Volume 17 Number 15 1989 Nucleic Acids Research

Identification of new human repetitive sequences: characterization of the corresponding cDNAs and their expression in embryonal carcinoma cells

Girolama La Mantia*, Gina Pengue, Domenico Maglione, Antonio Pannuti, Anna Pascucci and Luigi Lania

Dipartimento di Genetica, Biologia Generale e Molecolare, University of Naples, Via Mezzocannone 8, 80134 Naples, Italy

Received June 8, 1989; Accepted July 6, 1989 EMBL accession nos X15673–X15675 (incl.)

ABSTRACT

We have identified new repeated interspersed DNA sequences by analysis of homologous RNA transcripts from a human teratocarcinoma cell line (NTERA-2 clone D1). The abundance of transcripts varies upon retinoic acid induced differentiation of NTERA-2/D1 cells, and it is highest when the cells display the embryonal carcinoma phenotype. The expression of these novel repeated sequences appears to be tissue specific as no detectable expression was found in various cell lines of different embryological derivation. Characterization of the RNA transcripts by analysis of recombinant cDNA clones indicated that transcripts of different genomic units are present in undifferentiated embryonal teratocarcinoma cells. Nucleotide sequencing of the cloned cDNAs reveals a complex structure composed by unique and tandemly repeated sub-elements.

INTRODUCTION

Several families of dispersed, moderately repeated sequences have been identified in mammalian genome (1,2). SINE sequences are defined as short (70–300 bp) repetitive elements present in over 100 000 copies per genome. The best studied of all short repeated families in humans is the Alu family, accounting for the 3–6% of the genome (3).Alu elements appear to be retropseudogenes derived from RNA Pol III transcripts (4,5), as they contain a functional internal promoter (6).

Only two families of long, interspersed repetitive sequences (LINE) have been found in human genome, the L1 and the THE-1 families. The THE-1 repeats, present in about 10.000 copies per genome, show features of retrotransposons. The prototype 2.3 Kb element contains two flanking terminal repeats 350 bp long, generates 5 bp duplication of the target site, and a 2 Kb polyadenylated RNA homologous to this element has been found in Hela cells (7). L1 family includes about 20 000 copies of a 6–7 Kb long interspersed element, and more than 80% of the family members are truncated and internally rearranged (8,9). The structural properties of L1 units support the possibility that they are processed retrogenes. Analysis of their open-reading frame suggest that they may encode for proteins homologous to reverse transcriptase of retroviruses (10–13). The L1 elements terminate at the 3′ end with an A-rich tract, and flanking direct repeats representing target site duplications have often been found (14). Heterogeneous transcripts, reflecting probably a readtrough transcription from unrelated promoters, can be found in the nucleus of various cell types (15). However, in the human embryonal carcinoma cell line (NTERA-2/D1) the putative retroposition intermediate, a 6.5 Kb polyadenylated transcript has been found (16).

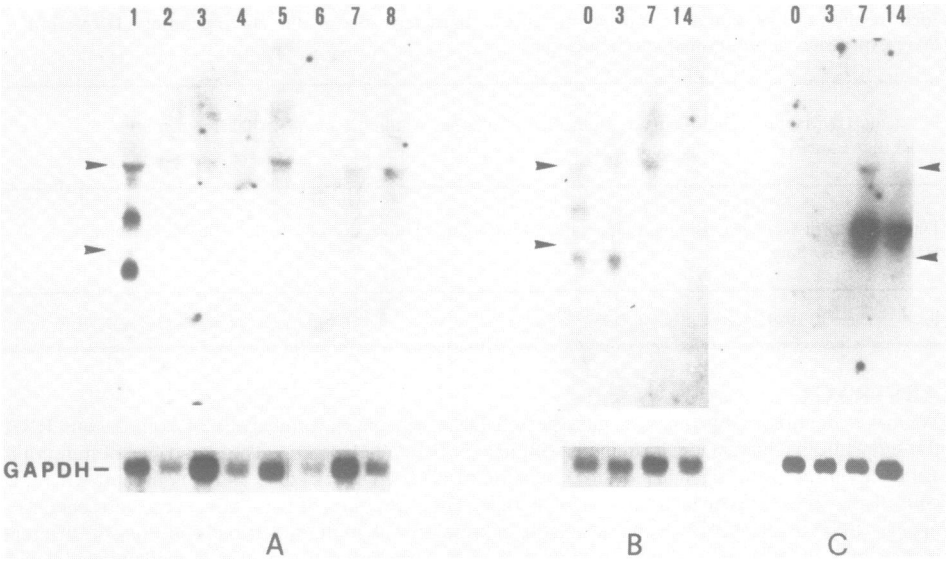


Figure 1.
Panel A) Northern blot analysis of total RNA isolated from various human cell lines hybridized to a probe containing three 85bp repeats. The RNA samples are: 1, NTERA-2/D1 (teratocarcinoma); 2, HeLa (cervical carcinoma); 3, HL60 (promyelocitic); WI-38 (lung fibroblast); 4, PEER (T-cell lynphoma); 5, NAMALVA (B-cell); 6, MOLT-4 (T-cell lymphoma); 7, CRL1220 (normal fibroblast); 8, HTB10 (neuroblastoma). Fifteen micrograms aliquots were electrophoresed and processed as described in the text, the blot was exposed at −80°C with an intensifying screen for 5 days. In overexposed autoradiograms, some hybridization to 28S RNA is detectable. Ribosomal RNA (28S and 18S) are indicated by arrowheads. Re-hybridization of the same blot to GAPDH (24) is shown at the bottom.
Panel B and C) Expression of repeated sequences (B) and HOX-3 (C) in undifferentiated (0) and differentiated (3-7-14) NTERA-2/D1 teratocarcinoma cells. NTERA-2/D1 cells were treated with retinoic acid and RNA extracted at day 3, day 7 and day 14. The same blots were re-hybridized to the GAPDH probe. RNA extraction and Northern blot were performed as described in the text. Ribosomal RNAs (28S and 18S) are indicated by arrowheads.

We have identified cDNA clones homologous to sequences that are members of a novel human repetitive family. Transcriptional studies indicate that the expression of the repetitive sequences is easily detectable as discrete sizes of polyadenylated cytoplasmic RNA in embryonal carcinoma cells, and their expression is negatively regulated during in vitro induced differentiation. The expression of these novel repeated elements appear to be restricted in the embryonal carcinoma cells since we fail to observe detectable expression in more highly differentiated cell types. Characterization of the transcripts by analysis of cloned cDNA, revealed a peculiar structure consisting of a complex array of different repeated elements and unique sequences.

MATERIALS AND METHODS

cDNA library screening and DNA sequencing

A cDNA library from undifferentiated human teratocarcinoma cell line NTERA-2/D1 cytoplasmic poly(A)$^+$RNA (17) was kindly provided by J. Skowronski. 3×10^5 plaques were screened with a 350 bp DNA fragment containing three 85bp repeat. This fragment was originally found in the 5′-end portion of a cDNA (pR17) encoding finger domains.

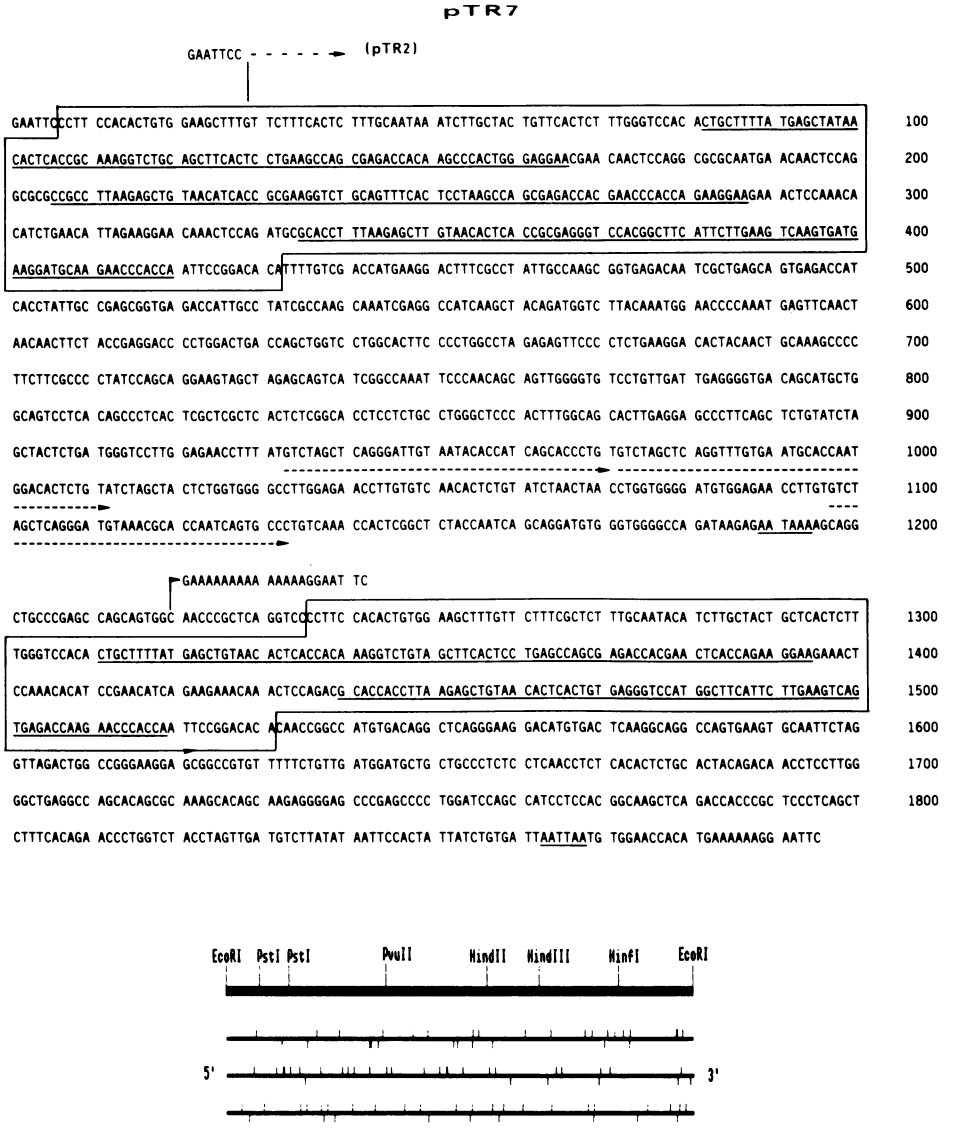
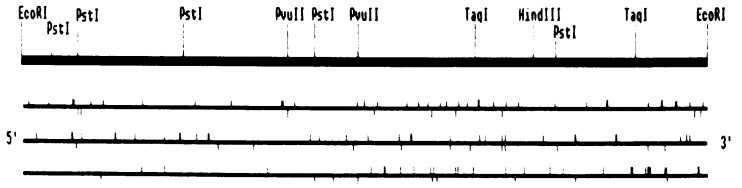


Figure 2. Nucleotide sequences of pTR7 and pTR2 cDNA clones. The sequence of pTR2 is identical to pTR7 from nt. 29 to nt.1228. The putative polyadenylation signals are underlined. The 400 bp repetitive sequences are boxed, and the internal 85bp repeats are underlined. The position and direction of the 41 bp repeats are indicated by a dotted lines. The restriction map of relevant enzymes is indicated at the bottom. The three frames of the pTR7 sequence are represented as horizontal lines; vertical lines above and below indicate translational stop codons and ATG codons respectively.

DNAs from positive clones were digested with endonuclease EcoRI and the resulting fragments were cloned into pGEM-3 plasmid (Promega, Biotech) for restriction enzyme mapping. Appropriate fragments were subcloned in the same vector for DNA sequencing.

pTR5

```
GAATTCGAA GCTTTGTTCT TTAGCTCTTT GCAATAAATC TTGCTGCTGC TCACTCTTTG GGTCCACACT GCCTTTACGA GCTGTAACAC TCACCGCGAA   100
GGTCTGCAGC TTCACTCCTG AAGCCAGTGA GACCACGAAC CCACCAGGAG GAACAAACAA CTCCAGATGT GCTGCCTTAA GAGCTGTAAC ACTCACCGCG   200
AAGGTCTGCA GCTTCAGTCC TAAGCCAGCA AGACCACGAA CCCACCAGAA GGAAGAAACT CTGAACACAT CCAAACATCA GAAGGAACAA ACTCTGGACA   300
TGCCGCCTTT AAGAACTGTG ACACTCACCA CGAGGGTCCG TGGCTTCATT CTTGAAGTCA GTGAGACCAA GAACCCACCA ATTCCAGACA CATTTTGGCA   400
ACCACAAAGG GACTATCACC TATTGCCAAG CGGTGAGACT ATCGCCAAGT GGTGAGACCA TCGTCTATCG CTGAGCAGCA AAACAATCGC CTATCGCCAA   500
GCAAATCGAA GCCATCAAGC TACAGATGGT CTTACAAATG GAACCCCAAA TGAGTTCAAC TAACAACTTC TACCGAGGAC CCCTGGACCG ACCCGCTGGC   600
ACTTCCACTG GCCTAGAGAG TTCCCCCTTG AGACACTACA ACTGCAGGGC CCCTTCTTCA CTTCTATCCA GCAGGAAGTA GCTAGAGCAG TCATCGGCCA   700
AATTCCCAAC AGCAGTTGGG GTGTCCTGTT TAGAGGGGGG ATTGAGAGGT GGCAGTGTGC TGGCAGCCCT CGCAGCCCTC GCTCACTCTT GGCTCCTCCT   800
CGGCCTTGGC GCCCACTCTG GCCATGCTTA AGGAGCCCTT CAGCCTGCCA CTGCACTGTG GGAACACTGG CCAAGGCCAG AGCCAGCTCA CTCAGCTTGT   900
GGGGAAGTGT GGAGGGAGAG GCGCAGGCGG GAACTGGGGC TCGATGCGGC GCTTGCGGGC CAGCGTGAGT TCCGGTGGTT GGTGGCTTGG CGGGCCCCAG  1000
ACTCGGAGCG GCTGGCCTGC CCTGCTGGCC CCAGGCAGTG AGGGGCTTAG CACCCAGGCC AGCAGCTGTG GATGGTGCGC TGAGTCTCCC AGCAGTGCTG  1100
GCCCACCCCA CCAGCACTGC ACTCGATTTC TCGCCAGGCC TTAGCTGCCT CCCCGCAGGC AGGCCTTGGG ACCTGCAGCC CACCATGCCT GAGTCTCCCC  1200
TGCTGCCGCC GTGGGCTCCT GCGTGGCCCA AGCCTCCCCG ACGAGCACTG CCCCCTGCTC CCCGGTGCCT GGTCCCATCG ACCCCAAGGG CTGAGGAGTG  1300
CAGGTGCATG CTGCGGGACT GGCAGGCAGC TCCACCTGCG GCCCCATGCA GGAAGCCAGC TGGGCTCCTG AGTCTGGTGG GGACTTGGAG AATTTTTATG  1400
TCTAGCTAAG GGATTGTAAA TACACCAATC AGCACTCTGT ATCTAGCTCA AGGATTGTAA ATACACCAAT CAGCACCCTG TGTCTAGCTC AGGGTTTGTC  1500
AATGCACCAA TCGGCACTCT GTATCTAGTT AATCTGGTGG GGAGTTGGAG AATCTTTATG TCTAGCTAAG GGATTGTAAA TACACCAATC AGCACTCTGT  1600
ATCTAGCTCA AGGTTTGTAA ACACACCAAT CAGAACCCTG TGTCTAGCTC AGGGTTTGTG AATGCACCAA TCAGCACTCT GTATCTAGTT AATCTGGTGG  1700
GGACTTGGAG AATCTTTATG TCTAGCTAAG GGATTGTGAA TGCACCAATC GGCACTCTGT ATCTAGCTCA AGGTTGTAAA TGGACCAATC AGCACTCTAT  1800
GTCTAGCTCA GTGTTTGTAA ATACACCAAT CGACACTCTA TCTAGCTAAT CTAGTGGGGA CGTGGAGAAC TTTTGTGTCT AACTCAGGGA TTGTAAATGC  1900
ACCAATCAGA ACCCTGTCAA AATGGACCAA TAGCTCTCTG TAAAACAGAC TGACTTTCTG TAAAATGGAC CAATCAGCAG GATGTGGGTG GGGCCAGATA  2000
AGAGAAGAAA AGCAGGCTGC CTGAGCCAGC AGTGGCAACC CGCTTGGGTC CCCTTCCACA CTGTGGAAGC TTTGTTCTTT CGCTCTTTGC AATAAATCTT  2100
GCTGCTGCTC ACTCTTTGGG TCCACACTGC CTTTATGAGC TGTAACACTC ACCGCGAAGG TCTGCAGCTT CATCTGAAGC AGCGAGACAC GAACCCACCA  2200
GGAGGAACAA ACAACTCCGA TGCGCTGCCT TAAGAGCTGT AACACCGCGA GGTCTCGAGC TTCACTCCTG AGCCAGCAAG ACCACAAACC CACCAGAAGG  2300
AAGAAACTCA GGAAAACAGC CGAACATCAG AAGGAATAAA CTCTGGACAC ACCACCTTTA AGAACTGTGA CACTCGCTGC GAGCGTCTGT GGCTTCATTC  2400
TTGAAGTCAG TGAGACCAAG AACCCACCAA TTCTGGACAC AATTGCACTC TAGCCTGGGC AACAAGAGTG AAACTCATCT CGAAAAAAAA AAGAGTCTGG  2500
AAGGGCCTTC TGAGTTGTAA TATTATGAAC TTAAATTTTG AGGTCCTGAA GTTTTGTTGC AGTGTGAATG ATAAGAGTAG TCTTTTTTTG ATATCATCTG  2600
TAGAAGACCC AATTTTGGGG TTTCAGATTA GGAAGGGTTT GATGGTTTTT AGTGGACTAT GAAAAGGTTT TTTTTGGTGA AAATATATTT TGGTATAATG  2700
TACTAAAGTC TTGTAGTATT TAGTCATAAT AGAGTTTAAC AGTGGAAGGG AATTC
```



The DNA sequences were independently determined by two different methods using the GemSeq K/RT system as recommended by Promega Biotech. Nucleotide sequencing data were analyzed using the MicroGenie Sequence Analysis Program software (18).Cell cultures Cell lines used were obtained from the American Type Cell Collection. NTERA-2, clone D1 (NT2/D1) cells (19) were kindly provided by E. Boncinelli. Stock cultures, predominantly exhibiting a human embryonal carcinoma phenotype, were mantained at high density in Dulbecco's modified MEM supplemented with 10% fetal calf serum. Retinoic acid induced differentiation was performed as described (20). Retinoic acid was added to a final concentration of 10 μM. Total RNA was isolated at various time after induction.

Southern and Northern analysis

Total RNA was extracted from the various cell lines by the guanidine isothiocyanate method (21). Fifteen micrograms aliquots were electrophoresed on 1.2% agarose-formaldheyde gels and transferred to nitrocellulose filters. The blots were hybridized (50% formamide, 5× SSPE, 5× Denhardt's, 0.1% SDS and 100 μg/ml salmon sperm DNA). The filters were washed at 60°C in 0.2× SSC and 0.2% SDS for 2 hrs. The DNA probes were labelled by the multiprime labelling kit (Amersham) to a specific activity of 5×10^8 c.p.m./μg.

RESULTS

Expression of repeated sequences is regulated in NTERA-2/D1 cells.

In the course of characterizing human cDNA clones encoding multi-finger proteins (22), we identified in the 5′ terminal portion of a clone, pR17, sequences containing three 85 bp direct adjacent repeats. Southern blot analysis revealed that these sequences are higly repeated in human genome, and no homology between the 85bp repeat and known SINE and LINE sequences was found (see below).

To study the expression of these sequences we analyzed total RNA from various human cell lines by Northern blotting (Fig 1, panel A), using as probe a 350 bp fragment containing three 85 bp direct adjacent repeats. Two bands of discrete size (1.8Kb and 2.8Kb) are visible only in the RNA derived from NT2/D1 cell line, which is a human pluripotent teratocarcinoma cell line, representative of a very early stage of the embryonic development (23). We failed to observe detectable expression in more highly differentiated cell types (Fig 1A). Northern blot of nuclear and cytoplasmic polyadenylated RNA, indicated that the 1.8kb and 2.8kb RNA are polyadenylated and cytoplasmic (data not shown, and see below).

Treatment of NT2/D1 cells with retinoic acid results in the loss of undifferentiated phenotype. Cells go through differentiation into specific cell types, including neurons (19). We examined total RNA from NT2/D1 cells at 3−7 and 14 days after exposure to retinoic acid. Fig 1 B shows that the relative amount of the two transcripts decreased upon retinoic acid treatment. After 3 days the transcript of 2.8Kb disappeared, while the 1.8 kb transcript become undetectable after 7 days, showing that the two transcripts are differentially regulated.

The efficiency of differentiation upon retinoic acid treatment was monitored after probing a similar Northern blot with HOX-3, a human homeobox gene, whose expression was

Figure 3. Nucleotide sequence of pTR5 cDNAS clones. The boxed areas indicate the large tandem repeat, containing three 85 bp repeats (underlined). The 41 bp repeats are underlined with dotted lines. The restriction map of relevant enzymes is indicated at the bottom. The three frames of the pTR5 sequence are represented as horizontal lines; vertical lines above and below indicate translational stop codons and ATG codons respectively.

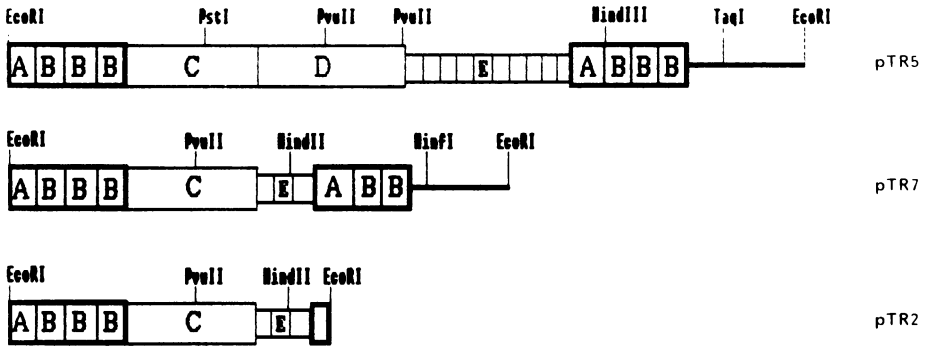


Figure 4. Schematic representation of the structure of the pTR5, pTR7 and pTR2 cDNA clones. The solid bar areas indicated the large tandem duplication containing both A and B elements. The C, D and E elements are described in the text. Thin lines indicate unique sequences.

previously shown to increase during *in vitro* induced differentiation (25).

Isolation and characterization of cDNA clones from NTERA-2/D1 cells

To characterize the observed transcripts, we screened a cDNA library derived from polyadenylated cytoplasmic RNA from NT2/D1 cells (17), using a fragment containing the 85 bp repeated sequences as a probe. About 5O phages out of 3×10^5 annealed to the probe, and 10 of them were selected for further analysis. The size of the cDNA inserts varied from about 1 kb to 3 kb. Restriction map analysis and partial DNA sequencing showed that five of them represent different transcripts.

The restriction map and the entire nucleotide sequence of the three longest clones are shown in Fig.2 and 3. In Fig. 2 the DNA sequence of the clones pTR7 and pTR2 are shown. The pTR2 is identical to pTR7 throughout from nt. 29 to nt. 1228; in this position a poly A tail is present in pTR2 30 bp downstream from a canonical polyadenylation site. A short poly A tail is also present in pTR7 with a non canonical polyadenylation site (AATTAA) located 20 bp upstream. Thus, the pTR2 and pTR7 clones can represent different transcripts of the same genomic unit by alternative usage of the polyadenylation sites; conversely they are different transcripts of partially duplicated genomic units. The pTR5 is 2755bp long and lacks a poly A tail. Conceptual translation of cDNA clones sequences revealed that no open reading frames longer than 105 and 187 codons are present in pTR7 and pTR5 respectively. The distribution of stop and ATG codons in the three reading frames is shown under the nucleotide sequence in Fig.2 and 3. Moreover we could not find any significant homology with retrovirus sequences.

A schematic representation of the three cDNA clones is shown in Fig 4. Sequence comparison of the cDNA clones shows common features that are most easily described in relation to the longest cDNA pTR5. Both pTR5 and pTR7 have a tandem repeat of about 400bp (boxed sequences in Fig.2 and 3), and in each repeat two different elements (A and B) can be identified. The 5′ end of all three clones lies in the first A element. Element B is a 85 bp long repeat similar to the 85bp repeat originally found in the finger-coding cDNA pR17. The B element is present 6 times in pTR5, 5 times in pTR7 and 3 times in pTR2; pairwise comparison of the 85 bp repeats show that the B element sequences are 85−88% identical to each other and a consensus sequence has been derived as shown in Fig. 5.

```
pTR5  B1    69..CTGCCTTTACGAGCTGTAACACTCACCGCGAAGGTCTGCAGCTTCACTCCTGAAGCCAGTGAGACCACGAACCCACCAGGAGGAA...153
      B2   171..CTGCCTT-AAGAGCTGTAACACTCACCGCGAAGGTCTGCAGCTTCAGTCCT-AAGCCAGCAAGACCACGAACCCACCAGGAGGAA...254
      B3   300..ATGCCTTTAAGAACTGTGACACTCACCACGAGGGTCCGTGGCTTCATTCTTGAAGTCAGTGAGACCAAGAACCCACCA.........380
                    /\
                   GCC
      B4  2127..CTGCCTTTATGAGCTGTAACACTCACCGCGAAGGTCTGCAGCTTCATCT-TGAAG-CAGCGAGAC-ACGAACCCACCAGGAGGAA..2207
      B5  2225..CTGCCTTTAAGAGCTGTAACA----CCGCGA-GGTCTCGAGCTTCACTCCTGA-GCCAGCAAGACCACAAACCCACCAGAAGGAA..2302
      B6  2353...CACCTTTAAGAACTGTGACACTCGCTGCGAGCGTCTGTGGCTTCATTCTTGAAGTCAGTGAGACCAAGAACCCACCA.........2429

pTR7  B1    82..CTGCTTTTATGAGCTATAACACTCACCGCAAAGGTCTGCAGCTTCACTCCTGAAGCCAGCGAGACCACAAGCCCACTGGGAGGAA...166
      B2   206..CCGCCTT-AAGAGCTGTAACATCCACCGCGAAGGTCTGCAGTTTCACTCCT-AAGCCAGCGAGACCACGAACCCACCAGAAGGAA...287
      B3   334..GCACCTTTAAGAGCTGTAACACTCACCGCGAGGGTCCACGGCTTCATTCTTGAAGTCAA-GTGAGCAAGAACCCACCA.........420
                         /\                                                  /\
                         T                                                   TGAAGGAT
      B4  1311..CTGCTTTTATGAGCTGTAACACTCACCACAAAGGTCTGTAGCTTCACTCCTGAGGCCAGCGAGACCACGAACTCACCAGAAGGAA..1394
      B5  1440..GCACCCTTAAGAGCTGTAACACTCACTGTGAGGGTCCATGGCTTCATTCTTGAAGTCAGTGAGACCAAGAACCCACCA.........1519
                    /\
                    AC

pR17  B1    117..CTGCTTTTACCAGCTGTAACACTAACCGCAAACGTCTGCAGCTTCACTCCTGAAGCCAGCGAGACCACGAGCCCACCGGGAGGAA...201
      B2    220..CCGCCTT-AAGAGCTGTAACACTCACCACGAAGGTCTGCAGCTTCACTCCTGA-GCCAGCGAGACCACGAACCCGCCAGAAGGAA...310
                       /\
                      CCTTA
      B3    354..GCGCCTT-AAGAGCTGTAACACTCACCACAAGGGTCCGCGGCTTCATTCTTGAAGTCAGTGAGACCAAGAACCCACCA.........444
                    /\
                    ACC
Consensus         CTGCCTTTAAGAGCTGTAACACTCACCGCGAGGGTCTGCAGCTTCACTCCTGAAGCCAGCGAGACCACGAACCCACCAGGAGGAA
```

Figure 5. Comparison of the 85 bp repeats present in various cDNA clones. A consensus sequence is derived from the various 85 bp repeat (B elements) underlined in Fig 2 and 3 and from the 85 bp repeats originally found in pR17 clone.

The C element (from nt. 398 to nt. 822) is present in all three cDNA with a similarity of 85%. The D elememt (nt.823 to nt 1400) is absent in pTR7, but hybridization analysis suggest its presence in other cDNA clones (data not shown). The E element (from nt. 1401 to nt. 1914) contains a tandem array of a 41 bp repeat previously described as a member of a highly repeated non-mobile family (26). The 41 bp unit is repeated 10 times in pTR5 and 3 times in both pTR2 and pTR7. Finally no similarity was found in the 3′-end portion of pTR5 and pTR7 as indicated in Fig. 4 by a thin line.

Searching for related sequences using the GenBank and EMBL Data Base failed to reveal homologous sequences other than a short sequence of 70 bp found associated to the 41 bp repeat in a region of the human genome 3′ to the β-globin gene cluster, where the presence of members of different repeated families has been described (27). This sequence is present in pTR5 at nt. 1970−2040, in pTR7 at nt. 1154−1224 and in pTR2 at nt. 1124−1194.

Genomic abundance of different elements

To determine the genomic abundance of the different repeated subelements, we performed Southern blot experiments, using HeLa cell DNA digested with different restriction enzymes. We used as probes various fragments of pTR5 and pTR7 clones representing the different subunits (Fig 6). No significative difference in the hybridization patterns can be seen with the probes representative of the different sub-elements, all probes giving a smear throughout the lanes. Dot blot and genomic library screening hybridizations indicate that there are about 20.000 copies of each sub-elements in human genome. Moreover, the cDNA clones anneal in situ to many scattered chromosomal locations (data not shown)

When the 3′ portion of pTR7 was used as probe (Hinf-EcoR1 fragment, Fig. 6 probe 5), we obtained hybridization to unique fragments, moreover we noted a hybridization

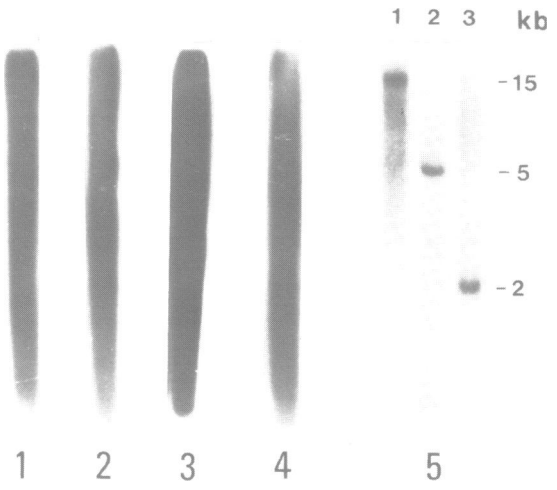


Figure 6. Southern blot analysis of HeLa cell DNA. The probe 1 is a 206 bp EcoRI-PstI fragment from pTR5 and it is specific for the long tandem repeat. Probe 2 is 421 bp PstI-PvuII of pTR5 (nt. 642 to nt. 1063) and it contains sequences of the C and D regions. Probe 3 is a 294 bp PvuII-PvuII (nt. 1063 to nt. 1357 of pTR5) and it is specific for the D element. Probe 4 is a 473 bp PvuII-TaqI (nt. 1357 to nt. 1830) specific for the region E. Probe 5 is a 322 bp HinfI-EcoRI fragment pf pTR7 (nt. 1568–1890). The DNA hybridized with probes 1,2,3,4, has been digested with EcoRI. The DNA hybridized with probe 5 has been digested with EcoRI (lane 1), PstI (lane 2) and HindIII (lane 3).

background suggesting the presence of related sequences in the human genome. Similar results were obtained using a 3'-end probe from pTR5 (data not shown). Thus, in both pTR5 and pTR7 clones the results show that the transcript units are composed by a patchwork of different repeated subunits followed by a unique 3'-end portion.

DISCUSSION

It has been suggested that amplification and dispersal of repeated families members in higher eucaryotes might be explained by a model based on RNA-mediated transposition events (28). Many structural features of known repeated elements support this suggestion, even if no consistent structural similarities between different retrotransposons exist. The identification of repeated units transcripts may substantiate the hypothesis of a transit through a RNA intermediate.

We have identified a novel family of repetitive DNA in humans composed of distinct modules. Two discrete polyadenylated cytoplasmic transcripts of this repeated family are present in the human pluripotent embryonal carcinoma cell line NT2/D1. The expression of the repeated sequences appears to be tissue specific as no detectable expression was found in various cell lines of different embrylogical derivation. Furthermore, both RNA transcripts are negatively regulated upon *in vitro* induced differentiation of NT2/D1 cells. The human L1 transcripts are also negatively regulated during NT2/D1 differentiation. However, no structural similarity between L1 cDNA (17) and our cDNA clones was found.

Transcription of repeated sequences is often derepressed in undifferentiated cells (16,29). This fact is considered a support for the retroposition model, as amplification and transposition are expected to occur in germ line cells or in their progenitors (16).

Characterization of three cDNA clones from teratocarcinoma cells showed that transcripts of different genomic units are expressed in these cells. Conceptual translation of cDNA clones sequences revealed that no open reading frames longer than 105 and 187 codons are present in pTR7 and pTR5, respectively. The absence of long open reading frames in the sequence of cDNA clones cannot be considered conclusive, as other members of the polymorphic family may code for a functional product.

Characterization of the transcripts by nucleotide sequencing of cloned cDNA, revealed a peculiar structure consisting in a complex array of different repeated and unique elements. Direct tandem repeat of about 400bp have been found in pTR5 and pTR7 clones, and in each repeat two different elements (A and B) can be identified. Tandem repeats in the 5′ end region of some mouse (30) and primate (31) L1 elements have been described. Sequence comparison analysis failed to detect any significative similarity of our cDNA clones to both mouse and primate 5′ tandem repeats.

Different sub-elements can be associated in the genome with unrelated sequences. The 41 bp elements have been found in a region 3′-end of the human β-globin cluster, however no transcription of this region has been detected (26). Moreover we found that the relative genomic abundance of various subfragments are comparable in Hela cells DNA. The isolated cDNA clones can indeed represent transcripts of colinear genomic units; alternatively, genomic elements can have a more complex structure, due to the presence of intron sequences. Interestingly, the 3′ portion of cDNA clones appear to be a unique sequence. A possibility is that this sequence can represent the unique site of insertion of the mobile element. We are presently involved in the isolation and characterization of genomic sequences to elucidate the genomic structure of this new family of repeated sequences.

ACKNOWLEDGEMENTS

We would like to thank Dr. L. Lanfrancone for RNA preparations, Dr. J. Skowronsky for the cDNA library and Dr. E. Boncinelli for providing the NT2/D1 cell line. We also thank E. Boncinelli and P. Di Nocera for advice in the preparation of this manuscript. This work was supported by grants from M.P.I., from Progetto Finalizzato 'Oncologia' CNR, and by the Italian Association for Cancer Research (AIRC) to L.L.

*To whom correspondence should be addressed

REFERENCES

1. Jelinek, W.R. & Schmid, C.W. (1982) Ann.Rev.Biochem. 51, 813–844
2. Singer, M.F. (1982) Int. Rev. Cytol. 76, 67–112
3. Rinehart, F.P., Ritch, T.G., Deininger, P.L., Schimd, C.W. (1981) Biochemistry 20, 3003–3010.
4. Deininger, P.L. & Daniels, G.R. (1986) Trends Genet. 2, 76–78.
5. Ullu, E. & Tschudi, C. (1984) Nature 312, 171–172.
6. Weiner, A.M., Deininger, P.L. & Efstratiadis, A.(1986) Ann.Rev.Biochem. 55, 631–661.
7. Paulson, K.E., Deka, N., Schimd, C.W., Misra, R., Schindler, C.W., Rush, M.G., Kadyk, L. & Leinwand, L. (1985) Nature 316, 359–361.
8. Hattori, M., Hidaka, S. & Sakaki, Y. (1985) Nucleic Acids Res. 13, 7813–7817.
9. Grimaldi, G., Skowronski, J. & Singer, M.F. (1984) EMBO J. 3, 1753–1759.
10. Hattori, M., Kuhara, S., Takenaga, O. and Sakaki, Y. (1986) Nature 321, 625–628.
11. Skowronski, J. & Singer, M.F. (1986) Cold Spring Harbor Symp. Quant. Biol. 51, 457–464.
12. Sakaki,Y., Hattori,M., Fujita,A., Yioshioka, K.,Kuhara,S. & Takenaka O. (1986) Cold Spring Harbor Symp. Quant.Biol. 51, 465–469.
13. Doolittle, R.F., Feng, D.F. Johnson, M.S. & McClure M.A. (1989) The Quarterly Rev. of Biol. in press
14. Thayer, R.E. & Singer,M.F. (1983) Mol. Cell. Biol. 3, 967–973.

15. Kole, L.B., Haynes,S.R. & Jelinek, W.R. (1983) J. Mol. Biol. 165, 5739–5745.
16. Skowronski, J. & Singer, M.F. (1985) Proc. Natl. Acad. Sci. USA 82,6050–6054.
17. Skowronski, J., Fanning, T.G. & Singer, M.F. (1988) Mol. Cell. Biol. 3, 1385–1397.
18. Queen, C. & Korn, L.J. (1984) Nucleic Acids Res. 12,581–599.
19. Andrews, P.W. (1984) Dev. Biol. 103, 285–293.
20. Mavilio, F., Simeone, A., Boncinelli, E. & Andrews, P. W. (1988) Differentiation 37, 73–79.
21. Chirgwin, J.M., Przybyla,A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294–5299.
22. Pannuti, A., Lanfrancone, L., Pascucci, A., Pelicci, P.G., La Mantia, G. & Lania, L. (1988) Nucleic Acids Res. 16, 4227–4237.
23. Andrews, P. W., Damjanov, I., Simon, D., Banting, G., Dracopoli,N.C. & Fogh, J. (1984) Lab. Invest. 50, 147–162.
24. Piechaczyk, M., Blanchard, J.M., Marty, L., Dani, C., Panabieres, F.,El Sabouty, S., Fort, P. & Jeanteur, P. (1984) Nucleic Acids Res. 12, 6951–6963.
25. Simeone, A., Acampora, D., D'Esposito, M., Faiella, A., Pannese, M., Scotto, L., Montanucci, M., D'Alessandro, G., Mavilio, F. & Boncinelli, E. (1989) Mol. Reprod. and Dev. 1, 107–115.
26. Yang, R., Fristensky, B., Deutch, A. H., Huang, R. C., Tan,Y.H., Narag, S.A. & Wu, R. (1983) Gene 25, 59–66.
27. Henthorn, P.S., Mager, D. L., Huisman, T. H. J. & Smithies, O. (1986) Proc. Natl. Acad. Sci. 83, 5194–5198.
28. Rogers, J. (1985) Int. Rev. Cytol. 93, 187–279.
29. Bennet, K.L.H., Hill,R.E., Pietras, D. F., Woodworth-Gutai, M., Kane-Haas, C., Houston, I.M., Hearth, I.K. and Hastie, M.D. (1984) Mol. Cell. Biol. 4, 1561–1571.
30. Loeb, D.D., Padgett, R.W., Hardies, S.C., Shehee, W.R., Comer, M., Edgell, M.H. & Hutchison III, C.A. (1986) Mol. Cell. Biol. 6, 168–182.
31. Loyid, J.A. & Potter S.S. (1988) Nucleic Acids Res. 16, 6147–1156

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.