[Frontiers in Bioscience E4, 1663-1668, January 1, 2012]

LINE-1: creators of neuronal diversity

Charles A. Thomas[1], Alysson R. Muotri[1]

[1]Department of Pediatrics, Rady Children's Hospital San Diego and Department of Cellular and Molecular Medicine, University of California, San Diego School of Medicine, La Jolla, CA 92093

TABLE OF CONTENTS

1. ABSTRACT

Long interspersed nucleotide element 1 (L1) is a family of non-LTR retrotransposons that can replicate and reintegrate into the host genome. L1s have considerably influenced mammalian genome evolution by retrotransposing during germ cell development or early embryogenesis, leading to massive genome expansion. In humans, over 30% of the genome can be attributed to L1-mediated retrotransposition. Historically, L1s were thought to only retrotranspose during gametogenesis and in neoplastic processes, but recent studies have shown that L1s are extremely active in the mouse, rat, and human neuronal progenitor cells (NPCs). In fact, it is estimated that the hippocampus and other regions of the brain may have multiple insertions per cell. These insertions can dramatically impact neuronal transcriptional expression, creating unique transcriptomes of individual neurons. Furthermore, transcriptional activation of L1 elements mimics the transcription activation of the NeuroD1 gene, suggesting a prominent role of L1 expression during neurogenesis.

2. INTRODUCTION

Discovered by Barbara McClintock in 1940s (1), mobile DNA elements are repetitive sequences capable of "moving" and reinserting into the host genome. Class II DNA transposons, the mobile DNA elements discovered by McClintock, utilize a "cut-and-paste" mechanism to confer mobility. The element can excise itself from the host's genome and integrate in a new location (2). Such activity can impact gene expression, affecting color in maize kernels for example. Retrotransposons, on the other hand, use a "copy-and-paste" mechanism, which enables the retroelement to duplicate each integration event (3). Retrotransposons are transcribed into RNA intermediates and must utilize a reverse transcriptase activity to reintegrate into the host genome. Similar to DNA transposons, new retrotransposon insertions can also affect gene expression, generating a genetic mosaicism in multicellular organisms. The impact of these de novo insertions in the genome is probably underestimated due to the paucity of data. However, it is clear the multiplicity effect of the L1 replication mechanism, retrotransposons

compromise an overwhelming majority of mobile DNA elements in eukaryotic genomes (4). Of particular interest is the LINE family, the most prominent autonomous mobile element in mammalian genomes.

2.1. L1 prominence in mammalian genomes

In mammals, the LINE family has expanded immensely and is the single greatest constituent of genomic DNA, accounting for approximately 20% of the entire genome in humans, chimpanzees, mice, and platypuses alike (5, 6). Small Interspersed Nuclear Elements (SINEs) account for another ~10% of mammalian genomes. SINEs are short sequence retroposons (~300 bp) that hijack the proteins expressed by L1 in order to retrotranspose. Therefore, about 30% of mammalian genomes can be attributed to LINE-mediated expansion, signifying the importance of L1 in mammalian genome evolution.

Although mammalian genomes are highly abundant in LINE and SINE sequences, only a small subset of these transposable elements are full length, free of mutations and able to retrotranspose. In humans, it is postulated that of the 516,000 L1 sequences, approximately 100 are currently active (7). In comparison, approximately 3,000 putatively active L1 elements exist within the mouse genome (8). It should be noted, however, the rate of generational L1 integration for human is the same as the mouse, even though the mouse has 30 times more putatively active elements (9).

2.2. L1 structure

An L1 element is approximately 6 kb long, composed of 910 bp 5'UTR with promoter-like properties, two open reading frames (ORF1 and ORF2), and a 205 bp 3'UTR that encodes a functional polyadenylation signal (10). ORF1 is a RNA chaperone and probably assists on annealing of complementary strands, and removal of mispaired DNA duplexes (11). ORF2 contains a functional endonuclease (EN) (12) and reverse transcriptase (RT) domain (13). Translation of ORF1 and ORF2 is necessary for L1 to integrate back into the genome (14). When expressed and translated, L1s can integrate into the host genome by the target-primed reverse transcription mechanism (15). In general, L1s nick and integrate at 5'-TTTT/AA-3' canonical sites in the host's genome. Newly inserted L1 elements typically consist of an intact 3' region including a variable length in the polyA tail, but are often truncated at the 5' end, possibly due to incomplete RT activity, and always flanked in between target site duplications (16). For more detailed information on the structure, biology, retrotransposition mechanism, reviewed in 17, 18.

2.3. L1 retrotransposition modulates gene function

When L1 elements reinsert into the host genome, they can alter gene transcription by a variety of mechanisms. Insertion within an exon can cause frameshift mutations, premature stop codons, or exon-skipping. In some instances, an integration event can cause a large genomic deletion at the site of insertion, thus removing part of an exon or regulatory sequence (19). Likewise, if L1 inserts within a transcribed portion of a gene, it can decrease the gene's mRNA levels by causing the transcriptional machinery to slow down or abort due to the high A/T content in ORF2 (20). L1s can create premature polyadenylation sites (21) or chromatin modifications due to methylation of the CpG islands in the 5'UTR (22). Alternatively, the promoter capabilities of L1 can increase gene transcription or create new transcription start sites when inserted into untranslated portions of genes (23). L1 insertions can also create new isoforms by introducing new splice sites (24). In summary, de novo L1 integrations can cause mutagenesis by a variety of mechanism, shaping the genome and impacting nearby gene expression.

3. L1 RETROTRANSPOSITION IN THE BRAIN

The human brain consists of approximately 86.1 billion neurons and 84.6 non-neuronal cells (25). In addition, the typical mature neuron makes 5,000-20,000 synaptic connections (26), creating an elaborate and complex network. Furthermore, the mammalian brain is diversified with many neuronal types and subtypes. Although the true definitions of a neuronal type and subtype are debated, the number of subtypes is far too great to be estimated with current knowledge and methodologies (27). However, current definitions of neuronal types and subtypes do not even take into consideration differences of molecular expression levels, connectivity, and electrophysiology of individual cells within cell types. It is difficult to attribute the origin of neuronal diversity solely on spatial and temporal gene regulation. In fact, the nervous system uses several molecular mechanisms to generate neuronal diversity (28). These mechanisms may include aneuploidy, promoter usage, alternative RNA splicing, alternative polyadenylation, RNA editing, epigenetics, and L1 retrotransposition. Combinations of these mechanisms would likely create individualistic functions in distinct neurons. L1 retrotransposition is extremely intriguing because it can significantly interfere with many of the other mechanisms listed above, including promoter usage, RNA splicing, polyadenylation, and epigenetics.

3.1. L1 retrotransposition in neuronal progenitor cells

It has been previously shown that L1 is highly expressed and capable of retrotransposing in neuronal progenitor cells (29). To visualize retrotransposition, a construct consisting of L1 and the eGFP indicator cassette was transfected into adult rat hippocampal neural progenitor cells (NPCs). This construct is designed to fluoresce only after the L1 construct retrotransposes back into the genome. Several NPCs fluoresced, whereas no fluorescence was seen in other somatic cell types, such as mesenchymal stem cells and fibroblasts. PCR of genomic DNA confirmed the retrotransposition. These data show the capacity of L1s to retrotranspose in rat NPCs, and has been confirmed both *in vivo* with transgenic mice (29) and *in vitro* using human NPCs derived from embryonic stem cells (30).

The major advantage of the transgenic mouse expressing the L1-eGFP indicator cassette is the ability to trace the lineage of which cells have undergone

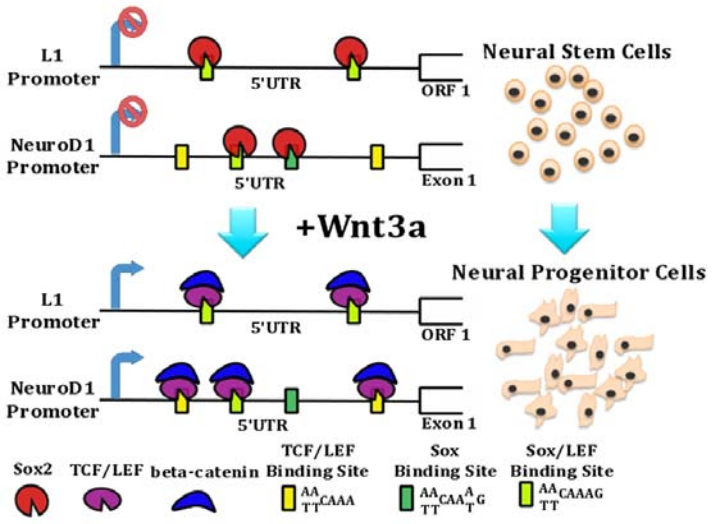

Figure 1. Schematic of the molecular mechanism regulating L1 retrotransposition during neurogenesis. NeuroD1 and L1 transcription are controlled by Sox2 and TCF/LEF sites of the 5'UTR. In the neural stem cell state, Sox2 binds to the Sox sites and inhibits transcription. Stimulation by Wnt3a induces the recruitment of beta-catenin and TCF/LEF. Binding of TCF/LEF induces the transcription and subsequent expression of NeuroD1 and L1, accompanying the neural stem cell

retrotransposition (31). Previous studies with the L1-eGFP transgenic mouse detected L1 retrotransposition in testis but did not closely examine the NPCs of the brain (32). In contrast, using a non-tissue specific reporter, L1 retrotransposition was detected at relatively frequent rates in the brain in addition to the testis (29). Other tissues were examined, including the kidney, intestine, heart, liver, spleen, lung, skin, and muscle, but no indication of retrotransposition existed. Analyses of the transgenic mice also show multiple neuron types derived at different times of development display evidence of retrotransposition, as evidenced by co-localization with the neuronal maker NeuN. No eGFP-positive cells co-localized with astrocyte marker S100-beta or oligodendrocyte marker glutathione S-transferase pi. By analysis of embryonic brains from the L1-eGFP transgenic mice and estimating from neuroatomical adult brain regions, it was possible to conclude that L1 retrotransposition begins around E10 and continues postnatally. These results demonstrate L1 retrotransposition is probably specific to the cells of neuronal lineage and can occur during either embryonic or adult neurogenesis.

Although the L1-eGFP transgenic mice display the ability of L1 to retrotanspose in NPCs, the question of whether endogenous LINEs retrotransposed in the brain still remained. To answer this query, a multiplex qPCR analysis on genomes derived from human postmortem tissue was performed (30). The experiment was set up to compare the copy number of L1 elements in brain, heart, and liver tissues. The results demonstrated brain tissue contained approximately 80 more L1 element sequences per cellular genome than tissue of the heart and liver. The copy number quantification of the qPCR results, generated by a standard plasmid-based comparison, does not completely reproduce the genomic structure. Moreover, an increase number in L1 sequences may not necessarily reflects and activity of L1 retrotransposition. It is still possible that other somatic mechanism could account for the expansion of L1 sequences in the brain. Definite proof of new L1 insertions from sequencing techniques will help to elucidate this question.

As noted earlier, L1 integration within or near a gene can dramatically change its expression pattern. For example, when an L1 element inserted within the 5'UTR of the post-synaptic density gene Psd-93 of a rat NPCs, the PSD-93 expression was altered, affecting cell fate (29). During neuronal differentiation, Psd-93 began expression earlier and at greater levels in the NPCs. Furthermore, the NPCs overexpressing PSD-93 differentiated almost exclusively into neurons, whereas normal differentiation produces similar amounts of neurons and glia. This data shows a single integration event can dramatically affect a cell's expression profile and fate. Considering the fact that an average of 80 L1 integration events occur per neural cell in humans (30), we hypothesize L1-mediated integration would create an extraordinary level of inter-neuronal diversity.

3.2. Control of L1 activation and expression in neuronal cells

Understanding the regulatory elements of the L1 transcript would likely yield insights into why L1 is specifically active in NPCs. Recently, Kuwubara *et al.* discovered that Wnt3a stimulation on neural stem cells increases L1 expression in about 10-fold via the beta-catenin pathway (33). Overlapping Sox and TCF/LEF (Sox/LEF) sites within the 5' UTR and ORF2 protein of LINE-1 mediate the upregulation of expression (Figure 1). Furthermore, the Sox/LEF sites exist in the mouse, rat, and human L1, suggesting a conserved role among all mammals.

Fibroblast growth factor 2 signaling keeps neural stem cells in a proliferative, undifferentiated state. Sox2 associates to the Sox/LEF sites, repressing L1 expression (29). Wnt3a stimulation induces the neural stem cell to differentiate into NPCs. As the neural stem cell progresses into the progenitor stage, Sox2 expression is lost concomitant with increase of L1 expression. Interestingly, Wnt3a activates transcription of NeuroD1 in a similar manner as L1. NeuroD1 is a major transcription factor that promotes neurogenesis by activating many proneural genes. The promoter region of NeuroD1 has Sox/LEF sites, similar to the 5'UTR of L1. Likewise, the temporal expression pattern of NeuroD1 and L1 mirror each other during neuronal differentiation. The correlation between NeuroD1 and L1 insinuates that L1 expression is of utmost importance during neurogenesis.

The chromatin modifier methyl-CpG binding protein 2 (MeCP2) is another key regulator of L1 in neurons (34). In a neural stem cell state, the MeCP2 protein occupies to the promoter of L1, preventing transcription. Upon differentiation, MeCP2 releases from the promoter, allowing L1 transcription to take place. In mature neurons, MeCP2 reoccupies the L1 promoter, preventing spurious expression (35). Mutations that lead to alterations in MeCP2 expression or loss of MeCP2 protein function are the cause of Rett Syndrome (36). In neurons derived from iPSCs of Rett patients, we have detected increased expression and retrotransposition of L1 (34). Therefore, we propose loss of regulation of the neuronal retrotransposition could contribute to the pathology or heterogeneity of Rett Syndrome (37).

4. SUMMARY AND PERSPECTIVE

Mammals, and especially humans, are often described as having individualistic behaviors. Monozygotic twins, although alike in appearance and physical build, typically have unique personalities even when reared in the same environment (38). The work of Tellegen *et al.*, and other similar studies, have deducted personality is 50% determined by genetics, 50% determined by "environment." However, rearing monozygotic twins in virtually the same environment yielded only 0-10% affect on personality. Therefore, 40-50% of the personality determined by "environment" is actually due to something other than the environment. In similar experiments, studies were performed to determine where behavior variability in inbred laboratory mice arises (39, 40). The studies concluded that behavior variability in inbred mice is due to a complex set of factors, including the possibility of differences in repetitive sequences, generating stochastic variability within cells. This variability phenomenon is known in genetics as "incomplete penetrance", indicating the existence of a mechanism that generates diversity even among genetically identical individuals exposed to exactly the same environment. Most of the times, such stochastic process is considered noise and frequently overlooked. However, more studies are realizing that the random process is an important part of biological complex organisms. In fact, random fluctuations in gene expression can reach some threshold level to cause an outcome (41, 42). Unfortunately, cellular outcomes are not always visible and experimental validation of stochastic processes is scarce, with few examples that lead to a more dramatic readout (43, 44). We hypothesize that mammalian behavioral variability between individuals is in part due to neuronal variability mediated by L1 somatic retrotransposition. Because de novo L1 integrations occur per neuron, each neuron is going to express a mildly distinct transcriptome. Furthermore, since L1 integration seems to insert within neuronal genes at random, each mammalian individual will have a distinct combination of neurons, which, in turn will dictate the unique networks and circuits, creating individual behaviors.

5. ACKNOWLEDGEMENTS

The work was supported by startup funds from the University of California San Diego Stem Cell Research Program, the Emmerald Foundation and by the National Institutes of Health through the NIH Director's New Innovator Award Program, 1-DP2-OD006495-01. We would like to thank members of the Muotri's lab and Fred H. Gage and Maria C.N. Marchetto, from the Salk Institute for a critical reading of our manuscript.

6. REFERENCES

1. B McClintock: Controlling elements and the gene. *Cold Spring Harb Symp Quant Biol*, 21, 197-216 (1956)

2. WS Reznikoff: Tn5 as a model for understanding DNA transposition. *Mol Microbiol*, 475, 1199-1206 (2003)

3. G Liu, S Zhao, JA Bailey, SC Sahinalp, C Alkan, E Tuzun, ED Green, EE Eichler: Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. *Genome Res*, 133, 358-368 (2003)

4. AF Smit: The origin of interspersed repeats in the human genome. *Curr Opin Genet Dev*, 66, 743-748 (1996)

5. PK Mandal, HH Kazazian Jr: SnapShot: Vertebrate transposons. *Cell*, 1351,192 (2008)

6. J Lee, R Cordaux, K Han, J Wang, DJ Hedges, P Liang, MA Batzer: Different evolutionary fates of recently integrated human and chimpanzee LINE-1 retrotransposons. *Gene*, 390(1-2),18-27 (2007)

7. B Brouha, J Schustak, RM Badge, S Lutz-Prigge, AH Farley, JV Moran, HH Kazazian Jr: Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci USA*, 1009, 5280-5285 (2003)

8. JL Goodier, EM Ostertag, K Du, HH Kazazian Jr: A novel active L1 retrotransposon subfamily in the mouse. *Genome Res*, 1110, 1677-1685 (2001)

9. S Boissinot, P Chevret, AV Furano: L1 retrotransposon evolution and amplification in recent human history. *Mol Biol Evol*, 176, 915-928 (2000)

10. DV Babushok, HH Kazazian Jr.: Progress in understanding the biology of the human mutagen LINE-1. *Hum Mutat*, 28, 527-539 (2007)

11. SL Martin, FD Bushman: Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol Cell Biol*, 212, 467-475 (2001)

12. Q Feng, JV Moran, HH Kazazian Jr, J D Boeke: Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*, 875, 905-916 (1996)

13. SL Mathias, AF Scott, HH Kazazian Jr, JD Boeke, A Gabriel: Reverse transcriptase encoded by a human transposable element. *Science*, 254(5039), 1808-1810 (1991)

14. JV Moran, SE Holmes, TP Naas, RJ DeBerardinis, JD Boeke, HH Kazazian Jr: High frequency retrotransposition in cultured mammalian cells. *Cell*, (875), 917-927 (1996)

15. DD Luan, MH Korman, JL Jakubczak, TH Eickbush: Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell*, 724, 595-605 (1993)

16. ST Szak, OK Pickeral, W Makalowski, MS Boguski, D Landsman, JD Boeke: Molecular archeology of L1 insertions in the human genome. *Genome Biol*, 3(10), (2002)

17. AR Muotri, MC Marchetto, NG Coufal, FH Gage: The necessary junk: new functions for transposable elements. *Hum Mol Genet*, 16, R159-R167 (2007)

18. JL Goodier, HH Kazazian Jr: Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell*, 1351, 23-35 (2008)

19. N Gilbert, S Lutz-Prigge, JV Moran: Genomic deletions created upon LINE-1 retrotransposition. *Cell*, 1103, 315-325 (2002)

20. JS Han, ST Szak, JD Boeke: Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature*, 429(6989), 268-274 (2004)

21. V Perepelitsa-Belancio, P Deininger: RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat Genet*, 354, 363-366 (2003)

22. WA Schulz, C Steinhoff, AR Florl: Methylation of endogenous human retroelements in health and disease. Curr Top *Microbiol Immunol*, 310, 211-250 (2006)

23. M Speek: Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol Cell Biol*, 216, 1973-1985 (2001)

24. VP Belancio, DJ Hedges, P Deininger: LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic Acids Res*, 345, 1512-1521 (2006)

25. FA Azevedo, LR Carvalho, LT Grinberg, JM Farfel, RE Ferretti, RE Leite, W Jacob Filho, R Lent, S Herculano-Houzel: Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J Comp Neurol*, 5135, 532-541 (2009)

26. B Pakkenberg, D Pelvig, L Marner, MJ Bundgaard, HJ Gundersen, JR Nyengaard, L Regeur: Aging and the human neocortex. *Exp Gerontol*, 38(1-2), 95-99 (2003)

27. RH Masland: Neuronal cell types. *Curr Biol*, 1413, R497-R500 (2004)

28. AR Muotri, FH Gage: Generation of neuronal variability and complexity. *Nature*, 441(7097), 1087-1093 (2006)

29. AR Muotri, VT Chu, MC Marchetto, W Deng, JV Moran, FH Gage: Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature*, 435(7044), 903-910 (2005)

30. NG Coufal, JL Garcia-Perez, GE Peng, GW Yeo, Y Mu, MT Lovci, M Morell, KS O'Shea, JV Moran, FH Gage: L1 retrotransposition in human neural progenitor cells. *Nature*, 460(7259), 1127-1131 (2009)

31. ET Prak, AW Dodson, EA Farkash, HH Kazazian Jr: Tracking an embryonic L1 retrotransposition event. *Proc Natl Acad Sci USA*, 1004, 1832-1837 (2003)

32. EM Ostertag, RJ DeBerardinis, JL Goodier, Y Zhang, N Yang, GL Gerton HH Kazazian Jr: A mouse model of human L1 retrotransposition. *Nat Genet*, 324, 655-660 (2002)

33. T Kuwabara, J Hsieh, A Muotri, G Yeo, M Warashina, DC Lie, L Moore, K Nakashima, M Asashima and FH Gage: Wnt-mediated activation of NeuroD1 and retro-elements during adult neurogenesis. *Nat Neurosci*, 129, 1097-1105 (2009)

34. AR Muotri, MC Marchetto, NG Coufal, R Oefner, G Yeo, K Nakashima, FH Gage: L1 retrotransposition in neurons is modulated by MeCP2. Nature, 468(7322):443-446 (2010)

35. PJ Skene, RS Illingworth, S Webb, AR Kerr, KD James, DJ Turner, R Andrews, AP Bird: Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Mol Cell*, 37(4), 457-468 (2010)

36. RE Amir, IB Van den Veyver, M Wan, CQ Tran, U Francke, HY Zoghbi: Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet*, 23(2), 185-188, (1999)

37. MC Marchetto, C Carromeu, A Acab, D Yu, GW Yeo, Y Mu, G Chen, FH Gage, AR Muotri: A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell*, 143(4), 527-539 (2010)

38. A Tellegen, DT Lykken, TJ Bouchard Jr, KJ Wilcox, NL Segal, S Rich: Personality similarity in twins reared

apart and together. *J Pers Soc Psychol*, 546, 1031-1039 (1988)

39. M Jakovcevski, M Schachner, F Morellini: Individual variability in the stress response of C57BL/6J male mice correlates with trait anxiety. *Genes Brain Behav*, 72, 235-243 (2008)

40. R Lathe: The individuality of mice. *Genes Brain Behav*, 36, 317-327 (2004)

41. MB Elowitz, AJ Levine, ED Siggia, PS Swain: Stochastic gene expression in a single cell. *Science*, 297(5584), 1183-1186 (2002)

42. JM Raser, EK O'Shea: Control of stochasticity in eukaryotic gene expression. *Science*, 304(5678), 1811-1814 (2004)

43. A Raj, SA Rifkin, E Andersen, A van Oudenaarden: Variability in gene expression underlies incomplete penetrance. *Nature*, 463(7283), 913-918 (2010)

44. T Shin, D Kraemer, J Pryor, L Liu, J Rugila, L Howe, S Buck, K Murphy, L Lyons, M Westhusin: A cat cloned by nuclear transplantation. *Nature*, 415(6874), 859 (2002)

Abbreviations: LINE: long interspersed nucleotide element, L1: long interspersed nucleotide element-1, NPC: neural progenitor cell, SINE: short interspersed nucleotide element

Key Words: LINE-1, L1, Neurogenesis, Neuronal Diversity, Retrotransposition, Review

Send correspondence to: Alysson R. Muotri, University of California San Diego, Department of Pediatrics, Rady Children's Hospital San Diego, Department of Cellular and Molecular Medicine, 9500 Gilman Drive, CMM-E, La Jolla, CA 92093, MC 0695, USA, Tel: 858- 34-9320, Fax: 858-822-3249, E-mail: muotri@ucsd.edu

http://www.bioscience.org/current/vol4E.htm