SRSRSR-------SSSGSRTPPYKSRSS-----RRRCKDFEERGYCIRGDQCPYYHGRDPVVVDENALGS 245
U. maydis XP_758625.1 ATTDGL-------DANGSQHEGGRREHR----PKQLCRDYHNKGFCPRGANCKKEHSADTMPGNADARNQ 219
C. neoformans AAW46129.1 SAGPST-------MQNGQQKDVIMQEMPVGVSGKRKCRDYHERGYCMRGANCQQEHSADMLIPTPEMMFQ 173
S. pombe CAB62098.1 NSMETN-------FNNSLPAVGKTNTFP------SQVPMMFGPPLYHPAATAPSEFMPSIPGFGNLPNPA 167
N. crassa EAA33049.1 QQQMGL-------SAPPGYPAPAQGQRR-----PARCRDYETKGYCSRGNTCLPEHGDDSIYVPPPSTSF 207
D. discoidum XP_642694.1 RNGSGANRFKRRSTNDEFNNNNNNNSFKKFPKSTRICKYIEKNGICTRGDSCNLSHDLSLLNNNNNNNNN 350
Included in Analysis ----------++++++++++++++++++++++++---------- -
Putative CCCH ZnFn


M. Musculus Se70-2
H. sapiens NP_071401.3
G. gallus XP_416998.1
X. laevis AAH43744.1
T. nigroviridis CAF99779.1
M. musculus XP_128924.6
H. sapiens XP_291128.2
G. gallus XP_414664.1
T. nigroviridis CAG11720.1
S. purpuratus XP_787282.1
D. melanogaster NP_609976.1
A. millifera XP_393406.2
C. elegans NP_498234.1
U. maydis XP_758625.1
C. neoformans AAW46129.1
S. pombe CAB62098.1
N. crassa EAA33049.1
D. discoidum XP_642694.1
Included in Analysis


M. Musculus Se70-2
H. sapiens NP_071401.3
G. gallus XP_416998.1
X. laevis AAH43744.1
T. nigroviridis CAF99779.1
M. musculus XP_128924.6
H. sapiens XP_291128.2
G. gallus XP_414664.1
T. nigroviridis CAG11720.1
S. purpuratus XP_787282.1
D. melanogaster NP_609976.1
A. millifera XP_393406.2
C. elegans NP_498234.1
U. maydis XP_758625.1
C. neoformans AAW46129.1
S. pombe CAB62098.1
N. crassa EAA33049.1
D. discoidum XP_642694.1
Included in Analysis
RRM
RRM


M. Musculus Se70-2
H. sapiens NP_071401.3
G. gallus XP_416998.1
X. laevis AAH43744.1
T. nigroviridis CAF99779.1
M. musculus XP_128924.6
H. sapiens XP_291128.2
G. gallus XP_414664.1
T. nigroviridis CAG11720.1
S. purpuratus XP_787282.1
D. melanogaster NP_609976.1
A. millifera XP_393406.2
C. elegans NP_498234.1
U. maydis XP_758625.1
C. neoformans AAW46129.1
S. pombe CAB62098.1
N. crassa EAA33049.1
D. discoidum XP_642694.1
Included in Analysis


1060 * 1080 * 1100 * 1120
M. Musculus Se70-2 872
H. sapiens NP_071401.3 845
G. gallus XP_416998.1 881
X. laevis AAH43744.1 922
T. nigroviridis CAF99779.1 467
M. musculus XP_128924.6 960
H. sapiens XP_291128.2 960
G. gallus XP_414664.1 673
T. nigroviridis CAG11720.1 796
S. purpuratus XP_787282.1 944
D. melanogaster NP_609976.1 912
A. millifera XP_393406.2 756
C. elegans NP_498234.1 597
U. maydis XP_758625.1 651
C. neoformans AAW46129.1 493
S. pombe CAB62098.1 477
N. crassa EAA33049.1 567
D. discoidum XP_642694.1 920
Included in Analysis -
* 1140 * 1160 * 1180 *
M. Musculus Se70-2 927
H. sapiens NP_071401.3 900
G. gallus XP_416998.1 939
X. laevis AAH43744.1 978
T. nigroviridis CAF99779.1 -
M. musculus XP_128924.6 1006
H. sapiens XP_291128.2 1006
G. gallus XP_414664.1 725
T. nigroviridis CAG11720.1 850
S. purpuratus XP_787282.1 1001
D. melanogaster NP_609976.1 961
A. millifera XP_393406.2 805
C. elegans NP_498234.1 650
U. maydis XP_758625.1 712
C. neoformans AAW46129.1 542
S. pombe CAB62098.1 524
N. crassa EAA33049.1 628
D. discoidum XP_642694.1 989
Included in Analysis -
1200 * 1220 * 1240 * 1260
M. Musculus Se70-2 958
H. sapiens NP_071401.3 931
G. gallus XP_416998.1 970
X. laevis AAH43744.1 1009
T. nigroviridis CAF99779.1 -
M. musculus XP_128924.6 1014
H. sapiens XP_291128.2 1014
G. gallus XP_414664.1 795
T. nigroviridis CAG11720.1 891
S. purpuratus XP_787282.1 1034
D. melanogaster NP_609976.1 999
A. millifera XP_393406.2 843
C. elegans NP_498234.1 673
U. maydis XP_758625.1 746
C. neoformans AAW46129.1 565
S. pombe CAB62098.1 547
N. crassa EAA33049.1 682
D. discoidum XP_642694.1 1032
Included in Analysis -


M. Musculus Se70-2
H. sapiens NP_071401.3
G. gallus XP_416998.1
X. laevis AAH43744.1
T. nigroviridis CAF99779.1
M. musculus XP_128924.6
H. sapiens XP_291128.2
G. gallus XP_414664.1
T. nigroviridis CAG11720.1
S. purpuratus XP_787282.1
D. melanogaster NP_609976.1
A. millifera XP_393406.2
C. elegans NP_498234.1
U. maydis XP_758625.1
C. neoformans AAW46129.1
S. pombe CAB62098.1
N. crassa EAA33049.1
D. discoidum XP_642694.1
Included in Analysis

APPENDIX G: SCIENTIFIC CONTRIBUTIONS

Everitt R, Minnema SE, Wride MA, Koster CS, Hance JE, Mansergh FC, Rancourt DE. *RED: the analysis, management and dissemination of expressed sequence tags.* Bioinformatics. 2002 Dec;18(12):1692-3.

Wride MA, Mansergh FC, Adams S, Everitt R, Minnema SE, Rancourt DE, Evans MJ. *Expression profiling and gene discovery in the mouse lens.* Molecular Vision. 2003 Aug 22;9:360-96.

Nicholls CD, Minnema SE, Rancourt DE, Beattie TL. *The alternative splicing of PIG3 exon 4 is mediated by multiple NAG elements near the 3' splice site.* In preparation.