UNIVERSITY OF CALIFORNIA, SAN DIEGO

Clearance of endogenous L1 retroelements in the cytosol by TREX1 prevents neuronal toxicity

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy

in

Biomedical Sciences

by

Charles Thomas

Committee in Charge:

Professor Alysson Muotri, Chair
Professor Frank Furnari
Professor Fred Gage
Professor Bruce Hamilton
Professor Gene Yeo

2014

The dissertation of Charles Thomas is approved, and acceptable in quality and form for publication on microfilm and electronically:

Chair

UNIVERSITY OF CALIFORNIA, SAN DIEGO

2014

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SUPPLEMENTAL FILES

ACKNOWLEDGEMENTS

I would like to acknowledge Alysson Muotri as mentor. I would also like to acknowledge Leon Tejwani as a major contributor on all aspects. I would like to acknowledge Cleber Trujillo and Roberto Herai for their contributions. Lastly, I would like to acknowledge the rest of the Muotri lab for their ideas and support.

Chapter 2, in part, is a reprint of the material as it appears in Annual Review of Cell and Developmental Biology, 2012. Thomas, C. A.; Paquola, A. C. M.; and Muotri, A. R. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in part, has been submitted for publication of the material as it may appear in Nature, 2014, Thomas, C.A.; Tejwani, L.; Herai, R.; Trujillo, C.A.; Crow, Y.J.; and Muotri, A.R. The dissertation author was the primary investigator and author of this paper.

VITA

2005	Bachelor of Science, Virginia Tech, Blacksburg, VA
2005-2008	Research Associate, GenVec, Inc, Gaithersburg, MD
2014	Doctor of Philosophy, University of California, San Diego, La Jolla, CA

PUBLICATIONS

Chen P., Hamilton M., **Thomas C.A.**, Kroeger K., Carrion M., Macgill R.S., Gehlbach P., Brough D.E., Wei L.L., King C.R., Bruder J.T. Persistent expression of PEDF in the eye using high-capacity adenovectors. *Mol. Ther.* **16**:1986-1994 (2008).

Bruder J.T., Semenova E., Chen P., Limbach K., Patterson N.B., Stefaniak M.E., Konovalova S., **Thomas C.**, Hamilton M., King C.R., Richie T.L., Doolan D.L. Modification of Ad5 hexon hypervariable regions circumvents pre-existing Ad5 neutralizing antibodies and induces protective immune responses. *PLoS One.* **7**:e33920 (2012).

Thomas C.A. and Muotri A.R. LINE-1: creators of neuronal diversity. *Front. Biosci.***4**:1663-1668 (2012).

Thomas C.A., Paquola A., Muotri A.R. LINE-1 retrotransposition in the nervous system. *Annu. Rev. Cell Dev. Biol.* **28**:555-573 (2012).

Bruder J.T., Chen P., Semenova E., **Thomas C.A**., Konovalova S., Ekberg G., Ettyreddy D., McVey D., Gall J.G., King C.R., Brough D.E.. Identification of a suppressor mutation that improves the yields of hexon-modified adenovirus vectors. *J Virol*. **87**:9661-9671 (2013).

ABSTRACT OF THE DISSERTATION

Clearance of endogenous L1 retroelements in the cytosol by TREX-1 prevents neuronal toxicity

by

Charles Thomas

Doctor of Philosophy in Biomedical Sciences

University of California, San Diego, 2014

Professor Alysson Muotri, Chair

This Dissertation is broken into four chapters. In chapter 1 is an introduction into the thesis. In chapter 2, I review the Long Interspersed Element-1 (L1), which is a repetitive DNA retrotransposon capable of duplication by a “copy-and-paste” genetic mechanism. Scattered throughout mammalian genomes, L1 is typically quiescent in most somatic cell types. In developing neurons, however, L1 can express and retrotranspose at high frequency.

In chapter 3, I record my findings on using zinc-finger repressors to inhibit murine L1. The mouse has an estimated 3100 putatively active L1 elements, split into three families, T_F, G_F, and A. Each family is distinguished by their promoter-like 5’UTR. Using zinc-finger repressors, we inhibit the expression of mouse L1 T_F family.

In chapter 4, I document my discoveries on ssDNA and L1 in three prime repair exonuclease 1 (TREX1)-deficient neural cells. Accumulation of

deoxyribonucleic acid (DNA) species in the cytosol leads to a type I interferon response and inflammation. TREX1 removes single-stranded (ss) and double-stranded (ds) DNA from the cytosol, preventing accumulation and a subsequent inflammatory response. Several autoinflammatory diseases, such as Aicardi-Goutières syndrome syndrome (AGS) and systemic lupus erythematosus, arise when the function of TREX1 is compromised. AGS is a severe autoinflammatory disease most typically affecting the brain and the skin, resulting in severe psychomotor retardation. Research on the neurological aspects of AGS at a cellular and molecular level has proven difficult, due to the lack of relevant animal models. In particular, although demonstrating a lethal inflammatory state, the *TREX1* knockout mouse does not exhibit a neurological phenotype, and is thus unsuitable to study AGS neuropathology. Here, we create pluripotent stem cell models devoid of TREX1 function, which we further differentiate into neural cells. We find an abundance of ssDNA in TREX1-deficient neural precursor cells, neurons, and astrocytes. Furthermore, we show the Long Interspersed Element-1 (LINE-1 or L1), a class of endogenous retrotransposons, is a major source of the ssDNA, and that inhibition of reverse transcription reduces ssDNA levels. TREX1-deficient neurons experience greater cytotoxicity, which can be rescued with reverse transcriptase inhibitors (RTi). Likewise, treating neurons with conditioned media from TREX1-deficient astrocytes increases neuronal toxicity, indicating the presence of toxic factors or a lack of neurotrophic factors. Our results demonstrate that TREX1 removes cytosolic ssDNA in neural cells created by L1 reverse transcription, thus preventing neurotoxicity.

CHAPTER 1. INTRODUCTION

1.1 History of retrontansposition

Barbara McClintock first described transposition in 1950, when she discovered the “mutable loci” in maize (McClintock, 1950). These “mutable loci” were able to move from one chromosome locus to another and could change the expression pattern of nearby genes. Her work on transposition was originally received with great skepticism by the scientific community, but research in the 1970’s proved her “mutable loci” were actually class II DNA transposons. Class II DNA transposons utilize a “cut-and-paste” mechanism to reposition itself within a genome (Reznikoff, 2003). Class II transposons have not been shown to be active in mammals. However, class I retrotransposons, such as the Long Interspersed Element 1 (L1), are very active. Utilizing a “copy-and-paste” mechanism, retrotransposons can insert into dispersed genomic locations that can change the chromatin and nearby gene expression (Kazazian et al., 1988; Lerman et al., 1983; Skowronski and Singer, 1985). Host cells use a variety of defenses against retrotransposons to prevent deleterious insertions (Goodier and Kazazian, 2008). In fact, virtually all somatic cells display little to no retransposition activity (Ostertag et al., 2002). Recent evidence, however, has determined neural precursor cells (NPCs) are an exception and actually support high levels of L1 retrotransposition (Muotri et al., 2005).

1.2 The L1 element

L1 is a DNA element of approximately 6-7kb that, when fully intact, is capable of retrotransposition. (Adams et al., 1980; Warren et al., 2008). From 5’ to 3’, the element consists of a promoter, two open reading frames, and polyA tail. These components enable the L1 retrotransposition by transcribing into RNA, translating into

two functional proteins, and reinserting into the genome by a mechanism called target-primed reverse transcription (Cost et al., 2002).

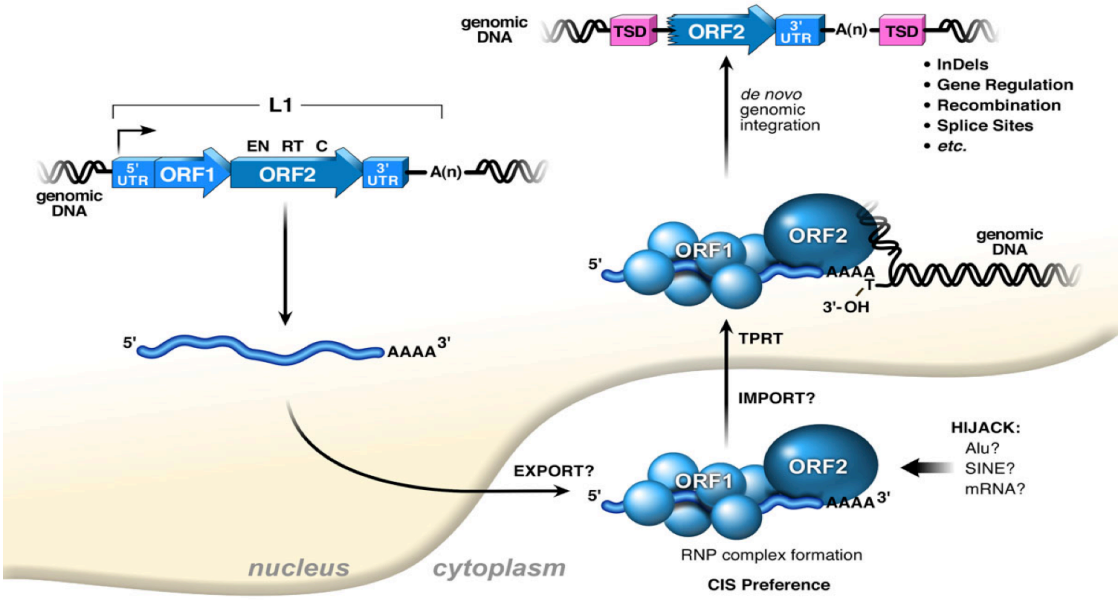


Figure 1. The L1 life cycle. The L1 element exists within the genome in all mammals. Full-length elements consist of a 5'UTR with promoter function, two open reading frames, a 3'UTR, and a short poly-A tail. Pol II transcribes the element into RNA, which is transported to the ribosome to make the proteins ORF1p and ORF2p. ORF1p is RNA chaperone while ORF2p has endonuclease and reverse transcriptase functions. ORF1p and ORF2p bind the L1 RNA with cis-preference and re-enter the nucleus, most likely when the nuclear envelope is broken down during cell division. ORF2p nicks a 5'-TTAAAA-3' site, and begins reverse transcribing the RNA. Reverse transcription is inefficient, and often the newly transcribed element is truncated. If the reverse transcription was efficient and the newly inserted element contains all functional features, this new element could express and begin the L1 life cycle again.

L1s comprise ~17% of the human genome (Lander et al., 2001a). L1 is abundant in all mammalian genomes; however, most are 5' truncated and thus no longer active. In humans, there are approximately 520,000 elements but only 3000 full-length elements (Grimaldi et al., 1984; Lander et al., 2001a). Of these full-length elements, only ~90 L1s had intact ORFs and only 6 were extremely active (Brouha et

al., 2003). These 6 elements comprised 84% of the retrotransposition capability of the intact L1s.

In 1988, it was discovered that two cases of hemophilia A were due to L1 mutagenesis into the *Factor VIII* gene (Kazazian et al., 1988; Woods-Samuels et al., 1989). This was the first evidence of L1 activity in humans. Since, then, L1 was discovered to be responsible for mutagenesis of several genes leading to disease, including Duchenne muscular dystrophy, retinis pigmentosa, and familial adenomatous polyposis Holmes et al. (1994); (Miki et al., 1992; Narita et al., 1993; Schwahn et al., 1998). From these data, we can conclude L1 is functional and active in human genomes. Furthermore, L1-mediated retrotransposition events may account for approximately 1 in 1000 spontaneous, disease-inducing insertions in humans (Beck et al., 2011; Chen et al., 2005; Kazazian and Moran, 1998).

All the active human L1 elements are from a subset of an L1 family, called Ta-1 (Boissinot et al., 2000). The Ta subset is characterized by a 3 base pair change (GAG → ACA) from older families at 92-94 bp upstream of the polyA tail (Sassaman et al., 1997; Skowronski and Singer, 1986). The mechanism how the ACA substitution contributes to better L1 retrotransposition activity remains unclear.

When compared with humans, the mouse has more elements and more activity. The mouse has approximately 3100 putatively active L1s in their genome, split into three families, T_F, G_F, and A. Of the 3100 elements, approximately 1800 are T_F, 400 G_F, and 900 A (DeBerardinis et al., 1998; Goodier et al., 2001). The L1 families are distinguished by their 5'UTR promoter. The 5'UTR of mouse L1 is typically composed of 2-11 "monomer" repeats, which are approximately 200 base pairs.

The murine L1 is very active in mouse and has contributed to approximately 2-3% of spontaneous mutations. Several mutant mouse lines arose from L1 retrotransposition, including the spastic mouse, the Orleans reeler mouse, the black-eyed white mouse, the beige mouse, and the med mouse (Kingsmore et al., 1994; Kohrman et al., 1996; Perou et al., 1997; Takahara et al., 1996; Yajima et al., 1999). Interestingly, the insertion in the spastic, Orleans reeler, and black-eyed white mouse were full insertions, where as the beige mouse and med mouse were severely truncated insertions.

While the bulk of L1 research has been focused on germ-line mutagenesis, somatic mutagenesis is becoming more appreciated. To study retrotransposition in a living animal, Ostertag *et. al.* created a mouse model expressing a human L1 retrotransposition cassette under the control of it's own promoter (Ostertag et al., 2002). Their studies found high retrotransposition levels in spermatocytes and oocytes, but not in any somatic cells. However, when Muotri *et. al* closely examined the brain of transgenic mouse and found very robust and high levels of retrotansposition in the brain and specifically in neurons (Muotri et al., 2005). From these data, we conclude somatic retrotransposition does occur; however, the significance of somatic retrotransposition is unknown.

1.3 Methods to inhibit L1 retrotransposition

One way to assess significance of somatic retrotransposition would be to inhibit L1 activity. Since retrotransposition requires reverse transcription, one could use RTis such as the nucleoside analogues used for the treatment of HIV-infected patients. In one such study, Jones *et. al* compared the effectiveness of 4 nucleoside analogues on inhibiting L1 retrotransposition from an L1-eGFP cassette in HeLa cells

(Jones et al., 2008). The study used the following nucleoside analogues stavudine, zidovudine, tenofovir disoproxil fumarate, and lamivudine. Stavudine was the most effective at inhibiting L1 retrotransposition, with an IC_{50} of 220 nM.

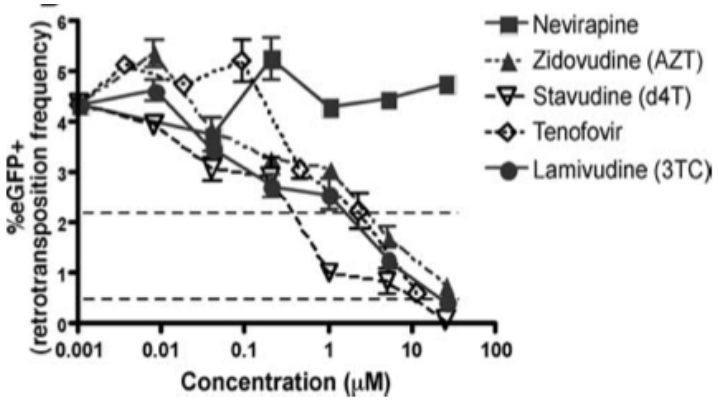


Figure 2. Reverse transcriptase inhibitors on L1 retrotranspostion. The nucleoside analogue reverse transcriptase inhibitors (NARTis) zidovudine, stavudine, tenovir disoproxil fumarate, and lamivudine inhibit L1 well, while the non-nucleoside analogue reverse transcriptase inhibitor nevirapine does not. The IC_{50} of the NARTis: zidovudine – 2.21 μM; stavudine – 0.22 μM; tenofire disoproxil fumarate – 1.82 μM; and lamivudine – 1.12 uM. Dashed lines represent 50% and 90% inhibition. Image retrieved from Jones et al., 2008.

Having a genetic model with inhibited L1 retrotransposition would be ideal for studying the role of somatic retrotransposition. Unfortunately, creating an L1-knockout would be near impossible with over 3000 putative active elements. However, one could conceivably create a transgenic mouse expressing a repressor that binds to L1-promoter. Using a DNA binding protein in conjunction with a RNA polymerase repressor has proven effective to inhibit expression (Bellefroid et al., 1991). Typically, the Kruppel-associated box (KRAB) domain is used in conjunction with zinc-fingers, transcription activator-like effectors (TALEs), or the clustered regularly interspaced short palindromic repeats- (CRISPR)/Cas system (Cong et al., 2012; Larson et al., 2013).

Zinc-finger repressors (ZFRs), which are highly expressed in mammals, are modular proteins with two main domains: the zinc-finger protein (ZFP)-DNA binding domain and the KRAB repression domain (Bellefroid et al., 1991). The DNA binding domain contains a tandem array of cysteine and histidine fingers (Miller et al., 1985; Wolfe et al., 2000). Each finger binds specifically to a tri-nucleotide sequence. In the human genome, approximately 1/3 of the zinc-finger proteins are associated with the KRAB repressor, indicating an important role in gene regulation (Rousseau-Merck et al., 2002). The function of ZFRs is to inhibit transcription by recruiting histone deacetylases and to promote DNA condensation (Ryan et al., 1999).

TALEs, like zinc-fingers, are modular DNA-binding proteins. However, TALEs derive from the plant pathogenic bacteria *Xanthomonas* and *Ralstonia* (Boch et al., 2009). These bacteria use TALEs to modulate host gene expression to promote it's own survival. The TALE DNA-binding domain consists of multiple monomer repeats of 34 amino acids. Amino acid positions 12 and 13 of the repeat are referred to as the repeat-variable di-residues (RVDs) and confer binding specificity for one specific DNA base (Deng et al., 2012; Streubel et al., 2012). Linking multiple monomers, one could create a TALE specific to any sequence. Fusing the TALE DNA-binding domain with an effector, such as KRAB, would yield a repressor specific to the user's desired genomic loci.

The CRISPR/Cas system originates from bacteria as an RNA-guided defense mechanism against viral pathogens by detecting and silencing their nucleic acid species (Wiedenheft et al., 2012). The bacteria use the CRISPR system to express RNA that is complementary to the pathogenic sequence. In addition to the complementary pathogenic sequence, the CRISPR RNA has a secondary structure to

bind to the endonuclease Cas9. Cas9, when in contact with the pathogenic sequence, will cleave the foreign nucleic acid species. This system has been modified and optimized for use in human systems (Cong et al., 2013; Mali et al., 2013). Furthermore, the Cas9 protein can be mutated to remove its endonuclease activity. When the mutated Cas9 protein is fused with a KRAB repressor, the CRISPR/Cas9 system can be used to inhibit transcription without damage to the genome (Gilbert et al., 2013; Larson et al., 2013; Qi et al., 2013).

1.4 L1 and neurological disorders

Another interesting aspect of L1 biology is the relationship to several neurological disorders, including Rett syndrome (RTT), ataxia-telangiectasia (A-T), schizophrenia, and Aicardi-Goutieres syndrome (AGS) (Bundo et al., 2014; Coufal et al., 2011; Muotri et al., 2010; Stetson et al., 2008). Each disease has unique pathophysiology relating to L1. For instance, RTT syndrome arises due to mutations in the methyl-CpG-binding protein 2 (*MECP2*) gene. The MeCP2 protein has been shown to control L1 expression by binding to its methylated promoter (Yu et al., 2001). Thus, loss MeCP2 protein yields higher retrotransposition events that are detected in the neurons of RTT patients (Muotri et al., 2010). Unlike RTT, A-T patients do not have higher rates of retrotransposition in neurons, but longer insertions of the L1 element, suggesting the ataxia-telangiectasia mutated protein is involved in the inhibition of reverse transcription (Coufal et al., 2011). The schizophrenia publication showed a similar phenotype as the RTT publication with greater level of L1 insertions in neurons (Bundo et al., 2014). However, it is not clear how or why schizophrenic patients have a greater level of neuronal L1 mutagenesis.

AGS is a severe autoinflammatory disease most typically affecting the brain and the skin, resulting in severe psychomotor retardation (Aicardi and Goutieres, 1984). The effects on the nervous system are dramatic, resulting in progressive microcephaly, intracranial calcification, leukodystrophy, and infiltration of leukocytes. The disease typically presents in the first year of life, and is often misdiagnosed with congentital infection. A hallmark of AGS is the high level of interferon alpha (IFNa) in the cerebral spinal fluid, signifying a type-I IFN response (Goutieres et al., 1998). Several studies have shown cytosolic nucleic acid species can induce a type-I IFN response (Ishii et al., 2006; Martin and Elkon, 2006; Stetson and Medzhitov, 2006).

Dr. Jean Aicardi and Dr. Francoise Goutieres first described the disease in 1984 (Aicardi and Goutieres, 1984). Due to the high prevalence among siblings both male and female, Aicardi and Goutieres deduced the disease was an autosomal recessive trait. We now know mutations in 6 different genes can give rise to AGS. The first gene to be described was the three prime repair exonuclease 1 (*TREX1*) (Crow et al., 2006a; Crow et al., 2000) In addition, mutations in the subunits (α,β, and γ) of Ribonucelease H2 (*RNase H2*), SAM domain and HD domain-containing protein 1 (*SAMHD1*), and adenosine deaminase acting on RNA 1 (*ADAR1*) can cause AGS (Crow et al., 2006b; Rice et al., 2009; Rice et al., 2012). Interestingly, all of the proteins associated with AGS are involved in the repression of retroviruses and retroelements, such as HIV or the endogenous L1 element.

TREX1 is a DNase that breaks down both single-stranded (ss) and double-stranded (ds) DNA in a 3'-5' fashion (Mazur and Perrino, 1999). The gene consists of a single coding exon with three conserved motifs that form the exonuclease domain Exo I, Exo II, and Exo III (Barnes et al., 1995). Furthermore, TREX1 has a

transmembrane domain that anchors the protein to the endoplasmic reticulum, permitting the protein to degrade DNA species in the cytosol (Richards et al., 2007). TREX1 plays an important role in immune defense, protecting the host cell from retrtoviruses (Yan et al., 2010).

The *Trex1* knockout mouse has systemic inflammation and dies of myocarditis (Morita et al., 2004). Unlike in AGS patients, the *Trex1* knockout mouse has no inflammation in the brain (Gall et al., 2012; Pereira-Lopes et al., 2013). However, murine embryonic fibroblasts from the *Trex1* knockout mouse have an increase of ssDNA in their cytosol (Yang et al., 2007). Furthermore, the *Trex1* knockout mouse has a type-I IFN signature (Stetson et al., 2008). Together, we can deduce the loss of Trex1 protein leads to an increase of cytosolic DNA and subsequent type-I IFN response. The source of cytosolic DNA has been debated, but Stetson *et al.* had convincing data that retroelements are a major source (Stetson et al., 2008).

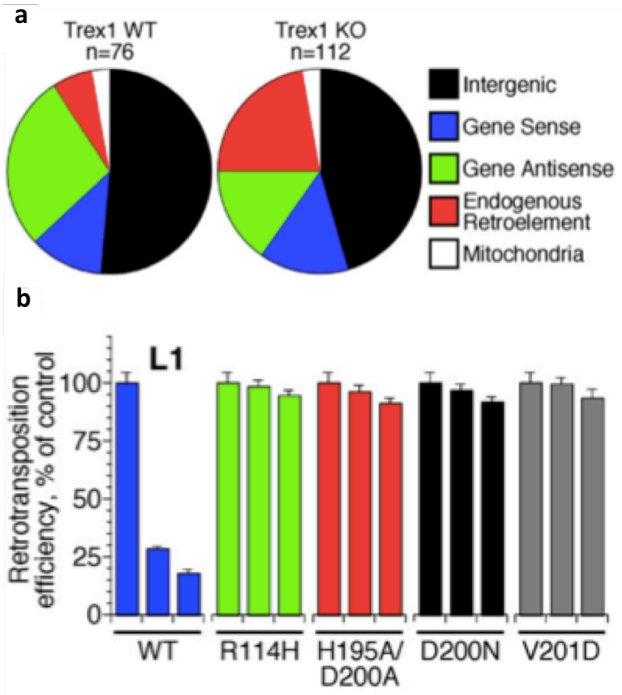


Figure 3. Endogenous retroelements, Trex1, and retrotransposition. a, DNA species from hearts of Trex1 KO and control WT mice (3 pooled hearts). Trex1 KO mice had 3-fold more endogenous retroelements, including 4-fold of L1 elements. **b**, TREX1 expression reduces retrotransposition from an L1-neomycin reporter assay in HeLa cells. L1 reporter was transfected with 0 ng (first blue bar), 250 ng (second blue bar), or 500 ng (third blue bar) of TREX1 overexpression vector. Mutant forms of TREX1 (green, red, black and grey bars), do not inhibit L1 retrotransposition. Image retrieved from Stetson et al., 2008

Research on the neurological aspects of AGS at a cellular and molecular level has proven difficult, due to the lack of relevant animal models. To study the role of TREX1 in neural cells, we created a human pluripotent stem cell models devoid of TREX1 function, which we further differentiate into neural cells. Using the neural cells, we have found a rise in L1 element DNA in the cytosol of NPCs, neurons and astrocytes. Furthermore, the loss of TREX1 in the neurons resulted in high levels of apoptosis. We also found the astrocytes expressed interferon response genes, signifying an immunologic response to the cytosolic DNA. Furthermore, we found

RTis stavudine and lamivudine can prevent the accumulation of ssDNA in the cytosol of neural cells and thus prevent neuronal apoptosis.

For the first time, we are able to create a system to study the neurological implications of the loss of TREX1 function in AGS-like cells on the molecular level. We were able to identify the molecular source of neuronal toxicity and subsequent inflammation. From this data, we also identified RTis as a possible medication that can be used to treat AGS patients. These findings amplify our understanding of AGS and pinpoint a possible method of intervention.

CHAPTER 2 LINE-1 RETROTRANSPOSITION IN THE NERVOUS SYSTEM

2.1 Form, function, and biology of L1

L1 is a 6 kilobase repeated DNA element that exists in all mammalian genomes (Adams et al., 1980; Warren et al., 2008). As an autonomous element, full-length, retrotransposition-competent L1s contain all necessary sequence information to express and re-insert themselves into the genome. From 5' to 3', the element consists of a promoter, two open reading frames, and polyA tail. Open reading frame 1 protein (Orf1p) is a nucleic acid chaperone that stabilizes the L1 RNA molecule (Martin and Bushman, 2001). Open reading frame 2 protein (Orf2p) binds the 3' end of the L1 RNA molecule and facilitates re-integration of the element back into the genome with three key domains. The zinc-knuckle domain likely assists the L1 RNA molecule to come into proximity to its integration site (Fanning and Singer, 1987). The endonuclease domain nicks a single strand of DNA at the consensus AA/TTTT sequence (Feng et al., 1996). The third domain is the reverse transcriptase (RT), which transcribes the RNA molecule into DNA sequence (Dombroski et al., 1994; Hattori et al., 1986; Mathias et al., 1991).

The life cycle of L1 begins with RNA polymerase II transcription and export into the cytoplasm where Orf1p and Orf2p are translated (Alisch et al., 2006; Kulpa and Moran, 2005). The Orf proteins and the L1 RNA assemble together to form ribonucleoprotein (RNP) complexes (Hohjoh and Singer, 1996; Kolosha and Martin, 2003; Martin and Bushman, 2001). To re-insert back in the genome, the RNPs must come into contact with genomic DNA, however it is not clear if the RNPs re-enter the nucleus via a transport mechanism or if nuclear envelope breakdown is required (Kubo et al., 2006). Target-primed reverse transcription (TPRT) is the suggested

mechanism L1 employs to integrate into the genome (Cost et al., 2002). This mechanism, first described for the R2 element, requires a nick in DNA at a target site, priming of the DNA, and reverse transcription (Luan et al., 1993). Orf2p provides all these functions (Feng et al., 1996; Mathias et al., 1991). To recreate another fully functional element, the entire L1 RNA must be reverse transcribed. However, the majority L1 insertions are 5'-truncated and average only 1 kilobase in length (Lander et al., 2001b). This observation suggests TPRT is extremely inefficient, likely due to mechanisms host cells employ to defend against TPRT.

2.2 L1 retrotransposition affects gene expression

When L1 inserts into an intragenic region, the element can disrupt expression via a variety of genetic mechanisms. Most obvious, insertions within a coding region will could change the codon code and create missense or nonsense mutations (Kazazian et al., 1988). L1 can also change a gene's splicing pattern with cryptic splice sites or by exon skipping (Mulhardt et al., 1994; Takahara et al., 1996). In addition, the L1 5'UTR has both sense and antisense promoter activity. These promoters can create new transcription start sites or enhance already existent promoters upon insertion (Speek, 2001; Wolff et al., 2010).

L1 insertion mutagenesis can also affect the genome globally by altering the epigenetic landscape and creating genomic rearrangements. Various sequences of the element influence the local chromatin architecture by recruiting chromatin modifiers. For instance, the 5'UTR promoter sequence of L1 is a CpG-rich sequence, which can be silenced by methylation (Coufal et al., 2009; Yu et al., 2001). Although the *ORF2* and the 3'UTR are AT-rich sequences, they can also signal for chromatin

condensation by unknown mechanisms (Chow et al., 2010; Garcia-Perez et al., 2010). We suspect silencing of L1 would also affect nearby genes, although this hypothesis has yet to be explored. The L1 sequence can also mediate genomic deletions and non-allelic homologous recombination. These genomic rearrangements have been demonstrated in transformed cell lines, certain spontaneous diseases, and during evolution (Burwinkel and Kilimann, 1998; Gilbert et al., 2002; Han et al., 2005).

2.3 Alternative mechanisms of L1 affecting gene transcription

The proteins expressed by L1 enable autonomous retrotransposition. However, these proteins can also be hijacked by small interspersed element (SINE) RNA to retrotranspose back into the genome. The most successful and active SINE in the human genome is the *Alu* repeats. The *Alu* element is approximately 300 bases and has over one million copies scattered throughout the genome (Lander et al., 2001b). In addition to *Alu*, the only other SINE still active in the human genome is the SVA element (Wang et al., 2005). Other RNAs, coding and non-coding, can also hijack Orf2p and insert into new locations within the genome, creating new pseudogenes and regulatory sites (Buzdin et al., 2002; Maestre et al., 1995).

Besides being hijacked by other RNA elements for retrotransposition, Orf2p can be involved disrupting the stability of chromosomes. A recent publication found Orf2p to be involved genomic rearrangement in prostate cancer (Lin et al., 2009). In conjunction with the androgen receptor and the activation-induced deaminase (AID), the Orf2p endonuclease induces DNA double-strand breaks in specific loci, which is followed by non-homologous end joining to produce chromosomal translocations.

Although somatically derived translocations and copy number variations (CNVs) have not been assessed in benign neural tissue, it is an interesting hypothesis to consider.

For a more detailed description of L1 and its biology, refer to the Beck et al. Annual Review (Beck et al., 2011).

2.4 The discovery of neuronal retrotransposition

Multipotent neural stem cells reside in neurogenic regions of the brain. They have three fundamental characteristics. First, neural stem cells can remain multipotent and continue to replicate in the neurogenic niche. The second characteristic is the ability to differentiate into glial progenitors which will mature into astrocytes or oligodendrocytes. The last characteristic of the neural stem cell is to differentiate into neuronal progenitor cells (NPCs). When a stem cell commits to the neuronal lineage, it triggers a specific gene expression profile. Interestingly, L1 is one of the genes that are upregulated upon neuronal commitment (Muotri et al., 2005). This observation led to several puzzling questions: Why would a repetitive element be one of the highest expressed genes in NPCs? Also, what is this repetitive element's function in neurogenesis? And since this repetitive element is an autonomous retrotransposon, does it actually retrotranspose in NPCs? The first two questions are hotly debated as research groups continually work to find function and purpose for L1 expression and retrotransposition in neurons. However, we now know L1 is capable of high levels retrotransposition in NPCs, generating a neuronal genetic mosaicism. We detail the data in subsequent paragraphs.

Because mammalian genomes are composed of approximately 20% of L1-derived sequences, it is extremely difficult to detect novel L1 retrotransposition events. The situation is analogous to finding a new straw of hay placed in the middle

of a haystack. To overcome this obstacle, an engineered active L1 retrotransposition cassette can utilize a marker that only expresses if the element inserts back into the genome (Freeman et al., 1994; Moran et al., 1996). For example, if the marker in the L1 cassette is eGFP, then a cell expressing eGFP indicates the engineered L1 successfully retrotransposed. Utilizing the L1-eGFP cassette, the ability of NPCs to support L1 retrotransposition was first reported in rat hippocampal NPCs in vitro (Muotri et al., 2005). Since then, the L1-eGFP cassette has proven L1 can also retrotranspose in human NPCs in vitro and mouse NPCs in vivo (Coufal et al., 2009; Muotri et al., 2005).

The creation of the L1-eGFP transgenic mouse allowed researchers to study somatic retrotransposition in all tissues. As expected, L1 is capable of retrotransposition in germ cells (Ostertag et al., 2002). Perhaps unexpectedly, L1 also retrotransposed in the brain as several cells expressed eGFP from the L1 retrotansposition cassette (Muotri et al., 2005). Upon closer review, it was discovered that the retrotransposed cells also expressed the neuronal marker NeuN. No eGFP expression was detected in S100-β and glutathione S-transferase positive cells, indicating that the glial cells did not support high-levels of retrotransposition. However, we cannot completely discard the possibility of eGFP transgene silencing in glia. The brains of the L1-eGFP mouse contained many eGFP-positive cells in virtually all regions, including the striatum, cortex, hypothalamus, hilus, cerebellum, ventricles, amygdala, and hippocampus. This result suggests that several NPC types, including those arising from different niches at different time points, are susceptible to L1 somatic retrotransposition.

L1 integration events in Rat NPCs have been shown to alter expression of nearby genes by promoter enhancement and epigenetic silencing (Muotri et al., 2005). The alteration in gene expression directly affected neuronal fate and function. Our group has also detected chromosomal rearrangements in our neuronal culture that we believe was mediated by L1 (unpublished data). Thus, we postulate that L1 retrotransposition would affect gene expression in vivo as well.

2.5 The magnitude of neuronal retrotransposition

The ability of L1 to retrotranspose in neurons is surprising. However, the magnitude of L1 retrotransposition in neurons is even more surprising. When using the L1-eGFP cassette, an eGFP positivecell signifies one engineered L1 was able to retrotranspose. However, we do not know how many endogenous L1s actually retrotransposed. The human genome and mouse genome have approximately 150 and 3,000 putatively active L1s in their genome, respectively. One way to determine the amount retrotransposition is to perform quantitative PCR on the genomic DNA and compare the level of L1 content. Coufal et al. performed this experiment using post-mortem human tissue. The authors compared the amount of L1 sequence content between neural tissue with the amount of L1 content in heart and liver tissue and found significantly greater L1 in the neural tissue (Coufal et al., 2009). The dentate gyrus, frontal lobe, and spinal cord contained the most L1 DNA copies. The authors went on to estimate that the cells of the hippocampus contained 80 more copies of L1 sequences in their genome when compared to the cells of the heart and liver. It should be noted, however, the qPCR strategy does not necessarily support the idea of retrotransposition. The increase in L1 sequences detected by qPCR could

be caused by other genetic mechanism, such as duplications or aneuploidy (chromosome gain).

Determining the integration sites of de novo L1 sequences in NPCs would yield insights into the potential effects of neuronal retrotransposition. Early sequencing efforts in neurons discovered L1 could integrate into introns of active genes (Coufal et al., 2009; Muotri et al., 2005). The sites of L1 integration included metabolic, DNA repair, cell adhesion, cell signaling, and neuronal-specific genes. More sophisticated and deeper sequencing techniques on human postmortem brain tissue have determined L1 integrates into intragenic sites approximately 47% of the time and within exons 3.5% (Baillie et al., 2011). Using the deep sequencing data, the authors performed gene ontology analysis on the sites of intronic L1 integration. Synaptic proteins, cell adhesion, and cell projection were the most enriched genes. Other gene groups that were enriched include phosphorylation, axonogenesis, axon development, and neuron development.

The results from the deep sequencing analysis of de novo L1 insertions are not very surprising if one considers gene size in the human genome. Many of the largest genes in the genome are neuronal and cell projection related. Interestingly, a simple computational experiment of picking random genomic positions and performing GO analysis on those that map into genes also revealed significant enrichment for terms like synapse, axon, cell adhesion, etc. in the human genome (**Supplementary Table 1**). The same experiment was performed in the mouse, fly and worm. Random positions in the mouse genome found enrichment in neuron-related genes, as well. The fly and worm genomes found enrichment in cell membrane and projection proteins. The enrichment of neuronal genes in human can

be explained by the fact that some neuron-related GO terms have average gene length larger than the global gene length average (**Fig. 1**). Therefore, neuronal retrotransposition would have the greatest effect on axonogenesis, dendritogenesis, and synaptogenesis.

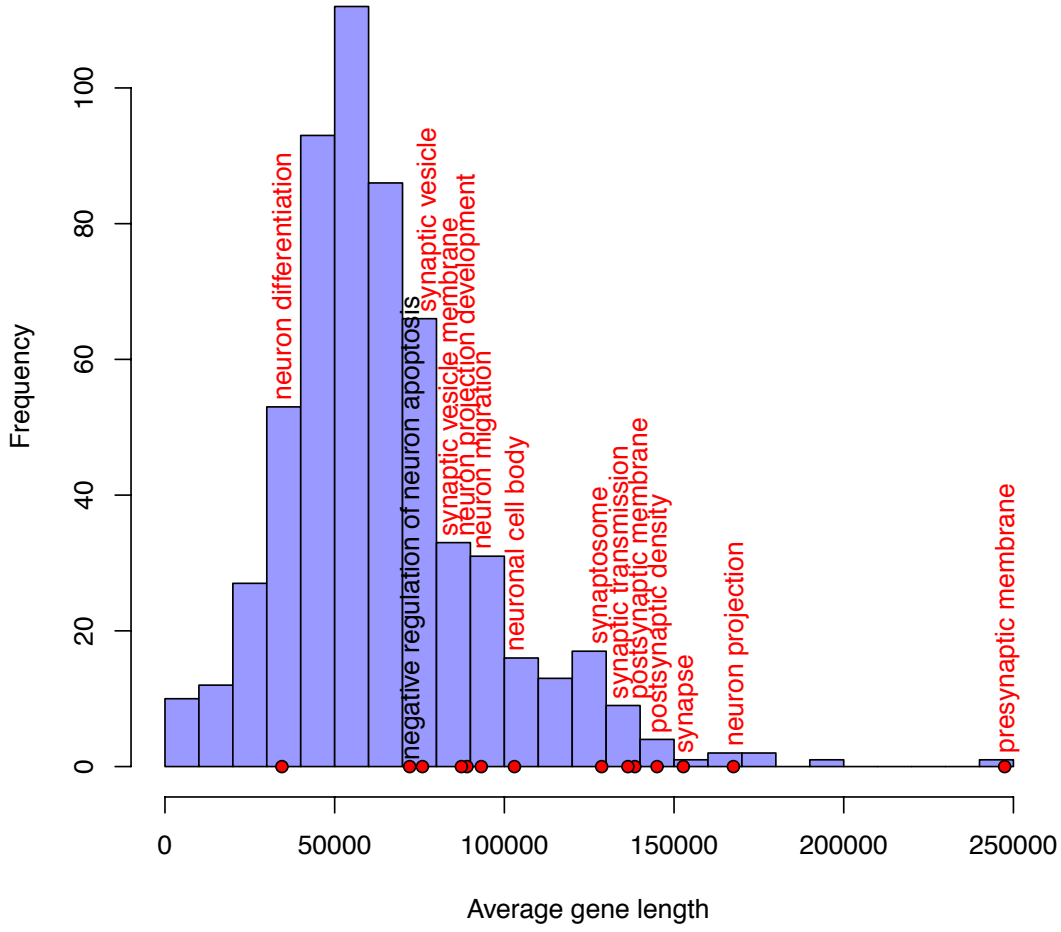


Figure 4. Ontology of human genes by length. The average human gene length of gene ontology groups is depicted above. Most neuronal genes are larger than the average gene size. Synaptic genes are some of the largest genes in the genome.

2.6 The mechanism of neuronal retrotransposition

The unique ability of L1 to retrotranspose in neuronal cells can partially be explained by its mechanism of expression. The promoter of L1 contains several transcription factor binding sites, including sites for Yin-Yang 1 (YY1), the Sex Determining-Region Y-box (Sox), and the Transcription Factor-Lymphoid Enhancer Factor (TCF-LEF) (Athanikar et al., 2004; Tchenio et al., 2000). The Sox2 protein is a key transcription factor to keep neural stem cells proliferative and multipotent (Bylund et al., 2003; Graham et al., 2003). Sox2 is also partially responsible for inhibiting L1 expression in neural stem cells, by associating with the L1 promoter. Sox2 forms a repressor complex with histone deactylase 1(HDAC1) and histone 3 methylated at Lysine 9 (H3K9) (Muotri et al., 2005). In addition, the numerous CpG sites of the L1 promoter are highly methylated in neural stem cells, which recruits methyl-CpG-binding protein 2 (MeCP2) to further repress the promoter (Muotri et al., 2010; Yu et al., 2001).

As the neural stem cell commits to the neuronal lineage, Sox2 expression decreases, the L1 promoter demethylates and the repressor complexes dissociate from the L1 promoter. Concurrent with the dissociation of repressors, a β-catenin activator complex binds to the TCF-LEF sites, stimulated by Wnt-3a (Kuwabara et al., 2009). These factors induce elevated L1 expression in NPCs. One of the most interesting aspects of L1 expression in NPCS is the close resemblance to NeuroD1 expression. NeuroD1 is a transcription factor that can drive neuronal maturation (Gao et al., 2009; Hsieh et al., 2004). In the mouse brain, L1 expression can only be found in neurogenic regions, co-localizing with the expression of NeuroD1 and β-tubulin III (Kuwabara et al., 2009). The colocalization of NeuroD1 and L1 expression

can be explained by their similar mechanisms of expression. In neural stem cells, the Sox2/HDAC1 complex associates with and represses NeuroD1 promoter, much like the complex associated with the L1 promoter. Furthermore, Wnt3a signaling activates NeuroD1 expression via β-catenin (Kuwabara et al., 2009).

2.7 Environmental influence on neuronal retrotransposition

Neuronal circuitry and synaptic activity are directly influenced by experience. As determined in experiments with laboratory rodents, the local environment can affect neuronal size, complexity of dendritic arborization, and the number of synapses (van Praag et al., 2000). Favorable environmental settings will contribute to enhanced cognition, learning, and memory. Surprisingly, the environment can also have an effect on the genomic level by influencing L1 retrotransposition. Mice in enhanced environments with running wheels had 3-fold more L1 retrotransposition than mice in sedentary environments (Muotri et al., 2009). Perhaps, one explanation for the increase in retrotransposition in running mice can be explained by increase in hormone levels. Testosterone was previously shown to upregulate L1 expression in tissue culture, which can lead to increased retrotransposition (Morales et al., 2002). Further investigation of environmental cues that affect L1 expression and retrotransposition in vivo warrants further testing.

2.8 L1 and Rett Syndrome

Rett syndrome (RTT) is a neurodevelopmental disorder caused by mutations in the *MECP2* gene (Amir et al., 1999). Because *MECP2* is on the X-chromosome, RTT typically only affects girls, but a few rare cases of affected males have been reported. Most RTT girls are born without symptoms and develop normally for 6-18

months (Chahrour and Zoghbi, 2007). At this time, development begins to stagnate and eventually regresses. Common attributes of RTT are autism, loss of speech, hand wringing, anxiety, and eventual motor deterioration. The role of MeCP2 in neurons and how *MECP2* mutations contribute to RTT pathology is still under great scrutiny and hotly debated. Nevertheless, several reports suggest one role of MeCP2 is to regulate repetitive elements such as L1 (Muotri et al., 2010; Skene et al., 2010; Yu et al., 2001).

In neural stem cells, MeCP2 was found in association with the CpG-methylated promoter of L1 (Muotri et al., 2010). As mentioned previously, MeCP2 forms a repressive complex with HDAC1, inhibiting L1 expression (Coufal et al., 2009). Not surprisingly, neuroepithelial cells from MeCP2 knockout (KO) mice express 4-fold more L1 RNA than wild-type matched controls and the MeCP2 KO mice also exhibit 3.5-fold increase in neuronal retrotransposition (Muotri et al., 2010). Another function of MeCP2 may be control spurious expression of L1s in mature neurons. One study found mature neurons of MeCP2 KO mice express greater amount of repeat elements, including L1, when compared to age-matched controls (Skene et al., 2010).

Induced pluripotent stem cells from RTT patients were created and differentiated into NPCs to examine if naturally occuring *MECP2* mutations also exhibit greater L1 retrotransposition (Muotri et al., 2010). RTT patient NPCs support 2.5-fold more retrotransposition compared to age matched controls. Furthermore, postmortem brain tissue from RTT patient brains displayed greater L1 DNA copies compared to age-matched control patients, suggesting RTT neurons likely underwent greater retrotransposition.

The impact of L1 overexpression and increased retrotransposition on the pathology of RTT is unknown. It is possible L1 overexpression and retrotransposition are simply byproducts of the loss of MeCP2 function. However, the greater mutagenesis of the neuronal genomes is likely to cause more diversification in "normal" neuronal function and could contribute to the heterogeneity of the disease.

Rescuing the RTT phenotype has come with variable success. Two separate studies have shown proper re-expression of MeCP2 in post-mitotic neurons rescues the life-span and certain physical phenotypes of RTT mouse model (Guy et al., 2007; Luikenhuis et al., 2004). However, the rescued mice were not stringently tested for cognitive defects. It would be interesting to compare the cognitive abilities and variability of cognition rescued RTT mice compared to wild-type mice.

2.9 L1 and Ataxia Telangiectasia

Ataxia Telangiectasia (A-T) is a rare hereditary neurodegenerative disorder caused by mutations in the *Ataxia Telangiectasia Mutated* (*ATM*) gene. As the name of the disease implies, patients experience a loss of motor function and a noticeable dilation of their capillaries. Complications of the disease often result with patients acquiring diabetes, lymphomas, immunodeficiency and paralysis due to cerebellar degeneration. Most patients die in their teens and early twenties, although life expectancy can vary and some live past mid-age. ATM is a serine/threonine protein kinase that senses and responds to DNA damage (Savitsky et al., 1995). ATM can detect DNA double-strand breaks in DNA, starting a signaling cascade by phosphorylating its multiple substrates, including p53. This response activates DNA-damage checkpoint and arrests the cell cycle until the damage is repaired (Bar-Shira

et al., 2002). In cells deficient of ATM function, double-strand breaks in DNA can go unnoticed and DNA mutagenesis increases with cell cycle.

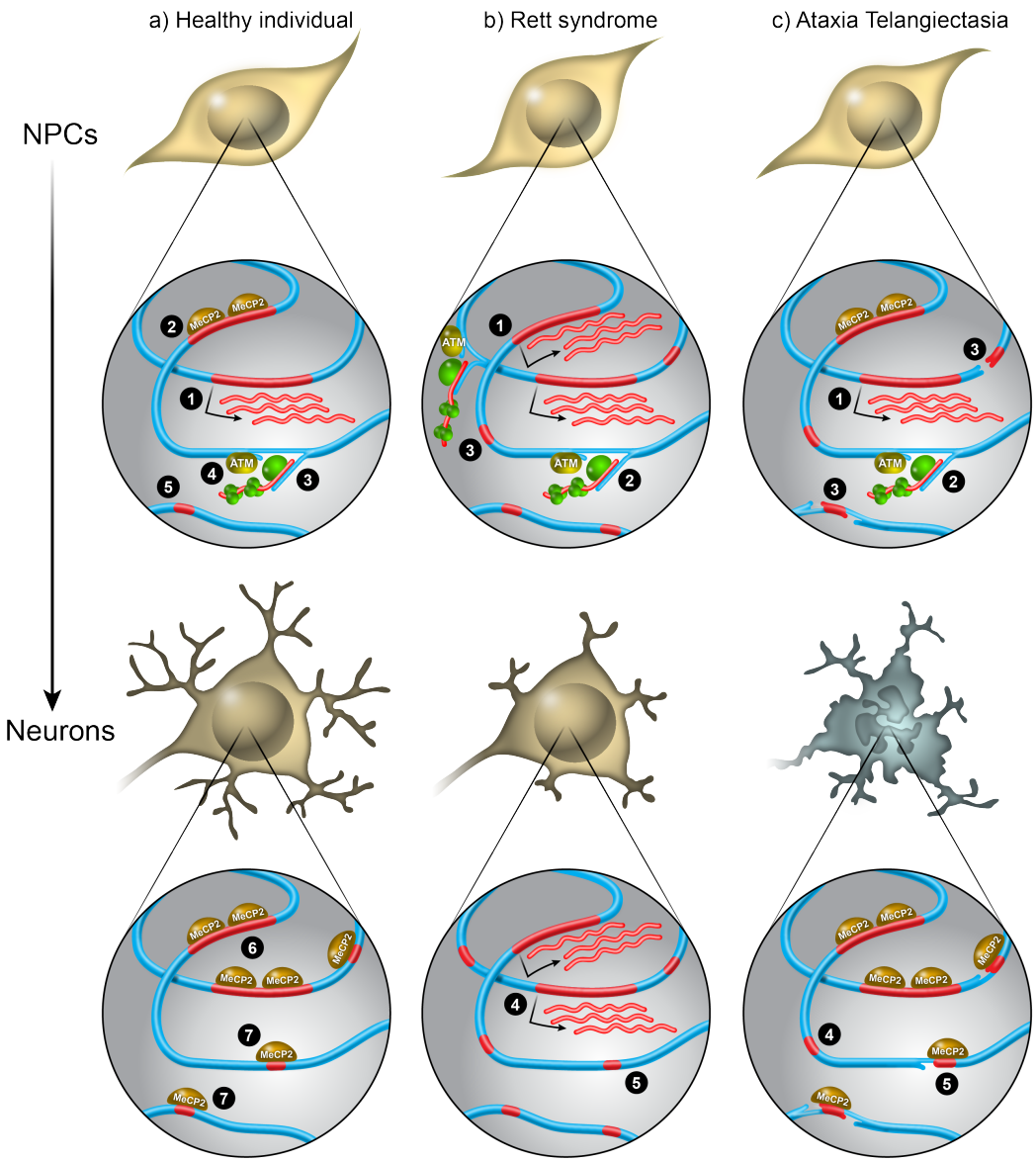
An interesting discovery was the enhanced ability of L1 to retrotranspose in NPCs devoid of ATM function (Coufal et al., 2011). The brains of ATM KO mice displayed significantly greater levels of L1 retrotransposition compared to their sibling wild-type controls. Likewise, when comparing the L1 copy number in postmortem tissue, A-T patients contained significantly more L1 copies than age-matched controls. These data suggest ATM has a role in preventing retrotransposition. This finding was in stark contrast with a previous report, which claimed ATM was required for L1 retrotranspostion (Gasior et al., 2006). The different conclusions are likely due to the L1-retrotransposition cassette used in the two studies. Gasior et al used a neomycin-resistance cassette which requires G418 selection. G418 selection is fairly toxic and could have influenced the retrotransposition results in ATM-deficient cells.

To determine why ATM-deficient cells harbored more L1 insertions, Coufal et al. initially hypothesized that L1 was able to take advantage of the double-strand breaks in DNA to retrotranspose in an endonuclease-independent fashion. However, when the authors realized the loss of ATM function was not enough for endonuclease-independent retrotransposition, they began to explore other possible hypotheses. They discovered ATM deficient neurons had longer L1 insertions, suggesting ATM might participate on the detection of the reverse transcription activity of L1's Orf2p directly and signal to inhibit the insertion.

Using postmortem brain tissues, authors also detected that patients with A-T have greater levels of L1 retrotransposition and longer L1 insertions in their neurons. It is unclear, however, whether the enhanced retrotransposition ability is related to the

neurodegeneration that is common in the pathology of A-T. An emerging hypothesis in the A-T field suggests differentiating neuronal cells lacking ATM are unable to induce apoptosis, and thus ATM may be required to prevent damaged cells to enter into the neuronal circuitry (McConnell et al., 2004; Rass et al., 2007). As DNA damage builds over time, combined with the high oxidative stress in the neuronal environment, neurons cannot handle the damage and degenerate. Considering L1 retrotransposition damages DNA upon integration, high levels of retrotransposition could possibly contribute to cerebellar degeneration displayed in A-T.

Figure 5. Depiction of L1 retrotransposition in the development of healthy, RTT and A-T neurons. a, Typical retrotrasposition in healthy neuronal progenitor cells begins when L1 expression is induced (1). MeCP2 keeps tight regulation of L1 expression(1), preventing dangerous levels of L1 retrotransposition. After the L1 RNP formation occurs in the cytoplasm, the L1 RNP re-enters the nucleus. Once in contact with DNA, the RNPs initiate target-primed reverse transcription (3). ATM may be involved in protecting against L1 integration by initiating a DNA damage response (4). L1 insertions are typically truncated around 1 KB (5). As the neuron matures, MeCP2 forms a repressor complex on the promoter of inherited L1 sequences, preventing spurious expression (6). MeCP2 also binds new de novo L1 insertions and likely regulates their impact (7). All DNA is repaired at new L1 integration sites
b, Retrotransposition in Rett Syndrome neuronal progenitor cells is significantly greater. A lack of functional MeCP2 increases L1 expression is 2.5-4 fold compared to unaffected controls (1). The higher level of L1 expression induces more L1 integration events (2). The L1 insertions are truncated around 1KB, but are supernumerous (3). The DNA of mature RTT neurons have many more L1 integration events and loss of MeCP2 regulation, which permits spurious L1 expression (4). The new L1 insertions are left unregulated (5).
c, Retrotransposition in A-T neuronal progenitor cells is unnoticed, due to a lack of ATM signaling. The L1 expression is comparable to healthy NPCs (1). The lack of functional ATM delays the DNA damage response during L1 integration (2). Dysfunctional DNA repair leaves some nicks in the DNA (3). The DNA of A-T mature neurons display longer L1 integration events compared to unaffected controls(4) The DNA damage persists in maturation, leading to degeneration (5).


a) Healthy individual
b) Rett syndrome
c) Ataxia Telangiectasia
NPCs
Neurons
MeCP2
ATM

2.10 Influence of L1 on mammalian genome evolution

L1s, in combination with SINEs, have been speculated to be major drivers of mammalian evolution. All throughout mammalian history, L1 and SINEs have persistently been actively retrotransposing (Cordaux and Batzer, 2009). No other taxonomic class has such a high correlation with a transposable element. Furthermore, LINEs and SINEs can mediate genome evolution by amplification and creating genomic instability, generating the necessary variation for evolution to act. As mammals evolved, their genomes continually expand due to increased copy number of retroelements (Liu et al., 2003). Even after speciation, the human genome has continued to expand, accumulating approximately 2,000 LINES and 8,000 SINES over the past 6 million years (Ma) (Cordaux and Batzer, 2009). At the same time of genome expansion, the stability of the genome has decreased. As mentioned earlier, L1s and SINEs can always threaten to insert within a gene or regulatory, changing the expression pattern.

2.11 A correlation of L1 families and hominin evolution

The many features specific to great apes and humans developed over the past 15 Ma of evolution. Here we take a speculative view of the correlation between major events in hominin evolution alongside L1 evolution. In the human genome, only five major L1 families have been active and amplified over the past 25 Ma (Smit et al., 1995). These families, sequentially denoted L1PA5-L1PA1, have emerged from a single lineage with successive activity and amplification rates (Khan et al., 2006). In other words, the arrival of activity of a new L1 family was concurrent with the quick demise of activity from the preceding family. Looking at the arrival of some of the families, it is interesting to note the correlation with hominin evolution (Figure

3). The divergence of ancestral apes from Old World monkeys occurs around the origin of L1PA5 (~25 Ma) (Rhesus Macaque Genome et al., 2007). At that time, ancestors of hominids were diverse, but restricted to tropical forests and woodlands of Africa and Arabian Peninsula (Reed, 1997). Within the hominids, the deviations of last common ancestors closely correlate to the emergence of new L1 families. The gorilla/chimp/human ancestor deviated from the orangutan ancestor at the arrival of L1PA3 (~14 Ma). Likewise, the chimp/human ancestor deviated during the arrival of and L1PA2 (~8 Ma) (Goodman et al., 1998; Lee et al., 2007). It was between 17-14 Ma that Africa become drier and seasonal. This environmental change reduced the diversity and the hominids became dominant, most likely by developing a range of specialized locomotors behaviors and dietary adaptations. Beginning around 10 Ma, gradual cooling and drying shifted the African environment, creating more diversified climates and selective pressures, contributing to the divergence of the hominin ancestors (Chaline, 1996).

Perhaps the most interesting correlative comparisons are the arrivals of L1PA1-preTa (~3 Ma) and L1PA1-Ta1 (~2 Ma) (Lee et al., 2007), which closely correspond to the speciation of *Australopithecus africanus* and *Homo ergaster* (Strait et al., 1997), respectively. Anthropologists believe that both of these species made great advances in human cognition, affecting both behavior and intelligence. These were changes likely induced by environment. For instance, *A. africanus* is the first species suggested to be completely bipedal, with an upright posture, living in non-arboreal environment (Harcourt-Smith and Aiello, 2004; Robinson, 1972). Likewise, *H. ergaster* had a major expansion of cranial size, was the first to begin use of complex tools, and the first to control fire (Goren-Inbar et al., 2004; Stout et al., 2008).

Furthermore, *H. ergaster* is believed to be the first to use a primitive language and to form societies of hunters and gatherers (Stout et al., 2008). While these are simply correlations, it is intriguing to question L1's role in hominin evolution.

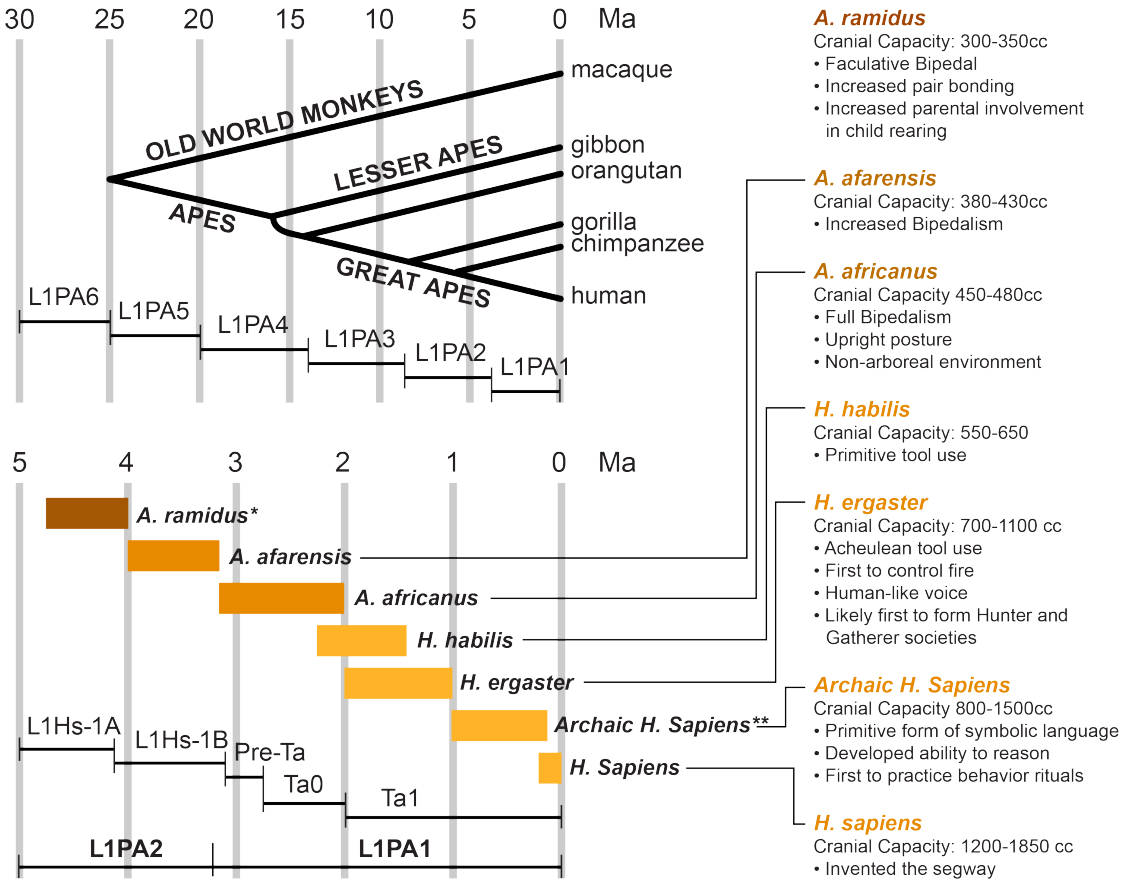


Figure 6. Schematic of the correlation of L1 and hominin evolution. The evolution of L1 is linear with each family subsequently replacing the previous. **a**, Depiction of L1 family and hominid evolution over the past 30 Ma. **b**, Depiction of L1 subfamily evolution and hominid speciation over the last 5 Ma.
*It is debated whether *A. ramidus* is a true ancestor of *H. sapiens*. However, studies on *A. ramidus* will help with understanding what the true human ancestor of that time was like.
***Archaic H. sapiens* include *H. antecessor* and *H. heidelbergensis*.

2.12 Conclusion and perspectives

For many years, L1 retroelements were viewed as a selfish DNA parasite that had no function in somatic cells. The evidence for this hypothesis was based on a

few facts. One, classical L1 retrotransposition assays could only detect L1 insertions in neoplastic cells and not in primary cells. Furthermore, somatic cells have many different mechanisms to inhibit L1 retrotransposition. If the purpose of L1 was to replicate and expand itself, it would only need to retrotranspose in germ cells and very early embryogenesis, as these are the only genomes to transfer to the next generation. Thus, somatic retrotransposition would be of no benefit. So, why does L1 retrotranspose in neurons?

We propose two reasonable hypotheses to explain neuronal retrotransposition. The first hypothesis is simple; neurons are able to permit high levels of retrotransposition because the mutagenesis has negligible effect on neuronal function. The second hypothesis requires consideration of the heterogeneity of the mammalian brain. Because high levels of retrotransposition would mutagenize the genome, the neuronal transcriptome would be slightly altered. Furthermore, each neuron would have its own unique set of integration events and thus each neuron's transcriptome would be unique. The result would be a heterogeneous nervous system composed of individually distinct neurons.

The hypotheses of neuronal retrotransposition are extremely difficult to test when one considers the following experimental issues. The mouse genome is composed of approximately 19% L1 DNA and 3000 L1s are putatively active, thus creating an L1 knockout mouse is not possible. Each post-mitotic neuron has a unique set of retrotransposition events, thus finding a particular insertion and correlating the mutation with a change in transcription is not possible. The mammalian nervous system is extremely complex and plastic, thus assigning a "normal" function to an individual neuron is arbitrary. Although these facts make it

extremely difficult to test the above hypotheses, we will provide evidence and argue for the “heterogeneity hypothesis.”

If one were to defend the “negligible effect hypothesis,” one could argue mamalian genomes are nearly 20% L1 sequence and neuronal genes have L1 sequences throughout introns; therefore, a few more L1 sequences would have a negligible effect. However, L1 insertions within neuronal genes have given rise to several spontaneous mouse mutants, including the spastic mouse and Orleans reeler mouse (Mulhardt et al., 1994; Takahara et al., 1996). The spastic mouse arose from an L1 integration event into intron 5 of the Glycine Receptor β subunit gene. Likewise, the Orleans reeler mouse arose from and L1 insertion into the 8th exon of the reeler gene. Both of these mutations induce exon skipping, changing the splicing pattern. These mouse mutants are examples of how L1 mutations can have a large impact on neuronal function.

One interesting aspect of the mammalian nervous system is the magnitude of programmed cell death (PCD) in the brain. In the mouse cortex, it is estimated that 50-70% of cells undergo PCD (Blaschke et al., 1996). Although the reason for the high level of cell death is unknown, several hypotheses have been suggested. The three predominant hypotheses are the following: $^{1)}$ to regulate the progenitor population, $^{2)}$ to remove of neurons with errors, and $^{3)}$ to ensure efferent and afferent systems are properly connected (Buss et al., 2006; Kim and Sun, 2011). Neuronal retrotransposition could partially explain the genetic basis of these hypotheses.

Inhibition of neuronal PCD results in an enlarged telencephalon size, supernumerary neurons, and embryonic lethality (Hakem et al., 1998). Thus, PCD functions to control brain size. Furthermore, another possible function of PCD is to

remove neurons that migrated incorrectly (Heckroth et al., 1989; Jung et al., 2008). From these data, we can conclude PCD is necessary to prevent overgrowth and ectopic brain development, but the genetic basis of determining which cells are selected for PCD is still unclear. We propose neurons with detrimental L1 mutagenesis are negatively selected and neurons with beneficial L1 insertions are positively selected.

The positive and negative selection of neurons can be further explained at the time of connectivity and synaptogenesis. Several reports have confirmed target-guidance molecules are neurotrophic factors (Lotto et al., 2001; Vogel et al., 1989; Von Bartheld and Johnson, 2001). Therefore, proper efferent and afferent connections keep neurons alive, whereas improper or lack of connection leads to neuronal PCD. Could L1 retrotransposition affect neuronal projection and synaptic connection? Considering L1 most commonly inserts in the genes of synaptic proteins, cell adhesion, and cell projection, we hypothesize L1 can alter a neuron's axonal pathway and synaptic connections. Due to the random nature of L1 integration, we predict the connectome of mammals would be different between individuals. Interestingly, Lu et al. recently published the connectome of the interscutularis neuromuscular junction of the mouse, the most complete mammalian connectome yet to be described (Lu et al., 2009). The authors discovered that no two connectomes were the same between genetically identical siblings. Furthermore, the connectome between the left and right muscles of individual animals varied significantly.

The benefit of neuronal heterogeneity and unique connectivity is still unknown. However, the advantage of neuronal variability may be easier to define in the context

of evolution. As external environments kept changing over the course of time, adaptability was necessary for species survival. Variability in neurons, networks, and behaviors could have generated outliers within a mammalian species. Some of those outliers could have adapted to hostile environmental challenges, favoring the survival of the species.

The remarkable number of somatic L1 insertions estimated in the mammalian brain indicates that L1 retrotransposition may have the capacity to generate. This idea is still consistent with the view suggesting that L1s are selfish genes, caring for their own survival. Somatic insertions represent a novel strategy, remarkably distinct from mobility in germ cells. In the nervous system, L1s were able to escape strong cellular silencing strategies, such as DNA methylation, RNA editing and siRNAs. Parasites that manipulate host behavior by changing neuronal networks can be extremely successful in survival, as exemplified by the protozoan *Toxoplasma gondi*. *T. gondi* can change rat behavior, inducing a fearless reaction to cat odors (Vyas et al., 2007a; Vyas et al., 2007b). High levels of the protozoan can be found in the rat, especially in the amygdala, but *T. gondi* can only sexually reproduce in the cat gut. Thus, it is hypothesized that *T. gondi* manipulates the rat amygdala to transfer from one cat to another to enhance the species' fitness. It has also been postulated that *Toxoplasma* may also induce personality changes in humans (Holliman, 1997).

While L1 retrotransposition may be beneficial to mammalian behavior, it is detrimental when misregulated. Recent studies have shown misregulation of L1 is an aspect in Rett syndrome and Ataxia Telangiectasia and may contributes to their pathologies (Coufal et al., 2011; Muotri et al., 2010). As of yet, only two neurological disorders have been shown to involve L1, however, it is likely that many others are

impacted by L1 retrotransposition. In some disorders, unregulated L1 retrotransposition could affect the pathology directly. In others, the neuronal heterogeneity created by L1 retrotransposition will influence the course, spectrum, and treatment.

The impact of L1 retrotransposition on neural tissue is only now being recognized. It is apparent that neuronal retrotransposition is not a rare, aberrant event, but a natural part of development. As with any biological phenomena, misregulation of retrotransposition can have detrimental effects and possibly contributes to neuropathological diseases. With the neuronal retrotransposition field still in its infancy, many future directions are ripe for discovery. One of the biggest questions in the field is what is the function of retrotransposition in neurons? To answer that question, we must develop ways to inhibit neuronal retrotransposition. Another important issue is to determine which neurological diseases involve retrotransposition and how it affects the pathology and or treatment.

Chapter 2, in whole, is a reprint of the material as it appears in Annual Review of Cell and Developmental Biology, 2012. Thomas, C. A.; Paquola, A. C. M.; and Muotri, A. R. The dissertation author was the primary investigator and author of this paper.

CHAPTER 3. INHIBITING MOUSE LINE-1 WITH ZINC-FINGER REPRESSORS

3.1 Introduction

L1 retrotransposes in the human brain approximately 80 times in each neuron (Coufal 2010). In MeCP2 deficient neurons, we can detect 3.5-fold greater rate of retrotransposition. However, we still do not understand the consequences of the retrotransposition in a normal or disordered state.

We have formulated 2 distinct hypotheses to the role L1 retrotransposition plays in neural development: 1) L1 retrotransposition has no consequence on neuronal gene expression and thus retrotransposition has negligible effects on neuronal activity and behavior. Or, 2) L1 retrotransposition can alter gene expression in neurons and thus create a variety of slightly variant neurons. The variety of expression between neurons would create a heterogeneous population of neurons that would function differently to different responses. Furthermore, in a disordered state such as RTT where retrotransposition levels are much greater, we hypothesize the increased variability may create a chaotic environment causing the neuronal function to shut down.

To ascertain the role L1 retrotransposition in the nervous system, we sought to create a system to inhibit L1 expression and retrotransposition. Our ultimate goal was to create a genetic mouse model with limited L1 retrotransposition. We choose to use zinc-finger repressors to modulate L1 gene expression for the following reasons: 1) we could create a transgenic mouse overexpressing the zinc finger; 2) zinc fingers were the best DNA binding proteins at the time we devised the project, before the advent of TALEs and CRISPRs; 3) we could easily design a zinc finger to bind to the

murine L1 promoter; and 4) we could easily assess the function of the zinc finger in *in vitro* assays before creating transgenic mice.

3.2 Zinc-finger repression of L1

To inhibit murine L1 expression, we created a ZFR to bind to the T_F family 5' UTR (**Fig. 4**). The ZFR consists of DNA binding zinc-finger domain, transcription inhibiting KRAB domain, and a cMyc tag. The ZFR was also tagged with a 2A-mCherry to ensure robust expression of the vector. Because of the differences of 5'UTR sequence of the T_F with the G_F and A families, the ZFR was expected to only bind and inhibit T_F family transcription. As a control, we also created a mutant L1-ZFR with alanine substitutions in the DNA binding domain.

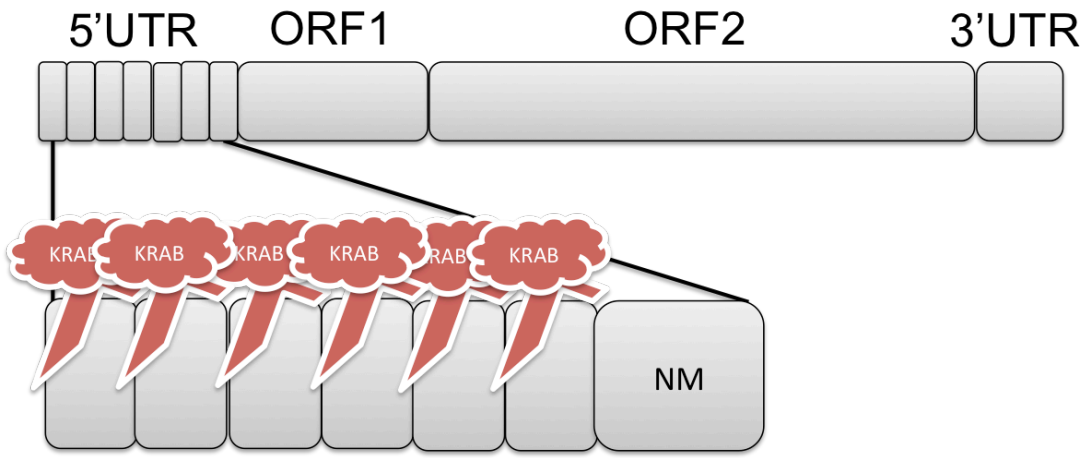


Figure 7. **Schematic of Mouse L1 and ZFR.** The mouse L1 is composed of 4 major regions: the 5'UTR with promoter activity, ORF1 that encodes the RNA chaperone protein, ORF2 that encodes the reverse transcriptase, and 3'UTR with poly adenylation signal. The ZFR binds to the "monomer" repeat of the 5'UTR to promote chromosome condensation and inhibit transcription.

We tested the ability of the L1-ZFR to inhibit the L1 T_F promoter with a plasmid expressing luciferase driven by the T_F promoter. We expressed the luciferase and L1ZFR constructs at equal molarity in 293T cells. We collected the lysates of the cells

at 72 hours and measured luciferase intensity. The L1-ZFR was able to inhibit luciferase significantly, reducing the expression 14 fold compared to the mutant L1-ZFR (**Fig. 5a**). We also tested the ability of the L1-ZFR to inhibit L1 retrotransposition. We expressed the L1 retrotransposition cassette and the L1-ZFR at equal molarity in 293T cells. The L1 retrotransposition cassette contains spastic mouse L1, which is driven by the T_F promoter. In the 3'UTR, there is an eGFP cassette, split by the β-globin intron. The eGFP cassette only expresses after the L1 construct undergoes proper retrotransposition. We calculated the percentage of eGFP positive cells, relative to mCherry positive cells. The L1-ZFR significantly inhibited $L1_{spa}$ retrotransposition cassette, compared to the mutant L1-ZFR, reducing the retrotransposition 3-fold (**Fig. 5b**). Together, we conclude the L1-ZFR can inhibit T_F expression and retrotransposition in an exogenous setting.

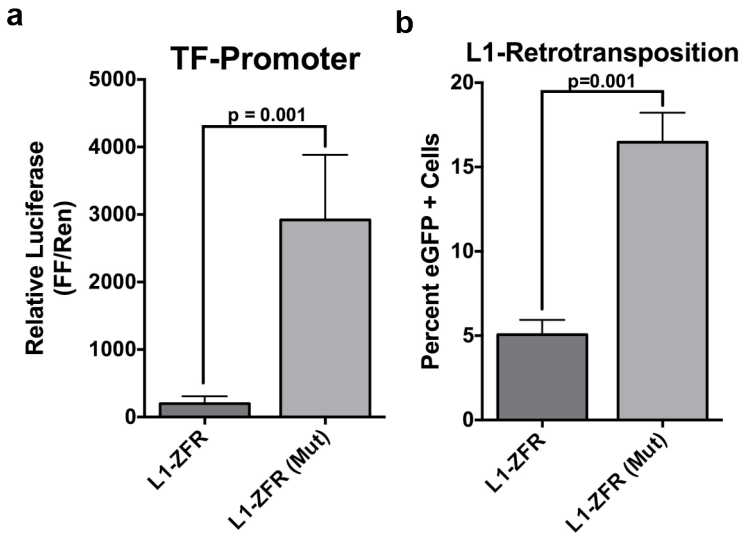


Figure 8. The L1-ZFR inhibits L1 expression and retrotransposition exogenously. a, Graphical representation of luciferase assay. Values are averages of firefly (FF) luciferase normalized to Renilla (Ren) luciferase. Student's t-test was performed to determine significance (n=4). **b**, Graphical representation of L1 retrotransposition assay. Values are averages of percent of eGFP cells as determined by FACs, normalized to mCherry expression. Student's t-test was performed to determine significance (n=3).

Our ultimate goal was to create an *in vivo* model in which we could significantly inhibit mouse L1 expression and retrotransposition. Thus, we next sought to determine if the L1-ZFR could inhibit L1 expression in a murine cell line with high L1 expression. To test the ability of the L1-ZFR to inhibit endogenous expression, we transfected the L1-ZFR in the mouse embryonal carcinoma cell line, F9. We confirmed high expression of the L1-ZFR by western blot (**Fig. 6a**). We then used an optimized qPCR assay to look at the expression of each murine L1-family. Unfortunately, we did not see a robust decrease of L1 T_F expression (**Fig. 6b**). This result was confirmed with a western blot against the L1 Orf1p (**Fig. 6a**).

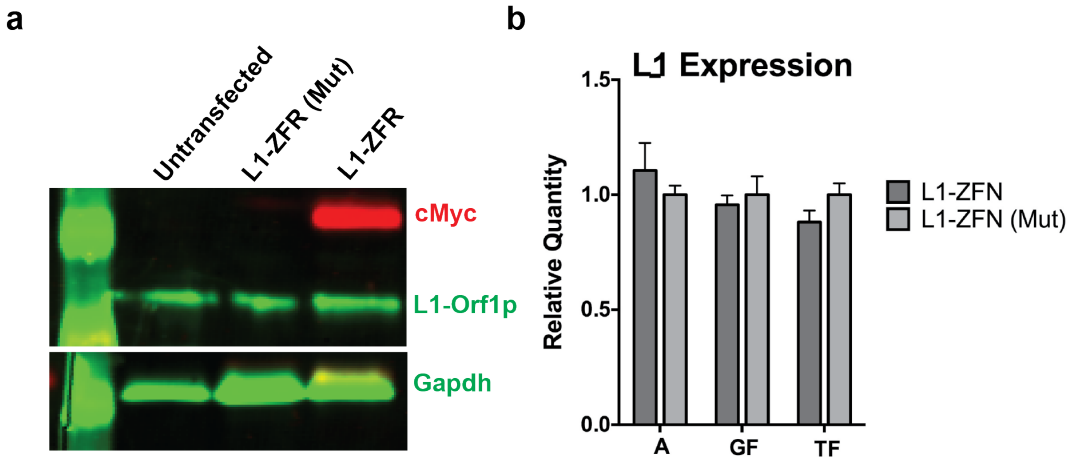


Figure 9. **L1-ZFR does not robustly inhibit L1 expression in F9 embryonal carcinoma cells. a**, Western blot of F9 cells 72 hours post transfection with L1-ZFR. cMyc tag represents L1-ZFR. **b**, qPCR expression of L1 families in F9 cells 72 hours post transfection with L1-ZFR.

3.3 Discussion

The ultimate goal of the project was to create an *in vivo* model to assess the significance of L1 expression and retrotransposition in the murine brain. As shown in previous publications, a robust level of L1 expression and retrotransposition is present in mouse neurons (Muotri et al., 2005; Muotri et al., 2010). We hypothesize

the retrotransposition observed in the neurons of the mouse brain could create a heterogeneity of behavior, due to a heterogeneity of genetic material. Unfortunately, the ZFR approach did not work in on endogenous L1 expression and thus would not work to create an *in vivo* model.

The ZFR failed for a multitude of reasons. One reason was the extreme difficulty to obtain high level of expression in mouse cells. We codon optimized the ZFR for mouse expression and tried an assortment of plasmid vectors with different expression cassettes. We settled on the pMAX vector from AMAXA, which contains the minimum CMV promoter and a small intron. The advantage of the pMAX vector was the size (2.9 kb), and thus we could transfect an extremely high amount of copies into the F9 cells. When we transfected 30ug of plasmid into the F9s using the AMAXA technology. Only at this level, could we detect robust expression of the cMyc tag. Perhaps, using the TALE or CRISPR technology would improve expression.

Another reason the ZFR failed was the ZFR was only bound to the T_F family. Because the ZFR only bound to about 50% of the mouse L1, we could not analyze the western blot data with confidence. If the ZFR was working on the T_F family, it is possible we could not see a reduction in Orf1p of due to expression of the G_F and A families. To overcome this issue, I would suggest designing a DNA binding domain that recognizes some sequence of ORF1, which is highly conserved between all three families. By doing so, one could create a DNA-binding protein that binds over 95% of full length L1.

A third possible reason the ZFR failed was the transient nature in which we performed all the experiments. All experiments performed on the F9 cells had harvest times of 72 hours, chosen because the highest level of ZFR expression was at 48-72

hours. However, it is possible that to see robust L1 expression reduction, we would need to ensure expression of the ZFR over a long period of time. Perhaps the use of a different vector with a selectable marker such as hygromycin, one could create an F9 line with chronic ZFR expression.

If I was to attempt at this experiment again, I suggest the following ideas to ensure a better outcome. First, I would suggest using the TALE or CRISPR systems to create the repressor proteins. The TALE and CRISPR systems are proving to be much easier and less expensive to create and superior in performance. I would also suggest designing the DNA-binding domain against a conserved region of ORF1. This will ensure binding to over 95% of L1s based on sequence. Lastly, I would suggest expressing the ZFR from the pCEP4 (Life Technologies) or similar plasmid that contains a selectable marker like hygromycin. One could conceivably create a high expressing ZFR mouse cell line, which would definitively answer if the ZFR could repress endogenous L1 elements.

In conclusion, we created a ZFR that was specific to the L1 T_F family of mouse. The ZFR was able to robustly inhibit the L1 T_F promoter expression and retrotransposition in exogenous assays. However, the ZFR did not show ability to inhibit L1 expression from the mouse genome. Perhaps, a better design could yield the desired results on the endogenous L1.

3.4 Materials and methods

Constructs and plasmid construction

The ZFR was codon optimized and cloned into pMAX vector with a 2A-mCherry. The T_F promoter was cloned from the $L1_{spa}$ sequence and cloned into pGL3 luciferase vector (Promega).

Luciferase assay

The ZFR or mutant expression vector was co-transfected with the pGL3-T_F-Luciferase and Renilla luciferase vector in 293T cells using PEI, 4ug/ 1ug of DNA. 4ug of total DNA was used, in equal molarities. Lysates were collected and luciferase levels using Dual-luciferase kit (Promega).

Retrotransposition assay

The ZFR or mutant expression vector was co-transfected with the $L1_{spa}$-eGFP retrotransposition cassette into 293Ts using PEI, 4ug/ 1ug of DNA. 4ug of total DNA was used, in equal molarities. After 96 hours, the cells were collected and eGFP cell percentage was determined by FACs.

Western blot

F9 embryonal carcinoma cells were transfected with 30 ug of the ZFR or mutant expression vector using AMAXA program C-30 (Lonza). Protein lysates were collected at 72 hours. cMyc, Orf1p and Gapdh antibodies were used at the recommended dilution. Odyssey secondary bodies were used 1:10,000. Westerns were imaged on Odyssey CfX machine.

qPCR

F9 embryonal carcinoma cells were transfected 30 ug of the ZFR or mutant expression vector using AMAXA program C-30 (Lonza). RNA was collected at 72

hours using the RNeasy Plus kit (Qiagen), with DNase digestion. RNA was reverse transcribed into cDNA with gene-specific primers using the Quantitech kit (Qiagen). Primers for the T_F, G_F, and A families were created using the monomer sequence of the 5'UTR

CHAPTER 4. CLEARANCE OF ENDOGENOUS L1 RETROELEMENTS IN THE CYTOSOL BY TREX1 PREVENTS NEURONAL TOXICITY

4.1 Introduction

Aicardi Goutieres Syndrome arises from mutations in proteins that regulate the level of exogenous nucleic acid species. One of these proteins is the three-prime repair exonuclease I. The function of TREX1 is to remove single-stranded and double-stranded DNA species from the cytosol. This role is especially important when unwanted virus infect host cells. Interestingly, mutations in TREX1 lead to an autoimmflammatory disease in the nervous system and skin when no exogenous viruses are present. Here, we examine human neural cells to try and understand how mutations in a ubiquitously expressed exonuclease can give rise to autoinfllammation.

4.2 Generation of pluripotent cells devoid of TREX1 function

To understand the role of TREX1 function in human neural cells we created three human pluripotent stem cell lines with distinct *TREX1* mutations (**Fig. 7a**). In two of the cell lines, we mutagenized H9 human embryonic stem cells (hESCs) using the CRISPR/Cas9 genome-editing system, generating co-isogenic cell line pairs (**Fig 7b**)(Cong et al., 2013; Mali et al., 2013). We dubbed these co-isogenic lines V63fs and E83fs. In addition to the mutant lines, we expanded two other lines that went through the mutagenesis procedure but had no mutations in *TREX1* to use as controls alongside the H9 hESC line. We dubbed these lines WT63 and WT83. We created a third TREX1-deficient cell line by inducing pluripotency from fibroblasts of one AGS patient with the stereotypical V201D mutation in the homozygous state (**Fig.**

7b) (Crow et al., 2006a). In summary, we have three mutant lines, V63fs, E83fs, and V201D and three control lines, H9, WT63, and WT83.

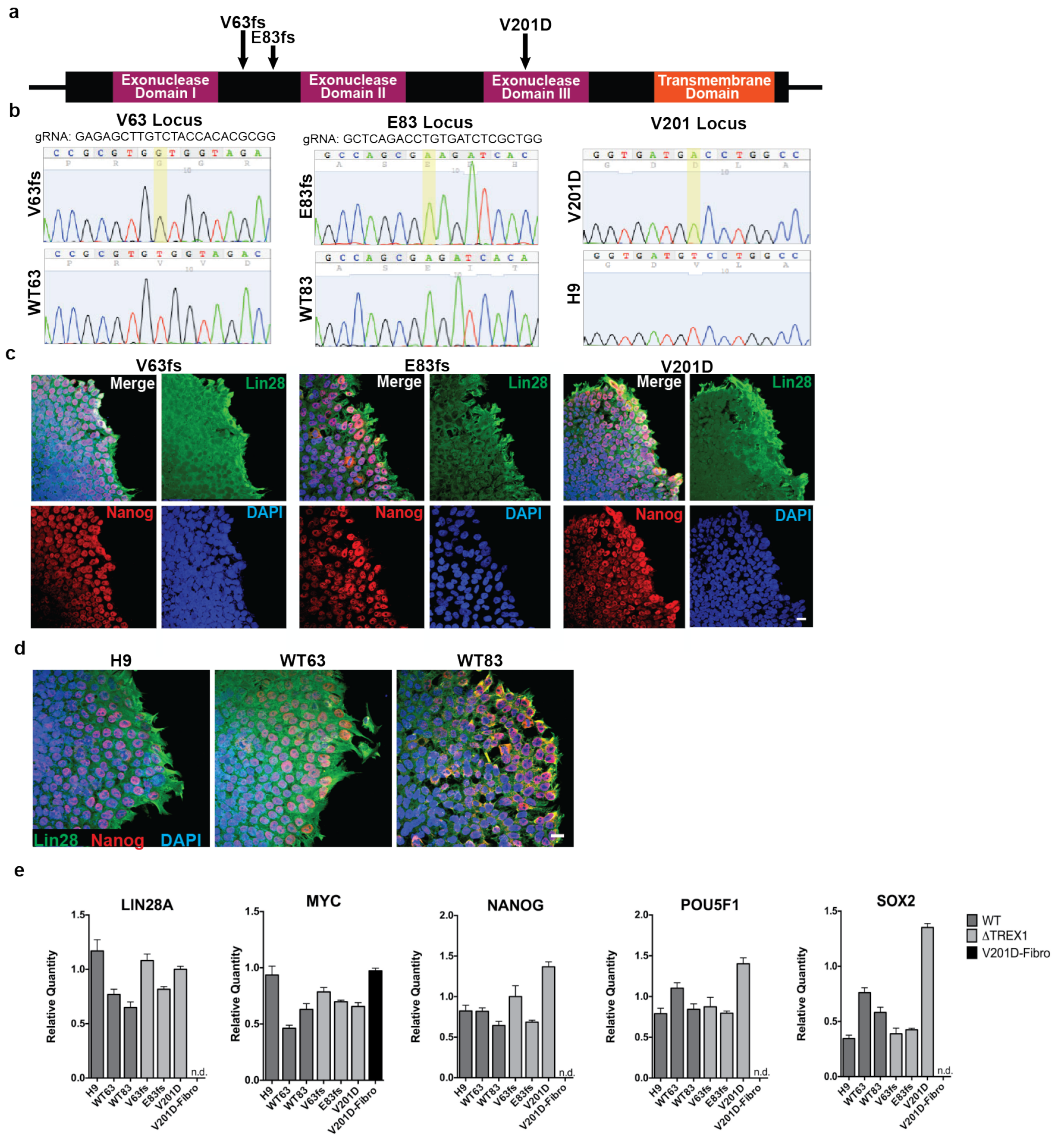


Figure 10. Creation of human pluripotent lines devoid of TREX1 function. **a**, Schematic representation of the *TREX1* gene with mutations of derived pluripotent lines. **b**, DNA sequence chromatogram displaying the nucleic acid change in *TREX1* sequence in mutant lines. Golden box denotes nucleic acid mutation. Amino acid sequence is denoted in white ribbon underneath nucleic acid sequence. **c-d**, Immunofluorescence of mutant TREX1 cell lines, displaying the pluripotent markers Lin28 and Nanog. **e**, Expression of pluripotent markers determined by qPCR. Expression was normalized to zero and HPRT1 was used as an internal reference.

Both the V63fs and E83fs lines arose with a homozygous single nucleotide insertion in the *TREX1* gene, creating frame-shift mutations. The V63fs line incorporated an additional guanine in the codon of amino acid 63, valine (**Fig. 7b**). The E83fs line incorporated an additional adenosine in the codon of amino acid 83, glutamate (**Fig. 7b**). The frame-shift in each of the V63fs and E83fs lines create an early stop codon at amino acid 100, rendering the TREX1 protein nonfunctional. Because there is only one coding exon in the *TREX1* gene, the RNA does not undergo nonsense-mediated decay and maintains high RNA expression (**Fig. 8a**) (Zhang et al., 1998). Although we could not determine protein expression due to a lack of specificity of available antibodies, we did confirm the mutation was stably present at the RNA level.

We performed exome sequencing on the genomic DNA of the V63fs, E83fs, WT63, WT83, and parental H9 hESCs. Using the exome data, we first confirmed the mutagenesis of the V63fs and E83fs lines. All of the reads of the exome data across the mutated locus of both V63fs and E83fs indicated that the single nucleotide insertion was present (**Fig. 8b**). With over 100 reads on each locus, we concluded that the V63fs and E83fs lines contain a pure population of cells with a homozygous frame-shift mutation.

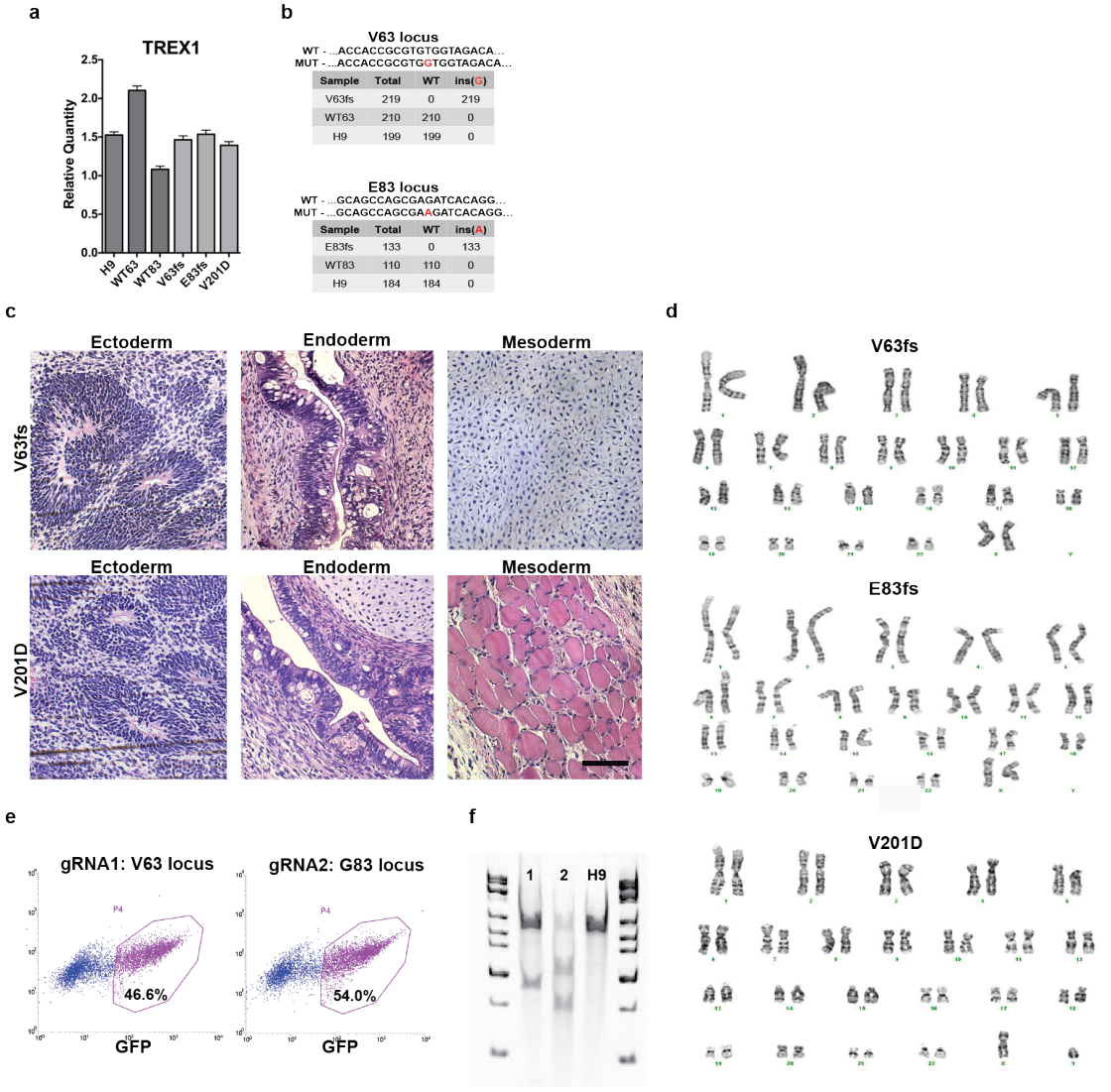


Figure 11. Characterization of human pluripotent cells devoid of TREX1 function. **a**, Expression of TREX1 in neural precursor cells determined by qPCR. Expression was normalized to zero and both B2M and HPRT1 were used as dual internal references. **b**, Summary of exome sequencing across the V63 and E83 locus. Inserted nucleotide causing frame-shift is marked in red. Numbers indicate the amount of reads across the locus. **c**, Hematoxylin and eosin stains on teratomas. All three germ layers are present. **d**, Karyotypes of the TREX1-deficient cell lines. **e**, Dot plots of eGFP expression of the Cas9-2A-eGFP vector used in conjunction with the guide RNA (gRNA) to mutagenize the *TREX1* locus in H9 ESCs. **f**, Surveyor nuclease activity of genomic DNA extracted from H9 ESCs post FACs sort. The PCR product was 538 bp long. The cleavage site of gRNA1 and gRNA2 were 266 bp and 327 bp from the 5' end of PCR product, respectively.

We also used the exome data to check off-targets predicted from the CRISPR design website(Hsu et al., 2013; Hsu, 2013). None of the predicted off-targets were mutagenized in any of the clonally derived lines (**Supplementary Table 2**). In order to find additional potential off-target lesions, we took an unbiased approach to identify all insertions and deletions (InDels) in the exome of each of these lines. Comparing the V63fs, E83fs, WT63, and WT83 exomes against the parental H9 exome, we found between 20 and 35 homozygous InDels in each line that would lead to a frame-shift mutation, including at the expected position in the *TREX1* locus in the V63fs and E83fs lines (**Supplementary Table 3**). We examined the sequence of each InDel and found that none of the sequences of these homozygous InDels matched sequence to the guide RNA. Therefore, we concluded that each nonspecific InDel already existed in subpopulations in the parental H9 hESCs and arose during clonal expansion. Our off-target analysis data are consistent with previous findings (Smith et al., 2014; Veres and Talkowski, 2014).

We performed immunofluorescence on Lin28 and Nanog and found all mutagenized and control pluripotent lines generated here expressed the pluripotent genes (**Fig. 7c, d**). Furthermore, we found all the pluripotent lines *LIN28*, *NANOG*, *MYC*, *POU5F1*, and *SOX2*, as determined by qPCR (**Fig. 7e**). Furthermore, when injected into nude mice to form teratomas, each pluripotent line was able to generate the three germ layers (**Fig. 8c**). The karyotype of each cell line was normal (**Fig. 8d**). Therefore, we concluded that the mutagenized and control pluripotent lines were completely satisfactory to differentiate into neural cells for experimentation.

4.3 Differentiation of pluripotent cells into neural cells

Next, we differentiated the pluripotent cells into neural precursor cells (NPCs) (**Fig. 9a**). To differentiate the pluripotent cells into NPCs, we lifted the pluripotent cells as colonies to form embriod bodies (EBs). Using neuronal media, we grew the EBs in suspension on a shaker. At 6 days, we plated the EBs onto matrigel to select for rosettes. Rosettes were selected, dissociated into NPCs, and expanded. As determined by immunofluorescence, all lines robustly expressed the NPC markers Nestin and SOX2 (**Fig. 9b, c**). Furthermore, we performed qPCR and determined that all the NPC lines expressed *PAX6*, *SOX1*, and *Musashi1* (**Fig. 9d**). We expanded and differentiated these NPCs into neurons by removing bFGF from the neuronal media. (**Fig. 10a**). To purify neurons, we fluorescently sorted 3-week old differentiated neuronal cultures, collecting the $CD184^-$, $CD44^-$ and $CD24^+$ cells (**Fig. 10b**). We performed immunofluorescence and found sorted neurons exhibited expression of the neuronal markers MAP2, Synapsin 1 (**Fig. 10c,d**). Furthermore, we detected expression of *DLG4*, *RBFOX3*, and *TUBB3*, as determined and qPCR (**Fig. 10e**). We differentiated the NPCs into astrocytes via neurospheres in suspension with astrocyte media (**Fig. 11a**). To purify the astrocytes, we labeled the NPCs with a lentivirus expressing the *tdTomato* reporter gene under the activity of the astrocyte GFAP promoter before beginning differentiation. We sorted the astrocytes by fluorescence (**Fig. 11b**). By immunofluorescence, we determined the resulting astrocytes expressed markers such as GFAP, S100 (**Fig. 11c,d**). Furthermore, we determined the astrocytes also chondroitin sulfate, and vimentin by qPCR (**Fig. 11e**).

Figure 12. Differentiation of pluripotent cell lines into neural precursor cells. a, Phase contrast images overviewing NPC differentiation protocol. Pluripotent stem cells (PSCs) were grown as colonies until large, and lifted to form embryoid bodies (EBs). Embryoid bodies were grown in suspension on a shaker until plated, with rosettes forming. Rosettes were picked and dissociated to form neural precursor cells (NPCs). Scale bar in μm, as indicated within. **b**,**c**, Representative immunofluorescence images of NPC markers, Nestin and SOX2. Scale bar, 20 μm. **d**, Expression of NPC markers determined by qPCR. Expression was normalized to zero and both B2M and HPRT1 were used as dual internal references.

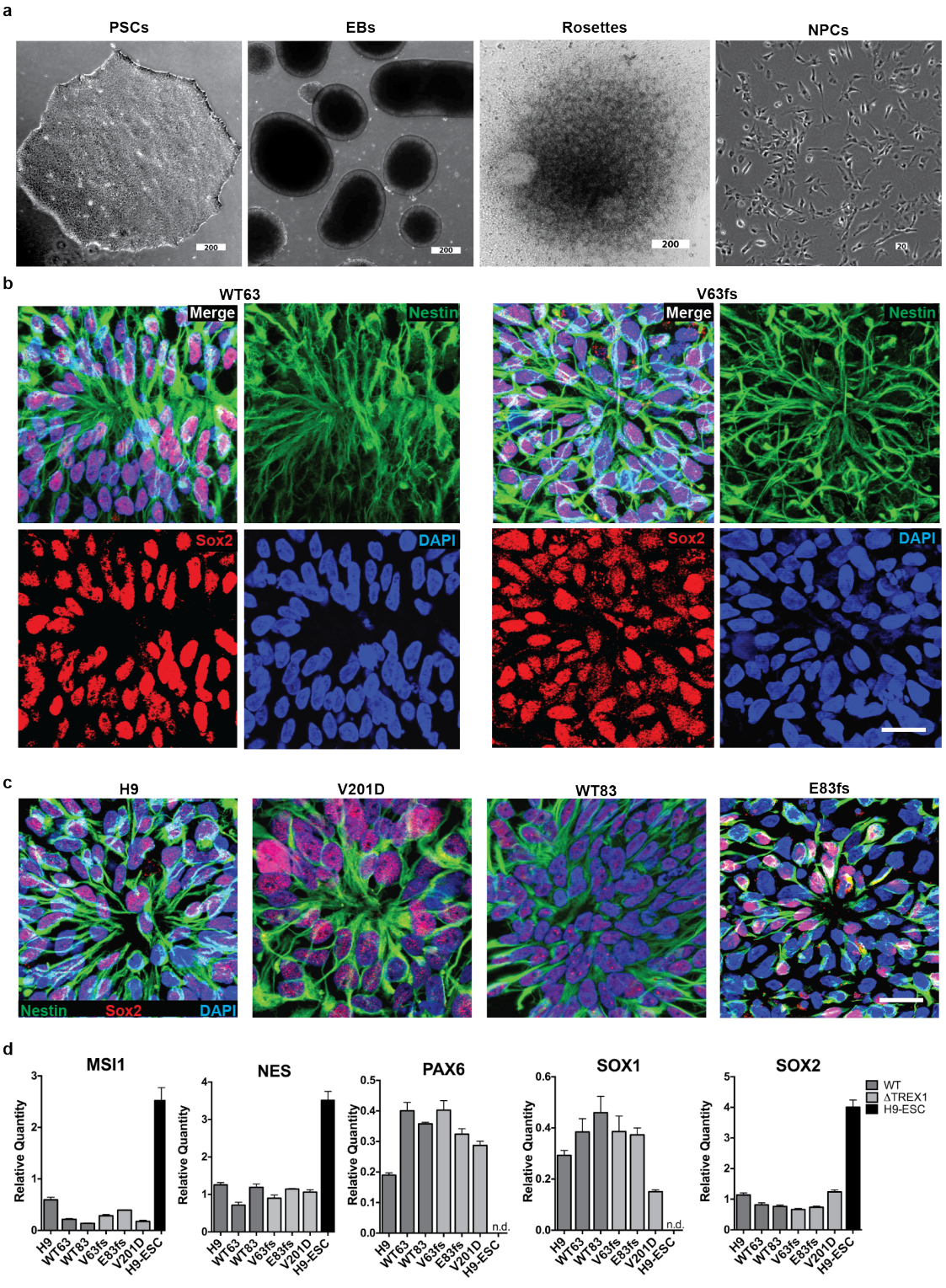

a
PSCs
EBs
Rosettes
NPCs
200
200
200
200
b
WT63
V63fs
Merge
Nestin
Merge
Nestin
Sox2
DAPI
Sox2
DAPI
c
H9
V201D
WT83
E83fs
Nestin
Sox2
DAPI
d
MSI1
NES
PAX6
SOX1
SOX2
Relative Quantity
WT
ΔTREX1
H9-ESC
n.d.
n.d.
H9
WT63
WT83
V63fs
E83fs
V201D
H9-ESC

Figure 13. Differentiation and purification of NPCs into neurons. a, Phase contrast images overviewing neuronal differentiation. In neuronal differentiation, bFGF was removed from the media of NPCs to form neurons. After 21 days, neurons were purified by FACs. **b**, Dot plot of FACs purification of neurons. Neurons were collected by selecting for $CD184^{-}$ $CD44^{-}$ $CD24^{+}$ cells. **c**, **d**, Representative immunofluorescence images of neuronal markers, MAP2 and Synapsin. Scale bar, 20 μm. **e**, Expression of neuronal markers determined by qPCR. Expression was normalized to zero and both B2M and HPRT1 were used as dual internal references. **g**, Representative immunofluorescence images of astrocytic markers, GFAP and S100. Scale bar, 20 μm. **h**, Expression of astrocytic markers determined by qPCR. Expression was normalized to zero and both *B2M* and *HPRT1* were used as dual internal references.

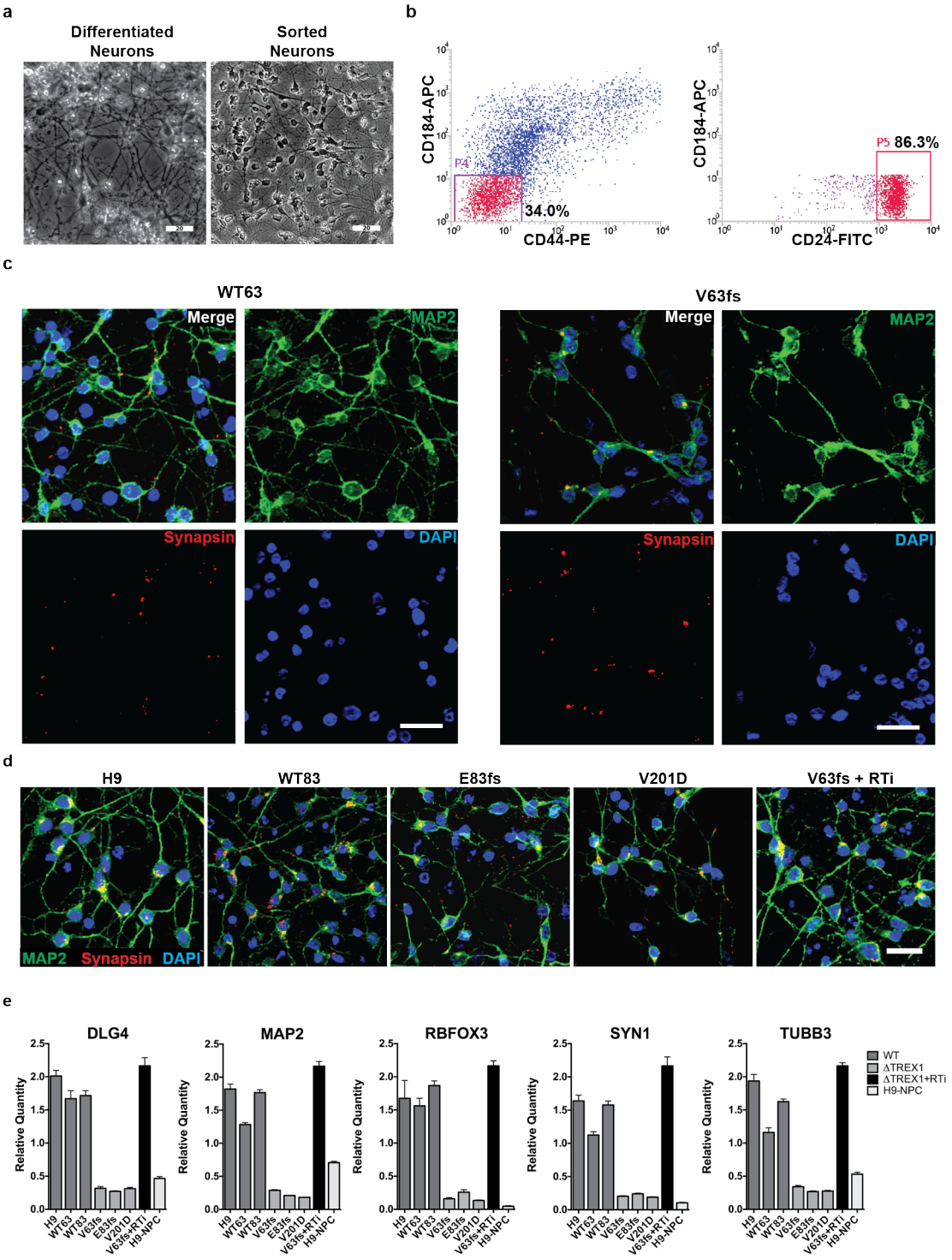
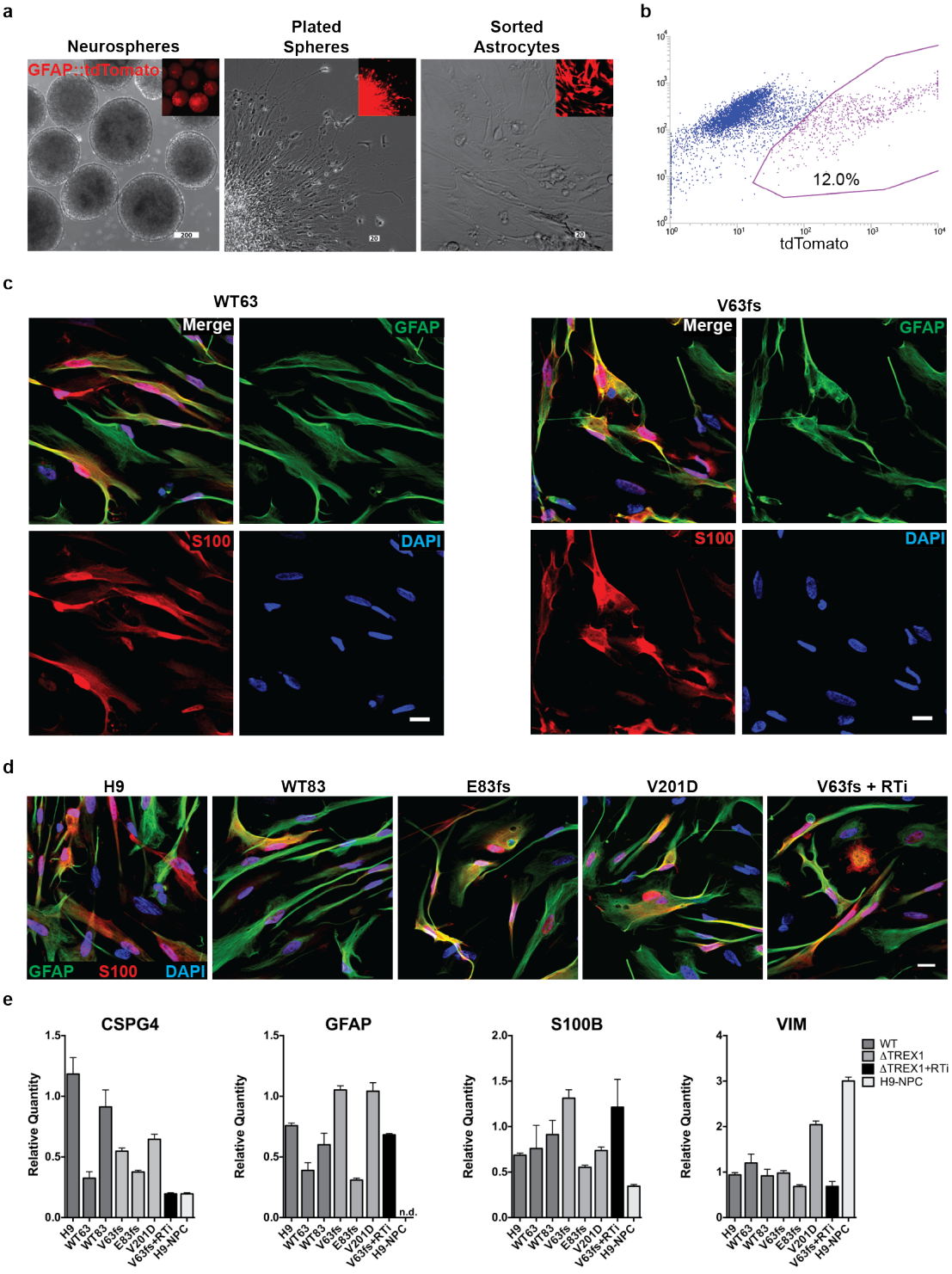

a
Differentiated Neurons
Sorted Neurons
b
CD184-APC
CD44-PE
P4
34.0%
CD24-FITC
P5 86.3%
c
WT63
V63fs
Merge
MAP2
Synapsin
DAPI
d
H9
WT83
E83fs
V201D
V63fs + RTi
MAP2 Synapsin DAPI
e
DLG4
MAP2
RBFOX3
SYN1
TUBB3
Relative Quantity
WT
ΔTREX1
ΔTREX1+RTi
H9-NPC

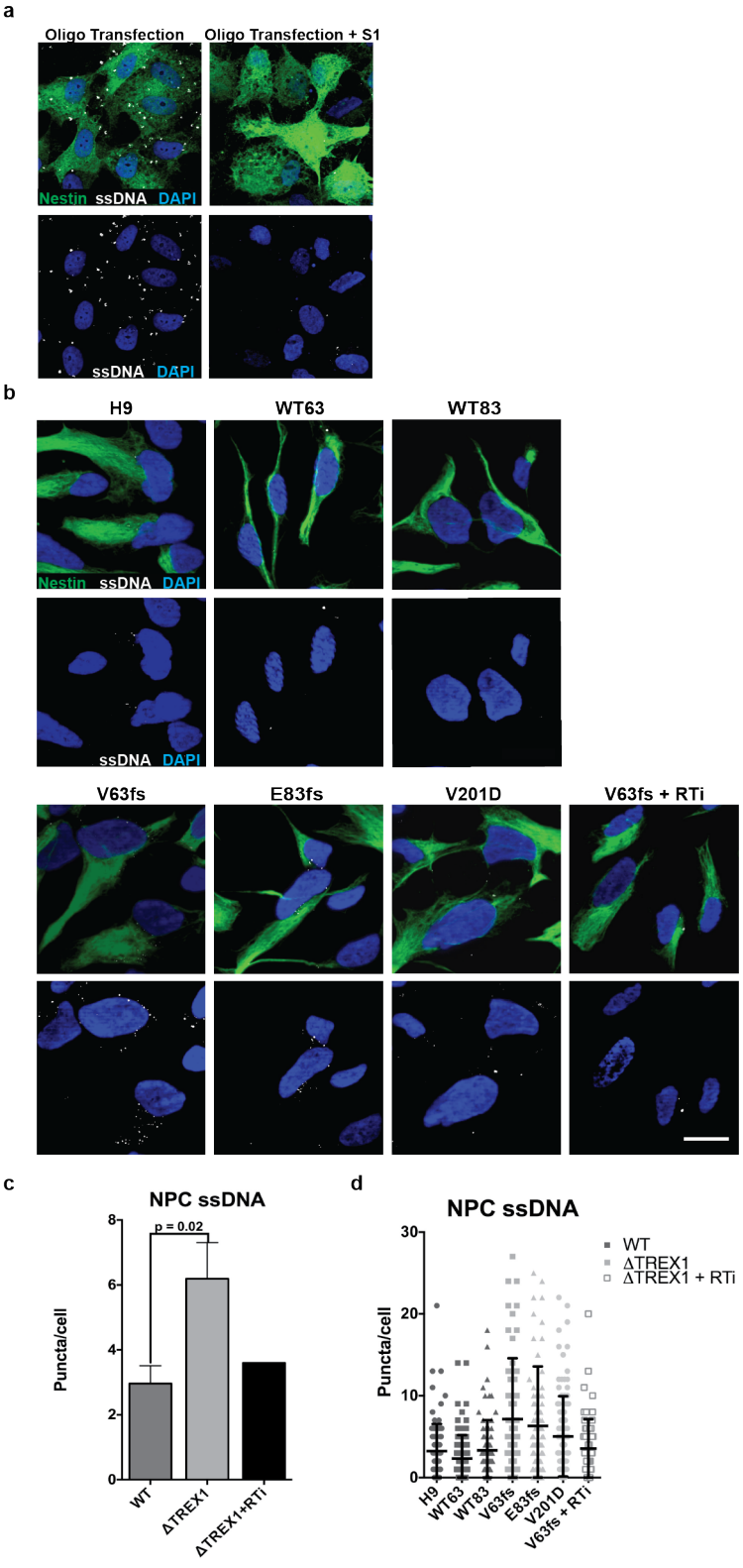
Figure 14. Differentiation and purification of NPCs into astrocytes. a, In astrocytic differentiation, the NPCs were transduced with a lentivirus expressing tdTomato from the GFAP promoter. The NPCs were lifted to form neurospheres, grown in suspension with and pushed towards astrocytes. After 21 days, the neurospheres were plated, and the astrocytes proliferated, before FACS purification. Scale bar in μm, as indicated within. **b**, Dot plot of FACs purification of astrocytes. Astrocytes were collected by selecting high tdTomato fluorescence. **c**,**d**, Representative immunofluorescence images of astrocytic markers, GFAP and S100. Scale bar, 20 μm. **e**, Expression of astrocytic markers determined by qPCR. Expression was normalized to zero and both *B2M* and *HPRT1* were used as dual internal references.


a
Neurospheres
Plated Spheres
Sorted Astrocytes
GFAP::tdTomato
b
12.0%
tdTomato
c
WT63
V63fs
Merge
GFAP
S100
DAPI
d
H9
WT83
E83fs
V201D
V63fs + RTi
GFAP S100 DAPI
e
CSPG4
GFAP
S100B
VIM
Relative Quantity
n.d.
WT
ΔTREX1
ΔTREX1+RTi
H9-NPC
H9
WT63
WT83
V63fs
E83fs
V201D
V63fs+RTi
H9-NPC

4.4 Loss of TREX1 function leads to accumulation of reverse-transcribed ssDNA in the cytosol of neural cells

Because a fundamental role of TREX1 is to degrade ssDNA in the cytosol (Chowdhury et al., 2006; Mazur and Perrino, 1999; Yang et al., 2007), we hypothesized that TREX1-deficient neural cells would accumulate ssDNA. Using an antibody specific to ssDNA, we examined ssDNA levels in NPCs, neurons, and astrocytes of TREX1-deficient cells and respective controls. To ensure the protocol was effective, we performed two transfections on control NPCs with added oligonucleotide, and treated one transfection with the S1 nuclease to remove ssDNA (**Fig. 12a**). A high-level of distinct ssDNA puncta was observed in the transfected NPCs, and these puncta were not apparent in S1-treated cells. When compared to control NPCs, TREX1-deficient NPCs demonstrated significantly greater numbers of ssDNA puncta per cell (**Fig. 12b-d**) We also examined neurons, and noticed that both control and TREX1-deficient neurons showed less total ssDNA puncta compared to NPCs. Astrocytes, on the other hand, had a greater total of ssDNA puncta compared to NPCs. In both neurons and astrocytes, TREX1-deficient cells exhibited significantly more puncta per cell than the respective controls (**Fig. 12 and Fig. 13**).

Figure 15. Accumulation of ssDNA in TREX1-deficient NPCs. a, Immunofluorescence image of ssDNA in oligo transfected and S1 treated NPCs. **b**, Representative immunofluorescence images of ssDNA in NPCs. **c-d**, Graphical representation of ssDNA puncta levels of each line in NPCs. All ssDNA images were acquired and ssDNA puncta were quantified blindly. Puncta per cell levels of each line were averaged and graphed accordingly to genotype (n=3). The V63fs line was chronically treated with reverse-transcriptase inhibitors (RTi), imaged and graphed (n=1). Scale bar, 20μm.


a
Oligo Transfection
Oligo Transfection + S1
Nestin ssDNA DAPI
ssDNA DAPI
b
H9
WT63
WT83
Nestin ssDNA DAPI
ssDNA DAPI
V63fs
E83fs
V201D
V63fs + RTi
c
NPC ssDNA
p = 0.02
Puncta/cell
WT
ΔTREX1
ΔTREX1+RTi
d
NPC ssDNA
Puncta/cell
WT
ΔTREX1
ΔTREX1 + RTi
H9
WT63
WT83
V63fs
E83fs
V201D
V63fs + RTi

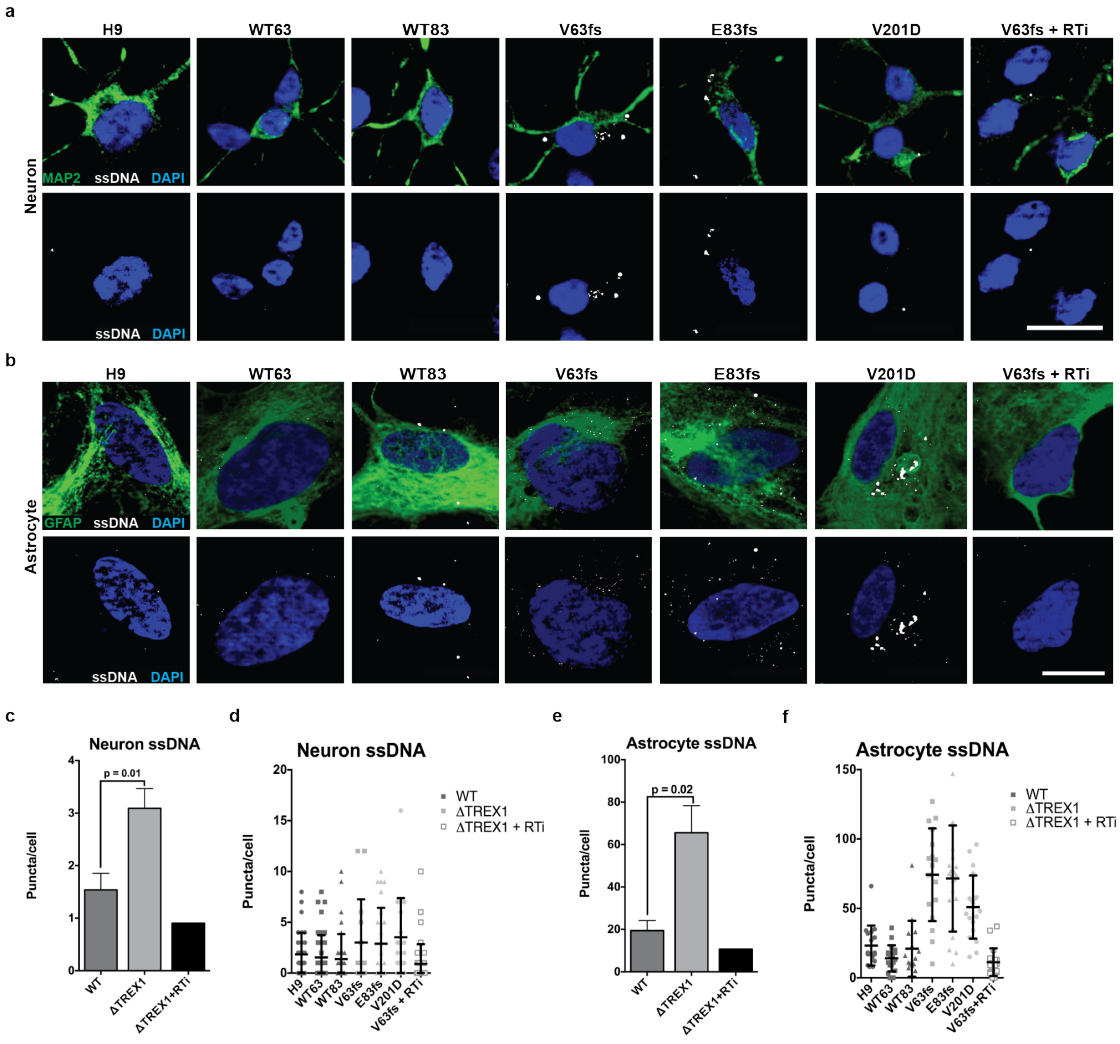


Figure 16. TREX-1 deficient neural cells have greater levels of ssDNA. a, Representative ssDNA immunofluorescence images of neurons. **b,** Representative ssDNA immunofluorescence images of astrocytes. **c-d**, Puncta quantification of ssDNA in the cytosol of neurons. e-**f**, Puncta quantification of ssDNA in the cytosol of astrocytes. All ssDNA images were acquired and ssDNA puncta were quantified blind to genotype. Puncta per cell levels of each line were averaged and graphed accordingly to genotype (n=3). The V63fs line was chronically treated with reverse-transcriptase inhibitors (RTi), imaged and graphed (n=1). Scale bar, 20μm.

Previous reports have suggested that retroelements comprise a large proportion of cytosolic DNA(Stetson et al., 2008). In humans, L1 is the major active retroelement(Brouha et al., 2003; Kazazian et al., 1988). As previously shown, L1 elements are expressed at high levels in neural cells(Coufal et al., 2009; Muotri et al., 2005). To determine if the human L1 retrotransposon was abundant in the cytosolic fraction of our experimental neural cells, we performed qPCR on NPC extrachromosomal DNA. We confirmed that the extrachromosomal DNA extract was high in mitochondrial DNA compared to genomic DNA extracts (**Fig. 14a**). Using primers specific to the L1 ORF1, ORF2, and 3'UTR-ORF2 junction, we detected more L1 DNA in the NPCs of TREX1-deficient cells, suggesting that L1 is a major source of DNA species in the cytosol (**Fig. 14b**). To validate this observation, we differentiated one of the TREX1-deficient cell line with reverse transcriptase inhibitor (RTi) HIV drugs Lamivudine (3TC) and Stavudine (D4T), which have been shown to inhibit L1 reverse transcription(Jones et al., 2008). Treatment of TREX1-deficient cells with 3TC and D4T reduced the level of ssDNA punctate to near-control levels in NPCs, neurons, and astrocytes (**Fig. 11, Fig. 12, and Fig. 13**). Furthermore, in the qPCR assay on extrachromosomal DNA, we detected fewer L1 DNA copies per cell in 3TC and D4T treated NPCs (**Fig. 14b**). Thus, we conclude that a significant source of the ssDNA in human neural cells arises from reverse transcription activity of endogenous L1 elements.

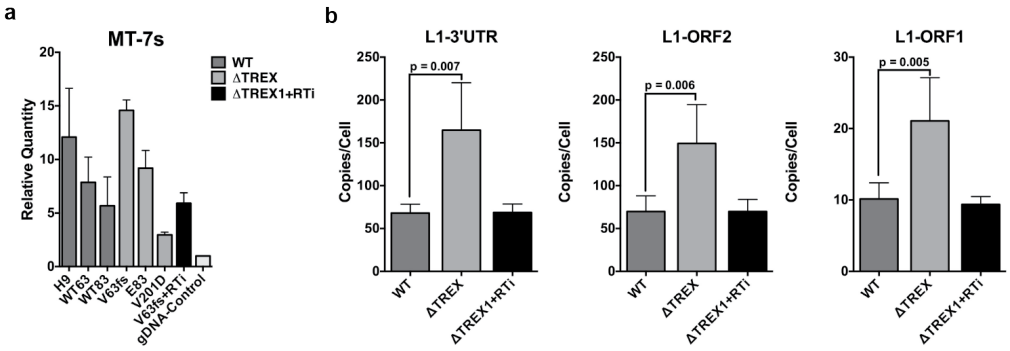


Figure 17. TREX-1 deficient neural cells have greater levels of L1 sequences in extrachromosomal DNA. a, Relative mitochondrial DNA levels as represented by *7S* gene and determined by qPCR. Quantities relative to NPC gDNA. **g**, Extrachromosomal L1 copies were quantified and graphed with three different primers corresponding to different regions of the L1. L1 copies were acquired in duplicate and normalized to cell number at time of extrachromosomal DNA extraction. L1 copies of each line were averaged and graphed accordingly to genotype (n=6). The V63fs line was chronically treated with RTi, quantified in duplicate and graphed (n=2). Values are means with standard deviation. Student's t-tests with Welch's correction were performed to compare genotypes.

4.5 Loss of TREX1 function results in increased neuronal toxicity

When differentiating TREX1-deficient NPCs into neurons, we noticed an apparent difference in cell death and toxicity. Furthermore, when we analyzed the gene expression of purified neurons, we found a significant decrease of neuronal markers (**Fig. 15a**). To test if TREX1-deficiency increases apoptosis in NPCs and neurons, we performed cleaved caspase-3 (CC3) and terminal deoxynucleotidyl transferase dUTP nick end-labeling (TUNEL) assays. When evaluating the NPCs, we did not detect a significant difference in the percentage of CC3 positive cells between TREX1-deficient lines and control lines (**Fig 15b**). However, when we performed the apoptosis assays on purified neurons, we identified a significantly greater percentage of CC3 and TUNEL positive cells in the TREX1-deficient line (**Fig. 15c-f**). Furthermore, treatment with 3TC and D4T reduced the percentage of CC3 and TUNEL positive cells in a TREX1-deficient line (**Fig. 15c-f**). This finding suggests that

human neurons are especially sensitive to increases in ssDNA, leading to apoptosis and cell death. Preventing the accumulation of ssDNA from reverse transcription with inhibitors promotes neuronal survival. This data is concordant with Trex1-KO mouse studies, which show RTi can ameliorate myocarditis and delay pre-mature death (Beck-Engeser et al., 2011).

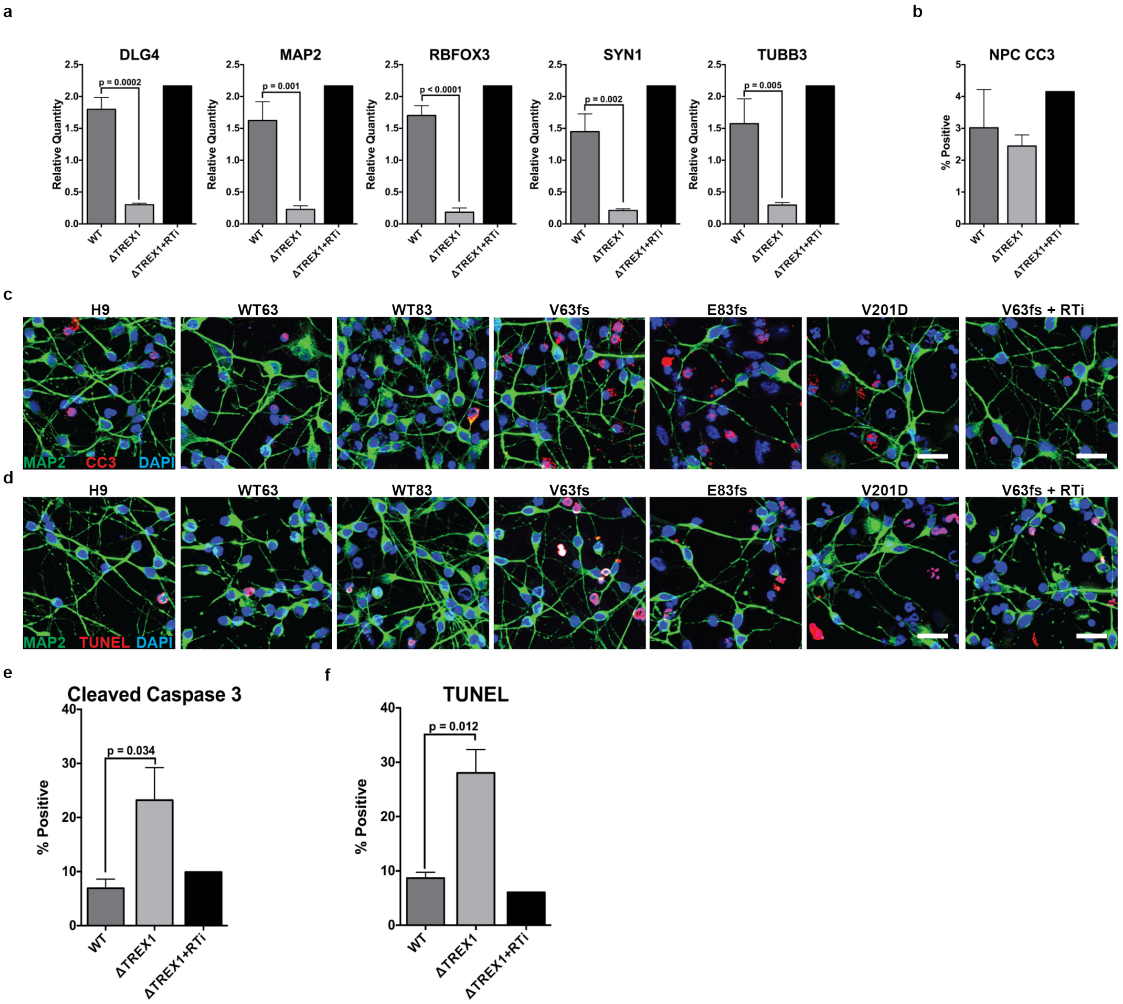


Figure 18. Trex1-deficiency promotes neuronal apoptosis. a, Expression of neuronal markers determined by qPCR. Values are means with standard deviation (n=3). Student's t-tests were performed to compare genotypes. Expression was normalized to zero and both B2M and HPRT1 were used as dual internal references. **b**, Graphical representation of percent of NPCs with cleaved caspase 3. Values are means with standard deviations. Student's t-test with Welch's correction determined no significant difference between the TREX1-deficient and control lines. **c-f**, Representative images of cleaved Caspase 3 (CC3) and TUNEL and corresponding graphs. CC3 and TUNEL images of purified neurons were acquired and percent of apoptotic cells calculated, averaged, and graphed accordingly to genotype (n=3) (c-**f**). The V63fs line was chronically treated with reverse-transcriptase inhibitors (RTi), imaged and graphed (n=1)(**c-f**). Scale bar, 20μm. Values are means with standard deviation. Student's t-tests with Welch's correction were performed to compare genotypes.

Previous reports have shown that astrocytes can induce toxicity when stressed (Campbell et al., 1999; Cuadrado et al., 2013). Considering the high level of ssDNA in the cytosol of TREX1-deficient astrocytes, we hypothesized the astrocytes could also contribute to the neurotoxicity associated with AGS. To test this hypothesis, we conditioned media on FACs-purified astrocytes for 48 hours. The astrocyte-conditioned media was overlaid on top of H9-purified neurons for 48 hours and tested for toxicity by CC3 and TUNEL. The neurons overlaid with the conditioned media from TREX1-deficient astrocytes exhibited significantly greater percentage of CC3 and TUNEL positive cells (**Fig. 16a-d**). Neurons overlaid with the conditioned media from TREX1-deficient astrocytes treated with RTi demonstrated similar percentages of CC3 and TUNEL positive cells compared to controls (**Fig. 16a-d**). The increased cell death in the astrocyte-conditioned treated neurons could be explained by two non-exclusive hypotheses. Either the TREX1-deficient astrocytes are secreting a neurotoxic factor, or the TREX1-deficient astrocytes are not capable of producing neurotrophic factors provided by control astrocytes. Further investigation is required to determine the mechanisms and contribution of astrocytes to the observed non-cell autonomous toxicity in TREX1-deficient neurons.

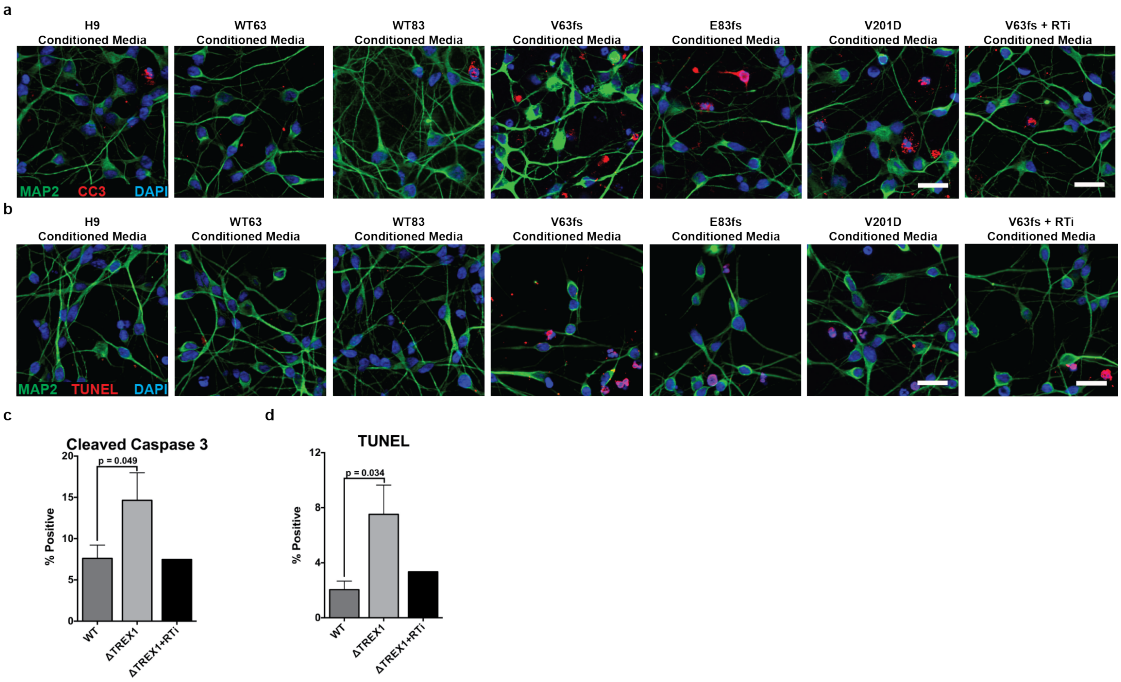


Figure 19. Trex1-deficiency in astrocytes promotes neuronal apoptosis via secreted factors. a-d, Representative images of cleaved Caspase 3 (CC3) and TUNEL and corresponding graphs. CC3 and TUNEL images of H9 purified neurons overlaid with astrocyte-conditioned media of different genotypes (n=3)(RTi n=1)(**f-i**). Scale bar, 20μm. Values are means with standard deviation. Student's t-tests with Welch's correction were performed to compare genotypes.

4.6 Type I interferon and interferon stimulated genes in astrocytes

Reports on AGS patients have indicated Interferon-α (IFNa) as a part of it's pathophysiology(Lebon et al., 1988). Thus, we examined IFNa mRNA levels in astrocytes using primers to identify several of the IFNa genes, including IFNa 1, 2, 4, 6, 10, 13, and 21. We could only detect IFNa 13, which was not significantly different between TREX1-deficient and control astrocytes (**Fig 17a**). A recent study showed interferon stimulated genes (ISGs) could be upregulated independently of interferon signaling (Hasan et al., 2013). We examined seven ISGs and found Trex1-deficient cells upregulated all seven. From this data, we conclude that ISGs are upregulated in astrocytes independent of IFNa. Furthermore, we suggest the source of intrathecal IFNa in AGS patients does not originate from astrocytes, but perhaps from the

resident microglia and the infiltrating leukocytes. Further examination is necessary to determine the IFNa source.

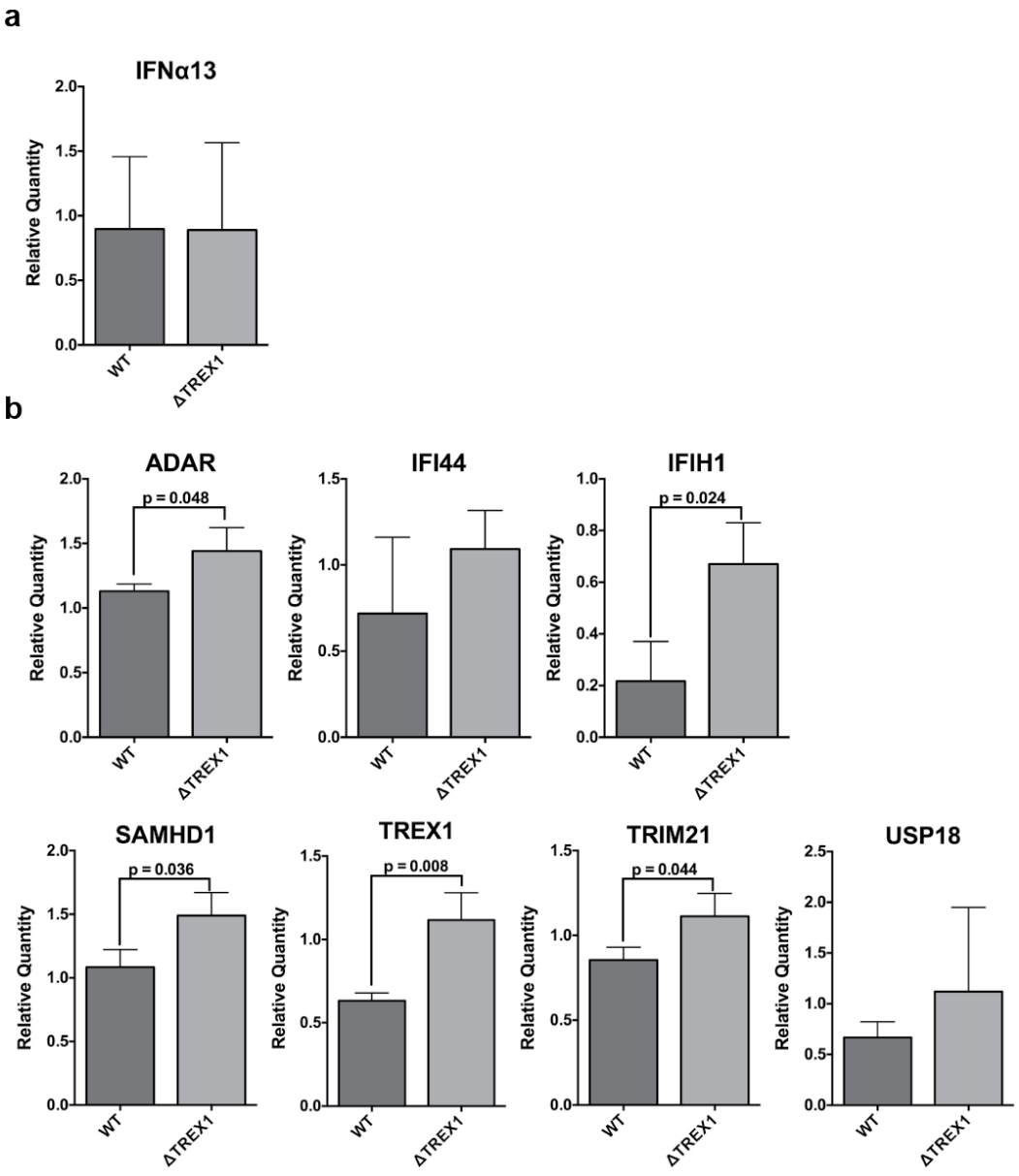


Figure 20. Trex1-deficienct astrocytes upregulate interferon-stimulated genes. **a**, Expression of IFNα13 determined by qPCR. **b**, Expression of ISGs determined by qPCR. Values are means with standard deviation (n=3). Student's t-tests were performed to compare genotypes. Expression was normalized to zero and both B2M and HPRT1 were used as dual internal references.

4.7 Discussion

Here, we created a human neural cell model to study the autoinflammatory disorder AGS. The advantages of a human cell based model are numerous. For the first time, we can examine AGS-like human neural cells to explore the molecular phenotypes. We could delineate phenotypes specific to each neurons and astrocytes, which was not possible studying the *Trex1* knockout mouse. Futhermore, we can use the pluripotent system to study the molecular AGS pathophysiology as the neural cells differentiate.

Two AGS neural cell models in this report were differentiated from human pluripotent cells that were mutagenized by the CRISPR/Cas9 system. A major caveat to the CRISPR/Cas9 system is the potential of mutating unintended genomic sequences. Previous reports have indicated a CRISPR/Cas9 system can cleave target sites with 4 mismatches to the gRNA, suggesting high level of off-target mutagenesis could occur (Fu et al., 2013). However, these reports used an exogenous plasmid to determine off-target efficiencies, not actual genomic sequence. In our work, we performed two separate analyses to determine off-target mutagenesis. The first was biased approach, checking the exonic loci most similar to our guide RNA. Using the CRISPR-design website we found 15 and 17 potential off-targets for the V63 and E83 loci, respectively (Hsu, 2013). The V63 gRNA had two predicted off-targets with only 3 mismatches. In the E83 gRNA, *TREX2* was predicted to be the most likely off-target with only 2 mismatches. However, none of the predicted off-targets were mutagenized, suggesting, at least in the context of exonic DNA, that the gRNA is highly specific.

The second approach to find off-targets was unbiased. We examined the entire exome, comparing the clonally derived lines to the parental H9 line. Each line had between 20-35 homozygous InDels, with 13-18 inducing a frameshift. We closely assessed the sequences of each InDel locus. None of the loci matched the gRNA sequence when allowing up to 50% mismatch. Furthermore, some of the InDels in each line were exactly the same mutation, regardless of which gRNA was used. For example, all four clonally derived lines (V63fs, E83fs, WT63, and WT83) had a 1 base pair cytosine insertion in a specific locus within *NCAM1.* It is extremely unlikely that disparate gRNAs induced a mutation in the exact same locus. In addition, it is extremely unlikely that all four lines would arise with the exact same insertion if repaired by nonhomologous end joining, the common mechanism of Cas9 endonuclease repair. Therefore, we concluded that these InDels existed in the subpopulations of the late-passage H9 cells and arose during clonal expansion.

Our off-target results correspond well with the recent publications (Smith et al., 2014; Veres and Talkowski, 2014). Similar to our approach, Veres *et al.* performed whole genome sequencing on pluripotent stem cells undergoing CRISPR/Cas9 endonuclease activity and clonal expansion. The authors assessed 6 cell lines, finding between 2 and 5 InDels per line. None of the InDels had sequence homology with the gRNA sequence. The authors also sequenced 3 pluripotent clonal lines that were cut with TALENs. Although TALENS are thought to be more specific, the authors did find one cell line with a putative off-target matching the TALEN binding sequence with only 4-mismatches on each arm (Veres and Talkowski, 2014). These publications strongly corroborate our off-target sequencing data.

We used neural cell-based system and delineate the inflammatory pathology from different neural cell types. We found astrocytes can tolerate extremely high levels of ssDNA species in their cytosol. Neurons could not, suggesting that neurons are undergoing apoptosis if ssDNA is not degraded by TREX1. We can also use this system to perform controlled experiments to determine how one neural cell type can affect another. In this report, we used conditioned media from astrocytes to determine the effect of astrocyte secretion on neurons.

The neural cell model enabled us to attribute new phenotypes to the pathophysiology of AGS. We found that TREX1-deficient neurons are undergoing massive apoptosis as they differentiate. The neuronal cell death was self-evident when comparing control and TREX1-deficient differentiating neurons in a phase-contrast microscope and in the apoptosis assays. We found it surprising that the neurons, without mature astrocytes, were undergoing apoptosis. The data suggests that the rise of ssDNA is the intrinsic cause of neuronal death. However, we cannot discount possible influence from the “contaminant” cell population that inevitably arises during neuronal differentiation as contributors to the neurotoxicity.

We also found that the neurons overlaid with conditioned media from TREX1-deficient astrocytes underwent apoptosis at an increased rate when compared to neurons overlaid with conditioned media from control astrocytes. We propose two possible hypotheses to explain the greater level of apoptosis in the neurons overlaid with conditioned media from TREX1-deficient astrocytes. One possibility is that the TREX1-deficient astrocytes are secreting a neurotoxic factor, such as an inflammatory signal. Alternatively, the TREX1-deficient astrocytes are not secreting enough neurotrophic factors, such as glial derived neurotrophic factor (GDNF) or

brain derived neurotrophic factor (BDNF). Further experimentation is necessary to delineate these two hypotheses.

We determined that L1 elements were a major source of nucleic acid species in the cytosol. Furthermore, we found inhibiting reverse transcription with nucleoside analogues, we could rescue many of the phenotypes observed in the TREX1-deficient neural cells. This data provide strong support to consider using RTi to treat AGS patients. It should be noted, however, in our study, the RTi was present before neuronal differentiation took place, preventing the high level of apoptosis observed. Thus, when treating AGS patients, it would be extremely important to begin treatment as soon as possible to limit neuron loss.

One hallmark of AGS is the increased amount of IFNa in the cerebral spinal fluid. The source of the IFNa is not conclusively know, but one study showed a histological image of IFNa expression overlapping with GFAP, suggesting astrocytes were the source of IFNAa. When examining the TREX1-deficient astrocytes for interferon levels, we could not detect expression of IFNb and IFNa 1, 2, 6, 10 and 21 by qPCR. Detecting IFNa expression is made more difficult by the fact there are 13 different proteins, each with a fairly unique DNA sequence. Furthermore, no one has performed a study to determine with IFNas are produced in the human brain or by human astrocytes in culture. However, one group has found that SIV induces the expression of IFNa 2, 6 and 13 in the brains of rhesus macaques. Assuming the human system is similar to macaque system, we explored the expression of these IFNa genes plus a few others in cultured astrocytes. Only IFNa13 was detected by qPCR. In addition, we did not find any difference between the TREX1-deficient and control astrocytes suggesting the ssDNA did not stimulate an IFNa response or the

response was somehow muted. We could not corroborate the data from the previous study and further research is necessary to determine the IFNa source.

It is important to consider the fact that all experiments performed were done *in vitro*, absent of immunological cells. Although our study focused on the process of intrinsic autoinflammation, it is highly likely that immunological cells exacerbate the inflammation response. Therefore, we could miss the extrinsic cross talk between neural cells and immune cells that are could be necessary for different processes in AGS pathophysiology, such as IFNa production. Experiments using TREX1-deficient neural cells and TREX1-deficient microglia or macrophages could yield some insight on this debate.

Our results suggest a model in which the neurotoxicity in AGS arises in part due to reverse transcription of endogenous L1 elements. The loss of TREX1 function permits L1-derived ssDNA species to accumulate in the cytosol of neural cells. Neurons are sensitive to the excessive DNA and undergo apoptosis. In contrast, astrocytes are able to tolerate extraordinary levels of ssDNA. However, the lack of TREX1-function induces astrocytes to further contribute to the neuronal cell death independent of IFNa. Results from this *in vitro* model allowed us to show, for the first time, that TREX1-deficiency leads to neurotoxicity in the absence of a fully intact immune system. Furthermore, our results indicate that RTi might be capable of reducing the neurotoxicity observed in AGS, thus opening a new therapeutic opportunity by treatment with FDA-approved HIV reverse transcriptase inhibitors.

4.8 Materials and methods

Media composition for tissue culture

Pluripotent stem cell media (MT): mTeSR^{TM}1 (Stem Cell Technologies). Differentiation media for Embryoid Bodies (N2): Dulbecco's Modification of Eagle's Medium/ Ham's F12 (DMEM/F12 50/50; Corning Cellgro) with 1 x HEPES, 1 x pen-strep, glutamax (Life Technologies), and N2 NeuroPlex (Gemini Bio-products). Supplemented with 1μM dorsomorphin (Tocris) and Stemolecule SB431542 (StemGent). Neural progenitor cell media (NGF): DMEM/F12 50/50 with 1 x HEPES, 1x pen-strep, glutamax (Life Technologies), N2 NeuroPlex (Gemini Bio-products), Gem21 Neuroplex (Gemini Bio-products). Supplemented with 20ng/mL basic fibroblast growth factor (bFGF; Life Technologies). Neuronal media (NG): DMEM/F12 50/50 with 1 x HEPES, 1x pen-strep, glutamax (Life Technologies), N2 NeuroPlex (Gemini Bio- products), Gem21 Neuroplex (Gemini Bio-products). cAMP, GDNF, and BDNF were added to the neuronal media of purified neurons. Astrocytes were cultured in Astrocyte Growth Medium (Lonza). Treatment with reverse transcriptase inhibitor (RTi) was with 1μM Stavudine (D4T) and 10μM Lamivudine (3TC) (Sigma-Aldrich). Astrocyte conditioned media was created by overlaying neurobasal with B27 and 5% FBS over purified astrocytes for 48 hours.

Maintenance of iPSC and hESC culture

Reprogrammed iPSCs, H9 ESCS, and mutagenized H9 ESCs were propagated in mTeSR and passaged manually as small clumps onto matrigel (BD Bioscences) coated plates.

Mutagenesis of H9 ESC with CRISPR/Cas9

The CMV::Cas9-2A-eGFP plasmid was purchased from addgene (#44719). The guide RNA (gRNA) sequences correspond to two loci of the *TREX1* gene(Mali et al., 2013). The gRNA vector was created according to the gRNA synthesis protocol(Mali et al., 2013) using DNA Strings (Life Technologies) cloned into a pUC57 plasmid.

gRNA1 – GAGAGCTTGTCTACCACACGCGG

gRNA2 – GCTCAGACCTGTGATCTCGCTGG

H9 ESCs were lifted off the plate with accutase (Stem Cell Technologies), and passed through a 40 μm nylon mesh (BD Biosciences) to ensure single cells before transfection. Using hESC Kit 2 and program B16 From Amaxa Nucleofection (Lonza), 1.5 million cells were transfected with the 7 μg of the CMV::Cas9-2A-eGFP vector and 3μg of the U6::gRNA vector. Transfected cells were plated with mTeSR and 5 μM Rock Inhbitor (Ri) (Tocris) for 48 hours before FAC sorting. $eGFP^+$ cells were collected using the BD Influx (**Fig. 2a**), and plated communally to recover for 48 hours with 5uM Ri. After 48 hours, 40 thousand single-cells were seeded onto 10cm plates with 5 μM Ri, and allowed to grow into colonies (3-4 weeks). Genomic DNA from the unused transfected cells was assayed for endonuclease efficiency using the Surveyor Nuclease Kit (Transgenomic) (**Fig. 2b**). Isolated colonies were selected, expanded, and genomic DNA extracted for sequencing with the DNeasy Blood and Tissue Kit (Qiagen). To ensure the mutation was stable, RNA was collected from early and late passage ESCs and NPCs and reverse transcribed into cDNA. The cDNA was topo cloned and 6 clones of each passage were sequenced.

Primers for V63 and E83 locus sequencing and Surveyor Nuclease:

For – CTTCGGATCTTAACACTGGGC

Rev – CCACACAGAAGGCACCATCC

Primers for V201 locus sequencing:

For – CAGCGAGATCACAGGTCTGAG

Rev - GCCAGGGATAGTCCATACAGT

Creation of induced-pluripotent stem cells from patient fibroblasts

Fibroblasts of an AGS patient with a stereotypical V201D mutation were reprogrammed using episomal Yamanaka factors as described previously(Yu et al., 2009). Five isolated clones were selected and expanded. Three clones were used for differentiation and experimentation.

Teratoma formation and karyotyping

Teratomas were formed, extracted and stained as described previously(Marchetto et al., 2010). Karyotyping was outsourced and performed by Children's Hospital Los Angeles.

Differentiation of pluripotent cells into NPCs, neurons and astrocytes

Pluripotent cells were differentiated into NPCs and neurons as described previously(Marchetto et al., 2010). Neurons were purified by FACs as described previously(Yuan et al., 2011). Purified neurons were plated onto a 96 well plate coated with poly-ornithine and laminin at a cell density of 100,000 cells per well. To differentiate into astrocytes, NPCs were transduced with a lentivirus expressing tdTomato from the GFAP promoter. The transduced NPCs were lifted into suspension

and kept on a shaker (95 rpm) to form neurospheres and maintained for 3 weeks. After 1 week in neuronal media, the neurospheres were overlaid with astrocyte media for the remaining two weeks. The neurospheres were plated onto poly-ornithine and laminin coated plates, and expand for 2-3 passages before FAC sorting. Purified astrocytes were selected by high tdTomato expression and plated onto poly-ornithine and laminin coated 96 well plates with 10,000 cells per well.

Immunocytochemistry

Unless otherwise noted, cells were fixed in 4% paraformaldehyde (PFA, Electron Microscopy Sciences) for 20 minutes, permeabilized with 0.25% Triton X-100 for 15 minutes, blocked with 3% bovine serum albumin (BSA) (Gemini Bio) and treated with the primary antibodies diluted in 3% BSA overnight at 4°C. The next day, cells were incubated with the secondary antibodies, DAPI, and mounted. Primary antibody dilutions were used as follows: anti-Nanog (R&D, *AF1997*, 1:500), anti-Lin28 (Abcam, *ab46020,* 1:500), anti-Nestin (Millipore, *AB5922,* 1:1000), anti-Sox2 (Abcam, *ab75485,* 1:250), anti-GFAP (Abcam, *ab4674,* 1:2000), anti-S100 (Abcam, *ab4066,* 1:200), anti-Map2 (Abcam *ab3392,* 1:2000*),* anti-Syn1 (Millipore *AB1543,* 1:500), anti-ssDNA (Millipore, *MAB3299*, 1:20), anti-Cleaved Caspase 3 (Asp175) (Cell Signaling, *9661*, 1:400). Secondary antibodies conjugated to Alexa Fluors 488, 555, and 647 were used with a dilution of 1:500 (Life Technologies). TUNEL assay was performed using the Click-iT TUNEL assay kit (Life Technologies). To perform the ssDNA staining, cells were fixed on ice with 4% PFA for 20 minutes, then with methanol overnight at -20^{o}C. The next day, cells were treated with 200 µg/mL of RNase (Sigma R4642) for 4 hours at 37^{o}C. The cells were blocked with 3% BSA and

stained as indicated above. Images of ssDNA were acquired blindly and ssDNA puncta were quantified blindly.

Oligonucleotide transfection and S1 nuclease treatment

5ug of a random 60-mer oligonucleotide was tranfected into 1 million H9 NPCs using 20 ug of polyethylenimine. The cells were fixed 30 minutes post-transfection. For S1 nuclease treatment, 500 U/mL of nuclease was added during the RNase treatment with the S1 buffer supplied (Promega M5761).

qRT-PCR analysis

RNA was obtained from cells using the RNeasy Plus Mini Kit (Qiagen). One microgram of RNA used to make cDNA using Qiagen's Quantitect Reverse Transcriptase Kit. 10 ng of cDNA was used in each qPCR reaction using TaqMan probes and the Taqman Universal Master Mix II (Life Technologies). Reactions were performed in triplicate.

Extrachromosomal DNA extraction and qPCR

Extrachromosomal DNA was extracted using the modified Hirt protocol(Arad, 1998). 5ng of DNA was used in each qPCR reaction using TaqMan probes and the Taqman Universal Master Mix II (Life Technologies). Two biological replicates were tested per cell line. Each qPCR reaction was performed in triplicate. Detection of L1 copies in the extrachromosomal DNA was quantified with qPCR using TaqMan probes designed by Life Technologies. The L1 probes amplify within the following 100 bp sequences.

ORF1 –

ATGGGGAAAAAACAGAACAGAAAAACTGGAAACTCTAAAACGCAGAGCGCCTCT
CCTCCTCCAAAGGAACGCAGTTCCTC

ORF2 –

GCTCATGGGTAGGAAGAATCAATATCGTGAAAATGGCCATACTGCCCAAGGTAA
TTTACAGATTCAATGCCATCCCCATC

ORF2-3'UTR –

TGGAAACCATCATTCTCAGTAAACTATCGCAAGAACAAAAAACCAAACACCGCAT
ATTCTCACTCATAGGTGGGAATTGA

Exome sequencing data analysis

Genomic DNA was extracted from ESCs of passage 42 H9 line and passage 3 of the WT63, WT83, V63fs, and E83fs lines with the DNeasy Blood & Tissue Kit (Qiagen). The exome DNA sequencing was outsourced and performed by the UCSD IGM Genomics center.

Raw exome data were filtered to recover high quality (phread-scaled quality score higher than 22) sequencing reads using IlluQC software(Patel and Jain, 2012). High quality reads of each individual sequenced cell lines were aligned to human reference genome (build Hg19) using BWA-MEM software(Li and Durbin, 2009) with default parameters. Unrelated alignments to human exome were then filtered out of the analysis. Using SAMTOOLS software(Li et al., 2009) over valid alignments, duplicated reads as well as redundant alignments were removed to produce sorted alignment files for each sequenced exome library. We then applied FreeBayes over sorted alignments to perform genomic InDel detection for each H9 control and H9-induced CRISPR/cas9 mutations. For InDel detection, alignments of sequenced cell

lines where restricted to a minimum of 50 non-repetitive reads of coverage. All InDel mutations found were considered homozygous when a minimum of 95% of covered reads have the same mutation, and heterozygous, when a minimum of 40% of covered reads shares the same mutation. Only those mutations not shared by H9-induced CRISPR/cas9 mutated cell lines and H9 control are valid variations between H9 control and mutated cell lines. Using Annovar(Wang et al., 2010) and reference transcriptome annotation (UCSC Hg19) of human genome (build Hg19).

To detect off-target genomic mutations caused by CRISPR/cas9 in the transformed H9 cell lines, gDNA reads were aligned against human reference genome (build Hg19) with Bowtie software(Langmead et al., 2009), allowing up to 13 mismatches, but requiring the dinucleotide GG of the PAM motif.
A list of *in-silico* genomic off-target sites detected by the CRISPR Design Tool (http://www.genome-engineering.org/) (Hsu, 2013) was compiled. The sequences of the predicted off-target sites were examined in the exome data.

Chapter 4, in part, has been submitted for publication of the material as it may appear in Nature, 2014, Thomas, C.A.; Tejwani, L.; Herai, R.; Trujillo, C.A.; Crow, Y.J.; and Muotri, A.R. The dissertation author was the primary investigator and author of this paper.

REFERENCES

Adams, J.W., Kaufman, R.E., Kretschmer, P.J., Harrison, M., and Nienhuis, A.W. (1980). A family of long reiterated DNA sequences, one copy of which is next to the human beta globin gene. Nucleic acids research *8*, 6113-6128.

Aicardi, J., and Goutieres, F. (1984). A progressive familial encephalopathy in infancy with calcifications of the basal ganglia and chronic cerebrospinal fluid lymphocytosis. Annals of neurology *15*, 49-54.

Alisch, R.S., Garcia-Perez, J.L., Muotri, A.R., Gage, F.H., and Moran, J.V. (2006). Unconventional translation of mammalian LINE-1 retrotransposons. Genes & development *20*, 210-224.

Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., and Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nature genetics *23*, 185-188.

Arad, U. (1998). Modified Hirt procedure for rapid purification of extrachromosomal DNA from mammalian cells. BioTechniques *24*, 760-762.

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 research *32*, 3846-3855.

Baillie, J.K., Barnett, M.W., Upton, K.R., Gerhardt, D.J., Richmond, T.A., De Sapio, F., Brennan, P.M., Rizzu, P., Smith, S., Fell, M.*, et al.* (2011). Somatic retrotransposition alters the genetic landscape of the human brain. Nature *479*, 534-537.

Bar-Shira, A., Rashi-Elkeles, S., Zlochover, L., Moyal, L., Smorodinsky, N.I., Seger, R., and Shiloh, Y. (2002). ATM-dependent activation of the gene encoding MAP kinase phosphatase 5 by radiomimetic DNA damage. Oncogene *21*, 849-855.

Barnes, M.H., Spacciapoli, P., Li, D.H., and Brown, N.C. (1995). The 3'-5' exonuclease site of DNA polymerase III from gram-positive bacteria: definition of a novel motif structure. Gene *165*, 45-50.

Beck, C.R., Garcia-Perez, J.L., Badge, R.M., and Moran, J.V. (2011). LINE-1 elements in structural variation and disease. Annual review of genomics and human genetics *12*, 187-215.

Beck-Engeser, G.B., Eilat, D., and Wabl, M. (2011). An autoimmune disease prevented by anti-retroviral drugs. Retrovirology *8*, 91.

Bellefroid, E.J., Poncelet, D.A., Lecocq, P.J., Revelant, O., and Martial, J.A. (1991). The evolutionarily conserved Kruppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. Proceedings of the National Academy of Sciences of the United States of America *88*, 3608-3612.

Blaschke, A.J., Staley, K., and Chun, J. (1996). Widespread programmed cell death in proliferative and postmitotic regions of the fetal cerebral cortex. Development *122*, 1165-1174.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science *326*, 1509-1512.

Boissinot, S., Chevret, P., and Furano, A.V. (2000). L1 (LINE-1) retrotransposon evolution and amplification in recent human history. Molecular biology and evolution *17*, 915-928.

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. Proceedings of the National Academy of Sciences of the United States of America *100*, 5280-5285.

Bundo, M., Toyoshima, M., Okada, Y., Akamatsu, W., Ueda, J., Nemoto-Miyauchi, T., Sunaga, F., Toritsuka, M., Ikawa, D., Kakita, A.*, et al.* (2014). Increased l1 retrotransposition in the neuronal genome in schizophrenia. Neuron *81*, 306-313.

Burwinkel, B., and Kilimann, M.W. (1998). Unequal homologous recombination between LINE-1 elements as a mutational mechanism in human genetic disease. Journal of molecular biology *277*, 513-517.

Buss, R.R., Sun, W., and Oppenheim, R.W. (2006). Adaptive roles of programmed cell death during nervous system development. Annual review of neuroscience *29*, 1-35.

Buzdin, A., Ustyugova, S., Gogvadze, E., Vinogradova, T., Lebedev, Y., and Sverdlov, E. (2002). A new family of chimeric retrotranscripts formed by a full copy of U6 small nuclear RNA fused to the 3' terminus of l1. Genomics *80*, 402-406.

Bylund, M., Andersson, E., Novitch, B.G., and Muhr, J. (2003). Vertebrate neurogenesis is counteracted by Sox1-3 activity. Nature neuroscience *6*, 1162-1168.

Campbell, I.L., Krucker, T., Steffensen, S., Akwa, Y., Powell, H.C., Lane, T., Carr, D.J., Gold, L.H., Henriksen, S.J., and Siggins, G.R. (1999). Structural and functional neuropathology in transgenic mice with CNS expression of IFN-alpha. Brain research *835*, 46-61.

Chahrour, M., and Zoghbi, H.Y. (2007). The story of Rett syndrome: from clinic to neurobiology. Neuron *56*, 422-437.

Chaline, J.D., A.; Marchand, D.; Dambricourt Malassé, A.; Deshayes, M. J. (1996). Chromosomes and the origins of apes and australopithecins. Journal of human evolution *11*, 43-60.

Chen, J.M., Chuzhanova, N., Stenson, P.D., Ferec, C., and Cooper, D.N. (2005). Meta-analysis of gross insertions causing human genetic disease: novel mutational mechanisms and the role of replication slippage. Human mutation *25*, 207-221.

Chow, J.C., Ciaudo, C., Fazzari, M.J., Mise, N., Servant, N., Glass, J.L., Attreed, M., Avner, P., Wutz, A., Barillot, E.*, et al.* (2010). LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. Cell *141*, 956-969.

Chowdhury, D., Beresford, P.J., Zhu, P., Zhang, D., Sung, J.S., Demple, B., Perrino, F.W., and Lieberman, J. (2006). The exonuclease TREX1 is in the SET complex and acts in concert with NM23-H1 to degrade DNA during granzyme A-mediated cell death. Molecular cell *23*, 133-142.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A.*, et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.

Cong, L., Zhou, R., Kuo, Y.C., Cunniff, M., and Zhang, F. (2012). Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. Nature communications *3*, 968.

Cordaux, R., and Batzer, M.A. (2009). The impact of retrotransposons on human genome evolution. Nature reviews Genetics *10*, 691-703.

Cost, G.J., Feng, Q., Jacquier, A., and Boeke, J.D. (2002). Human L1 element target-primed reverse transcription in vitro. The EMBO journal *21*, 5899-5910.

Coufal, N.G., Garcia-Perez, J.L., Peng, G.E., Marchetto, M.C., Muotri, A.R., Mu, Y., Carson, C.T., Macia, A., Moran, J.V., and Gage, F.H. (2011). Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells. Proceedings of the National Academy of Sciences of the United States of America *108*, 20382-20387.

Coufal, N.G., Garcia-Perez, J.L., Peng, G.E., Yeo, G.W., Mu, Y., Lovci, M.T., Morell, M., O'Shea, K.S., Moran, J.V., and Gage, F.H. (2009). L1 retrotransposition in human neural progenitor cells. Nature *460*, 1127-1131.

Crow, Y.J., Hayward, B.E., Parmar, R., Robins, P., Leitch, A., Ali, M., Black, D.N., van Bokhoven, H., Brunner, H.G., Hamel, B.C.*, et al.* (2006a). Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 cause Aicardi-Goutieres syndrome at the AGS1 locus. Nature genetics *38*, 917-920.

Crow, Y.J., Jackson, A.P., Roberts, E., van Beusekom, E., Barth, P., Corry, P., Ferrie, C.D., Hamel, B.C., Jayatunga, R., Karbani, G.*, et al.* (2000). Aicardi-Goutieres syndrome displays genetic heterogeneity with one locus (AGS1) on chromosome 3p21. American journal of human genetics *67*, 213-221.

Crow, Y.J., Leitch, A., Hayward, B.E., Garner, A., Parmar, R., Griffith, E., Ali, M., Semple, C., Aicardi, J., Babul-Hirji, R.*, et al.* (2006b). Mutations in genes encoding

ribonuclease H2 subunits cause Aicardi-Goutieres syndrome and mimic congenital viral brain infection. Nature genetics *38*, 910-916.

Cuadrado, E., Jansen, M.H., Anink, J., De Filippis, L., Vescovi, A.L., Watts, C., Aronica, E., Hol, E.M., and Kuijpers, T.W. (2013). Chronic exposure of astrocytes to interferon-alpha reveals molecular changes related to Aicardi-Goutieres syndrome. Brain : a journal of neurology *136*, 245-258.

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. Nature genetics *20*, 288-290.

Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.K., Shi, Y., and Yan, N. (2012). Structural basis for sequence-specific recognition of DNA by TAL effectors. Science *335*, 720-723.

Dombroski, B.A., Feng, Q., Mathias, S.L., Sassaman, D.M., Scott, A.F., Kazazian, H.H., Jr., and Boeke, J.D. (1994). An in vivo assay for the reverse transcriptase of human retrotransposon L1 in Saccharomyces cerevisiae. Molecular and cellular biology *14*, 4485-4492.

Fanning, T., and Singer, M. (1987). The LINE-1 DNA sequences in four mammalian orders predict proteins that conserve homologies to retrovirus proteins. Nucleic acids research *15*, 2251-2260.

Feng, Q., Moran, J.V., Kazazian, H.H., Jr., and Boeke, J.D. (1996). Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell *87*, 905-916.

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.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature biotechnology *31*, 822-826.

Gall, A., Treuting, P., Elkon, K.B., Loo, Y.M., Gale, M., Jr., Barber, G.N., and Stetson, D.B. (2012). Autoimmunity initiates in nonhematopoietic cells and progresses via lymphocytes in an interferon-dependent autoimmune disease. Immunity *36*, 120-131.

Gao, Z., Ure, K., Ables, J.L., Lagace, D.C., Nave, K.A., Goebbels, S., Eisch, A.J., and Hsieh, J. (2009). Neurod1 is essential for the survival and maturation of adult-born neurons. Nature neuroscience *12*, 1090-1092.

Garcia-Perez, J.L., Morell, M., Scheys, J.O., Kulpa, D.A., Morell, S., Carter, C.C., Hammer, G.D., Collins, K.L., O'Shea, K.S., Menendez, P.*, et al.* (2010). Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. Nature *466*, 769-773.

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

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A.*, et al.* (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442-451.

Gilbert, N., Lutz-Prigge, S., and Moran, J.V. (2002). Genomic deletions created upon LINE-1 retrotransposition. Cell *110*, 315-325.

Goodier, J.L., and Kazazian, H.H., Jr. (2008). Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell *135*, 23-35.

Goodier, J.L., Ostertag, E.M., Du, K., and Kazazian, H.H., Jr. (2001). A novel active L1 retrotransposon subfamily in the mouse. Genome research *11*, 1677-1685.

Goodman, M., Porter, C.A., Czelusniak, J., Page, S.L., Schneider, H., Shoshani, J., Gunnell, G., and Groves, C.P. (1998). Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. Molecular phylogenetics and evolution *9*, 585-598.

Goren-Inbar, N., Alperson, N., Kislev, M.E., Simchoni, O., Melamed, Y., Ben-Nun, A., and Werker, E. (2004). Evidence of hominin control of fire at Gesher Benot Ya'aqov, Israel. Science *304*, 725-727.

Goutieres, F., Aicardi, J., Barth, P.G., and Lebon, P. (1998). Aicardi-Goutieres syndrome: an update and results of interferon-alpha studies. Annals of neurology *44*, 900-907.

Graham, V., Khudyakov, J., Ellis, P., and Pevny, L. (2003). SOX2 functions to maintain neural progenitor identity. Neuron *39*, 749-765.

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

Guy, J., Gan, J., Selfridge, J., Cobb, S., and Bird, A. (2007). Reversal of neurological defects in a mouse model of Rett syndrome. Science *315*, 1143-1147.

Hakem, R., Hakem, A., Duncan, G.S., Henderson, J.T., Woo, M., Soengas, M.S., Elia, A., de la Pompa, J.L., Kagi, D., Khoo, W.*, et al.* (1998). Differential requirement for caspase 9 in apoptotic pathways in vivo. Cell *94*, 339-352.

Han, K., Sen, S.K., Wang, J., Callinan, P.A., Lee, J., Cordaux, R., Liang, P., and Batzer, M.A. (2005). Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages. Nucleic acids research *33*, 4040-4052.

Harcourt-Smith, W.E., and Aiello, L.C. (2004). Fossils, feet and the evolution of human bipedal locomotion. Journal of anatomy *204*, 403-416.

Hasan, M., Koch, J., Rakheja, D., Pattnaik, A.K., Brugarolas, J., Dozmorov, I., Levine, B., Wakeland, E.K., Lee-Kirsch, M.A., and Yan, N. (2013). Trex1 regulates lysosomal biogenesis and interferon-independent activation of antiviral genes. Nature immunology *14*, 61-71.

Hattori, M., Kuhara, S., Takenaka, O., and Sakaki, Y. (1986). L1 family of repetitive DNA sequences in primates may be derived from a sequence encoding a reverse transcriptase-related protein. Nature *321*, 625-628.

Heckroth, J.A., Goldowitz, D., and Eisenman, L.M. (1989). Purkinje cell reduction in the reeler mutant mouse: a quantitative immunohistochemical study. The Journal of comparative neurology *279*, 546-555.

Hohjoh, H., and Singer, M.F. (1996). Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. The EMBO journal *15*, 630-639.

Holliman, R.E. (1997). Toxoplasmosis, behaviour and personality. The Journal of infection *35*, 105-110.

Holmes, S.E., Dombroski, B.A., Krebs, C.M., Boehm, C.D., and Kazazian, H.H., Jr. (1994). A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. Nature genetics *7*, 143-148.

Hsieh, J., Nakashima, K., Kuwabara, T., Mejia, E., and Gage, F.H. (2004). Histone deacetylase inhibition-mediated neuronal differentiation of multipotent adult neural progenitor cells. Proceedings of the National Academy of Sciences of the United States of America *101*, 16659-16664.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O.*, et al.* (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nature biotechnology *31*, 827-832.

Hsu, P.D.Z., F. (2013). CRISPR Design.

Ishii, K.J., Coban, C., Kato, H., Takahashi, K., Torii, Y., Takeshita, F., Ludwig, H., Sutter, G., Suzuki, K., Hemmi, H.*, et al.* (2006). A Toll-like receptor-independent antiviral response induced by double-stranded B-form DNA. Nature immunology *7*, 40-48.

Jones, R.B., Garrison, K.E., Wong, J.C., Duan, E.H., Nixon, D.F., and Ostrowski, M.A. (2008). Nucleoside analogue reverse transcriptase inhibitors differentially inhibit human LINE-1 retrotransposition. PloS one *3*, e1547.

Jung, A.R., Kim, T.W., Rhyu, I.J., Kim, H., Lee, Y.D., Vinsant, S., Oppenheim, R.W., and Sun, W. (2008). Misplacement of Purkinje cells during postnatal development in Bax knock-out mice: a novel role for programmed cell death in the nervous system?

The Journal of neuroscience : the official journal of the Society for Neuroscience *28*, 2941-2948.

Kazazian, H.H., Jr., and Moran, J.V. (1998). The impact of L1 retrotransposons on the human genome. Nature genetics *19*, 19-24.

Kazazian, H.H., Jr., Wong, C., Youssoufian, H., Scott, A.F., Phillips, D.G., and Antonarakis, S.E. (1988). Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature *332*, 164-166.

Khan, H., Smit, A., and Boissinot, S. (2006). Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. Genome research *16*, 78-87.

Kim, W.R., and Sun, W. (2011). Programmed cell death during postnatal development of the rodent nervous system. Development, growth & differentiation *53*, 225-235.

Kingsmore, S.F., Giros, B., Suh, D., Bieniarz, M., Caron, M.G., and Seldin, M.F. (1994). Glycine receptor beta-subunit gene mutation in spastic mouse associated with LINE-1 element insertion. Nature genetics *7*, 136-141.

Kohrman, D.C., Harris, J.B., and Meisler, M.H. (1996). Mutation detection in the med and medJ alleles of the sodium channel Scn8a. Unusual splicing due to a minor class AT-AC intron. The Journal of biological chemistry *271*, 17576-17581.

Kolosha, V.O., and Martin, S.L. (2003). High-affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1). The Journal of biological chemistry *278*, 8112-8117.

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. Proceedings of the National Academy of Sciences of the United States of America *103*, 8036-8041.

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

Kuwabara, T., Hsieh, J., Muotri, A., Yeo, G., Warashina, M., Lie, D.C., Moore, L., Nakashima, K., Asashima, M., and Gage, F.H. (2009). Wnt-mediated activation of NeuroD1 and retro-elements during adult neurogenesis. Nature neuroscience *12*, 1097-1105.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W.*, et al.* (2001a). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W.*, et al.* (2001b). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.

Larson, M.H., Gilbert, L.A., Wang, X., Lim, W.A., Weissman, J.S., and Qi, L.S. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nature protocols *8*, 2180-2196.

Lebon, P., Badoual, J., Ponsot, G., Goutieres, F., Hemeury-Cukier, F., and Aicardi, J. (1988). Intrathecal synthesis of interferon-alpha in infants with progressive familial encephalopathy. Journal of the neurological sciences *84*, 201-208.

Lee, J., Cordaux, R., Han, K., Wang, J., Hedges, D.J., Liang, P., and Batzer, M.A. (2007). Different evolutionary fates of recently integrated human and chimpanzee LINE-1 retrotransposons. Gene *390*, 18-27.

Lerman, M.I., Thayer, R.E., and Singer, M.F. (1983). Kpn I family of long interspersed repeated DNA sequences in primates: polymorphism of family members and evidence for transcription. Proceedings of the National Academy of Sciences of the United States of America *80*, 3966-3970.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Lin, C., Yang, L., Tanasa, B., Hutt, K., Ju, B.G., Ohgi, K., Zhang, J., Rose, D.W., Fu, X.D., Glass, C.K.*, et al.* (2009). Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. Cell *139*, 1069-1083.

Liu, G., Zhao, S., Bailey, J.A., Sahinalp, S.C., Alkan, C., Tuzun, E., Green, E.D., and Eichler, E.E. (2003). Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. Genome research *13*, 358-368.

Lotto, R.B., Asavaritikrai, P., Vali, L., and Price, D.J. (2001). Target-derived neurotrophic factors regulate the death of developing forebrain neurons after a change in their trophic requirements. The Journal of neuroscience : the official journal of the Society for Neuroscience *21*, 3904-3910.

Lu, J., Tapia, J.C., White, O.L., and Lichtman, J.W. (2009). The interscutularis muscle connectome. PLoS biology *7*, e32.

Luan, D.D., Korman, M.H., Jakubczak, J.L., and Eickbush, T.H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell *72*, 595-605.

Luikenhuis, S., Giacometti, E., Beard, C.F., and Jaenisch, R. (2004). Expression of MeCP2 in postmitotic neurons rescues Rett syndrome in mice. Proceedings of the National Academy of Sciences of the United States of America *101*, 6033-6038.

Maestre, J., Tchenio, T., Dhellin, O., and Heidmann, T. (1995). mRNA retroposition in human cells: processed pseudogene formation. The EMBO journal *14*, 6333-6338.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823-826.

Marchetto, M.C., Carromeu, C., Acab, A., Yu, D., Yeo, G.W., Mu, Y., Chen, G., Gage, F.H., and Muotri, A.R. (2010). A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. Cell *143*, 527-539.

Martin, D.A., and Elkon, K.B. (2006). Intracellular mammalian DNA stimulates myeloid dendritic cells to produce type I interferons predominantly through a toll-like receptor 9-independent pathway. Arthritis and rheumatism *54*, 951-962.

Martin, S.L., and Bushman, F.D. (2001). Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. Molecular and cellular biology *21*, 467-475.

Mathias, S.L., Scott, A.F., Kazazian, H.H., Jr., Boeke, J.D., and Gabriel, A. (1991). Reverse transcriptase encoded by a human transposable element. Science *254*, 1808-1810.

Mazur, D.J., and Perrino, F.W. (1999). Identification and expression of the TREX1 and TREX2 cDNA sequences encoding mammalian 3'-->5' exonucleases. The Journal of biological chemistry *274*, 19655-19660.

McClintock, B. (1950). The origin and behavior of mutable loci in maize. Proceedings of the National Academy of Sciences of the United States of America *36*, 344-355.

McConnell, M.J., Kaushal, D., Yang, A.H., Kingsbury, M.A., Rehen, S.K., Treuner, K., Helton, R., Annas, E.G., Chun, J., and Barlow, C. (2004). Failed clearance of aneuploid embryonic neural progenitor cells leads to excess aneuploidy in the Atm-deficient but not the Trp53-deficient adult cerebral cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience *24*, 8090-8096.

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 research *52*, 643-645.

Miller, J., McLachlan, A.D., and Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. The EMBO journal *4*, 1609-1614.

Morales, J.F., Snow, E.T., and Murnane, J.P. (2002). Environmental factors affecting transcription of the human L1 retrotransposon. I. Steroid hormone-like agents. Mutagenesis *17*, 193-200.

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.

Morita, M., Stamp, G., Robins, P., Dulic, A., Rosewell, I., Hrivnak, G., Daly, G., Lindahl, T., and Barnes, D.E. (2004). Gene-targeted mice lacking the Trex1 (DNase III) 3'-->5' DNA exonuclease develop inflammatory myocarditis. Molecular and cellular biology *24*, 6719-6727.

Mulhardt, C., Fischer, M., Gass, P., Simon-Chazottes, D., Guenet, J.L., Kuhse, J., Betz, H., and Becker, C.M. (1994). The spastic mouse: aberrant splicing of glycine receptor beta subunit mRNA caused by intronic insertion of L1 element. Neuron *13*, 1003-1015.

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., Marchetto, M.C., Coufal, N.G., Oefner, R., Yeo, G., Nakashima, K., and Gage, F.H. (2010). L1 retrotransposition in neurons is modulated by MeCP2. Nature *468*, 443-446.

Muotri, A.R., Zhao, C., Marchetto, M.C., and Gage, F.H. (2009). Environmental influence on L1 retrotransposons in the adult hippocampus. Hippocampus *19*, 1002-1007.

Narita, N., Nishio, H., Kitoh, Y., Ishikawa, Y., Ishikawa, Y., Minami, R., Nakamura, H., and Matsuo, M. (1993). Insertion of a 5' truncated L1 element into the 3' end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. The Journal of clinical investigation *91*, 1862-1867.

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. Nature genetics *32*, 655-660.

Patel, R.K., and Jain, M. (2012). NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PloS one *7*, e30619.

Pereira-Lopes, S., Celhar, T., Sans-Fons, G., Serra, M., Fairhurst, A.M., Lloberas, J., and Celada, A. (2013). The exonuclease Trex1 restrains macrophage proinflammatory activation. Journal of immunology *191*, 6128-6135.

Perou, C.M., Pryor, R.J., Naas, T.P., and Kaplan, J. (1997). The bg allele mutation is due to a LINE1 element retrotransposition. Genomics *42*, 366-368.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173-1183.

Rass, U., Ahel, I., and West, S.C. (2007). Defective DNA repair and neurodegenerative disease. Cell *130*, 991-1004.

Reed, K.E. (1997). Early hominid evolution and ecological change through the African Plio-Pleistocene. Journal of human evolution *32*, 289-322.

Reznikoff, W.S. (2003). Tn5 as a model for understanding DNA transposition. Molecular microbiology *47*, 1199-1206.

Rhesus Macaque Genome, S., Analysis, C., Gibbs, R.A., Rogers, J., Katze, M.G., Bumgarner, R., Weinstock, G.M., Mardis, E.R., Remington, K.A., Strausberg, R.L.*, et al.* (2007). Evolutionary and biomedical insights from the rhesus macaque genome. Science *316*, 222-234.

Rice, G.I., Bond, J., Asipu, A., Brunette, R.L., Manfield, I.W., Carr, I.M., Fuller, J.C., Jackson, R.M., Lamb, T., Briggs, T.A.*, et al.* (2009). Mutations involved in Aicardi-Goutieres syndrome implicate SAMHD1 as regulator of the innate immune response. Nature genetics *41*, 829-832.

Rice, G.I., Kasher, P.R., Forte, G.M., Mannion, N.M., Greenwood, S.M., Szynkiewicz, M., Dickerson, J.E., Bhaskar, S.S., Zampini, M., Briggs, T.A.*, et al.* (2012). Mutations in ADAR1 cause Aicardi-Goutieres syndrome associated with a type I interferon signature. Nature genetics *44*, 1243-1248.

Richards, A., van den Maagdenberg, A.M., Jen, J.C., Kavanagh, D., Bertram, P., Spitzer, D., Liszewski, M.K., Barilla-Labarca, M.L., Terwindt, G.M., Kasai, Y.*, et al.* (2007). C-terminal truncations in human 3'-5' DNA exonuclease TREX1 cause autosomal dominant retinal vasculopathy with cerebral leukodystrophy. Nature genetics *39*, 1068-1070.

Robinson, J.T. (1972). Early hominid posture and locomotion (Chicago,: University of Chicago Press).

Rousseau-Merck, M.F., Koczan, D., Legrand, I., Moller, S., Autran, S., and Thiesen, H.J. (2002). The KOX zinc finger genes: genome wide mapping of 368 ZNF PAC clones with zinc finger gene clusters predominantly in 23 chromosomal loci are confirmed by human sequences annotated in EnsEMBL. Cytogenetic and genome research *98*, 147-153.

Ryan, R.F., Schultz, D.C., Ayyanathan, K., Singh, P.B., Friedman, J.R., Fredericks, W.J., and Rauscher, F.J., 3rd (1999). KAP-1 corepressor protein interacts and colocalizes with heterochromatic and euchromatic HP1 proteins: a potential role for

Kruppel-associated box-zinc finger proteins in heterochromatin-mediated gene silencing. Molecular and cellular biology *19*, 4366-4378.

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. Nature genetics *16*, 37-43.

Savitsky, K., Bar-Shira, A., Gilad, S., Rotman, G., Ziv, Y., Vanagaite, L., Tagle, D.A., Smith, S., Uziel, T., Sfez, S.*, et al.* (1995). A single ataxia telangiectasia gene with a product similar to PI-3 kinase. Science *268*, 1749-1753.

Schwahn, U., Lenzner, S., Dong, J., Feil, S., Hinzmann, B., van Duijnhoven, G., Kirschner, R., Hemberger, M., Bergen, A.A., Rosenberg, T.*, et al.* (1998). Positional cloning of the gene for X-linked retinitis pigmentosa 2. Nature genetics *19*, 327-332.

Skene, P.J., Illingworth, R.S., Webb, S., Kerr, A.R., James, K.D., Turner, D.J., Andrews, R., and Bird, A.P. (2010). Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. Molecular cell *37*, 457-468.

Skowronski, J., and Singer, M.F. (1985). Expression of a cytoplasmic LINE-1 transcript is regulated in a human teratocarcinoma cell line. Proceedings of the National Academy of Sciences of the United States of America *82*, 6050-6054.

Skowronski, J., and Singer, M.F. (1986). The abundant LINE-1 family of repeated DNA sequences in mammals: genes and pseudogenes. Cold Spring Harbor symposia on quantitative biology *51 Pt 1*, 457-464.

Smit, A.F., Toth, G., Riggs, A.D., and Jurka, J. (1995). Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. Journal of molecular biology *246*, 401-417.

Smith, C., Gore, A., Yan, W., Abalde-Atristain, L., Li, Z., He, C., Wang, Y., Brodsky, R.A., Zhang, K., Cheng, L.*, et al.* (2014). Whole-Genome Sequencing Analysis Reveals High Specificity of CRISPR/Cas9 and TALEN-Based Genome Editing in Human iPSCs. Cell stem cell *15*, 12-13.

Speek, M. (2001). Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Molecular and cellular biology *21*, 1973-1985.

Stetson, D.B., Ko, J.S., Heidmann, T., and Medzhitov, R. (2008). Trex1 prevents cell-intrinsic initiation of autoimmunity. Cell *134*, 587-598.

Stetson, D.B., and Medzhitov, R. (2006). Recognition of cytosolic DNA activates an IRF3-dependent innate immune response. Immunity *24*, 93-103.

Stout, D., Toth, N., Schick, K., and Chaminade, T. (2008). Neural correlates of Early Stone Age toolmaking: technology, language and cognition in human evolution. Philosophical transactions of the Royal Society of London Series B, Biological sciences *363*, 1939-1949.

Strait, D.S., Grine, F.E., and Moniz, M.A. (1997). A reappraisal of early hominid phylogeny. Journal of human evolution *32*, 17-82.

Streubel, J., Blucher, C., Landgraf, A., and Boch, J. (2012). TAL effector RVD specificities and efficiencies. Nature biotechnology *30*, 593-595.

Takahara, T., Ohsumi, T., Kuromitsu, J., Shibata, K., Sasaki, N., Okazaki, Y., Shibata, H., Sato, S., Yoshiki, A., Kusakabe, M.*, et al.* (1996). Dysfunction of the Orleans reeler gene arising from exon skipping due to transposition of a full-length copy of an active L1 sequence into the skipped exon. Human molecular genetics *5*, 989-993.

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

van Praag, H., Kempermann, G., and Gage, F.H. (2000). Neural consequences of environmental enrichment. Nature reviews Neuroscience *1*, 191-198.

Veres, A.G., B.S.; Ding, Q.; Collins, R.; Ragavendran, A.; Brand, H.; Erdin, S.;, and Talkowski, M.E.M., K (2014). Low Incidence of Off-Target Mutations in Individual CRISPR-Cas9 and TALEN Targeted Human Stem Cell Clones Detected by Whole-Genome
Sequencing. Cell stem cell *15*, 27-30.

Vogel, M.W., Sunter, K., and Herrup, K. (1989). Numerical matching between granule and Purkinje cells in lurcher chimeric mice: a hypothesis for the trophic rescue of granule cells from target-related cell death. The Journal of neuroscience : the official journal of the Society for Neuroscience *9*, 3454-3462.

Von Bartheld, C.S., and Johnson, J.E. (2001). Target-derived BDNF (brain-derived neurotrophic factor) is essential for the survival of developing neurons in the isthmo-optic nucleus. The Journal of comparative neurology *433*, 550-564.

Vyas, A., Kim, S.K., Giacomini, N., Boothroyd, J.C., and Sapolsky, R.M. (2007a). Behavioral changes induced by Toxoplasma infection of rodents are highly specific to aversion of cat odors. Proceedings of the National Academy of Sciences of the United States of America *104*, 6442-6447.

Vyas, A., Kim, S.K., and Sapolsky, R.M. (2007b). The effects of toxoplasma infection on rodent behavior are dependent on dose of the stimulus. Neuroscience *148*, 342-348.

Wang, H., Xing, J., Grover, D., Hedges, D.J., Han, K., Walker, J.A., and Batzer, M.A. (2005). SVA elements: a hominid-specific retroposon family. Journal of molecular biology *354*, 994-1007.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research *38*, e164.

Warren, W.C., Hillier, L.W., Marshall Graves, J.A., Birney, E., Ponting, C.P., Grutzner, F., Belov, K., Miller, W., Clarke, L., Chinwalla, A.T.*, et al.* (2008). Genome analysis of the platypus reveals unique signatures of evolution. Nature *453*, 175-183.

Wiedenheft, B., Sternberg, S.H., and Doudna, J.A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. Nature *482*, 331-338.

Wolfe, S.A., Nekludova, L., and Pabo, C.O. (2000). DNA recognition by Cys2His2 zinc finger proteins. Annual review of biophysics and biomolecular structure *29*, 183-212.

Wolff, E.M., Byun, H.M., Han, H.F., Sharma, S., Nichols, P.W., Siegmund, K.D., Yang, A.S., Jones, P.A., and Liang, G. (2010). Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer. PLoS genetics *6*, e1000917.

Woods-Samuels, P., Wong, C., Mathias, S.L., Scott, A.F., Kazazian, H.H., Jr., and Antonarakis, S.E. (1989). Characterization of a nondeleterious L1 insertion in an intron of the human factor VIII gene and further evidence of open reading frames in functional L1 elements. Genomics *4*, 290-296.

Yajima, I., Sato, S., Kimura, T., Yasumoto, K., Shibahara, S., Goding, C.R., and Yamamoto, H. (1999). An L1 element intronic insertion in the black-eyed white (Mitf[mi-bw]) gene: the loss of a single Mitf isoform responsible for the pigmentary defect and inner ear deafness. Human molecular genetics *8*, 1431-1441.

Yan, N., Regalado-Magdos, A.D., Stiggelbout, B., Lee-Kirsch, M.A., and Lieberman, J. (2010). The cytosolic exonuclease TREX1 inhibits the innate immune response to human immunodeficiency virus type 1. Nature immunology *11*, 1005-1013.

Yang, Y.G., Lindahl, T., and Barnes, D.E. (2007). Trex1 exonuclease degrades ssDNA to prevent chronic checkpoint activation and autoimmune disease. Cell *131*, 873-886.

Yu, F., Zingler, N., Schumann, G., and Stratling, W.H. (2001). Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. Nucleic acids research *29*, 4493-4501.

Yu, J., Hu, K., Smuga-Otto, K., Tian, S., Stewart, R., Slukvin, II, and Thomson, J.A. (2009). Human induced pluripotent stem cells free of vector and transgene sequences. Science *324*, 797-801.

Yuan, S.H., Martin, J., Elia, J., Flippin, J., Paramban, R.I., Hefferan, M.P., Vidal, J.G., Mu, Y., Killian, R.L., Israel, M.A.*, et al.* (2011). Cell-surface marker signatures for the isolation of neural stem cells, glia and neurons derived from human pluripotent stem cells. PloS one *6*, e17540.

Zhang, J., Sun, X., Qian, Y., LaDuca, J.P., and Maquat, L.E. (1998). At least one intron is required for the nonsense-mediated decay of triosephosphate isomerase

mRNA: a possible link between nuclear splicing and cytoplasmic translation. Molecular and cellular biology *18*, 5272-5283.