Title
Evidence for L1 retrotransposition in the human nervous system

Permalink
https://escholarship.org/uc/item/1fk4891f

Author
Coufal, Nicole Gabriele

Publication Date
2008

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Evidence for L1 Retrotransposition in the Human Nervous System

A Dissertation submitted in partial satisfaction of the

Requirements for the Degree of Doctor of Philosophy

in

Neurosciences

by

Nicole Gabriele Coufal

Committee in charge:

    Professor Fred H. Gage, Chair
    Professor Lawrence Goldstein, Co-chair
    Professor Anirvan Ghosh
    Professor Leanne Jones
    Professor Adjit Varki

2008

The Dissertation of Nicole Gabriele Coufal is approved, and it is acceptable in quality
and form for publication on microfilm:

_____

_____

_____

_____
Co-Chair

_____
Chair

University of California, San Diego

2008

DEDICATION

To my father
Hans Juergen Coufal
(1945-2006)

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

LINE or L1 – Long Interspersed Nuclear Element, ubiquitous human transposons which replicate through an RNA intermediate.

hESC – human embryonic stem cells, cells which are cultured from the inner blastocyst mass of the human embryo and continue to proliferate in culture, which are totipotent and therefore capable of differentiating to all three germ lineages.

NSC – Neural Stem Cells; cells which are committed to a neural lineage but continue to proliferate. They can differentiate into neurons or astrocytes (and perhaps oligodendrocytes).

HUES6 - Human Embryonic Stem (HUES) Cell Collection, human embryonic stem cell line number 6 from the HUES cell collection, laboratory of Doug Melton, Harvard.

AT – Ataxia Telangiectasia; an autosomal recessive disease.

ATM - ataxia telangiectasia mutated protein, a mutation in ATM results in ataxia telangiectasia.

UTR – Untranslated region, the portion of a gene either 3' (before) or 5' (after) the coding region which is also transcribed to RNA.

ORF1 and **ORF2** – Open reading frame 1 and Open reading frame 2, the two proteins encoded for and expressed by the L1 retrotransposon.

qPCR – quantitative polymerase chain reaction, a real time measurement of nucleic acid content.

DAPI – 4',6-diamidino-2-phenylindole; intercalates between nucleotide bases to stain cell nuclei

SOX2 – the transcription factor Sry-related homeobox (HMG-box) 2, used as a marker of neural stem cells and embryonic stem cells.

SOX1 - the transcription factor Sry-related homeobox (HMG-box) 1, used as a marker of neural stem, one of the earliest transcription factors to be expressed in ectodermal cells committed to the neural lineage.

GFAP – Glial fibrillary acidic protein; an intermediate filament protein that is found in glial cells such as astrocytes.

Map2a+2b – Microtubule associated protein 2a and 2b; encodes a protein that belongs to the microtubule associated protein family and is expressed in dendrites and cell bodies of neurons.

PFA – paraformaldehyde

Chr - Chromosome

Tuj1 - neuron specific beta (β) III Tubulin, a marker for neurons.

CpG – a cytosine followed 3' by a guanine base, connected by a phosphate group. The cytosine is a methylation target.

TSD – target site duplication, a duplication of 3-20 bases of genomic DNA at the LINE insertional site.

TPRT – target primed reverse transcription, the mechanism by which an L1 inserts into a new genomic location.

RNP – ribonucleoprotein complex; the complex of L1 RNA associated with ORF1 and ORF2 proteins.

DSB – double strand break

HCNS-SCns – human central nervous system stem cells grown as neurospheres.

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

VITA

<u>EDUCATION</u>

University of California San Diego School of Medicine, La Jolla, CA
M.D. estimated 2011.

University of California San Diego Graduate Neuroscience Program, La Jolla,
CA. PhD 2008.

University of California Berkeley, Berkeley, CA.
BA in Molecular and Cellular Biology, emphasis in Neuroscience.
BA in Cognitive Science, 2000

<u>RESEARCH EXPERIENCE</u>   DATE

Salk Institute for Biological Studies   2004-present
        Graduate student in Neurosciences
        Thesis Advisor: Fred Gage
        Graduate school GPA: 4.0/4.0

University of California, San Diego   2004
        Research rotation in the lab of Dr. Leon Thal

Salk Institute for Biological Studies   2003
        Research rotation in the lab of Dr. Greg Lemke

Salk Institute for Biological Studies & UC San Diego   2000-2002
        Research Assistant I
        Advisors: Fred Gage & Leon Thal

University of California, San Francisco   1999-2000
        Student Researcher in the lab of Dr. Lawrence Tecott

University of California, San Francisco   1997-1999
        Student Researcher in the lab of Dr. Steven McIntire

University of California, San Francisco   1997
        Lab Assistant I, Lung Biology Department

<u>HONORS & AWARDS</u>

UCSD Neurosciences Graduate, GPA 4.0   2004-2008
UCSD School of Medicine, Honors   2002-2003

Honors in Neurology, Physiology, Pharmacology, Pathology, Laboratory Medicine, Histology, Epidemiology/Biostatistics, Endocrinology & Reproductive Medicine.

International Center Fellowship – Clinical work abroad	2003, 2004
High Honors in Cognitive Science, UC Berkeley	2000

Honor's Thesis Title: *Alterations in Cocaine Responses in 5HT2CR Null Receptor Mice*

Regent's and Chancellor's Scholar, UC Berkeley	1996-2000
Stanley Foundation Fellowship, UC San Francisco	1999
Dean's List, College of Arts & Sciences	1999, 2000
Thomas J. Watson Scholar, I.B.M.	1996-2000

PROFESSIONAL MEMBERSHIPS

Society for Neuroscience	2001-present
American Medical Association Student Member	2002-present
American Association for the Advancement of Sciences	2004-present

TEACHING EXPERIENCE

University of California, San Diego, Neurology   2005
Teaching Assistant for Medical School neurology laboratory

University of California, San Diego, Neurology   2006, 2007, 2008
Teaching Assistant for Medical School neurology clinical cases
Prof. Mark Kritchevsky

PUBLICATIONS

Yeo GW, Xu X, Liang TY, Muotri AR, Carson CT, Coufal NG, Gage FH. Alternative splicing events identified in human embryonic stem cells and neural progenitors. PLoS Comput Biol. 2007 Oct;3(10):1951-67.

Muotri AR, Marchetto MC, Coufal NG, Gage FH. The necessary junk: new functions for transposable elements. Hum Mol Genet. 2007 Oct 15;16 Spec No. 2:R159-67.

Garcia-Perez JL, Marchetto MC, Muotri AR, Coufal NG, Gage FH, O'Shea KS, Moran JV. LINE-1 retrotransposition in human embryonic stem cells. Hum Mol Genet. 2007 Jul 1;16(13):1569-77.

Coufal, NG and Farnaes L. "Vinca Alkaloids" In: Cancer Growth and Progression Book Series, Springer, In Press, 2008

Coufal, NG and Farnaes L. "Anthracyclines and Anthracenediones" In: Cancer Growth and Progression Book Series, Springer, In Press, 2008

Pizzo DP, Coufal NG, Lortie MJ, Gage FH, Thal LJ. Regulatable Acetylcholine-Producing Fibroblasts Enhance Cognitive Performance. Mol Ther. 2006 Jan;13(1):175-82.

Singer O, Marr RA, Rockenstein E, Crews L, Coufal NG, Gage FH, Verma IM, Masliah E. Targeting BACE1 with siRNAs ameliorates Alzheimer disease neuropathology in a transgenic model. Nat Neurosci. 2005 Oct;8(10):1343-9.

Pizzo DP, Paban V, Coufal NG, Gage FH, Thal LJ. Long-term production of choline acetyltransferase in the CNS after transplantation of fibroblasts modified with a regulatable vector. Brain Res Mol Brain Res. 2004 Jul 5;126(1):1-13.

Brandon EP, Mellot T, Pizzo DP, Coufal N, D'Amour KA, Gobeske K, Lortie M, Lopez-Coviella I, Berse B, Thal LJ, Gage FH, Blusztajn JK. Choline transporter 1 maintains cholinergic function in choline acetyltransferase haploinsufficiency. J Neurosci. 2004 Jun 16;24(24):5459-66.

Rocha BA, Goulding EH, O'Dell LE, Mead NA, Coufal NG, Parsons LH, Tecott LH, Enhanced locomotor, reinforcing, and neurochemical effects of cocaine in serotonin 5-hydroxytryptamine 2C receptor mutant mice. J. Neuroscience; Nov.15, 2002. 22(22): 10039-10045.

POSTER PRESENTATIONS

Gordon Conference March 2008.
Farnaes L, Coufal NG, Kauffman CA, Jensen PR, Fenical W. Napyradiomycin Redux.

Mobile Elements in Mammalian Genomes, FASEB. June 2007.
Coufal NG, Peng GE, Carson CT, Muotri AR, Gage FH. L1 Retrotransposition in human neural stem cells.

Young Neuroscientists Workshop, Children's Neurobiological Solutions (CNS) Foundation. August 2007.
Coufal NG, Peng GE, Carson CT, Muotri AR, Gage FH. L1 Retrotransposition in human neural stem cells.

12th International C. elegans conference, June 1999.
Davis AG, Eastman CL, Coufal NG, Kim HK, McIntire SL; Mutants with altered sensitivity to the intoxicating effects of ethanol.

COURSES ATTENDED

Frontiers in Human Embryonic Stem Cells at Stanford University by Gerald Schatten and Roger Pederson June 2005.

VOLUNTEER

Clinic Manager - UCSD Free Clinic School of Medicine, University of California, San Diego	2003-present
Manager, UCSD Medical School Homeless Outreach Project	2003-present
Editor in Chief - The Thinker: Undergraduate Journal of Cognitive Science University of California, Berkeley	1998-2000
Judge – San Diego Science Fair	2005, 2006

PERSONAL INFORMATION

Citizenship: USA
Languages: English (fluent), German (conversational)
Experimental: molecular biology, cloning, human primary tissue culture, human ES cell tissue culture, RNA and DNA preparation, RT-PCR and qPCR, Animal handling, surgical techniques including intracranial grafting and *in utero* injections, tissue preparation and histology, basic microscopy.
Activities: competitive horseback riding, yoga, kickboxing, piano, cooking, backpacking, traveling, medicine abroad.

ABSTRACT OF THE DISSERTATION

Evidence for L1 Retrotransposition in the Human Nervous System

by

Nicole Gabriele Coufal

Doctor of Philosophy in Neurosciences

University of California, San Diego, 2008

Professor Fred H. Gage, Chair
Professor Lawrence Goldstein, Co-chair

This work explores whether human transposons from the L1 family of retrotransposons are active in human neural stem cells derived from various sources. This work explores both the characteristics of cells which allow L1 retrotransposition and the insertional characteristics of the L1 within those cells. Not only did human neural stem cells allow for retrotransposition, they reproducibly exhibited robust rates of insertion. In these studies I also develop a technique for analyzing the copy number of endogenous L1s and discover evidence for endogenous retrotransposition in the human brain, notably in the hippocampus. To begin to investigate the mechanism of L1 retrotransposition, methylation analysis of the L1 promoter within the 5' untranslated region (UTR) was undertaken using human developmental tissues. The outcome indicated decreased methylation of the L1 promoter in the brain as compared to the skin across many genomic L1s. Moreover, this work seeks to investigate how

L1 retrotransposition in the brain might be affected in a neurological disease process. Ataxia telangiectasia (AT) is an autosomal recessive disorder characterized by neurological degeneration and a mutation in the ATM (ataxia telangiectasia mutated) gene. In these studies we found that a transgenic mouse model of AT exhibits increased retrotransposition, as do human neural stem cells in the absence of ATM. These findings further support earlier studies showing L1 retrotransposition in the mouse brain and contribute to our understanding of the activity of L1 in nervous system. Furthermore, they propose that L1 retrotransposition may be aberrantly increased in the context of neurological disease. Finally, this work suggests multiple additional research pathways for investigating the mechanism and potential consequences of L1 retrotransposition in the nervous system.

CHAPTER 1

INTRODUCTION

In the course of evolution, mobile gene elements known as transposons have propagated in the genomes of creatures as diverse as humans and corn. There has been a widespread scientific belief that transposons or mobile elements are junk; selfish or parasitic DNA of no importance for health, development, or disease, existing merely to propagate themselves. To some extent, this reputation is deserved as mobile elements collectively inhabit at least 45% and 37% of human and mouse genomes, respectively(Lander et al., 2001; Waterston et al., 2002). These elements have participated in shaping our genome, increasing genome plasticity and providing fertile soil for evolution to act (Babushok et al.). Newer data supports the theory that transposons may move in a highly restricted subset of cells during a narrow window of development, and may take advantage of a situation in which DNA damage is occurring (Athanikar et al.). Mammalian transposons fall into two large families: DNA transposons and retrotransposons. DNA transposons move through a cut-and-paste transposase mechanism from one genomic location to another. They represent only approximately 3% of the human genome, and have largely accumulated mutations, rendering them immobile (Smit and Riggs;Deininger et al., 2003). Retrotransposons overall comprise the majority of mammalian transposable elements, they constitute 42% of the genome and move through an RNA intermediate. Retrotransposons encode a reverse transcriptase allowing for a second novel insertion while the original transposon is preserved. By far the vast majority of retrotransposons in the genome are inactive or degenerate.

Retrotransposons consist of autonomous elements, which encode certain proteins necessary for transposition, and non-autonomous elements which require

additional proteins hijacked from other retrotransposons for insertional ability. Among the autonomous retrotransposons there is evidence that only a single sub-family, the Long Interspersed Nuclear Elements (LINE or L1) retain mobility in humans (Sassaman et al.).

L1s are highly abundant retrotransposons, and constitute approximately 17% of the genome (Gibbs et al., 2004; Lander et al., 2001; Waterston et al., 2002). Most L1s are retrotransposition-defective because they are 5' truncated, contain internal rearrangements, or harbor mutations within their open reading frames(Grimaldi et al., 1984). There are approximately 11,000 within the genome that are full-length (approximately 6,000 bp,) and of these roughly 151 contain two open reading frames and are considered retrotransposition-competent (Penzkofer et al.). By comparison, the mouse genome is estimated to contain at least 3,000 active L1s (Goodier et al., 2000; DeBerardinis et al., 1998). Human L1s structurally resemble those present in rodent genomes, and a recently developed retrotransposition assay has revealed that human L1s can retrotranspose in a variety of mammalian cell lines (Han et al., 2004; Moran et al., 1996; Morrish et al., 2002; Wei et al., 2000). Eighty-four full length retrotransposition competent L1s were cloned and investigated as to their activity in a cultured cell assay, and it was found that approximately 10% of these elements are classified as highly active or "hot" and likely represent the majority of L1 insertional ability (Brouha et al., 2003). L1 insertions can have varying downstream effects depending on their genomic location, including up- or down-regulation of nearby genes (Babushok and Kazazian ; Chen et al.). Germline retrotransposition events are predicted to occur in approximately one per 80 individuals (Deininger et al.).

Retrotransposition-competent L1s encode the machinery necessary to support their own replication through an RNA intermediate. L1s contain a 5' untranslated region (UTR) with internal promoter activity, two open reading frames (ORF1 and ORF2) and a 5' UTR followed by a polyadenylation signal (AATAAAA) (see Fig. 1.1). The 5'UTR includes binding sites for a number of regulatory elements, including a CpG island, two SRY box binding domains, a YY-box transcriptional initiation start site, and a RUNX binding site(Becker et al., ; Kurose et al., ; Tchenio et al.). ORF1 encodes for an RNA chaperone protein which is necessary for retrotransposition activity. ORF2 encodes a protein that has both an endonuclease domain (EN) which nicks DNA at the insertional target site, and a reverse transcriptase (RT) domain which reverse transcribes the L1 into the *de novo* location. Although L1s encode the majority of protein necessary for retrotransposition, it is doubtful that they are stringently autonomous, as some host factors such as DNA repair enzymes are likely also required (Athanikar et al.).

The mechanism of L1 retrotransposition requires a number of steps to result in a *de novo* insertion (Figure 1.1). First the L1 is transcribed from its own internal promoter, beginning at the YY-1 transcriptional start site, and generating a bicistronic mRNA. The mRNA exits the nucleus through an unknown mechanism. In the cytoplasm ORF1 and ORF2 proteins are translated and exhibit *cis* preference, such that they preferentially associate with their mRNA of origin. The formation of the ribonucleoprotein (RNP) complex is believed to involve multiple copies of ORF1 and only a few copies of ORF2 associated with a given L1 mRNA. Through a mechanism which likely involves nuclear breakdown or cell division, the RNP complex regains

entry to the nucleus. Subsequently the ORF2 endonuclease nicks the DNA and the L1 RNA is reverse transcribed in a new genomic location through a process called target primed reverse transcription (TPRT). Insertion begins with the polyA tail and proceeds from 3' to 5'. Insertions are frequently 5' truncated, although the mechanism resulting in truncation is poorly understood. Frequently the process of TPRT results in a 3-20bp target site duplication (TSD) (Moran and Gilbert).

Retrotransposition can affect the genome in a variety of ways (Speek, 2001; Weiner, 2002)·(Kashkush et al., 2003). The L1 5'UTR region (approximately 1,000 bp) encompasses both a sense and an antisense promoter which can affect nearby gene expression(Speek). Deleterious retrotransposition events in the germ line or during development have resulted in a variety of genetic disorders (Kazazian, 1998; Ostertag and Kazazian, 2001a). There have been at least 35 reported sporadic insertions of retrotransposons leading to human disease as diverse as neurofibromatosis and hemophilia (Ostertag and Kazazian). A single documented human somatic L1 retrotransposition has resulted in a sporadic case of colon cancer, which was clearly somatic as neighboring colonic tissue did not harbor the insertion (Miki et al.). L1 retrotransposition events in or near genes may also act to fine-tune gene expression (Belancio et al., 2006; Han et al., 2004). L1-encoded proteins can also function *in trans* to mobilize non-autonomous retrotransposons (e.g., Alu elements) and cellular mRNAs. Mobilization of cellular mRNAs has lead to the generation of processed pseudogenes, mRNAs which have been reinserted into the genome without their introns or other regulatory regions (Dewannieux et al., 2003; Esnault et al., 2000; Wei

et al., 2001). L1-mediated retrotransposition events can also lead to the genesis of new genes both *in vitro* and *in vivo* (Moran, 1999; Sayah et al., 2004).

Until recently, active retrotransposition was postulated to occur rarely in germ cells and to be almost unheard of in somatic tissues (Branciforte and Martin). L1 retrotransposons were thought to be heavily methylated, which represses transcription, in all tissues except germ cells. A transgenic mouse model, in which a human L1 is driven by its endogenous promoter, was discovered to support retrotransposition events occurring in late-meiotic and post-meiotic male germ cells but not in other tissues (Ostertag and Kazazian). A report of active retrotransposition in human somatic non-dividing cells used a retrotransposon which has been altered in two ways: it was driven by the ubiquitously active PGK-1 promoter and was engineered to have increased activity (Kubo et al.). However, evidence is emerging that L1 retrotransposons may be weakly active in other cell types. For instance, L1 retrotransposition has recently been described to occur at a low rate in human embryonic stem cells (Garcia-Perez et al.). There has also been a single report of a low level of retrotransposition in primary fibroblasts lines (Shi et al.).

The recent report from Muotri and colleagues found that rat neural stem cells (NSCs) and the developing mouse nervous system can support retrotransposition of a human L1 driven by its own internal promoter providing the first report of somatic retrotransposition of an unaltered L1 in somatic tissues (Muotri et al.). The authors did not find retrotransposition in any other rodent somatic tissue. They also found that the L1 5'UTR was activated during differentiation and inhibited by the SRY-box-2 (SOX2) transcription factor, a factor which is necessary for self-renewal of stem cells,

was found bound to the L1 5'UTR. All the clones selected for retrotransposition remain multipotent, indicating that L1 can retrotranspose during early stages of neuronal differentiation. In addition, they found that retrotransposon mobilization in NSC's can influence not only gene expression but also cell fate. They showed that retrotransposition of an engineered human L1 into the neuronal specific *Psd-93* gene can lead to its over-expression, which influenced cell fate and differentiation pattern of NSCs in culture to a primarily neuronal fate. These data suggest that if L1 retrotransposition does occur in the nervous system, it has the potential to influence gene expression and cell fate (Muotri et al.). This raises the question whether retrotransposition during neural stem cell differentiation contributes to the generation of variability within the nervous system.

NSCs are self-renewing, multipotent cells that are able to generate neurons, astrocytes, and oligodendrocytes in the central nervous system. These stem cells exist not only during development of the nervous system but also in discrete areas of the adult brain such as the ventricular zone and the subgranular zone of the dentate gyrus of the hippocampus (Gage, 2000; Reynolds and Weiss, ; Vescovi et al.). Mitogens such as epidermal and fibroblast growth factors (EGF and FGF-2, respectively) influence the extent and rate of proliferation of neural stem and progenitor cell populations *in vitro* and *in vivo* (Gage et al., 1995; Kuhn et al., 1997; Palmer et al., 1999; Palmer et al., 1995; Reynolds and Weiss, 1992; Roy et al., 2000; Shihabuddin et al., 1997; Vescovi et al., 1993). During brain development, an excess of neural cells is produced; the cells migrate toward their target, where a limited supply of trophic factors, produced by the target cells, regulates their survival (Oppenheim). The post-

mitotic development of most neuronal populations in vertebrate animals is characterized by a period of cell death during which 40-60% of the neurons are eliminated (Oppenheim, 1991). The degree of cell death has been shown to be controlled by the target tissues that the neurons innervate, via the restricted availability of neurotrophic factors produced by target cells. Neurogenesis is regulated by a variety of stimuli, including steroid hormones, aging, environmental enrichment, genetic background, stresses, and physical activity (Gould et al., 1992; Gould et al., 1998; Kempermann et al., 1997a, b; Kuhn et al., 1996; van Praag et al., 1999). However, the molecules and mechanisms controlling neural stem or progenitor cell proliferation remain only partially understood.

Single cell culture studies show that EGF and FGF-2 are mitogens for rodent-derived neurospheres (Reynolds and Weiss, 1996 ; Gritti et al., 1996). Previous studies investigate the establishment of long-term cultures, in the presence of FGF-2, of adult rat hippocampus-derived neural progenitor cells (rat NSCs) (Gage et al., 1995) and also of adult rat neural progenitor cells derived from other areas of the adult brain (Palmer et al., 1999; Palmer et al., 1995; Shihabuddin et al., 1997). Cloning of rat NSCs has shown that all three lineages are generated from single genetically marked cell and that the cloned NSCs display stem cell properties, including the capacity for self-renewal and multi-lineage differentiation (Palmer et al., 1995). Grafting of rat NSCs in the brain results in integration and specification into terminally differentiated neurons appropriate for the area into which they are grafted (Gage et al., 1995; Suhonen et al., 1996).

Although cultured NSCs from the adult rodent brain has long been routine (Palmer et al., ; Reynolds and Weiss), it has been more challenging to isolate long-term proliferating cultures from the human brain (Palmer et al.). Cultures from Human Central Nervous System Stem Cell neurospheres (HCNS-SCns) are more robust than adult NSC cultures and can be cultured in the presence of FGF-2 and EGF (Svendsen et al.). Addition of leukemia inhibitor factor (LIF) to the FGF-2/EGF containing growth media results in cultures that can be maintained up to 30 passages. LIF has recently been shown to maintain telomere length and prevent senescence in HCNS-SCns (Wright et al.). Upon transplantation into immunosuppressed animals, cultured human neurospheres can engraft, migrate, and exhibit site-specific differentiation similar to rodent NSCs (Uchida et al. ; Svendsen et al.).

Neurospheres, non-adherent spherical cultures of primary neural stem cells, from both rodents and humans consist of a mixture of multipotent stem cells and more restricted progenitors and are considered to comprise a heterogeneous populations of NSC's (Reynolds and Weiss). Previous studies have shown that only one in 200 HCNS-SCns is a new "sphere forming cell" which is capable of both self-renewal and differentiation into multipotent lineages (Uchida et al.). Uchida and colleagues have identified a number of surface cell markers and have coupled them to fluorescence-activated cell sorting (FACs) to isolate a population of HCNS-SCns which are CD133+, 5E12+, CD34-, and CD45- (Uchida et al.). These cells can initiate new neurosphere cultures at a much higher frequency than previously possible (1 in 30 cells are sphere forming), differentiate into both neurons and astrocytes *in vitro* (Uchida et al.), and differentiate robustly when grafted into the immunosuppressed

rodent brain and spinal cord (Kelly et al., ; Tamaki et al., ; Uchida et al.). Although grown as a neurosphere culture, they are relatively more homogeneous (based on marker expression and differentiation potential) than cells isolated directly and grossly from the fetal brain.

Even though HCNS-SCns represent a vital tool, their culture and expansion potential is limited. The potential to generate virtually any differentiated cell type from embryonic stem cells (hESCs), including neural progenitors and differentiated neurons, offers both a second model for neuronal differentiation and the possibility of an easily renewable resource for NSCs. There are two major methods to differentiate hESCs into NSCs: co-culturing hESCs with a stromal feeder layer such as PA6 (Kawasaki et al.) and MS5 (Perrier et al.), or isolating NSC-containing neuroectoderm from embryoid bodies (EBs). Methods have been developed to enrich and promote the neuroectodermal population in EBs, including incubating EBs in a defined neural promoting media (Zhang et al.), in conditioned media from stromal HepG2 cells (Schulz et al.) or in the presence of the BMP antagonist Noggin (Itsykson et al., ; Pera et al., ; Sonntag et al.). With the exception of complete neural differentiation by co-culture on PA6 cells, nearly all of these methods converge at the manual isolation of neural rosettes that can be enriched and expanded as NSCs in the presence of FGF2. Rosettes consist of radial arrangements of columnar cells that express many markers also found in the developing neural tube. NSCs from rosettes can be propagated as neurospheres (Zhang et al.) or as adherent cultures (Yeo et al.) for multiple passages before further differentiation into neurons and glia. Moreover, engraftment of hESC-derived NSCs into rodents can result in functional neuronal integration (Uchida et al., ;

Zhang et al.). Together, these achievements have demonstrated that hESC-derived NSCs have utility in both *in vitro* and *in vivo* model systems, making them a viable model in which to study L1 retrotransposition.

Previous studies have indicated that L1 retrotransposition can occur in germ cells or in early embryogenesis, before the germ line becomes a distinct lineage (Ostertag et al., 2003; Prak et al., 2003), and studies using a retrotransposition assay has revealed that human and mouse L1 elements can retrotranspose in a variety of transformed or immortalized cultured cell lines (Han and Boeke, 2004; Moran et al., 1996; Morrish et al., 2002). Data from Muotri and colleagues show that rat NSCs, unlike a variety of other rodent somatic cell types tested, can support retrotransposition of engineered human L1 elements. These data provide proof in principle that new L1 retrotransposition events can modulate the expression of neuronal genes. Their data also indicate that a human L1 element can undergo somatic retrotransposition in the mouse brain and suggest that mammalian neurons may be mosaic with respect to L1 content. Notably, their experiments only followed the retrotransposition of a single L1 element, whereas the average human and mouse genomes contain approximately 150 and 3,000 retrotransposition-competent L1 elements, respectively (Brouha et al., 2003; Goodier et al., 2000). If a single L1 element can retrotranspose in at least one in every 100 developing neurons, and 35% of the resultant insertions occur into genes expressed during neuronal development, it is possible that brain development could be significantly affected by L1 retrotransposition (Muotri and Gage).

Therefore, the central hypothesis of this research is that retrotransposon mobility can occur in the human nervous system. This work seeks to address this question using different systems and different methods to build a body of evidence toward confirming the hypothesis.

- To demonstrate L1 retrotransposition of a tagged human L1 in both human fetal NSCs and in hESC derived NSCs using multiple different methods.

- To examine the copy number of endogenous L1s in the human brain as opposed to other somatic tissues to determine if there is evidence for endogenous L1 retrotransposition. Additionally, to begin to investigate a possible mechanism allowing retrotransposition to occur in the human brain.

- To determine if there are changes in brain levels of L1 retrotransposition in the context of neurological disease, specifically using a model of Ataxia Telangiectasia.

Figure 1.1
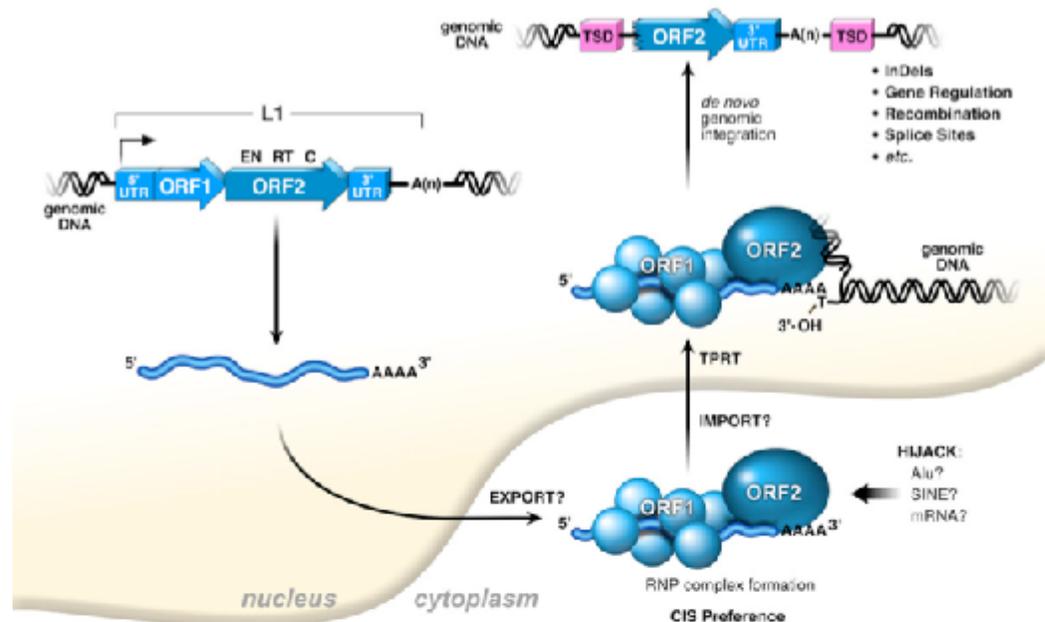
Figure 1.1
Schematic of the mechanism of L1 retrotransposition.
A full length L1 with two open reading frames is transcribed from its internal promoter
beginning at the YY-1 transcriptional start site and resulting in a bicistronic mRNA. Through
an unknown mechanism the L1 mRNA is exported from the nucleus. In the cytoplasm the
ORF1 and ORF2 proteins are translated and associate with the RNA, exhibiting cis prefer-
ence for their RNA of origin. Current evidence indicates that many copies of ORF1 and only
a few copies of ORF2 protein associate with the RNA to form the ribonucleoprotein com-
plex (RNP). Through a mechanism which likely involves nuclear membrane breakdown or
cell division the RNP complex reenters the nucleus, where the RNA is reverse transcribed
and integrated into the new genomic location by a process called target primed reverse tran-
scription (TPRT). The insertion proceeds from 3' to 5' beginning with the polyA tail. There-
fore, due to low processivity of the L1 RT, the insertion is frequently 5' truncated, resulting
in an inactive copy. The TPRT process frequently results 3-20 base target site duplication
(TSDs) on either side of the L1.

WORKS CITED

Athanikar, J.N., Morrish, T.A., and Moran, J.V. (2002). Of man in mice. Nat Genet *32*, 562-563.

Babushok, D.V., and Kazazian, H.H., Jr. (2007). Progress in understanding the biology of the human mutagen LINE-1. Human mutation *28*, 527-539.

Babushok, D.V., Ostertag, E.M., and Kazazian, H.H., Jr. (2007). Current topics in genome evolution: molecular mechanisms of new gene formation. Cell Mol Life Sci *64*, 542-554.

Becker, K.G., Swergold, G.D., Ozato, K., and Thayer, R.E. (1993). Binding of the ubiquitous nuclear transcription factor YY1 to a cis regulatory sequence in the human LINE-1 transposable element. Hum Mol Genet *2*, 1697-1702.

Belancio, V.P., Hedges, D.J., and Deininger, P. (2006). LINE-1 RNA splicing and influences on mammalian gene expression. Nucleic Acids Res *34*, 1512-1521.

Branciforte, D., and Martin, S.L. (1994). Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. Mol Cell Biol *14*, 2584-2592.

Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V., and Kazazian, H.H., Jr. (2003). Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci U S A *100*, 5280-5285.

Chen, J.M., Stenson, P.D., Cooper, D.N., and Ferec, C. (2005). A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. Human genetics *117*, 411-427.

DeBerardinis, R.J., Goodier, J.L., Ostertag, E.M., and Kazazian, H.H., Jr. (1998). Rapid amplification of a retrotransposon subfamily is evolving the mouse genome. Nat Genet *20*, 288-290.

Deininger, P.L., Moran, J.V., Batzer, M.A., and Kazazian, H.H., Jr. (2003). Mobile elements and mammalian genome evolution. Curr Opin Genet Dev *13*, 651-658.

Dewannieux, M., Esnault, C., and Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. Nat Genet *35*, 41-48.

Esnault, C., Maestre, J., and Heidmann, T. (2000). Human LINE retrotransposons generate processed pseudogenes. Nat Genet *24*, 363-367.

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

Gage, F.H., Coates, P.W., Palmer, T.D., Kuhn, H.G., Fisher, L.J., Suhonen, J.O., Peterson, D.A., Suhr, S.T., and Ray, J. (1995). Survival and differentiation of adult neuronal progenitor cells transplanted to the adult brain. Proc Natl Acad Sci U S A *92*, 11879-11883.

Garcia-Perez, J.L., Marchetto, M.C., Muotri, A.R., Coufal, N.G., Gage, F.H., O'Shea, K.S., and Moran, J.V. (2007). LINE-1 retrotransposition in human embryonic stem cells. Hum Mol Genet *16*, 1569-1577.

Gibbs, R.A., Weinstock, G.M., Metzker, M.L., Muzny, D.M., Sodergren, E.J., Scherer, S., Scott, G., Steffen, D., Worley, K.C., Burch, P.E., Okwuonu, G., Hines, S., Lewis, L., DeRamo, C., Delgado, O., Dugan-Rocha, S., Miner, G., Morgan, M., Hawes, A., Gill, R., Celera, Holt, R.A., Adams, M.D., Amanatides, P.G., Baden-Tillson, H., Barnstead, M., Chin, S., Evans, C.A., Ferriera, S., Fosler, C., Glodek, A., Gu, Z., Jennings, D., Kraft, C.L., Nguyen, T., Pfannkoch, C.M., Sitter, C., Sutton, G.G., Venter, J.C., Woodage, T., Smith, D., Lee, H.M., Gustafson, E., Cahill, P., Kana, A., Doucette-Stamm, L., Weinstock, K., Fechtel, K., Weiss, R.B., Dunn, D.M., Green, E.D., Blakesley, R.W., Bouffard, G.G., De Jong, P.J., Osoegawa, K., Zhu, B., Marra, M., Schein, J., Bosdet, I., Fjell, C., Jones, S., Krzywinski, M., Mathewson, C., Siddiqui, A., Wye, N., McPherson, J., Zhao, S., Fraser, C.M., Shetty, J., Shatsman, S., Geer, K., Chen, Y., Abramzon, S., Nierman, W.C., Havlak, P.H., Chen, R., Durbin, K.J., Egan, A., Ren, Y., Song, X.Z., Li, B., Liu, Y., Qin, X., Cawley, S., Cooney, A.J., D'Souza, L.M., Martin, K., Wu, J.Q., Gonzalez-Garay, M.L., Jackson, A.R., Kalafus, K.J., McLeod, M.P., Milosavljevic, A., Virk, D., Volkov, A., Wheeler, D.A., Zhang, Z., Bailey, J.A., Eichler, E.E., Tuzun, E., Birney, E., Mongin, E., Ureta-Vidal, A., Woodwark, C., Zdobnov, E., Bork, P., Suyama, M., Torrents, D., Alexandersson, M., Trask, B.J., Young, J.M., Huang, H., Wang, H., Xing, H., Daniels, S., Gietzen, D., Schmidt, J., Stevens, K., Vitt, U., Wingrove, J., Camara, F., Mar Alba, M., Abril, J.F., Guigo, R., Smit, A., Dubchak, I., Rubin, E.M., Couronne, O., Poliakov, A., Hubner, N., Ganten, D., Goesele, C., Hummel, O., Kreitler, T., Lee, Y.A., Monti, J., Schulz, H., Zimdahl, H., Himmelbauer, H., Lehrach, H., Jacob, H.J., Bromberg, S., Gullings-Handley, J., Jensen-Seaman, M.I., Kwitek, A.E., Lazar, J., Pasko, D., Tonellato, P.J., Twigger, S., Ponting, C.P., Duarte, J.M., Rice, S., Goodstadt, L., Beatson, S.A., Emes, R.D., Winter, E.E., Webber, C., Brandt, P., Nyakatura, G., Adetobi, M., Chiaromonte, F., Elnitski, L., Eswara, P., Hardison, R.C., Hou, M., Kolbe, D., Makova, K., Miller, W., Nekrutenko, A., Riemer, C., Schwartz, S., Taylor, J., Yang, S., Zhang, Y., Lindpaintner, K., Andrews, T.D., Caccamo, M., Clamp, M., Clarke, L., Curwen, V., Durbin, R., Eyras, E., Searle, S.M., Cooper, G.M., Batzoglou, S., Brudno, M., Sidow, A., Stone, E.A., Payseur, B.A., Bourque, G., Lopez-Otin, C., Puente, X.S., Chakrabarti, K., Chatterji, S., Dewey, C., Pachter, L., Bray, N., Yap, V.B., Caspi, A., Tesler, G., Pevzner, P.A., Haussler, D., Roskin, K.M., Baertsch, R., Clawson, H., Furey, T.S., Hinrichs, A.S., Karolchik, D., Kent, W.J., Rosenbloom, K.R., Trumbower, H., Weirauch, M., Cooper, D.N., Stenson, P.D., Ma, B., Brent, M., Arumugam, M., Shteynberg, D., Copley, R.R., Taylor, M.S., Riethman, H., Mudunuri,

U., Peterson, J., Guyer, M., Felsenfeld, A., Old, S., Mockrin, S., and Collins, F. (2004). Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature *428*, 493-521.

Goodier, J.L., Ostertag, E.M., and Kazazian, H.H., Jr. (2000). Transduction of 3'-flanking sequences is common in L1 retrotransposition. Hum Mol Genet *9*, 653-657.

Gould, E., Cameron, H.A., Daniels, D.C., Woolley, C.S., and McEwen, B.S. (1992). Adrenal hormones suppress cell division in the adult rat dentate gyrus. J Neurosci *12*, 3642-3650.

Gould, E., Tanapat, P., McEwen, B.S., Flugge, G., and Fuchs, E. (1998). Proliferation of granule cell precursors in the dentate gyrus of adult monkeys is diminished by stress. Proc Natl Acad Sci U S A *95*, 3168-3171.

Grimaldi, G., Skowronski, J., and Singer, M.F. (1984). Defining the beginning and end of KpnI family segments. Embo J *3*, 1753-1759.

Gritti, A., Parati, E.A., Cova, L., Frolichsthal, P., Galli, R., Wanke, E., Faravelli, L., Morassutti, D.J., Roisen, F., Nickel, D.D., and Vescovi, A.L. (1996). Multipotential stem cells from the adult mouse brain proliferate and self-renew in response to basic fibroblast growth factor. J Neurosci *16*, 1091-1100.

Han, J.S., and Boeke, J.D. (2004). A highly active synthetic mammalian retrotransposon. Nature *429*, 314-318.

Han, J.S., Szak, S.T., and Boeke, J.D. (2004). Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. Nature *429*, 268-274.

Itsykson, P., Ilouz, N., Turetsky, T., Goldstein, R.S., Pera, M.F., Fishbein, I., Segal, M., and Reubinoff, B.E. (2005). Derivation of neural precursors from human embryonic stem cells in the presence of noggin. Mol Cell Neurosci *30*, 24-36.

Kashkush, K., Feldman, M., and Levy, A.A. (2003). Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nat Genet *33*, 102-106.

Kawasaki, H., Suemori, H., Mizuseki, K., Watanabe, K., Urano, F., Ichinose, H., Haruta, M., Takahashi, M., Yoshikawa, K., Nishikawa, S., Nakatsuji, N., and Sasai, Y. (2002).
Generation of dopaminergic neurons and pigmented epithelia from primate ES cells by stromal cell-derived inducing activity. Proc Natl Acad Sci U S A *99*, 1580-1585.

Kazazian, H.H., Jr. (1998). Mobile elements and disease. Curr Opin Genet Dev *8*, 343-350.

Kelly, S., Bliss, T.M., Shah, A.K., Sun, G.H., Ma, M., Foo, W.C., Masel, J., Yenari, M.A., Weissman, I.L., Uchida, N., Palmer, T., and Steinberg, G.K. (2004). Transplanted human fetal neural stem cells survive, migrate, and differentiate in ischemic rat cerebral cortex. Proc Natl Acad Sci U S A *101*, 11839-11844.

Kempermann, G., Kuhn, H.G., and Gage, F.H. (1997a). Genetic influence on neurogenesis in the dentate gyrus of adult mice. Proc Natl Acad Sci U S A *94*, 10409-10414.

Kempermann, G., Kuhn, H.G., and Gage, F.H. (1997b). More hippocampal neurons in adult mice living in an enriched environment. Nature *386*, 493-495.

Kubo, S., Seleme Mdel, C., Soifer, H.S., Perez, J.L., Moran, J.V., Kazazian, H.H., Jr., and Kasahara, N. (2006). L1 retrotransposition in nondividing and primary human somatic cells. Proc Natl Acad Sci U S A *103*, 8036-8041.

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

Kuhn, H.G., Winkler, J., Kempermann, G., Thal, L.J., and Gage, F.H. (1997). Epidermal growth factor and fibroblast growth factor-2 have different effects on neural progenitors in the adult rat brain. J Neurosci *17*, 5820-5829.

Kurose, K., Hata, K., Hattori, M., and Sakaki, Y. (1995). RNA polymerase III dependence of the human L1 promoter and possible participation of the RNA polymerase II factor YY1 in the RNA polymerase III transcription system. Nucleic Acids Res *23*, 3704-3709.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N.,

Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., and Chen, Y.J. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K.W., Vogelstein, B., and Nakamura, Y. (1992). Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. Cancer Res *52*, 643-645.

Moran, J., and Gilbert, N. (2002). Mammalian LINE-1 retrotransposons and related elements (Washington, DC, ASM Press).

Moran, J.V. (1999). Human L1 retrotransposition: insights and peculiarities learned from a cultured cell retrotransposition assay. Genetica *107*, 39-51.

Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D., and Kazazian, H.H., Jr. (1996). High frequency retrotransposition in cultured mammalian cells. Cell *87*, 917-927.

Morrish, T.A., Gilbert, N., Myers, J.S., Vincent, B.J., Stamato, T.D., Taccioli, G.E., Batzer, M.A., and Moran, J.V. (2002). DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat Genet *31*, 159-165.

Muotri, A.R., Chu, V.T., Marchetto, M.C., Deng, W., Moran, J.V., and Gage, F.H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature *435*, 903-910.

Muotri, A.R., and Gage, F.H. (2006). Generation of neuronal variability and complexity. Nature *441*, 1087-1093.

Oppenheim, R.W. (1991). Cell death during development of the nervous system. Annu Rev Neurosci *14*, 453-501.

Ostertag, E.M., Goodier, J.L., Zhang, Y., and Kazazian, H.H., Jr. (2003). SVA elements are nonautonomous retrotransposons that cause disease in humans. Am J Hum Genet *73*, 1444-1451.

Ostertag, E.M., and Kazazian, H.H., Jr. (2001a). Biology of mammalian L1 retrotransposons. Annu Rev Genet *35*, 501-538.

Ostertag, E.M., and Kazazian, H.H., Jr. (2001b). Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. Genome Res *11*, 2059-2065.

Palmer, T.D., Markakis, E.A., Willhoite, A.R., Safar, F., and Gage, F.H. (1999). Fibroblast growth factor-2 activates a latent neurogenic program in neural stem cells from diverse regions of the adult CNS. J Neurosci *19*, 8487-8497.

Palmer, T.D., Ray, J., and Gage, F.H. (1995). FGF-2-responsive neuronal progenitors reside in proliferative and quiescent regions of the adult rodent brain. Mol Cell Neurosci *6*, 474-486.

Palmer, T.D., Schwartz, P.H., Taupin, P., Kaspar, B., Stein, S.A., and Gage, F.H. (2001). Cell culture. Progenitor cells from human brain after death. Nature *411*, 42-43.

Penzkofer, T., Dandekar, T., and Zemojtel, T. (2005). L1Base: from functional annotation to prediction of active LINE-1 elements. Nucleic Acids Res *33*, D498-500.

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

Perrier, A.L., Tabar, V., Barberi, T., Rubio, M.E., Bruses, J., Topf, N., Harrison, N.L., and Studer, L. (2004). Derivation of midbrain dopamine neurons from human embryonic stem cells. Proc Natl Acad Sci U S A *101*, 12543-12548.

Prak, E.T., Dodson, A.W., Farkash, E.A., and Kazazian, H.H., Jr. (2003). Tracking an embryonic L1 retrotransposition event. Proc Natl Acad Sci U S A *100*, 1832-1837.

Reynolds, B.A., and Weiss, S. (1992). Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. Science *255*, 1707-1710.

Reynolds, B.A., and Weiss, S. (1996). Clonal and population analyses demonstrate that an EGF-responsive mammalian embryonic CNS precursor is a stem cell. Dev Biol *175*, 1-13.

Roy, N.S., Wang, S., Jiang, L., Kang, J., Benraiss, A., Harrison-Restelli, C., Fraser, R.A., Couldwell, W.T., Kawaguchi, A., Okano, H., Nedergaard, M., and Goldman, S.A. (2000). In vitro neurogenesis by progenitor cells isolated from the adult human hippocampus. Nat Med *6*, 271-277.

Sassaman, D.M., Dombroski, B.A., Moran, J.V., Kimberland, M.L., Naas, T.P., DeBerardinis, R.J., Gabriel, A., Swergold, G.D., and Kazazian, H.H., Jr. (1997). Many human L1 elements are capable of retrotransposition. Nat Genet *16*, 37-43.

Sayah, D.M., Sokolskaja, E., Berthoux, L., and Luban, J. (2004). Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. Nature *430*, 569-573.

Schulz, T.C., Palmarini, G.M., Noggle, S.A., Weiler, D.A., Mitalipova, M.M., and Condie, B.G. (2003). Directed neuronal differentiation of human embryonic stem cells. BMC neuroscience *4*, 27.

Shi, X., Seluanov, A., and Gorbunova, V. (2007). Cell divisions are required for L1 retrotransposition. Mol Cell Biol *27*, 1264-1270.

Shihabuddin, L.S., Ray, J., and Gage, F.H. (1997). FGF-2 is sufficient to isolate progenitors found in the adult mammalian spinal cord. Exp Neurol *148*, 577-586.

Smit, A.F., and Riggs, A.D. (1996). Tiggers and DNA transposon fossils in the human genome. Proc Natl Acad Sci U S A *93*, 1443-1448.

Sonntag, K.C., Pruszak, J., Yoshizaki, T., van Arensbergen, J., Sanchez-Pernaute, R., and Isacson, O. (2007). Enhanced yield of neuroepithelial precursors and midbrain-like dopaminergic neurons from human embryonic stem cells using the bone morphogenic protein antagonist noggin. Stem Cells *25*, 411-418.

Speek, M. (2001). Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Mol Cell Biol *21*, 1973-1985.

Suhonen, J.O., Peterson, D.A., Ray, J., and Gage, F.H. (1996). Differentiation of adult hippocampus-derived progenitors into olfactory neurons in vivo. Nature *383*, 624-627.

Svendsen, C.N., ter Borg, M.G., Armstrong, R.J., Rosser, A.E., Chandran, S., Ostenfeld,
T., and Caldwell, M.A. (1998). A new method for the rapid and long term growth of human neural precursor cells. J Neurosci Methods *85*, 141-152.

Tamaki, S., Eckert, K., He, D., Sutton, R., Doshe, M., Jain, G., Tushinski, R., Reitsma, M., Harris, B., Tsukamoto, A., Gage, F., Weissman, I., and Uchida, N. (2002). Engraftment of sorted/expanded human central nervous system stem cells from fetal brain. J Neurosci Res *69*, 976-986.

Tchenio, T., Casella, J.F., and Heidmann, T. (2000). Members of the SRY family regulate the human LINE retrotransposons. Nucleic Acids Res *28*, 411-415.

Uchida, N., Buck, D.W., He, D., Reitsma, M.J., Masek, M., Phan, T.V., Tsukamoto, A.S., Gage, F.H., and Weissman, I.L. (2000). Direct isolation of human central nervous system stem cells. Proc Natl Acad Sci U S A *97*, 14720-14725.

van Praag, H., Christie, B.R., Sejnowski, T.J., and Gage, F.H. (1999). Running enhances neurogenesis, learning, and long-term potentiation in mice. Proc Natl Acad Sci U S A *96*, 13427-13431.

Vescovi, A.L., Reynolds, B.A., Fraser, D.D., and Weiss, S. (1993). bFGF regulates the proliferative fate of unipotent (neuronal) and bipotent (neuronal/astroglial) EGF-generated CNS progenitor cells. Neuron *11*, 951-966.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S.E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., Bork, P., Botcherby, M., Bray, N., Brent, M.R., Brown, D.G., Brown, S.D., Bult, C., Burton, J., Butler, J., Campbell, R.D., Carninci, P., Cawley, S., Chiaromonte, F., Chinwalla, A.T., Church, D.M., Clamp, M., Clee, C., Collins, F.S., Cook, L.L., Copley, R.R., Coulson, A., Couronne, O., Cuff, J., Curwen, V., Cutts, T., Daly, M., David, R., Davies, J., Delehaunty, K.D., Deri, J., Dermitzakis, E.T., Dewey, C., Dickens, N.J., Diekhans, M., Dodge, S., Dubchak, I., Dunn, D.M., Eddy, S.R., Elnitski, L., Emes, R.D., Eswara, P., Eyras, E., Felsenfeld, A., Fewell, G.A., Flicek, P., Foley, K., Frankel, W.N., Fulton, L.A., Fulton, R.S., Furey, T.S., Gage, D., Gibbs, R.A., Glusman, G., Gnerre, S., Goldman, N., Goodstadt, L., Grafham, D., Graves, T.A., Green, E.D., Gregory, S., Guigo, R., Guyer, M., Hardison, R.C., Haussler, D., Hayashizaki, Y., Hillier, L.W., Hinrichs, A., Hlavina, W., Holzer, T., Hsu, F., Hua, A., Hubbard, T., Hunt, A., Jackson, I., Jaffe, D.B., Johnson, L.S., Jones, M., Jones, T.A., Joy, A., Kamal, M., Karlsson, E.K., Karolchik, D., Kasprzyk, A., Kawai, J., Keibler, E., Kells, C., Kent, W.J., Kirby, A., Kolbe, D.L., Korf, I., Kucherlapati, R.S.,

Kulbokas, E.J., Kulp, D., Landers, T., Leger, J.P., Leonard, S., Letunic, I., Levine, R., Li, J., Li, M., Lloyd, C., Lucas, S., Ma, B., Maglott, D.R., Mardis, E.R., Matthews, L., Mauceli, E., Mayer, J.H., McCarthy, M., McCombie, W.R., McLaren, S., McLay, K., McPherson, J.D., Meldrim, J., Meredith, B., Mesirov, J.P., Miller, W., Miner, T.L., Mongin, E., Montgomery, K.T., Morgan, M., Mott, R., Mullikin, J.C., Muzny, D.M., Nash, W.E., Nelson, J.O., Nhan, M.N., Nicol, R., Ning, Z., Nusbaum, C., O'Connor, M.J., Okazaki, Y., Oliver, K., Overton-Larty, E., Pachter, L., Parra, G., Pepin, K.H., Peterson, J., Pevzner, P., Plumb, R., Pohl, C.S., Poliakov, A., Ponce, T.C., Ponting, C.P., Potter, S., Quail, M., Reymond, A., Roe, B.A., Roskin, K.M., Rubin, E.M., Rust, A.G., Santos, R., Sapojnikov, V., Schultz, B., Schultz, J., Schwartz, M.S., Schwartz, S., Scott, C., Seaman, S., Searle, S., Sharpe, T., Sheridan, A., Shownkeen, R., Sims, S., Singer, J.B., Slater, G., Smit, A., Smith, D.R., Spencer, B., Stabenau, A., Stange-Thomann, N., Sugnet, C., Suyama, M., Tesler, G., Thompson, J., Torrents, D., Trevaskis, E., Tromp, J., Ucla, C., Ureta-Vidal, A., Vinson, J.P., Von Niederhausern, A.C., Wade, C.M., Wall, M., Weber, R.J., Weiss, R.B., Wendl, M.C., West, A.P., Wetterstrand, K., Wheeler, R., Whelan, S., Wierzbowski, J., Willey, D., Williams, S., Wilson, R.K., Winter, E., Worley, K.C., Wyman, D., Yang, S., Yang, S.P., Zdobnov, E.M., Zody, M.C., and Lander, E.S. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Wei, W., Gilbert, N., Ooi, S.L., Lawler, J.F., Ostertag, E.M., Kazazian, H.H., Boeke, J.D., and Moran, J.V. (2001). Human L1 retrotransposition: cis preference versus trans complementation. Mol Cell Biol *21*, 1429-1439.

Wei, W., Morrish, T.A., Alisch, R.S., and Moran, J.V. (2000). A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events. Anal Biochem *284*, 435-438.

Weiner, A.M. (2002). SINEs and LINEs: the art of biting the hand that feeds you. Curr Opin Cell Biol *14*, 343-350.

Wright, L.S., Prowse, K.R., Wallace, K., Linskens, M.H., and Svendsen, C.N. (2006). Human progenitor cells isolated from the developing cortex undergo decreased neurogenesis and eventual senescence following expansion in vitro. Exp Cell Res *312*, 2107-2120.

Yeo, G.W., Xu, X., Liang, T.Y., Muotri, A.R., Carson, C.T., Coufal, N.G., and Gage, F.H. (2007). Alternative splicing events identified in human embryonic stem cells and neural progenitors. PLoS computational biology *3*, 1951-1967.

Zhang, S.C., Wernig, M., Duncan, I.D., Brustle, O., and Thomson, J.A. (2001). In vitro differentiation of transplantable neural precursors from human embryonic stem cells. Nat Biotechnol *19*, 1129-1133.

CHAPTER 2

EVIDENCE FOR L1 RETROTRANSPOSITION IN DIFFERENT NEURAL STEM CELL SYSTEM

Abstract

Long interspersed nuclear (L1) elements are highly abundant in the human genome; however, their impact on the level of the individual is largely unknown. Here we show that human neural stem cells (NSCs) derived from both fetal brain and from human embryonic stem cells (hESCs) can support retrotransposition of an engineered L1 *in vitro*. These events occur in NSCs with the potential to differentiate to both neuronal or a glial lineages. Furthermore, we show that in hESC derived NSCs L1 insertional events can occur into genes. Our data suggest that *de novo* L1 retrotransposition events occur in human neural progenitors from various sources at an appreciable rate.

Introduction

Long-interspersed element-1 (L1) is a highly abundant retrotransposon that comprises approximately 17% of the human genome (Lander et al.). The majority of L1 elements are retrotransposition incompetent due to 5' truncation or mutations in their open reading frames (Moran and Gilbert). However, the average human genome is estimated to carry approximately 150 full-length, retrotransposition-competent L1s, of which only about 10% are highly active and responsible for the majority of mobility (Brouha et al.). Active L1 insertion has been shown to occur in germ cells and can result in a variety of genetic disorders dependent on the integration site, such as muscular dystrophy and hemophilia A (Kazazian). Mouse models indicate that retrotransposition is found primarily in the testes (Ostertag et al.).

It is estimated that up to 5% of people may carry a germline *de novo* L1-mediated event(Kazazian); however, relatively little is known about the developmental timing or cell types that accommodate L1 retrotransposition in humans. The only documented case of somatic retrotransposition involves an insertion into the adenomatous polyposis coli gene (APC), resulting in colon cancer(Miki et al.). *In vitro* models have shown that retrotransposition occurs readily in transformed cells lines (Han and Boeke, ; Moran et al., ; Wei et al.). However, retrotransposition has generally been thought to occur only at low levels outside the gonads, and primary tissue cultures such as fibroblasts have shown only extremely limited retrotransposition capability(Kubo et al., ; Shi et al.). Recent research, using a human L1 in both adult rat hippocampal progenitors *in vitro* and a L1 transgenic mouse model, indicates that L1 retrotransposition also occurs somatically in the rodent brain (Muotri et al.). However, the existence of *de novo* retrotransposition events in the human brain has not been reported.

Neural stem cells (NSCs) give rise to the three main lineages of the nervous system: neurons, astrocytes, and oligodendrocytes. The majority of cells are formed during embryonic development; however, numerous studies suggest that neurogenesis occurs in some brain regions throughout life(Eriksson et al.) (Gage, ; Taupin and Gage). The human nervous system is hugely complex, containing approximately 10^{15} synapses (Tang et al.). This vast diversity of neuronal cell types, subtypes, and connections is influenced by a complex and incompletely understood combination of environment and genetics. What effect L1 elements may have in this system is an open question.

Methods

Cell Culture

hESC lines HUES6 and H9 were cultured as previously described (http://www.mcb.harvard.edu/melton/HUES/) in the Gage lab (Thomson et al.). Briefly, cells were grown on mitomycin C-treated mouse embryonic fibroblast (MEF) feeder layers (Chemicon) in DMEM media (Invitrogen) supplemented with 10% KO serum replacement, 1 mM L-glutamine, 50 uM β-mercaptoethanol, 0.1 mM nonessential amino acids, and 10 ng/mL β-FGF2 (fibroblast growth factor 2), and passaged by manual dissection. For EB formation, cells were dissociated from the underlying MEF layer with Dispase (0.2 mg/mL; Stem Cell Technologies) and grown for 7 days in DMEM-F12 Glutamax media (Invitrogen) with N2 supplement (Gibco) and 500 ng/mL Noggin (Fitzgerald). Subsequently, EBs were plated onto laminin/poly ornithine (Sigma) -coated plates and grown for 7-10 days more. At this point rosettes were manually dissected and dissociated in 0.1% trypsin and plated in DMEM-F12 media supplemented with N2 and B-27, 1 ug/mL laminin, and 20 ng/mL FGF2. Resulting neural progenitors could be maintained for multiple passages before induction of differentiation. Differentiation conditions involved withdrawal of mitogens and treatment with 20 ng/ml BDNF, 20 ng/ml GDNF (Peprotech), 1 mm dibutyrl-cyclicAMP (Sigma), and 200 nm ascorbic acid (Sigma) for 4-12 weeks.

Registered hESC lines H7, H9, H13B, and BG01 were cultured as previously described in the Moran lab(Garcia-Perez et al.). Briefly, cells were grown on mitotically inactivated MEFs and passaged by manual dissection using the *StemPro EZPassage* passaging tool (Invitrogen). For NSC formation, a protocol based on

Zhang et al. was used (Zhang et al.). For EB formation, hESCs (see methods above) were seeded in a Suspension Culture Dish (Corning) in hESC media lacking FGF2. After 4-6 days, EBs were seeded in a Suspension Culture Dish coated with gelatin and cultured in NeuroSphere (NS) media for 14-16 days. NS culture medium consists of DMEM F12 (Invitrogen) supplemented with 20 ng/ml β-FGF, N2 supplement and 2 mg/ml Heparin (Sigma). After 14-16 days, rosettes were manually picked and trypsinized, then plated to form neurospheres. Neurospheres were passaged by single cell dissociation using a pulled Pasteur pipette once a week. For differentiation, a single cell suspension of NSCs was plated on polyornithine-coated plates in DMEM/F12 with N2 and 1% FBS and allowed to differentiate for 6 days.

HCNS-SCns FBR 1664, 1651, and 1673 were a kind gift from Stem Cells, Inc. (Palo Alto, CA) and were cultured as previously described (Uchida et al.). Briefly, cells were cultured in X-Vivo 15 media (Lonza Bioscience) supplemented with 20 ng/mL FGF-2, 20 ng/mL epidermal growth factor (EGF), 10 ng/mL leukemia inhibitor factor (LIF), N2 supplement, 0.2 mg/mL heparin, and 60 ug/mL N-acetylcysteine. For differentiation, cells were dissociated using Liberase Blendzymes (Roche) and plated on laminin/polyornithine-coated plates. Mitogens were withdrawn and cells were differentiated by retrovirus-mediated transduction with Neurogenin1 (NGN1), a pan-neuronal helix-loop-helix transcription factor. NGN1 was a kind gift from Dr. David Turner and was cloned into a murine Moloney leukemia retrovirus-based plasmid and expressed under the control of the ubiquitously expressed CAG promoter as previously described (Zhao et al.). Virus was made as previously described in human embryonic kidney 293T cells and collected by ultracentrifugation. HCNS-SCns were

infected 48 hrs before differentiation at an approximate efficiency of 70% and allowed to differentiate for 3-4 weeks.

Primary human fibroblasts and primary astrocytes were obtained and cultured as per manufacturer's instructions (Lonza Bioscience). Karyotyping and fluorescence *in situ* hybridization were performed by Cell Line Genetics, Madison, WI.

Constructs, Transfection, and Retrotransposition Assay

Cells were transfected with LINE-1 elements containing the EGFP retrotransposition cassette in the pCEP4 (Invitrogen) plasmid backbone, with the hygromycin selection gene replaced with a puromycin selection gene. Prior to transfection, DNAs were checked for superhelicity by electrophoresis on 0.7% agarose-ethidium bromide gels. Only highly supercoiled preparations of DNA (>90%) were used.

The $L1_{RP}$ element is an active full-length element under the control of the native 5'UTR (Kimberland et al.), and has been previously described (Ostertag et al.). LRE3 is an active full-length LRE3 element that has been previously described (Brouha et al.). JM111 is a derivative of $L1_{RP}$ containing the double missense mutation RR261-262AA in the ORF1 protein, rendering it retrotransposition incompetent (Garcia-Perez et al.). The UB-LRE3 and UB-JM111 constructs are modified such that the LRE3 and JM111 elements are strongly driven by the ubiquitin C promoter (a 1.2-kb fragment of the human UBC gene, nucleotides 123964272-123965484 from chromosome 12). All constructs contain the CMV-EGFP expression cassette (Ostertag et al.). The LRE3-neo and LRE3-blasticidin constructs contains the *mneoI* or

blasticidin retrotransposition cassettes rather than the EGFP cassette (Freeman et al., ; Moran et al.).

HUES6- and H9-derived NSCs one passage after neural rosette selection, as well as HCNS-SCns, were transfected by Nucleofection using the Amaxa rat NSC nucleofector solution and program A-31. Cells were cultured as progenitors in the presence of mitogens. For differentiation studies, cells were dissociated and plated for differentiation 18 days after initial transfection. H7-, H13B-, H9-, and BG01-derived NSCs were transfected using the Amaxa mouse NSC nucleofector solution and program A-33 and cultured as progenitors. When puromycin selection was used, puromycin 0.2 ug/mL was added two days post transfection for 5-7 days. Primary human fibroblasts and astrocytes were transfected using Fugene6 (Roche) as per manufacturer's instructions. Cells were monitored for GFP expression by fluorescence microscopy. For FACS analysis, cells were dissociated and analyzed on a Becton-Dickinson LSR I in the presence of 1 ug/mL propidium iodide for live/dead cell gating. All assays were performed in triplicate, and JM111 transfected cells were used as a negative control for gating purposes.

NSCs transfected with LINE-1 elements containing the *mneoI* or *blasticidin* retrotransposition indicator cassettes were subjected to either geneticin or blasticidin selection beginning 4-7 days after transfection. Cells were selected with 50 μg/ml of geneticin (G-418, Invitrogen) for 1 week and with 100 μg/ml of G418 the following week, or with Blasticidin 2 μg/mL (InvivoGen) for 2 weeks.

Immunohistochemistry and Imaging

Cells were fixed in 4% paraformaldehyde, and immunocytochemistry was performed as previously described (Gage et al., ; Garcia-Perez et al.). Antibodies and dilutions were as follows: β-III tubulin mouse monoclonal 1:400 or rabbit polyclonal 1:500 (both Babco/Covance), Map (2a+2b) mouse monoclonal 1:500 (Sigma), GFAP rabbit polycolonal 1:300 (DAKO), GFAP guinea pig polyclonal 1:1000 (Advanced Immunochemical), Nestin mouse monoclonal 1:800 (Chemicon), Musashi rabbit polyclonal 1:200 (Chemicon), Sox1 1:200 rabbit polyclonal (Chemicon), Sox1 goat polyclonal 1:200 (R&D), TH rabbit polyclonal 1:500 (Pel-Freez), Ki-67 rabbit monoclonal 1:500 (VectorLabs), Sox2 rabbit polyclonal 1:500 (Sigma), Sox3 rabbit polyclonal 1:500 (a generous gift from Dr. M.W. Klymkowsky, Denver, CO). Secondary antibodies were purchased from Jackson ImmunoResearch or Invitrogen and were all used at 1:250. Cells were imaged using a CARVII spinning disk confocal imaging system (BD).

Luciferase Assay

Luciferase activity was measured with the Dual-Luciferase reporter assay system (Promega) in accordance with manufacturer's instructions. A plasmid containing the Renilla *luciferase* gene was used as an internal control in all assays. All assays were replicated at least three times independently. The L1 5'UTR luciferase construct has been previously described (Muotri et al.). The Synapsin-1 promoter region was a kind gift from G. Thiel. All promoters were in the pGL3-basic vector (Promega).

Southern blot

Southern blotting was done as previously described (Sambrook et al.). Briefly, 10 ug of gDNA was digested with *ClaI*, a restriction enzyme that digests the L1 both at 6006 bp (200 bp from the 5' end of the retrotransposition cassette) and at 8553 bp (in the 3'UTR). Therefore the L1.3 plasmid with indicator cassette would yield a 2547 bp band, whereas a retrotransposed L1 integrated into a genomic sequence lacking the intron would yield a 1645 bp band. The probe was a full-length GFP DNA fragment that was radioactively labeled with dCTP using the Random Prime Labeling Kit (Roche) as per manufacturer's instructions.

Cell lysates and Western blot analysis

hESC or NSCs were harvested and lysed with 1 ml of 1.5 mM KCl, 2.5 mM MgCl$_2$, 5 mM Tris-Hcl pH 7.4, 1% deoxycolic acid, 1% Triton X-100, and 1X Complete Mini EDTA-free Protease Inhibitor cocktail (Complete) (Roche) as previously described (Kulpa and Moran). Cell debris was removed by centrifugation at 3,000 x g at 4°C for 5 minutes, and a 10% fraction of the supernatant was saved (i.e., Whole Cell Lysate or WCL fraction). A sucrose cushion was then prepared with 8.5% and 17% w/v sucrose in 80 mM NaCl, 5 mM MgCl$_2$, 20 mM Tris-Hcl pH 7.5, 1 mM DTT, with Complete. WCLs were spun at 39,000 rpm at 4°C for 2 hours using a Sorvall SW-41 rotor. After centrifugation, the pellet material (i.e., Ribonucleoprotein Particle or RNP) was resuspended in 50 μl of purified water with Complete. Total protein concentration was determined by Bradford assay (BioRad) according to the manufacturer's instructions, and 8 μg of each sample (either WCL or RNPs) were load

on 10% SDS-PAGE gels (BioRad). Rabbit anti-ORF1 antibody (a generous gift from Dr. Thomas Fanning) was used at a 1:10,000 dilution; rabbit anti-S6-Ribosomal protein antibody (Cell Signaling) was used at a 1:1000 dilution; rabbit anti-Sox3 antibody (a generous gift from Dr. M. Klymkowsky) was used at a 1:1000 dilution; goat anti-Sox1 antibody (R&D) was used at a 1:500 dilution. All HRP conjugated secondary antibodies were used at a 1:20,000 dilution (abcam). For Western blotting, WCL was collected in RIPA buffer (Upstate) with Complete. Antibodies used were rabbit anti-ORF2 1:1000 (a generous gift of Dr. J. Boeke) and mouse monoclonal β-actin 1:1000 (Ambion). HRP conjugated secondary was used at 1:5000 (GE Healthcare).

PCR

Genomic DNA from transfected NSCs and HCNS-SCns was isolated using the DNeasy Blood & Tissue kit (Qiagen) following manufacture's instruction. To analyze removal of the intron upon retrotransposition, 200 ng of gDNA was used in a 25 ul reaction with primers GFP968s and GFP1013as (in experiments conducted with EGFP-tagged L1s), NEO437s and NEO1808as (in experiments conducted with *mneoI*-tagged L1s), or Blast Fw and Rv primers (with Blasticidin-tagged L1s), using the conditions described previously (Moran et al., ; Ostertag et al.).

Insertional characterization by inverse PCR

Cells were transfected with LRE3 pCEP4 plasmid and allowed to proliferate under NSC conditions for 18-20 days. Subsequently cells were dissociated with trypsin and sorted on a Becton-Dickenson FACscan. A total of 40,000 GFP-positive

cells were sorted and expanded in culture for three passages, with a corresponding GFP-negative control of 50,000 cells. This experiment was replicated with a second set of independently derived NSCs. Cells were harvested and gDNA was isolated using standard phenol-chloroform techniques. Alternatively, single GFP-positive cells were sorted into 96-well plates and allowed to proliferate for 6-8 weeks. Cells were then trypsinized in 10 uL Tryple. Since DNA yield for single colonies was very low, whole genome amplification was performed using the Genomiphi kit (GE Life Sciences). Analysis was performed as previously described (Garcia-Perez et al.). Briefly, 5-10 ug of DNA was digested overnight with either *SspI* or *XbaI* and self-ligated in a final volume of 600 μl with 3,200 U of T4 DNA ligase (NEB) overnight at 4°C. Ligated DNA was concentrated to 50 uL using a Microcon 100 column (Millipore). Inverse PCR was performed as previously described (Muotri et al.) (Garcia-Perez et al.) and PCR products were gel-isolated and cloned into TOPO-TA for sequencing. DNA sequence analysis was performed using the UCSC genome browser (http://genome.ucsc.edu, March 2006 assembly).

Results

Human fetal-derived NSCs support LINE retrotransposition

To determine if human NSCs are capable of supporting L1 retrotransposition, we transfected human fetal NSCs (HCNS-SCns) with an active engineered human L1(Brouha et al.). The HCNS-SCns (a gift from Stem Cells, Inc., Palo Alto, CA) were derived from fetal brain using fluorescence-activated cell sorting (FACS) for cell surface markers, specifically (CD133)$^+$, 5E12$^+$, CD34$^-$, CD45$^-$, and CD24$^{-/lo}$. This

combination has been shown to enrich for progenitor neurosphere-initiating cells and to be capable of differentiating into both neuronal and glial lineages (Uchida et al.). The L1$_{RP}$, LRE3, and L1.3 elements are highly active L1s whose expression is driven from the native 5'UTR and contains a retrotransposition indicator cassette in the 3'UTR. The indicator cassette is in the reverse orientation with regards to the L1 and consists of the EGFP gene, interrupted by an intron (IVS2 of the γ–globin gene), complete with its own poly adenylation signal and driven by the ubiquitous CMV promoter (Moran et al., ; Ostertag et al.). This configuration ensures GFP expression only after mRNA processing of the intron, followed by successful retrotransposition of the LINE element (Figure 2.1A).

Active L1 elements L1$_{RP}$ or L1.3 were transfected into three different HCNS-SCns lines (FBR 1664, FBR 1673 and FBR 1651; Fig. 2.1A) which were allowed to proliferate as NSCs. A low level of retrotransposition was reproducibly observed in HCNS-SCns transfected with the active L1s. We never detected retrotransposition events from an L1 that contained two missense mutations in the ORF1 RNA-binding domain, rendering it immobile (JM111/ L1$_{RP}$ /RR261-262AA) (Fig. 1B, Fig. S1E).

In order to confirm intron removal and genomic insertion events, PCR experiments utilizing genomic DNA derived from individual neurospheres indicated that a retrotransposition event had occurred with precise splicing of the intron from the indicator cassette (Fig. 2.1B). Successive subculture of transfected neurospheres indicates that the insertions were stable and could be detected by PCR after more than 3 months of subculture.

Southern blotting performed with a probe toward full-length GFP indicated integration of $L1_{RP}$ into the genome even after 3 months of neurosphere subculture (Fig. 2.1C). Since we expect numerous random L1 genomic insertions, a standard southern protocol produced a smear with a full length GFP probe (data not shown). In order to specifically assay for insertion of the GFP, we digested genomic DNA with restriction enzymes *ClaI* which specifically digests the L1-GFP construct on either side of the γ-globin intron, yielding two different sized bands dependent on insertional status: a 2547bp band with intron, and a 1645bp band when an L1 insertional event occurred. This method detects insertions of at least 2500 bp as it requires the L1 to insert through the *ClaI* restriction site in ORF2.

Detection of L1 retrotransposition by FACs reproducibly required 18-21 days for maximal detection in all three cell lines (Fig. 2.1D), in contrast to transformed cell lines such as 293T or 3T3, where GFP is evident 4-6 days post-transfection (Moran et al.). It is unknown whether this difference in the time required for GFP expression is due to differences in cell division time (doubling time for HCNS-SCns ~4 days as opposed to 293T, which is 33 hrs) or due to GFP expression and silencing effects. The rapid doubling time of transformed cell lines may allow for more rapid retrotransposition or perhaps another mechanism is at work. In addition, the HCNS-SCns expressed the endogenous L1 ORF2 protein, with expression decreasing sharply with the induction of differentiation (Fig. 2.1E).

Furthermore, undifferentiated, GFP-positive HCNS-SCns cells expressed canonical NSC markers, including Sox2, Nestin (Fig. 2.1F), Musashi and Sox1 (Fig. S1A-B). They also continued to proliferate in culture and co-labeled with the mitotic

marker, Ki-67 (Fig. S2.1C). These cells could be differentiated and resulted in cells of both neuronal and glial lineages, as shown with immunocytochemistry towards the neuronal markers β-III tubulin and Map2(a+b) (Fig. 2.1G) and the glial marker GFAP (Fig. 2.1H). Results were independent of differentiation method, such that differentiation using either a retrovirus expressing the pan-neuronal beta-helix-loop-helix Neurogenin 1 or by means of a mix of brain-derived and glial-derived neurotrophic factors (BDNF & GDNF) resulted in similar outcomes (data not shown). As controls, we tested L1$_{RP}$ in primary human astrocytes and primary human fibroblasts (Fig. S2.1D). We never detected retrotransposition events in either line (Fig. 2.1 D) at any time point (FACs, Figure S2.1E). Additionally, frequent karyotyping indicated that all three HCNS-SCns derivations were karyotypically stable (Fig. S2.4 A-C).

Together these data demonstrate that an native human L1, engineered with an indicator cassette, whose expression is driven by its own internal promoter, can retrotranspose in primary HCNS-SCns *in vitro*. Furthermore, HCNS-SCns also express the endogenous L1 ORF2 protein, indicating the possibility that other L1s may also be active.

NSCs derived from human embryonic stem cells support LINE retrotransposition

To further investigate L1 retrotransposition both in a second model and in a model that is more renewable than one derived from primary fetal brain tissue, we utilized a system based on human embryonic stem cells (hESC) -derived NSCs and neurons. These data were obtained by two different laboratories independently and

utilized different hESC lines, different transfection methods, and different differentiation protocols.

hESC-derived NSCs were transfected with engineered $L1_{RP}$ or LRE3 elements and exhibited a rate of retrotransposition significantly higher than that of HCNS-SCns (Fig. 2.2B). HUES6 and H9 ES lines were differentiated using a modified protocol as previously described (Yeo et al.) (Fig. 2.2A). Briefly, hESC were differentiated into embryoid bodies (EBs) in the presence of Noggin and replated to form neural rosettes. Rosettes were manually dissected and dissociated with trypsin, and resultant cultures were NSCs and grown as adherent cultures in the presence of FGF2. hESC cultures were frequently karyotyped, and fluorescence in situ hybridization was performed to rule out common hESC translocations (Fig. S2.3D-F). Cells were transfected and analyzed for GFP expression secondary to L1 insertion at 7-8 days post transfection (see Table 1, lab G). Independent experiments were performed using NIH-approved hESC lines H7, H9, H13B and BG01, using an adapted NSC differentiation protocol (Zhang et al.) (Fig. S2.2A). hESCs were differentiated to EBs and replated to form neural rosettes, which were subsequently manually dissected, physically dissociated, and cultured as neurospheres. Cells were transfected, selected for plasmid incorporation for days 2-7 post transfection using 0.2 ug/mL puromycin, and analyzed for GFP expression at day 7-8 (see Table 1, lab M). As with HCNS-SCns, intron splicing could be detected by FACS (Fig. S2.3A),

Retrotransposition in hESC-derived NSCs was variable, depending on the differentiation method, on whether puromycin selection was used, and on the individual NSC preparation. However, multiple experiments using multiple hESC

lines indicated that the effect was robust, reproducible, and frequently higher than in any other cell line assayed. Whereas HCNS-SCns exhibited a retrotransposition rate of approximately 0.0001%, HUES6-derived NSCs reproducibly exhibited a retrotransposition rate with at least a 10-fold increase, with some increases being greater than 1,000-fold. Using a ubiquitously expressed L1 driven by the ubiquitin promoter, retrotransposition rates as high as 16.25% were seen (Table 2.1). This rate was higher than what has previously been documented in other cell lines, including transformed lines such as HEK293 and 3T3 cells.

To confirm that L1 retrotransposition events occurred in legitimate NSCs, immunocytochemistry established that L1-GFP-positive cells co-labeled for NSCs markers such as SOX1 and SOX3 (Fig. 2.2B, S2.2D) in H13B-derived NSCs and for SOX1, SOX2, and Nestin in HUES6-derived NSCs (Fig. 2.2E, S2.2C). Additionally, L1-GFP-positive, HUES6-derived NSCs differentiated to a neuronal fate, as indicated by the neuronal markers βIII tubulin and Map2a+2b (Fig. 2.2F, S2.2E). L1-GFP-positive cells also differentiated to a glial fate and co-labeled with the marker GFAP (S2.2F). Moreover, some L1-GFP-positive neurons exhibited subtype-specific markers, such as tyrosine hydroxylase (TH, Fig. 2.2G). To confirm active retrotransposition, genomic DNA was used for PCR to confirm the loss of the γ-globin intron (Fig. S2.3F-G).

In addition to the L1-GFP construct, retrotransposition was observed using L1-neomycin and L1-blasticidin constructs where the GFP γ-globin was replaced with neomycin and blasticidin γ-globin constructs (Moran et al.) (Fig. S2.3C-E). Neomycin-resistant colonies were positive for the NSC marker, SOX3 with some cells

differentiating to a neuronal lineage (βIII tubulin) but with minimal marker overlap (Fig. 2.2B).

We also sought to investigate the activity of the endogenous L1 promoter and L1 expression. First we examined whether the ORF1 protein was expressed in hESC-derived NSCs. To this end, we isolated ribonucleoprotein particles (RNPs) from hESC and NSCs and compared them by Western analysis (Fig. 2.2D). RNPs have previously been shown to contain the L1 RNA in addition to ORF1 and ORF2 proteins (Kulpa and Moran). Data indicated that L1 ORF1 protein is expressed in NSCs, albeit at a lower level than in hESCs when compared to the ribosomal S6 protein-loading control. NSCs also expressed SOX1 and SOX3. Secondly, we studied the activity of the intact human L1 promoter (the 5'UTR) in hESC-derived NSCs. The human L1 5' UTR contains two SOX-binding sites (Tchenio et al.), a YY1-binding site needed for transcriptional initiation (Athanikar et al., ; Swergold), and a runt-domain transcription factor 3 (RUNX3) binding site (Yang et al.). Previous luciferase data in rat NSCs showed that the activity of the L1 5'UTR increased sharply with the start of differentiation and fell off thereafter (Muotri et al.). We found a similar pattern of human L1 5'UTR activation in hESC-derived NSCs, with initiation of differentiation resulting in a 25-fold increase in L1 5'UTR activity over 2 days, which fell off thereafter (Fig. 2.2C). As a control, we found that the luciferase activity of the synapsin promoter increased ~250 fold during the same period (Fig. S2.2E), indicating robust neuronal differentiation.

These data are consistent with the experiments with HCNS-SCns and with previous findings in rodent NSCs indicating increased L1 retrotransposon activity in

early neural progenitors from multiple sources. Not only is L1 retrotransposition increased, but the rate of retrotransposition is also greater than in any other cell line assayed.

GFP expression and L1 insertions

To further characterize cells harboring L1 retrotransposition events, we sorted GFP-positive and-negative, HUES6-derived NSCs by FACs 18 days after electroporation. Although single cell clones were attempted, <2% of single cells survived and continued to proliferate (both GFP positive and negative), and none grew to confluency in a 96 well plate. Seven GFP positive and four GFP negative clones were collected for genomic DNA analysis and a single insertion was obtained (see Table 2.2). Due to this inability to expand individual cell clones in culture, we sorted 40,000 GFP-positive into a single culture (with a matched GFP-negative control). Genomic DNA analysis yielded a PCR product corresponding to the retrotransposed GFP gene (Fig. 2.3A). These cells proliferated in culture and were positive for NSC markers similar to GFP-negative control cultures (Fig. 2.3C). Upon differentiation to a mixed neuronal and glial lineage, both GFP-positive and -negative cultures differentiated to both lineages at identical rates (Fig. 2.3D).

To characterize post-integration sites from both hESC-derived NSC clones and sorted cultures, we performed inverse PCR (IPCR) as previously described (Myers et al.). From the sorted GFP-positive cultures, we derived 18 insertions either fully or partially (Table 2.2). In some cases, full insertions were not recoverable due to long 5' insertions (Table 2.2, column 2). A second replication of this experiment yielded a similar culture, but different insertions (Table 2.2, column 1). Interesting, multiples of

these insertions are near neuronally expressed genes. Furthermore, several of the insertions were recovered with multiple independent IPCR reactions from the same culture, suggesting the insertion may contribute a growth advantage, allowing the NSC to proliferate preferentially (Table 2.1, column 3). These insertions were near membrane-expressed proteins, G-protein-coupled receptors or tyrosine kinase receptors. The majority of insertions were <100 kB from the nearest gene (16 out of 18) (Muotri et al., 2005). All insertional events occurred into the canonical endonuclease sites AA/TTTT or a derivative thereof, and a number of these exhibited target site duplications, as expected (Morrish et al.) (Fig. 2.3B). Due to the pooling of cells in this study, it is impossible to tell whether cells incurred a single insertional event or multiple events during the course of isolation and expansion. These data are consistent with previous studies in rat NSCs (Muotri et al.) and transformed cell lines (Gilbert et al., ; Symer et al.), demonstrating that L1 can retrotranspose into genes and may insert into neuronal genes.

Discussion

Previous studies have focused primarily on rodent models and have shown that retrotransposition of a human L1 can occur in rodent systems both in the brain and in germ cells or early embryogenesis (Muotri et al., ; Ostertag et al., ; Prak et al.). Recent evidence indicates that hESCs can also support a very low level of retrotransposition events (Garcia-Perez et al.), as compared to transformed or immortalized cell lines (Moran et al., ; Morrish et al.). Our data indicate that human NSCs derived from multiple different embryonic and fetal sources can support retrotransposition of a native human L1 at varying levels. We detected active retrotransposition using a

variety of different methods: FACs, immunocytochemistry, PCR from genomic DNA, and southern blotting. Conversely, we did not detect active retrotransposition events in other human primary cultures. Additionally, our data suggest that these insertional events need not be deleterious, in that NSCs with insertional events can continue to propagate in successive subculture successfully for at least 3 months. In fact, the burden of L1-GFP insertions appears to increase such that the intron-removed band is prominent at 3 months post-transfection. This may be due to successive insertions occurring as the plasmid vector continues to be propagated and the L1 continues to retrotranspose.

There is also some evidence that endogenous L1s may be active in these NSCs. The large number of degenerate L1s present in the human genome makes it extremely difficult to detect *de novo* endogenous insertions. However, the expression of ORF1 and ORF2 in NSCs in addition L1 promoter activation with the onset of neuronal differentiation strongly suggests that the possibility exists that endogenous L1s could be active.

Our results are consistent with previous analyses showing that retrotransposition events secondary to an engineered human L1 frequently insert into genes or their promoters (Garcia-Perez et al., ; Moran et al., ; Symer et al.). In fact, we found that some L1 retrotransposition events in NSCs integrated into or near neuronally expressed genes. The majority of insertions isolated here were proximal to known genes, and several others were into L1 or other repeat regions near genes, another common feature of L1 insertions. Given that some insertions were isolated multiple times despite the large number of initially sorted cells, it is likely that this

selection protocol, which yields true insertional events, also biases as to which ones are identified. In addition, these data indicate that L1 retrotransposition is occurring through the canonical endonuclease dependent mechanism, as all of the insertions were into a derivative of the degenerate AA/TTTT endonuclease recognition site. Similarly, a number of insertions contained target site duplications.

Overall, our findings indicate L1s are active in NSCs derived from multiple different sources, and also provide some evidence that endogenous L1s may also be active   in   these   NSCs.

Figure 2.1

L1 retrotransposition occurs in HCNS-SCns. A. Fetal brain tissue was collected and dissociated, and the resulting cells were sorted by FACs for stem cell markers CD133, 5E12, CD34, and CD45 to enrich for neural progenitor cells. Resulting cells were propagated as neurospheres in culture and were transfected with the human L1 plasmid. Cells were either propagated after transfection as neural progenitors and analyzed at varying time points or differentiated to a primarily neuronal fate. B. PCR analysis of gDNA isolated from different cell populations. The 1,243bp product corresponds to the original L1 vector harboring the intron-containing GFP cassette, and the 343bp product is diagnostic for loss of the intron, indicating a retrotransposition event. Lane 1, molecular mass standards; Lane 2, HCNS-SCns transfected with JM111 negative control; Lanes 3, 4, 5 correspond to three different HCNS-SCns lines 3 months post transfection with L1RP; Lane 6, primary astrocytes transfected with L1RP; Lane 7, primary fibroblasts transfected with L1RP; Lane 8, positive control; Lane 9, water control. C. Southern blot analysis of HCNS-SCns line FBR 1664, 3 months post transfection with the L1RP plasmid showing insertional events. D. Time course analysis of three HCNS-SCns lines, primary fibroblasts, and astrocytes for 25 days post transfection. E. Western blot towards ORF2 indicating expression of ORF2 in NSCs but not in differentiating neurons, with actin loading control. F. L1-GFP-positive cells were also positive for the NSC markers Nestin and Sox2. G-H. Immunocytochemistry of differentiated HCNS-SCns indicates L1-GFP-positive cells differentiate to both neuronal (β-III tubulin and Map2A+B) and glial (GFAP) lineages. Scale bar = 25μm.

Figure 2.1

A

Dissociate brain

Fetus

FACS sort

CD133
5E12 CD34 CD45

+FGF2
+EGF
+LIF

Neurosphere

Transfect with
L1 pCEP plasmid

pCEP

Analysis, day 2-25

green cell

+NGN
retrovirus
–Growth factors

Neurons

B

1 2 3 4 5 6 7 8 9

GFP+ intron

GFP

C

50ng   L1RP   C

GFP+ intron

GFP

D

Number of GFP events p 100,000

12
10
8
6
4
2
0

1673 L1RP
1664 L1RP
1651 L1RP
JM111
Fibro L1RP
Astro L1RP

12 15 18 21 25

Time (Days)

E

Differentiation

NSCs   3hrs   2D   4D

ORF2   150kD

Actin   42kD

F

GFP
Nestin
Sox2

G

GFP
Map2a+b
BIII Tubulin

H

GFP
GFAP
BIII Tubulin

46

Figure 2.2

L1 retrotransposition occurs in hES-derived NSCs. A. hESC colonies were plated to form embryoid bodies in the presence of Noggin and then were allowed to form neural rosettes on a coated substrate. Neural rosettes were manually dissected and dissociated and the resulting NSCs were propagated in the presence of FGF2. NSCs were transfected with the L1 plasmid and then either propagated in an undifferentiated state for analysis at varying times or differentiated to a primarily neuronal fate. B. Top: L1 retrotransposition occurs in NSCs derived from HB13 ES cells using the L1-GFP construct. Bottom: Using the L1-Neo construct, neo resistant colonies can exhibit both NSCs markers such as SOX3 and markers of neuronal differentiation such as βIII tubulin C. Luciferase assay conducted with the L15'UTR driving the luciferase gene exhibits a 25-fold increase in expression with the induction of differentiation. D. To analyze endogenous L1 expression, we isolated ribonucleoprotien particles (RNP) from hESC and NSCs and probed for endogenous L1 expression using an antibody toward ORF1, as compared to the ribosomal S6 loading control. NSCs also expressed SOX1 and SOX3 proteins. WCL = whole cell lysate. E. L1-GFP-positive, HUES6-derived NSCs were also positive for the NSC markers SOX2 and Nestin. F-G. Immunocytochemistry of differentiated hES-derived NSCs indicates L1-GFP positive cells differentiate to a neuronal fate (β-III tubulin) and can be positive for more subtype-specific neuronal markers such as tyrosine hydroxylase (TH). Scale bar = 25 μm.

Figure 2.2

A

ESC → EB → Neural rosette → +FGF2 → Neural progenitor cell → Analysis, day 2-25 → FACS analysis

Transfect with L1.3 pCEP plasmid

+ Differentiation factors, BDNF, GDNF → Neurons

B

BF | L1-GFP

SOX1

SOX3
BIII Tubulin
Hoecht

C

L1 5'UTR-Luciferase

Differentiation (Days)

D

H13B | H7 | NSCs H13B

αORF1 — RNPs

αS6

αSox3 — WCLs

αSox1

E

GFP | Nestin | Sox2

F

GFP
BIII Tubulin
DAPI

G

GFP
TH
B-III Tubulin
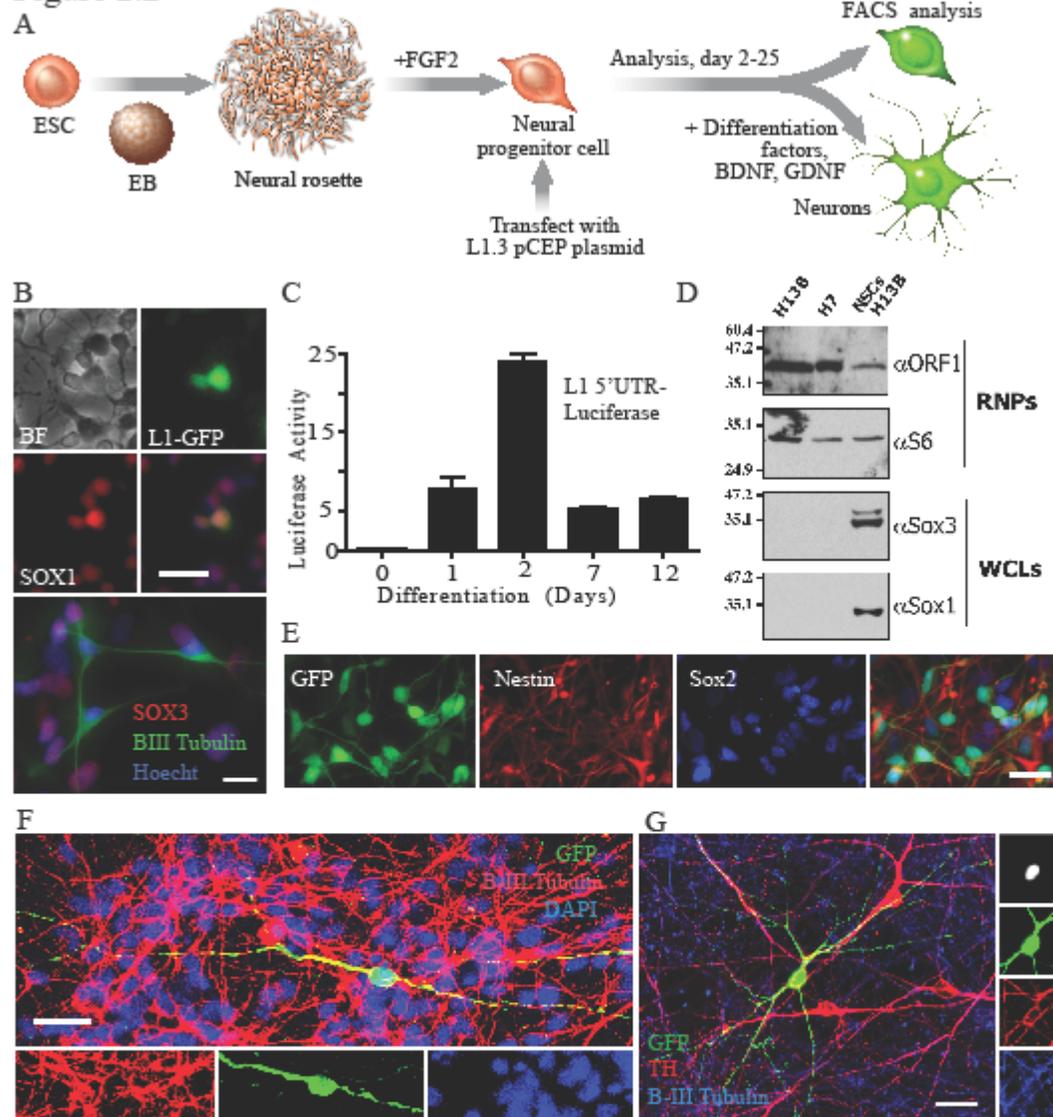
Figure 2.3

Characterization of L1 retrotransposition in hESC-derived NSCs A. FACs analysis of transfected cells indicates GFP-positive cells with the L1RP plasmid, and no positive events with the JM111 negative control. B. Luciferase assay using the synapsin promoter driving luciferase shows a >250-fold induction of synapsin during neuronal differentiation of hESC-derived NSCs. C-D. Transfection of both HUES6- (C) and H7- (D) derived NSCs with the L1-Neo construct indicates that neo-resistant colonies have experienced a retrotransposition event resulting in removal of the γ-globin intron. E. The L1-blasticidin construct also retrotransposes in HUES6-derived NSCs. F-G. PCR from gDNA of L1-GFP transfected cells indicates that removal of the intron and insertion of the L1 occur in both H13B- (F) and HUES6- (G) derived NSCs.

Figure 2.3

A

B

Chromosome 7, insertion in an intron of zinc finger protein 804B (ZNF804B)

Post-integration: 6266|

5'AATAACATAGAATGAAG L1 GFP A~137 AAGATTTAGTGTTACCG3'

Pre-integration site:

TTGAATAACATAGAATG AAGATTTAGTGTTACCGTTAACA
AACTTATTGTAT CTT AC/TTCT AAATCACAATGGCAATTGT
              endonuclease recognition site

C

D
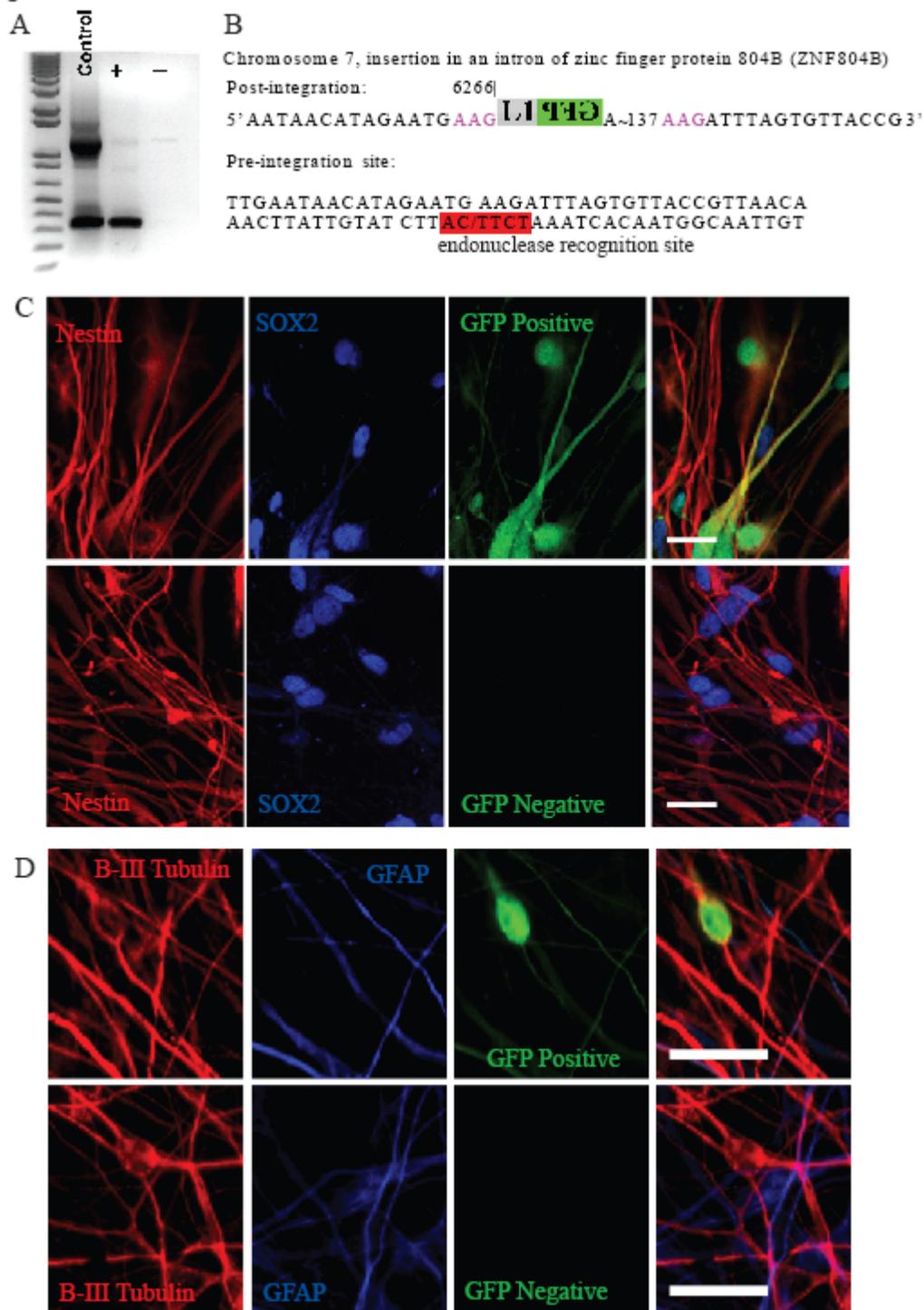
Supplement 2.1
Characteristics of retrotransposition in HCNS-SCns. A-B. L1-GFP-positive cells are also positive for the NSC markers musashi-1 and SOX1. C. L1-GFP-positive HCNS-SCns are still capable of cell division and can be positive for the cell division marker Ki-67. D. Brightfield images of primary astrocytes and fibroblasts. E. FACs analysis of transfected cells indicates a low but reproducible rate of GFP-positive cells in HCNS-SCns with the L1RP plasmid, and no positive events with other controls: JM111 in HCNS-SCns, and L1RP in primary human astrocytes and fibroblasts.

Supplement 2.1
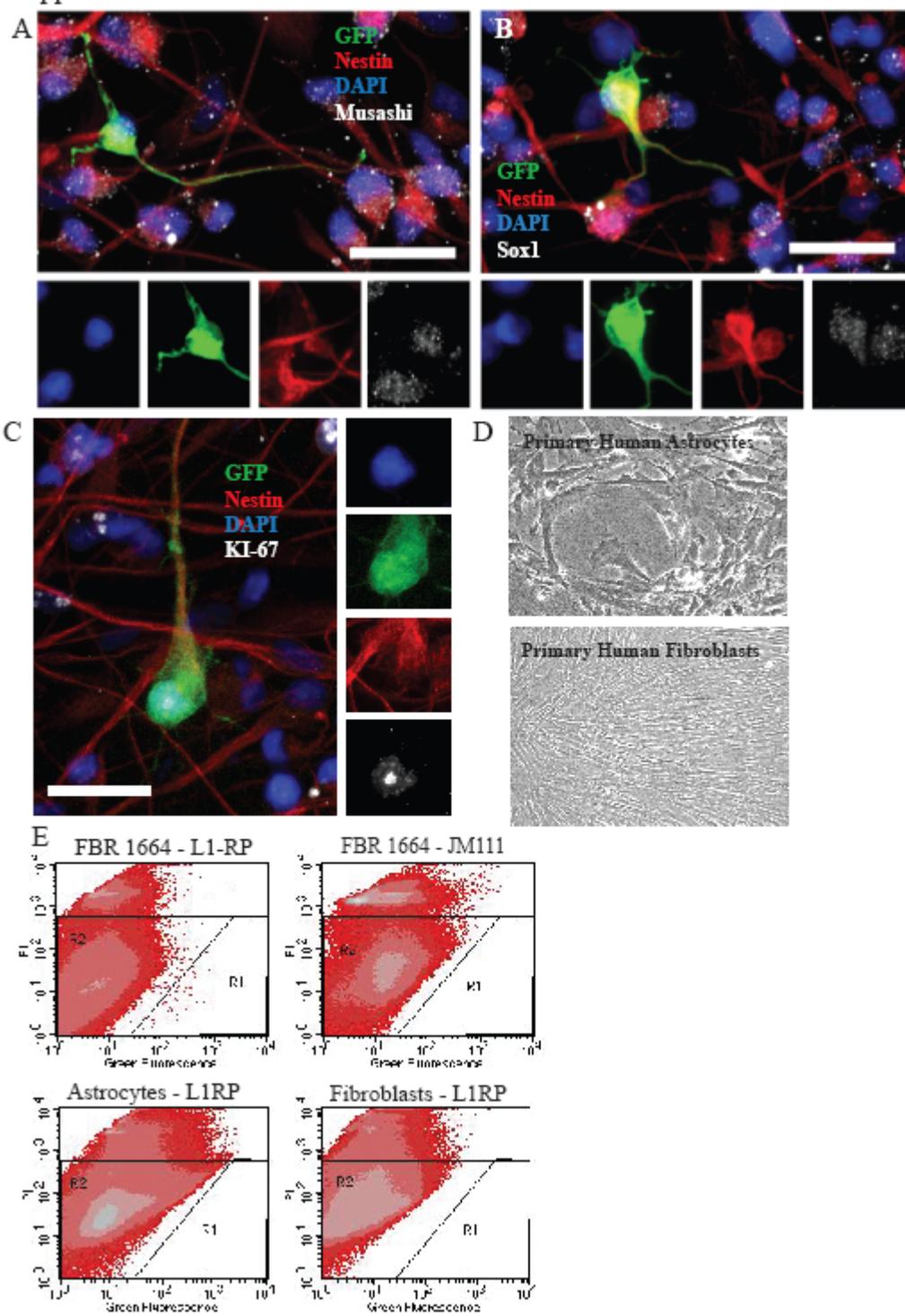
Supplement 2.2
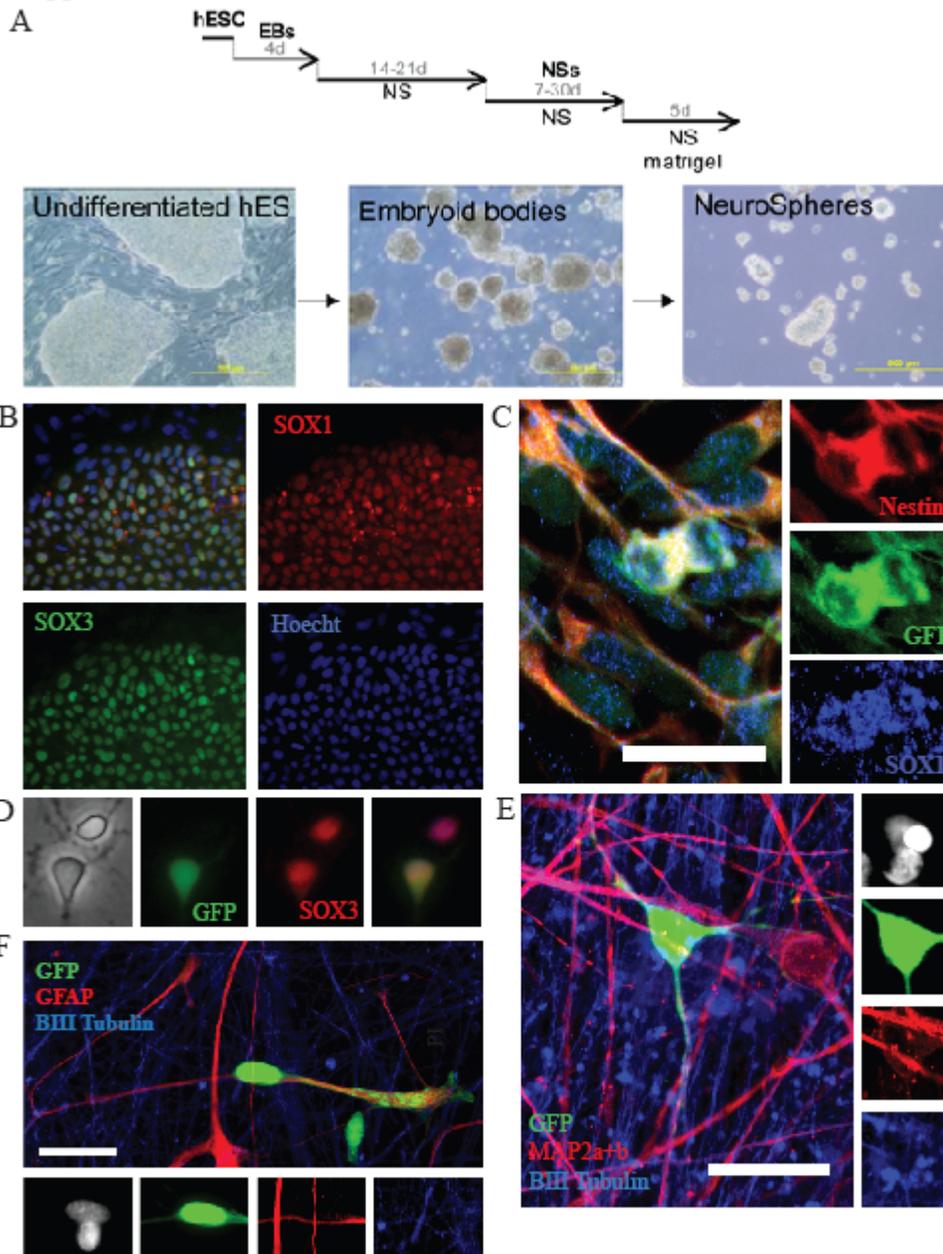
Supplement 2.2

Retrotransposition in hES-derived NSCs. A. Schematic overview of neural differentiation of hESC lines H7, H9, H13B and BG01. B. When dissociated, the resulting neurospheres are positive for the NSC markers SOX1 and SOX3. C. L1-GFP-positive, HUES6-derived NSCs are also positive for the NSCs markers SOX1 and Nestin, D. H13B NSCs support L1-GFP retrotransposition and express the NSCs marker SOX3. E. L1-GFP- -positive, HUES6-derived neurons co-label for the neuronal markers β-III tubulin and Map2(A+B). F. L1-GFP-positive, HUES6-derived NSCs can differentiate to a glial lineage (GFAP-positive).

Supplement 2.3
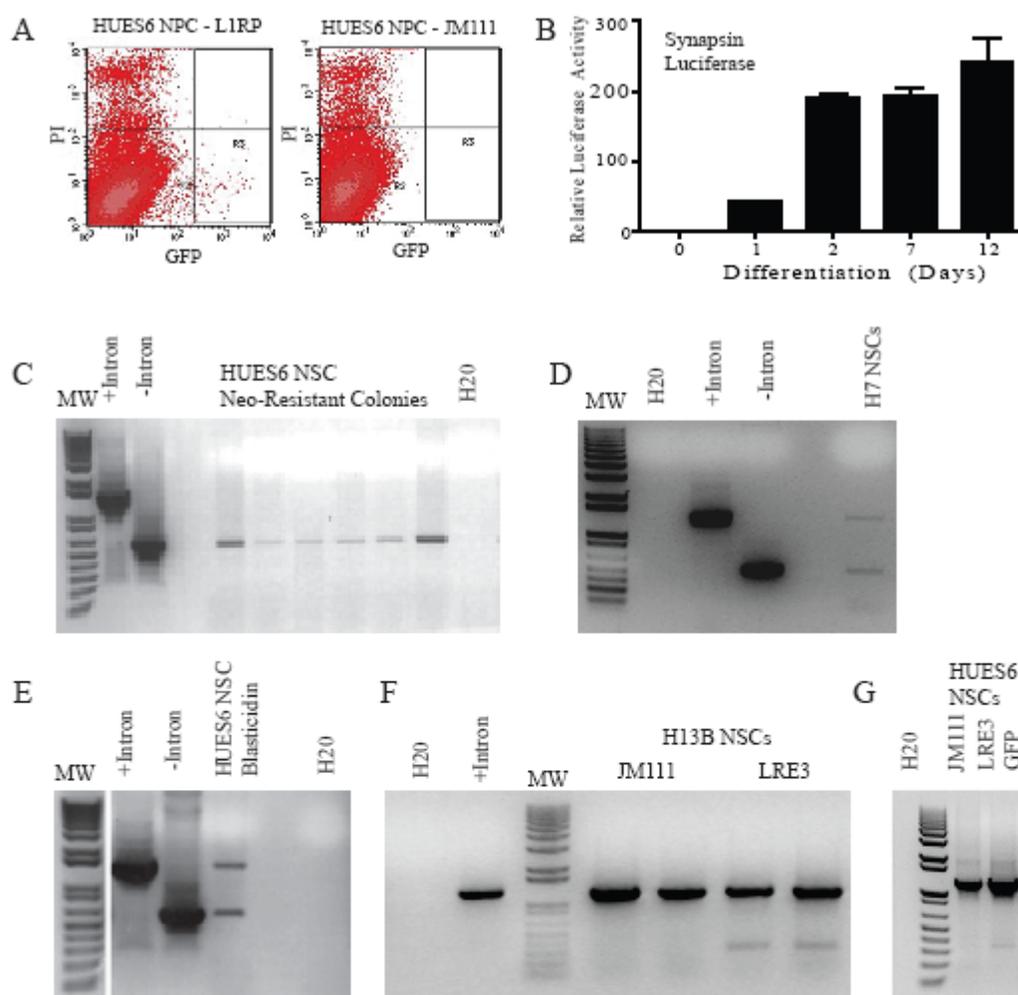
Supplement 2.3

Characterization of L1 retrotransposition in hESC-derived NSCs A. FACs analysis of transfected cells indicates GFP-positive cells with the L1RP plasmid, and no positive events with the JM111 negative control. B. Luciferase assay using the synapsin promoter driving luciferase shows a >250-fold induction of synapsin during neuronal differentiation of hESC-derived NSCs. C-D. Transfection of both HUES6- (C) and H7- (D) derived NSCs with the L1-Neo construct indicates that neo-resistant colonies have experienced a retrotransposition event resulting in removal of the γ-globin intron. E. The L1-blasticidin construct also retrotransposes in HUES6-derived NSCs. F-G. PCR from gDNA of L1-GFP transfected cells indicates that removal of the intron and insertion of the L1 occur in both H13B- (F) and HUES6- (G) derived NSCs.

54

Supplement 2.4
NSC lines exhibit a stable karyotype. A-C. All three HCNS-SCns lines are karyotypically normal and stable. D. Two hundred interphase nuclei were examined by FISH (fluorescence in situ hybridization) using a probe cocktail specifically designed to identify very small populations of cells with changes in chromosome 12 and 17 copy number, a common karyotypic problem with abnormal hES. All cells demonstrated a normal signal pattern for the ETV6 BAP (TEL) gene located on chromosome 12, and all cells demonstrated a normal signal pattern for the 17 centromere on chromosome 17. In summary, no evidence of trisomy 12 and/or trisomy 17 could be detected. HUES6 (E) and H9 ES (F) lines are karyotypically normal.

Supplement 2.4

A

Cell Line CGR1854 Passage 0
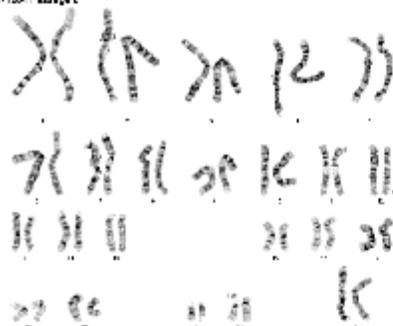
Results	
Number of Metaphase Counted	100
Band Level	450
Number of Metaphase Karyotyped	8
Number of Chromosomes	46
Sex	Female
Comments	None
Karyotype	46,XX

B

Line CGR1854 Passage 15

Results	
Number of Metaphase Counted	20
Band Level	450
Number of Metaphase Karyotyped	6
Number of Chromosomes	46
Sex	Male
Comments	None
Karyotype	46,XY

C

Cell Line CGR1873

Results	
Number of Metaphase Counted	20
Band Level	450
Number of Metaphase Karyotyped	5
Number of Chromosomes	46
Sex	Female
Comments	None
Karyotype	46,XX

D

ETV6 (TEL) BAP/17 Centromere cocktail probe
Normal Signal Pattern

CLG-734
HUES-6 p38

ETV6 gene

17 Cen

12p13 ETV6 (TEL) gene
Chromosome 17 centromere
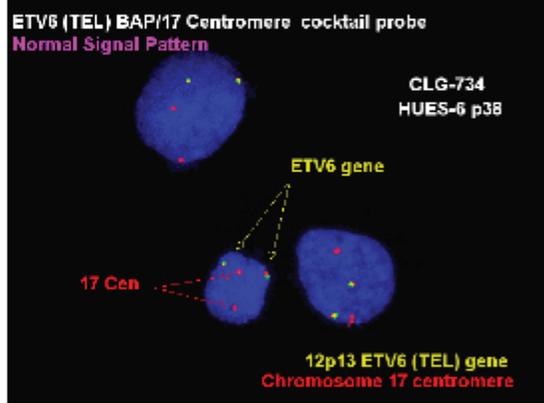
E
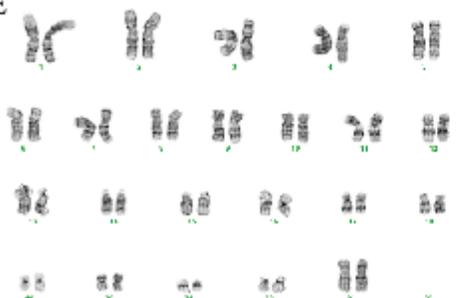
F
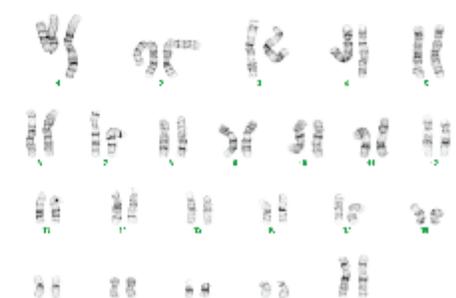
Table 2.1 FACS Analysis of hESC-derived NSCs

Cell Line	Lab[a]	Selection[b]	Plasmid-%EGFP(*)
HUES6	G	NO	0.93 +/- 0.11
HUES6	G	NO	0.25 +/- 0.04
HUES6	G	NO	0.39 +/- 0.03
HUES6	G	NO	0.01 +/- 0.02
H9	G	NO	0.13 +/- 0.01
H7	M	NO	0.42 +/- 0.2
H7	M	YES	9.80 +/- 2.82
H7	M	YES	5.70+/- 0.46
H7	M	YES	2.85 +/- 0.86
H7	M	YES	**UB** 4.65 +/- 0.21
H13B	M	YES	**UB** 3.25 +/ -0.26
H13B	M	YES	**UB** 16.25 +/- 3.6
H13B	M	YES	5.15 +/- 0.54
H13B	M	YES	0.82 +/- 0.1
H13B	M	YES	0.60 +/- 0.2
H13B	M	YES	0.23 +/- 0.05
H9	M	YES	4.21+/-0.84
BG01	M	YES	7.73+/-1.94

[a] Gage or Moran Lab

[b] puromicin, 0.2 ug/ml

(*) From triplicates. Either a JM111 construct or untransfected samples (in triplicates) were used to determine the baseline background fluorescence. UB = ubiquitin promoter, all the remainder LRE3.

Table 2.2 Analysis of L1 Insertions in hESC-derived NSCs

NSC Derivation [a]	Insertional Analysis [b]	Multiple Occurrences [b]	Locus	L1 insertional target site
Clone	Full	NA	3p24	13 kB upstream from several mRNAs, into a LINE element.
1	Full	N	15p11	15 kB upstream from APCDD1 precursor (Adenomatosis polyposis coli down-regulated 1 protein) involved in beta catenin signaling
1	Full	Y	7q21	In an intron of zinc finger protein 804B (ZNF804B)
1	Full	N	2q24	100 kB upstream from EST AA319772
2	Full	Y	10q25	Into an intron of SLC18A2, a synaptic vesicular monoamine transporter
2	Full	N	5q21	100 kB upstream from EST DA377288, in a LINE element
2	Full	N	11q24	10 kB upstream from several ESTs, in an LTR repeat
2	Full	Y	5p13	In an exon of C7, complement component 7 precursor (a component of immune complement system)
2	Full	Y	12q13	5 kB upstream from olfactory receptor OR6C1, in an intron of EST AK127862
1	Partial	N	7p15	Intron of pleckstrin homology domain containing family A (PLEKHA8)
1	Partial	Y	3p14	5 kB upstream from PRICKLE 2, prickle-like protein 2 (nuclear membrane protein expressed in brain, eye and testes)
1	Partial	N	3q22	7 kB downstream from RYK receptor-like tyrosine kinase isoform 1 (growth factor receptor)
1	Partial	Y	Xq21	2.5 kB downstream from GPR174, putative purinergic receptor FKSG79 (G-protein coupled receptor)
1	Partial	N	10q23	90 kB upstream from NGR3 (Neuregulin 3). Brain expressed direct ligand for the ERBB4 tyrosine kinase receptor
1	Partial	N	5p14	180 kB downstream from PRDM9 (involved in transcriptional regulation)
1	Partial	N	10q25	Into a region of ESTs
1	Partial	N	19p13	11 kB downstream from KIAA0892 (secreted protein in the mau-2 family)
1	Partial	N	16q21	70 kB upstream from GOT2, aspartate aminotransferase 2 precursor (role in amino acid metabolism)

[a] Clone, or from either the first or second independent NSC derivation.

[b] Full insertions - both 5' and 3' documented, partial insertions,- only 3' genomic location isolated.

[c] Multiple occurrences were isolated with independent IPCR reaction from same sample

WORKS CITED

Athanikar, J.N., Badge, R.M., and Moran, J.V. (2004). A YY1-binding site is required for accurate human LINE-1 transcription initiation. Nucleic Acids Res *32*, 3846-3855.

Brouha, B., Meischl, C., Ostertag, E., de Boer, M., Zhang, Y., Neijens, H., Roos, D., and Kazazian, H.H., Jr. (2002). Evidence consistent with human L1 retrotransposition in maternal meiosis I. Am J Hum Genet *71*, 327-336.

Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V., and Kazazian, H.H., Jr. (2003). Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci U S A *100*, 5280-5285.

Eriksson, P.S., Perfilieva, E., Bjork-Eriksson, T., Alborn, A.M., Nordborg, C., Peterson, D.A., and Gage, F.H. (1998). Neurogenesis in the adult human hippocampus. Nat Med *4*, 1313-1317.

Freeman, J.D., Goodchild, N.L., and Mager, D.L. (1994). A modified indicator gene for selection of retrotransposition events in mammalian cells. BioTechniques *17*, 46, 48-49, 52.

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

Gage, F.H., Coates, P.W., Palmer, T.D., Kuhn, H.G., Fisher, L.J., Suhonen, J.O., Peterson, D.A., Suhr, S.T., and Ray, J. (1995). Survival and differentiation of adult neuronal progenitor cells transplanted to the adult brain. Proc Natl Acad Sci U S A *92*, 11879-11883.

Garcia-Perez, J.L., Marchetto, M.C., Muotri, A.R., Coufal, N.G., Gage, F.H., O'Shea, K.S., and Moran, J.V. (2007). LINE-1 retrotransposition in human embryonic stem cells. Hum Mol Genet *16*, 1569-1577.

Gilbert, N., Lutz, S., Morrish, T.A., and Moran, J.V. (2005). Multiple fates of L1 retrotransposition intermediates in cultured human cells. Mol Cell Biol *25*, 7780-7795.

Han, J.S., and Boeke, J.D. (2004). A highly active synthetic mammalian retrotransposon. Nature *429*, 314-318.

Kazazian, H.H., Jr. (1999). An estimated frequency of endogenous insertional mutations in humans. Nat Genet *22*, 130.

Kazazian, H.H., Jr. (2004). Mobile elements: drivers of genome evolution. Science *303*, 1626-1632.

Kimberland, M.L., Divoky, V., Prchal, J., Schwahn, U., Berger, W., and Kazazian, H.H., Jr. (1999). Full-length human L1 insertions retain the capacity for high frequency retrotransposition in cultured cells. Hum Mol Genet *8*, 1557-1560.

Kubo, S., Seleme Mdel, C., Soifer, H.S., Perez, J.L., Moran, J.V., Kazazian, H.H., Jr., and Kasahara, N. (2006). L1 retrotransposition in nondividing and primary human somatic cells. Proc Natl Acad Sci U S A *103*, 8036-8041.

Kulpa, D.A., and Moran, J.V. (2005). Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. Hum Mol Genet *14*, 3237-3248.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson,

L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., and Chen, Y.J. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K.W., Vogelstein, B., and Nakamura, Y. (1992). Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. Cancer Res *52*, 643-645.

Moran, J., and Gilbert, N. (2002). Mammalian LINE-1 retrotransposons and related elements (Washington, DC, ASM Press).

Moran, J.V., DeBerardinis, R.J., and Kazazian, H.H., Jr. (1999). Exon shuffling by L1 retrotransposition. Science *283*, 1530-1534.

Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D., and Kazazian, H.H., Jr. (1996). High frequency retrotransposition in cultured mammalian cells. Cell *87*, 917-927.

Morrish, T.A., Gilbert, N., Myers, J.S., Vincent, B.J., Stamato, T.D., Taccioli, G.E., Batzer, M.A., and Moran, J.V. (2002). DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat Genet *31*, 159-165.

Muotri, A.R., Chu, V.T., Marchetto, M.C., Deng, W., Moran, J.V., and Gage, F.H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature *435*, 903-910.

Myers, J.S., Vincent, B.J., Udall, H., Watkins, W.S., Morrish, T.A., Kilroy, G.E., Swergold, G.D., Henke, J., Henke, L., Moran, J.V., Jorde, L.B., and Batzer, M.A. (2002). A comprehensive analysis of recently integrated human Ta L1 elements. Am J Hum Genet *71*, 312-326.

Ostertag, E.M., DeBerardinis, R.J., Goodier, J.L., Zhang, Y., Yang, N., Gerton, G.L., and Kazazian, H.H., Jr. (2002). A mouse model of human L1 retrotransposition. Nat Genet *32*, 655-660.

Ostertag, E.M., Prak, E.T., DeBerardinis, R.J., Moran, J.V., and Kazazian, H.H., Jr. (2000). Determination of L1 retrotransposition kinetics in cultured cells. Nucleic Acids Res *28*, 1418-1423.

Prak, E.T., Dodson, A.W., Farkash, E.A., and Kazazian, H.H., Jr. (2003). Tracking an embryonic L1 retrotransposition event. Proc Natl Acad Sci U S A *100*, 1832-1837.

Sambrook, J., Fritsch, E., and Maniatis, A. (1989). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, Cold Spring Harbor Press).

Shi, X., Seluanov, A., and Gorbunova, V. (2007). Cell divisions are required for L1 retrotransposition. Mol Cell Biol *27*, 1264-1270.

Swergold, G.D. (1990). Identification, characterization, and cell specificity of a human LINE-1 promoter. Mol Cell Biol *10*, 6718-6729.

Symer, D.E., Connelly, C., Szak, S.T., Caputo, E.M., Cost, G.J., Parmigiani, G., and Boeke, J.D. (2002). Human l1 retrotransposition is associated with genetic instability in vivo. Cell *110*, 327-338.

Tang, Y., Nyengaard, J.R., De Groot, D.M., and Gundersen, H.J. (2001). Total regional and global number of synapses in the human brain neocortex. Synapse (New York, NY *41*, 258-273.

Taupin, P., and Gage, F.H. (2002). Adult neurogenesis and neural stem cells of the central nervous system in mammals. J Neurosci Res *69*, 745-749.

Tchenio, T., Casella, J.F., and Heidmann, T. (2000). Members of the SRY family regulate the human LINE retrotransposons. Nucleic Acids Res *28*, 411-415.

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

Uchida, N., Buck, D.W., He, D., Reitsma, M.J., Masek, M., Phan, T.V., Tsukamoto, A.S., Gage, F.H., and Weissman, I.L. (2000). Direct isolation of human central nervous system stem cells. Proc Natl Acad Sci U S A *97*, 14720-14725.

Wei, W., Morrish, T.A., Alisch, R.S., and Moran, J.V. (2000). A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events. Anal Biochem *284*, 435-438.

Yang, N., Zhang, L., Zhang, Y., and Kazazian, H.H., Jr. (2003). An important role for RUNX3 in human L1 transcription and retrotransposition. Nucleic Acids Res *31*, 4929-4940.

Yeo, G.W., Xu, X., Liang, T.Y., Muotri, A.R., Carson, C.T., Coufal, N.G., and Gage, F.H. (2007). Alternative splicing events identified in human embryonic stem cells and neural progenitors. PLoS computational biology *3*, 1951-1967.

Zhang, S.C., Wernig, M., Duncan, I.D., Brustle, O., and Thomson, J.A. (2001). In vitro differentiation of transplantable neural precursors from human embryonic stem cells. Nat Biotechnol *19*, 1129-1133.

Zhao, C., Teng, E.M., Summers, R.G., Jr., Ming, G.L., and Gage, F.H. (2006). Distinct morphological stages of dentate granule neuron maturation in the adult mouse hippocampus. J Neurosci *26*, 3-11.

CHAPTER 3

EVIDENCE FOR ENDOGENOUS L1 RETROTRANSPOSITION IN HUMAN BRAIN TISSUES

Abstract

Although active L1 retrotransposition in fetal and hESC derived neural stem cells (NSCs) is a tantalizing suggestion that L1 activity may occur in the nervous system, by itself this is insufficient evidence. Therefore, we sought to determine whether there is increased ORF2 copy number in primary human brain tissues as compared to somatic tissues from the same individual. Although the increases in L1 copy number were variable between the brain tissues tested, we found a consistent increase in hippocampal ORF2 content. To investigate a component of the mechanism responsible for L1 retrotransposition in the brain, we studied methylation of the L1 promoter in human developmental tissues and found decreased methylation in the brain in first trimester fetal tissue as compared to skin tissue.

Introduction

Active retrotransposition of an engineered human L1 in various types of NSCs is a suggestion that L1 retrotransposition may also be occurring in the human nervous system. Using L1-GFP as a visible marker allows us to easily recognize cells which are permissive for retrotransposition and identify the insertional point. Nevertheless, these experiments only reflect the activity of a single retrotransposon, and there are 150 within the genome whose activity remains unknown. Additionally, they investigate the role of an episomal plasmid, which is not subject to contextual chromosomal epigenetic regulation, such as methylation, or the effects of nearby genes on transcriptional regulation.

The challenge was to develop a technique for investigating the activity of endogenous retrotransposons. Since 17% of the human genome consists of L1 fragments and various open reading frames (Lander et al.), distinguishing the activity of the ~150 endogenous full length theoretically retrotransposition competent L1's is a daunting task. Without an engineered tag such as GFP to mark the location of a *de novo* insertion, elucidating a novel insertion point is extremely difficult.

The main approaches for studying novel L1 insertions to date are PCR based, and have been developed to identify human dimorphic retrotransposon insertions. For instance, suppression PCR (Lavrentieva et al.) has been combined with both subtractive as well as with differential hybridization, to identify human-specific endogenous retrovirus insertions (Buzdin et al.). Additionally, L1 display, a technique which employs PCR primers specific to L1 subfamilies, has identified dimorphic L1 insertions (Ovchinnikov et al., ; Sheen et al.). The most current approach, termed ATLAS (for amplification typing of L1 active subfamilies), merges suppression PCR with selective linker usage in order to selectively amplify and display DNA fragments with human-specific L1s and their flanking sequences (Badge et al.). This technique has been successfully used to identify L1s which are polymorphic in human populations, but germline within the individual. These techniques are practical if all cells sampled carry the *de novo* L1, so that it is present in equal amount compared to other genomic L1s. However if the sample is heterogeneous these techniques will amplify more penetrant genomic locations preferentially. Although these techniques have worked well in other settings, we required a technique to measure whole genome L1 copy number for the active families of L1 which would not be biased by

prevalence of a given insertion in the sample. Therefore, we developed a quantitative PCR method for estimating the overall copy number of young, active L1s, those in the Ta-1 and Ta-0 L1 families.

The structure of the L1 promoter within the 5' untranslated region (UTR) identifies a CpG island just 3' to several regulatory sites within the promoter. Immediately downstream to the CpG island there are binding sites for a number of regulatory elements, including two SRY box binding domains, and a RUNX binding site(Becker et al., ; Kurose et al., ; Tchenio et al.). This suggests that methylation of the CpG island may influence transcription factor binding to the regulatory sites. L1 CpG island methylation has previously been identified as crucial for repression of L1 activity in transformed cells(Hata and Sakaki, ; Thayer et al.). Additionally, hypomethylation of the L1 5'UTR has been associated with tumor cultures as well as leukemia progression (Roman-Gomez et al.). Data suggest that although CpG island methylation successfully inhibits L1 activity, it is not through inhibiting the binding of the YY-1 transcription factor at the L1 initiation start site which is 5' to the CpG island(Hata and Sakaki). This implies that CpG island methylation works through other mechanisms to inhibit L1 retrotransposition, perhaps by the recruitment of SRY-Box-2 (SOX2), a transcription factor which has previously been shown to inhibit L1 retrotransposition in NSCs (Muotri et al.) and whose binding sites fall at the 3' end of the L1 CpG island. It is difficult to assess the effect of SOX2 on L1 activity in a comparable human *in vitro* situation because SOX2 changes much more slowly in human NSCs, such that decreases in expression require 2-3 weeks rather than less than 24 hours (unpublished observation).

Methods

Bisulfite analysis

Fetal tissues were obtained from donations resulting from voluntary pregnancy terminations and were collected by the Birth Defects Research Lab at the University of Washington, Seattle, WA (NIH HD 000836). Brain and skin gDNA from 80-day-old female fetal tissue was isolated using standard phenol-chloroform extraction techniques. Subsequently, DNA was digested with the restriction enzyme *DraI* and the bisulfite conversion reaction was performed as per manufacturer's instructions using the Epitect kit (Qiagen). The bisulfite conversion was performed twice consecutively to achieve a conversion rate >90% in LINE-1 repeat regions. Primers were designed using Methyl Primer Express; primers for LINE-1 converted 5'UTR region: forward 5-'AAGGGGTTAGGGAGTTTTTTT-3' and reverse 5'-TATCTATACCCTACCCCCAAAA-3'. PCR products were cloned into TOPO TA 2.1 plasmids (Invitrogen) and 100 bacterial colonies were analyzed by sequencing for each tissue sample.

L1 5'UTR sequences after bisulfite treatment were aligned using blastn to a database of full-length L1's with two open reading frames, extracted from the May 2004 assembly of the human genome (hg17). The blastn alignment used a mismatch penalty of -1 with a reward for match hESC of +1. The best match for each brain or skin sequence to the genomic L1 database was determined. The alignment neglected cytosine nucleotides in the L1 database to prevent conversion bias. The fraction of CpG sites that are unmethylated was calculated by comparing CpG dinucleotides in the L1 database to the corresponding sequences from brain and skin; the fraction

converted is the proportion of TG dinucleotides in brain and skin sequences at CpG sites in the genomic L1 database to total number of CpG sites in the region. Next, to determine differences in methylation between brain and skin L1s, a CDF plot was generated for all the sequences that aligned above an alignment cutoff. The alignment cutoff was one standard deviation below the mean of the alignment score for all sequences aligned. Conversion efficiency was assessed by analyzing the conversion rate at genomic cytosine nucleotides that were not upstream of a guanine nucleotide. The same analysis was carried out for all possible dinucleotides and possible conversions of the first nucleotide.

Quantitative PCR

Oligonucleotide PCR primers and TaqMan-MGB probes were designed using Primer Express software (Applied Biosystems). Primers were purchased from Allele Biotech, and probes were purchased from Applied Biosystems. L1 primers were verified using the L1 database (http://l1base.molgen.mpg.de/), and matched a minimum of 140 of 145 full-length L1s with two open reading frames in the database. Human tissues were obtained from the NICDH Brain and Tissue Bank for Developmental Disorders at the University of Maryland, Baltimore, MD. Patients were between 17 and 22 years old. Human gDNA was extracted and purified from human tissues using the DNeasy Blood & Tissue kit (Qiagen), according to the manufacturer's instructions. PCR reactions were carried out using 80 pg of DNA, and were verified empirically as amplifying with a CT between 20 and 25 (n=16). Quantitative PCR experiments were performed using an ABI Prism 7000 sequence

detection system and Taqman Gene Expression Mastermix from Applied Biosystems. Data analysis was performed using the SDS 2.3 software (Applied Biosystems). The multiplexing reaction was optimized by limiting reaction components until both reactions amplified as completely as each individual reaction. Primer efficiency was verified using a PCR standard curve of plasmid DNA to have a slope of near -3.32. Standard curves of genomic DNA ranging from 2 ng to16 pg were performed to verify the 80 pg dilution used is within the linear range. ORF2 probes were conjugated to the fluorophore label VIC and all other probes were conjugated with 6FAM. Primers and probes are listed below: L1 ORF2 – 1: Matches 4,560 genomic L1's. Probe – CTGTAAACTAGTTCAACCATT, Fw – TGCGGAGAAATAGGAACACTTTT. Rv – TGAGGAATCGCCACACTGACT. L1 ORF2 – 2: Matches 2,918 genomic L1's. Probe – AGGTGGGAATTGAAC, Fw – CAAACACCGCATATTCTCACTCA, Rv – CTTCCTGTGTCCATGTGATCTCA. L1 5'UTR – 1: Matches 965 genomic L1's. Probe – AAGGCTTCAGACGATC, Fw – GAATGATTTTGACGAGCTGAGAGAA Rv – GTCCTCCCGTAGCTCAGAGTAATT. L1 5'UTR-2: Matches 876 genomic L1's. Probe – TCCCAGCACGCAGC, Fw – ACAGCTTTGAAGAGAGCAGTGGTT, Rv -AGTCTGCCCGTTCTCAGATCT. SATA: Matches millions of tandem copies in genome with little variability in sequence. Probe – TCTTCGTTTCAAAACTAG, Fw – GGTCAATGGCAGAAAAGGAAAT, Rv – CGCAGTTTGTGGGAATGATTC. 5S RNA: Matches 35 copies in the genome. Probe – AGGGTCGGGCCTGG, Fw – CTCGTCTGATCTCGGAAGCTAAG, Rv – GCGGTCTCCCATCCAAGTAC. HERV-H: Matches 99 copies in the genome (primers). Probe –

CCCTTCGCTGACTCTC, Fw – AATGGCCCCACCCCTATCT, Rv – GCGGGCTGAGTCCGAAA.

Results

Evidence for endogenous L1 insertions in human brain

We asked whether it would be possible to obtain data about endogenous L1 elements in the human brain. Although human NSC models combined with an engineered L1-GFP are useful *in vitro* models for probing L1 retrotransposition, this approach represents only the activity of a single L1 element. The human genome is estimated to carry 80-100 active L1s, which are regulated by contextual chromatin structure (Brouha et al.). To investigate the activity of endogenous L1s in the human brain, we developed a quantitative multiplexing PCR strategy to probe L1 quantity in the genome. We hypothesized that active retrotransposition in the human brain would result in increased L1 DNA content as compared with other somatic tissues, such as heart and liver (Fig. 3.1A).

To detect L1 content, we designed two different Taqman probes (conjugated with the VIC fluorophore) that match conserved regions in ORF2 nearest the 3' end of the L1. We also designed a variety of different probes to control for DNA content (all conjugated with the 6FAM fluorphore). These controls include other non-mobile repeat regions in the genome, including SATα, HERVH, and 5sRNA. In addition, we used two different probes to conserved regions in the L1 5'UTR. Since the majority of novel L1 retrotransposition events are 5' truncated, the quantity of 5'UTR should remain unchanged despite increases in the ORF2 regions (Grimaldi et al., ; Moran and

Gilbert). All probes amplified a single-sized product as expected (Fig S3.1B). The strength of the multiplex approach is the presence of an internal control within each PCR reaction, allowing for detailed measurement of the ratio of ORF2 content to various internal controls.

We compared genomic DNA derived from three adult humans, and in each case we analyzed four different tissues per individual: hippocampus, cerebellum, liver and heart. For each tissue, we derived genomic DNA from three independent samples and ran all samples in triplicate. We found in all comparisons a statistically significant increase in ORF2 content relative to control DNA (Fig 3.1B) in the brain as compared to somatic tissues (n = 9; 3 independent samples per individual). Interestingly, the ORF2 content was statistically significantly higher in the hippocampus than in the cerebellum in all cases, independent of ORF2 or control probes (data from individual donors; S3.1A). Whether this higher level is related to continued neurogenesis in the subgranular zone of the hippocampus during adult life, with L1 retrotransposition occurring in stem cells as they undergo cell division, remains to be determined. These data strongly suggest that more L1 retrotransposition may occur in the brain developmentally and/or during adulthood than in other somatic tissues.

Based on these data, the ratio of ORF2/control indicates a 5-8% increase in L1 ORF2 in hippocampus and a 1-2% increase in the cerebellum. Since the L1 ORF2 probes match ~3,500 L1s in the UCSC genome database, and the efficiency of the qPCR probes according to the genomic DNA standard curve is approximately 80-85%, the number of insertions per cell can be estimated. The estimate suggests that

there are 30-60 insertions per cerebellar cell and 175-225 insertions per hippocampal cell.

To corroborate this estimate, we sought to verify the copy number increase by spiking known numbers of L1 plasmid copies into liver and heart genomic DNA (gDNA) to validate how many copies would be needed to bring liver and heart gDNA equal to hippocampal DNA. Individual cell gDNA content was estimated based on the equation, cell gDNA $= 3*10^9$(# bp's) * 2(dipoid) *660 (MW 1 bp) * $1.67*10^{12}$ (weight 1 dalton), resulting in the approximation of one cell containing 6.6 pg gDNA (Forslund et al.). qPCR reactions were run with 80 pg of genomic DNA, which is approximately equal to 12.12 genomes worth of gDNA. L1 plasmid copy number was estimated using the information that copy # = 650 daltons (weight 1 bp) * 18230 (# bp of plasmid)* 1.7 x 10^{-24} grams (weight 1 dalton), results in the estimation that 1 pg of plasmid = approximately 50,000 copies of L1. Plasmid was linearized using *NotI* restriction digest.

Consequently, we spiked 80 pg of liver and heart gDNA with 10, 100, 1000, and 10000 copies of L1 plasmid and multiplexed the L1 ORF2 primer with the internal control 5sRNA (Fig. 3.1D). The data showed that the hippocampus contained between 1,000 and 10,000 copies of L1 more than the heart or liver DNA (when considering 12 genomes of DNA). Replication of this experiment three times indicated that the increase in copy number was between 80 and 500 copies per genome in the hippocampus, roughly supporting the above estimate.

Methylation analysis of L1 5'UTR

To investigate a possible mechanism by which increased L1 retrotransposition may occur in the human brain as opposed to other somatic tissues, we investigated the methylation status of the L1 5'UTR. Previous studies in L1 transgenic mice indicated that L1-GFP positive neurons are found throughout the brain, likely indicating that they occurred early during embryonic development (Muotri et al.). L1 is unlikely to be active in the adult brain, except perhaps in neurogenic regions such as the dentate gyrus of the hippocampus or the subventricular zone, since neurons are post-mitotic and L1 retrotransposition requires nuclear breakdown and therefore cell division (Shi et al.). Therefore, we hypothesized that differences in brain and skin L1 promoter methylation would likely be found in early embryonic developmental tissues. To this end we analyzed fetal brain and skin tissues from two late first trimester (day 80-82) fetuses, one male and one female. Quantitative PCR from these tissues indicates that there is likely to be some increase in L1 ORF2 gDNA in the brain tissue, although n=1 sample in each group precludes any statistics (Fig S3.2) Therefore, we used these samples to investigate the methylation of a large number of L1 5'UTR promoters from Ta-1 and Ta-0 L1 families across the entire genome.

The L1 5'UTR contains a CpG island meeting the definition criteria of >60% G+C content and an observed CpG frequency >0.6. We hypothesized that the L1 5'UTR in fetal brain tissue would be less methylated than in somatic tissue from the same donor. We performed bisulfite conversion on genomic DNA derived from brain and matched skin tissue and analyzed a 363 bp region of the L1 5'UTR containing 20 CpG sites (Fig. 3.2A). Primers were designed to match a conserved region of the

5'UTR matching a maximum number of L1-Ta elements that are full length and retrotransposition competent.

We found that the L1 5'UTR promoter was significantly less methylated in the fetal brain sample as compared to the matched skin sample (P = 0.0079 d80 female, P = 0.0034 d82 female) (Fig. 3.2B). This comparison collapsed 20 CpG sites and included all L1 elements with an inclusion threshold of one standard deviation below the mean (68% identity) to the consensus active L1 sequence (Supp 3.3A). A comparison of all dinucleotide pairs within the L1 5'UTR sequences showed a statistically significant difference between brain and skin only at the CpG sites and not at any other nucleotides, suggesting there was no sampling bias (Fig. S3.3C).

Individual analyses of sequences with the highest sequence identity to a full length open reading frame genomic L1 showed a clear bias towards more unmethylated sequences in the brain samples, such that one brain sequence was unmethylated at all CpG sites (Fig. 3.2C). Analysis of each individual CpG site within the L1 5'UTR showed a greater degree of hypomethylation at the 3' end of the sequences in both brain and skin samples, with the 5' CpG's exhibiting 70-90% methylation, whereas the more 3' CpG's exhibited greater variability between brain and skin but were lower in both cases, with 40-70% methylation. In all cases, the bisulfite conversion efficiency was >90% (Fig. S3.3C).

Discussion

Although the engineered L1 is a powerful tool for investigating and following the activity of a single L1, we were also interested in the activity of endogenous L1s, of which there are approximately 100 in the human genome (Brouha et al.). We found evidence for increased L1 content in the human brain, specifically for the 3' portion of ORF2, which would be inserted before premature 5' truncation. In all comparisons, we found a statistically significant increase in L1 ORF2 content in the brain, most notably in the hippocampus. The increase in L1 content here is slight: 1-2% for the cerebellum and 8-11% for the hippocampus.

Given that the L1 ORF2 probes match approximately 3,000-4,500 L1s in the UCSC genome browser, this increase could indicate as many as 250 insertions per genome, a seemingly non-trivial quantity. Spiking of liver and heart gDNA with a known number of copies of L1 indicates that the estimate of 80-250 copies of L1 per cell genome of DNA is a valid estimate.

To address a possible mechanism for L1 retrotransposition in the brain as opposed to other somatic tissues, we used early human developmental tissue to study the pattern of methylation of the L1 promoter in the brain as opposed to skin tissue from the same individual. We chose these tissues based on the L1 mouse data which indicated insertions throughout the mouse brain, suggesting embryonic insertion events (Muotri et al.). We found a statistically significant hypomethylation of the L1 promoter in the brain as compared to skin. Although many analyzed sequences in both the brain and skin contained a small, seemingly sporadic number of unmethylated CpG sites (of the 20 that we assayed), only in the brain were there multiple L1

promoter sequences that were entirely unmethylated. These data, together with luciferase assay data indicating increased activation of the L1 promoter during the early part of NSC differentiation, suggest a specific developmental time course during both human development and the individual NSC's development that may be regulating whether retrotransposition occurs.

Overall, our findings indicate L1s are active in human brain cells and contribute to a non-trivial, human genomic neuronal mosaicism. Future experiments will focus on determining whether the insertions are random in the genome, and what the function of neuronal mosaicism might be.

Figure 3.1

Multiplex quantitative PCR analysis of L1 sequences in human tissues.

A. Samples of hippocampus, cerebellum, heart, and liver were derived from the same individuals. After DNA extraction, a Taqman multiplex qPCR approach was used to compare the number of L1 ORF2 sequences in the human genome, using 80 ng per reaction. Primers for L1 ORF2 were multiplexed with primers for control sequences. We hypothesized that our methods would detect a higher number of L1 ORF2 sequences in the brain compared to somatic tissues due to *de novo* L1 neuronal retrotransposition.

B. Abbreviations, Hi = Hippocampus, C = Cerebellum, H = Heart, and L = Liver. The inverse ratio of ORF2/5'UTR represents the amount of L1 ORF2 DNA sequence in each sample relative to the amount of L1 5'UTR. Under these conditions, L1 ORF2 sequences are more frequent in brain (hippocampus + cerebellum) when compared to somatic (heart + liver) tissue from three individuals. Similar results were obtained when different primers/probe for ORF2 (See Methods) were multiplex/normalized to other control sequences. Controls include the human endogenous retrovirus-H sequences (HERV) and 5S ribosomal RNA gene (5sRNA), both of which are non-mobile conserved repetitive regions in the genome. The satellite alpha sequence exists as highly repetitive tandem repeats (SATA). Two different probes towards the L1 T'UTR were also used as internal controls. These graphics were originated by grouping different individuals, represented in Supplement 3.1. Error bars in all panels show S.E.M.

C. Multiplexing of control 5sRNA primers with control SATα primers indicated no change in copy number.

D. To estimate changes in L1 copy number, hippocampal tissue was compared to liver and heart samples from a single individual that were spiked with estimated plasmid copy number of L1 (10, 100, 1000, and 10000 copies of L1 plasmid). 80 pg of genomic DNA were run per qPCR reaction.

Figure 3.1

Figure 3.2

A

B

C

D
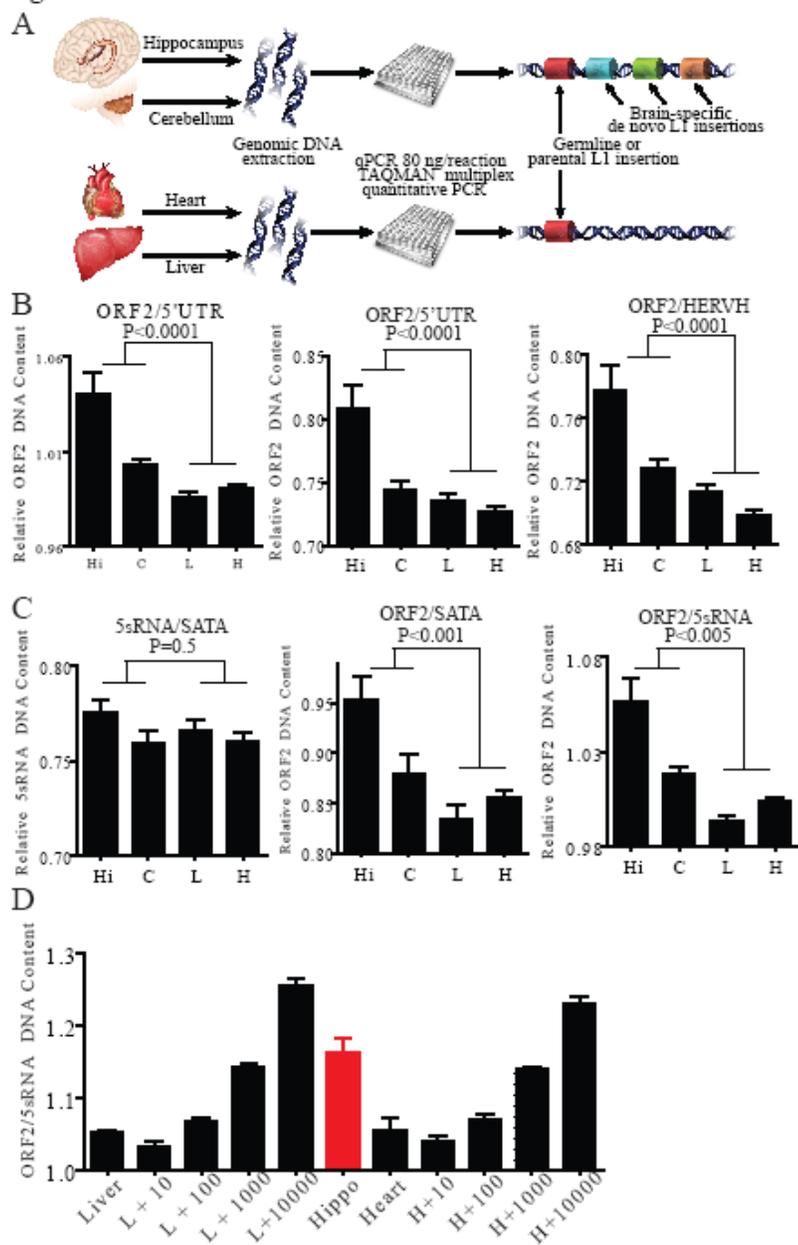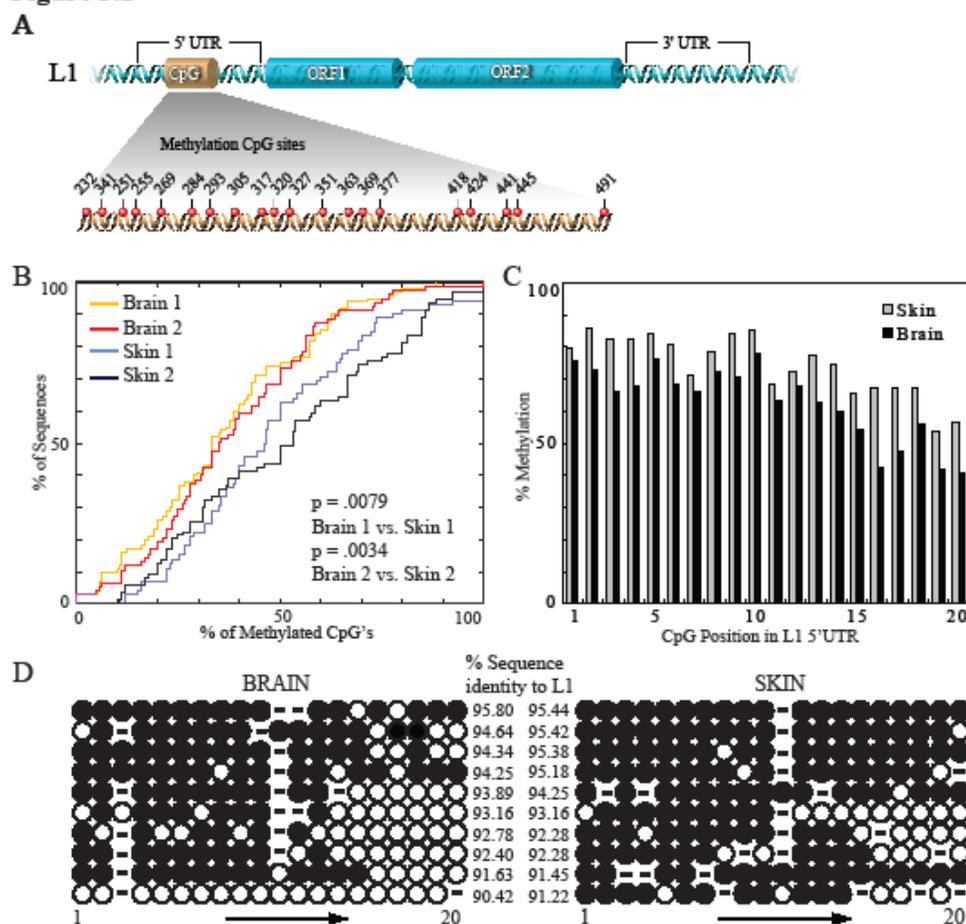
Figure 3.2

Methylation analysis of the L1 5'UTR in human development.

A. Analysis of 20 CpG sites, all found within a CpG island in the 5' UTR of L1.

B. CDF plot, the overall analysis of 72 sequences from brain and 90 samples from skin from individual 1 (d80 female) and 78 skin and 90 brain samples from a individual 2 (d82 male). The analysis was collapsed across the CpG sites and compares the overall level of methylation of the sequences.

C. Analysis of each of the 20 CpG sites analyzed in the 5'UTR. Generally CpG sites were more methylated at the 5' end and less methylated at the 3' end of the CpG island, with a greater degree of methylation seen at all sites in the skin.

D. Individual methylation analysis of those sequences with the highest similarity to the consensus active L1. Open circles = unmethylated CpG site, closed circles = methylated CpG sites. One brain sequence exhibited a completely unmethylated L1 5'UTR.

Supplement 3.1
A. Multiplex qPCR data from each of three individuals across four analyzed tissues. These data are collapsed across individuals in Figure 3.1
B. Each primer set amplified only a single PCR product, as expected.

Supplement 3.1

A
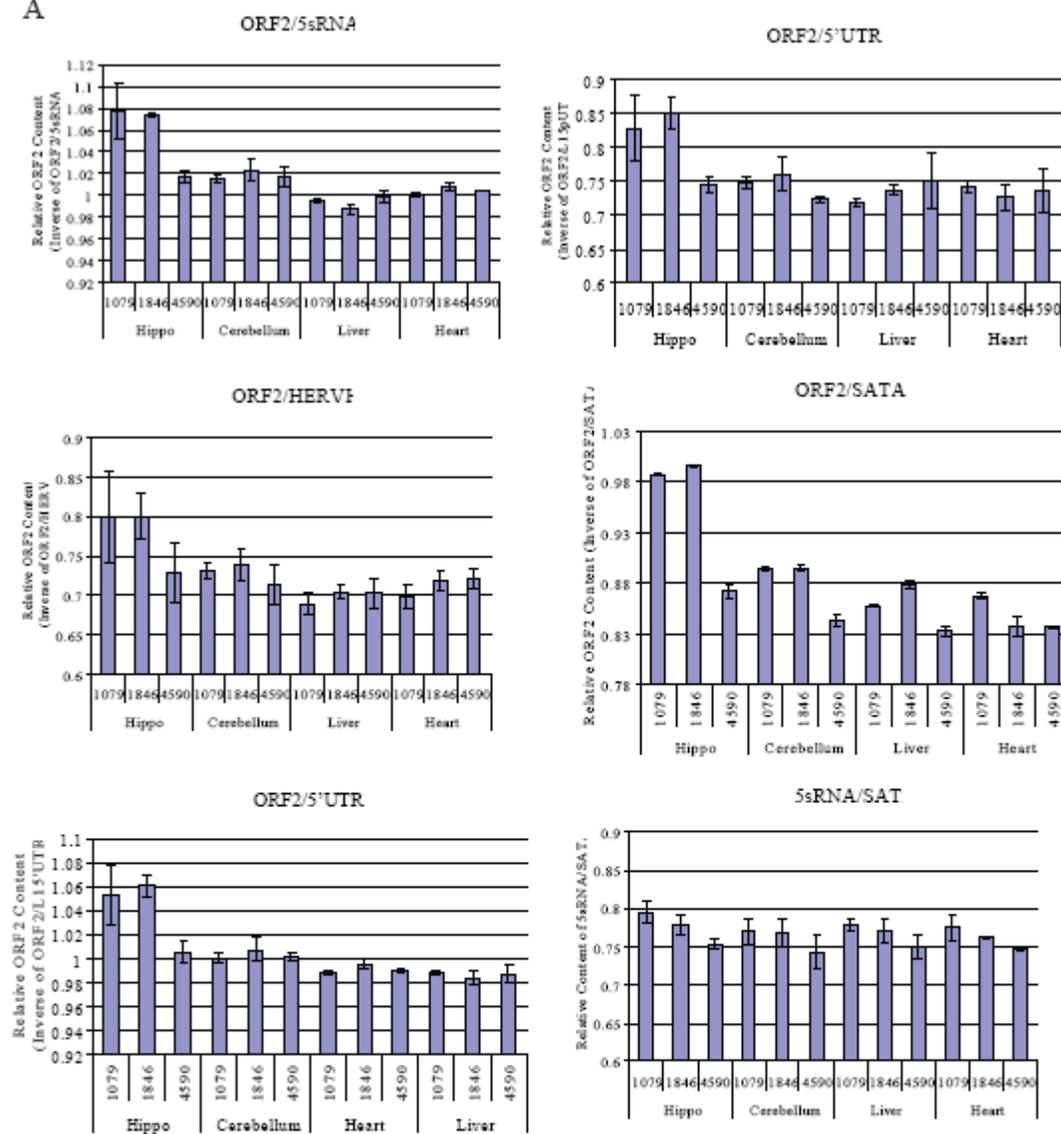
B
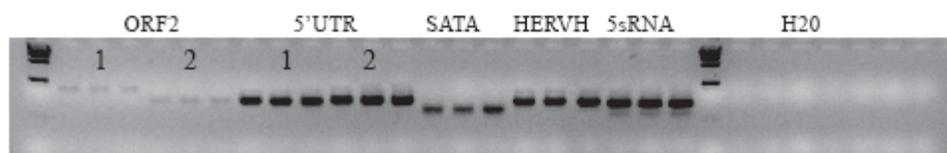
Supplement 3.2
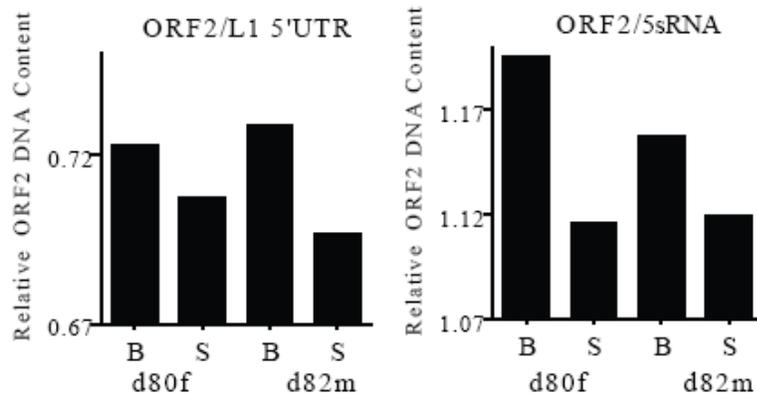
Supplement 3.2

QPCR data from human embryonic tissues. Multiplexing data from ORF2 primers combined with either L1 5'UTR or 5sRNA primers. B = brain, S = skin. Samples are d80f = gestational day 80 female tissues from a single individual, d82m = gestational day 82 male also from a single individual. n = 1 per group. These tissues were utilized for the methylation analysis.

Supplement 3.3

Methylation analysis of the L1 5'UTR.

A-B. Percent sequence identity of each brain and skin 5'UTR analyzed (X axis) as compared to the consensus active L1 5'UTR. The cutoff for analysis was the mean minus one standard deviation. The Y axis is the percent of unmethylated CpG's.

C. Percent conversion of isolated C's (not a part of a CpG dinucleotide) in the sequence. This measures the completeness of the bisulfite conversion reaction and indicates a conversion efficiency of >90% for all analyzed sequences, with no statistically significant difference between samples. (left, D80 female, right, D82 male)

D. Analysis of all dinucleotide pairs in the L1 5'UTR analyzed sequences. We compared all dinucleotides of analyzed sequences as compared to the consensus active L1 sequence and looked for changes in the second dinucleotide. The only statistically significant difference between brain and skin is in the conversion of CG sites with the first nucleotide changing to T (more unmethylated sequences in brain as compared to skin). Looking at all other dinucleotides changes to all other possible combinations for the first in the pair showed no other significant changes. These data indicate no large-scale sampling bias of different L1 subtypes between data sets.

Supplement 3.3

A D82 MALE

B D80 FEMALE

C

D

Dinucleotide Changes	P Value	Dinucleotide Changes	P Value	Dinucleotide Changes	P Value
CG..T	7.9359510e -03	GG..C	9.7995262e -01	GA..C	1.0000000e+00
AC..G	1.6797540e -01	TG..G	9.8041167e -01	AT..C	1.0000000e -00
TG..C	1.6976372e -01	GA..T	9.9037454e -01	TA..G	1.0000000e+00
CA..T	2.6006163e -01	AG..T	9.9338105e -01	AG..C	1.0000000e -00
CG..G	3.4166445e -01	AA..G	9.9670056e -01	AA..C	1.0000000e -00
CT..T	3.7374136e -01	AC..T	9.9833303e -01	GT..A	1.0000000e+00
AG..G	6.6612273e -01	AT..G	9.9836229e -01	TA..C	1.0000000e -00
GG..A	6.7431382e -01	GA..A	9.9853156e -01	GT..C	1.0000000e+00
CG..A	7.2164350e -01	CA..A	9.9881696e -01	TT..C	1.0000000e -00
CT..A	7.5173357e -01	CC..G	9.9878307e -01	TT..A	1.0000000e -00
AT..T	7.6674550e -01	TC..A	9.9977063e -01	CC..A	1.0000000e -00
AA..T	8.4693846e -01	CA..G	9.9980407e -01	GC..C	1.0000000e -00
GC..A	8.6640707e -01	CT..G	9.9990581e -01	TT..G	1.0000000e+00
GG..T	9.5007025e -01	CC..T	9.9999745e -01	TA..A	1.0000000e -00
TC..G	9.5601742e -01	TG..A	9.9999997e -01	GC..T	1.0000000e -00
GT..T	9.5984078e -01	AC..C	1.0000000e -00	TC..C	1.0000000e+00

WORKS CITED

Badge, R.M., Alisch, R.S., and Moran, J.V. (2003). ATLAS: a system to selectively identify human-specific L1 insertions. Am J Hum Genet *72*, 823-838.

Becker, K.G., Swergold, G.D., Ozato, K., and Thayer, R.E. (1993). Binding of the ubiquitous nuclear transcription factor YY1 to a cis regulatory sequence in the human LINE-1 transposable element. Hum Mol Genet *2*, 1697-1702.

Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V., and Kazazian, H.H., Jr. (2003). Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci U S A *100*, 5280-5285.

Buzdin, A., Khodosevich, K., Mamedov, I., Vinogradova, T., Lebedev, Y., Hunsmann, G., and Sverdlov, E. (2002). A technique for genome-wide identification of differences in the interspersed repeats integrations between closely related genomes and its application to detection of human-specific integrations of HERV-K LTRs. Genomics *79*, 413-422.

Forslund, O., Antonsson, A., Higgins, G., Ly, H., Delius, H., Hunziker, A., and de Villiers, E.M. (2003). Nucleotide sequence and phylogenetic classification of candidate human papilloma virus type 92. Virology *312*, 255-260.

Grimaldi, G., Skowronski, J., and Singer, M.F. (1984). Defining the beginning and end of KpnI family segments. Embo J *3*, 1753-1759.

Hata, K., and Sakaki, Y. (1997). Identification of critical CpG sites for repression of L1 transcription by DNA methylation. Gene *189*, 227-234.

Kurose, K., Hata, K., Hattori, M., and Sakaki, Y. (1995). RNA polymerase III dependence of the human L1 promoter and possible participation of the RNA polymerase II factor YY1 in the RNA polymerase III transcription system. Nucleic Acids Res *23*, 3704-3709.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra,

M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., and Chen, Y.J. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lavrentieva, I., Broude, N.E., Lebedev, Y., Gottesman, II, Lukyanov, S.A., Smith, C.L., and Sverdlov, E.D. (1999). High polymorphism level of genomic sequences flanking insertion sites of human endogenous retroviral long terminal repeats. FEBS letters *443*, 341-347.

Moran, J., and Gilbert, N. (2002). Mammalian LINE-1 retrotransposons and related elements (Washington, DC, ASM Press).

Muotri, A.R., Chu, V.T., Marchetto, M.C., Deng, W., Moran, J.V., and Gage, F.H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature *435*, 903-910.

Ovchinnikov, I., Troxel, A.B., and Swergold, G.D. (2001). Genomic characterization of recent human LINE-1 insertions: evidence supporting random insertion. Genome Res *11*, 2050-2058.

Roman-Gomez, J., Jimenez-Velasco, A., Agirre, X., Cervantes, F., Sanchez, J., Garate, L., Barrios, M., Castillejo, J.A., Navarro, G., Colomer, D., Prosper, F., Heiniger, A., and Torres, A. (2005). Promoter hypomethylation of the LINE-1 retrotransposable elements activates sense/antisense transcription and marks the progression of chronic myeloid leukemia. Oncogene *24*, 7213-7223.

Sheen, F.M., Sherry, S.T., Risch, G.M., Robichaux, M., Nasidze, I., Stoneking, M., Batzer, M.A., and Swergold, G.D. (2000). Reading between the LINEs: human genomic variation induced by LINE-1 retrotransposition. Genome Res *10*, 1496-1508.

Shi, X., Seluanov, A., and Gorbunova, V. (2007). Cell divisions are required for L1 retrotransposition. Mol Cell Biol *27*, 1264-1270.

Tchenio, T., Casella, J.F., and Heidmann, T. (2000). Members of the SRY family regulate the human LINE retrotransposons. Nucleic Acids Res *28*, 411-415.

Thayer, R.E., Singer, M.F., and Fanning, T.G. (1993). Undermethylation of specific LINE-1 sequences in human cells producing a LINE-1-encoded protein. Gene *133*, 273-277.

Chapters 2 and 3 will be submitted together for publication: Coufal NG, Garcia-Perez J, Peng GE, Yeo GW, Lovci MT, Morell M, O'Shea KS, Moran JV, Gage FH. "Somatic L1 Retrotransposition in Human Brain and Neural Stem Cells." The dissertation author was the primary author and investigator of this paper.

CHAPTER 4

INCREASES IN L1 RETROTRANSPOSITION IN MODELS OF ATAXIA TELANGIECTASIA

Abstract

Ataxia Telangiectasia (A-T) is a rare autosomal recessive disorder resulting from a loss of function mutation in Ataxia Telangiectasia Mutated (ATM), a protein kinase involved in multiple pathways necessary for sensing DNA damage. Patients with A-T exhibit characteristic progressive neurodegneration as well as immunodeficiency and a predisposition to cancer. We discovered a substantial increase in L1 retrotransposition rates in the brains of ATM knockout mice as compared to wild-type littermates. We sought to establish an *in vitro* model for ATM deficiency in order to further investigate and characterize this increase. Utilizing lentivirally delivered short hairpin RNAs (shRNAs) targeting ATM, we established stable human embryonic stem cell (hESC) lines substantially hypomorphic for ATM. Using hESC derived neural stem cells (NSCs) made from these lines we verified an increase in L1 retrotransposition in ATM hypomorphic NSCs as compared to controls and validated this model for investigating the connection between ATM and L1. We begin to address what mechanism may associate ATM and L1 and propose further studies.

Introduction

Ataxia Telangiectasia, or A-T, is a rare, neurodegenerative, autosomal recessive disorder characterized by a complete loss of function mutation in the Ataxia Telangiectasia Mutated (ATM) protein (Mavrou et al.). The neurodegeneration exhibited in A-T is characterized by a progressive loss of movement coordination due to degeneration of the cerebellar cortex (Biton et al.). Ataxia is canonical for A-T, but progressive neurological symptoms also include dystonia, choreoathetosis, oculomotor

apraxia, dysarthria and difficulty swallowing(Biton et al.). Other non-neurological symptoms in A-T include immunodeficiency, sensitivity to ionizing radiation, and a predisposition to cancer. The role of ATM in the brain and the consequences of its loss on neuronal function and survival are incompletely understood.

One question which has been actively debated in the DNA repair scientific community is whether the neurodegeneration seen in A-T is due to the DNA damage sensing property of ATM or whether a different activity of ATM is implicated (Biton et al., ; Lukas et al.). There are several studies which suggest that perhaps ATM in neurons is cytoplasmically located, whereas in other cell types ATM is nuclear and its activity sensing double stranded breaks (DSB) must be nuclear (Barlow et al.). However, numerous studies have since dismissed this hypothesis and have shown that ATM is expressed in the nucleus in neurons and its role in neurons is a DSB sensing role similar to its role in other tissues (Barzilai et al., ; Uziel et al.).

Several studies have suggested a connection between L1 retrotransposition and deficiencies in DNA repair pathways. Morrish and colleagues investigated the role of DNA damage proteins *XRCC4* and DNA protein kinase catalytic subunit (DNA-PKcs) in mediating L1 retrotransposition (Morrish et al.). They found that chinese hamster ovary (CHO) cells deficient in XRCC4 or DNA-PKcs and therefore with decreased non-homologous end joining repair, exhibited high rates of endonuclease independent L1 retrotransposition. The L1 ORF2 protein encodes both an endonuclease domain, which recognizes a degenerate consensus target sequence (AA/TTTT) and nicks the DNA for insertion, and a reverse transcriptase domain which inserts the L1 RNA into the new genomic location. L1s with a mutation in the endonuclease domain (EN-)

retrotranspose at only 1-5% the efficiency of the wild-type L1. However, in the background of XRCC4 or DNA-PKcs deficiency, the EN- L1 retrotransposed at near wild-type levels (Morrish et al.). Insertions were not into A/T rich regions, and there was evidence that in addition to inserting L1, these EN- events also included pieces of cellular mRNA.

However, other studies have shown that DNA damage need not depend on an EN- pathway. For instance, gamma radiation increases overall L1 retrotransposition, but not in an EN- fashion (Farkash et al.). Similarly, a recent report indicated that decreases in ERCC1 or XPF1, protein involved in sensing DNA "flap" intermediates, also increased L1 retrotransposition (Gasior et al.). The traditional L1 insertion complex with a free 3' hydroxyl and poly T site binding the L1 polyA tail is considered a "flap" intermediate and it may be that DNA damage proteins are activated in these situations and act to decrease or truncate L1 insertion (Farkash and Prak). Other studies have shown that the DSB repair machinery is necessary to repair DNA after transposon integration (Izsvak et al., ; Yant and Kay). Additionally, it has been shown that the L1 insertions can create double strand breaks and co-localize with γ-H2AX foci (Gasior et al.).

Given this connection between DNA damage pathways and L1 retrotransposition, we investigated whether deficiency in ATM would change L1 retrotransposition rates or increase the amount of EN- retrotransposition. ATM is the activator of the cellular response to DSBs and orchestrates the repair mechanisms, cell cycle checkpoints, and apoptosis pathways that result from DSBs (Shiloh) (Biton et

al.). Therefore, we hypothesized that a deficiency in ATM would result in increased opportunity for retrotransposition and increased DSBs.

We found evidence for increased L1 retrotransposition in the brains of ATM deficient mice *in vivo* and in ATM deficient human embryonic stem cell (hESC) derived neural stem cells (NSCs) *in vitro*. We also found no evidence for EN-retrotransposition events in either ATM deficient hESC or derived NSCs. However, we found that L1 retrotransposition events occurred in progenitors which could differentiate to neuronal and glial fates normally in ATM deficient cells. Further investigation into the mechanism which allows for increased L1 retrotransposition will examine this process.

Methods

Animals and tissue preparation

ATM mice (Barlow et al.) were kindly provided by Dr. Carolee Barlow. The generation of the L1-EGFP animals has been previously described (Muotri et al.). Six gender-matched mice, from the same C57BL/6J background, were used per group. Tissues were prepared from adult animals (8 weeks old) as previously described (Muotri et al.). Primers used for genotyping were as follows: To analyze L1 transgene primers GFP968s (GCACCATCTTCTTCAAGGAC) and GFP1013 (TCTTTGCTCAGGGCGGACTC) as were used. A 1243 bp band indicates the L1 transgene with intron, a 343 bp product from tail genomic DNA indicates a germline or early embryonic insertional event. To determine ATM allele status, primers ATMf (GACTTCTGTCAGATGTTGCTGCC), ATMr

(CGAATTTGCAGGAGTTGCTGAG), and ATMneo (GGGTGGGATTAGATAAATGCCTG) were utilized. The wild-type allele amplifies a 162 bp product, whereas the neomycin insertion transgene is a 441 bp product. All experimental procedures and protocols were approved by the Animal Care and Use Committees of The Salk Institute, La Jolla, CA.

Cell Culture

hESC lines HUES6 was cultured as previously described (http://www.mcb.harvard.edu/melton/HUES/)(Thomson et al.). Briefly, cells were grown on mitomycin C-treated mouse embryonic fibroblast (MEF) feeder layers (Chemicon) in DMEM media (Invitrogen) supplemented with 20% KO serum replacement, 1 mM L-glutamine, 50 uM β-mercaptoethanol, 0.1 mM nonessential amino acids, and 10 ng/mL β-FGF2 (fibroblast growth factor 2), and passaged by manual dissection. For lentiviral infection and blasticidin selection cells were grown on GFR matrigel coated plates in MEF-conditioned medium and 20 ng/mL β-FGF2. For EB formation, cells were either dissociated from the underlying MEF layer with Dispase (0.2 mg/mL; Stem Cell Technologies) or scraped directly from matrigel plates, then grown for 7 days in DMEM-F12 Glutamax media (Invitrogen) with N2 supplement (Gibco) and B-27 supplement (Invitrogen). Subsequently, EBs were plated onto laminin/poly ornithine (Sigma) coated plates and grown for 7-10 days more in N2 containing media alone. Rosettes were then manually dissected and dissociated in 0.1% trypsin and plated in DMEM-F12 media supplemented with N2 and B-27, 1 ug/mL laminin, and 20 ng/mL FGF2. Resulting neural progenitors could be

maintained for multiple passages before induction of differentiation. Differentiation conditions involved withdrawal of mitogens and treatment with 20 ng/ml BDNF, 20 ng/ml GDNF (Peprotech), 1 mm dibutyrl-cyclicAMP (Sigma), and 200 nm ascorbic acid (Sigma) for 4-12 weeks.

Lentiviral knockdown of ATM

Lentiviral siRNA constructs towards ATM were commercially available and purchased from Invitrogen. Hairpin sequences were as follows ATM1: GCAACATTTGCCTATATCAGCAATTTTCGAATTGCTGATATAGGCAAATGT TGCGGTG; ATM2: GCGCAGTGTAGCTACTTCTTCTATTTTCGAATAGAAGAAGTAGCTACACGC GCGGTG. ATM3: GCACTGACCTCTGTGACTTTTCGAAGTCACAGAGGTCAGTGCGGTG.

siRNAs were transfected into hESCs using the Amaxa electroporation system, nucleofector solution V and program A-24. Cells were harvested at three days for western blotting. Once specific siRNAs were confirmed to decrease ATM expression, shRNAS were constructed using 70 bp oligonucleotides matching the siRNAs along with complimentary oligos (all purchased from IDT) and annealed by boiling and slow cooling. Annealed products were cloned into the lentiviral pDEST backbone (Invitrogen) and confirmed by sequencing.

Lentivirus was produced as previously described (Singer et al., 2005). Briefly, recombinant lentiviruses were produced by transient transfection of HEK293T with three packaging vectors in addition to the pDEST shRNA vector. Control virus used is

pDEST containing a shRNA towards GFP. Infectious media was harvested, filtered and concentrated by ultracentrifugation. Efficacy of lentivirus was tested by infection of HUES6 hESC at varying viral concentrations and subsequent Western blotting with an antibody against ATM. Western blotting was performed using standard protocols; whole cell lysates were prepared in RIPA buffer with protease inhibitors (Roche). Antibodies utilized were: ATM (1:2500, rabbit polyclonal, Epitomics), SOX2 (1:2000, rabbit polyclonal, Sigma).

hESCs were grown for at least one passage on matrigel before lentiviral infection. Post infection, cells were selected with 2 μg/mL blasticidin beginning 4 days after infection and lasting for a minimum of 10 days. hESCs were re-selected with blasticidin after every three passages. hESCs were plated back onto MEF feeder layers to recover appropriate morphology before initiating EB formation.

Constructs, Transfection, and Retrotransposition Assay

Cells were transfected with L1 elements containing the EGFP retrotransposition cassette in the pCEP4 (Invitrogen) plasmid backbone, with the hygromycin selection gene replaced with a puromycin selection gene. Prior to transfection, DNAs were checked for superhelicity by electrophoresis on 0.7% agarose-ethidium bromide gels. Only highly supercoiled preparations of DNA (>90%) were used.

The LRE3 element is an active full-length element under the control of the native 5'UTR(Brouha et al.), and has been previously described (Ostertag et al.). JM111 is a derivative of $L1_{RP}$ containing the double missense mutation RR261-262AA in the ORF1 protein, rendering it retrotransposition incompetent (Garcia-Perez

et al.). The H230A mutation in the endonuclease domain has been previously described (Morrish et al.). All constructs contain the CMV-EGFP expression cassette (Ostertag et al.).

HUES6-derived NSCs one passage after neural rosette selection were transfected by Nucleofection using the Amaxa rat NSC nucleofector solution and program A-31. Cells were cultured as progenitors in the presence of mitogens. For differentiation studies, cells were dissociated and plated for differentiation 18 days after initial transfection. Cells were monitored for GFP expression by fluorescence microscopy. For FACS analysis, cells were dissociated and analyzed on a Becton-Dickinson LSR I in the presence of 1 ug/mL propidium iodide for live/dead cell gating. All assays were performed in triplicate, and JM111 transfected cells were used as a negative control for gating purposes.

Immunohistochemistry and Imaging

Cells were fixed in 4% paraformaldehyde, and immunocytochemistry was performed as previously described (Gage et al., ; Garcia-Perez et al.). Antibodies and dilutions were as follows: β-III tubulin mouse monoclonal 1:400 or rabbit polyclonal 1:500 (both Babco/Covance), Map (2a+2b) mouse monoclonal 1:500 (Sigma), GFAP rabbit polycolonal 1:300 (DAKO), GFAP guinea pig polyclonal 1:1000 (Advanced Immunochemical), Nestin mouse monoclonal 1:800 (Chemicon), TH rabbit polyclonal 1:500 (Pel-Freez), Sox2 rabbit polyclonal 1:500 (Sigma). Secondary antibodies were purchased from Jackson ImmunoResearch or Invitrogen and were all used at 1:250. Cells were imaged using a CARVII spinning disk confocal imaging system (BD).

Luciferase Assay

Luciferase activity was measured with the Dual-Luciferase reporter assay system (Promega) in accordance with the manufacturer's instructions. A plasmid containing the Renilla *luciferase* gene was used as an internal control in all assays. All assays were replicated at least three times independently. The L1 5'UTR luciferase construct has been previously described (Muotri et al.). The Synapsin-1 promoter region was a kind gift from G. Thiel. All promoters were in the pGL3-basic vector (Promega).

Cell cycle

Cell cycle staining was performed as previously described(Crissman and Steinkamp, 1973)'(Krishan, 1975). Briefly, cells were trypsinized, washed, and resuspended in PBS buffer, then fixed by addition of 3:1 ratio of ice cold 100% ethanol in PBS, overnight at -20°C. Subsequently cells were washed and resuspended in a solution containing 50 ug/mL propidium iodide and 500 ng/mL RNase A for 1 hr at 37°C before analysis by fluorescent-activated cell sorting (FACS) on a Becton-Dickinson FACScan.

Results

Human L1 retrotransposition in ATM/L1 transgenic mice

Mice carrying a complete loss of function mutation in ataxia telangiectasia mutated (ATM) protein have previously been described (Barlow et al.). We crossed ATM knockout mice with L1 transgenic mice (Muotri et al.). We subsequently compared those littermates which were homozygous either for loss of ATM or wild-

type littermates, both of whom also carried the human L1 transgene replete with the CMV-GFP indicator cassette (containing the γ-globin intron). We confirmed that animals did not contain a germline L1 insertion using tail clips (Fig 4.1B). Analysis of the brains of ATM knockouts and wild-type littermates indicated a substantial increase in L1 retrotransposition in the hippocampus (Fig 4.1 A)

L1 retrotransposition events in hESC

In order to confirm this phenotype in an *in vitro* system, we constructed multiple lentiviral shRNA vectors targeting ATM, as well as a control targeted toward GFP. HUES6 hESCs were infected and subsequently selected with blasticidin 2 ug/mL for 7 days. These lentiviruses successfully decreased ATM long term in hESC (Fig. 4.2A). ATM depleted hESC continued to proliferate and to express markers of pluripotency (Fig. 4.2B).

We proceeded to investigate the retrotransposition capacity of ATM deficient hESC lines, since hESCs have previously been shown to support a low level of L1 retrotransposition (Garcia-Perez et al.). L1 retrotransposition assays in ATM knockdown and control hESC indicated a very low level of retrotransposition. As a negative control, L1 JM111 was utilized, which is derived from LRE3 but contains the RR261-262AA mutation in the ORF1 sequence rendering the L1 retrotransposition incompetent (Moran et al.). The CMV promoter, which drives GFP expression in this model, has previously been shown to silence rapidly in hESC (Xia et al.). Therefore, we also treated cultures with 10 μm trichostatin A (TsA), an inhibitor of histone deacetylase (HDAC), for 24 hours before fluorescence activated cell sorting (FACs) to

reactivate silenced insertions (Fig. 4.2C). Treatment with TsA indicated an increase in L1 retrotransposition rates in all three ATM deficient cells as compared to control.

Previous studies have shown that L1 retrotransposition activity is increased in cell lines which contain deficiencies in non-homologous-end joining pathways (Morrish et al.). Deficiency of XRCC4 or DNA protein kinase catalytic subunit (DNA-PKcs) in chinese hamster ovary cells not only increased L1 retrotransposition, but also increased the rate of endonuclease independent (EN-) events in these cells. In order to investigate whether increases in L1 retrotransposition in ATM deficient hESCs occurred as a result of increases in EN- event, we utilized an L1 with the H230A mutation (histidine to alanine) in the endonuclease domain. We found no increase in EN- events with knockdown of ATM (Fig. 4.2C).

L1 retrotransposition in ATM deficient NSCs

We hypothesized that perhaps ATM deficient NSCs would also exhibit an increase in L1 retrotransposition, similar to the brains of ATM deficient L1 transgenic mice. In order to investigate this question, we derived NSCs from ATM deficient hESCs using the same methodology as is presented in Chapter 2. Briefly, hESC were allowed to form EBs, then plated down to form neural rosettes in the presence of Noggin. Rosettes were manually dissected, dissociated, and propagated in the presence of FGF2. Both ATM deficient and control hESC formed NSCs cultures which expressed canonical NSCs markers SOX2 and Nestin. In addition, control NSCs expressed ATM in the nucleus, whereas ATM deficient cells expressed ATM at only very low levels (Fig. 4.3A and 4.4A). Differentiation of these NSCs using standard techniques (Chapter 2) indicated that both could differentiate to a neuronal phenotype

successfully, as indicated by the neuronal marker map2(a+b) (Fig. 4.3B). ATM has previously been shown to exhibit nuclear expression in differentiated neurons(Shiloh), which is also true of the hESC derived NSCs (Fig 4.3B).

We investigated the capacity of ATM deficient and control NSCs cultures to allow L1 retrotransposition. We found a statistically significant increase in L1 retrotransposition in ATM deficient cultures as compared to control (p=0.004)(Fig. 4.4B). To investigate whether changes in transgene silencing might account for this difference, rather than actual changes in retrotransposition rates, we treated cultures with 10μm TsA for 24 hours before analysis. This treatment did not decrease the difference in GFP positive cells between ATM and control, instead increased it such that ATM deficient NSCs had 2.5% GFP positive cells, whereas matched controls only a 0.8% GFP positive cells (p=0.00000003). When this experiment was replicated with independently derived NSCs cultures (also matched ATM hypomorph and control), the rate of L1 retrotransposition is somewhat variable, but the fold difference between ATM and control NSCs remains approximately 3-fold higher in ATM deficient cells (Fig. 4.4C). However, we did not see an increase in retrotransposition using the H230A construct in multiple independent experiments. NSCs which experienced an L1 retrotransposition event differentiate normally to both a neuronal and glial phenotype (Fig. 4.5A-B).

In order to confirm that increased rates of L1 retrotransposition are not due to changes in L1 promoter activity, we performed luciferase assays using the L1 5' untranslated region (UTR) driving luciferase. We found a 25-fold increase in L1 promoter activity during the course of differentiation with both ATM deficient cells

and control, with no difference between them (Supp 4.1A). As a control, synapsin luciferase activity was used to assess differentiation, which also did not differ between cultures and indicated robust differentiation for both conditions.

A comparison of cell cycle between ATM deficient NSCs and controls indicated a slight S to G2 shift in ATM deficient NSCs. This data is comparable to previous work which has shown increased cell cycle for adult NSCs derived from ATM deficient mice (Allen et al.). L1 retrotransposition assay was performed at 8 days post transfection in order to attempt to limit the effect of cell cycle such that cells which have experienced an L1 retrotransposition event have limited time to proliferate.

Discussion

ATM is a serine-threonine kinase in the PI3/PI4 family of kinases and acts to coordinate the cellular response to double strand breaks (DSB). ATM kinase activity acts to phosphorylate multiple proteins involved in sensing DNA damage to activate cell cycle checkpoints, apoptotic pathways, and DNA repair (Biton et al.) (Lavin). Ionizing radiation leads to DSBs, which are sensed through ATM-dependent pathways. Without ATM, cells are inefficient at detecting DSB, although they appear to have no deficiency in repairing DSB when they are detected (Riballo et al.). The response to DSB is dependent on the Mre11 complex, which senses DSB and recruits ATM, which in turn phosphorylates not only Mre11, but also a variety of proteins targets such as p53 to lead to cell cycle arrest (Lavin). Studies have shown that ATM acts through activation of the MRN/mre11 complex to both auto-phosphorylate and to

phosphorylate numerous proteins necessary to respond to DNA damage, such as p53, Chk2, and NBS1.

The first evidence for increased retrotransposition in the context of ATM deficiency came from crosses of the human L1 transgenic mouse with ATM knockout mice. These mice have shown an appreciable increase in L1 retrotransposition in the brain as compared to control littermates. Further study will be necessary to more quantitatively address which brain areas have an increase, but preliminary studies indicate a substantial increase in the hippocampus. Future questions to address include quantitative studies both of ATM knockouts and controls to quantify L1 retrotransposition in various brain areas, as well as studies addressing whether this increase occurs early in embryogenesis or during adult life. This study would require comparing ATM knockout and wildtype littermates at different embryonic times and quantifying L1 retrotransposition.

A second question which arises from the preliminary ATM data is whether there is increased L1 copy number in the brains of AT patients. Future studies will investigate L1 ORF2 copy number in a study analogous to that in Chapter 3 here, to inquire whether the finding of increased L1 retrotransposition in ATM knockout mouse brain can be translated to represent human endogenous L1 retrotransposition rates. Provided that increased L1 insertional rates are identified, no doubt a high throughput sequencing approach will ultimately be applied to not only identify increased insertional rates, but also to elucidate the structure and location of those insertions.

Lastly, the question of how ATM deficiency and L1 retrotransposition interact remains to be shown. The most straightforward explanation is that ATM senses DNA damage and in the absence of ATM this damage is undetected and therefore unrepaired (Biton et al.). In the context of decreased DNA break repair, L1 would act to patch these DNA breaks by engaging in endonuclease independent insertions. Previous studies have shown that mutations in the non-homologous end joining. repair pathway (NHEJ), such as mutations in the XRCC4 or DNA protein kinase catalytic subunit genes in rodent cells, lead to large increases in endonuclease independent insertions (Morrish et al.). It remains to be shown if the same mutations in NHEJ in human cells would also lead to increases in EN- L1 retrotransposition. The studies undertaken here suggest that EN- retrotransposition is not responsible for increases in L1 retrotransposition in the ATM deficient cells. However, studies in ES derived NSCs need to be completed with more ATM shRNAs. In order to ensure this pathway is not utilized, numerous insertions from ATM deficient NSCs need to be identified and the insertional location characterized. However, if studies continue to find that this pathway is not responsible for increased retrotransposition, there are a number of other pathways either upstream of ATM, such as Mre11 which binds to the DNA breakage directly, or downstream, such as ATR, DNA-protein kinase catalytic subunit which might be involved.

Further studies will focus on elucidating the pathways involved in increasing L1 retrotransposition in an ATM negative background. A second possible hypothesis is that ATM acts in a pathway which senses the L1 insertional intermediate, with its single strand DNA breaks and a free 3' hydroxyl, as DNA damage and mediates a

response. This response may lead to truncation of the L1 insertion. Given the high processivity of the L1 reverse transcriptase (Bibillo and Eickbush), it has been an open question as to why the majority of L1 insertions are truncated. However, given the high retrotransposition ability of de novo full length insertions such as LRE3 or $L1_{RP}$, the advantage of truncating L1 insertions is clear. If ATM is involved in sensing L1 insertion as a DNA break, one would predict that L1 insertions would not occur in an endonuclease independent fashion, and that L1 insertions in an ATM deficient model would be longer.

To this end we propose to investigate the role of ATM and related pathways in another human cell model. Previous studies have shown that complete knockdown of numerous members of the DNA damage response pathways, including DNA-PKcs, ATR, and MRN, is fatal to human cells but not to comparable rodent cultures, suggesting that pathways interact differently in these models (Morrish et al.). Similarly, rodent L1s and human L1s are different in their number and regulation. Whereas humans have ~150 retrotransposition competent L1s, mice have more than 3000, and mouse element contain numerous repeats regions for that are lacking in human (Ostertag and Kazazian). Therefore, we propose to model ATM deficiency in a transformed human cell line model such as HCT116 colorectal cancer cells and compare retrotransposition rates for native and EN- L1s in ATM deficient and control cells. We will also use shRNAs towards XRCC4 to confirm that EN- events are increased in a human model similar to the Morrish model. HCT116 cells are an excellent model because they are karyotypically stable but allow for homologous recombination. Therefore, HCT116 lines are available which are homozygous

knockouts for p53, or hypomorphs for DNA-pk or Ligase IV (Bunz et al.). Therefore, we will utilize these lines, in combination with our ATM shRNAs and available shRNAs towards ATR and Mre11 to investigate whether L1 or L1 EN-retrotransposition events are altered as compared to the parental line. Provided these results are complementary to the ATM deficient NSCs, we can pursue shRNAs towards other members of the DNA damage pathway, such as MRN complex member/Mre11, ATR, DNA-PKcs. Those pathways which are found to change L1 retrotransposition activity will then be pursed in the hESC derived NSCs model.

Figure 4.1

A

WT

B

ATM-

C

ATM-
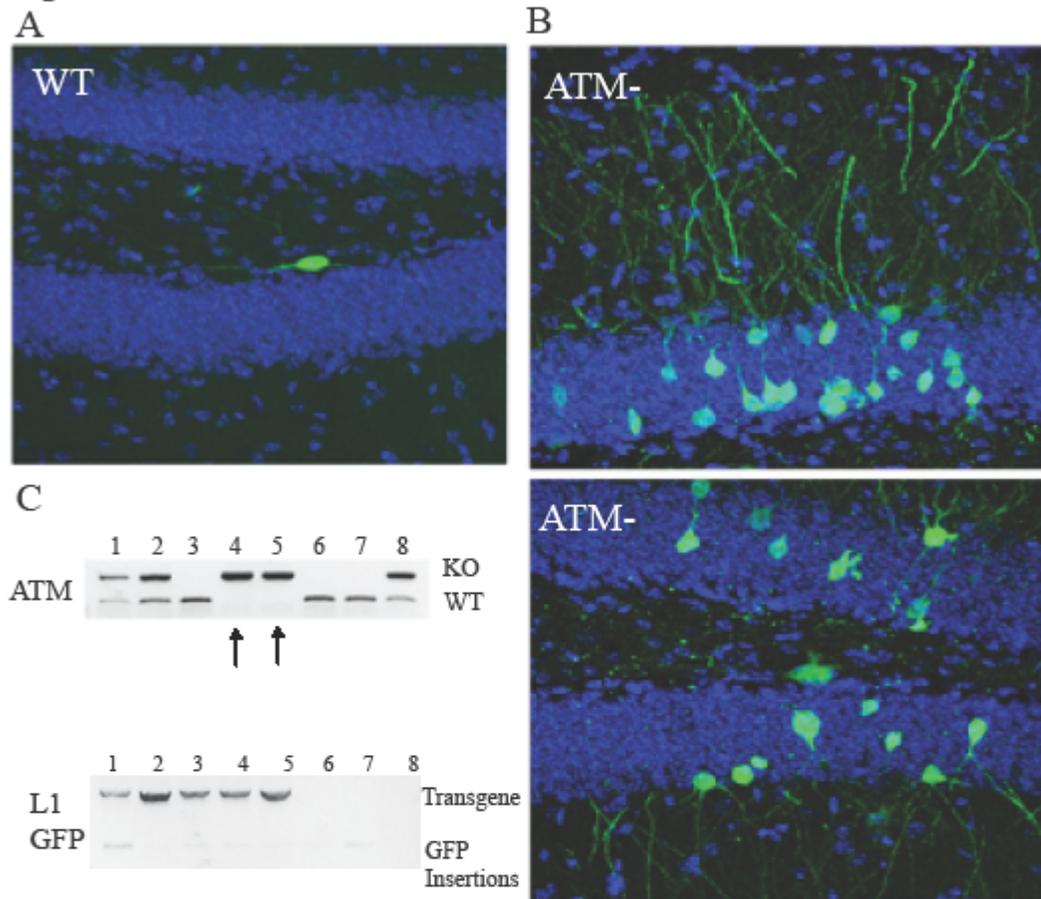
Figure 4.1
A. L1-GFP wildtype littermates exhibit a low level of retrotransposition in the hippocampus
B. ATM knockout homozygotes which also carry the L1-GFP trangene exhibit an increase in L1 retrotranspostion, which is especially noteable in the hippocampus.
C. Sample genotyping of transgenic mice. Top, ATM genotyping, the smaller bands indicates ATM knockout allele and the larger band the wild-type allele. Those animals exhibiting both alleles (#1, 2, 8) are heterozygotes. Bottom, L1 genotyping. Top band is the L1 transgene, if a lower band is present in this indicates a germline L1 insertional events. Animals exhibiting only the L1 transgene who are also homozygous knockouts for ATM were used in this study (#4 and #5) and compared to wild-type litter-mates (#3).

Figure 4.2
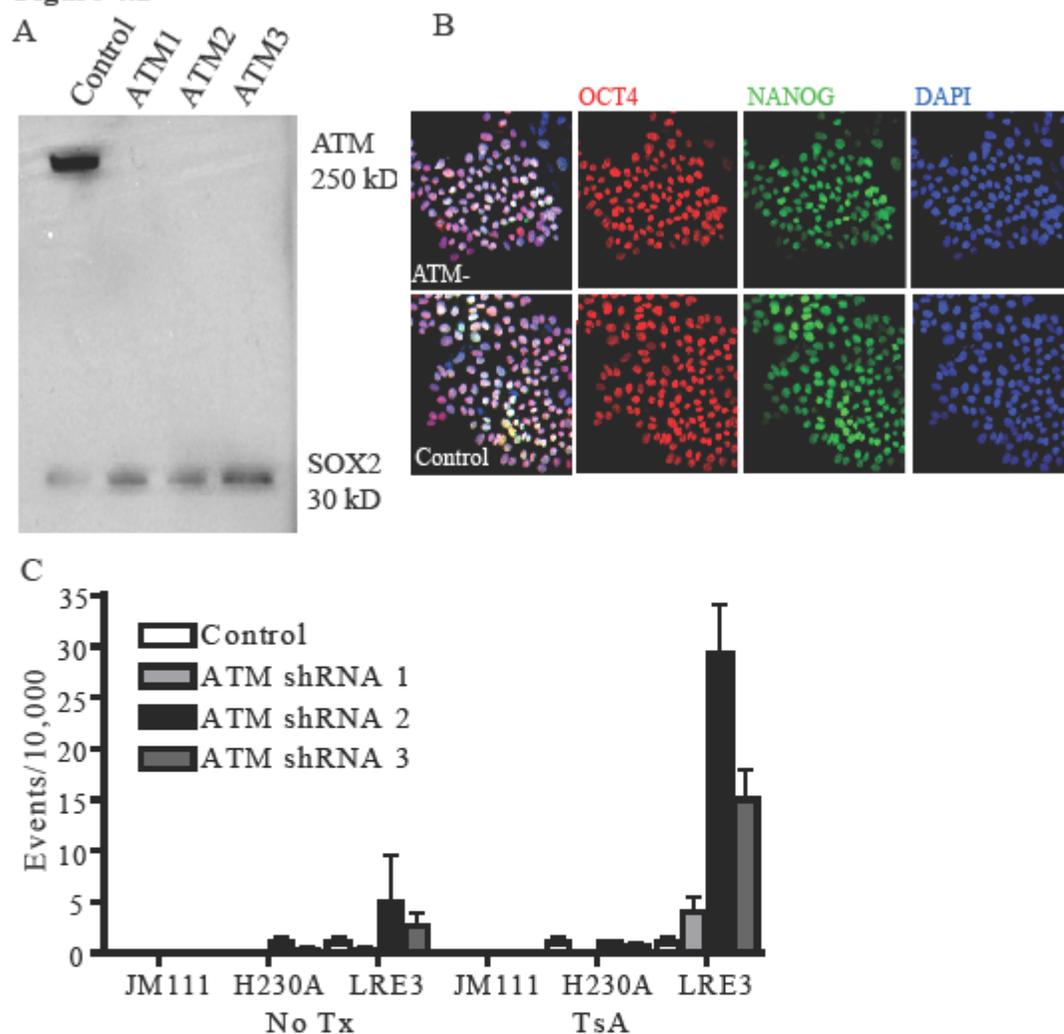
Figure 4.2

Characterization of ATM shRNA infected hESCs

A. Western blotting exhibits decreased ATM protein expression in hESCs with lentiviral infection carrying each of three different shRNAs toward ATM as compared to a control shRNA. The loading control is SOX2.

B. Immunocytochemistry of control and ATM hypomorphs indicates that cells retain normal morphology and express the hESC pluripotency markers Oct4, and Nanog.

C. L1 retrotransposition assay with control and ATM shRNA infected hESC. FACs analysis indicates low levels of L1 retrotransposition in hESC, with an increased rate exhibited by two of three ATM hESC lines. The addition of TsA, an HDAC inhibitor which reverses some silencing of the CMV promoter, reveals an increase in L1 retrotransposition in all ATM negative lines as compared to control.

Figure 4.3

A
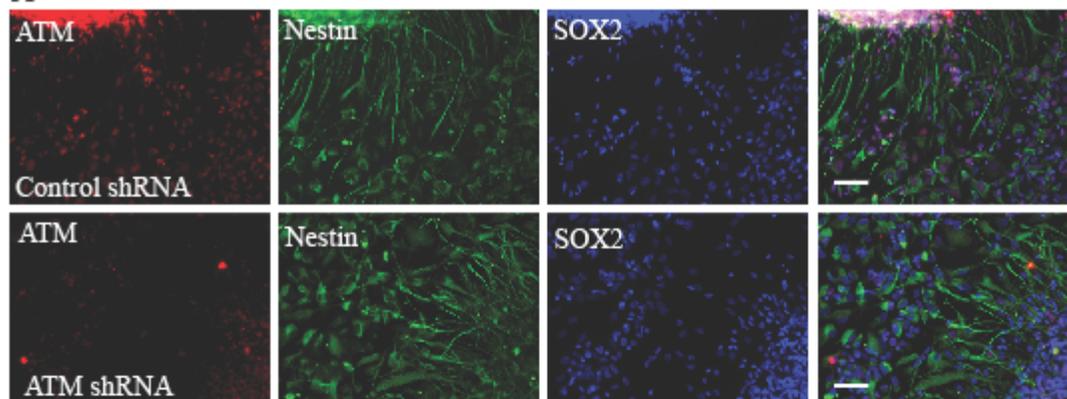
B
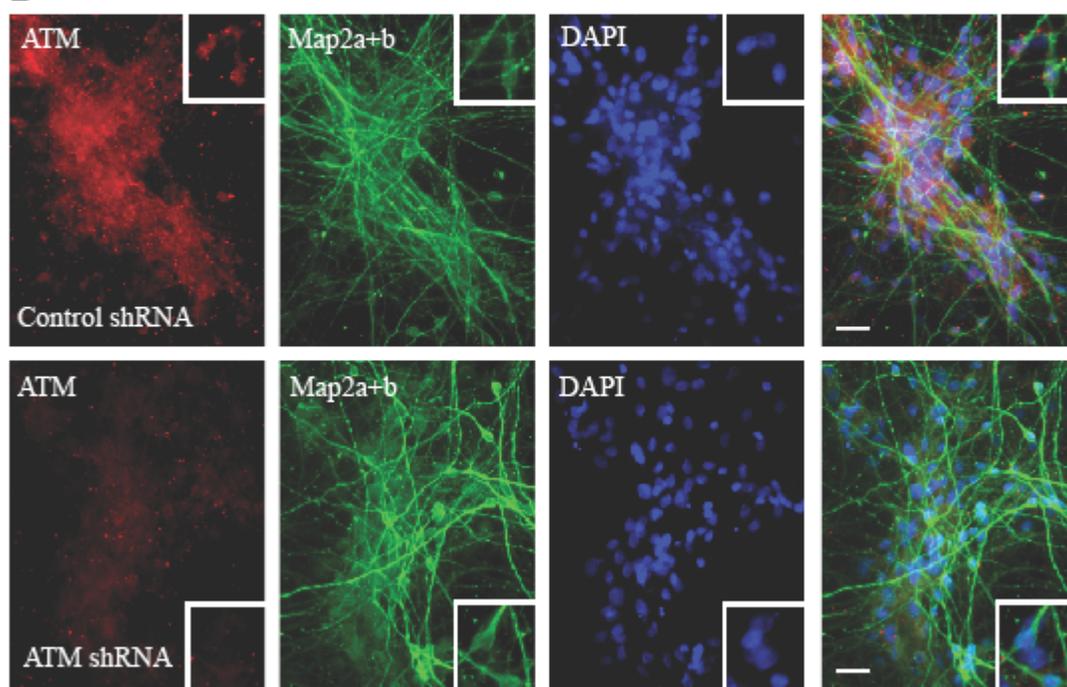
Figure 4.3
A. Immunohistochemistry of ATM and control NSCs. hESCs infected with the ATM or control shRNA lentiviruses were made into NSCs. Both ATM and control NSCs express normal progenitor markers such as Nestin and SOX2. ATM is expressed in the nucleus and is visibly decreased in ATM hypomorphs.
B. Immunohistochemistry of ATM and control neurons indicating that cells differentate at approximately equal rates to Map2a+b positive neurons. ATM is expressed in the nucleus and is decreased in ATM hypomorphs. Scale bar = 25microns

Figure 4.4

L1 retrotransposition n hESC derived NSCs

A. L1 retrotransposition assay in hES derived NSCs indicates increased rates of L1 retrotransposition in ATM hypomorphs as compared to controls (p=0.004). Addition of TsA, an HDAC inhibitor which reverses some silencing of the CMV promoter, indicreased this difference (p=0.00000003) is robust and not a result of decreased silencing of the promoter in ATM hypomorphs.

B. Western blotting indicates decreased ATM expression in ATM shRNA infected NSCs as compared to the loading control, SOX2. No retrotransposition was seen using the H230A endonuclease negative L1 construct either with or without addition of TsA.

C. Sample FACs analysis of LRE3 transfected ATM and control NSCs as compared to the negative retrotransposition control, JM111, indicating a robust increase in L1 retrotransposition in ATM hypomorphs.

Figure 4.4

A

B

C

Figure 4.5

A

B
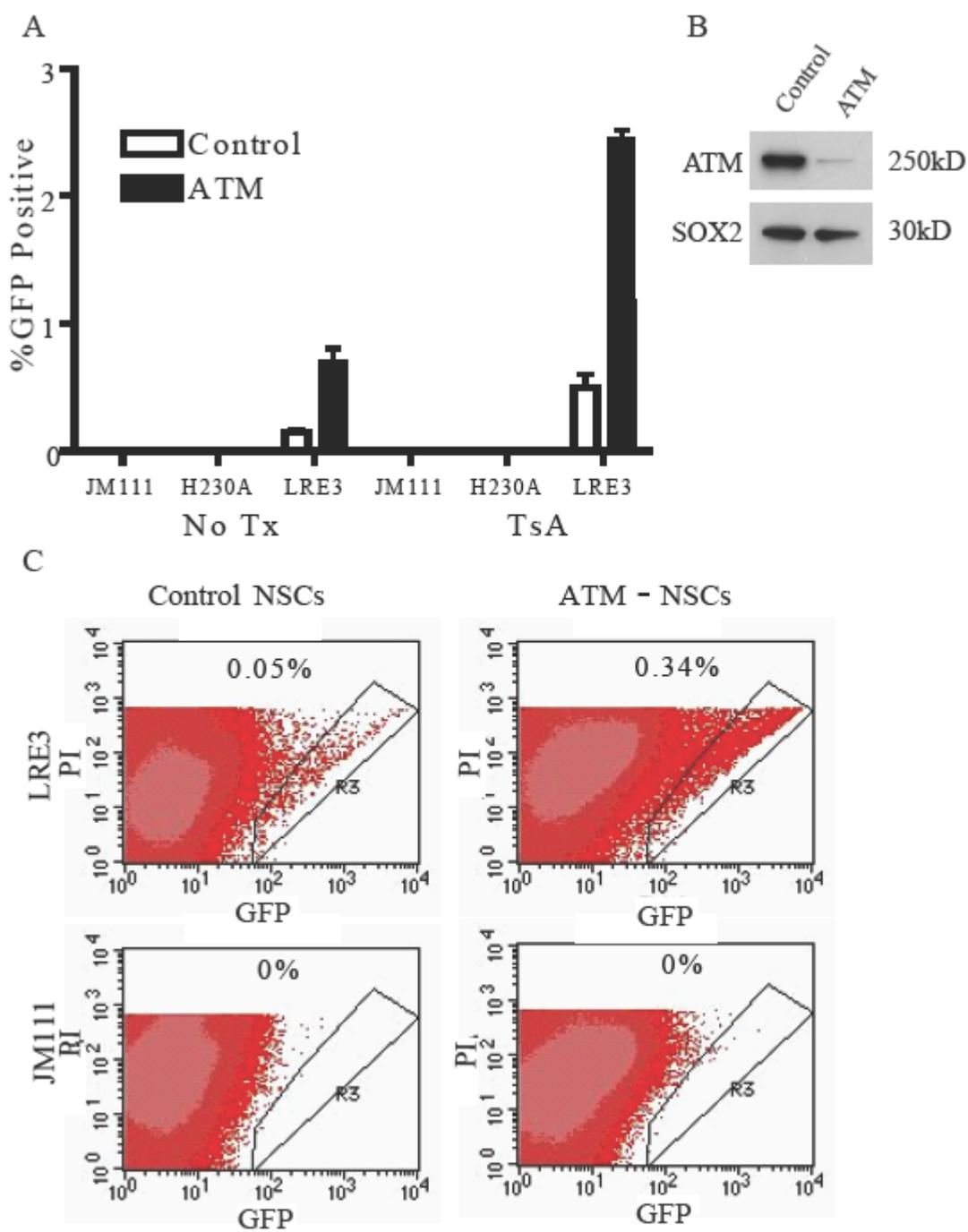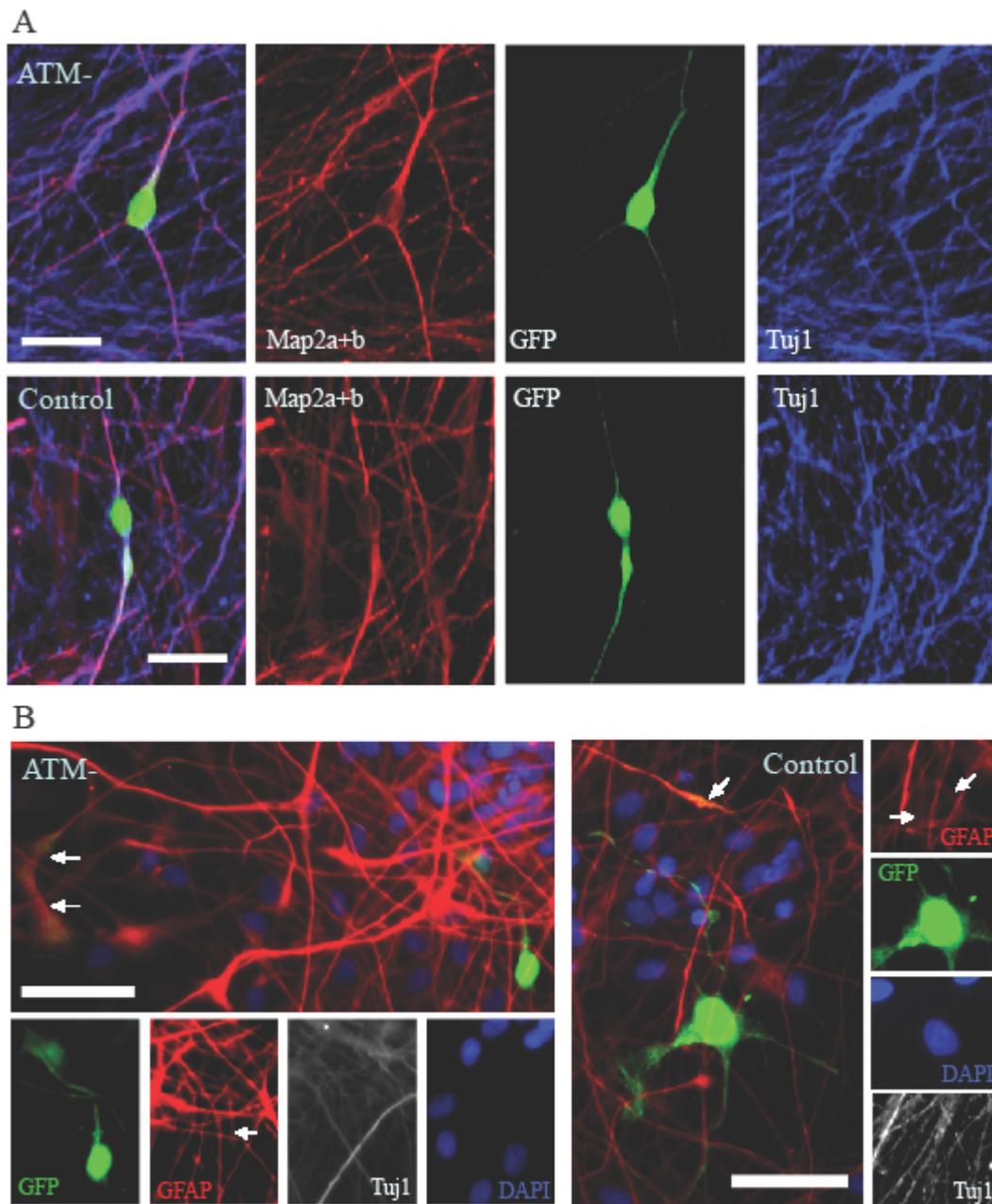
Figure 4.5
Characterization of L1 retrotranposition events in both ATM hypomorph and matched control cultures.
A. NSCs which are permissive for L1 retrotransposition events can differentiate into neurons which co-express neuronal markers Map2a+b and Tuj1.
B. NSCs which experience L1 retrotransposition events can differentiate into GFAP positive astrocytes in both ATM hypomorph and control cultures.

Supplement 4.1

A
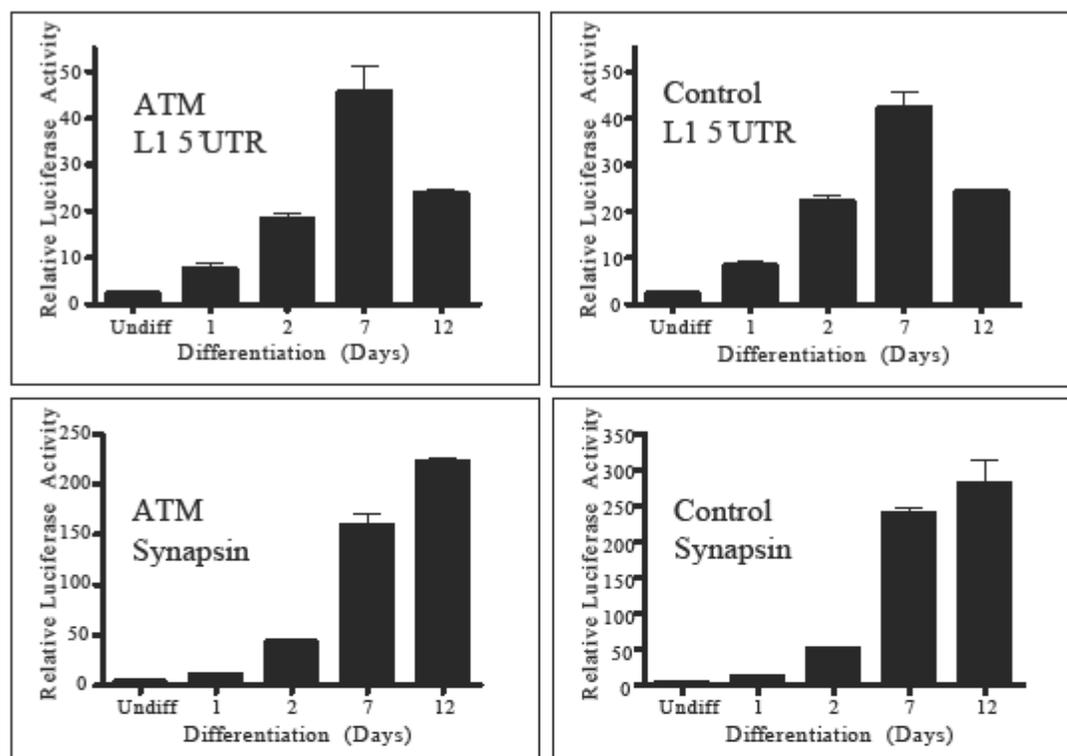
B
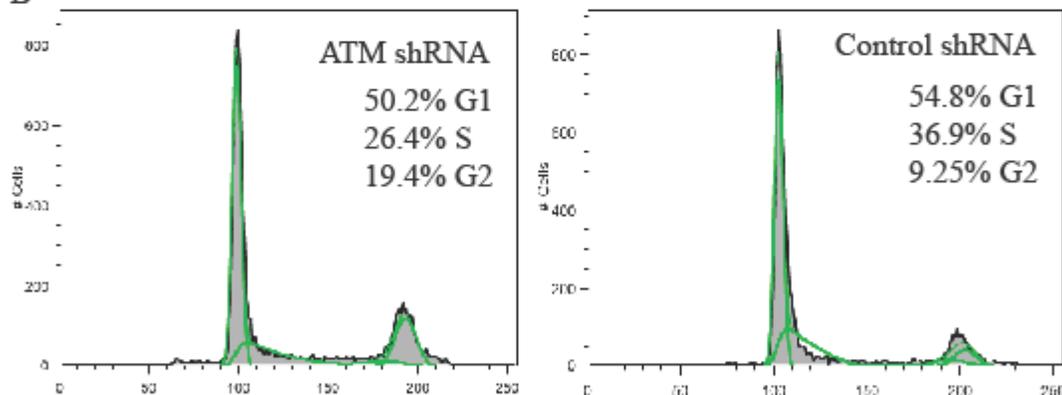
Supplement 4.1

A. Luciferase assays using the L1 5'UTR to drive luciferase indicate a 25 fold increase in L1 activity with the onset of differentiation. This rate is not changed in ATM hypmorphs as compard to controls. The differentiation control is the synapsin promoter driving luciferase, which indicates that both ATM hypomorphs and controls differentiated to a neuronal phenotype at statistically similar rates.

B. Cell cycle assays of ATM hypomorphs and controls indicates a mild S to G2 shift in ATM hypomorphs as compared to controls.

WORKS CITED

Allen, D.M., van Praag, H., Ray, J., Weaver, Z., Winrow, C.J., Carter, T.A., Braquet, R., Harrington, E., Ried, T., Brown, K.D., Gage, F.H., and Barlow, C. (2001). Ataxia telangiectasia mutated is essential during adult neurogenesis. Genes & development *15*, 554-566.

Barlow, C., Hirotsune, S., Paylor, R., Liyanage, M., Eckhaus, M., Collins, F., Shiloh, Y., Crawley, J.N., Ried, T., Tagle, D., and Wynshaw-Boris, A. (1996). Atm-deficient mice: a paradigm of ataxia telangiectasia. Cell *86*, 159-171.

Barlow, C., Ribaut-Barassin, C., Zwingman, T.A., Pope, A.J., Brown, K.D., Owens, J.W., Larson, D., Harrington, E.A., Haeberle, A.M., Mariani, J., Eckhaus, M., Herrup, K.,

Bailly, Y., and Wynshaw-Boris, A. (2000). ATM is a cytoplasmic protein in mouse brain required to prevent lysosomal accumulation. Proc Natl Acad Sci U S A *97*, 871-876.

Barzilai, A., Biton, S., and Shiloh, Y. (2008). The role of the DNA damage response in neuronal development, organization and maintenance. DNA repair *7*, 1010-1027.

Bibillo, A., and Eickbush, T.H. (2002). High processivity of the reverse transcriptase from a non-long terminal repeat retrotransposon. J Biol Chem *277*, 34836-34845.

Biton, S., Barzilai, A., and Shiloh, Y. (2008). The neurological phenotype of ataxia-telangiectasia: Solving a persistent puzzle. DNA repair *7*, 1028-1038.

Biton, S., Gropp, M., Itsykson, P., Pereg, Y., Mittelman, L., Johe, K., Reubinoff, B., and
Shiloh, Y. (2007). ATM-mediated response to DNA double strand breaks in human neurons derived from stem cells. DNA repair *6*, 128-134.

Brouha, B., Meischl, C., Ostertag, E., de Boer, M., Zhang, Y., Neijens, H., Roos, D., and Kazazian, H.H., Jr. (2002). Evidence consistent with human L1 retrotransposition in maternal meiosis I. Am J Hum Genet *71*, 327-336.

Bunz, F., Dutriaux, A., Lengauer, C., Waldman, T., Zhou, S., Brown, J.P., Sedivy, J.M.,
Kinzler, K.W., and Vogelstein, B. (1998). Requirement for p53 and p21 to sustain G2 arrest after DNA damage. Science *282*, 1497-1501.

Crissman, H.A., and Steinkamp, J.A. (1973). Rapid, simultaneous measurement of DNA, protein, and cell volume in single cells from large mammalian cell populations. J Cell Biol *59*, 766-77.

Farkash, E.A., Kao, G.D., Horman, S.R., and Prak, E.T. (2006). Gamma radiation increases endonuclease-dependent L1 retrotransposition in a cultured cell assay. Nucleic Acids Res *34*, 1196-1204.

Farkash, E.A., and Prak, E.T. (2006). DNA damage and l1 retrotransposition. J Biomed Biotechnol *2006*, 37285.

Gage, F.H., Coates, P.W., Palmer, T.D., Kuhn, H.G., Fisher, L.J., Suhonen, J.O., Peterson, D.A., Suhr, S.T., and Ray, J. (1995). Survival and differentiation of adult neuronal progenitor cells transplanted to the adult brain. Proc Natl Acad Sci U S A *92*, 11879-11883.

Garcia-Perez, J.L., Marchetto, M.C., Muotri, A.R., Coufal, N.G., Gage, F.H., O'Shea, K.S., and Moran, J.V. (2007). LINE-1 retrotransposition in human embryonic stem cells. Hum Mol Genet *16*, 1569-1577.

Gasior, S.L., Roy-Engel, A.M., and Deininger, P.L. (2008). ERCC1/XPF limits L1 retrotransposition. DNA repair *7*, 983-989.

Gasior, S.L., Wakeman, T.P., Xu, B., and Deininger, P.L. (2006). The human LINE-1 retrotransposon creates DNA double-strand breaks. J Mol Biol *357*, 1383-1393.

Izsvak, Z., Stuwe, E.E., Fiedler, D., Katzer, A., Jeggo, P.A., and Ivics, Z. (2004). Healing the wounds inflicted by sleeping beauty transposition by double-strand break repair in mammalian somatic cells. Molecular cell *13*, 279-290.

Krishan, A. (1975). Rapid flow cytofluorometric analysis of mammalian cell cycle by propidium iodide staining. J Cell Biol *66*, 188-193.

Lavin, M.F. (2007). ATM and the Mre11 complex combine to recognize and signal DNA double-strand breaks. Oncogene *26*, 7749-7758.

Lukas, J., Lukas, C., and Bartek, J. (2004). Mammalian cell cycle checkpoints: signalling pathways and their organization in space and time. DNA repair *3*, 997-1007.

Mavrou, A., Tsangaris, G.T., Roma, E., and Kolialexi, A. (2008). The ATM gene and ataxia telangiectasia. Anticancer research *28*, 401-405.

Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D., and Kazazian, H.H., Jr. (1996). High frequency retrotransposition in cultured mammalian cells. Cell *87*, 917-927.

Morrish, T.A., Gilbert, N., Myers, J.S., Vincent, B.J., Stamato, T.D., Taccioli, G.E., Batzer, M.A., and Moran, J.V. (2002). DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat Genet *31*, 159-165.

Muotri, A.R., Chu, V.T., Marchetto, M.C., Deng, W., Moran, J.V., and Gage, F.H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature *435*, 903-910.

Ostertag, E.M., and Kazazian, H.H., Jr. (2001). Biology of mammalian L1 retrotransposons. Annu Rev Genet *35*, 501-538.

Ostertag, E.M., Prak, E.T., DeBerardinis, R.J., Moran, J.V., and Kazazian, H.H., Jr. (2000). Determination of L1 retrotransposition kinetics in cultured cells. Nucleic Acids Res *28*, 1418-1423.

Riballo, E., Kuhne, M., Rief, N., Doherty, A., Smith, G.C., Recio, M.J., Reis, C., Dahm, K., Fricke, A., Krempler, A., Parker, A.R., Jackson, S.P., Gennery, A., Jeggo, P.A., and Lobrich, M. (2004). A pathway of double-strand break rejoining dependent upon ATM, Artemis, and proteins locating to gamma-H2AX foci. Molecular cell *16*, 715-724.

Shiloh, Y. (2006). The ATM-mediated DNA-damage response: taking shape. Trends in biochemical sciences *31*, 402-410.

Singer, O., Marr, R.A., Rockenstein, E., Crews, L., Coufal, N.G., Gage, F.H., Verma, I.M., and Masliah, E. (2005). Targeting BACE1 with siRNAs ameliorates Alzheimer disease neuropathology in a transgenic model. Nat Neurosci *8*, 1343-1349.

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

Uziel, T., Lerenthal, Y., Moyal, L., Andegeko, Y., Mittelman, L., and Shiloh, Y. (2003). Requirement of the MRN complex for ATM activation by DNA damage. Embo J *22*, 5612-5621.

Xia, X., Zhang, Y., Zieth, C.R., and Zhang, S.C. (2007). Transgenes delivered by lentiviral vector are suppressed in human embryonic stem cells in a promoter-dependent manner. Stem cells and development *16*, 167-176.

Yant, S.R., and Kay, M.A. (2003). Nonhomologous-end-joining factors regulate DNA repair fidelity during Sleeping Beauty element transposition in mammalian cells. Mol Cell Biol *23*, 8505-8518

CHAPTER 5

CONCLUSION

L1 retrotransposons have been ascribed many different roles in genome evolution, from parasitic, selfish elements to fine tuners of gene expression. Only recently has there been any evidence that L1 retrotransposition occurs actively on the level of the individual rather than at the population level. In a recent report Muotri and colleagues demonstrated that a human L1 is active both in rat derived NSCs *in vitro* and in the brains of L1 transgenic mice *in vivo* (Muotri et al.). They showed the regulation of L1 may be through the action of the transcription factor SRY-Box2 (SOX2) repressing L1 expression in self-renewing progenitors and that insertion of an L1 near a neuronally expressed gene could influence gene expression and cell fate choice in an *in vitro* system (Muotri et al.).

This thesis responds to the question of whether L1 retrotransposition occurs in the human nervous system; and data presented here supports this hypothesis. We showed that both fetal and hESC derived NSCs support active L1 retrotransposition. Both of these models have been shown to lead to functional neurons in the mouse brain after grafting (Muotri et al., ; Uchida et al.). Additionally, we investigated whether we could detect an increase in L1 copy number in adult human brains as compared to other tissues from the same individual, and found a robust increase in the hippocampus. Spiking known copies of L1 into heart or liver tissue confirmed this L1 copy number increase to roughly equal 80-300 copies per cell, a significant quantity.

Although this work addresses some questions about human L1 retrotransposition, it underscores a large array of other questions remaining to be addressed. The first proof in principle required is to isolate novel endogenous L1

insertions in the brain of an individual, and to show that this insertion is not present in other tissues from the same individual.

The advances in Solexa high throughput sequencing are making this a possibility, indicating this question will likely be answered in the coming years (Smith et al.). Solexa sequencing currently yields approximately 35 bases of sequencing data per read, with millions of reads per sample. Due to the L1 canonical long 3' polyA tail and the highly variable 5' truncations, L1 is a difficult target to isolate with such an approach. However, the advent of paired-ends sequencing, which yields 35 bases of read from *both* sides of a DNA fragment, will allow L1 DNA fragments to be sequenced from the L1 side as well as from the genomic insertion and allow *de novo* insertions to be identified. Not only will this approach fulfill the proof in principle of a single novel CNS insertion, it represents a high throughput method of isolating many insertions. In addition, this approach will address the loci of insertion in order to discover whether insertion sites are patterned in some way.

The data presented here suggests that retrotransposition occurs in the human brain and can be detected by qPCR in the adult and seen in both hESCs and NSCs. There are multiple research avenues opened through these studies. Firstly, only the hippocampus and cerebellum were analyzed here, leaving open the question of what a survey of many brain regions might indicate about L1 insertions. Additionally, the tissues analyzed here consisted of a mix of many cell types: neurons, astrocytes, oligodendrocytes. It is currently unknown whether L1 retrotransposition occurs equally in all progenitors, or whether it occurs in more lineage restricted NSCs. Future studies will address insertions on a single cell level. Do multiple insertions occur in a

single cell and are insertions cell type specific? Do many brain cells contain the same insertion or are insertions restricted so a single cell? In the course of selecting fetal brain tissues for methylation analysis, we performed some preliminary qPCR studies on these tissues as well, which suggested that perhaps L1 retrotransposition has already occurred by gestational day 80. The question of when L1 retrotransposition occurs in the developing nervous system is another study that remains to be pursued, but would ideally investigate multiple gestational ages as well as younger and older adult individuals.

Additionally, there is a need to investigate whether adult NSCs would support L1 retrotransposition. The large increase in ORF2 copy number in the hippocampus as compared to the cerebellum suggests that perhaps ongoing adult neurogenesis supports L1 retrotransposition and contributes to ORF2 copy number in the hippocampus. Unpublished transcriptional profiles from hESC, fetal NSCs, hESC derived NSCs, and adult brain have suggested that fetal NSCs cluster more with adult brain whereas hESC and derived NSCs cluster together. These data support our observation that the differentiation potential of hESC derived NSCs is greater and more robust than that of fetal NSCs. For instance, TH positive neurons are easily identifiable in hESC derived NSCs, but are found only in passage 1 or passage 0 fetal NSCs (Wright et al.). Furthermore, we found a 1000x increase in retrotransposition in hESC derived NSCs as compared to fetal NSCs. Therefore, we might expect that adult NSCs would support retrotransposition at a lower rate than fetal NSCs, with the majority of L1 retrotransposition events occurring during early neural development. This question has yet to be addressed.

A significant question concerns the insertions sites of L1 in the brain. It has previously been shown that the L1 endonuclease prefers a six base pair degenerate consensus sequence, AA/TTTT (Jurka). This sequence occurs frequently throughout the genome, and provides no hint as to L1 insertional specificity. Recent data from Muotri and colleagues isolated 17 insertions in cultured rat hippocampal NSCs, six of which were into neuronally expressed genes. Similarly, of the 18 insertions that we isolated from hESC derived NSCs in this work, eight were into or near neuronally expressed genes. One possible explanation for this might be that neuronal genes are actively expressed in NSCs and therefore their heterochromatin is open and relatively accessible as compared to other genomic locations (Muotri et al.). Only high throughput sequencing identifying large numbers of L1 insertions will resolve this question.

Another question which arises as a result of this work is what possible effect L1 retrotransposition may have in the human nervous system. The first possibility that arises is that L1 retrotransposition occurs randomly, a byproduct of evolution. It has been suggested that L1 retrotransposition in the gonads is evolutionally selected (Kazazian). This hypothesis indicates that L1s, as selfish parasites, continue to insert into the germline to propagate to the next generation (Branciforte and Martin). It has previously been noted that the transcriptome of the testes is remarkably similar to that of the brain, suggesting the possibility that brain expression of L1 is the result of "leaky" transcription (Guo et al.). A second possibility is that L1 retrotransposition in the brain may contribute to diversity, on the neuronal or transcriptional level. This hypothesis indicates that L1 retrotransposition, in a process which likely began by

chance, has been adapted to be of some utility in generating brain diversity (Muotri and Gage). Exploring this could greatly increase our understanding of neural development.

Lastly, we investigated whether L1 retrotransposition could be altered in the context of neurological disease. To this end we investigated L1 retrotransposition in models of Ataxia Telangiectasia, or A-T, a rare, neurodegenerative, autosomal recessive disorder characterized by a complete loss of function mutation in the Ataxia Telangiectasia Mutated (ATM) protein, a kinase essentially involved in sensing DNA damage (Mavrou et al.).

One question which has been actively debated in the DNA repair scientific community is whether the neurodegeneration seen in A-T is due to the DNA damage sensing property of ATM or whether a different activity of ATM is implicated (Biton et al., ; Lukas et al.). There are several studies which suggest that perhaps ATM in neurons is cytoplasmically located, whereas in other cell types ATM is nuclear and its DSB sensing properties must be nuclear (Barlow et al.). However, numerous studies have since dismissed this hypothesis and have shown that ATM is expressed in the nucleus in neurons and its role in neurons is a DSB sensing role similar to its role in other tissues (Barzilai et al., ; Uziel et al.). These studies have shown that ATM acts through activation of the MRN/mre11 complex to both auto-phosphorylate and to phosphorylate numerous proteins necessary to respond to DNA damage, such as p53, Chk2, and NBS1.

The first evidence for increased retrotransposition in the context of ATM deficiency came from crosses of the human L1 transgenic mouse with ATM knockout

mice. These mice have shown an appreciable increase in L1 retrotransposition in the brain as compared to control littermates. Preliminary studies indicate a substantial increase in the hippocampus but further study will be necessary to more quantitatively address which brain areas have an increase, . Future questions to address include quantitative studies both of ATM knockouts and controls to quantify L1 retrotransposition in various brain areas, as well as studies addressing whether this increase occurs early in embryogenesis or during adult life. This study would require comparing ATM knockout and wildtype littermates at different embryonic times and quantifying L1 retrotransposition.

A second question which arises from the preliminary ATM is data is whether there is increased L1 copy number in the brains of AT patients. Future studies will investigate L1 ORF2 copy number in a study analogous to that in Chapter 3 here, to inquire whether the finding of increased L1 retrotransposition in ATM knockout mouse brain can be translated to represent human endogenous L1 retrotransposition rates. Provided that increased L1 insertional rates are identified, no doubt a high throughput sequencing approach will ultimately be applied to not only identify increased insertional rates, but also to elucidate the structure and location of those insertions.

Lastly, the question of how ATM deficiency and L1 retrotransposition interact remains to be shown. The most straightforward explanation is that ATM senses DNA damage and in the absence of ATM this damage is undetected and therefore unrepaired (Biton et al.). In the context of decreased DNA break repair, L1 would act to patch these DNA breaks by engaging in endonuclease independent insertions.

Previous studies have shown that mutations in the non-homologous end-joining repair pathway (NHEJ), such as mutations in the XRCC4 or DNA protein kinase catalytic subunit genes in rodent cells, lead to large increases in endonuclease independent insertions (Morrish et al.). It remains to be shown if the same mutations in human cells would also lead to increase in L1 retrotransposition.

One hypothesis is that the act of L1 insertions, at an A/T rich consensus site and in an endonuclease dependent fashion, might activate DNA damage pathways due to the free 3' hydroxyl and single strand DNA breaks. If the endonuclease pathway is not responsible for increased retrotransposition, there are a number of other pathways either upstream of ATM, such as Mre11 which binds to the DNA breakage directly, or downstream, such as ATR, DNA-protein kinase, or p53, which might be responsible for mediating increased L1 retrotransposition in the context of decreased ATM (Matsuoka et al.). A recent study showned that DNA damage proteins ERCC1 and XPF1, which form a heterodimer and are involved in sensing "flap" intermediates (as opposed to DSBs), are involved in L1 retrotransposition (Gasior et al.). Decreases in these proteins results in increased endonuclease dependent (but *not* independent) retrotransposition. If ATM is involved in sensing and responding to flap intermediates, since it has been shown to interact with ERCC1/XPF (Matsuoka et al.), decreases in ATM would result in increased EN dependent retrotransposition. This hypothesis would be that ATM senses L1 integration as DNA damage and responds, resulting in truncation of the L1 insertion, and without ATM more insertions occur. This data is supported by the finding that majority of L1 insertions are truncated, however the L1 reverse transcriptase is highly processive, leading to the question of what causes

truncation (Bibillo and Eickbush). Given that full length insertions are known to be highly active (eg, LRE3 and L1$_{RP}$) and therefore mutagenic, ATM may be inhibiting full length insertions. It remains to be seen whether insertion sizes are increased in ATM deficient NSCs.

L1 retrotransposition has been variably viewed as parasitic and selfish, but alternatively also as a modulator of genome evolution, contributing to processes such as modulation of gene expression. This work has furthered the ideas of which cell types and processes may support active L1 retrotransposition. However, the question of L1 function remains to be addressed. Although there is now evidence that L1 retrotransposition may occur in the human brain, this work has not answered the question of whether this is an evolutionary boon or a barely contained deleterious process. The finding that L1 retrotransposition is increased in ataxia telangiectasia, a severe neurodegenerative disease, suggests that regardless of the possible utility of L1, it can be distorted in the context of neurological disease and may possibly contribute in part to the difficulties experienced by through excessive mutagenesis. Again, the function of L1 in the setting of A-T remains to be addressed.

WORKS CITED

Barlow, C., Ribaut-Barassin, C., Zwingman, T.A., Pope, A.J., Brown, K.D., Owens, J.W., Larson, D., Harrington, E.A., Haeberle, A.M., Mariani, J., Eckhaus, M., Herrup, K., Bailly, Y., and Wynshaw-Boris, A. (2000). ATM is a cytoplasmic protein in mouse brain required to prevent lysosomal accumulation. Proc Natl Acad Sci U S A *97*, 871-876.

Barzilai, A., Biton, S., and Shiloh, Y. (2008). The role of the DNA damage response in neuronal development, organization and maintenance. DNA repair *7*, 1010-1027.

Bibillo, A., and Eickbush, T.H. (2002). High processivity of the reverse transcriptase from a non-long terminal repeat retrotransposon. J Biol Chem *277*, 34836-34845.

Biton, S., Gropp, M., Itsykson, P., Pereg, Y., Mittelman, L., Johe, K., Reubinoff, B., and Shiloh, Y. (2007). ATM-mediated response to DNA double strand breaks in human neurons derived from stem cells. DNA repair *6*, 128-134.

Branciforte, D., and Martin, S.L. (1994). Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. Mol Cell Biol *14*, 2584-2592.

Gasior, S.L., Roy-Engel, A.M., and Deininger, P.L. (2008). ERCC1/XPF limits L1 retrotransposition. DNA repair *7*, 983-989.

Guo, J.H., Huang, Q., Studholme, D.J., Wu, C.Q., and Zhao, Z. (2005). Transcriptomic analyses support the similarity of gene expression between brain and testis in human as well as mouse. Cytogenetic and genome research *111*, 107-109.

Jurka, J. (1997). Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc Natl Acad Sci U S A *94*, 1872-1877.

Kazazian, H.H., Jr. (2004). Mobile elements: drivers of genome evolution. Science *303*, 1626-1632.

Lukas, J., Lukas, C., and Bartek, J. (2004). Mammalian cell cycle checkpoints: signalling pathways and their organization in space and time. DNA repair *3*, 997-1007.

Matsuoka, S., Ballif, B.A., Smogorzewska, A., McDonald, E.R., 3rd, Hurov, K.E., Luo, J., Bakalarski, C.E., Zhao, Z., Solimini, N., Lerenthal, Y., Shiloh, Y., Gygi, S.P., and Elledge, S.J. (2007). ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. Science *316*, 1160-1166.

Mavrou, A., Tsangaris, G.T., Roma, E., and Kolialexi, A. (2008). The ATM gene and ataxia telangiectasia. Anticancer research *28*, 401-405.

Morrish, T.A., Gilbert, N., Myers, J.S., Vincent, B.J., Stamato, T.D., Taccioli, G.E., Batzer, M.A., and Moran, J.V. (2002). DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat Genet *31*, 159-165.

Muotri, A.R., Chu, V.T., Marchetto, M.C., Deng, W., Moran, J.V., and Gage, F.H. (2005a). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature *435*, 903-910.

Muotri, A.R., and Gage, F.H. (2006). Generation of neuronal variability and complexity. Nature *441*, 1087-1093.

Muotri, A.R., Marchetto, M.C., Coufal, N.G., and Gage, F.H. (2007). The necessary junk: new functions for transposable elements. Hum Mol Genet *16 Spec No. 2*, R159-167.

Muotri, A.R., Nakashima, K., Toni, N., Sandler, V.M., and Gage, F.H. (2005b). Development of functional human embryonic stem cell-derived neurons in mouse brain. Proc Natl Acad Sci U S A *102*, 18644-18648.

Smith, A.D., Xuan, Z., and Zhang, M.Q. (2008). Using quality scores and longer reads improves accuracy of Solexa read mapping. BMC bioinformatics *9*, 128.

Uchida, N., Buck, D.W., He, D., Reitsma, M.J., Masek, M., Phan, T.V., Tsukamoto, A.S., Gage, F.H., and Weissman, I.L. (2000). Direct isolation of human central nervous system stem cells. Proc Natl Acad Sci U S A *97*, 14720-14725.

Uziel, T., Lerenthal, Y., Moyal, L., Andegeko, Y., Mittelman, L., and Shiloh, Y. (2003). Requirement of the MRN complex for ATM activation by DNA damage. Embo J *22*, 5612-5621.

Wright, L.S., Prowse, K.R., Wallace, K., Linskens, M.H., and Svendsen, C.N. (2006). Human progenitor cells isolated from the developing cortex undergo decreased neurogenesis and eventual senescence following expansion in vitro. Exp Cell Res *312*, 2107-2120.