

Synchronous Expression of LINE-1 RNA and Protein in Mouse Embryonal Carcinoma Cells

SANDRA L. MARTIN* AND DAN BRANCIFORTE

Department of Cellular and Structural Biology, Box B111, University of Colorado School of Medicine, 4200 E. Ninth Avenue, Denver, Colorado 80262

Received 5 April 1993/Returned for modification 19 May 1993/Accepted 1 June 1993

L1, or LINE-1, is a repetitive DNA family found in all mammalian genomes that have been examined. At least a few individual members of the L1 family are functional transposable elements. Expression of these active elements leads to new insertions of L1 into the genomic DNA by the process of retrotransposition. We have detected coexpression of full-length, sense-strand L1 RNA transcripts and L1-encoded protein in mouse embryonal carcinoma cell lines. Both of these L1 expression products are candidates for intermediates in the retrotransposition process. L1 protein is found in what appear to be cytoplasmic aggregates and is not localized to any known cytoplasmic organelles. The six embryonal carcinoma cell lines tested were chosen to represent commitment to different developmental pathways in early mouse embryogenesis. The only two cell lines that express L1 are unique among the six in that they have a strong predilection to differentiate into extraembryonic endoderm. This observation is consistent with L1 expression and transposition in primordial germ cells of the mouse. An important implication of these studies is that L1 expression may provide a new marker for use in determining the origin of primordial germ cells during mouse embryogenesis.

L1, or LINE-1, has achieved its status as a middle repetitive DNA family in mammalian genomes by the process of duplicative transposition. For LINES, duplicative transposition appears to proceed via an RNA intermediate and reverse transcription; hence, it is termed retrotransposition (see reference 17 and references therein). Little is known about the process of retrotransposition in any of the LINE-like elements, although they are widely distributed throughout eucaryotic phyla and appear to have an ancient evolutionary origin (6, 49). Unlike retroviruses and long terminal repeat-containing retrotransposons such as *Ty*, which also use RNA intermediates and reverse transcription in their replication cycle, LINE-like elements have no long terminal repeat structures. For this reason, only the general features of the intermediates involved in L1 retrotransposition may be accurately modeled on these more familiar systems.

Retrotransposition of L1 must begin with a full-length sense-strand transcript that moves to the cytoplasm for translation. Proteins are translated from both open reading frames (ORFs), and these proteins are hypothesized to preferentially associate with their own mRNA (7, 27, 29), forming a ribonucleoprotein (RNP) complex. The RNA is then converted into cDNA, which is integrated into the genome, most often as a truncated product of the original RNA. Among the details missing from this model are the following: the specifics of the interaction of L1-encoded proteins with L1 "genomic" RNA, where on the RNA template and how cDNA synthesis is primed, where in the cell cDNA synthesis occurs, how the RNA or cDNA gets back into the nucleus, and how integration occurs (28). There are essentially two models that address the site of reverse transcription: *in situ* at breaks in chromosomal DNA (see reference 17 and references therein) or in cytoplasmic RNPs, as with long terminal repeat-containing retrotrans-

posons and retroviruses (1, 47). Recent evidence in support of the *in situ* model has been provided by using the ORF from the LINE-like element of *Bombyx mori*, R2Bm (25). However, previous experiments provide evidence for reverse transcriptase in a virus-like particle containing human L1 RNA (4). These results are not necessarily contradictory; in eucaryotic cells, either the cDNA or the RNA template and reverse transcriptase must return to the nucleus following translation in the cytoplasm before completion of the retrotransposition cycle by integration into chromosomal DNA. Therefore, all of the necessary components may be packaged into a cytoplasmic particle which is an intermediate in retrotransposition, regardless of whether reverse transcription occurs in that cytoplasmic particle or following translocation to the nucleus.

Studies of the intermediates involved in the process of L1 transposition in mammalian cells have been hindered by two aspects of the intrinsic biology of the system: the high copy number of L1 elements in the genome and the low expression of potential functional intermediates in most cell types. The copy number of L1 elements in the mouse genome, for example, approaches 10^5 , although most (>90%) of the copies are truncated versions of the full-length >6.5-kb element. Since a large proportion of the full-length L1 element is apparently protein encoding (Fig. 1) and those proteins are likely to be important for the retrotransposition process, all of the truncated L1 elements are presumed to be defective for subsequent transposition. In addition to the obvious loss of protein-coding function because of truncation, most of the elements have frameshift and/or termination codons in the two ORFs encoded by L1. Thus, L1 is best thought of as a large multigene family which is composed of primarily truncated pseudogenes and a small number of functional, active, transposable elements (13, 14, 17).

Full-length copies of L1 (~6.5 kb in length) with intact ORFs have been isolated and characterized from both mouse and human cells (5, 24, 41). For mouse L1, comparison of several such elements led to the prediction of a consensus functional sequence which has been used to reconstruct a

* Corresponding author. Electronic mail address: martin_s%maui@vaxf.colorado.edu.

putative parental (active) element. This ancestral sequence was used to provide preliminary evidence in support of the long-held hypothesis that L1 transposes via an RNA intermediate (8); similar experiments have also been reported for rat L1 (40). An active human element was identified and isolated by working backwards from an L1 insertional mutagenesis event in a patient with hemophilia. In this case, the full-length progenitor of a truncated L1 that recently landed in the factor VIII gene was found by taking advantage of a sequence polymorphism that distinguished the particular copy of L1 causing the hemophilia from the high background of L1 copies in the human genome (5). A sequence taken from this element was used to provide direct evidence that ORF 2 in human L1 encodes a reverse transcriptase (31).

Largely on the basis of arguments regarding the pattern of sequence evolution among individual elements belonging to the L1 family, it is believed that L1 retrotransposition must occur in germ cells or early embryos (44). In the case of the human patient, the L1 transposition event that gave rise to hemophilia must have occurred in the germ line or early embryo leading to the affected individual, because neither parent carried a factor VIII gene with an L1 insertion (20). The difficulty of studying biochemical events in mammalian germ cells and early embryos has prompted the use of embryonal carcinoma (EC) cells as a model system for the study of L1 RNA and protein expression and, hence, the necessary intermediates in L1 retrotransposition.

A screen of several human cell lines, representing a variety of cell types, revealed that full-length L1 transcripts were detected only in an EC line, NTera2D1. Furthermore, the full-length L1 transcripts disappeared when the cells were differentiated with retinoic acid (44). These transcripts were characterized further and found to contain a subset of the genomic L1 sequences that belong to the same subfamily as the active L1 that caused hemophilia (5, 43). The ORF 1 protein product of human L1 is also detected in these cells (23), as is a complex that may include full-length L1 RNA, ORF 1 protein, and reverse transcriptase (4). Full-length, sense-strand transcripts of mouse L1 have been found in an EC line, F9, in an RNP particle (27). In an independent study, L1 transcripts isolated from F9 were found to contain only the youngest (active) subfamily of L1 sequences that are present in the mouse genome (39). Taken together, these results are consistent with the hypothesis that the full-length L1 transcripts and protein products seen in EC cells represent intermediates in L1 transposition. Furthermore, the unique expression patterns seen in EC cells may provide important clues regarding L1 expression during development.

In order to explore these ideas further, in this study we examined six different mouse EC lines for the expression of L1 RNA and ORF 1 protein. These cell lines were chosen because they possess and express different developmental potentials. Although all cell types express RNA with L1-homologous sequences as detected by Northern (RNA) blot analysis, only two of the six lines tested were found to express a full-length, sense-strand L1 RNA. Significantly, the same two cell lines were also the only ones found to express L1 ORF 1 protein. In both of these cell lines, L1 ORF 1 protein is located primarily in the cytoplasm. Immunolabelling of cells and examination by both light and electron microscopy reveal an unusual, punctate staining pattern that distinguishes L1 from cytoplasmic organelles including the endoplasmic reticulum, Golgi apparatus, lysosomes, and the mitochondria. This pattern is consistent with our previous finding that ORF 1 protein and full-length L1

RNA form a cytoplasmic RNP and the expectation that this RNP complex contains several ORF 1 protein molecules for each RNA (27). Finally, the differential expression of L1 in these different mouse EC cell lines is discussed in light of the implications of these results for L1 expression and transposition during mouse development.

MATERIALS AND METHODS

Cell lines. F9, C44, P19, ECa247, 1003, and PCC4 mouse cell lines were obtained from E. Barry Pierce, Department of Pathology, University of Colorado School of Medicine. All lines except C44 were grown as monolayers in modified Eagle's medium (GIBCO) supplemented with 10% fetal bovine serum (HyClone) at 37°C in 5% CO₂; F9 and ECa247 were maintained on gelatin-coated (0.1%) tissue culture dishes. C44 grows as embryoid bodies and was maintained in the ascites of strain 129 mice as previously described (34).

Extraction and analysis of RNA. Cells were removed from the culture dishes by scraping with a rubber policeman and washed once with ice-cold PBS (10 mM phosphate, [pH 7.4]–0.15 M NaCl). RNA was prepared from either whole cells or cytoplasmic extracts (27) by acid-phenol extraction (2). Agarose formaldehyde gels (1%) (38) were run with 5 µg of RNA. Following electrophoresis, the RNA was transferred to nitrocellulose. Northern blots were hybridized either to random-primed ³²P-labelled (Boehringer Mannheim) human β-actin fragment (12) or to in vitro transcripts corresponding to various regions of L1 (Fig. 1) made with SP6 RNA polymerase (Boehringer Mannheim). Blots were hybridized at 42°C by using standard conditions (48) with either 40% (actin) or 50% formamide (L1) and then washed in 2× SSC (2× SSC is 0.3 M NaCl–30 mM sodium citrate)–0.1% sodium dodecyl sulfate (SDS) three times for 5 min (each) at room temperature and two times for 15 min (each) at either 42°C in 0.8× SSC–0.1% SDS (actin) or 55°C in 0.1× SSC–0.1% SDS (L1).

Affinity-purified antibody. Two independent affinity-purified antibodies against either the entire ORF 1 or its carboxy-terminal two-thirds were prepared (Fig. 1). FP1 contains nearly all of ORF 1 (note that ORF 1 contains 22 codons 5' of the first AUG). An 1,845-bp *NheI*-*NcoI* restriction fragment from L1Md-A2 (24) was cloned into a modified pET-3 vector (45) for expression in *Escherichia coli*. The mobility of this fusion protein by SDS-polyacrylamide gel electrophoresis (PAGE) is in good agreement with its predicted molecular weight of 44,623. The protein contains 24 amino acids from the vector fused to 364 amino acids from L1 ORF 1, beginning six amino acids upstream of the first methionine codon in the sequence of L1Md-A2. The construct expressing FP2 contains a 2,232-bp *Bgl*III fragment of L1Md-A2 (24) cloned into pET-3a (45) and leads to production of a truncated version of ORF 1. This fusion protein contains 12 amino acids from the vector and the last 237 amino acids of ORF 1 from L1; it has a predicted molecular weight of 29,800. Both of these fusion proteins are expressed at high levels in induced *E. coli* and both are purified from the insoluble, inclusion body fraction by a detergent wash procedure (32). For affinity purification of antibodies, the fusion protein was further purified after solubilization in buffer containing 25 mM Tris-HCl (pH 8.0), 10 mM EDTA, 10 mM dithiothreitol, and 8 M urea by chromatography on carboxymethyl Sepharose (Pharmacia). Proteins were eluted with a 60 to 270 mM NaCl gradient in the same buffer. For immunization, the proteins were further purified by preparative SDS-PAGE (22) and electroelution (Elutrap; Schle-

cher & Schuell). Rabbits were immunized with 200 μ g of protein in Freund's complete adjuvant and boosted with 100 μ g of protein in Freund's incomplete adjuvant (15). The immunoglobulin G (IgG) fraction was prepared from preimmune serum from the same rabbits by using Affi-Gel Blue chromatography (Bio-Rad). For L1 antibodies, crude serum was purified on antigen affinity columns containing the appropriate (FP1 or FP2) fusion proteins coupled to agarose (AminoLink; Pierce). Antibodies were eluted from the antigen columns with 0.1 M glycine, pH 2.8.

Extraction and analysis of protein. Cells were harvested by scraping, and cytoplasmic extracts were prepared as described previously (27). Protein (20 μ g) (bicinchoninic acid assay; Pierce) was fractionated by SDS-PAGE in 10% acrylamide gels (22) and then electrophoretically transferred to nitrocellulose (Novablot; Pharmacia LKB). L1 ORF 1 proteins were detected by using the affinity-purified FP1 and FP2 antibodies. Blots were incubated in TBST (10 mM Tris-HCl [pH 8.0], 150 mM NaCl, 0.05% Tween 20) containing 2% bovine serum albumin for 15 min at room temperature, and then antibody (or preimmune IgG) was added to a final concentration of 1.02 μ g/ml. The antibody was allowed to bind for 2 h at room temperature, and then blots were washed three times for 5 min (each) in an excess of TBST and incubated 1 h in goat anti-rabbit IgG conjugated to alkaline phosphatase. The conditions used for secondary antibody binding and detection were as described by the manufacturer (Promega).

Immunofluorescence microscopy. F9 cells were grown in eight-chamber Permanox slides (Nunc) by seeding them onto a thin coating of Matrigel (Collaborative Research). The Matrigel was diluted 1/10 in serum-free medium, applied to the wells, and then allowed to gel for 10 min at room temperature prior to adding the cells. At 70 to 100% confluence, the cells were washed in PBS and then fixed in 4% paraformaldehyde-0.1 M phosphate, pH 7.4, for 20 min. Cells were permeabilized in ice-cold PBS containing 0.1% Triton X-100-10 mM ethylene glycol-bis(β -aminoethyl ether)-*N,N,N',N'*-tetraacetic acid (EGTA) for 4 min and then washed three times for 10 min (each) in PBS. L1 antibodies were added at a concentration of 10 μ g/ml in PBS containing 2% normal goat serum and incubated with the cells 1 h at room temperature. After being washed in PBS, the secondary antibody, Texas Red-conjugated goat anti-rabbit IgG (Jackson ImmunoResearch Laboratories) was added at a concentration of 10 μ g/ml in PBS with 2% normal goat serum. Following a 1-h incubation at room temperature, the cells were washed and mounted in Mowiol 4.88 (Calbiochem) with 2% *N*-propyl gallate (Sigma) added as an antileaching agent. Cells were observed and photographed with an Axiophot (Zeiss) microscope by using T-Max 400 (Kodak) film. Identical procedures were used to stain proteins localized to the Golgi apparatus (TGN-38 [26]), the endoplasmic reticulum (protein disulfide isomerase [9, 46]), or the lysosome (LGP-120 [21]). These three polyclonal antisera were used at a 1/100 dilution and were the generous gift of K. Howell.

C44 embryoid bodies were isolated from mouse ascites fluid by sequential centrifugation at $20 \times g$ for 5 min; the cell pellet was washed until the supernatant was clear. Embryoid bodies were resuspended in PBS and transferred to 1.5-ml microcentrifuge tubes for subsequent processing; all solution changes were accomplished by a 5-min centrifugation at $20 \times g$ and removal of the supernatant by aspiration. The embryoid bodies were fixed in solution as described for F9 cells above, by adding an equal volume of 8% paraformal-

dehyde in 0.2 M phosphate, pH 7.4, to the cell suspension. Embryoid bodies were washed three times for 5 min (each) in PBS and once for 10 min and then incubated with antibody at 10 μ g/ml in PBS with 1.5% normal goat serum and 0.08% saponin for 1 h at room temperature. The cells were washed with PBS and then incubated with secondary, Texas Red-conjugated goat anti-rabbit IgG, washed, mounted, and photographed on a Zeiss Axioskop equipped for difference interference microscopy.

Immunogold labelling of ultrathin cryosections and electron microscopy. F9 cells were removed from dishes with trypsin-EDTA (Sigma; 37°C, 1 min) and transferred into 30 ml of fresh medium in T-50 flasks and then incubated in suspension for 2 h to allow recovery from trypsinization. The cells were pelleted, and the medium was removed. Procedures described by Griffiths et al. (11) were modified as follows. The cell pellet was resuspended in 1.0 ml of PHEM buffer (60 mM piperazine-*N,N'*-bis(2-ethanesulfonic acid [PIPES], 25 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid [HEPES], pH 7.0, 10 mM EGTA, 2 mM MgCl₂), transferred to a 1.5-ml microcentrifuge tube, and pelleted again. Without resuspending the cells, the supernatant was removed and replaced with 8% paraformaldehyde-5% sucrose in PHEM buffer. Cells were fixed for 24 h at 4°C, and then the pellet was removed, infiltrated with 2.1 M sucrose, transferred to a copper specimen stub, and frozen rapidly in liquid nitrogen. Thin cryosections (90 to 100 nm) were cut on a Reichert-Jung Ultramicrotome equipped with an FC4-D cryostage. Sections were transferred to Formvar- and carbon-coated copper grids.

For immunolabelling, all procedures were carried out at room temperature. Grids were treated for 10 min with 5% calf serum to block nonspecific antibody binding sites and then reacted with primary antibody at 10 μ g/ml for 30 min. The grids were washed in PBS and then reacted for an additional 30 min with 10-nm colloidal gold-protein A (100 ng/ml; E-Y Laboratories). Grids were stained for 20 min with 0.2% uranyl acetate in 2% methylcellulose, picked up in wire loops, and allowed to air dry. Specimens were viewed and photographed with a Philips CM-10 transmission electron microscope.

RESULTS

The structure of a typical L1 element from the mouse genome is shown in Fig. 1. The locations of the hybridization probes and the fusion proteins that were used for antibody production are indicated.

L1 RNA expression in mouse EC cells. In RNA extracted from the cytoplasm of six mouse EC cell lines and screened for L1 expression by Northern hybridization, only F9 and C44 clearly express a full-length, sense-strand RNA. A hybridization signal that corresponds to an RNA of ~7.5 kb is observed with both the 5' (8.1) and 3' (5.1) sense-strand-specific probes (Fig. 1) for L1 with RNA isolated from F9 and C44 cells (Fig. 2). A species of this size is not detected with either the 8.2 or the 5.2 (antisense-specific) probes and is not detected in RNA prepared from any of the other four cell lines. This signal is readily detected above a background smear that is visible in all of the RNA preparations with the same two sense-strand probes as well as in their antisense counterparts (Fig. 2). Although cytoplasmic extracts are shown for all cell types, total cell RNA extracts were prepared and examined for L1 RNA expression as well. As shown for C44 cells in Fig. 2, all of the full-length, sense-strand L1 RNA appears to be in the cytoplasm. Some of the

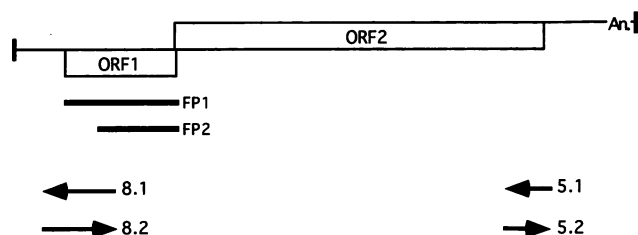


FIG. 1. Structure of mouse L1 and locations of probes. Typical full-length genomic copies of L1 in the mouse genome contain two ORFs (ORF 1 and ORF 2). The 3' end of the elements are characterized by an A-rich tail (A_n), and both ends are bounded by duplications of several bases of target site DNA (vertical bar). Both of these features are presumed to be the result of retrotransposition. Construction of fusion proteins 1 and 2 (FP1 and FP2) is described in detail in Materials and Methods. The single-stranded probes 8.1 and 8.2 and 5.1 and 5.2 are transcribed from clones of an 800-bp *Bam*HI fragment from L1Md-A9 (41) or a 500-bp *Bam*HI fragment of L1Md-A2 (24).

larger, hybridizing material at the top of the gels is due to a small amount of contamination of the RNA preparations with genomic DNA. However, the majority of the smear that is observed with all L1 probes is most likely due to fortuitous transcription originating from non-L1 promoters in the vicinity of individual, defective, variously truncated elements. In contrast, transcription giving rise to the discrete, full-length, sense-strand-specific signal on the Northern blot is a candidate for bona fide L1 expression and thus the first intermediate in L1 retrotransposition.

L1 ORF 1 protein expression in mouse EC cells. Protein extracts from the same six cell lines were examined for L1 ORF 1 protein expression by Western blot (immunoblot) analysis. Four specific protein bands with an apparent molecular weight of $\sim 41,500$ to $44,000$ are detected with anti-FP1 on Western blots of extracts prepared from F9 cells. Two of these proteins are also readily detected in the C44 extract, and the other two are present, but in lesser

amount. None of them, however, are seen in extracts from P19, ECa247, 1003, or PCC4 (Fig. 3). Identical results were obtained with an antibody to the truncated form of ORF 1, FP2, that was independently prepared (data not shown). None of the protein extracts showed any reactivity with the preimmune serum (Fig. 3). A larger band, corresponding to a molecular weight of $\sim 53,000$, is variably observed (see PCC4, P19, and F9 in Fig. 3). This protein varies in intensity from experiment to experiment and is blocked by the addition of 2% milk to the antibody-binding solution. Therefore, the 53,000-apparent-molecular-weight protein is not considered to be a bona fide L1 ORF 1 product.

Translation of ORF 1 from the published nucleotide sequence of L1Md-A2 (24) predicts a protein with a molecular weight of 41,226 if translation begins with the first methionine in ORF 1. Since the smallest of the quartet of polypeptides reacting with the anti-ORF 1 antibody in F9 and C44 cells has an apparent molecular weight of $\sim 41,500$, ORF 1 apparently is translated from the first methionine through the TAA termination codon and not further processed by proteolytic cleavage into smaller polypeptides. The multiple forms of ORF 1 protein are always seen in F9 and C44 extracts, but the relative intensity of the four bands is variable in different extracts and with storage of the same extract. Although experiments have not been done specifically to address this point, it appears that at least one of the lower bands intensifies at the expense of the upper bands. Significantly, as with the full-length, sense-strand RNA, only F9 and C44 cell lines show evidence of expression of ORF 1 protein.

L1 RNA and ORF 1 protein cosediment in sucrose gradients. Previous studies have shown that full-length, sense-strand L1 RNA is found in a RNP particle in the cytoplasm of F9 cells and that L1 ORF 1 protein cosediments with the L1 RNP in sucrose gradients (27). Similar fractionation experiments have been done with cytoplasmic extracts from C44 cells, shown above to be positive on Northern (Fig. 2) and Western (Fig. 3) blots for L1 RNA and ORF 1 expression, as well as on extracts from ECa247 cells, negative for

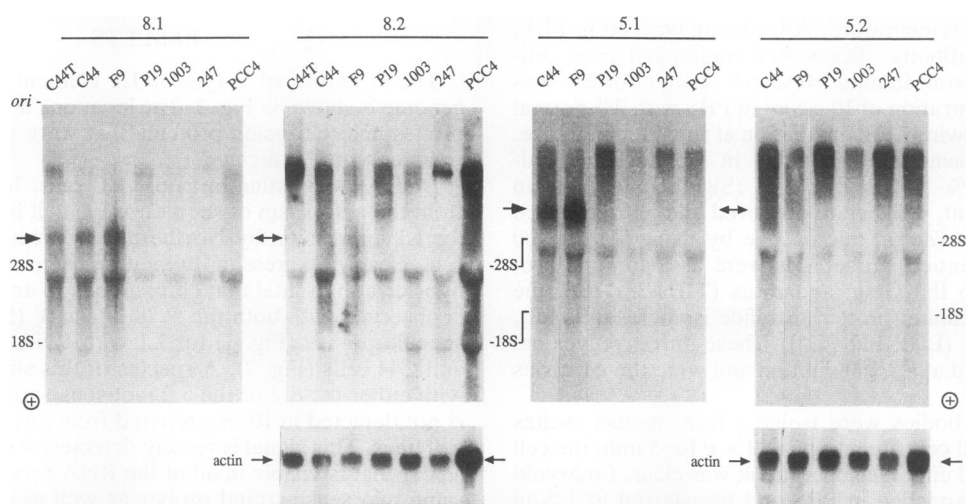


FIG. 2. Northern blots of EC RNAs. Total cellular RNA ($5 \mu\text{g}$) from C44 (C44T, left set only) and cytoplasmic RNA ($5 \mu\text{g}$) from each cell line were loaded onto two gels. Lanes are labelled by the name of the cell line; 247 is ECa247. Following electrophoresis, the RNA was transferred to nitrocellulose and each blot was hybridized three times in succession to three probes: 8.2, 8.1, and actin (left set) or 5.2, 5.1, and actin (right set). Blots were washed and exposed to film; autoradiograms are shown here. The positions of rRNA are indicated. Arrows mark the positions of the expected and/or detected RNA species for each probe.

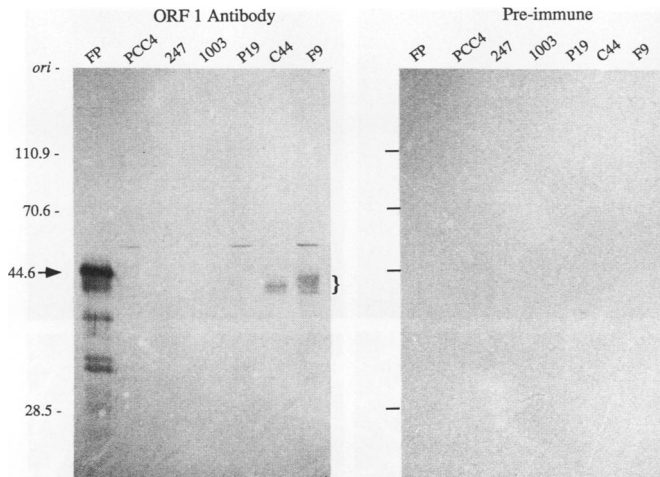


FIG. 3. Western blots of EC proteins. Protein (20 μ g) from the cell lines indicated at the tops of the lanes (247 is ECa247) was fractionated by electrophoresis and blotted to nitrocellulose. Affinity-purified antibody prepared against FP1 or the preimmune IgG from the same rabbit was used to locate the ORF 1 protein(s). The positions of protein size standards are indicated along with their molecular masses (in kilodaltons). The position of the L1 fusion protein, FP1, in lane FP, with a calculated molecular mass of 44.6 kDa, is indicated by an arrow. The bracket delineates the position of the cluster of ORF 1 protein bands in C44 and F9 extracts; the smallest is \sim 41.5 kDa.

both L1 RNA (Fig. 2) and ORF 1 (Fig. 3) expression. Extracts prepared from the C44 cell line behaved as those prepared previously from F9 cells. That is, L1 RNA is found in an RNP complex in the cytoplasm of C44 cells, and ORF 1 protein cofractionates with the full-length, sense-strand RNA in sucrose gradients. Fractionated material from ECa247 cells showed no evidence of full-length L1 RNA or

ORF 1 protein, as expected from the results with unfractionated extracts with this cell line (data not shown).

Immunolocalization of ORF 1 protein in F9 cells and C44 embryoid bodies. On the basis of the results of Northern and Western blot analysis, both full-length, sense-strand L1 RNA (Fig. 2) and ORF 1 protein (Fig. 3) are found in the cytoplasm of F9 cells. Localization of L1 ORF 1 protein in the cytoplasm by indirect immunofluorescence confirms these results. As shown in Fig. 4, the cytoplasm of F9 cells shows the identical, punctate distribution of L1 ORF 1 protein when labelled with affinity-purified antibodies made against either FP1 or FP2. Neither of the preimmune antibody preparations shows this staining pattern, indicating that it is specific to L1 ORF 1.

In an effort to assess whether this punctate staining pattern is due to the association of L1 ORF 1 protein with known cytoplasmic organelles, F9 cells were fixed and stained as described above, with antibodies that recognize a specific protein in the Golgi apparatus (TGN-38), the endoplasmic reticulum (protein disulfide isomerase), and the lysosome (LGP-120). As shown in Fig. 5, none of these antigens, when detected by indirect immunofluorescence, colocalizes with the staining pattern of L1 ORF 1 protein in F9 cells. These data indicate that L1 ORF 1 is not uniformly distributed throughout the Golgi apparatus, the endoplasmic reticulum, or the lysosomes. However, these immunofluorescence results do not rule out a nonuniform distribution of L1 ORF 1 protein in the endoplasmic reticulum. Another noteworthy aspect of L1 ORF 1 staining that is apparent from the results presented in Fig. 5 is the marked variation in the intensity of staining with the anti-L1 antibodies from cell to cell. This intensity variation does not appear to be correlated with cell morphology but could be due to subtle differences in the differentiation state or position in the cell cycle that are not detectable morphologically. In spite of this variation in amount of L1 ORF 1 protein in different cells, the pattern of staining is always the same: punctate and

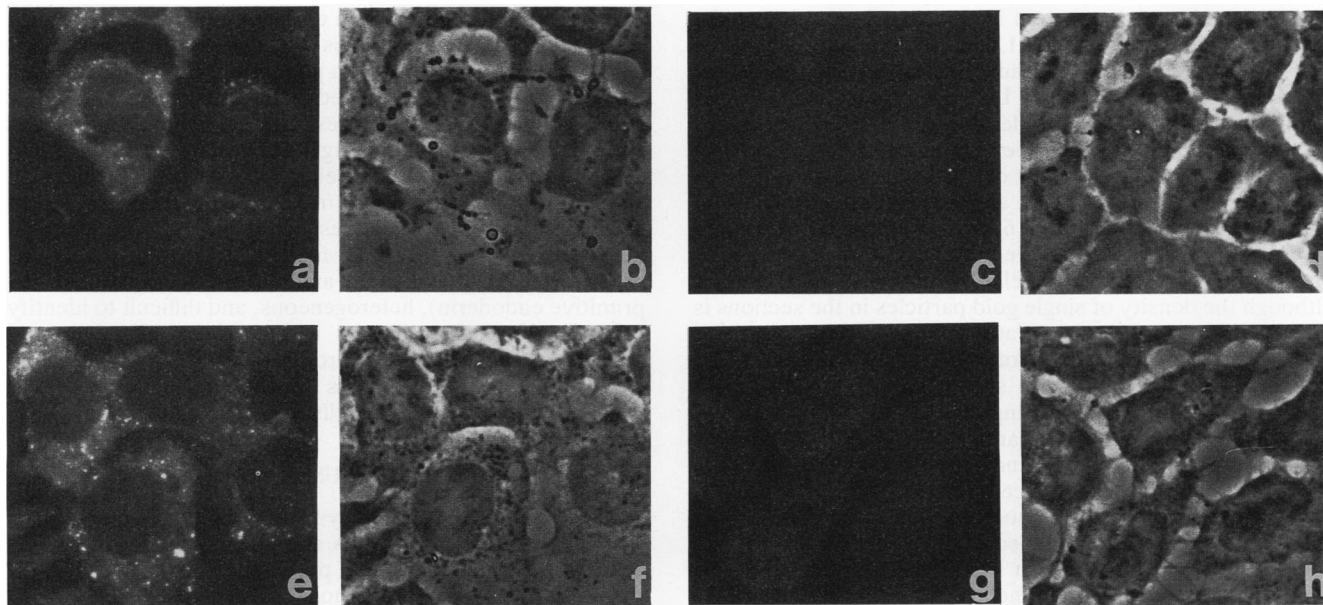


FIG. 4. Specific cytoplasmic staining of F9 cells with antibodies against FP1 and FP2. Micrographs were taken of indirect immunofluorescence (a, c, e, and g) or the corresponding fields in phase (b, d, f, and h), with affinity-purified antibodies against FP1 (a and b), its preimmune IgG (c and d), FP2 (e and f), or its preimmune IgG (g and h). Magnification, \times 2,652.

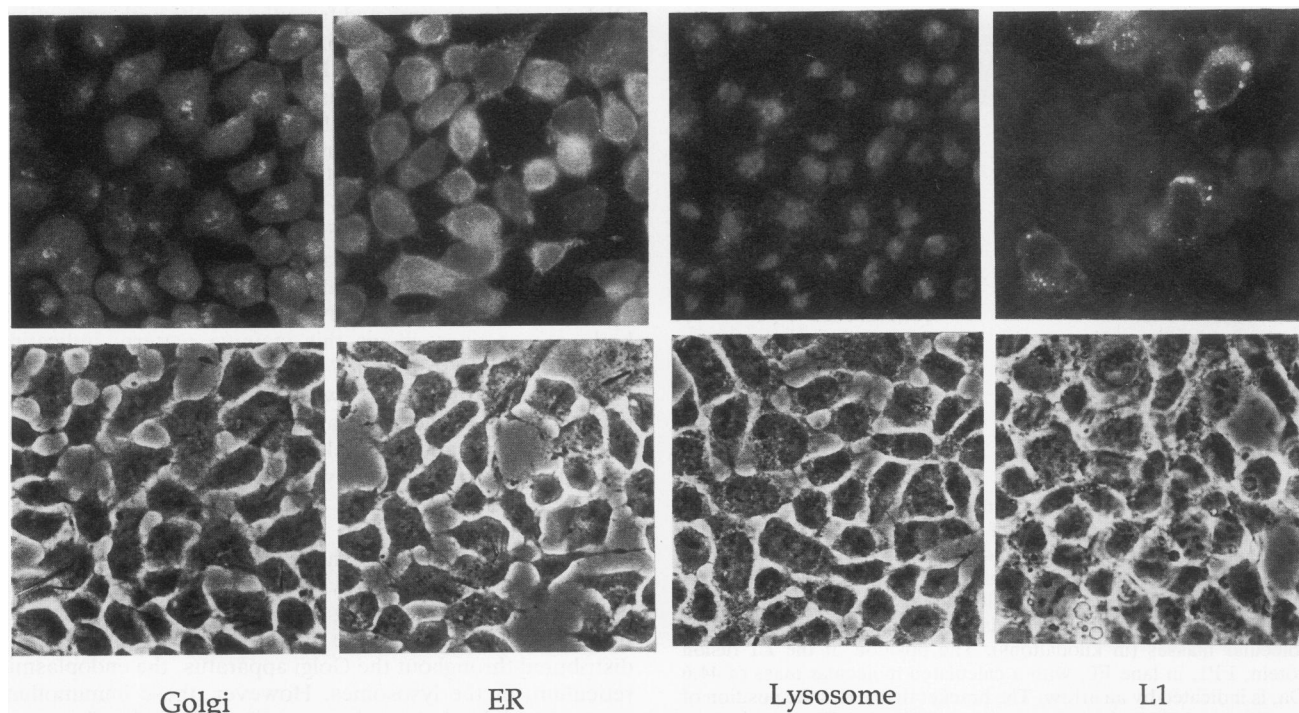


FIG. 5. Comparison of the localization of L1 ORF 1 in F9 cells to other cytoplasmic organelles. Localization of L1 (affinity-purified FP1 antibody) within the cytoplasm of F9 cells is compared with the indirect immunofluorescence staining pattern observed with antibodies against proteins localized to the Golgi apparatus (Golgi, TGN-38), the endoplasmic reticulum (ER; protein disulfide isomerase), or the lysosome (LGP-120). Magnification, $\times 1,860$.

cytoplasmic. This indicates that the expression of L1 ORF 1 protein varies considerably within the population of F9 cells, in contrast to the other three organellar markers, where the intensity of staining (and thus their expression) is relatively uniform from cell to cell. The intense punctate staining observed with antibodies against L1 ORF 1 suggests that ORF 1 proteins are present as aggregates that contain multiple copies of the protein in the cytoplasm of F9 cells. Such aggregates could be the L1 RNPs (or clumps of RNPs) that fractionate in sucrose gradients and may be a complex containing multiple copies of L1 ORF 1 protein in association with each other and full-length, sense-strand L1 RNA.

Immunogold labelling was performed on ultrathin cryosections to visualize and further define the cytoplasmic structure that is labelled by the L1 antibodies, using the electron microscope. As shown in Fig. 6, L1 labelling is characterized by clustering of the gold-labelled protein A. This clustering is not seen in cells similarly labelled with the preimmune IgG, although the density of single gold particles in the sections is the same. For this reason, we attribute the clusters of gold to aggregates of L1 ORF 1 protein. It is often possible to distinguish an electron-dense structure in association with the clusters of antibody labelling (Fig. 6A, D, E, and F). This electron-dense structure appears to have a roughly spherical shape, ranging between 37 and 50 nm in diameter, and is consistent with the appearance of RNA-protein complexes. Finally, the results of immunoelectron microscopy confirm and extend the results obtained by indirect immunofluorescence; the L1 ORF 1 protein is not specifically associated with known cytoplasmic organelles, including the mitochondria or the endoplasmic reticulum, and appears as dispersed aggregates consistent with the appearance of RNP particles that are randomly distributed throughout the cytoplasm.

L1 ORF 1 antibodies did not stain P19, 1003, ECa247, or PCC4 cells by indirect immunofluorescence. This result is in good agreement with the results obtained by Western blotting (Fig. 3) analysis of extracts prepared from these four cell lines. In contrast, C44 embryoid bodies showed strong specific staining with the L1 antibodies when examined by indirect immunofluorescence, again consistent with the Western blotting results from extracts of these cells (Fig. 3). As with F9 cells, L1 ORF 1 is localized to the cytoplasm in individual cells and leads to a punctate staining pattern, and none of the cells are labelled with preimmune antibodies (Fig. 7). The embryoid bodies recovered from C44 ascites are either solid or cystic aggregates of tumor cells and resemble mouse embryos at either the morula or blastocyst stage (34). The inner cells in either of these multicelled structures resemble the corresponding inner cells in normal morula and blastocysts (the inner cell mass) and also EC cells, while the outer cells are more differentiated (often primitive endoderm), heterogeneous, and difficult to identify (34). Significantly, only the inner cells of the embryoid bodies contain L1 ORF 1 protein (Fig. 7), indicating that expression of L1 ORF 1 is highly dependent upon the differentiation state of the cells.

DISCUSSION

These data demonstrate synchronous expression of full-length, sense-strand L1 RNA and ORF 1 protein in two EC cell lines, F9 and C44. ORF 1 protein expressed from both of these lines has an apparent molecular weight by SDS-PAGE that is consistent with expression of the entire ORF 1 protein from the first AUG and no further processing by proteolysis as is the rule with the *gag* proteins of retroviruses (47).

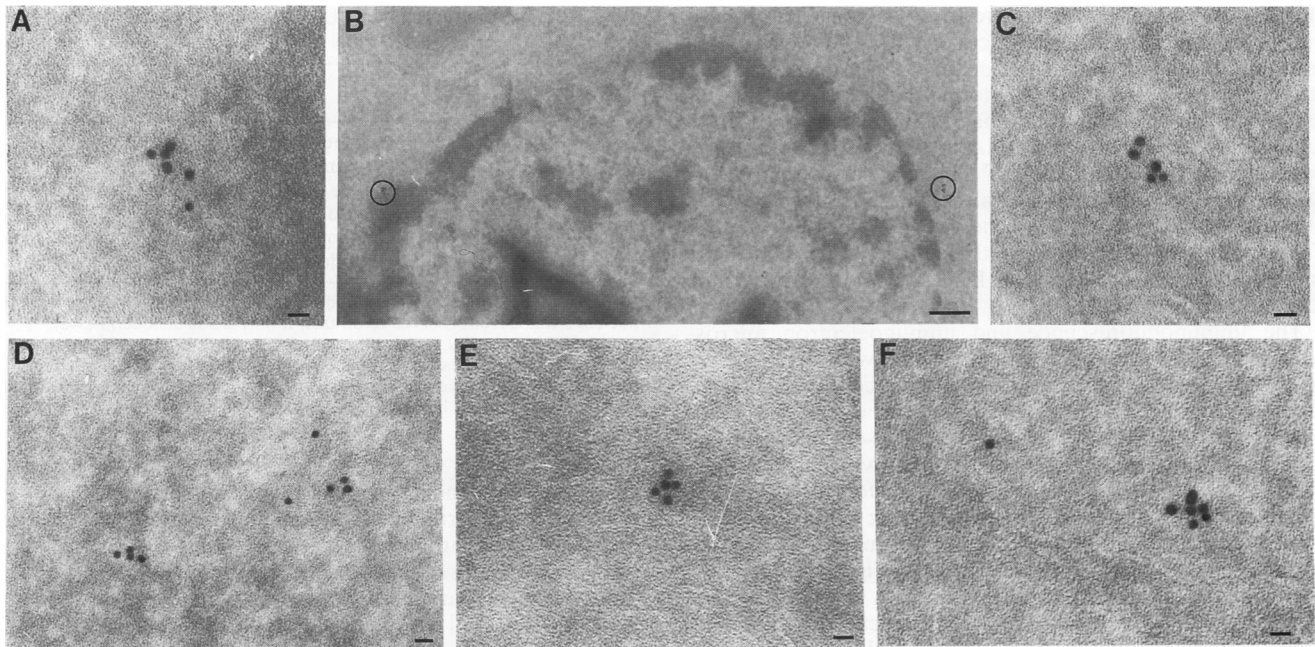


FIG. 6. Detection of L1 ORF 1 expression in F9 cells by immunolabelling on ultrathin cryosections. Series of panels showing electron micrographs of a number of regions from F9 cells that demonstrate small clusters of labelling. (B) Lower magnification for orientation of two clusters (circled) visible on either side of the nucleus; these two are shown at higher magnifications in the adjacent panels (A and C). (E and F) Representative labelling from different cells. Magnification, $\times 139,650$ (bar, 25 nm) in panels A, C, E, and F; $\times 21,280$ (bar, 250 nm) in panel B; and $\times 106,400$ (bar, 25 nm) in panel D.

Similar findings have been reported for the ORF 1 protein from human (23) and rat L1 (18); in no case is there evidence for a fusion protein with ORF 2. Both of the ORF 1 antibodies bind to a cluster of bands by Western analysis, rather than a single protein species. The apparent molecular weights of the individual bands within this cluster of proteins are similar to, but slightly larger than, the size predicted by translation of the ORF 1 protein sequence from L1Md-A2 (24) that react with the antibody. Further studies will be needed to ascertain whether they result from a cytoplasmic posttranslational modification of a single type of L1, such as phosphorylation, or expression of different-size polymorphic

subtypes of L1. We think the former explanation is more likely, because characterization of cDNA sequences from F9 cells has demonstrated that the phylogenetically young L1Md-A2 subtype predominates in the mRNA population (39); the ORF 1 proteins produced from such an RNA population would not be expected to show the observed size heterogeneity. Also, the variation in intensity that we observe between different extracts and over time with the same extract is suggestive of a labile modification, like phosphorylation.

The process of L1 retrotransposition must require a full-length, sense-strand L1 RNA as a transposition intermedi-

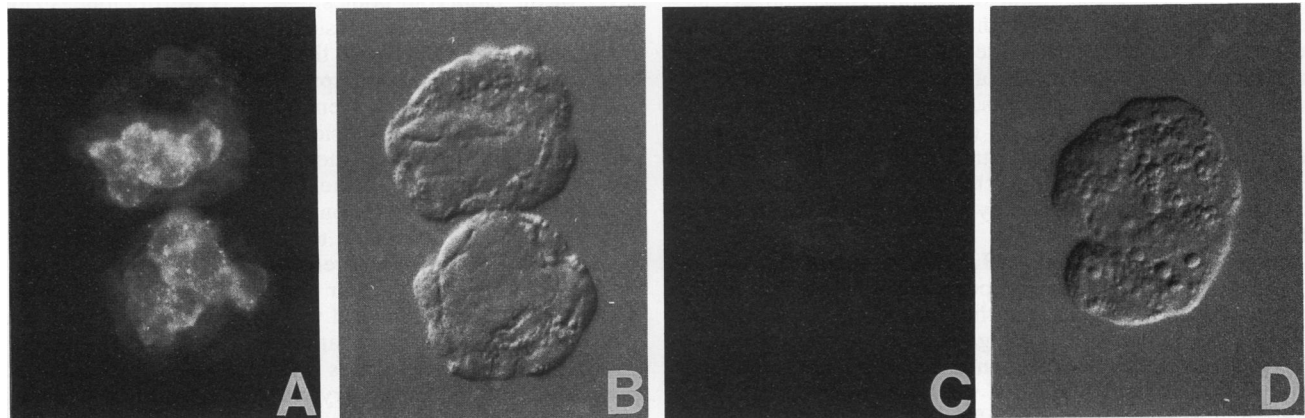


FIG. 7. Restricted expression of L1 ORF 1 protein in C44 embryoid bodies. Micrographs of indirect immunofluorescence of L1 ORF 1 antibody (FP1 [A]) and the preimmune IgG (C) and the corresponding embryoid bodies viewed by differential-interference contrast microscopy (B and D). Magnification, ca. $\times 2,000$.

ate. By analogy with other retrotransposons, it is also probable that the protein products of ORF 1 and ORF 2 are essential intermediates in the transposition process and that they associate with the RNA intermediate. Previous studies using sucrose gradient fractionation have shown that the full-length, sense-strand L1 RNA in F9 cells is present in a complex with protein or an RNP. There is cofractionation of ORF 1 protein with the RNA, although absolute proof of a physical association between the two remains elusive (27). Treatment of C44 extracts in the same way provides the same evidence of packaging of L1 RNA into a cytoplasmic RNP and cofractionation of this RNP with ORF 1 protein.

The immunofluorescence and immunoelectron microscopy results presented here are consistent with the presence of a cytoplasmic L1 RNP. These data indicate that L1 ORF 1 protein is not uniformly distributed throughout the cytoplasm of F9 or C44 cells but, rather, is present in aggregates of multiple protein molecules associated with one another. Furthermore, there may be a tendency for these smaller aggregates or particles to cluster together within the cell. This is suggested by the larger spots observed by indirect immunofluorescence (Fig. 4) as well as in Fig. 6A and D. The aggregates of ORF 1 protein in conjunction with an electron-dense structure are consistent with expectations based on sucrose gradient fractionation of an L1 RNP in which many copies of ORF 1 protein appear to be associated with each L1 RNA. The clusters of these aggregates may explain the presence of a significant amount of L1 RNA and ORF 1 protein that can be recovered from a pellet following sucrose gradient fractionation. The aggregates could be organized by protein-protein interactions between molecules of ORF 1, by RNA-protein interactions between full-length, sense-strand L1 RNA and ORF 1 protein, or by a combination of the two. In addition, protein products of L1 ORF 2 and/or other cellular proteins may be present in the L1 RNP complexes.

We hypothesize that the coexpression of full-length, sense-strand L1 RNA and ORF 1 protein and their cosedimentation through sucrose density gradients under conditions in which L1 RNA is present as an RNP complex (27) provide our first glimpse of the early intermediates involved in L1 transposition. If this is the case, then it is not surprising that EC cells appear to be the best natural source of these putative intermediates (44). The argument for this is largely based on evolutionary considerations and can be summarized as follows: the L1 multigene family is known to undergo concerted evolution (19, 30), and this concerted evolution is largely the result of retrotransposition events (reviewed in reference 17); therefore, retrotransposition must occur in cells that are passed to the next generation. In mammals, this means that retrotransposition must occur in germ cells or their predecessors, the primordial germ cells, or in the early embryo prior to differentiation and commitment. Evidence that L1 retrotransposition occurs during development is provided directly by the human patients whose hemophilia was caused by L1 insertion into the factor VIII gene (20).

Interestingly, in this study of L1 RNA and protein expression in six mouse EC cell types, two of the lines were positive (F9 and C44), but the other four cell types examined (PCC4, ECa247, P19, and 1003) gave no evidence for expression of either the full-length, sense-strand L1 RNA or ORF 1 protein. Although this result was surprising at first, the differential expression in the different lines examined here may provide useful clues about the timing of expression of these L1 products during mouse development. It has been proposed that EC cells represent a window of normal

TABLE 1. Characteristics of the mouse EC cell lines used in this study

Cell line	Developmental potential	Differentiation	Reference
F9	Unipotent	Extraembryonic endoderm ^a	16, 37
C44	Unipotent	Proximal and distal endoderm ^b	34
P19	Multipotent	Three germ layers	36, 37 ^c
1003	Multipotent	Neural ectoderm	33, 42
ECa247	Multipotent	Pretrrophectoderm ^d	34, 35
PCC4	Multipotent	Ectoderm	3

^a There are reports that F9 can differentiate into neural ectoderm, but the line used was never observed to do this; it only made endoderm (34a).

^b Makes endodermal vesicles (proximal and distal) in vitro (34a).

^c Never differentiates into extraembryonic endoderm.

^d ECa247 only expresses trophodermal potential when injected into blastocysts. It has not been observed to make endoderm (34a).

embryo development, such that each independent line is a caricature of a certain cell type of a normal embryo (34, 37). Care was taken throughout this study to ensure that the cells were maintained such that they retained their EC phenotype. This was necessary because earlier work with the human Ntera2D line had demonstrated that induction of differentiation abolished expression of the full-length L1 RNA (44). At the outset of this study, we tested all of the mouse EC cell lines for the expression of full-length, sense-strand RNA and found no expression in any of the lines that were differentiated with retinoic acid, including F9.

The fact that all six of the lines were maintained in their undifferentiated, EC phenotype intimates that these cells are similar. On the basis of a number of criteria, however, these different EC cell lines almost certainly represent different stages or levels of commitment towards different developmental pathways. For example, C44 has been described as a line that represents a late inner cell mass that has expressed endodermal potential. This characterization contrasts to ECa247, which is more like early inner cell mass cells with trophodermal potential (34). A list of the EC lines used for this study and some of their relevant characteristics appear in Table 1. The one obvious link between the two lines that express L1, C44, and F9, is the close relationship of both of these to primitive endoderm. F9 cells, when induced to differentiate in culture, form the two cell types derived from primitive endoderm, visceral and parietal endoderm (16). C44 has cells characteristic of endoderm on the outer cell layer of the embryoid bodies (34). It is worth emphasizing that products of L1 expression are not detected in these differentiated derivatives of the EC phenotype in either of the two lines. Thus, L1 is expressed in mouse EC cells with the potential to make early endoderm and its derivatives.

The restriction of expression of L1 to these EC cells is intriguing in light of normal mouse development, because the primordial germ cells of the mouse are detected first in extraembryonic mesoderm on about day 7 of gestation (10). It has not yet been possible to trace the origin of the eight cells seen at this stage to even earlier stages; thus, it is not possible to be certain of their origins. In light of the requirement for L1 expression and transposition in cells that are destined for the next generation and these results demonstrating that L1 expression is restricted to EC cells with the potential to make extraembryonic endoderm, it is tempting to suggest that the primordial germ cells of the mouse arise from the inner cell mass of the blastocyst from cells with the potential to make extraembryonic endoderm and that L1 expression will provide a valuable marker for these cells. A

further prediction of these results is that L1 expression, and hence transposition, occurs in the subset of cells in the early mammalian embryo that are destined to form the primordial germ cells.

ACKNOWLEDGMENTS

We thank G. B. Pierce and A. Lewellyn for cell lines, M. Ladinski and the Hepatobiliary Core (NIH DK34914) for electron microscopy, K. Howell for antibodies, and P. Bandyopadhyay, K. Howell, V. Kolosha, M. Lohka, R. Lasher, G. B. Pierce, and H. Wichman for helpful discussion.

This work was supported by NIH grant GM40367 to S.L.M.

ADDENDUM IN PROOF

Since the submission of this article, Packer et al. (A. I. Packer, K. Manova, and R. F. Bachvarova, *Dev. Biol.* 157:281–283, 1993) have reported the detection of an 8-kb sense-strand L1 RNA in mouse blastocysts.

REFERENCES

- Boeke, J. D. 1989. Transposable elements in *Saccharomyces cerevisiae*, p. 335–374. In D. E. Berg and M. M. Howe (ed.), *Mobile DNA*. American Society for Microbiology, Washington, D.C.
- Chomczynski, P., and N. Sacchi. 1987. Single step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162:156–159.
- Darmon, M., J. Bottenstein, and G. Sato. 1981. Neural differentiation following culture of embryonal carcinoma cells in a serum-free defined medium. *Dev. Biol.* 85:463–473.
- Deragon, J.-M., D. Sinnett, and D. Labuda. 1990. Reverse transcriptase activity from human embryonal carcinoma cells Ntera2D1. *EMBO J.* 9:3363–3368.
- Dombroski, B. A., S. L. Mathias, E. Nanthakumar, A. F. Scott, and H. H. Kazazian, Jr. 1991. Isolation of an active human transposable element. *Science* 254:1805–1808.
- Doolittle, R. F., D. F. Feng, M. S. Johnson, and M. A. McClure. 1989. Origins and evolutionary relationships of retroviruses. *Q. Rev. Biol.* 64:1–30.
- Dudley, J. P. 1987. Discrete high molecular weight RNA transcribed from the long interspersed repetitive element L1Md. *Nucleic Acids Res.* 15:2581–2592.
- Evans, J. P., and R. D. Palmiter. 1991. Retrotransposition of a mouse L1 element. *Proc. Natl. Acad. Sci. USA* 88:8792–8795.
- Freedman, R. 1987. Folding into the right shape. *Nature (London)* 329:196–197.
- Ginsberg, M., M. H. Snow, and A. McClaren. 1990. Primordial germ cells in the mouse embryo during gastrulation. *Development* 110:521–528.
- Griffiths, G., K. Simons, G. Warren, and K. T. Tokuyasu. 1983. Immunoelectron microscopy using thin, frozen sections: application to studies of the intracellular transport of Semliki Forest virus spike glycoproteins. *Methods Enzymol.* 96:466–485.
- Gunning, P., P. Ponte, H. Okayama, J. Engel, H. Blau, and L. Kedes. 1983. Isolation and characterization of full-length cDNA clones for human α , β , and γ actin mRNAs: skeletal but not cytoplasmic actins have an amino-terminal cysteine that is subsequently removed. *Mol. Cell. Biol.* 3:787–795.
- Hardies, S. C., S. L. Martin, C. F. Voliva, C. A. Hutchison III, and M. H. Edgell. 1986. An analysis of replacement and synonymous changes in the rodent L1 repeat family. *Mol. Biol. Evol.* 3:109–125.
- Hardies, S. C., and B. A. Rikke. 1990. A selfish retrotransposition model for rodent LINE-1. *UCLA Symp. Mol. Cell. Biol. New Ser.* 122:127–134.
- Harlow, E., and D. Lane. 1988. *Antibodies: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Hogan, B. L. M., D. P. Barlow, and R. Tilly. 1983. F9 terato carcinoma cells as a model for the differentiation of parietal and visceral endoderm in the mouse embryo. *Can. Sur.* 2:115–140.
- Hutchison, C. A., III, S. C. Hardies, D. D. Loeb, W. R. Shehee, and M. H. Edgell. 1989. LINES and related retrotransposons: long interspersed repeated sequences in the eucaryotic genome, p. 593–617. In D. E. Berg and M. M. Howe (ed.), *Mobile DNA*. American Society for Microbiology, Washington, D.C.
- Ives, H., O. Kahre, and M. Speck. 1992. Translation of the rat LINE bicistronic RNAs *in vitro* involves ribosomal reinitiation instead of frameshifting. *Mol. Cell. Biol.* 12:4242–4248.
- Jubier-Maurin, V., B. J. Dod, J. Bellis, M. Piechaczyk, and G. Roizes. 1985. Comparative study of the L1 family in the genus *Mus*: possible role of retrotransposition and conversion events in its concerted evolution. *J. Mol. Biol.* 184:547–564.
- Kazazian, H. H., Jr., C. Wong, H. Yousoufian, A. F. Scott, D. G. Phillips, and S. E. Antonarakis. 1988. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature (London)* 332:164–166.
- Kornfeld, S., and I. Mellman. 1989. The biogenesis of lysosomes. *Annu. Rev. Cell Biol.* 5:483–525.
- Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* 227:680–685.
- Liebold, D. M., G. D. Swergold, M. F. Singer, R. E. Thayer, B. A. Dombroski, and T. F. Fanning. 1990. Translation of LINE-1 DNA elements *in vitro* and in human cells. *Proc. Natl. Acad. Sci. USA* 87:6990–6994.
- Loeb, D. D., R. W. Padgett, S. C. Hardies, W. R. Shehee, M. B. Comer, M. H. Edgell, and C. A. Hutchison III. 1986. The sequence of a large L1Md element reveals a tandemly repeated 5' end and several features found in retrotransposons. *Mol. Cell. Biol.* 6:168–182.
- Luan, D. D., M. H. Korman, J. L. Jakubczak, and T. H. Eickbush. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 72:595–605.
- Luzio, J. P., B. Brake, G. Banting, K. E. Howell, P. Braghetta, and K. K. Stanley. 1990. Identification, sequencing and expression of an integral membrane protein of the trans-Golgi network (TGN38). *Biochem. J.* 270:97–102.
- Martin, S. L. 1991. Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Mol. Cell. Biol.* 11:4804–4807.
- Martin, S. L. 1991. LINES. *Curr. Opin. Genet. Dev.* 1:505–508.
- Martin, S. L., C. F. Voliva, F. H. Burton, M. H. Edgell, and C. A. Hutchison III. 1984. A large interspersed repeat found in mouse DNA contains a long open reading frame that evolves as if it encodes a protein. *Proc. Natl. Acad. Sci. USA* 81:2308–2312.
- Martin, S. L., C. F. Voliva, S. C. Hardies, M. H. Edgell, and C. A. Hutchison III. 1985. Tempo and mode of concerted evolution in the L1 repeat family of mice. *Mol. Biol. Evol.* 2:127–140.
- Mathias, S. L., A. F. Scott, H. H. Kazazian, Jr., J. D. Boeke, and A. Gabriel. 1991. Reverse transcriptase encoded by a human transposable element. *Science* 254:1808–1810.
- Nagai, K., and H. C. Thorgersen. 1987. Synthesis and sequence-specific proteolysis of hybrid proteins produced in *E. coli*. *Methods Enzymol.* 153:461–481.
- Nicolas, J. F., P. Avner, J. Gaillard, J. L. Guenet, H. Jakob, and F. Jacob. 1976. Cell lines derived from teratocarcinomas. *Cancer Res.* 36:4224–4231.
- Parchment, R. E., R. A. Gramzinski, and G. B. Pierce. 1990. Neoplastic embryoid bodies of embryonal carcinoma C44 as a source of blastocoele-like fluid. *Differentiation* 43:51–58.
- Pierce, B. Personal communication.
- Pierce, G. B., J. Arechaga, A. Jones, A. Lewellyn, and R. S. Wells. 1987. The fate of embryonal-carcinoma cells in mouse blastocysts. *Differentiation* 33:247–253.
- Pruitt, S. C., and T. A. Natoli. 1992. Inhibition of differentiation by leukemia inhibitory factor distinguishes two induction path-

- ways in P19 embryonal carcinoma cells. *Differentiation* **50**:57-65.
37. Rudnicki, M. A., and M. W. McBurney. 1987. Cell culture methods and induction of differentiation of embryonal carcinoma cell lines, p. 19-49. *In* E. J. Robertson (ed.), *Teratocarcinomas and embryonic stem cells: a practical approach*. IRL Press, Oxford.
 38. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
 39. Schichman, S. A., D. M. Severynse, M. H. Edgell, and C. A. Hutchison III. 1992. Strand-specific LINE-1 transcription in mouse F9 cells originates from the youngest phylogenetic subgroup of LINE-1 elements. *J. Mol. Biol.* **224**:559-574.
 40. Segal-Bendirdjian, E., and T. Heidmann. 1991. Evidence for a reverse transcription intermediate for a marked LINE transposon in tumoral rat cells. *Biochem. Biophys. Res. Commun.* **181**:863-870.
 41. Shehee, W. R., S.-F. Chao, D. D. Loeb, M. B. Comer, C. A. Hutchison III, and M. H. Edgell. 1987. Determination of a functional ancestral sequence and definition of the 5' end of A type mouse L1 elements. *J. Mol. Biol.* **196**:757-767.
 42. Sherman, M. I. 1975. Differentiation of teratoma cell line PCC4.azal *in vitro*, p. 191-205. *In* M. I. Sherman and D. Solter (ed.), *Teratomas and differentiation*. Academic Press, Inc., New York.
 43. Skowronski, J., T. G. Fanning, and M. F. Singer. 1988. Unit-length LINE-1 transcripts in human teratocarcinoma cells. *Mol. Cell. Biol.* **8**:1385-1397.
 44. Skowronski, J., and M. F. Singer. 1985. Expression of a cytoplasmic LINE-1 transcript is regulated in a human teratocarcinoma cell line. *Proc. Natl. Acad. Sci. USA* **82**:6050-6054.
 45. Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**:60-89.
 46. Tooze, J., J. F. Kern, S. D. Fuller, and K. E. Howell. 1989. Condensation-sorting events in the rough endoplasmic reticulum of exocrine pancreatic cells. *J. Cell Biol.* **109**:35-50.
 47. Varmus, H., and P. Brown. 1989. Retroviruses, p. 53-108. *In* D. E. Berg and M. M. Howe (ed.), *Mobile DNA*. American Society for Microbiology, Washington, D.C.
 48. Wahl, G. M., M. Stern, and G. R. Stark. 1979. Efficient transfer of large DNA fragments from agarose gels to diazobenzoyloxymethyl-paper and rapid hybridization using dextran sulfate. *Proc. Natl. Acad. Sci. USA* **76**:3683-3687.
 49. Xiong, Y., and T. H. Eickbush. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* **9**:3353-3362.