Retrotransposons – Natural and Synthetic

Jef D. Boeke[1], *Wenfeng An*[1], *Lixin Dai*[1], *Edward S. Davis*[1], *Jeffrey S. Han*[1], *Kathryn A. O'Donnell*[1], *Lisa Z. Scheifele*[1], and *Sarah J. Wheelan*[1]

Transposable elements are ubiquitous among sequenced genomes. The host genomes roughly subdivide into two types: 1) streamlined, that is, small, with little space between genes and lacking large introns, or 2) bulky, with lots of space between genes and many large introns. Most microorganisms, along with selected vertebrates like the pufferfish, fall into the first class, whereas mammals and most plants fall into the second class. As can be seen from Fig. 1, transposable element abundance mirrors the genome type of the host, with mobile elements comprising half or more of many of these bulky genomes! Mobile elements are of two basic types: DNA transposons, which predominantly mobilize via a cut and paste mechanism, and retrotransposons, which move by a copy and paste mechanism involving reverse transcription of an RNA intermediate (Fig. 1 right panel; Curcio and Derbyshire 2003). Retrotransposons are found in virtually all eukaryotes, from yeast (Kim et al. 1998) to human (Lander

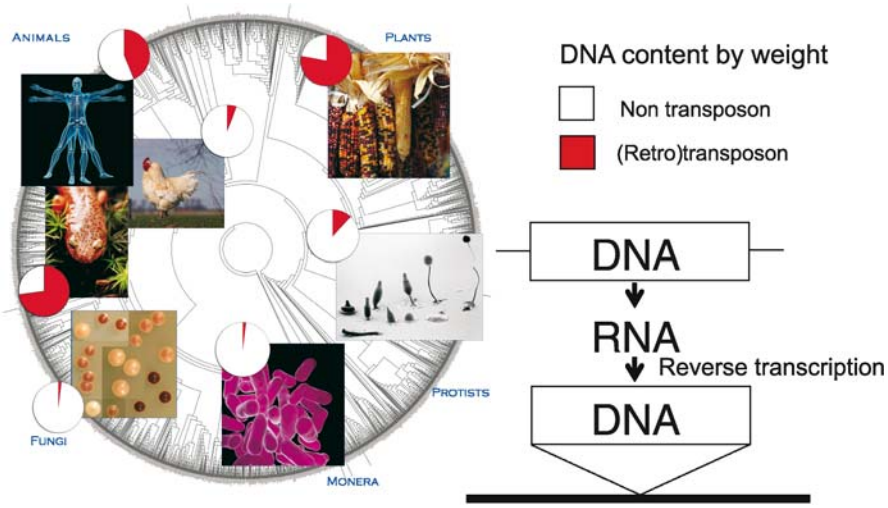

Fig. 1. *Left panel* shows the phylogenetic tree of life as determined by rDNA sequence alignments. Selected organisms are shown, along with the fraction of their genome made up of mobile elements as pie charts. On the *right* is the basic information flow used in the retrotransposition process

[1] High Throughput Biology Center, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA, e-mail: jboeke@jhmi.edu

Gage et al.
Retrotransposition, Diversity and the Brain

et al. 2001). Remarkably, in a yeast cell, the number of retrotransposon copies can be changed rather dramatically without a major impact on the phenotype of the host. The change in copy number can be seen using a new tiling array technique by which it is possible to comprehensively map the unique genomic regions adjacent to all transposable element copies probed (Fig. 2; Wheelan et al. 2006). The ability of yeast strains to tolerate very high copy numbers of transposons is due in part to the fact that, in yeast, most insertions are targeted to non-essential genomic regions, even though most of the genome is protein-coding (Chalker and Sandmeyer 1990; Devine and Boeke 1996; Ji et al. 1993; Zou et al. 1996). This property and many others suggest that retrotransposons are highly coevolved with their hosts.

L1 retrotransposons or LINE-1s are ubiquitous mammalian mobile elements. Each mammalian species' genome is littered with copies of an L1 species that has coevolved with its genome (Gibbs et al. 2004; Kirkness et al. 2003; Lander et al. 2001; Waterston et al. 2002). L1 elements directly make up about 17% of our genome and are responsible

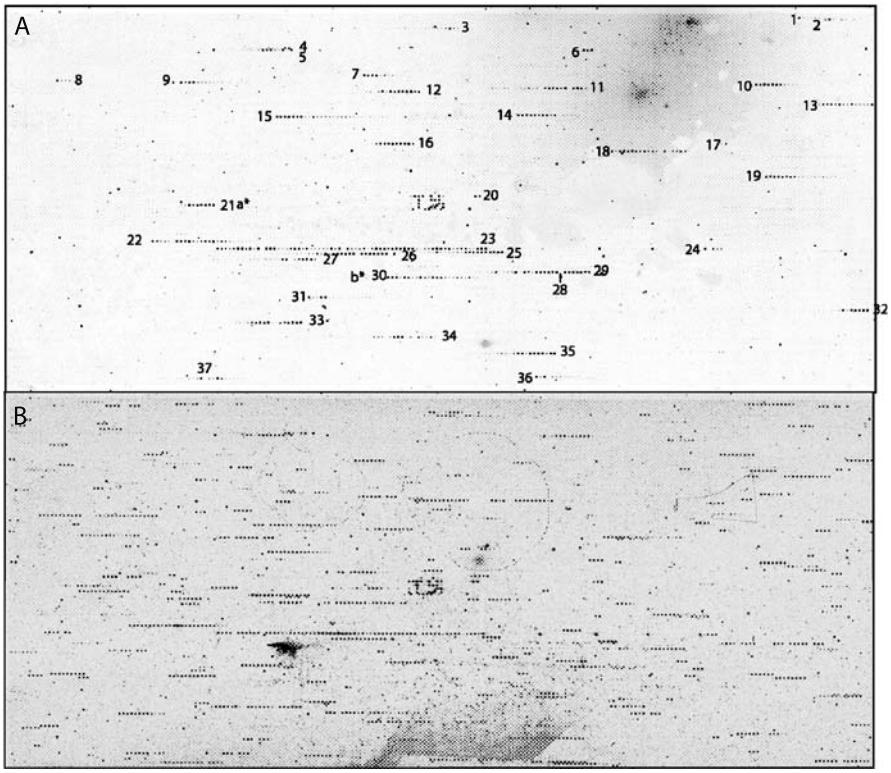

Fig. 2. Mapping transposon insertion sites. Genomic regions adjacent to transposons are PCR amplified and then identified by hybridization to a tiling array. Positive hybridization controls produce a visible "TY" signal. Because features on the array are ordered by chromosomal location, hybridization to adjacent features can be used to identify insertion sites in a wild-type yeast strain (**A**) or a strain with high transposon copy number (**B**)

for at least a third of our DNA by weight because they provide the molecular machinery for mobilizing not only their own sequences but also the highly abundant Alu sequences (Dewannieux et al. 2003), as well as the less abundant processed pseudogenes (Esnault et al. 2000). The latter "retrotranscripts" are simply cellular mRNAs that have been reverse transcribed by the L1 machinery and inserted into the genome, very much like L1 itself is inserted. Retrotransposons move in the genome via a replicative process (Fig. 3). After being transcribed into a full length RNA by host RNA polymerase, the RNA can be translated to produce two proteins, ORF1p and ORF2p. Together with the RNA, these form a ribonucleoprotein (RNP) complex (Martin 1991), which is imported into the nucleus. ORF2p has endonuclease (Feng et al. 1996) and reverse transcriptase (Mathias et al. 1991) functions essential for retrotransposition (Moran et al. 1996). The endonuclease selects and cleaves the target site (Cost and Boeke 1998), and the RNA is ultimately reverse transcribed to make a new retrotransposon copy, a process known as target-primed reverse transcription (TPRT; Fig. 4; Luan et al. 1993).

In somatic and tissue culture cells, L1 expression and hence transposition appear to be tightly regulated transcriptionally, and so the promoter that drives this expression

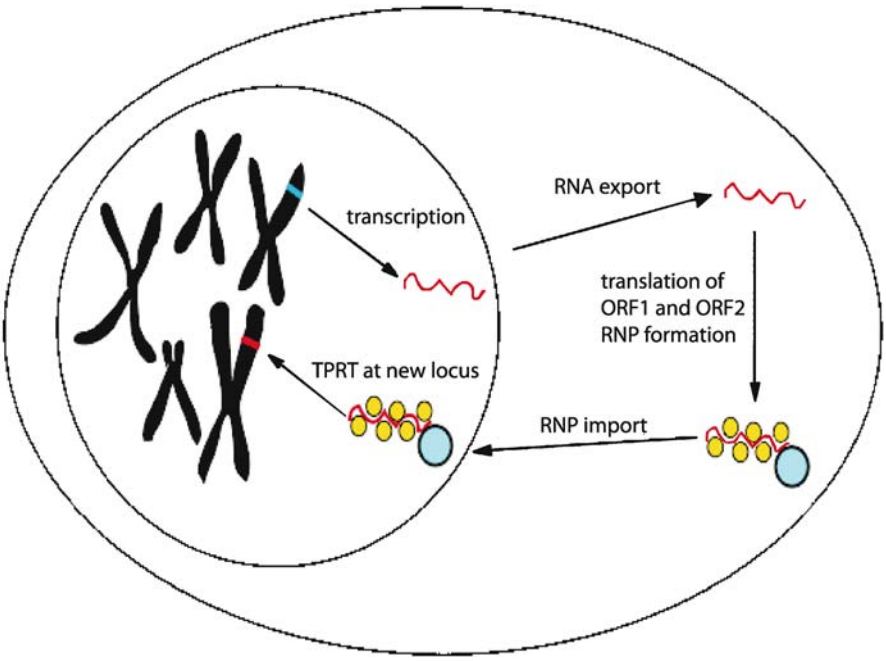

Fig. 3. Replicative cycle of L1 elements. A "donor" element (*blue band* on chromosome) is transcribed in the nucleus. The RNA (*red waved line*) is exported to the cytoplasm, where it is translated into ORF1 (*yellow spheres*) and ORF2 (*blue sphere*) proteins. The ribonucleoprotein (RNP) complex is imported into the nucleus and used as the machinery to drive target-primed reverse transcription (TPRT)/integration of a new copy of the element at a new locus (*red band* on chromosome)

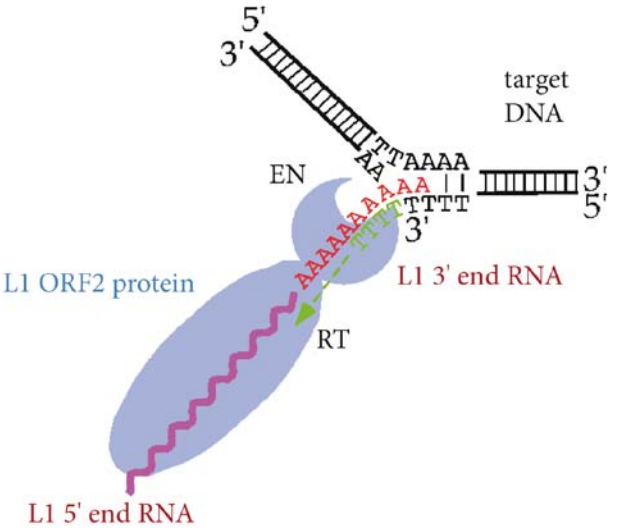

Fig. 4. Mechanism of LINE-1 integration by TPRT. The endonuclease (EN) domain of ORF2 creates a single-strand nick in the target DNA. The L1 RNA anneals with the DNA and ORF2's reverse transcriptase (RT) activity uses the target DNA's 3'-OH to prime synthesis of first strand cDNA

has been an object of considerable interest and scrutiny. Interestingly, although L1 elements in primates and rodents encode relatively similar proteins (percentage of amino acid identity ranges from 20% at the N-terminus of ORF1 to >60% in ORF2), the promoters not only lack sequence homology entirely but also have very different structures (Fig. 5). Most mouse L1 promoters (in the A, F, T^F and G^F subfamilies of mouse L1 elements), like those of other rodent L1s, are made up of a series of tandem repeats of ~200 bp, called monomers, followed by a short non-monomeric region (Goodier et al. 2001; Padgett et al. 1988). Both subfamilies are relatively ancient and most members are inactive. T^F is a young, expanding subfamily containing ~3 000 full-length members and ~1 800 of them are active. G^F is the most recently discovered subfamily that contains ~400 active elements. Both the T^F and G^F monomer are 70% identical to F-type monomer but differ from each other by 33%. In addition to the differences among monomer sequences, the numbers of monomer repeats and monomer lengths vary among individual element copies. In contrast, the human L1 promoter sequence in transpositionally and transcriptionally active (Ta) elements is

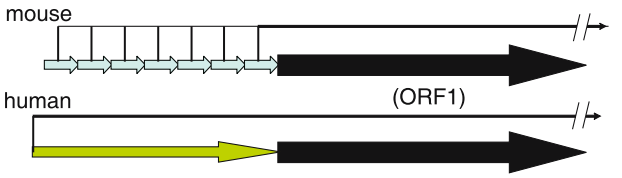

Fig. 5. Comparison of mouse and human L1 promoters. The 5' UTR region of most mouse L1 contains several tandem repeats (monomer) in the length of ~200 bp. Each *blue arrow* represents a monomer sequence. The 5' UTR of human L1 contains a ~900 bp, non-repetitive region (*yellow arrow*) that drives the transcription of L1 element. *Black arrow* denotes the first open reading frame of L1 (ORF1) and *fine line arrow* indicates the transcription direction

about 900 bp long, nonrepetitive and well-conserved in length, and it contains all of the elements required for transcription downstream of the transcription start site (Swergold 1990).

Selfish Gene?

In his book "The Selfish Gene," Richard Dawkins outlines the idea that evolution is driven at the level of individual genes. There is no more compelling example of this than mobile genetic elements like retrotransposons, to which host genomes/organisms are

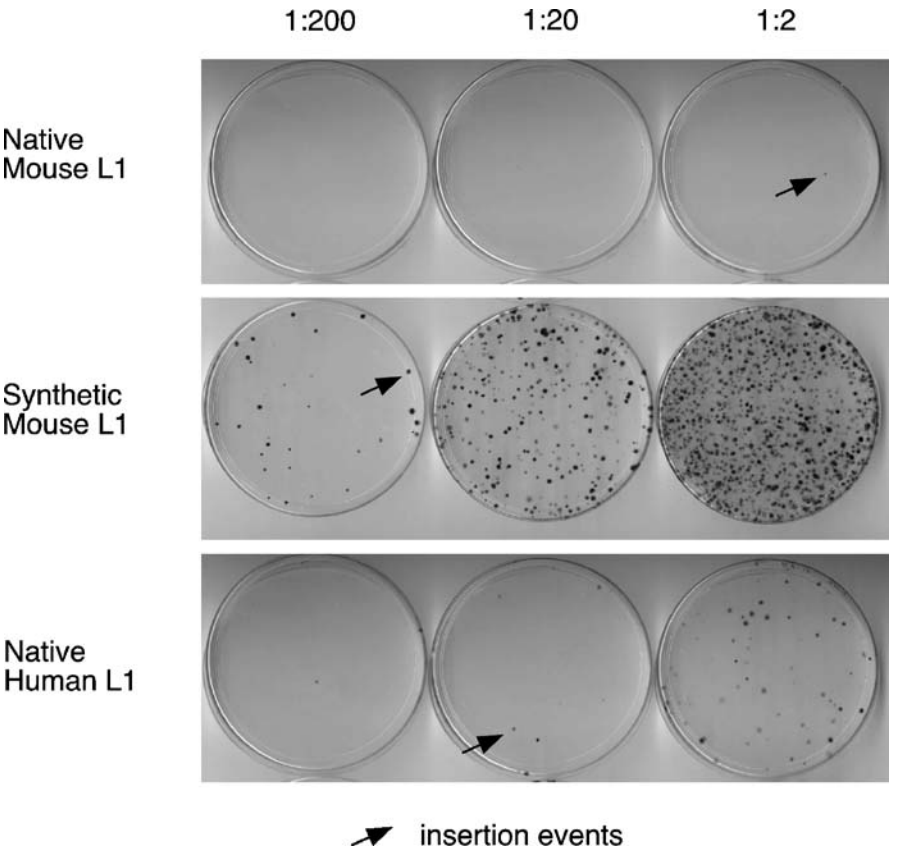

Fig. 6. Synthetic mouse L1 is much more active for retrotransposition than native mouse and human L1 elements. Retrotransposition assay was performed in HeLa cells for native mouse L1, synthetic mouse L1 and native human L1 elements, all of which were tagged with an intron-interrupted neomycin resistance gene reporter. L1 function is scored as the number of G418-resistant colonies because only when L1 completes one round of retrotransposition does a cell become G418-resistant. Cells were diluted at ratios as indicated prior to G418 selection

nothing more than "bags of genes" to exploit (Dawkins 1976). Like virtually all transposable elements found in metazoans, L1 element transposition was until relatively recently thought to be entirely germ-line specific, as predicted from strict "selfish gene" theory. However, recent findings indicate that L1s are highly active transcriptionally in mouse neuronal progenitor cells, and engineered human elements retrotranspose in mouse brain in a neuron-specific manner (Muotri et al. 2005).

Fully Synthetic Retrotransposons Are Highly Active

L1 retrotransposons are potential tools for in vivo mutagenesis; however, native L1 elements are relatively inactive transpositionally in mice. To this end, we have constructed a synthetic L1 element, referred to as *ORFeus*, consisting of two synonymously recoded open reading frames (Han and Boeke 2004). The sequence is based on a native mouse L1 element sequence, L1*spa* (Mulhardt et al. 1994) and can be controlled by either generic (e.g., CMV or CAG promoter) or native L1 5' end transcriptional control sequences. Such donor element constructs can be marked by a transposition indicator gene, which is inserted in the antisense orientation relative to the transcription of the *ORFeus* element. The reporter, either *neo* or *gfp*, is interrupted by an intron in the same sense as the *ORFeus* donor element. In this way, the donor element does not express the reporter because its coding region is disrupted by an inverse intron, but upon retrotransposition, the intron is removed during the RNA step and an active reporter gene is generated.

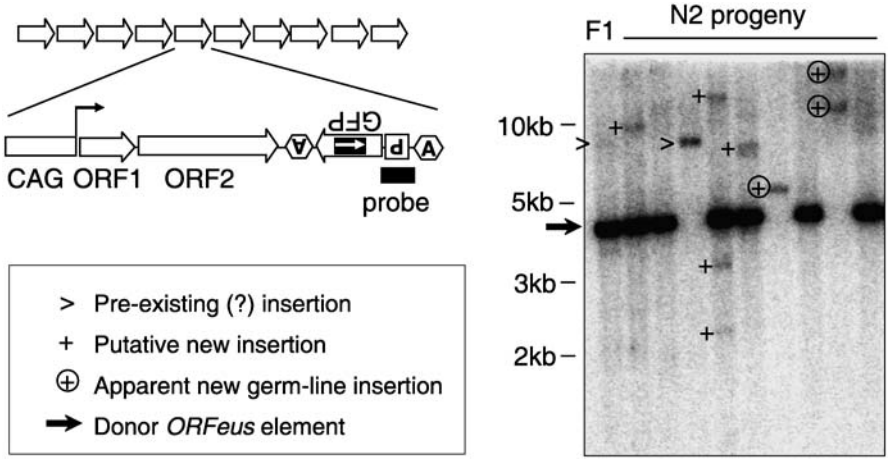

Fig. 7. Estimating germ-line insertion frequency by Southern blot analysis. The *top left panel* is a schematic of the 10-copy *ORFeus* donor transgene concatemer with a detailed view of the structure of a single-copy transgene under the regulation of CAG promoter and marked by an intron-disrupted GFP reporter cassette driven by its own promoter and polyadenylation site. The position of the Southern probe is indicated. The *right panel* is a Southern blot for nine N2 progeny mice from breeding their F1 transgenic parent (the first lane) to a wild type mouse

Using a *neo* intron removal assay, *ORFeus* was found to be ~200-fold more active for retrotransposition in cell culture than native mouse L1 elements and was even more active than the most active human elements studied previously (Fig. 6). To study *ORFeus* activity in vivo, we developed transgenic mouse models in which *ORFeus* expression was controlled by the constitutively active heterologous CAG promoter, and we measured *ORFeus* retrotransposition activity both in germ-line and somatic tissues (An et al. 2006). Germ-line retrotransposition frequencies resulting in 0.3–0.4 insertions per animal were seen among progeny of *ORFeus* donor element heterozygotes, as determined by Southern blotting (Fig. 7). This germ-line retrotransposition frequency compares favorably with previously observed retrotransposition frequencies with native elements driven by heterologous promoters (Babushok et al. 2006). Interestingly, we also observed somatic transposition events in 100% of these *ORFeus* donor-containing animals, and many different insertions were readily recovered from each animal using a modified inverse PCR protocol. Modeling exercises suggest that the numbers of somatic insertions per animal could be as high as millions, suggesting that these animals could provide important new models for cancer, as has recently been reported for the *Sleeping Beauty* DNA transposon (Collier et al. 2005; Dupuy et al. 2005). Somatic retrotransposition was observed in all tissues tested, including brain, but was not particularly elevated in any specific tissue in these mice driven by the CAG promoter. Nearly 200 insertions were precisely mapped, and their distribution in the mouse genome appeared random relative to transcription units and GC content (Fig. 8). Constitutive *ORFeus* may be extraordinarily useful for in vivo mouse mutagenesis. Gene traps are being developed for these purposes.

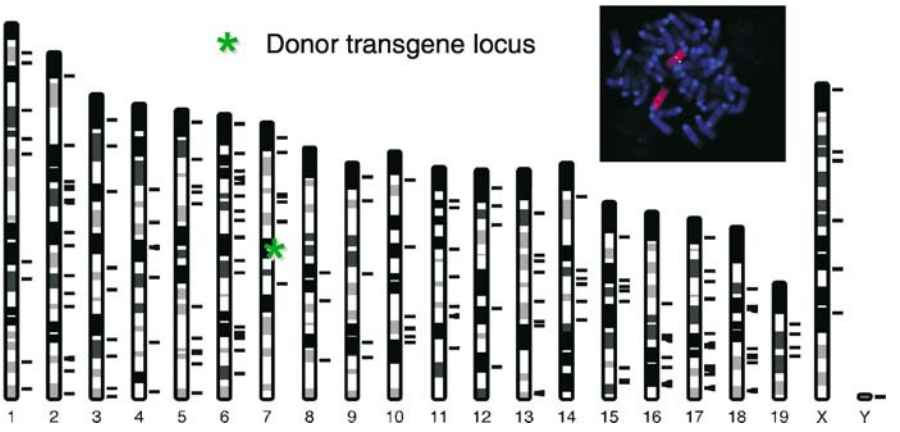

Fig. 8. Chromosomal distribution of mapped insertions. A total of 171 mappable insertions were charted to mouse genome build 36 (*short black lines* to the right of individual chromosomes). The approximate position of the donor concatemer on chromosome is marked (*green asterisk*), which was located by fluorescent in situ hybridization (shown in the insert). *Insert*: Metaphase spreads of splenocytes from donor-containing mice were probed with fluorescently labeled full-length transgene cDNA probe (*green*) and subsequently with a whole-chromosome paint probe for chromosome 7 (*red*). Chromosomes were counterstained with DAPI (*blue*)

Neural-specific Retrotransposition – Could it be Conserved from Rodents to Primates?

L1 elements in mouse and human, like most metazoan retrotransposons, show evidence for germ-line-specific expression (Branciforte and Martin 1994; Ergun et al. 2004; Trelogan and Martin 1995). There is evidence that native rodent L1s are active in neural progenitor cells stimulated to differentiate in response to FGF-2 and are upregulated transcriptionally. On this basis, Muotri et al. (2005) introduced a human L1 (driven by a human L1 promoter) marked with a retrotransposition indicator gene into such cells and into transgenic mice. Interestingly, when the cells were differentiated in tissue culture into astrocytes, glia and neurons, retrotransposition of the human constructs was only seen in those cells that differentiated into neuron-like cells. In some of these cases, insertion of the new retrotransposons was into transcriptionally active target genes in the differentiating neurons. Furthermore, significant retrotranspositional activity of this element (as inferred from GFP staining) was observed in a wide variety of neuronal cells in the brains of these mice. These results can be interpreted to suggest that the highly divergent promoters of primate and rodent LINEs, as well as the divergent proteins encoded by these elements, might be under genetic selection for retrotranspositional activity in the brain. Not only are these promoters highly divergent structurally, but there is also good reason to believe they have an independent genetic origin. The "promoter capture" model (Khan et al. 2006) posits that, as the host inactivates L1 promoters by various mechanisms, L1s can capture novel cellular promoters by TPRT followed by incomplete reverse transcription (Fig. 9). This would then put the element under control of a new promoter. The divergent structures of primate and rodent elements support the idea that at least one such event occurred between rodent and primate lineages. The hypothesis that the promoters are independently derived yet retain germ-line- and neuron-specific activities could be tested by

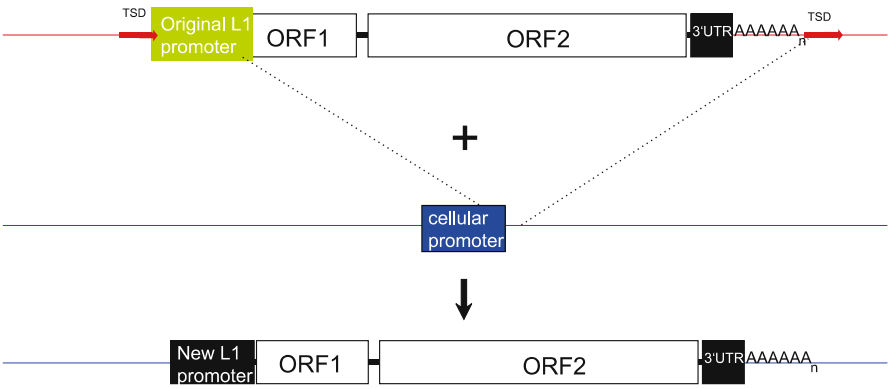

Fig. 9. Promoter capture model. L1 may capture cellular promoters during evolution by transposing a partially truncated element. During the TPRT reaction, the reverse transcription of L1 RNA may extend through ORF1 but fail to copy its own promoter. If this incomplete element is inserted downstream of a cellular promoter, then the L1 might capture this sequence as its own novel promoter

examining the retrotranspositional activity of native mouse elements or *ORFeus* driven by the native mouse L1 promoter. Such experiments are in progress (collaboration with A. Muotri and F. Gage).

References

An W, Han JS, Wheelan SJ, Davis ES, Coombes CE, Ye P, Triplett C, Boeke JD (2006) Active retrotransposition by a synthetic L1 element in mice. Proc Natl Acad Sci USA 103:18662–18667

Babushok DV, Ostertag EM, Courtney CE, Choi JM, Kazazian HH (2006) L1 integration in a transgenic mouse model. Genome Res 16:240–250

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

Chalker DL, Sandmeyer SB (1990) Transfer RNA genes are genomic targets for de novo transposition of the yeast retrotransposon Ty3. Genetics 126:837–850

Collier LS, Carlson CM, Ravimohan S, Dupuy AJ, Largaespada DA (2005) Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. Nature 436:272–276

Cost GJ, Boeke JD (1998) Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. Biochemistry 37:18081–18093

Curcio MJ, Derbyshire KM (2003) The outs and ins of transposition: from mu to kangaroo. Nature Rev Mol Cell Biol 4:865–877

Dawkins R (1976) The selfish gene. Oxford University Press, Oxford

Devine SE, Boeke JD (1996) Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. Genes Dev 10:620–633

Dewannieux M, Esnault C, Heidmann T (2003) LINE-mediated retrotransposition of marked Alu sequences. Nature Genet 35:41–48

Dupuy AJ, Akagi K, Largaespada DA, Copeland NG, Jenkins NA (2005) Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. Nature 436:221–226

Ergun S, Buschmann C, Heukeshoven J, Dammann K, Schnieders F, Lauke H, Chalajour F, Kilic N, Stratling WH, Schumann GG (2004) Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult human tissues. J Biol Chem 279:27753–27763

Esnault C, Maestre J, Heidmann T (2000) Human LINE retrotransposons generate processed pseudogenes. Nature Genet 24:363–367

Feng Q, Moran JV, Kazazian HH, Jr., Boeke JD (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell 87:905–916

Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE, Okwuonu G, Hines S, Lewis L, DeRamo C, Delgado O, Dugan-Rocha S, Miner G, Morgan M, Hawes A, Gill R, Celera, Holt RA, Adams MD, Amanatides PG, Baden-Tillson H, Barnstead M, Chin S, Evans CA, Ferriera S, Fosler C, Glodek A, Gu Z, Jennings D, Kraft CL, Nguyen T, Pfannkoch CM, Sitter C, Sutton GG, Venter JC, Woodage T, Smith D, Lee HM, Gustafson E, Cahill P, Kana A, Doucette-Stamm L, Weinstock K, Fechtel K, Weiss RB, Dunn DM, Green ED, Blakesley RW, Bouffard GG, De Jong PJ, Osoegawa K, Zhu B, Marra M, Schein J, Bosdet I, Fjell C, Jones S, Krzywinski M, Mathewson C, Siddiqui A, Wye N, McPherson J, Zhao S, Fraser CM, Shetty J, Shatsman S, Geer K, Chen Y, Abramzon S, Nierman WC, Havlak PH, Chen R, Durbin KJ, Egan A, Ren Y, Song XZ, Li B, Liu Y, Qin X, Cawley S, Cooney AJ, D'Souza LM, Martin K, Wu JQ, Gonzalez-Garay ML, Jackson AR, Kalafus KJ, McLeod MP, Milosavljevic A, Virk D, Volkov A, Wheeler DA, Zhang Z, Bailey JA, Eichler EE, Tuzun E, Birney E, Mongin E, Ureta-Vidal A, Woodwark C, Zdobnov E, Bork P, Suyama M, Torrents D, Alexandersson M, Trask BJ, Young JM, Huang H, Wang H, Xing H, Daniels S, Gietzen D, Schmidt J, Stevens K, Vitt U, Wingrove J, Camara F, Mar Alba M, Abril JF, Guigo R, Smit A, Dubchak I, Rubin EM, Couronne O, Poliakov A, Hubner N, Ganten D, Goesele C,

Hummel O, Kreitler T, Lee YA, Monti J, Schulz H, Zimdahl H, Himmelbauer H, Lehrach H, Jacob HJ, Bromberg S, Gullings-Handley J, Jensen-Seaman MI, Kwitek AE, Lazar J, Pasko D, Tonellato PJ, Twigger S, Ponting CP, Duarte JM, Rice S, Goodstadt L, Beatson SA, Emes RD, Winter EE, Webber C, Brandt P, Nyakatura G, Adetobi M, Chiaromonte F, Elnitski L, Eswara P, Hardison RC, Hou M, Kolbe D, Makova K, Miller W, Nekrutenko A, Riemer C, Schwartz S, Taylor J, Yang S, Zhang Y, Lindpaintner K, Andrews TD, Caccamo M, Clamp M, Clarke L, Curwen V, Durbin R, Eyras E, Searle SM, Cooper GM, Batzoglou S, Brudno M, Sidow A, Stone EA, Payseur BA, Bourque G, Lopez-Otin C, Puente XS, Chakrabarti K, Chatterji S, Dewey C, Pachter L, Bray N, Yap VB, Caspi A, Tesler G, Pevzner PA, Haussler D, Roskin KM, Baertsch R, Clawson H, Furey TS, Hinrichs AS, Karolchik D, Kent WJ, Rosenbloom KR, Trumbower H, Weirauch M, Cooper DN, Stenson PD, Ma B, Brent M, Arumugam M, Shteynberg D, Copley RR, Taylor MS, Riethman H, Mudunuri U, Peterson J, Guyer M, Felsenfeld A, Old S, Mockrin S.Collins F (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428:493–521

Goodier JL, Ostertag EM, Du K, Kazazian HH, Jr. (2001) A novel active L1 retrotransposon subfamily in the mouse. Genome Res 11:1677–1685

Han JS, Boeke JD (2004) A highly active synthetic mammalian retrotransposon. Nature 429:314–318

Ji H, Moore DP, Blomberg MA, Braiterman LT, Voytas DF, Natsoulis G, Boeke JD (1993) Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes and LTR sequences. Cell 73:1007–1018

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

Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF (1998) Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence. Genome Res 8:464–478

Kirkness EF, Bafna V, Halpern AL, Levy S, Remington K, Rusch DB, Delcher AL, Pop M, Wang W, Fraser CM, Venter JC (2003) The dog genome: survey sequencing and comparative analysis. Science 301:1898–1903

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D,

Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, Szustakowki J, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Luan DD, Korman MH, Jakubczak JL, Eickbush TH (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

Martin SL (1991) Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. Mol Cell Biol 11:4804–4807

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

Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH, Jr. (1996) High frequency retrotransposition in cultured mammalian cells. Cell 87:917–927

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

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

Padgett RW, Hutchison CA, 3rd, Edgell MH (1988) The F-type 5' motif of mouse L1 elements: a major class of L1 termini similar to the A-type in organization but unrelated in sequence. Nucleic Acids Res 16:739–749

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

Trelogan SA, Martin SL (1995) Tightly regulated, developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis. Proc Natl Acad Sci USA 92:1520–1524

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyras E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigo R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, Karolchik D, Kasprzyk A, Kawai J, Keibler E, Kells C, Kent WJ, Kirby A, Kolbe DL, Korf I, Kucherlapati RS, Kulbokas EJ, Kulp D, Landers T, Leger JP, Leonard S, Letunic I, Levine R, Li J, Li M, Lloyd C, Lucas S, Ma B, Maglott DR, Mardis ER, Matthews L, Mauceli E, Mayer JH, McCarthy M, McCombie WR, McLaren S, McLay K, McPherson JD, Meldrim J, Meredith B, Mesirov JP, Miller W, Miner TL, Mongin E, Montgomery KT, Morgan M, Mott R, Mullikin JC, Muzny DM, Nash WE, Nelson JO, Nhan MN, Nicol R, Ning Z, Nusbaum C, O'Connor MJ, Okazaki Y, Oliver K, Overton-Larty E, Pachter L, Parra G, Pepin KH, Peterson J, Pevzner P, Plumb R, Pohl CS, Poliakov A, Ponce TC, Ponting CP, Potter S, Quail M, Reymond A, Roe BA, Roskin KM, Rubin EM, Rust AG, Santos R, Sapojnikov V, Schultz B, Schultz J, Schwartz MS, Schwartz S, Scott C, Seaman S, Searle S, Sharpe T, Sheridan A, Shownkeen R, Sims S, Singer JB,

Slater G, Smit A, Smith DR, Spencer B, Stabenau A, Stange-Thomann N, Sugnet C, Suyama M, Tesler G, Thompson J, Torrents D, Trevaskis E, Tromp J, Ucla C, Ureta-Vidal A, Vinson JP, Von Niederhausern AC, Wade CM, Wall M, Weber RJ, Weiss RB, Wendl MC, West AP, Wetterstrand K, Wheeler R, Whelan S, Wierzbowski J, Willey D, Williams S, Wilson RK, Winter E, Worley KC, Wyman D, Yang S, Yang SP, Zdobnov EM, Zody MC, Lander ES (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562

Wheelan SJ, Scheifele LZ, Martinez-Murillo F, Irizarry RA, Boeke JD (2006) Transposon insertion site profiling chip (TIP-chip). Proc Natl Acad Sci USA 103:17632–17637

Zou S, Ke N, Kim JM, Voytas DF (1996) The Saccharomyces retrotransposon Ty5 integrates preferentially into regions of silent chromatin at the telomeres and mating loci. Genes Dev 10:634–645