Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 5641–5645, September 1983
Developmental Biology

Early differential tissue expression of transposon-like repetitive DNA sequences of the mouse

(retrovirus/embryonal carcinoma/differentiation)

PHILIPPE BRÛLET*, MOURAD KAGHAD*, YI-SHENG XU*, ODILE CROISSANT†, AND FRANÇOIS JACOB*

*Service de Génétique Cellulaire du Collège de France et de l'Institut Pasteur and †Service des Papillomavirus, Institut Pasteur, 25, rue du Dr. Roux, 75724 Paris Cedex 15, France

Contributed by François Jacob, May 31, 1983

ABSTRACT **Another family of long moderately repetitive and dispersed sequences has been identified in the mouse genome. These sequences have a transposon-like structure. A 6-kilobase RNA transcript is detected in undifferentiated embryonal carcinoma cell lines but not in any of the differentiated cell types tested. By R-loop formation, the RNA is colinear with a DNA fragment from a randomly selected genomic clone.**

Analysis at the molecular level of the first cellular differentiation in the mouse embryo, the formation of the blastocyst, has met considerable difficulty. For the trophectoderm, some intermediate filament protein can be used as a specific marker: a monoclonal antibody and a cDNA clone have been obtained that allow study of the expression of the proteins (1, 2). For the undifferentiated totipotent state, we describe here the presence of a RNA resulting from the specific transcription, in these cells, of some component of a new family of transposon-like or integrated retrovirus-like elements.

MATERIALS AND METHODS

Cells. Most teratocarcinoma cell lines used in this study are referred to in Kemler *et al.* (3). C17-S1-1003 is a C3H-derived EC line (4). DB1 and DB2 are cell lines obtained by D. Boullier after *in vitro* induction of differentiation by retinoic acid (5). Mouse and rat 3T3, neuroblastoma N18 and NS20 (6), and hypoxanthine/aminopterin/thymidine-sensitive L-cell lines were cultured in Dulbecco's modified Eagle's medium/10% fetal calf serum.

Nucleic Acids. RNA extraction and purification and the construction of a cDNA library were carried out as described (7). High molecular weight DNAs were prepared from cell lines, livers, or lungs, according to Blin and Stafford (8). Colony, plaque, and blot hybridizations, were done according to Hanahan and Meselson (9) and Maniatis *et al.* (10). Washing stringency was 30 mM NaCl/3 mM Na citrate, pH 7/0.1% $NaDodSO_4$ at 66°C. Enzymes were obtained from Bethesda Research Laboratories, Boehringer Mannheim, and New England BioLabs.

Nucleotide Sequence Analysis. [α-^{32}P]dCTP (Amersham) and the Klenow polymerase were used to label the *Pst* I ends of plasmid pMS3 (10). Single strands were fractionated on a 7% acrylamide gel and electrophoretically recovered; sequences were determined by the Maxam and Gilbert method (11). Computer analysis was carried out by Intelligenetics (Palo Alto, CA).

R-Loop Hybridization. R loops (12) were formed for 90 min at 58°C in 10 μl of 70% formamide/0.4 M NaCl/50 mM Pipes, pH 6.8/1 mM EDTA. Temperature was lowered to 25°C in 3 hr. The mixture was then treated with glyoxal for 2 hr at 12°C (13) and dialyzed. Samples for electron microscopy were spread from 40% formamide/100 mM Tris base/10 mM EDTA, pH 8.5, containing 0.06 mg of cytochrome *c* per ml over 10% formamide in 10 mM Tris base (pH 8.5). Circular pBR322 was used as a length standard.

RESULTS

Isolation and Characterization of cDNA Clone MS3. From poly(A)-RNA extracted from an EC line, PCC3/A/1, a cDNA library was prepared onto nitrocellulose filters. Duplicated filters were screened by *in situ* hybridization with [^{32}P]cDNA from EC line F9 and from trophoblastoma TDM-1. One clone, MS3, hybridizing with F9 cDNA but not with trophoblastoma cDNA, was further analyzed. It contains a 250-base-pair (bp) DNA insertion into the *Pst* I site of pBR322.

Cellular RNA and poly(A)-RNA from various undifferentiated cell lines were fractionated by gel electrophoresis and, after transfer onto nitrocellulose, analyzed with the MS3 probe. MS3 hybridizes to a RNA species migrating at about 6 kilobases (kb) in various undifferentiated EC lines. No RNA was detected in several differentiated lines, in particular TDM-1, a trophoblastoma line, and PYS-2, an endodermal line (Fig. 1). Controls involved probing the blots with a plasmid containing an actin sequence.

The genomic organization of the MS3 sequence was studied by Southern blot analysis. High molecular weight DNA from F9 was digested with various restriction enzymes and hybridized to the pMS3 probe (Fig. 2*A*). After 1 hr of exposure, many bands could be distinguished.

A genomic DNA library constructed with Charon 4A phage, a gift from François Rougeon, was screened with the MS3 probe. The library was prepared by partial *Eco*RI digestion of DNA from late embryos of BALB/c mice. Positive signals originated from 0.5 to 1% of the plaques, indicating that the MS3 sequence is repeated one to a few thousand times and dispersed in the mouse genome.

The presence of those repeated elements was investigated in various cell lines and mouse strains by the Southern blotting technique. Sequences hybridizing with pMS3 were found in all four groups belonging to the *Mus* genus—1 (*musculus* C3H, DW, SJL, DDK, BALB/c, and 129), 2 (PWK), 3 (*spretus* SPE), and 4A (*spicilegus* SBS).

Characterization of a Genomic DNA Sequence. From the BALB/c genomic DNA library, three plaques hybridizing to pMS3 were picked and the recombinant molecules were iso-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: bp, base pair(s); kb, kilobase(s); IAP, intracisternal A particle; ETn, early transposon.

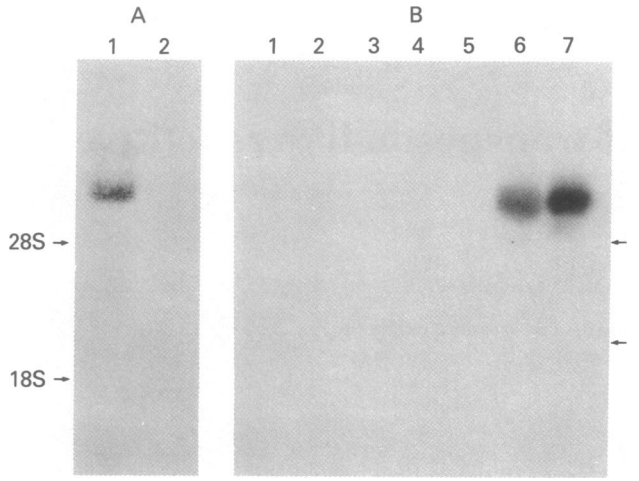

FIG. 1. Analysis of RNA extracted from various cell lines. (*A*) Lanes 1 and 2: 5-μg samples of F9 and TDM-1 poly(A)-RNA, denatured with glyoxal, were fractionated on a 1.1% agarose gel, transferred to nitrocellulose filters and hybridized to nick-translated pMS3 (10) (specific activity, 1.8×10^7 cpm/μg). Fluorographic exposure was for 5 hr with Kodak XR film. (*B*) Five-microgram samples of RNA from various cell lines were denatured with formaldehyde and fractionated on a 7-cm-long 1.1% agarose minigel (Bethesda Research Laboratories). Hybridization was done at 42°C in the presence of 50% formamide with nick-translated pMAC-2 (specific activity, 3×10^7 cpm/μg). Fluorographic exposure was 3 hr. Lanes: 1, DB2; 2, hypoxanthine/aminopterin/thymidine-sensitive L; 3, PYS-2; 4, mouse 3T3; 5, C17-S1-D1 984; 6, PCC4; 7, F9.

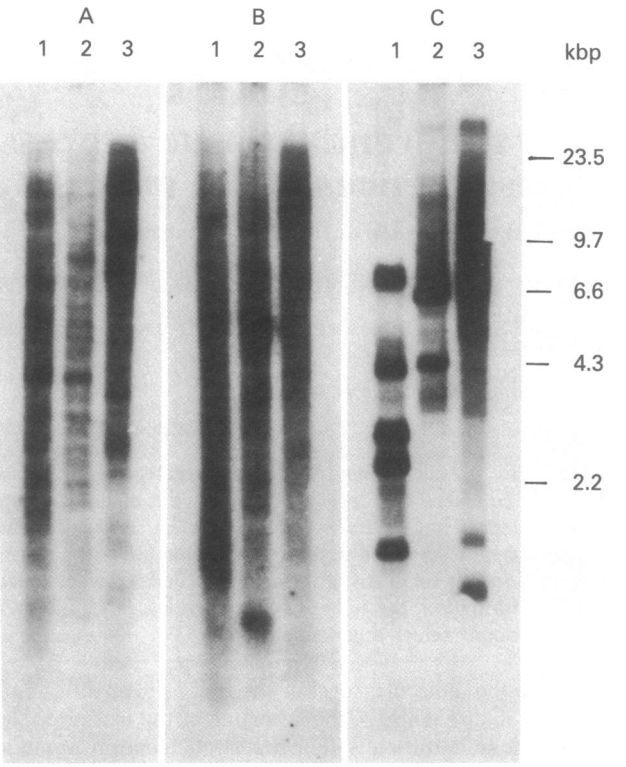

FIG. 2. Southern blot analysis of F9 DNA. High molecular weight DNA extracted from F9 cells was digested to completion with *Pst* I (lanes 1), *Hin*dIII (lanes 2), or *Eco*RI (lanes 3). The restricted DNAs (10 μg per lane) were subjected to electrophoresis on an 0.8% agarose gel and blotted onto nitrocellulose filters. Hybridization was for 19 hr at 42°C with 0.5 μg of nick-translated DNA of pMS3 at a specific activity of 7.5×10^7 cpm/μg (*A*), MG1 at a specific activity of 6.5×10^7 cpm/μg (*B*), or IAP genomic probe at a specific activity of 7×10^7 cpm/μg (*C*). Fluorographic exposures were 1 hr for *A*, 5 hr for *B*, and 30 min for *C*. Size markers displayed on the side are *Hin*dIII-digested λ DNA fragments. kbp, Kilobase pairs.

lated. On two of those molecules, the sequence MS3 is duplicated. One of them, MG1, was further analyzed in detail by restriction analysis. A *Pvu* II/*Pvu* II 6-kb fragment containing the two MS3 sequences was subcloned into the *Eco*RI site of pBR322 (plasmid pMAC-2) after addition of *Eco*RI linkers (Fig. 3). The restriction map of the fragment and the position of the two MS3 repeats are shown in Fig. 4*A*. They are located at the extremities of the 6-kb fragment. The DNA contained between these repeats was further investigated by several methods. (*i*) The internal *Pst* I/*Pst* I fragment was purified by electrophoresis and, after labeling, used as a probe for Southern and RNA blot analyses (Fig. 4). Many bands of genomic DNA were detected with this probe and they are rather similar in five different cellular DNAs. In the RNA blot analysis of RNA populations, again, a RNA species with an identical mobility was detected with this internal fragment in EC cells but not in trophoblastoma cells. (*ii*) pMAC-2 was digested with various enzymes and transferred onto nitrocellulose, and the blots were probed with labeled cellular DNA or poly(A)-RNA from different cell types (data not shown). The fragments were labeled with the DNA and RNA from EC cells (F9) but not with RNA from trophoblastoma. (*iii*) R-loop electron micrographs were obtained after hybridizing the poly(A)-RNA extracted from the EC line PCC4 with pMAC-2 plasmid DNA that had been digested with *Bam*HI, which cuts only once in the pBR322 sequence. The RNA and the genomic insert in pMAC-2 are colinear (Fig. 5). Since this genomic recombinant molecule MG1 was picked at random out of a few thousand, it is likely that we are dealing with a family of long repeated sequences that, by heteroduplex formation, are rather well conserved in the mouse genome. We have observed, however, a certain amount of polymorphism at the restriction site level: some sites are present in pMS3 but not in the repeated sequences in pMAC-2.

One sequence hybridizing with pMS3 lies on a *Bgl* I/*Sst* I fragment. From the restriction map, the 5′ end of the pMS3 200-bp fragment is located somewhere between 725 and 1,115 bp downstream from the pBR322 *Bam*HI site. Its 3′ end should be between 925 and 1,315 bp downstream from the *Bam*HI site. Analysis of the R-loop pictures indicates that the RNA starts somewhere between 1,170 and 1,300 bp downstream from this *Bam*HI site. The 5′ end of the RNA is therefore either located in the repeated sequence or just on the 3′ side of it. We already know that the 3′ end of the RNA contains this sequence because the original pMS3 plasmid was obtained from oligo(dT) priming of the reverse transcription of the RNA. Also, pMS3 contains a polyadenylylation site and an adenosine tail (see below).

Finally, the analysis of denatured pMAC-2 does not indicate the presence of a hairpin sequence at the position corresponding to those regions of pMAC-2 that hybridize with MS3. It seems therefore that the repeats at the ends of the pMAC-2 molecule are direct repeats. The length of these direct repeats has not yet been determined.

These results indicate that the genome contains repeats about 6 kb long, terminated by direct repeats several hundred base pairs long. The RNA is colinear with most of the DNA. It starts in, or just downstream of, one of the terminal sequences MS3 and carries the same sequence at its 3′ end. These structural features are reminiscent of transposable elements and integrated retroviruses.

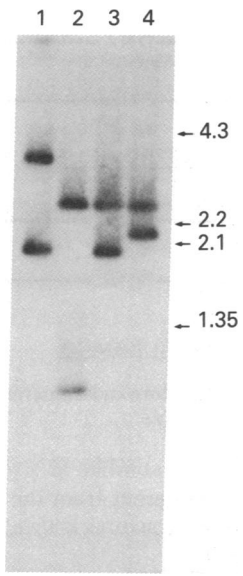

FIG. 3. MS3 sequences are redundant in plasmid pMAC-2. The genomic insert of pMAC-2 was excised by digestion with *Eco*RI and then digested with other restriction enzymes. The fragments were fractionated on a 1% agarose gel before Southern analysis. Lanes: 1, *Hin*dIII; 2, *Pst* I/*Sst* I; 3, *Pst* I/*Hin*dIII; 4, *Pst* I. Plasmid pMS3 was digested with *Pst* I and end labeled with [α-^{32}P]dCTP (Amersham) and the Klenow polymerase (Boehringer Mannheim). The ^{32}P-labeled insertion MS3 was purified from pBR322 sequences on a 30-cm-long 7% acrylamide gel and electrophoretically recovered from the acrylamide.

Nucleotide Sequences and Comparison with Known Sequences. The nucleotide sequence of pMS3 is as follows:

```
1          11         21         31         41
CGGCCCCCTC  CCCTGGCCTG  AAGCCTGCCT  GCTCGGGGTG  GAGCTTCCTG

51         61         71         81         91
TTCACCCGTT  CTGCCACGCC  CACTGCTGGA  ACCTGAGGAG  CCACACACGT

101        111        121        131        141
GCACCTTTCT  ACTGGACCAG  AGATTATTCG  GCGGGAATCG  GGTCCCCTCC

151        161        171        181        191
CCCTTCCTTC  ATAACTAGTG  TCGCAACAAT  AAAATTTGAG  CCTTGATCA
```

The main features of this sequence are (*i*) a terminal C-A followed by an adenosine tail, (*ii*) a polyadenylylation site A-A-T-A-A-A 20 bp upstream, (*iii*) a 19-bp inverted repeat (with two mismatches) 110 bp apart and, (*iv*) between the inverted repeats, a 13-bp-long series of alternating purines and pyrimidines with a potential capacity to switch to the Z-DNA form. This sequence was compared with the DNA sequences contained in a gene bank. No significant homology was found with reported eukaryotic, prokaryotic, or viral sequences.

Two families of long repeated sequences have been reported in the mouse: the VL30 (14–16) and the intracisternal A particle (IAP) families (17–19). Both have a transposon-like or retrovirus-like structure similar to the one reported in this article. The VL30 sequences are repeated 50 to 100 times in the mouse genome—i.e. one order of magnitude lower than for the MS3 family. The long terminal repeat DNA sequence (sent to us by M. J. Getz) was compared with the MS3 sequence. Only three local homologies, eight to nine bp long, were found.

A family of IAP sequences has been described by Kuff *et al.* (17). A cDNA probe was prepared from a 35S RNA extracted from an IAP-containing myeloma. This probe led to the characterization of a family of long well-conserved genomic DNA

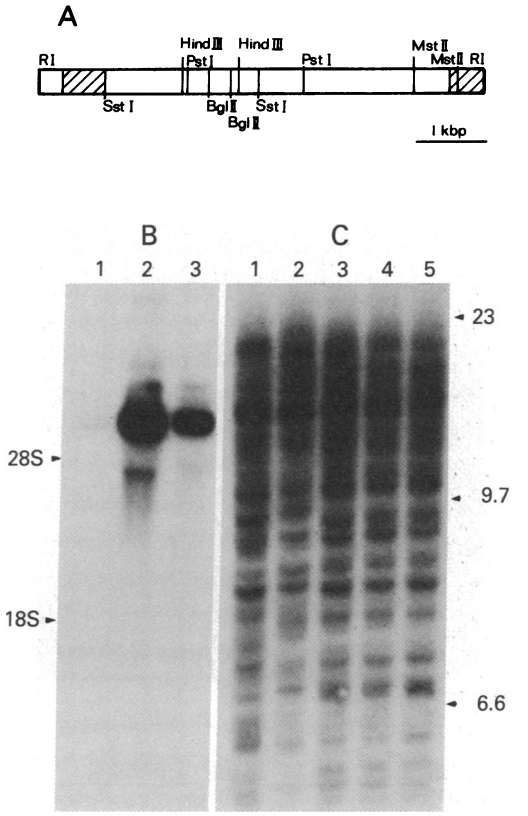

FIG. 4. Analysis of the genomic fragment of plasmid pMAC-2. (*A*) Restriction map of this fragment. Hatched areas indicate restriction fragments hybridizing with MS3. (*B and C*) RNA blot and Southern analyses of RNA and DNA from various cell lines, with the *Pst* I/*Pst* I internal fragment shown in *A*. The 1.65-kb fragment was purified by agarose gel electrophoresis, recovered by electroelution, and labeled by nick-translation with [α-^{32}P]dATP/[α-^{32}P]dCTP (Amersham). (*B*) Ten-microgram samples of poly(A)-RNA, denatured with formaldehyde, were fractionated on a 1.1% agarose gel. Fluorographic exposure was for 3 hr. Lanes: 1, TDM-1 cells; 2, F9 cells; 3, PCC4 cells. (*C*) Ten-microgram samples of *Eco*RI-digested DNA from various cell lines were electrophoresed for 48 hr on a 40-cm-long 0.8% agarose gel prior to Southern analysis. Size markers are in kilobase pairs (kbp). Lanes: 1, F9; 2, TDM-1; 3, PSA-1-NG2; 4, lung from a DDK mouse; 5, lung from a BALB/c mouse. The two latter DNAs were a gift from H. Condamine.

sequences homologous to the 35S RNA, present at about 1,000 copies in the genome of *Mus musculus*. The DNA sequence of the long terminal repeat of one member, MIA14 (20), was compared with the MS3 sequence. Again only short common fragments were found, one of 28 nucleotides with eight mismatches. Furthermore, the MS3 DNA was hybridized with a 5.5-kb genomic probe containing most of the information from the 35S RNA. No cross-hybridization was detected. The IAP genomic clone was also used to probe the RNA prepared from various cell types. Three RNA species, two of which have a higher molecular weight than MS3 RNA, were detected in the various cell lines used, irrespective of their developmental state. These RNA species are particularly abundant in trophoblastoma TDM-1 and endodermal PYS cell lines, which contain no RNA complementary to pMS3 and to the 6-kb genomic DNA plasmid pMAC-2. Patterns of genomic DNA fragments hybridizing either to pMS3, or to the original MG1 18-kb genomic DNA recombinant in λ, or to the 6-kb genomic sequences in pMAC-2, were analyzed (Fig. 2). They are different from those observed with an IAP genomic recombinant for which characteristic patterns have been reported after digestion with *Pst* I, *Hin*dIII, and *Eco*RI (17). Finally, the genomic DNA recombinant pMAC-2, which

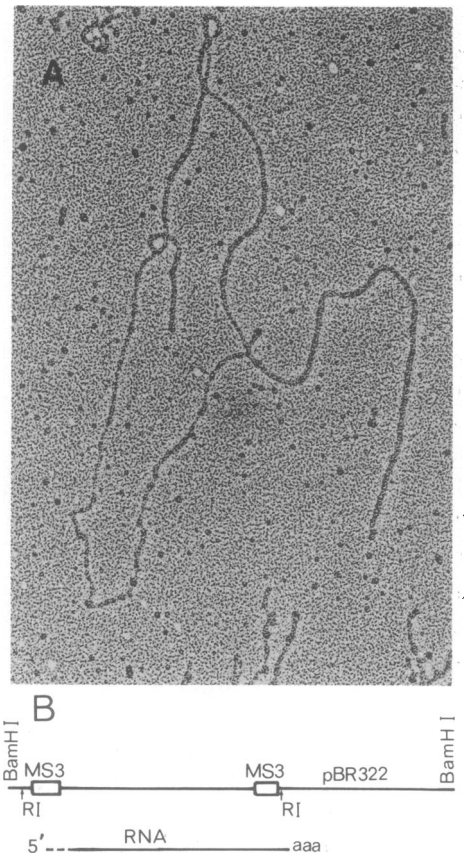

FIG. 5. (*A*) R-loop electron micrograph. pMAC-2 was digested with *Bam*HI, which cuts at only one point in the pBR322 sequence (*B*). A 0.25-μg sample of DNA was hybridized at 58°C for 90 min with 5 μg of poly(A)-RNA extracted from the EC line PCC4. (×48,000.)

is homologous by the R-loop technique to the RNA in EC cells, does not contain fragments corresponding to the IAP characteristic *Pst* I, *Hin*dIII, and *Eco*RI internal fragments. The MS3 family and the IAP family reported by Kuff *et al.* (17) are therefore clearly distinct.

Differential Tissue Expression. RNA populations were fractionated by agarose gel electrophoresis under denaturing conditions and probed with the pMAC-2 plasmid containing the 6-kb genomic DNA insert. The 6-kb RNA and two additional much less abundant RNA species are detected in the EC cell lines tested—namely, PCC4 Azal, PCC3/A/1, PSA-1-NG2, F9-41, and the C3H-derived C17-S1-1003. They were not found at a detectable rate in the differentiated cell lines tested—namely, trophoblastoma 3/TDM 1, parietal endoderm PYS-2, embryonic fibroblast 3/A/1-D3, osteogenic mesenchyme 3/A/1-D1, myoblastic cell C17-S1-D1 984, embryonic fibroblast 3T3, differentiated F9-derived cells DB1 and DB2, hypoxanthine/aminopterid/thymidine-sensitive L cells, and neuroblastomas N18 and NS20. Liver RNA from 129 mouse does not contain species hybridizing with pMAC-2.

DISCUSSION

The results reported here show the presence of a 6-kb-long sequence, repeated roughly a thousand times, in the mouse genome and terminated by direct repeats several hundreds base pairs long. At its 5′ end, the RNA transcript of this sequence starts into the 5′ repeat or downstream 3′ to it; it is colinear with the genomic DNA and its 3′ end contains, after the repeat, a poly(A) tail (Fig. 6). By these characteristics, the 6-kb sequence appears as a transposon-like or a retrovirus-like structure (21, 22). Yet, it is different from the two reiterated elements with retrovirus-like structures known in the mouse, VL30 (14–16) and IAP (17–19).

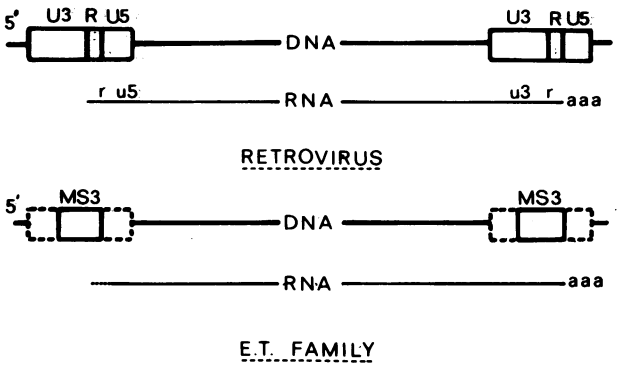

FIG. 6. Similarities of structure and transcription between ETn-like elements and retroviral elements.

Remarkably enough, the 6-kb sequence is transcribed in undifferentiated EC cells but not in a variety of differentiated cell types, in contrast with IAP RNA, which is found in abundance in various cell types, trophoblastoma, parietal endoderm, myoblasts, fibroblasts, and EC cells. We propose to call the 6-kb family an early transposon-like, or ETn-like, element. If the differential transcription of this element observed in cell lines is also found in early embryos, as expected, then it should provide a useful marker specific for ICM cells in blastocyst differentiation.

It is not yet known whether ETn-like RNA is translated into protein and associated with virus-like particles in the way that expression of the IAP family seems to be associated with intracisternal A-type particles (17). In addition to A-type particles, another virus-like structure, the ε particle, has been observed in electron micrographs of early mouse embryos. These ε particles are abundant at the two-cell stage, decrease thereafter, and disappear after the eight-cell stage (23–25). They have not been found in somatic tissues of other stages nor in cell lines. This distribution is in some way reminiscent of that found with ETn-like RNA.

We wish to thank Ms. D. Boullier for technical assistance, Drs. A. Minty and M. Buckingham for the gift of actin probe, Dr. J. Périès for the genomic IAP probe, Drs. R. Ollo and F. Rougeon for the genomic DNA library in λ, Drs. M. Getz and E. Kuff for the VL30-LTR and IAP-LTR nucleotide sequences prior to publication, and Dr. M. Katinka for advice about DNA sequence analysis. This work was supported by grants from the Centre National de la Recherche Scientifique (ATP 955133 and LA 269), the Foundation pour la Recherche Médicale Française, the Ligue Nationale Française contre le Cancer, and the Fondation André Meyer.

1. Brûlet, P., Babinet, C., Kemler, R. & Jacob, F. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 4113–4117.
2. Brûlet, P. & Jacob, F. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2328–2332.
3. Kemler, R., Brûlet, P., Schnebelen, M. T., Gaillard, J. & Jacob, F. (1981) *J. Embryol. Exp. Morphol.* **64**, 45–60.
4. Darmon, M., Bottenstein, J. & Sato, G. (1981) *Dev. Biol.* **85**, 463–473.
5. Strickland, S. & Mahdavi, V. (1978) *Cell* **15**, 393–403.
6. Vimard, C., Jeantet, C., Netter, Y. & Gros, F. (1976) *Biochimie* **58**, 473–478.
7. Brûlet, P. & Roskam, W. (1983) *Gene* **22**, 59–68.
8. Blin, N. & Stafford, D. W. (1976) *Nucleic Acids Res.* **3**, 2303–2308.
9. Hanahan, D. & Meselson, M. (1980) *Gene* **10**, 63–67.
10. Maniatis, T., Fritsch, E. E. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).

11. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
12. Homes, D. S., Cohn, R. H., Kedes, L. & Davidson, N. (1977) *Biochemistry* **16**, 1504–1512.
13. Kaback, D. B., Angerer, L. M. & Davidson, N. (1979) *Nucleic Acids Res.* **6**, 2499–2517.
14. Howk, R. S., Troxler, D. H., Lowy, D., Duesberg, P. H. & Scolnick, E. M. (1978) *J. Virol.* **25**, 115–123.
15. Keshet, E., Shaul, Y., Kaminchik, J. & Aviv, H. (1980) *Cell* **20**, 431–439.
16. Foster, D. N., Schmidt, L. J., Hodgson, C. P., Moses, H. L. & Getz, M. J. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7317–7321.
17. Kuff, E. L., Smith, L. A. & Lueders, K. K. (1981) *Mol. Cell. Biol.* **1**, 216–227.
18. Cole, M. D., Ono, M. & Huang, R. C. (1981) *J. Virol.* **38**, 680–687.
19. Shen-Ong, G. L. C. & Cole, M. D. (1982) *J. Virol.* **42**, 411–421.
20. Kuff, E. L., Feenstra, A., Lueders, K., Smith, L., Hawley, R., Hozumi, N. & Shulman, M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1992–1996.
21. Temin, H. M. (1981) *Cell* **27**, 1–3.
22. Temin, H. M. (1982) *Cell* **28**, 3–5.
23. Chase, D. G. & Piko, L. (1973) *J. Natl. Cancer Inst.* **51**, 1971–1975.
24. Yotsuyanagi, Y. & Szöllösi, D. (1981) *J. Natl. Cancer Inst.* **67**, 677–685.
25. Huang, T. T. F., Jr., & Calarco, P. G. (1981) *Dev. Biol.* **82**, 388–392.