[Frontiers in Bioscience E4, 2203-2225, January 1, 2012]

L1 expression and regulation in humans and rodents

James M. Rosser[1], Wenfeng An[1]

[1]*School of Molecular Biosciences and Center for Reproductive Biology, Washington State University, Pullman, WA 99164, USA*

TABLE OF CONTENTS

1. ABSTRACT

Long interspersed elements type 1 (LINE-1s, or L1s) have impacted mammalian genomes at multiple levels. L1 transcription is mainly controlled by its 5' untranslated region (5'UTR), which differs significantly among active human and rodent L1 families. In this review, L1 expression and its regulation are examined in the context of human and rodent development. First, endogenous L1 expression patterns in three different species—human, rat, and mouse—are compared and contrasted. A detailed account of relevant experimental evidence is presented according to the source material, such as cell lines, tumors, and normal somatic and germline tissues from different developmental stages. Second, factors involved in the regulation of L1 expression at both transcriptional and posttranscriptional levels are discussed. These include transcription factors, DNA methylation, PIWI-interacting RNAs (piRNAs), RNA interference (RNAi), and posttranscriptional host factors. Similarities and differences between human and rodent L1s are highlighted. Third, recent findings from transgenic mouse models of L1 are summarized and contrasted with those from endogenous L1 studies. Finally, the challenges and opportunities for L1 mouse models are discussed.

2. INTRODUCTION

Mammalian genomes are replete with transposable elements. Long interspersed elements type 1 (LINE-1s, or L1s) belong to the most abundant class of autonomous transposable elements in the sequenced human, mouse, and rat genomes, and account for roughly 17%, 20%, and 23% of their genomic DNA, respectively (1-3). Humans and rodents diverged about 75 million years ago. A comparative analysis of the age (i.e., the extent of sequence divergence) distribution of L1 sequences indicates that, in the past 25 million years, L1 activity has undergone significant decline in the human but not in the mouse and rat genomes (1-3). This overall decline of L1 activity in the human lineage is also evident by comparing the relative contribution of lineage-specific L1 sequences to each genome. Since the primate-rodent split, human-specific L1 subfamilies contribute to a mere 8% of the human genome as compared to over 16% by mouse-specific L1s (3). L1 activity is slightly higher in rat than mouse since the mouse-rat split (12–24 million years ago): rat-specific L1 subfamilies contribute to 12% of the rat genome, while 10% of the mouse genome is derived from mouse-specific L1s (1). However, highly active L1s do exist in the modern human genome. In the last 25 million years, five distinct human L1 families have emerged and generated more than 35,000 copies (4, 5). The youngest family, termed Ta (transcriptionally active), arose approximately 5 million years ago after the divergence of humans and chimpanzees. They have accumulated, per generation, at about the same rate in humans as the currently active L1 families in mice and rats (6). The distribution of polymorphic Ta1 elements in human populations suggests that the Ta1 subfamily continues to decrease the fitness of modern humans (7).

L1 transcription is mainly controlled by its unique, internal promoter (8). The promoter activity resides in the 5' untranslated region (5'UTR) of a full-length L1 sequence (6–8 kb), i.e., the promoter sequence is transcribed (9-12). The 3' untranslated region (3'UTR) contains a polyadenylation signal (13, 14). Two open reading frame proteins (ORF1p and ORF2p) are encoded by the L1 body, and both are essential for L1 retrotransposition (15). The vast majority of genomic L1 sequences are 5' truncated (13, 14) and thus cannot be autonomously expressed. The number of full-length elements is in the order of 6,000 to 7,000 in human and mouse genomes (16-18). Among these, approximately 100 and 3,000 copies are estimated to be potentially active for retrotransposition in human and mouse genomes, respectively (19, 20). Proteins expressed from active full-length L1s preferentially mobilize the source template L1 mRNA via a phenomenon termed *cis*-preference (21-23). However, non-autonomous retrotransposons, such as human Alu elements (24) and mouse B1 and B2 elements (25), can be efficiently mobilized by L1 retrotransposition machinery. L1 can also generate processed pseudogenes but at a relatively lower frequency (21).

Non-LTR (long terminal repeat) retrotransposons are drivers of mammalian genome evolution (26). L1, as the sole autonomous member, exerts its impact on mammalian genomes at multiple levels. First, existing L1 sequences can profoundly modulate the quantity and quality of mammalian gene expression (27). On the genome scale, there exists an inverse correlation between gene expression level and the amount of L1 sequence present (28). Human genes carry an average of 8 kb of L1 sequences, but the amount of L1 sequence is drastically different among highly and poorly expressed genes (approximately 3 kb and 14 kb, respectively) (28). Second, the expression of L1 proteins causes multiple types of toxicity in cultured cells (29-32). For example, ectopic expression of a full-length L1 results in high levels of double-strand DNA breaks (DSBs) (30, 31), cell cycle arrest at G2/M (30, 31), induction of apoptosis in p53-competent cancer cells (31, 32), and induction of a senescence-like cellular state (29). The endonuclease domain of ORF2p is critical for DSB formation (30), but both endonuclease and reverse transcriptase domains of ORF2p contribute to deleterious effects on cellular proliferation (29). Third, L1-mediated insertional and post-insertional mutagenesis is responsible for a wide spectrum of human diseases (33, 34). While the majority of observed L1 cases are germline events, two somatic L1 insertions have been reported (35, 36). The first case is an insertion in the intron of one proto-oncogene *myc* allele; the insertion is tumor-specific because it is only found in the ductal adenocarcinoma sample but not in normal breast tissue of the same patient (36). The second case is an insertion that disrupts the last exon of the tumor suppressor gene *APC* in a colon cancer. In contrast to the *myc* case, the insertion in the *APC* gene bears all hallmarks of *de novo* L1 insertion and is very likely the causative event (35). More recently, several somatic L1 insertions have been found in human lung cancer genomes through next-generation sequencing but their contribution to tumorigenesis is unknown (37).

In this review, we examine L1 expression and its regulation in the context of human and rodent development. In Section 3, endogenous L1 expression patterns in three different species—human, rat, and mouse—are compared and contrasted. A detailed account of relevant experimental evidence is presented according to the source material, such as cell lines, tumors, and normal somatic and germline tissues at different developmental stages. In Section 4, factors involved in the regulation of L1 expression at both transcriptional and posttranscriptional levels are discussed, including transcription factors, DNA methylation, PIWI-interacting RNAs (piRNAs), RNA interference (RNAi), and posttranscriptional host factors. Similarities and differences between human and rodent L1s are highlighted. In Section 5, recent findings from transgenic mouse models of L1 are summarized and contrasted with those from endogenous L1 studies. Finally, the challenges and opportunities for transgenic L1 studies are discussed.

3. ENDOGENOUS L1 EXPRESSION

The expression of endogenous L1 RNA and proteins is spatiotemporally regulated during animal growth and development. Our understanding of its spatial and temporal expression patterns originates mostly from studies

of endogenous human, rat, and mouse L1 sequences. In this section, we will provide an overview of experiments that profile endogenous L1 expression in the form of L1 RNA and/or proteins in cell lines and tissues. Two different techniques, i.e. reverse transcriptase-polymerase chain reaction (RT-PCR) and Northern blot analysis, have been featured in L1 RNA detection studies. RT-PCR is a very sensitive method and has been frequently used to quantify the level of L1 RNA. However, Northern blot analysis has the advantage in discerning full-length L1 transcripts from abundant read-through or truncated transcripts. Further, the interpretation of RT-PCR can be confounded by L1-related RNA species resulting from premature polyadenylation (28, 38), aberrant splicing (39), and cellular transcripts from an antisense promoter located in the 5'UTR (40, 41). Thus, an emphasis will be placed on studies that provide evidence for full-length L1 transcripts.

3.1. Endogenous human L1 expression

The examination of human L1 expression is limited by the lack of accessibility for normal tissue samples, and thus most work has focused on L1 expression in cell lines and tumor biopsies, such as germ cell tumors (GCTs), GCT-derived embryonal carcinoma cell (ECC) lines, and breast cancers. Recently, fetal/adult testes and a panel of poly (A)+ RNA from normal tissues have also been evaluated for endogenous L1 expression. Overall, L1 expression is predominantly detected in tumors, and cell lines derived from malignant tumors. Additionally, L1 expression is detected by immunohistochemistry in both germ cells and selected somatic cells of human fetal and adult testes, and by Northern blot analysis from selected normal tissue RNA samples.

3.1.1. Human cells lines

L1 expression is prevalent in GCT-derived ECC lines, such as 2102Ep, GH, JEG-3, NCCIT, NTera2D1, and PA1 (42-45). They are derived from two types of GCTs: teratocarcinoma and choriocarcinoma (GCTs are detailed in Section 3.1.2.). However, the first glimpse of potential full-length L1 transcripts was provided by gradient sedimentation of the cytoplasmic RNA from Jurkat cells, an immortalized human T lymphocyte cell line (46). Later, a discrete 6.5 kb L1 RNA signal was detected in the cytoplasmic fraction of a human teratocarcinoma cell line (NTera2D1) (44). This RNA species only hybridizes to antisense probes, is polyadenylated (44), and has been subsequently characterized at the sequence level (47). Most interestingly, its expression is regulated by the cellular differentiation process as evidenced by a strong correlation between its abundance and the differentiation status of these cells: the 6.5 kb L1 RNA signal is present at high levels in high-density NTera2D1 culture when the cells display the characteristic appearance of ECCs, and is absent when cells are differentiated by retinoic acid (44). Full-length L1 transcripts are also detected in additional teratocarcinoma cell lines, including NCCIT, GH and PA1 (42). The functional relevance of these full-length L1 transcripts is corroborated by the detection of L1 protein products. Western blot analyses with a polyclonal antibody reveal the presence of human L1 ORF1p in two teratocarcinoma cell lines (NTera2D1 and 2102Ep) and one choriocarcinoma cell line (JEG-3) (43). This antibody has served as a key reagent in numerous subsequent studies on L1 expression in both cell lines and formalin-fixed tissues (Sections 3.1.2., 3.1.3., and 3.1.4.).

Besides cell lines derived from GCTs, several other tumor cell lines also show expression of L1 RNA and/or ORF1p. At least seven breast cancer cell lines have been tested positive for ORF1p by Western blot analysis, including three infiltrating ductal carcinoma cell lines, three adenocarcinoma cell lines, and one carcinoma cell line (48, 49). Recently, a full-length L1 signal has also been detected in the poly (A)+ RNA fraction from a cervical adenocarcinoma cell line (HeLa; detailed in Section 3.1.4.) (50). However, L1 expression is not restricted to tumors of epithelial cell origin as previously claimed, because ORF1p can be detected in cell lines derived from human bladder carcinoma (J82), colon carcinoma (K12M), melanoma (A375), and fibrosarcoma (HT-1080) (49). Two other bladder carcinoma cell lines (HT1376 and 5637) and one T-cell lymphoma cell line (Molt4) show full-length RNA transcripts, albeit at a much reduced level as compared to teratocarcinoma cell lines (43).

In contrast to tumor cell lines, human embryonic stem cells (ESCs) are the only non-transformed human cell lines displaying endogenous L1 expression. ESCs are derived from the inner cell mass of blastocyst-stage embryos, and are pluripotent in that they can differentiate into cell types of all three primary germ layers. ESCs represent the normal counterpart of ECCs. Undifferentiated human H9 ESCs show levels of ORF1p comparable to ECCs (51).

3.1.2. Human germ cell tumors

Human GCTs are a heterogeneous group of neoplasms with germ cell origin. They mainly occur in gonads and in the midline of a human body. These locations represent the primordial germ cell (PGC) migration route from the yolk sac to the genital ridge (52). Several histotypes are observed in the testis: teratomas and yolk sac tumors are often found in neonates and infants, seminomas and nonseminomas in adolescents and adults, and spermatocytic seminomas in the elderly. Histologically, seminomas are homogeneous tumors. In contrast, nonseminomas are often heterogeneous, and can contain different histological elements, including embryonal carcinoma (the undifferentiated stem cell component), teratoma (the differentiated somatic component), yolk sac tumor and choriocarcinoma (differentiated extraembryonic components) (52). The precursors of infant teratomas and yolk sac tumors are probably early PGCs or gonocytes that carry partially-erased genomic imprinting. These precursors are arrested before meiosis I and become pluripotent embryonic germ cells (EGCs) after reprogramming. Both seminomas and nonseminomas originate from neoplastic PGCs/gonocytes, which have undergone complete erasure of genomic imprinting. These precursors occupy the spermatogonial niche of the seminiferous tubule in adult testes, and their default pathway is to form seminomas. Nonseminomas result from the reprogramming of a seminomatous tumor cell to a pluripotent ECC (the malignant counterpart of EGC) (53).

Several studies have examined ORF1p in adult and pediatric GCTs (54-56). Earlier results indicate that the prevalence of ORF1p-positive cells is approximately 5-10% in both adult and pediatric germ cell tumors (59 and 58 cases, respectively) (54, 55). However, a more recent study has detected variable ORF1p expression in all 162 cases of pediatric germ cell tumors (56). Among them, 18.5%, 37.7%, and 43.8% show strong, moderate, and weak immunoreactivity to ORF1p antibody, respectively. The differences between these studies may be explained by the more sensitive immunohistochemistry detection method adopted in the latter study (56); it is likely that only cells expressing high levels of ORF1p were detectable in the earlier studies (54, 55).

Some common themes emerge from these studies regardless of the methodological differences. First, strong ORF1p expression is frequently detected in embryonal carcinoma, choriocarcinoma, and yolk-sac tumors, but not in teratoma and germ cell tumors without embryonal carcinoma and yolk sac tumor components (55). Second, embryonal carcinoma or yolk sac tumor cells are the reactive component in all positive cases for both adult and pediatric germ cell tumors (54, 55). Normal cells within the testicular tumor mass do not react with ORF1p antibody (54). Normal testicular tissue adjacent to tumors show negative or weaker ORF1p expression than tumor cells (56). Third, most immunoreactive cells appear to be undifferentiated epithelial cells. Multivariate analyses show that strong expression of ORF1p is associated with poor differentiation, with extragonadal sites, and with the yolk sac tumor histotype (56). Fourth, metastases to the lung and lymph nodes display an ORF1p expression profile similar to that seen in primary tumors (54). Lastly, it is noted that an immunohistochemical screening of over twenty different normal tissues fail to detect any ORF1p signal (54).

3.1.3. Human breast cancers

ORF1p is also expressed in a significant number of human breast cancers (48, 57, 58). In one study, nine out of twelve infiltrating ductal carcinomas are positive for ORF1p by immunohistochemistry, but nonmalignant breast epithelial cell lines and normal breast tissues are all negative (48). In a separate study, the epithelia of all twelve breast cancers express ORF1p by immunohistochemistry, although one tissue is very weakly reactive (57). In contrast, the epithelia of four out of six nontumor tissues adjacent to matched cancers are negative or very weakly positive (57). For most positive cases, the ORF1p signal is heterogeneous among tumor cells, varying from negative to strongly positive (48, 57). Interestingly, Western blot analysis appears to be much more sensitive than immunohistochemistry in detecting ORF1p from breast tissues. Western blot analysis, with the same ORF1p antibody, detects a specific signal from all human breast tissues examined, including tumors, matched nontumor breast tissues, and reduction mammoplasty tissues. Nonetheless, matched nontumor breast tissues and reduction mammoplasties have significantly lower levels of expression than primary tumors (57). In general, there is no correlation between ORF1p levels and clinicopathologic outcomes (57, 58), but nuclear localization of L1 ORF1p is frequently associated with a poor prognosis of breast tumors (58).

3.1.4. Human somatic and germline tissues

Due to the lack of an effective ORF2p antibody, the examination of ORF2p expression was not possible until very recently (59). Using immunohistochemistry, both ORF1p and ORF2p are found in fetal and adult testes (the latter are from orchiectomized patients with prostate cancer). Germ cells positive for both ORF1p and ORF2p include gonocytes from fetal testes, secondary spermatocytes, and immature spermatids with residual bodies from adult testes (59). In addition, L1 proteins are found in some somatic cells, including Leydig cells in both fetal and adult testes and Sertoli cells of the adult testis. L1 proteins are also found in vascular endothelial cells of both fetal testis and adult epididymis and in syncytiotrophoblasts of the placenta (59). Similar but slightly differing findings were previously reported in mouse studies (60, 61) (discussed in Sections 3.3.2. and 3.3.3.). In contrast to males, L1 protein expression in female germ cells has not been reported. Recently, using an RNA-fluorescence *in situ* hybridization technique, L1 transcripts have been detected in human immature diploid oocytes from *in vitro* fertilization donors (62). However, it is unclear whether such transcripts are full-length.

L1 expression in normal human tissues has recently been evaluated at the RNA level by a sensitive Northern blot protocol (50). This protocol was previously used to reveal full-length, as well as variably processed, L1 transcripts from two cancer cell lines that have shown evidence for ORF1p expression (MCF7 and SK-BR-3) (39). A similar L1 RNA expression pattern is demonstrated in HeLa cells, although a higher proportion of L1 signals in HeLa cells are from processed RNA species as compared to MCF7 cells (50). Importantly, the improved sensitivity and the rational design of 5'UTR probes enable the detection of specific spliced L1 transcripts. Two of such spliced transcripts are predicted to encode a functional ORF2p (39). When the spliced 5'UTR sequences are cloned between an SV40 promoter and an active ORF2 sequence, the expressed product displays functions that are expected from an authentic ORF2p (50). For example, their overexpression supports Alu mobilization in HeLa cells and induces DNA damage and a senescence-like phenotype in normal human fibroblasts (50). However, the most provocative finding comes from the examination of L1 RNA in a poly (A)+ RNA panel of normal human tissues (50). Full-length L1 RNA is readily detected in several tissues, including esophagus and placenta, at a level comparable to that found in HeLa cells. Processed L1 RNA species are detected in nearly all tissues examined. The highest level of L1-related transcripts is found in the testis, but interestingly it lacks a discernable full-length L1 signal (50). In addition, the bulk of these "processed" L1-related RNA species are significantly shorter than the spliced ORF2 transcripts. This L1 RNA profile in the human testis sample is reminiscent of that in the mouse testis (Section 3.3.2.). The identity and functional significance of these short L1 RNAs warrant further investigation.

3.2. Endogenous rat L1 expression

There are only a limited number of studies on L1 expression in rats and rat cell lines. An elevated level of ORF1-containing transcripts is observed in rat neural progenitor cells (NPCs) by microarray and RT-PCR studies (63). A peculiar pattern of L1 expression is found in rat chloroleukemia (RCL) cells. Chloroleukemia is an uncommon condition in both humans and rats. Rat chloroleukemia can be experimentally induced by gastric instillation of a chemical carcinogen, methylcholanthrene (64). The chloroma cells consist of an immature population of myeloblasts, promyelocytes, and neutrophilic myelocytes (65). These cells represent the first three morphologic stages in the development of mature neutrophils, and, in contrast to later stages of neutrophil development, are able of self-renew. They can be maintained by repeated transplantation into young rats, or in suspension cell culture. It has been reported that L1 transcription is suddenly activated when RCL cells reach a critical population density (i.e., at approximately half of the maximal population density) (66). A ten-fold increase of cytoplasmic L1 transcripts over the course of four hours is detected by an RNA dot blot analysis (66), although it is not clear to what extent full-length L1 transcripts contribute to such an increase. Cells continue to grow beyond the critical population density while L1 transcript levels remain high (66). However, in less than thirteen hours, cell growth is inhibited, and all cells die within two days due to apoptosis. DNA dot blot analysis suggests a sudden incorporation of about 300,000 L1 copies into the genome of dying cells. Interestingly, this growth inhibition can be released by subculturing. If subculturing is performed at or near the maximal cell density, the level of L1 transcripts quickly return to the baseline level typical of the exponential growth phase. If the subculturing is delayed after reaching the maximal density, the level of L1 transcripts does not decrease (66). In addition to cell density, ultraviolet and ionizing radiation can induce rapid, premature activation of L1 during the exponential growth phase (67). In fact, discrete, full-length, and poly (A)+ L1 transcripts can be readily detected from ultraviolet-irradiated RCL cells as well as RCL cells reaching maximal population density; one such complementary DNA clone is competent in retrotransposition in cell culture (68). L1 expression in RCL cells has also been confirmed at the protein level by Western blot analysis of both ORF1p and ORF2p. In particular, two ORF1 polypeptides are detected, and they correspond to the predicted molecular masses of two ORF1p classes, due to differences in the length polymorphism region (LPR) (68). The molecular mechanism underlying the observed L1 expression dynamics has not been explored. Nevertheless, RCL cells represent a unique case that involves drastic changes in L1 transcription and retrotransposition activities.

Very little is known about L1 expression patterns in normal rat tissues. The presence of L1-related transcripts in nuclear RNA from rat liver, kidney, and a rat neural cell line was initially revealed by reverse Northern blot analysis, in which radio-labeled rat RNA is used as a probe and hybridized to restriction endonuclease-digested genomic DNA (69). Recently, the endogenous expression of rat L1 proteins has been detected in rat heart tissue maintained *ex vivo* (70). ORF1p is localized predominantly in the nucleus of cardiomyocytes, and its level increases in response to ischemia (70). The antibody against ORF2p detects a strong signal exclusively in the endothelial cells of the heart (70); however, this could simply be an artifact as the same antibody cross-reacts strongly to HSP90 in Western blot analysis (68).

3.3. Endogenous mouse L1 expression

Human and mouse L1 sequences are discovered in tandem (71), and many seminal discoveries about L1 structure and function originate from studies on mouse L1 retrotransposons. In fact, the mouse has served as an indispensable model for studying L1 expression in the past two and a half decades, and much we know about L1 expression patterns *in vivo* comes from mouse studies. Specifically, endogenous mouse L1 is expressed in different stages of mouse development in a tightly regulated manner.

3.3.1. Mouse cell lines

The initial evidence of full-length mouse A-type L1 transcripts was found in several mouse lymphoid cell lines (72). These transcripts are sense-strand specific and enriched in poly (A)+ cytoplasmic RNA fractions. Subsequently, full-length, sense-stranded mouse L1 transcripts were detected in two mouse teratocarcinoma cells: F9 (73, 74) and C44 (75). They are found to be associated with ORF1p in sucrose gradient fractions (73, 75). In addition, the younger A-type L1 transcripts are more abundant than F-type transcripts when hybridized with type-specific 5'UTR probes (74). Indirect immunofluorescence reveals a punctate distribution of ORF1p in the cytoplasm of F9 cells with a marked variation in the intensity of staining from cell to cell (75). Unlike F9 cells, C44 embryonal carcinoma cells are maintained in cell culture as embryoid bodies (EBs) (75). Normally, embryoid bodies are formed when ES cells undergo differentiation in cell culture. They are three-dimensional multicellular aggregates consisting of differentiated and undifferentiated cells. C44 cells form either simple or cystic EBs, which resemble mouse embryos at either the morula or blastocyst stage, respectively. Two distinct cell lineages are present in C44 EBs: the inner cells resemble the inner cell mass of a normal blastocyst and ECCs, and are the only cells positive for ORF1p; in contrast, the outer cells are histologically heterogeneous, with the majority reminiscent of endoderm cells. Similar to studies with human teratocarcinoma cells, no full-length RNA and ORF1p can be detected when F9 and C44 cells are differentiated by retinoic acid (75).

3.3.2. Mouse germ cells

The pursuit of L1 expression in normal mouse tissues started with an effort to detect full-length transcripts in mouse oocytes and blastocysts (76). In this study, a discrete 8 kb A-type transcript was found in mouse blastocysts (76). Two subsequent studies investigated the developmental and cell type specificity of L1 expression during mouse embryogenesis by examining postimplantation embryos, prepuberal and adult testes, and

adult ovaries (60, 61). In the male germ line, L1 ORF1p is detected by immunohistochemistry in gonocytes of E15.5 (i.e., embryonic day 15.5), E16.5, and E17.5 embryonic testes but not in E14.5 or earlier stages (61). In addition, ORF1p is observed by immunohistochemistry in primary spermatocytes (only in leptotene and zygotene stages) of postnatal mouse testes from P10 (i.e., postnatal day 10), P14, P18, and P25 animals, and in round spermatids, elongating spermatids, and residual bodies of the adult mouse testis (60). Similar to the timeline of the male germ line, in the female germ line, L1 ORF1p is first detected at E15.5 in the primary oocytes and continues through E16.5 and E17.5 (61). However, no germ cell immunoreactivity is found in adult ovaries (61). Several recent studies have reexamined the expression of L1 ORF1p by indirect immunofluorescence with the same antibody and have extended their analyses to neonatal testes (P2, P6 and P10) (77). Similar to the early study, prominent ORF1p signals are detected in E14.5 and E16.5 wild-type embryonic testes; E18.5 is also positive but with a reduced signal. A minor ORF1p signal is observed in P2 gonocytes, and no ORF1p protein is present in P6 testes. A minor signal reappears in P10 testes (77). In adult testes, L1 ORF1p is detected in leptotene, zygotene, and pachytene stages of primary spermatocytes, but not in preleptotene and diplotene stages (78).

L1 ORF1p expression in mouse testes as demonstrated by immunohistochemistry has been confirmed by Western blot analysis. Additionally, Western blot analysis reveals both quantitative and qualitative details of mouse ORF1p expression during development from P10 to adult (60). Three ORF1p variants (42 kD, 43 kD, and 44.6 kD) appear in testis extracts in a developmentally regulated manner. Weak 42 kD and 43 kD ORF1p signals are detected at P10. Both signals are increased at P14. At P18, the 42 kD signal is no longer detectable, the 43 kD signal is at its peak level, and a faint 44.6 kD signal starts to appear. In both P25 and adult testis extracts, the 44.6 kD form is the predominant signal detected (60). Distribution of ORF1p variants has also been studied in specific cell fractions from the adult mouse testis. Spermatocytes at the pachytene stage display a weak 43 kD signal and a strong 44.6 kD signal. The 43 kD signal maintains its low intensity in round spermatids, but increases in elongating spermatids and peaks in residual bodies. In contrast, the 44.6 kD signal decreases rapidly in round spermatids and becomes undetectable in elongating spermatids and residual bodies. Consistent with round spermatids and elongating spermatids being the most abundant cell types in adult testes, homogenized seminiferous tubules show an ORF1p profile that is similar to round spermatids: 43 kD and 44.6 kD signals coexist, with the 43 kD signal being slightly stronger than 44.6 kD. Interstitial cells display a single strong 44.6 kD signal, equivalent to that detected in whole testis extract (60). Thus, it seems that different cell types in the mouse testis express distinct ORF1p variants. Studies of ORF1p expression in mouse teratocarcinoma cell lines indicate that members of A and T_F subfamilies of mouse L1 are responsible for the 42 kD and 43 kD signals, respectively (79). *In vitro* expression of the 44.6 kD ORF1p isoform has not yet been reported, but a third active L1 subfamily (20), termed G_F, is predicted to be its source. Thus, at least three mouse L1 subfamilies—A, G_F, and T_F—are being actively transcribed and translated during mouse development. How they are differentially regulated remains unexplored.

In general, there is a lack of correlation between full-length L1 RNA and ORF1p in male germ cells (60). A full-length L1 transcript can be readily detected in total RNA from P14 testes but not in any other developmental stages. Likewise, a specific 2.8 kb L1 RNA signal can only be detected in P25 and adult testes, but its identity and functional significance remain to be studied.

3.3.3. Mouse somatic cells

There is scant evidence that L1 is expressed in somatic tissues. The initial evidence is from immunohistochemical studies of gonads. Several somatic cells in the gonads are positive for mouse ORF1p, including Leydig cells (in E17.5 (61), prepuberal and adult testes (60)), myoid cells (prepuberal and adult (60)), but not in Sertoli cells (prepuberal and adult (60)) or Theca cells in the adult ovary (61). In addition, cyncytiotrophoblast cells of the placenta are ORF1p positive at all developmental stages examined (E13.5, E14.5, E15.5, E16.5, and E17.5) (61). A potential link between steroid hormone production and L1 expression has been suggested because, with the exception of myoid cells, all ORF1p-positive somatic cells (i.e., Theca cells, Leydig cells and cyncytiotrophoblast cells) are involved in active androgen production (61). Interestingly, a later study on L1 expression in corresponding human tissues reveals a partially overlapping set of somatic cells that express human ORF1p and ORF2p (Section 3.1.4.) (59). In light of the data from human gonads, this view has been revised since not all somatic cells are involved in androgen production; rather they all share receptors for steroid hormones, androgen and/or progesterone (59). The significance of this link remains to be determined.

4. REGULATION OF L1 EXPRESSION

Recent evidence suggests that L1 expression is regulated at multiple levels, both transcriptional and posttranscriptional. Mammalian L1 families have repeatedly acquired novel 5' sequences during evolution (8). 5'UTR differs significantly in both sequence and organization between rodent and primate L1 families. In contrast to human L1 5'UTR, both rat and mouse L1s carry a varied number of monomers. As unique internal promoters, L1 5'UTRs from all three species—human (9, 12), rat (10), and mouse (11)—are necessary and sufficient in directing reporter gene expression in transfected cells. In this section, we evaluate the role of transcription factors, DNA methylation, PIWI-interacting RNAs (piRNAs), small-interfering RNA (siRNAs), and posttranscriptional host factors in L1 regulation.

4.1. Transcription factors

Several transcription factors that interact with the human L1 promoter have been identified. Human L1 5'UTR contains binding sites for YY1 (9, 80-82), RUNX3

(83), SOX proteins (63, 84), and p53 (85). The Yin Yang-1 (YY1) binding site is located close to the 5' end of human L1 5'UTR on the antisense strand (+21 to +13) (80). Mutations in the YY1-binding site have only minor effects on L1 transcriptional activation; instead, the YY1-binding site functions as a component of the L1 core promoter to direct accurate transcription initiation and thus to maintain the integrity of the 5' end of human L1 (82). Notably, the YY1-binding site is conserved in the 5'UTR of all extant full-length L1 sequences in the modern human genome (17). In addition, a consensus YY1-binding site is located in monomers of the mouse F and T_F subfamily, although it is mutated in the G_F subfamily (20). Human L1 5'UTR has three putative runt-domain transcription factor 3 (RUNX3) binding sites: the first two on the sense strand (+83 to +101 and +389 to +407), and the third on the antisense strand (+526 to +508). Only the first RUNX3-binding site promotes L1 transcription and retrotransposition *in vitro*. The third site may regulate the antisense promoter activity (83). Despite its functional significance, however, it is unlikely that RUNX3 dictates the tissue specificity of L1 expression, because RUNX3 is expressed in a wide range of tissues. In addition, a RUNX3-binding site is absent in the most active mouse L1 subfamily T_F (83). The tumor suppressor p53 is a key regulator in maintaining genomic integrity. Human L1 5'UTR contains a three-quarter p53-binding site (+452 to +467) and a half p53-binding site (+489 to +498) (85). L1 expression is increased in a reporter assay when p53 is overexpressed; only the three-quarter site is required for this p53 response. The expression of an integrated full-length L1 sequence is also increased in response to p53 activation (85).

It is puzzling as to how different L1 families achieve similar levels of cell-specific expression with divergent promoters. It is commonly believed that cell- and stage-specific epigenetic mechanisms play a pivotal role (Section 4.2.). Recent studies indicate that transcription factors of the SOX (SRY-related HMG box containing) family are involved in L1 transcriptional regulation at least during neurogenesis. The 5'UTR of human L1 contains two functional binding sites for a recombinant protein encompassing the SRY HMG box domain: +472 to +477 and +572 to +577 (84). Ectopic expression of SOX11 results in a 10-fold increase of L1 promoter activity not only in human embryonic kidney 293 cells using a luciferase-based reporter assay, but also in an embryonal rhabdomyosarcoma cell line when endogenous full-length L1 transcripts are measured (84). However, the physiological relevance of SOX11 in regulating L1 expression has yet to be explored. Another SOX family member, SOX2, represses L1 transcription in cultured adult rat hippocampus neural (HCN) stem cells (63). When HCN cells are differentiated toward neuronal lineage, SOX2 expression is down-regulated, which is concomitant with a transient stimulation of L1 transcription (63). Paradoxically, SOX2 has no effect on full-length 5'UTR of human L1 in a luciferase-based reporter assay. Instead, it stimulates reporter gene expression when the first 100 bp are deleted from L1 5'UTR, suggesting that other factors are involved in SOX2-mediated transcription repression of human L1 promoters (63). Adding another layer of complexity, both SOX-binding sites in human L1 5'UTR overlap with T-cell factor/lymphoid enhancer factor (TCF/LEF)-binding sites. TCF/LEF family transcription factors are effectors of the canonical Wnt pathway and can promote specific gene expression when, in association with beta-catenin, bound to their cognate binding sites. Indeed, the overexpression of either WNT3 or beta-catenin stimulates L1 promoter activity in neural stem cells (NSCs) using a reporter assay (86). Thus, it seems that SOX2 and TCF/LEF antagonistically regulate L1 promoter activity in NSCs. L1 promoters from all three species—human, rat, and mouse—contain at least one overlapping SOX2 and TCF/LEF-binding site (86), possibly as the result of convergent evolution.

4.2. DNA methylation and PIWI/piRNA pathway

DNA methylation is correlated with transcriptional gene silencing. It acts by directly impeding the binding of transcription factors to their target sites, and/or altering chromatin structure via histone modification and nucleosome occupancy within the promoter regions of genes (87). L1 5'UTR promoters from human (47, 88), rat (10, 89, 90), and mouse (both A-type (91, 92) and F-type (93)) are readily recognized as CpG islands. The correlation between DNA methylation and L1 expression has been substantiated by a number of studies. An active role of piRNA in *de novo* methylation of L1 sequences has also been revealed.

4.2.1. L1 methylation in cell lines and its correlation with L1 expression

The implication of DNA methylation in regulating L1 expression originates from multiple studies. Expression of L1 full-length transcripts is correlated with differential methylation of the promoter region in various cell lines (45): in teratocarcinoma cells and medulloblastoma cells that express abundant L1 protein, 5'UTR sequences are unmethylated; in somatic cells expressing little or no L1 protein, most or all are methylated (45). When DNA methylation is inhibited by 5-azacytidine or 5-aza-2'-deoxycytidine, a full-length L1 RNA signal becomes detectable in 3T3-derived A cells (94), and the level of 5'UTR-containing transcripts is increased in human embryonic fibroblasts and Ntera2D1 cells (95). Furthermore, individual L1 5'UTR sequences may be differentially methylated in different tumors. For example, the methylation status of nine individual 5'UTR sequences varies not only from 5'UTR to 5'UTR but also among different tumor types (96). The majority of these 5'UTR sequences are heavily methylated in teratocarcinoma-derived cell lines (2102Ep, NTera2D1, and PA-1), but only a varying subset are methylated in cell lines derived from non-germ cell malignancies and non-malignant breast tissues (96). In cell-based reporter assays, human (97) and rat (10) L1 promoter activity is significantly inhibited by CpG methylation. It is possible that not all CpG sites on 5'UTR are required for transcriptional inhibition, because only a subset of these CpG sites affect L1 promoter activity in an *in vitro* transcription assay when mutated (97).

DNA methylation mediates transcriptional silencing by impeding transcription factor binding and/or altering chromatin modification. In the latter case,

methylated CpG sites are recognized by members of the methyl-CpG-binding domain (MBD) protein family, which in turn recruit histone deacetylases and generate a transcriptionally inactive chromatin structure (98). The effect of DNA methylation has only been shown for one of the L1-related transcription factors (Section 4.1.): YY1 binding is unaffected by promoter methylation in a gel shift assay (97). On the other hand, as a member of the MBD protein family, methyl-CpG-binding protein 2 (MeCP2) can bind to methylated L1 5'UTR and represses its transcription in reporter assays (99, 100).

4.2.2. L1 methylation in mouse tissues during development

Genomic methylation patterns undergo at least two cycles of reprogramming during development (101): in preimplantation embryos and during gametogenesis. In general, the genomes of mature sperm and oocytes are highly methylated. The paternal genome of the zygote is actively demethylated shortly after fertilization. Thereafter, the diploid genome is passively demethylated in a stepwise, DNA replication-dependent fashion up to the morula stage. Genomic methylation is restored by *de novo* methylation in a lineage-specific fashion in blastocysts: the inner cell mass is significantly more methylated than trophectoderm cells (destined to form extraembryonic tissues). PGCs originate from proximal epiblast cells at around E7.25, and undergo genome-wide demethylation as early as E8.0. A second wave of demethylation occurs around E11.5 when PGCs arrive at the genital ridge (termed gonocytes from then on). Subsequently, *de novo* methylation starts in the gonocyte genome and specific methylation patterns are reestablished. It should be noted that different classes of sequences might deviate from this general description of methylation dynamics (L1 as an example below).

L1 methylation was initially examined by using Southern blot and methylation-sensitive restriction enzymes (10, 102-106). This method surveys the methylation status of a limited number of CpG sites that overlap with the enzyme recognition sites. In addition, earlier studies employed a probe that corresponds to the body of mouse L1, and consequently, the dynamics of the observed methylation change may only reflect that of the body of endogenous L1 elements (102-106). The body of endogenous L1s is fully methylated in somatic tissues, such as brain (104), kidney (104, 105), liver (102, 103) and spleen (104, 106). Male and female germ cells differ in the degree of methylation. Mature sperm are fully methylated (102-106). In contrast, primary oocytes at the diplotene stage show little methylation (103, 106), secondary oocytes (ovulated but unfertilized) show intermediate level of methylation (42%) (106), and one-cell embryos (fertilized eggs) show higher but incomplete methylation (72%) (106). During the development of preimplantation embryos, L1 methylation gradually decreases (103, 106). When cultured *in vitro*, 2-cell embryos display 44% of methylation, 8-cell embryos 36%, calvitating embryos 32%, and expanded blastocysts 13–23% (106). By E7.5, L1s are fully methylated (102). As the result of demethylation in the developing germ cells, by E11.5, germ cells are nearly devoid of any methylation in both male and female fetuses (105). Remethylation can be seen as early as E12.5, and continues through E13.5 and E14.5 (105). By E16.5, female germ cells remain partially demethylated (103, 105), but male germ cells have become fully methylated (105). Specifically, the body of L1 sequences is fully methylated in pachytene spermatocytes, round spermatids, and epididymal sperm from the adult testis (103). Female germ cells are never fully methylated (103, 105, 106). In extraembryonic tissues, yolk sac mesoderm has a level of methylation similar to somatic tissues (102, 105), while yolk sac endoderm (102, 105), placental DNA (105), and the ectoplacental cone (102) are partially unmethylated.

The characterization of L1 5'UTR made it possible to specifically investigate the methylation status of L1 promoters. In addition, a large number of CpG sites in L1 5'UTR have been surveyed by bisulfite sequencing (107, 108). In general, the methylation patterns of 5'UTR during development are consistent with those of the body of L1. Specifically, 5'UTR sequences are heavily methylated in adult somatic tissues, such as rat liver (10) and mouse brain (109). For germ cells, sperm are highly methylated (98%), and ovulated oocytes are modestly methylated (29%) (107). Methylation of the zygote genome is comparable to oocytes (25%) (107), indicating active demethylation has occurred in the paternal genome. No significant methylation changes occur during preimplantation development, and the level of methylation persists through the blastocyst stage (27%) (107). The methylation status of 5'UTR between blastocyst and E11.5 is not known. An elevated level of methylation is observed in PGCs by E11.5 (65%). It rapidly decreases to 32% by E12.5, and further to 17% by E13.5 (107, 108). The dynamics of 5'UTR demethylation and remethylation differ between male and female. No difference is detected in 5'UTR methylation between male and female at E12.5 (109). In males, no further demethylation occurs beyond E16.5, and 5'UTR sequences are completely remethylated by E17.5 (109, 110). This high level of methylation is maintained in male germ cells from P0–P2 (110, 111), P8–P12 (110), P15 (112) and P17 (113) animals. In females, demethylation extends beyond E15.5, and some remethylation is observed at E17.5, although a substantial portion of L1s remain unmethylated (109).

4.2.3. Correlation of L1 methylation and expression during development

Methylation studies of mouse L1 reveal two developmental time windows when L1 promoters are substantially demethylated: first in preimplantation embryos, and second during gametogenesis (Section 4.2.2.). It is plausible that DNA methylation is not the sole mechanism in place for L1 repression, because, otherwise, the organism would have become very vulnerable during these two time periods due to uncontrolled L1 expression and activity. Possible candidates include repressive histone modifications that are normally associated with transcriptional gene silencing. An alternative, but not necessarily mutually exclusive, possibility is that the relaxed control allows a physiological level of expression that is tolerated by the host. In certain cases (i.e., during gametogenesis and neurogenesis), such level of expression

may even be beneficial to the host. Here we discuss the correlation or the lack thereof between DNA methylation and L1 expression during three stages of mouse development. First, no correlation can be drawn for preimplantation embryos (the first window of opportunity for L1 expression), as little is known about the level of endogenous L1 expression during early embryogenesis in wild-type animals. This developmental stage should be revisited when more information becomes available. In fact, recent transgenic animal studies detect transcripts of human L1 transgene by PCR-based techniques (Section 5.1.2.). Second, there is a strong correlation between DNA methylation and endogenous L1 expression during fetal germ cell development (the second window of opportunity; Section 3.3.2.). In the mouse, L1 5'UTR sequences undergo rapid demethylation in male germ cells between E11.5 and E13.5 (from 65% to 17%) (107, 108), and then become fully remethylated by E17.5 (109). In good agreement with DNA methylation dynamics, prominent ORF1p signals are detected in E14.5–E17.5 embryonic testes (61, 77). Third, there is in general a lack of correlation between DNA methylation and L1 expression in the male germ cell lineage beyond E17.5 (discussed below in the context of meiosis progression).

Typically, germ cells in the embryonic mouse testis enter mitotic arrest at around E14.5, resume cell proliferation at around P2, and develop into spermatogonia via mitotic division. The latter are subsequently transformed into preleptotene spermatocytes, starting at P8–P10 and marking the first wave of meiosis. Leptotene, zygotene, pachytene, and diplotene spermatocytes from the prophase of meiosis I start to appear approximately at P10, P12, P14, and P17, respectively. Haploid round spermatids first appear at around P20 after the second meiotic division. They further differentiate into mature spermatozoa at around P30 (114, 115). Spermatogenesis continues in the adult testis in an asynchronous fashion. The majority of L1 5'UTR sequences are presumably hypermethylated in postnatal and adult germ cells for the following reasons, although DNA methylation has not been examined in germ cells of specific meiotic stages. First, there is no evidence for additional genome-wide demethylation of mouse 5'UTR sequences during spermatogenesis once they become fully methylated by E17.5 (109). Second, mouse L1 5'UTR is highly methylated in male germ cells from P0–P2 (110, 111), P8–P12 (110), P15 (112) and P17 (113) animals, and in adult sperm cells (107). Third, the body of L1 sequences is known to be fully methylated in pachytene spermatocytes, round spermatids, and epididymal sperm from the adult testis (103).

Despite the apparent hypermethylation status of 5'UTR, endogenous L1 is expressed in postnatal and adult testes (Section 3.3.2., and summarized below). A minor ORF1p signal is observed in P2 gonocytes, and reappears in P10 preleptotene spermatocytes (77). Prominent ORF1p signals are observed in primary spermatocytes (leptotene and zygotene stages) of the postnatal mouse testis (P10, P14, P18 and P25) (60). In the adult testis, L1 ORF1p is detected by immunofluorescence in leptotene, zygotene, and pachytene spermatocytes (78), and by immunohistochemistry in round spermatids, elongating spermatids and residual bodies in a separate study (60). During meiosis the germ cell genome undergoes extensive chromatin remodeling (116). Thus, it is likely that other epigenetic changes such as histone modifications are responsible for the activation of L1 expression in postnatal and adult male germ cells, although DNA demethylation at 5'UTR cannot be excluded.

4.2.4. L1 expression in mice deficient in piRNA/DNA methylation pathways

DNA methylation is a key mechanism in controlling endogenous transposable elements, and multiple cellular factors have been recently identified (117). DNA methyltransferase 3-like (DNMT3L) appears to regulate L1 remethylation specifically in the male germ line. DNMT3L is expressed in the non-dividing gonocytes in a brief perinatal period, which coincides with *de novo* methylation of retrotransposons including L1 (113). In *Dnmt3l* knockout mice, an elevated level of full-length L1 RNA is observed by Northern blot analysis in the P25 testis, and it is corroborated by *in situ* hybridization on cross sections of the P2 testis (113). DNA methylation analyses reveal significant demethylation at L1 5'UTR in germ cell but not somatic cell fractions isolated from *Dnmt3l* knockout P0–P2 (111) and P17 testes (113).

piRNAs belong to a class of small RNAs that are highly expressed in germ cells, and are so termed because of their association with PIWI proteins (118). Two mouse PIWI family proteins, MILI and MIWI2, are expressed during germ cell development, and their associated piRNAs display a strong strand bias (119). Primary piRNAs correspond to sense-strand transcripts from retrotransposons, and predominantly associate with MILI in the cytoplasm. Secondary piRNAs are mainly antisense-strand relative to retrotransposon transcripts, and associate with MIWI2 in the nucleus. Together, MILI and MIWI2 act to amplify piRNAs in a Ping-Pong cycle (119). Loss-of-function studies in mice suggest that piRNAs guide *de novo* methylation to retrotransposons (110, 119). L1 RNAs are up-regulated in male germ cells of *Mili* and *Miwi2* knockout mice, and this change in L1 expression is correlated with significant demethylation of L1 5'UTR sequences (110, 120, 121). Furthermore, the PIWI/piRNA pathway likely acts upstream of known mediators of DNA methylation, because piRNAs are still produced in *Dnmt3l*-deficient mutants (119). Details of piRNA-meditated retrotransposon silencing have started to emerge. Multiple cofactors functioning in the piRNA/DNA methylation pathway have been recently revealed by mouse knockout experiments, including MAEL (a murine homolog of *Drosophila* maelstrom) (78), TDRD1 (tudor domain-containing protein-1) (112), TDRD9 (tudor domain-containing protein-9) (122), GASZ (a germ cell-specific protein with four ankyrin repeats, a sterile alpha motif, and a basic leucine zipper domain) (123), GTSF1 (gametocyte specific factor 1) (124), and MOV10L1 (Moloney leukemia virus 10 like-1) (125, 126). Mice deficient in any of these genes inevitably manifest meiotic arrest of spermatogenesis and male sterility. Such phenotype is commonly interpreted as a direct consequence of derepression of transposable

elements, but mechanistic insights on how L1 derepression begets meiotic arrest remain lacking.

4.3. Posttranscriptional regulation of L1 expression

The second layer of L1 regulation is posttranscriptional. The term posttranscriptional is broadly used here to describe any regulatory mechanisms that act downstream of RNA transcription. Candidates for posttranscriptional regulation of L1 expression include RNA processing (discussed in Section 3.1.4.), RNA interference (RNAi), piRNAs, and potential translational and post-translational mechanisms.

4.3.1. Host factors for posttranscriptional regulation

Several cytidine deaminases and TREX1 (3' repair exonuclease 1) are promising host factors for posttranscriptional regulation of L1 activity. APOBEC3 (apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3), APOBEC1 and AID (activation-induced deaminase) belong to a family of RNA or DNA cytidine deaminases, which can mutate RNA (or DNA) by converting cytidine to uridine (or deoxycytidine to deoxyuridine) (127). The mouse genome has a single APOBEC3 gene (mA3), while humans possess seven APOBEC3 genes (hA3A, hA3B, hA3C, hA3DE, hA3F, hA3G, and hA3H). The expansion of the APOBEC3 family during mammalian evolution coincides with a reduced level of overall retrotransposition activity from endogenous retrotransposons in primates (127). Several human APOBEC3 genes (128-133), but not mouse mA3 (129, 134), inhibit L1 retrotransposition *in vitro*. Similarly, the overexpression of APOBEC1 or AID from human, rat and mouse can inhibit the retrotransposition activity of a cloned human L1 in human cell lines (135, 136). It is puzzling, however, that L1 inhibition by all these deaminase appears to be independent of their deaminase activity. It is also noteworthy that the effect of these cytidine deaminases varies greatly from study to study (127, 137). TREX1 is an abundant 3' to 5' single-stranded DNA (ssDNA) exonuclease found in both human and mouse cells. Overexpression of TREX1 reduces L1 retrotransposition *in vitro* by 5-fold (138). In addition, *Trex1*-deficient mice display an approximately 6-fold accumulation of L1-related ssDNA in the heart. These L1-related ssDNAs potentially reflect a failure to degrade reverse-transcribed L1 ssDNAs in these *Trex1* knockout animals (138). It remains to be investigated whether any of these host factors plays a role in controlling endogenous L1 retrotransposition, especially in tissues that naturally have high levels of L1 expression. For example, AID is expressed in both mouse PGCs and early embryos at a time window when demethylation occurs. *Aid*-deficient mouse PGCs undergo inefficient genome-wide erasure of DNA methylation, but the level of L1 expression/retrotransposition has not been evaluated in these cells (139).

4.3.2. RNA interference (RNAi)

The RNAi pathway is a conserved eukaryotic mechanism for post-transcriptional regulation, and involves two types of small RNA molecules (140). MicroRNAs (miRNAs) are usually derived from endogenously expressed precursor transcripts. Small interfering RNAs (siRNAs) were initially found to be of exogenous origin, but endogenous siRNAs have been recently discovered in a variety of species. In mammals, both types of small RNAs are processed by Dicer, a member of the RNase III nuclease family. One of the two strands, termed the guide strand, is incorporated into the RNA-induced silencing complex, which subsequently targets and cleaves complementary RNA species.

Human L1 5'UTR contains both a sense and antisense promoter, and bidirectional transcription of L1 5'UTR has the potential to generate double-stranded RNA specific to L1 5'UTR. Indeed, both sense- and antisense-strand small RNAs of ~21 bp are detected by Northern blot analysis of human 293 cells (141). In a reporter assay, the presence of the antisense promoter in 5'UTR decreases the level of 5'UTR-derived transcripts through post-transcriptional mRNA degradation (141). These results suggest that human L1 may have evolved to modulate the abundance of its own transcripts by co-opting the conserved cellular RNAi pathway. However, whether or not the observed L1 siRNAs arise from double-stranded precursors has not been determined. Additionally, it remains to be investigated whether siRNAs can be generated at other regions of an L1 sequence (142). There appears to be an inverse correlation between Dicer function and L1 expression level in *Dicer*-deficient ESCs (143), but it is not known whether such a correlation is simply due to an indirect effect. When Dicer function is selectively knocked out in mouse PGCs, the level of L1 RNA expression is reduced by 2-fold at E13.5, and shows no significant change at P3 (144). The latter results suggest Dicer-mediated processes may be dispensable for early germ cell development.

4.3.3. piRNA-mediated posttranscriptional silencing in germ cells

In addition to its role in *de novo* DNA methylation (Section 4.2.4.), recent studies suggest that piRNA machinery is also involved in posttranscriptional silencing of transposable elements. piRNA pathway components localize to two distinct types of cytoplasmic granules in fetal gonocytes: pi-bodies contain the MILI-TDRD1 module, and piP-bodies contain the MIWI2-TDRD9-MAEL module (77). The latter are considered as a specialized form of P-bodies because piP-bodies and P-bodies share several mRNA degradation/translational repression components. MAEL plays a critical role in piP-body formation. In *Mael*-deficient gonocytes, pi-body composition is not affected, but MIWI2, TDRD9 and MVH (mouse VASA homolog protein) fail to localize to piP-bodies (77). As a consequence, secondary piRNA production is abrogated, and *de novo* DNA methylation of L1 5'UTR sequences is delayed in *Mael*-deficient gonocytes. Interestingly, the defect in piP-body formation is accompanied by a prolonged accumulation of L1 ORF1p signals in postnatal spermatogonia even after DNA methylation has been fully restored by E18.5. This temporal decoupling of *de novo* DNA methylation and L1 protein expression can be explained by a failure in piRNA-mediated postranscriptional/translational silencing (77), the remarkable stability of L1 ORF1p as demonstrated in

solution (145), or both. Additional support for piRNA-mediated posttranscriptional silencing of L1 comes from an experiment in which the expression of a single specific antisense L1 piRNA is down-regulated (146). In contrast to global piRNA knockout studies, the down-regulation of this specific L1 piRNA leads to L1 derepression in the absence of any overt effect on spermatogenesis and male fertility (146).

4.3.4. Cytoplasmic sequestration of L1 ribonucleoprotein (RNP) complexes

The formation of a cytoplasmic RNP complex between L1 proteins and L1 mRNA is a proposed intermediate step for mouse and human L1 retrotransposition (73, 75, 147, 148). Genetic and biochemical studies indicate that proteins expressed from active full-length L1s preferentially mobilize the source template L1 mRNA via a phenomenon termed *cis*-preference (21-23). Both ORF1p and ORF2p localize predominantly to cytoplasm when overexpressed in human cell lines. In addition, the colocalization of L1 RNA and proteins is often associated with markers of cytoplasmic stress granules (149-151). Stress granules represent sites of stalled translation initiation complexes that accumulate in response to a variety of stress conditions (152). Thus, it is possible that cells recognize overexpression of L1 as a stressor and consequently sequester L1 RNP in stress granules for degradation (150). However, it remains unknown whether transit through stress granules by an L1 RNP is a requisite step for productive L1 retrotransposition (150, 151).

5. L1 TRANSGENIC MOUSE MODELS

Endogenous L1 studies have provided a wealth of information about L1 expression and regulation, but they are not without limitations. First, high levels of background transcriptional signals are often produced from the vast number of preexisting L1 sequences in the genome. As a consequence, the reliable detection of endogenous L1 RNA relies heavily on methodology that can differentiate full-length, authentic L1 transcripts from abundant transcriptional noises. Northern blot analysis is well suited for this purpose but it lacks the level of sensitivity that quantitative RT-PCR offers. Low but physiologically significant levels of L1 expression might have escaped detection by Northern blot analysis in earlier studies. Second, *de novo* L1 insertions by endogenous L1s are intractable by conventional techniques, because it is equivalent to "finding a needle in a haystack" of preexisting insertions. Third, standard loss- or gain-of-function studies by targeted mutagenesis cannot be performed on endogenous L1s due to its repetitive nature. These limitations, along with the prospect of developing L1 as useful insertional mutagens, prompted transgenic studies of tagged L1s in rodents (153). So far, at least nine different L1 transgenes, regulated by endogenous L1 promoters and/or heterologous promoters, have been studied (63, 153-160). These transgenic models of L1 have started to provide exciting new insights in L1 biology, especially on how L1 functions *in vivo*. In this section we will highlight the progress made concerning these models, with a focus on transgenic mouse models using endogenous L1 promoters.

5.1. L1 expression from transgenes

Under the regulation of their native 5'UTR promoters, human L1 transgenes provide a rare opportunity for tracking L1 expression from a specific L1 locus (63, 153, 158). On one hand, the expression of L1 transgenes mimics that of their endogenous counterparts, i.e., in male and female gonads (153). On the other hand, these studies also reveal some pleasant surprises, such as the unexpected expression of L1 in the brain (63) and an abundance of L1 transgene transcripts in preimplantation embryos (158). However, a thorough survey of L1 transgene expression that parallels endogenous L1 studies is still lacking.

5.1.1. Expression in germ cells

Human L1 transgenes are expressed in both male and female germ cells. In a study featuring an active human L1 isolated from a patient with X-linked retinitis pigmentosa ($L1_{RP}$) (161), strand-specific transcripts from the 5'UTR-$L1_{RP}$ transgene are found only in the testis and ovary among all tissues examined (153). Specifically, L1 transcripts are detected in pachytene spermatocytes, round spermatids, and elongating spermatids from adult testes (153). These results are consistent with observations from endogenous L1 studies (Section 3.3.2.): the endogenous mouse ORF1p is detected in the pachytene spermatocytes by immunofluorescence (78), and in round spermatids and elongating spermatids by immunohistochemistry (60). The above germ cell expression pattern has been replicated by a subsequent study using independently derived $L1_{RP}$ transgenic mouse and rat lines (158). However, in contrast to endogenous L1 studies, L1 transgene expression has not been examined in germ cell populations from fetal and prepuberal mouse testes.

5.1.2. Expression in preimplantation embryos

One surprising finding from 5'UTR-driven L1 transgenes is the detection of abundant L1 transgene transcripts in preimplantation embryos (158). This particular study features several human L1 transgenes, including $L1_{RP}$ transgenic mouse and rat lines (the same mouse line reported in reference (63)), a transgene based on a different human L1 isolate (LRE3 for L1 retrotransposable element-3, isolated from a patient with chronic granulomatous disease (162)), and a transgene based on an active endogenous mouse L1 (G_F21, (20)). Transcripts of these L1 transgenes are readily detected by RT-PCR in donor-positive preimplantation embryos (morulae or early blastocysts) (158). Interestingly, a significant proportion of donor-negative embryos (~40%) are also positive for L1 expression. The latter finding suggests that these RNAs are expressed prior to the completion of gametogenesis and subsequently partitioned away from the donor transgene during meiosis (158). Such "carryover" RNAs are functional because retrotransposition events are observed in ~10% of donor-negative animals (158). This high prevalence of carry-over RNA in donor-negative animals is unprecedented. It may result from a higher level of expression from transgenes as compared to endogenous elements. A systematic survey of endogenous

L1 expression in preimplantation embryos should provide further insights about L1 activities during this important developmental stage, as full-length endogenous L1 transcripts have been previously detected in mouse blastocysts (76).

5.1.3. Expression in somatic tissues

Currently there is little evidence that 5'UTR-driven L1 transgenes are expressed in adult somatic tissues. Inconsistent and usually faint signals can be detected by RT-PCR in lung tissue from 5'UTR-driven $L1_{RP}$ lines, but their significance is not known (153, 158). Presumably, 5'UTR-driven human $L1_{RP}$ transgenes are expressed in the mouse brain during fetal and adult neurogenesis because of observed *de novo* retrotransposition events launched from the donor transgene (63). In this regard, the detection of L1 RNA in the head portion of E10.5–E11.5 embryos provides a necessary link between L1 expression and retrotransposition in the fetal brain (158). However, to date, no L1 RNA signal has been seen in adult brain samples (153, 158), although it is unknown whether or not regions active in adult neurogenesis, such as the subventricular zone of the lateral ventricle and the subgranular zone of the hippocampus (163), were sampled in these studies. Immunofluorescence staining of wild-type brain sections indicates that a very small number of cells in the ventricular zone and the dentate gyrus of hippocampus are reactive to an antibody against mouse ORF1p (63). It is yet to be demonstrated how the expression level of human L1 transgenes compares to endogenous expression.

5.2. L1 retrotransposition in mouse models

L1 transgenes are tagged by retrotransposition indicator elements, and thus *de novo* retrotransposition events can be specifically tracked. Most frequently, tracking is achieved by using GFP (green fluorescent protein) expression cassettes that are specifically designed for reporting retrotransposition events (153). However, the placement of an intron alone in the 3'UTR of an L1 transgene also allows the detection of retrotransposition by PCR-based techniques (158). In this subsection, findings from transgenic mouse models are divided according to different stages of mouse development.

5.2.1. Retrotransposition in germ cells

Consistent with L1 expression in germ cells, retrotransposition events by 5'UTR-driven L1 transgenes are detectable in testes (63, 153, 158). Using RT-PCR, an earlier study detected a strand-specific spliced EGFP signal, which is indicative of retrotransposition, in elongating spermatids from one out of four 5'UTR $L1_{RP}$ lines, as well as from two control lines driven by a mouse RNA polymerase II large subunit (PolII) promoter (153). The germline retrotransposition frequency is low in elongating spermatids (~0.1% as estimated by RT-PCR (153), and a very faint amplicon by PCR (158)). Different 5'UTR lines display significant variation of retrotransposition, which appears to be correlated with the transcription level. For example, the only 5'UTR line with detectable level of retrotransposition by RT-PCR also has the highest level of transgene expression (153). Retrotransposition in the testis of transgenic animals has also been demonstrated for an independent $L1_{RP}$ transgene by indirect immunofluorescence detection of GFP proteins (63). Thus, it is evident that *de novo* retrotransposition from 5'UTR-driven human L1 transgenes can occur in germ cells. Perhaps reflecting its low frequency, so far no heritable germline insertion has been reported in mice carrying 5'UTR-driven L1 human transgenes (153, 158). In contrast, heritable germline insertions have been observed in L1 transgenic lines carrying heterologous promoters, such as PolII promoter (153) and a composite chicken beta-actin (CAG) promoter (156, 157).

5.2.2. Retrotransposition in preimplantation embryos

De novo retrotransposition can occur during early human embryogenesis, as evidenced by a case study for a specific L1 insertion in a nuclear family with choroideremia ($L1_{CHM}$) (164). Genotyping data indicate that the mother was both germline mosaic and somatic mosaic for $L1_{CHM}$ insertion, and thus this retrotransposition event must have occurred during early embryogenesis, before the patient's germ cells became a distinct lineage (164) (in mouse this happens around E6.25 (165)). Since $L1_{CHM}$ is an isolated case, the magnitude of such early embryonal events has not been appreciated until recently. Several interesting observations are made when retrotransposition is measured by PCR in preimplantation embryos (morula or blastocyst stage) from transgenic rodent models of both human and mouse L1s (158). First, retrotransposition is much more frequent in preimplantation embryos than in male germ cells, although a similar, if not higher, level of L1 expression is observed in germ cells (158). This discrepancy between transcription and retrotransposition suggests a post-transcriptional block of L1 retrotransposition in developing germ cells, or a reduced fitness for germ cells that are affected by retrotransposition or simply by L1 protein products. Second, retrotransposition occurs in both donor-positive and donor-negative preimplantation embryos (158). The occurrence of retrotransposition in donor-negative preimplantation embryos is attributable to carryover RNA from germ cells (Section 5.5.1.). Interestingly, donor-positive and negative preimplantation embryos display a similarly low level of retrotransposition. The frequency of retrotransposition is estimated by semi-quantitative PCR to be approximately one in every fifty cells in late blastocysts (158). Thus, only one or two cells in a late blastocyst have an L1 insertion (the average number of cells in mouse late blastocysts is around eighty (166)). Indeed, no retrotransposition signal is detected from the majority of preimplantation embryos (158). Third, none of the insertions from donor-negative animals pass through the germ line, indicating that the majority of putative preimplantation insertion events occurred in cells not destined to become germ cells (158).

5.2.3. Retrotransposition during postimplantation and postnatal development

Several lines of evidence suggest that retrotransposition can occur in postimplantation embryos. First, GFP-positive cells, indicative of *de novo* retrotransposition events, are detected in the cephalic neural tube of E10.5 embryos (63). Second, retrotransposition in postimplantation embryos is supported by the comparison

of the timing and relative retrotransposition frequencies between donor-positive and donor-negative samples. As discussed above (Section 5.2.2.), retrotransposition signals are present at a similarly low level in either donor-positive or donor-negative preimplantation embryos (158). Presumably, L1 RNA is continuously transcribed in donor-positive embryos post-fertilization, because all donor-positive preimplantation embryos show positive L1 RNA signals by RT-PCR. L1 RNA can also be detected in the head portion of E10.5–E11.5 donor-positive embryos (158). In contrast, L1 RNA is no longer detectable in E10.5 donor-negative embryos (158). Accordingly, retrotransposition can be readily detected by PCR in 64% of donor-positive and 9% of donor-negative weanlings. Retrotransposition signals, when measured by quantitative PCR, are roughly 10-fold higher in donor-positive animals than in donor-negative animals. These data suggest that most somatic retrotransposition events in these transgenic models occur during postimplantation and/or postnatal development. A detailed timeline of L1 expression and retrotransposition is of particular interest to developmental biologists, but must await further studies.

The physiological consequence of such somatic retrotransposition events, if any, remains to be determined. Significant variation of retrotransposition signals is observed not only among different transgenic animals but also among distinct tissues from the same animal (158). The frequency of retrotransposition, as determined by quantitative PCR, is higher in somatic cells (ranging from 0.02 to 0.002 per cell for brain, kidney, liver, lung and tail tissues), and lower in germ cells (0.001 per sperm) (158). However, in a separate study using the same $L1_{RP}$ mouse line, no GFP-positive cells are detected by immunofluorescence in several adult somatic tissues (heart, kidney, intestine, liver, lung, muscle, skin, and spleen); brain is the only exception (63). Such discrepancy may be explained by the methodological difference between the two studies. It is possible that the low frequency of retrotransposition, as detected by PCR, evades immunofluorescence detection. Another fundamental difference between the methods is that immunofluorescence detection requires GFP expression from L1 insertions. Therefore, an alternative possibility is that most somatic tissues do not support GFP expression, likely as the result of transcriptional silencing of L1 insertions. The latter explanation is at least partially supported by data from *in vitro* experiments with rat HCN stem cells (63). HCN cells containing an L1 insertion are largely negative for GFP fluorescence unless they are differentiated into neurons (63). Inhibition of DNA methylation or histone deacetylation increases the number of GFP-positive HCN cells (63). Transcriptional silencing of L1 insertions has also been demonstrated in embryonal carcinoma cells (167).

5.2.4. Retrotransposition in the brain

Another exciting finding from transgenic L1 studies is the discovery of active L1 retrotransposition from L1 transgenes in the brain, an organ that is least suspected for L1 activity based on previous endogenous L1 expression patterns (63). In this study, retrotransposition events are detected by indirect immunofluorescence against GFP proteins, which can only be expressed if *de novo* retrotransposition has occurred from donor human $L1_{RP}$ transgenes. Surprisingly, scattered GFP-positive cells are observed in a wide-range of adult brain regions, such as amygdala, cerebellum, cortex, hilus, hippocampus, hypothalamus, striatum, and ventricles (63). Staining for cell-specific markers indicates these EGFP-positive cells are neurons rather than oligodendrocytes or astrocytes (63). In a follow-up study, quantitative PCR analysis indicates an increased number of L1 ORF2 sequences in several regions of adult human brain as compared to other somatic tissues, such as heart and liver (168). Such results provide further support that endogenous L1s might express and retrotranspose in the human brain. However, a formal proof for brain-specific retrotransposition necessitates the mapping of new L1 insertions that are absent in control organs. A potential link between DNA methylation and L1 expression has been investigated by bisulfite sequencing of 5'UTR from fetal brain and skin tissues. Overall, two brain samples display a subtle but consistently lower level of DNA methylation when compared to matched skin samples, although it is difficult to evaluate the significance of such subtle changes of DNA methylation on L1 expression. In addition to locating GFP-positive cells in the adult brain, GFP signals are also detected in the cephalic neural tube of E10.5 embryos (63). Thus, the determination of the timing and dynamics of *de novo* retrotransposition events in the developing brain should provide important insights about the regulation of L1 expression in the neuronal system.

5.3. Challenges and opportunities for L1 transgenic models

So far, transgenic models of L1 have brought forth quite a few new discoveries (detailed in Sections 5.1. and 5.2.). For example, L1 transcription is abundantly detected in germ cells, but germ cells display the least amount of retrotransposition activity. Retrotransposition can also occur in donor-negative animals via RNA carryover from parental germ cells. L1 expression from transgenes spawns insertions in preimplantation embryos (albeit at relatively low frequency) and during postimplantation and postnatal development. For the first time, L1 transgenes enabled the detection of retrotransposition events in the adult brain and the developing neural tube, representing an important step toward understanding potential roles that L1 plays in generating neuronal diversification. While the advent of these exciting discoveries merits celebration, their impact will not be fully appreciated without rigorous follow-up studies. Especially, some aspects of these findings have yet to be corroborated by alternative approaches. In this section, we discuss some of the opportunities and challenges that face L1 transgenic studies.

5.3.1. Are human and mouse L1 families similarly regulated *in vivo*?

It is debatable whether L1 families from different mammalian species are regulated in a similar fashion in their respective cellular environment *in vivo*. It appears that host regulation of L1 expression hinges on L1 promoters.

This control is primarily achieved at the transcriptional level, although some aspects of L1 regulation occur at the posttranscriptional phase. These regulatory mechanisms include DNA methylation, piRNAs, siRNAs, and transcription factors (Section 4.). As the result of repeated acquisition of novel 5' sequences during evolution, different L1 families possess 5'UTR promoters with little sequence conservation (8). Most of the known sequence motifs present in the 5'UTR of currently active human Ta subfamily are only found in more recent primate-specific L1 families (17, 85). These motifs include transcription factor binding sites for p53, RUNX3 and the SOX family proteins, and the human L1 antisense promoter. The SOX-binding site is an interesting case. Despite the overall lack of sequence conservation, both mouse and rat L1s have evolved to contain at least one copy of the overlapping consensus SOX and TCF/LEF binding site (Section 4.1.). Species-specific sequence motifs and the species-specific differences of host factors, such as transcription factors and APOBEC3 proteins, may underlie differential L1 activities among mammalian lineages. DNA methylation, as a key mechanism for heritable silencing of retrotransposons, may operate similarly among humans, rats, and mice. L1 promoters from all three species are CpG rich, and likely subject to similar cycles of demethylation and remethylation during development (Section 4.2.). piRNA biogenesis is intimately coupled with *de novo* DNA methylation during fetal germ cell development in mice. piRNA has also been isolated from rat testes (169). Primary piRNAs are likely generated as a response to an elevated level of L1 transcription due to the initial wave of demethylation in fetal germ cells. The processing of full-length L1 transcripts into short piRNAs may also reduce the overall level of functional L1 transcripts during this critical stage of germ cell development (Sections 4.2. and 4.3.). Nevertheless, L1 proteins can be readily detected (Section 3.1.4. for human samples, and Section 3.3.2. for mouse samples). The physiological consequences, if any, of such levels of L1 expression have not been evaluated.

5.3.2. Do L1 copy number and genomic location matter?

L1 transgenes differ from their endogenous counterparts in both copy number and genomic distribution. L1 transgenes are typically present in the genome as tandem-repeated copies in a single genomic locus, a byproduct of transgenesis (170). In contrast, endogenous L1s are present largely as dispersed individual copies. Both copy number and genomic location can impact transgene expression. In particular, tandem-repeated transgenes frequently display position effects, i.e., they show substantial variation of transgene expression among different transgenic mouse lines according to their location in the genome (171). In addition, transgenes are subject to repeat-induced gene silencing (RIGS), which typically manifests as an inverse correlation between transgene copy number and expression level from a specific genomic locus (172, 173). The potential effect of RIGS on L1 transgenic studies has been demonstrated by a recent study employing an L1 transgene under the regulation of a heterologous, constitutively expressing promoter (160). Specifically, the reduction of transgene copy number to a single copy at the same genomic locus shows no decrease in retrotransposition activity. An average frequency of 0.45 insertions per cell is observed for animals carrying a single-copy donor transgene, representing a nine-fold increase of retrotransposition activity on a per-copy basis. The observed retrotransposition activity is correlated with differential CpG methylation at the heterologous promoter: the promoter region is largely methylated in animals with the high-copy transgene array but significantly hypomethylated in animals with the single-copy counterpart; the body and 3' end of the transgene show high levels of methylation in both high-copy and single-copy animals (160).

It remains an open question whether RIGS applies to tandem-repeated L1 transgenes under the regulation of endogenous L1 promoters. At the transcriptional level, if the 5'UTR is the primary target for piRNA-guided *de novo* methylation and silencing of L1 retrotransposons, it is reasonable to expect that a single-copy L1 transgene containing a mouse L1 5'UTR promoter will be recognized by endogenous piRNAs, and regulated similarly as compared to endogenous mouse L1s. A human L1, when placed as a transgene in the mouse cellular environment, may elicit efficient piRNA generation if the piRNA biogenesis is purely a function of L1 expression. It remains to be determined whether the high number and dispersed status of endogenous L1s make these retrotransposons more vulnerable to DNA methylation than introduced L1 transgenes.

6. PERSPECTIVE

It is now abundantly clear that L1 is expressed during development and disease. However, in most cases, a definitive role for L1 expression has yet to be established. Currently, our knowledge about the developmental timing and cell type specificity of L1 expression and its regulation comes predominantly from studies of L1 in mice. In addition, both commonalities and contrarieties on L1 regulation have been revealed for human, rat, and mouse L1 retrotransposons. The field as a whole has just started to learn the molecular inner workings of L1 regulation. This leaves great opportunities as well as grand challenges for modeling human L1 as transgenes in mice. Especially, the similarities and differences between human and mouse L1 families, including both *cis*- and *trans*-acting factors, should always be considered. A comparative approach between humans and mice is thus recommended.

7. ACKNOWLEDGMENTS

We thank Drs. Ronald Brosemer, Cathryn Hogarth, and Yi Xie for critical reading of the manuscript. Research in our laboratory is supported by American Cancer Society (IRG-77-003-32) and start-up funds from Washington State University. J.M.R. is supported in part by NIH Biotechnology Training Program (5T32GM008336-22).

8. REFERENCES

1. R. A. Gibbs, G. M. Weinstock, M. L. Metzker, D. M. Muzny, E. J. Sodergren, S. Scherer, G. Scott, D. Steffen, K.

C. Worley, P. E. Burch, G. Okwuonu, S. Hines, L. Lewis, C. DeRamo, O. Delgado, S. Dugan-Rocha, G. Miner, M. Morgan, A. Hawes, R. Gill, Celera, R. A. Holt, M. D. Adams, P. G. Amanatides, H. Baden-Tillson, M. Barnstead, S. Chin, C. A. Evans, S. Ferriera, C. Fosler, A. Glodek, Z. Gu, D. Jennings, C. L. Kraft, T. Nguyen, C. M. Pfannkoch, C. Sitter, G. G. Sutton, J. C. Venter, T. Woodage, D. Smith, H. M. Lee, E. Gustafson, P. Cahill, A. Kana, L. Doucette-Stamm, K. Weinstock, K. Fechtel, R. B. Weiss, D. M. Dunn, E. D. Green, R. W. Blakesley, G. G. Bouffard, P. J. De Jong, K. Osoegawa, B. Zhu, M. Marra, J. Schein, I. Bosdet, C. Fjell, S. Jones, M. Krzywinski, C. Mathewson, A. Siddiqui, N. Wye, J. McPherson, S. Zhao, C. M. Fraser, J. Shetty, S. Shatsman, K. Geer, Y. Chen, S. Abramzon, W. C. Nierman, P. H. Havlak, R. Chen, K. J. Durbin, A. Egan, Y. Ren, X. Z. Song, B. Li, Y. Liu, X. Qin, S. Cawley, A. J. Cooney, L. M. D'Souza, K. Martin, J. Q. Wu, M. L. Gonzalez-Garay, A. R. Jackson, K. J. Kalafus, M. P. McLeod, A. Milosavljevic, D. Virk, A. Volkov, D. A. Wheeler, Z. Zhang, J. A. Bailey, E. E. Eichler, E. Tuzun, E. Birney, E. Mongin, A. Ureta-Vidal, C. Woodwark, E. Zdobnov, P. Bork, M. Suyama, D. Torrents, M. Alexandersson, B. J. Trask, J. M. Young, H. Huang, H. Wang, H. Xing, S. Daniels, D. Gietzen, J. Schmidt, K. Stevens, U. Vitt, J. Wingrove, F. Camara, M. Mar Alba, J. F. Abril, R. Guigo, A. Smit, I. Dubchak, E. M. Rubin, O. Couronne, A. Poliakov, N. Hubner, D. Ganten, C. Goesele, O. Hummel, T. Kreitler, Y. A. Lee, J. Monti, H. Schulz, H. Zimdahl, H. Himmelbauer, H. Lehrach, H. J. Jacob, S. Bromberg, J. Gullings-Handley, M. I. Jensen-Seaman, A. E. Kwitek, J. Lazar, D. Pasko, P. J. Tonellato, S. Twigger, C. P. Ponting, J. M. Duarte, S. Rice, L. Goodstadt, S. A. Beatson, R. D. Emes, E. E. Winter, C. Webber, P. Brandt, G. Nyakatura, M. Adetobi, F. Chiaromonte, L. Elnitski, P. Eswara, R. C. Hardison, M. Hou, D. Kolbe, K. Makova, W. Miller, A. Nekrutenko, C. Riemer, S. Schwartz, J. Taylor, S. Yang, Y. Zhang, K. Lindpaintner, T. D. Andrews, M. Caccamo, M. Clamp, L. Clarke, V. Curwen, R. Durbin, E. Eyras, S. M. Searle, G. M. Cooper, S. Batzoglou, M. Brudno, A. Sidow, E. A. Stone, B. A. Payseur, G. Bourque, C. Lopez-Otin, X. S. Puente, K. Chakrabarti, S. Chatterji, C. Dewey, L. Pachter, N. Bray, V. B. Yap, A. Caspi, G. Tesler, P. A. Pevzner, D. Haussler, K. M. Roskin, R. Baertsch, H. Clawson, T. S. Furey, A. S. Hinrichs, D. Karolchik, W. J. Kent, K. R. Rosenbloom, H. Trumbower, M. Weirauch, D. N. Cooper, P. D. Stenson, B. Ma, M. Brent, M. Arumugam, D. Shteynberg, R. R. Copley, M. S. Taylor, H. Riethman, U. Mudunuri, J. Peterson, M. Guyer, A. Felsenfeld, S. Old, S. Mockrin and F. Collins: Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, 428 (6982), 493-521 (2004)

2. E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi and Y. J. Chen: Initial sequencing and analysis of the human genome. *Nature*, 409 (6822), 860-921 (2001)

3. R. H. Waterston, K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, P. An, S. E. Antonarakis, J. Attwood, R. Baertsch, J. Bailey, K. Barlow, S. Beck, E. Berry, B. Birren, T. Bloom, P. Bork, M. Botcherby, N. Bray, M. R. Brent, D. G. Brown, S. D. Brown, C. Bult, J. Burton, J. Butler, R. D. Campbell, P. Carninci, S. Cawley, F. Chiaromonte, A. T. Chinwalla, D. M. Church, M. Clamp, C. Clee, F. S. Collins, L. L. Cook, R. R. Copley, A. Coulson, O. Couronne, J. Cuff, V. Curwen, T. Cutts, M. Daly, R. David, J. Davies, K. D. Delehaunty, J. Deri, E. T. Dermitzakis, C. Dewey, N. J. Dickens, M. Diekhans, S.

Dodge, I. Dubchak, D. M. Dunn, S. R. Eddy, L. Elnitski, R. D. Emes, P. Eswara, E. Eyras, A. Felsenfeld, G. A. Fewell, P. Flicek, K. Foley, W. N. Frankel, L. A. Fulton, R. S. Fulton, T. S. Furey, D. Gage, R. A. Gibbs, G. Glusman, S. Gnerre, N. Goldman, L. Goodstadt, D. Grafham, T. A. Graves, E. D. Green, S. Gregory, R. Guigo, M. Guyer, R. C. Hardison, D. Haussler, Y. Hayashizaki, L. W. Hillier, A. Hinrichs, W. Hlavina, T. Holzer, F. Hsu, A. Hua, T. Hubbard, A. Hunt, I. Jackson, D. B. Jaffe, L. S. Johnson, M. Jones, T. A. Jones, A. Joy, M. Kamal, E. K. Karlsson, D. Karolchik, A. Kasprzyk, J. Kawai, E. Keibler, C. Kells, W. J. Kent, A. Kirby, D. L. Kolbe, I. Korf, R. S. Kucherlapati, E. J. Kulbokas, D. Kulp, T. Landers, J. P. Leger, S. Leonard, I. Letunic, R. Levine, J. Li, M. Li, C. Lloyd, S. Lucas, B. Ma, D. R. Maglott, E. R. Mardis, L. Matthews, E. Mauceli, J. H. Mayer, M. McCarthy, W. R. McCombie, S. McLaren, K. McLay, J. D. McPherson, J. Meldrim, B. Meredith, J. P. Mesirov, W. Miller, T. L. Miner, E. Mongin, K. T. Montgomery, M. Morgan, R. Mott, J. C. Mullikin, D. M. Muzny, W. E. Nash, J. O. Nelson, M. N. Nhan, R. Nicol, Z. Ning, C. Nusbaum, M. J. O'Connor, Y. Okazaki, K. Oliver, E. Overton-Larty, L. Pachter, G. Parra, K. H. Pepin, J. Peterson, P. Pevzner, R. Plumb, C. S. Pohl, A. Poliakov, T. C. Ponce, C. P. Ponting, S. Potter, M. Quail, A. Reymond, B. A. Roe, K. M. Roskin, E. M. Rubin, A. G. Rust, R. Santos, V. Sapojnikov, B. Schultz, J. Schultz, M. S. Schwartz, S. Schwartz, C. Scott, S. Seaman, S. Searle, T. Sharpe, A. Sheridan, R. Shownkeen, S. Sims, J. B. Singer, G. Slater, A. Smit, D. R. Smith, B. Spencer, A. Stabenau, N. Stange-Thomann, C. Sugnet, M. Suyama, G. Tesler, J. Thompson, D. Torrents, E. Trevaskis, J. Tromp, C. Ucla, A. Ureta-Vidal, J. P. Vinson, A. C. Von Niederhausern, C. M. Wade, M. Wall, R. J. Weber, R. B. Weiss, M. C. Wendl, A. P. West, K. Wetterstrand, R. Wheeler, S. Whelan, J. Wierzbowski, D. Willey, S. Williams, R. K. Wilson, E. Winter, K. C. Worley, D. Wyman, S. Yang, S. P. Yang, E. M. Zdobnov, M. C. Zody and E. S. Lander: Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420 (6915), 520-62 (2002)

4. A. F. Smit, G. Toth, A. D. Riggs and J. Jurka: Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J Mol Biol*, 246 (3), 401-417 (1995)

5. S. Boissinot and A. V. Furano: The recent evolution of human L1 retrotransposons. *Cytogenet Genome Res*, 110 (1-4), 402-6 (2005)

6. S. Boissinot, P. Chevret and A. V. Furano: L1 (LINE-1) retrotransposon evolution and amplification in recent human history. *Mol Biol Evol*, 17 (6), 915-28 (2000)

7. S. Boissinot, J. Davis, A. Entezam, D. Petrov and A. V. Furano: Fitness cost of LINE-1 (L1) activity in humans. *Proc Natl Acad Sci U S A*, 103 (25), 9590-4 (2006)

8. A. V. Furano: The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. *Prog Nucleic Acid Res Mol Biol*, 64, 255-94 (2000)

9. R. Minakami, K. Kurose, K. Etoh, Y. Furuhata, M. Hattori and Y. Sakaki: Identification of an internal cis-element essential for the human L1 transcription and a nuclear factor (s) binding to the element. *Nucleic Acids Res*, 20 (12), 3139-45 (1992)

10. I. Nur, E. Pascale and A. V. Furano: The left end of rat L1 (L1Rn, long interspersed repeated) DNA which is a CpG island can function as a promoter. *Nucleic Acids Res*, 16 (19), 9233-51 (1988)

11. D. M. Severynse, C. A. Hutchison, 3rd and M. H. Edgell: Identification of transcriptional regulatory activity within the 5' A-type monomer sequence of the mouse LINE-1 retroposon. *Mamm Genome*, 2 (1), 41-50 (1992)

12. G. D. Swergold: Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol Cell Biol*, 10 (12), 6718-29 (1990)

13. C. F. Voliva, C. L. Jahn, M. B. Comer, C. A. Hutchison, 3rd and M. H. Edgell: The L1Md long interspersed repeat family in the mouse: almost all examples are truncated at one end. *Nucleic Acids Res*, 11 (24), 8847-59 (1983)

14. G. Grimaldi, J. Skowronski and M. F. Singer: Defining the beginning and end of KpnI family segments. *EMBO J*, 3 (8), 1753-9 (1984)

15. J. V. Moran, S. E. Holmes, T. P. Naas, R. J. DeBerardinis, J. D. Boeke and H. H. Kazazian, Jr.: High frequency retrotransposition in cultured mammalian cells. *Cell*, 87 (5), 917-27 (1996)

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

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

18. K. Akagi, J. Li, R. M. Stephens, N. Volfovsky and D. E. Symer: Extensive variation between inbred mouse strains due to endogenous L1 retrotransposition. *Genome Res*, 18 (6), 869-80 (2008)

19. B. Brouha, J. Schustak, R. M. Badge, S. Lutz-Prigge, A. H. Farley, J. V. Moran and H. H. Kazazian, Jr.: Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci U S A*, 100 (9), 5280-5 (2003)

20. J. L. Goodier, E. M. Ostertag, K. Du and H. H. Kazazian, Jr.: A novel active L1 retrotransposon subfamily in the mouse. *Genome Res*, 11 (10), 1677-85 (2001)

21. C. Esnault, J. Maestre and T. Heidmann: Human LINE retrotransposons generate processed pseudogenes. *Nat Genet*, 24 (4), 363-7 (2000)

22. W. Wei, N. Gilbert, S. L. Ooi, J. F. Lawler, E. M. Ostertag, H. H. Kazazian, J. D. Boeke and J. V. Moran: Human L1 retrotransposition: cis preference versus trans complementation. *Mol Cell Biol*, 21 (4), 1429-39 (2001)

23. D. A. Kulpa and J. V. Moran: Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat Struct Mol Biol*, 13 (7), 655-60 (2006)

24. M. Dewannieux, C. Esnault and T. Heidmann: LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet*, 35 (1), 41-8 (2003)

25. M. Dewannieux and T. Heidmann: L1-mediated retrotransposition of murine B1 and B2 SINEs recapitulated in cultured cells. *J Mol Biol*, 349 (2), 241-7 (2005)

26. H. H. Kazazian, Jr.: Mobile elements: drivers of genome evolution. *Science*, 303 (5664), 1626-32 (2004)

27. J. S. Han and J. D. Boeke: LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression? *Bioessays*, 27 (8), 775-84 (2005)

28. J. S. Han, S. T. Szak and J. D. Boeke: Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature*, 429 (6989), 268-74 (2004)

29. N. A. Wallace, V. P. Belancio and P. L. Deininger: L1 mobile element expression causes multiple types of toxicity. *Gene*, 419 (1-2), 75-81 (2008)

30. S. L. Gasior, T. P. Wakeman, B. Xu and P. L. Deininger: The human LINE-1 retrotransposon creates DNA double-strand breaks. *J Mol Biol*, 357 (5), 1383-93 (2006)

31. S. M. Belgnaoui, R. G. Gosden, O. J. Semmes and A. Haoudi: Human LINE-1 retrotransposon induces DNA damage and apoptosis in cancer cells. *Cancer Cell Int*, 6, 13 (2006)

32. A. Haoudi, O. J. Semmes, J. M. Mason and R. E. Cannon: Retrotransposition-Competent Human LINE-1 Induces Apoptosis in Cancer Cells With Intact p53. *J Biomed Biotechnol*, 2004 (4), 185-194 (2004)

33. V. P. Belancio, D. J. Hedges and P. Deininger: Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. *Genome Res*, 18 (3), 343-58 (2008)

34. J. M. Chen, P. D. Stenson, D. N. Cooper and C. Ferec: A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. *Hum Genet*, 117 (5), 411-27 (2005)

35. Y. Miki, I. Nishisho, A. Horii, Y. Miyoshi, J. Utsunomiya, K. W. Kinzler, B. Vogelstein and Y. Nakamura: Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res*, 52 (3), 643-5 (1992)

36. B. Morse, P. G. Rotherg, V. J. South, J. M. Spandorfer and S. M. Astrin: Insertional mutagenesis of the myc locus by a LINE-1 sequence in a human breast carcinoma. *Nature*, 333 (6168), 87-90 (1988)

37. R. C. Iskow, M. T. McCabe, R. E. Mills, S. Torene, W. S. Pittard, A. F. Neuwald, E. G. Van Meir, P. M. Vertino and S. E. Devine: Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell*, 141 (7), 1253-61 (2010)

38. V. Perepelitsa-Belancio and P. Deininger: RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat Genet*, 35 (4), 363-6 (2003)

39. V. P. Belancio, D. J. Hedges and P. Deininger: LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic Acids Res*, 34 (5), 1512-21 (2006)

40. M. Speek: Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol Cell Biol*, 21 (6), 1973-85 (2001)

41. P. Nigumann, K. Redik, K. Matlik and M. Speek: Many human genes are transcribed from the antisense promoter of L1 retrotransposon. *Genomics*, 79 (5), 628-34 (2002)

42. A. R. Florl, R. Lower, B. J. Schmitz-Drager and W. A. Schulz: DNA methylation and expression of LINE-1 and HERV-K provirus sequences in urothelial and renal cell carcinomas. *Br J Cancer*, 80 (9), 1312-21 (1999)

43. D. M. Leibold, G. D. Swergold, M. F. Singer, R. E. Thayer, B. A. Dombroski and T. G. Fanning: Translation of LINE-1 DNA elements *in vitro* and in human cells. *Proc Natl Acad Sci U S A*, 87 (18), 6990-4 (1990)

44. J. Skowronski and M. F. Singer: Expression of a cytoplasmic LINE-1 transcript is regulated in a human teratocarcinoma cell line. *Proc Natl Acad Sci U S A*, 82 (18), 6050-4 (1985)

45. R. E. Thayer, M. F. Singer and T. G. Fanning: Undermethylation of specific LINE-1 sequences in human cells producing a LINE-1-encoded protein. *Gene*, 133 (2), 273-7 (1993)

46. L. B. Kole, S. R. Haynes and W. R. Jelinek: Discrete and heterogeneous high molecular weight RNAs complementary to a long dispersed repeat family (a possible transposon) of human DNA. *J Mol Biol*, 165 (2), 257-86 (1983)

47. J. Skowronski, T. G. Fanning and M. F. Singer: Unit-length line-1 transcripts in human teratocarcinoma cells. *Mol Cell Biol*, 8 (4), 1385-97 (1988)

48. G. L. Bratthauer, R. D. Cardiff and T. G. Fanning: Expression of LINE-1 retrotransposons in human breast cancer. *Cancer*, 73 (9), 2333-6 (1994)

49. P. Nangia-Makker, R. Sarvis, D. W. Visscher, J. Bailey-Penrod, A. Raz and F. H. Sarkar: Galectin-3 and L1 retrotransposons in human breast carcinomas. *Breast Cancer Res Treat*, 49 (2), 171-83 (1998)

50. V. P. Belancio, A. M. Roy-Engel, R. R. Pochampally and P. Deininger: Somatic expression of LINE-1 elements in human tissues. *Nucleic Acids Res*, 38 (12), 3909-22 (2010)

51. J. L. Garcia-Perez, M. C. Marchetto, A. R. Muotri, N. G. Coufal, F. H. Gage, K. S. O'Shea and J. V. Moran: LINE-1 retrotransposition in human embryonic stem cells. *Hum Mol Genet*, 16 (13), 1569-77 (2007)

52. L. H. Looijenga and J. W. Oosterhuis: Pathobiology of testicular germ cell tumors: views and news. *Anal Quant Cytol Histol*, 24 (5), 263-79 (2002)

53. J. W. Oosterhuis and L. H. Looijenga: Testicular germ-cell tumours in a broader perspective. *Nat Rev Cancer*, 5 (3), 210-22 (2005)

54. G. L. Bratthauer and T. G. Fanning: Active LINE-1 retrotransposons in human testicular cancer. *Oncogene*, 7 (3), 507-10 (1992)

55. G. L. Bratthauer and T. G. Fanning: LINE-1 retrotransposon expression in pediatric germ cell tumors. *Cancer*, 71 (7), 2383-6 (1993)

56. Y. Su, S. Davies, M. Davis, H. Lu, R. Giller, M. Krailo, Q. Cai, L. Robison and X. O. Shu: Expression of LINE-1 p40 protein in pediatric malignant germ cell tumors and its association with clinicopathological parameters: a report from the Children's Oncology Group. *Cancer Lett*, 247 (2), 204-12 (2007)

57. H. L. Asch, E. Eliacin, T. G. Fanning, J. L. Connolly, G. Bratthauer and B. B. Asch: Comparative expression of the LINE-1 p40 protein in human breast carcinomas and normal breast tissues. *Oncol Res*, 8 (6), 239-47 (1996)

58. C. R. Harris, R. Normart, Q. Yang, E. Stevenson, B. G. Haffty, S. Ganesan, C. Cordon-Cardo, A. J. Levine and L. H. Tang: Association of Nuclear Localization of a Long Interspersed Nuclear Element-1 Protein in Breast Tumors with Poor Prognostic Outcomes. *Genes Cancer*, 1 (2), 115-124 (2010)

59. S. Ergun, C. Buschmann, J. Heukeshoven, K. Dammann, F. Schnieders, H. Lauke, F. Chalajour, N. Kilic, W. H. Stratling and G. G. Schumann: Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult human tissues. *J Biol Chem*, 279 (26), 27753-63 (2004)

60. D. Branciforte and S. L. Martin: Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. *Mol Cell Biol*, 14 (4), 2584-92 (1994)

61. S. A. Trelogan and S. L. Martin: Tightly regulated, developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis. *Proc Natl Acad Sci U S A*, 92 (5), 1520-4 (1995)

62. I. Georgiou, D. Noutsopoulos, E. Dimitriadou, G. Markopoulos, A. Apergi, L. Lazaros, T. Vaxevanoglou, K. Pantos, M. Syrrou and T. Tzavaras: Retrotransposon RNA expression and evidence for retrotransposition events in human oocytes. *Hum Mol Genet*, 18 (7), 1221-8 (2009)

63. A. R. Muotri, V. T. Chu, M. C. Marchetto, W. Deng, J. V. Moran and F. H. Gage: Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature*, 435 (7044), 903-10 (2005)

64. H. Shay, M. Gruenstein, H. E. Marx and L. Glazer: The development of lymphatic and myelogenous leukemia in Wistar rats following gastric instillation of methylcholanthrene. *Cancer Res*, 11 (1), 29-34 (1951)

65. W. C. Moloney, A. D. Dorr, G. Dowd and A. E. Boschetti: Myelogenous leukemia in the rat. *Blood*, 19, 45-59 (1962)

66. K. Servomaa and T. Rytomaa: Suicidal death of rat chloroleukaemia cells by activation of the long interspersed repetitive DNA element (L1Rn). *Cell Tissue Kinet*, 21 (1), 33-43 (1988)

67. K. Servomaa and T. Rytomaa: UV light and ionizing radiations cause programmed death of rat chloroleukaemia cells by inducing retropositions of a mobile DNA element (L1Rn). *Int J Radiat Biol*, 57 (2), 331-43 (1990)

68. A. Kirilyuk, G. V. Tolstonog, A. Damert, U. Held, S. Hahn, R. Lower, C. Buschmann, A. V. Horn, P. Traub and G. G. Schumann: Functional endogenous LINE-1 retrotransposons are expressed and mobilized in rat chloroleukemia cells. *Nucleic Acids Res*, 36 (2), 648-65 (2008)

69. F. R. Witney and A. V. Furano: Highly repeated DNA families in the rat. *J Biol Chem*, 259 (16), 10481-92 (1984)

70. E. Lucchinetti, J. Feng, R. Silva, G. V. Tolstonog, M. C. Schaub, G. G. Schumann and M. Zaugg: Inhibition of LINE-1 expression in the heart decreases ischemic damage by activation of Akt/PKB signaling. *Physiol Genomics*, 25 (2), 314-24 (2006)

71. M. F. Singer: SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. *Cell*, 28 (3), 433-4 (1982)

72. J. P. Dudley: Discrete high molecular weight RNA transcribed from the long interspersed repetitive element L1Md. *Nucleic Acids Res*, 15 (6), 2581-92 (1987)

73. S. L. Martin: Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Mol Cell Biol*, 11 (9), 4804-7 (1991)

74. S. A. Schichman, D. M. Severynse, M. H. Edgell and C. A. Hutchison, 3rd: Strand-specific LINE-1 transcription in mouse F9 cells originates from the youngest phylogenetic subgroup of LINE-1 elements. *J Mol Biol*, 224 (3), 559-74 (1992)

75. S. L. Martin and D. Branciforte: Synchronous expression of LINE-1 RNA and protein in mouse embryonal carcinoma cells. *Mol Cell Biol*, 13 (9), 5383-92 (1993)

76. A. I. Packer, K. Manova and R. F. Bachvarova: A discrete LINE-1 transcript in mouse blastocysts. *Dev Biol*, 157 (1), 281-3 (1993)

77. A. A. Aravin, G. W. van der Heijden, J. Castaneda, V. V. Vagin, G. J. Hannon and A. Bortvin: Cytoplasmic compartmentalization of the fetal piRNA pathway in mice. *PLoS Genet*, 5 (12), e1000764 (2009)

78. S. F. Soper, G. W. van der Heijden, T. C. Hardiman, M. Goodheart, S. L. Martin, P. de Boer and A. Bortvin: Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. *Dev Cell*, 15 (2), 285-97 (2008)

79. J. A. Saxton and S. L. Martin: Recombination between subtypes creates a mosaic lineage of LINE-1 that is expressed and actively retrotransposing in the mouse genome. *J Mol Biol*, 280 (4), 611-22 (1998)

80. K. G. Becker, G. D. Swergold, K. Ozato and R. E. Thayer: Binding of the ubiquitous nuclear transcription factor YY1 to a cis regulatory sequence in the human LINE-1 transposable element. *Hum Mol Genet*, 2 (10), 1697-702 (1993)

81. S. L. Mathias and A. F. Scott: Promoter binding proteins of an active human L1 retrotransposon. *Biochem Biophys Res Commun*, 191 (2), 625-32 (1993)

82. J. N. Athanikar, R. M. Badge and J. V. Moran: A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic Acids Res*, 32 (13), 3846-55 (2004)

83. N. Yang, L. Zhang, Y. Zhang and H. H. Kazazian, Jr.: An important role for RUNX3 in human L1 transcription and retrotransposition. *Nucleic Acids Res*, 31 (16), 4929-40 (2003)

84. T. Tchenio, J. F. Casella and T. Heidmann: Members of the SRY family regulate the human LINE retrotransposons. *Nucleic Acids Res*, 28 (2), 411-5 (2000)

85. C. R. Harris, A. Dewan, A. Zupnick, R. Normart, A. Gabriel, C. Prives, A. J. Levine and J. Hoh: p53 responsive elements in human retrotransposons. *Oncogene*, 28 (44), 3857-65 (2009)

86. T. Kuwabara, J. Hsieh, A. Muotri, G. Yeo, M. Warashina, D. C. Lie, L. Moore, K. Nakashima, M. Asashima and F. H. Gage: Wnt-mediated activation of NeuroD1 and retro-elements during adult neurogenesis. *Nat Neurosci*, 12 (9), 1097-105 (2009)

87. T. B. Miranda and P. A. Jones: DNA methylation: the nuts and bolts of repression. *J Cell Physiol*, 213 (2), 384-90 (2007)

88. A. F. Scott, B. J. Schmeckpeper, M. Abdelrazik, C. T. Comey, B. O'Hara, J. P. Rossiter, T. Cooley, P. Heath, K. D. Smith and L. Margolet: Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. *Genomics*, 1 (2), 113-25 (1987)

89. E. D'Ambrosio, S. D. Waitzkin, F. R. Witney, A. Salemme and A. V. Furano: Structure of the highly repeated, long interspersed DNA family (LINE or L1Rn) of the rat. *Mol Cell Biol*, 6 (2), 411-24 (1986)

90. A. V. Furano, S. M. Robb and F. T. Robb: The structure of the regulatory region of the rat L1 (L1Rn, long interspersed repeated) DNA family of transposable elements. *Nucleic Acids Res*, 16 (19), 9215-31 (1988)

91. D. D. Loeb, R. W. Padgett, S. C. Hardies, W. R. Shehee, M. B. Comer, M. H. Edgell and C. A. Hutchison, 3rd: The sequence of a large L1Md element reveals a tandemly repeated 5' end and several features found in retrotransposons. *Mol Cell Biol*, 6 (1), 168-82 (1986)

92. W. R. Shehee, S. F. Chao, D. D. Loeb, M. B. Comer, C. A. Hutchison, 3rd and M. H. Edgell: Determination of a functional ancestral sequence and definition of the 5' end of A-type mouse L1 elements. *J Mol Biol*, 196 (4), 757-67 (1987)

93. R. W. Padgett, C. A. Hutchison, 3rd and M. H. Edgell: 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 (2), 739-49 (1988)

94. T. Tchenio, E. Segal-Bendirdjian and T. Heidmann: Generation of processed pseudogenes in murine cells. *EMBO J*, 12 (4), 1487-97 (1993)

95. D. M. Woodcock, C. B. Lawler, M. E. Linsenmeyer, J. P. Doherty and W. D. Warren: Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon. *J Biol Chem*, 272 (12), 7810-6 (1997)

96. G. Alves, A. Tatro and T. Fanning: Differential methylation of human LINE-1 retrotransposons in malignant cells. *Gene*, 176 (1-2), 39-44 (1996)

97. K. Hata and Y. Sakaki: Identification of critical CpG sites for repression of L1 transcription by DNA methylation. *Gene*, 189 (2), 227-34 (1997)

98. A. P. Bird and A. P. Wolffe: Methylation-induced repression--belts, braces, and chromatin. *Cell*, 99 (5), 451-4 (1999)

99. F. Yu, N. Zingler, G. Schumann and W. H. Stratling: Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. *Nucleic Acids Res*, 29 (21), 4493-501 (2001)

100. A. R. Muotri, M. C. Marchetto, N. G. Coufal, R. Oefner, G. Yeo, K. Nakashima and F. H. Gage: L1 retrotransposition in neurons is modulated by MeCP2. *Nature*, 468 (7322), 443-6 (2010)

101. H. D. Morgan, F. Santos, K. Green, W. Dean and W. Reik: Epigenetic reprogramming in mammals. *Hum Mol Genet*, 14 Spec No 1, R47-58 (2005)

102. V. Chapman, L. Forrester, J. Sanford, N. Hastie and J. Rossant: Cell lineage-specific undermethylation of mouse repetitive DNA. *Nature*, 307 (5948), 284-6 (1984)

103. J. P. Sanford, H. J. Clark, V. M. Chapman and J. Rossant: Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse. *Genes Dev*, 1 (10), 1039-46 (1987)

104. J. Sanford, L. Forrester, V. Chapman, A. Chandley and N. Hastie: Methylation patterns of repetitive DNA sequences in germ cells of Mus musculus. *Nucleic Acids Res*, 12 (6), 2823-36 (1984)

105. M. Monk, M. Boubelik and S. Lehnert: Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development*, 99 (3), 371-82 (1987)

106. S. K. Howlett and W. Reik: Methylation levels of maternal and paternal genomes during preimplantation development. *Development*, 113 (1), 119-27 (1991)

107. N. Lane, W. Dean, S. Erhardt, P. Hajkova, A. Surani, J. Walter and W. Reik: Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis*, 35 (2), 88-93 (2003)

108. P. Hajkova, S. Erhardt, N. Lane, T. Haaf, O. El-Maarri, W. Reik, J. Walter and M. A. Surani: Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev*, 117 (1-2), 15-23 (2002)

109. D. J. Lees-Murdock, M. De Felici and C. P. Walsh: Methylation dynamics of repetitive DNA elements in the mouse germ cell lineage. *Genomics*, 82 (2), 230-7 (2003)

110. S. Kuramochi-Miyagawa, T. Watanabe, K. Gotoh, Y. Totoki, A. Toyoda, M. Ikawa, N. Asada, K. Kojima, Y. Yamaguchi, T. W. Ijiri, K. Hata, E. Li, Y. Matsuda, T. Kimura, M. Okabe, Y. Sakaki, H. Sasaki and T. Nakano: DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev*, 22 (7), 908-17 (2008)

111. Y. Kato, M. Kaneda, K. Hata, K. Kumaki, M. Hisano, Y. Kohara, M. Okano, E. Li, M. Nozaki and H. Sasaki: Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum Mol Genet*, 16 (19), 2272-80 (2007)

112. M. Reuter, S. Chuma, T. Tanaka, T. Franz, A. Stark and R. S. Pillai: Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. *Nat Struct Mol Biol*, 16 (6), 639-46 (2009)

113. D. Bourc'his and T. H. Bestor: Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature*, 431 (7004), 96-9 (2004)

114. A. R. Bellve, J. C. Cavicchia, C. F. Millette, D. A. O'Brien, Y. M. Bhatnagar and M. Dym: Spermatogenic cells of the prepuberal mouse. Isolation and morphological characterization. *J Cell Biol*, 74 (1), 68-85 (1977)

115. B. R. Nebel, A. P. Amarose and E. M. Hacket: Calendar of gametogenic development in the prepuberal male mouse. *Science*, 134, 832-3 (1961)

116. H. Sasaki and Y. Matsui: Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat Rev Genet*, 9 (2), 129-40 (2008)

117. S. K. Ooi, A. H. O'Donnell and T. H. Bestor: Mammalian cytosine methylation at a glance. *J Cell Sci*, 122 (Pt 16), 2787-91 (2009)

118. T. Thomson and H. Lin: The biogenesis and function of PIWI proteins and piRNAs: progress and prospect. *Annu Rev Cell Dev Biol*, 25, 355-76 (2009)

119. A. A. Aravin, R. Sachidanandam, D. Bourc'his, C. Schaefer, D. Pezic, K. F. Toth, T. Bestor and G. J. Hannon: A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell*, 31 (6), 785-99 (2008)

120. A. A. Aravin, R. Sachidanandam, A. Girard, K. Fejes-Toth and G. J. Hannon: Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science*, 316 (5825), 744-7 (2007)

121. M. A. Carmell, A. Girard, H. J. van de Kant, D. Bourc'his, T. H. Bestor, D. G. de Rooij and G. J. Hannon: MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev Cell*, 12 (4), 503-14 (2007)

122. M. Shoji, T. Tanaka, M. Hosokawa, M. Reuter, A. Stark, Y. Kato, G. Kondoh, K. Okawa, T. Chujo, T. Suzuki, K. Hata, S. L. Martin, T. Noce, S. Kuramochi-Miyagawa, T. Nakano, H. Sasaki, R. S. Pillai, N. Nakatsuji and S. Chuma: The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. *Dev Cell*, 17 (6), 775-87 (2009)

123. L. Ma, G. M. Buchold, M. P. Greenbaum, A. Roy, K. H. Burns, H. Zhu, D. Y. Han, R. A. Harris, C. Coarfa, P. H.

Gunaratne, W. Yan and M. M. Matzuk: GASZ is essential for male meiosis and suppression of retrotransposon expression in the male germline. *PLoS Genet*, 5 (9), e1000635 (2009)

124. T. Yoshimura, S. Toyoda, S. Kuramochi-Miyagawa, T. Miyazaki, S. Miyazaki, F. Tashiro, E. Yamato, T. Nakano and J. Miyazaki: Gtsf1/Cue110, a gene encoding a protein with two copies of a CHHC Zn-finger motif, is involved in spermatogenesis and retrotransposon suppression in murine testes. *Dev Biol*, 335 (1), 216-27 (2009)

125. K. Zheng, J. Xiol, M. Reuter, S. Eckardt, N. A. Leu, K. J. McLaughlin, A. Stark, R. Sachidanandam, R. S. Pillai and P. J. Wang: Mouse MOV10L1 associates with Piwi proteins and is an essential component of the Piwi-interacting RNA (piRNA) pathway. *Proc Natl Acad Sci U S A*, 107 (26), 11841-6 (2010)

126. R. J. Frost, F. K. Hamra, J. A. Richardson, X. Qi, R. Bassel-Duby and E. N. Olson: MOV10L1 is necessary for protection of spermatocytes against retrotransposons by Piwi-interacting RNAs. *Proc Natl Acad Sci U S A*, 107 (26), 11847-52 (2010)

127. Y. L. Chiu and W. C. Greene: The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annu Rev Immunol*, 26, 317-53 (2008)

128. H. P. Bogerd, H. L. Wiegand, A. E. Hulme, J. L. Garcia-Perez, K. S. O'Shea, J. V. Moran and B. R. Cullen: Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc Natl Acad Sci U S A*, 103 (23), 8780-5 (2006)

129. H. Chen, C. E. Lilley, Q. Yu, D. V. Lee, J. Chou, I. Narvaiza, N. R. Landau and M. D. Weitzman: APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr Biol*, 16 (5), 480-5 (2006)

130. M. Kinomoto, T. Kanno, M. Shimura, Y. Ishizaka, A. Kojima, T. Kurata, T. Sata and K. Tokunaga: All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. *Nucleic Acids Res*, 35 (9), 2955-64 (2007)

131. H. Muckenfuss, M. Hamdorf, U. Held, M. Perkovic, J. Lower, K. Cichutek, E. Flory, G. G. Schumann and C. Munk: APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J Biol Chem*, 281 (31), 22161-72 (2006)

132. M. D. Stenglein and R. S. Harris: APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. *J Biol Chem*, 281 (25), 16837-41 (2006)

133. A. M. Niewiadomska, C. Tian, L. Tan, T. Wang, P. T. Sarkis and X. F. Yu: Differential inhibition of long interspersed element 1 by APOBEC3 does not correlate with high-molecular-mass-complex formation or P-body association. *J Virol*, 81 (17), 9577-83 (2007)

134. C. Esnault, O. Heidmann, F. Delebecque, M. Dewannieux, D. Ribet, A. J. Hance, T. Heidmann and O. Schwartz: APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. *Nature*, 433 (7024), 430-3 (2005)

135. D. A. MacDuff, Z. L. Demorest and R. S. Harris: AID can restrict L1 retrotransposition suggesting a dual role in innate and adaptive immunity. *Nucleic Acids Res*, 37 (6), 1854-67 (2009)

136. T. Ikeda, K. H. Abd El Galil, K. Tokunaga, K. Maeda, T. Sata, N. Sakaguchi, T. Heidmann and A. Koito: Intrinsic restriction activity by apolipoprotein B mRNA editing enzyme APOBEC1 against the mobility of autonomous retrotransposons. *Nucleic Acids Res* (2011)

137. G. G. Schumann: APOBEC3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition. *Biochem Soc Trans*, 35 (Pt 3), 637-42 (2007)

138. D. B. Stetson, J. S. Ko, T. Heidmann and R. Medzhitov: Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell*, 134 (4), 587-98 (2008)

139. C. Popp, W. Dean, S. Feng, S. J. Cokus, S. Andrews, M. Pellegrini, S. E. Jacobsen and W. Reik: Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature*, 463 (7284), 1101-5 (2010)

140. M. Ghildiyal and P. D. Zamore: Small silencing RNAs: an expanding universe. *Nat Rev Genet*, 10 (2), 94-108 (2009)

141. N. Yang and H. H. Kazazian, Jr.: L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat Struct Mol Biol*, 13 (9), 763-71 (2006)

142. H. S. Soifer and J. J. Rossi: Small interfering RNAs to the rescue: blocking L1 retrotransposition. *Nat Struct Mol Biol*, 13 (9), 758-9 (2006)

143. C. Kanellopoulou, S. A. Muljo, A. L. Kung, S. Ganesan, R. Drapkin, T. Jenuwein, D. M. Livingston and K. Rajewsky: Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev*, 19 (4), 489-501 (2005)

144. K. Hayashi, S. M. Chuva de Sousa Lopes, M. Kaneda, F. Tang, P. Hajkova, K. Lao, D. O'Carroll, P. P. Das, A. Tarakhovsky, E. A. Miska and M. A. Surani: MicroRNA biogenesis is required for mouse primordial germ cell development and spermatogenesis. *PLoS One*, 3 (3), e1738 (2008)

145. S. L. Martin, D. Branciforte, D. Keller and D. L. Bain: Trimeric structure for an essential protein in L1

retrotransposition. *Proc Natl Acad Sci U S A*, 100 (24), 13815-20 (2003)

146. M. Xu, Y. You, P. Hunsicker, T. Hori, C. Small, M. D. Griswold and N. B. Hecht: Mice deficient for a small cluster of Piwi-interacting RNAs implicate Piwi-interacting RNAs in transposon control. *Biol Reprod*, 79 (1), 51-7 (2008)

147. H. Hohjoh and M. F. Singer: Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J*, 15 (3), 630-9 (1996)

148. H. Hohjoh and M. F. Singer: Ribonuclease and high salt sensitivity of the ribonucleoprotein complex formed by the human LINE-1 retrotransposon. *J Mol Biol*, 271 (1), 7-12 (1997)

149. J. L. Goodier, L. Zhang, M. R. Vetter and H. H. Kazazian, Jr.: LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. *Mol Cell Biol*, 27 (18), 6469-83 (2007)

150. J. L. Goodier, P. K. Mandal, L. Zhang and H. H. Kazazian, Jr.: Discrete subcellular partitioning of human retrotransposon RNAs despite a common mechanism of genome insertion. *Hum Mol Genet*, 19 (9), 1712-25 (2010)

151. A. J. Doucet, A. E. Hulme, E. Sahinovic, D. A. Kulpa, J. B. Moldovan, H. C. Kopera, J. N. Athanikar, M. Hasnaoui, A. Bucheton, J. V. Moran and N. Gilbert: Characterization of LINE-1 ribonucleoprotein particles. *PLoS Genet*, 6 (10), e1001150 (2010)

152. J. R. Buchan and R. Parker: Eukaryotic stress granules: the ins and outs of translation. *Mol Cell*, 36 (6), 932-41 (2009)

153. E. M. Ostertag, R. J. DeBerardinis, J. L. Goodier, Y. Zhang, N. Yang, G. L. Gerton and H. H. Kazazian, Jr.: A mouse model of human L1 retrotransposition. *Nat Genet*, 32 (4), 655-60 (2002)

154. E. T. Prak, A. W. Dodson, E. A. Farkash and H. H. Kazazian, Jr.: Tracking an embryonic L1 retrotransposition event. *Proc Natl Acad Sci U S A*, 100 (4), 1832-7 (2003)

155. D. V. Babushok, E. M. Ostertag, C. E. Courtney, J. M. Choi and H. H. Kazazian, Jr.: L1 integration in a transgenic mouse model. *Genome Res*, 16 (2), 240-50 (2006)

156. W. An, J. S. Han, C. M. Schrum, A. Maitra, F. Koentgen and J. D. Boeke: Conditional activation of a single-copy L1 transgene in mice by Cre. *Genesis*, 46 (7), 373-83 (2008)

157. W. An, J. S. Han, S. J. Wheelan, E. S. Davis, C. E. Coombes, P. Ye, C. Triplett and J. D. Boeke: Active retrotransposition by a synthetic L1 element in mice. *Proc Natl Acad Sci U S A*, 103 (49), 18662-7 (2006)

158. H. Kano, I. Godoy, C. Courtney, M. R. Vetter, G. L. Gerton, E. M. Ostertag and H. H. Kazazian, Jr.: L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev*, 23 (11), 1303-12 (2009)

159. A. R. Muotri, C. Zhao, M. C. Marchetto and F. H. Gage: Environmental influence on L1 retrotransposons in the adult hippocampus. *Hippocampus*, 19 (10), 1002-7 (2009)

160. J. M. Rosser and W. An: Repeat-induced gene silencing of L1 transgenes is correlated with differential promoter methylation. *Gene*, 456 (1-2), 15-23 (2010)

161. M. L. Kimberland, V. Divoky, J. Prchal, U. Schwahn, W. Berger and H. H. Kazazian, Jr.: Full-length human L1 insertions retain the capacity for high frequency retrotransposition in cultured cells. *Hum Mol Genet*, 8 (8), 1557-60 (1999)

162. B. Brouha, C. Meischl, E. Ostertag, M. de Boer, Y. Zhang, H. Neijens, D. Roos and H. H. Kazazian, Jr.: Evidence consistent with human L1 retrotransposition in maternal meiosis I. *Am J Hum Genet*, 71 (2), 327-36 (2002)

163. D. C. Lie, H. Song, S. A. Colamarino, G. L. Ming and F. H. Gage: Neurogenesis in the adult brain: new strategies for central nervous system diseases. *Annu Rev Pharmacol Toxicol*, 44, 399-421 (2004)

164. J. A. van den Hurk, I. C. Meij, M. C. Seleme, H. Kano, K. Nikopoulos, L. H. Hoefsloot, E. A. Sistermans, I. J. de Wijs, A. Mukhopadhyay, A. S. Plomp, P. T. de Jong, H. H. Kazazian and F. P. Cremers: L1 retrotransposition can occur early in human embryonic development. *Hum Mol Genet*, 16 (13), 1587-92 (2007)

165. K. Hayashi, S. M. de Sousa Lopes and M. A. Surani: Germ cell specification in mice. *Science*, 316 (5823), 394-6 (2007)

166. B. Plusa, A. Piliszek, S. Frankenberg, J. Artus and A. K. Hadjantonakis: Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. *Development*, 135 (18), 3081-91 (2008)

167. J. L. Garcia-Perez, M. Morell, J. O. Scheys, D. A. Kulpa, S. Morell, C. C. Carter, G. D. Hammer, K. L. Collins, K. S. O'Shea, P. Menendez and J. V. Moran: Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature*, 466 (7307), 769-73 (2010)

168. N. G. Coufal, J. L. Garcia-Perez, G. E. Peng, G. W. Yeo, Y. Mu, M. T. Lovci, M. Morell, K. S. O'Shea, J. V. Moran and F. H. Gage: L1 retrotransposition in human neural progenitor cells. *Nature*, 460 (7259), 1127-31 (2009)

169. N. C. Lau, A. G. Seto, J. Kim, S. Kuramochi-Miyagawa, T. Nakano, D. P. Bartel and R. E. Kingston:

Characterization of the piRNA complex from rat testes. *Science*, 313 (5785), 363-7 (2006)

170. R. D. Palmiter and R. L. Brinster: Germ-line transformation of mice. *Annu Rev Genet*, 20, 465-99 (1986)

171. D. I. Martin and E. Whitelaw: The vagaries of variegating transgenes. *Bioessays*, 18 (11), 919-23 (1996)

172. F. F. Assaad, K. L. Tucker and E. R. Signer: Epigenetic repeat-induced gene silencing (RIGS) in Arabidopsis. *Plant Mol Biol*, 22 (6), 1067-85 (1993)

173. D. Garrick, S. Fiering, D. I. Martin and E. Whitelaw: Repeat-induced gene silencing in mammals. *Nat Genet*, 18 (1), 56-9 (1998)

Abbreviations: 3'UTR: 3' untranslated region, 5'UTR: 5' untranslated region, AID: activation-induced deaminase, APOBEC3: apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3, DNMT3L: DNA methyltransferase 3-like, DSB: double-strand DNA break, EB: embryoid body, ECC: embryonal carcinoma cell, EGC: embryonic germ cell, ESC: embryonic stem cell, GCT: germ cell tumor, GFP: green fluorescent protein, HCN: hippocampus neural (stem cell), L1: long interspersed element type 1, $L1_{CHM}$: an L1 insertion in a family with choroideremia, $L1_{RP}$: L1 isolated from a patient with X-linked retinitis pigmentosa, LPR: length polymorphism region, LRE3: L1 retrotransposable element-3 isolated from a patient with chronic granulomatous disease, LTR: long terminal repeat, MBD: methyl-CpG-binding domain, MeCP2: methyl-CpG-binding protein 2, miRNA: microRNA, NPC: neural progenitor cell, NSC: neural stem cell, ORF1p: open reading frame 1 protein, ORF2p: open reading frame 2 protein, PGC: primordial germ cell, piRNA: PIWI-interacting RNA, RCL: rat chloroleukemia, RIGS: repeat-induced gene silencing, RT-PCR: reverse transcriptase-polymerase chain reaction, RUNX3: runt-domain transcription factor 3, siRNA: small-interfering RNA, SOX: SRY-related HMG box containing, ssDNA: single-stranded DNA, Ta: transcriptionally active, TCF/LEF: T-cell factor/lymphoid enhancer factor, TREX1: 3' repair exonuclease 1, YY1: Yin Yang-1.

Key Words: 5'UTR, Blastocyst, DNA methylation, Embryonal carcinoma cell, Embryonic stem cell, Gene expression, Gene regulation, Germ cell tumor, L1, LINE-1, Long interspersed element, Morula, Mouse model, Neurogenesis, Non-LTR retrotransposon, ORF1, ORF2, piRNA, Preimplantation, Postimplantation, Primordial germ cell, siRNA, Spermatogenesis, Transcription factor, Transgene, Tumor, Review

Send correspondence to: Wenfeng An, School of Molecular Biosciences, Washington State University, PO Box 647520, Pullman, WA 99164-7520, USA, Tel: 509-335-8296, Fax: 509-335-4159, E-mail: wenfeng_an@wsu.edu.

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