

Illegitimate transcription: Transcription of any gene in any cell type

(tissue-specific genes/gene expression/cDNA polymerase chain reaction)

JAMEL CHELLY, JEAN-PAUL CONCORDET, JEAN-CLAUDE KAPLAN, AND AXEL KAHN

Unité de Recherches en Génétique et Pathologie Moléculaires, Unité 129, Institut National de la Santé et de la Recherche Médicale, CHU Cochin, 24 Rue du Fg Saint-Jacques, 75014 Paris, France

Communicated by Jean Dausset, January 3, 1989

ABSTRACT Using *in vitro* amplification of cDNA by the polymerase chain reaction, we have detected spliced transcripts of various tissue-specific genes (genes for anti-Müllerian hormone, β -globin, aldolase A, and factor VIIIc) in human nonspecific cells, such as fibroblasts, hepatoma cells, and lymphoblasts. In rats, erythroid- and liver-type pyruvate kinase transcripts were also detected in brain, lung, and muscle. The abundance of these “illegitimate” transcripts is very low; yet, their existence and the possibility of amplifying them by the cDNA polymerase chain reaction provide a powerful tool to analyze pathological transcripts of any tissue-specific gene by using any accessible cell.

We have recently demonstrated that transcripts of the Duchenne muscular dystrophy (DMD) gene could be detected not only in muscle and brain as expected but also, at a very low level, in cultured fibroblasts, lymphoblasts, and Hep G2 hepatoma cell lines (1). Therefore, we asked whether this phenomenon—namely, low transcription of a tissue-specific gene in nonspecific cells—was particular to the DMD gene, or corresponded, in fact, to a general phenomenon of basal transcription of any gene in any cell type.

To answer this question, we examined the presence of transcripts of different tissue-specific genes in various human and rat nonspecific tissues and cultured cells.

Four human genes were chosen for this study: (i) the gene for the anti-Müllerian hormone (AMH), which is specifically expressed in embryonic testis Sertoli cells and is responsible for regression of Müllerian ducts in males (2–4); (ii) the gene for β -globin, which is specific to adult erythroid cells; (iii) the gene for factor VIIIc, which is mainly expressed in liver (deficiency of which is associated with the chromosome X-linked disease hemophilia A); and (iv) the gene for aldolase A, which possesses three optional promoters, one of them being exclusively active in adult skeletal muscle, while the others are ubiquitous (5). One rat gene was also chosen: the rat L-type pyruvate kinase gene, which possesses two alternative promoters, one specific to hepatocytes (L-type) and the other to erythroid cells (L'-type). The start sites of transcription controlled by these promoters define the 5' ends of two alternative coding first exons. L- and L'-type pyruvate kinase subunits differ, therefore, by their N-terminal ends (6, 7).

MATERIALS AND METHODS

Isolation of RNA. Total cellular RNA was extracted from three different human lymphoblastoid cell lines, from one strain of human fetal skin fibroblasts and Hep G2 hepatoma cells (8), and from various human and rat tissues by the method of Chirgwin *et al.* (9). Fibroblasts and Hep G2 cells

were harvested at confluency. Lymphoblasts were harvested in the exponential growth phase.

Oligonucleotide Primers. Oligonucleotide primers were synthesized according to sequences determined in our laboratory (5, 6) or published elsewhere (2). The oligonucleotide primers complementary and identical to mRNA sequences were chosen in different exons, as reported in Fig. 1.

Amplification by the cDNA Polymerase Chain Reaction (cDNA-PCR) (1). *Specific first-strand cDNA synthesis.* Ten to 20 μ g of total RNA was incubated at 42°C for 1 hr in 10 μ l of 50 mM Tris-HCl buffer (pH 8.3) containing 75 mM KCl, 3 mM MgCl₂, 10 mM dithiothreitol, and 5 pmol of the oligonucleotide primer complementary to the transcript and corresponding to the 3' part of the fragment to be amplified (see details in Fig. 1). The cDNA was synthesized by primer extension using 100 units of Moloney murine leukemia virus reverse transcriptase (BRL) per μ g of RNA incubated at 42°C for 2 hr in 100 μ l of the same buffer additionally containing 1 mM each dNTP (dATP, dCTP, dGTP, and dTTP), 50 units of placental ribonuclease inhibitor (RNAsin, Promega Biotech), 1 mM sodium pyrophosphate, and 50 μ g of bovine serum albumin per ml.

PCR. After NaOH hydrolysis of RNA and neutralization, the synthesized cDNA was coprecipitated with 0.25 to 0.5 μ g of the oligonucleotide primers, complementary and identical to the transcripts (see Fig. 1). cDNA and primers were resuspended in 50 μ l of *Thermus aquaticus* (Taq) DNA polymerase buffer: 16.6 mM ammonium sulfate/67 mM Tris-HCl, pH 8.8/6.7 mM MgCl₂/10 mM 2-mercaptoethanol/6.7 μ M EDTA/1 mM dNTP (dATP, dCTP, dGTP, and dTTP)/10% (vol/vol) dimethyl sulfoxide. After 10 min at 94°C and 2 min at room temperature, 2 units of Taq DNA polymerase (New England Biolabs) was added, and the second strand of the cDNA was synthesized for 5 min at 70°C; this step was followed by 25–30 cycles of amplification (denaturation, 1 min at 92°C; annealing, 1 min at 42°C; extension, 2 min at 70°C).

Analysis of the cDNA-PCR Amplified Products. The amplified products were separated by electrophoresis on 8–10% (wt/vol) polyacrylamide gels. After alkaline denaturation for 30 min by a 0.2 M NaOH/0.6 M NaCl solution and washing for 30 min with a 7% (vol/vol) formaldehyde solution, gels were blotted overnight onto nylon filter and then hybridized with specific probes.

AMH and L- and L'-type pyruvate kinase probes corresponding respectively to the first exon of the AMH gene and to the first common exon of the pyruvate kinase gene were labeled with [³²P]dCTP by random priming.

The β -globin probe was a 6-kilobase (kb) fragment of genomic DNA including the first two exons of the gene. The factor VIIIc probe was a 4.7-kb fragment of factor VIIIc

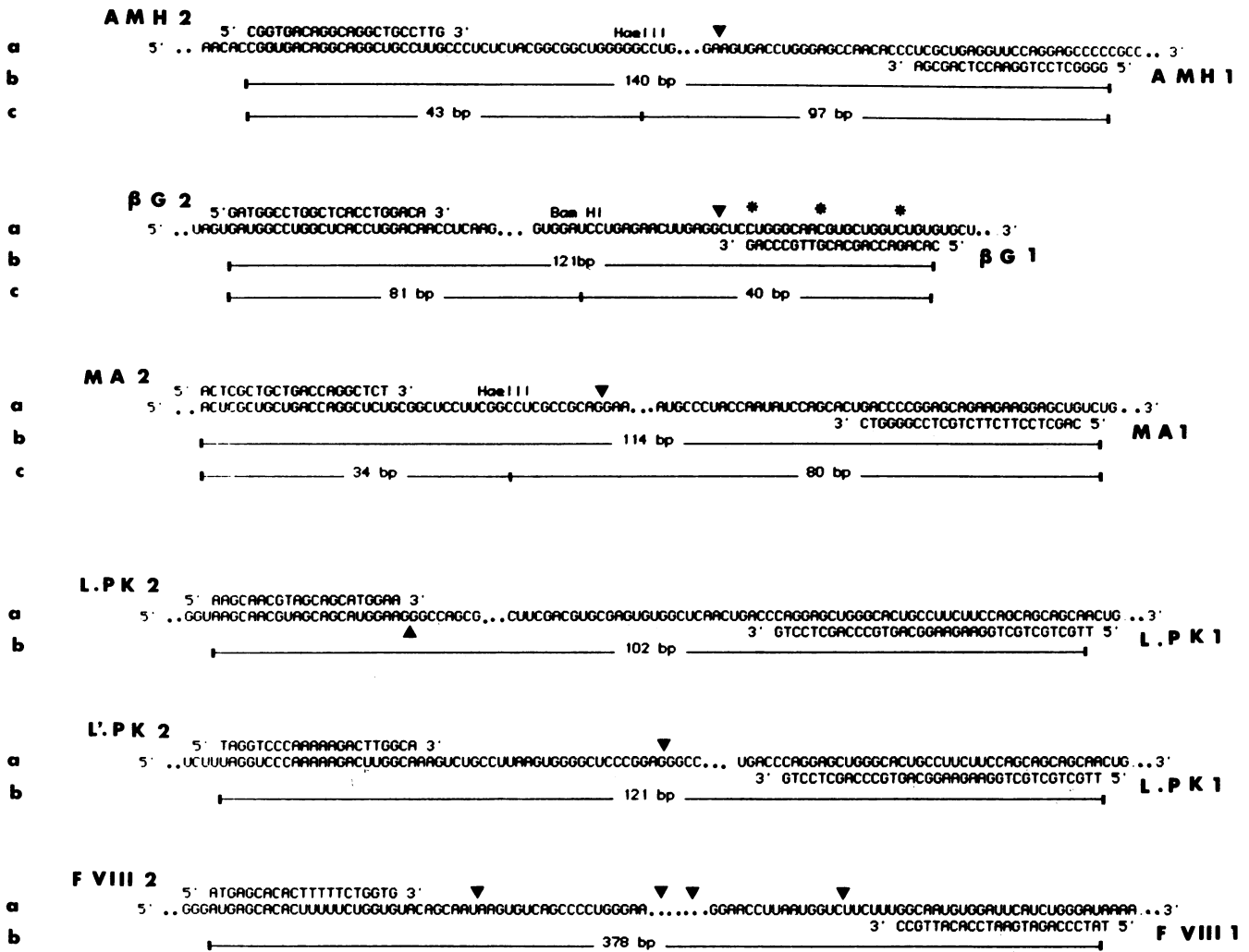


Fig. 1. Partial sequences of the different mRNAs analyzed by cDNA-PCR. Lines: a, sequence of synthetic oligonucleotide primers and their relation to the target mRNA region; b, size of the specific fragments amplified by cDNA-PCR; c, size of the fragments obtained after digestion of the amplified fragments with adequate restriction enzymes (indicated in line a). Arrowheads indicate the exon limits; asterisks indicate nucleotide differences between β -globin and δ -globin mRNA sequences. Primers AMH1, β G1, MA1, L.P.K1 (twice), and FVIII1 are complementary to a segment of the coding strands of, respectively: (i) the second exon of the AMH gene, (ii) the third exon of the β -globin gene, (iii) the first coding exon of the muscle-type aldolase A gene, (iv) the first L-type and L'-type common exon of the pyruvate kinase gene, and (v) exon 23 of the factor VIIIc gene. Primers AMH2, β G2, MA2, L-PK2, L'-PK2, and FVIII2 are identical to a segment of the coding strands of, respectively: (i) the first exon of the AMH gene, (ii) the second exon of the β -globin gene, (iii) the muscle-specific exon of the aldolase A gene, (iv) the first L-type exon of the pyruvate kinase gene, (v) the first L'-type exon of the pyruvate kinase gene, and (vi) exon 19 of the factor VIIIc gene.

cDNA encompassing exons 14–26 (gift of J. J. Toole, Genetics Institute). These probes were labeled by nick-translation.

The aldolase A probe, corresponding to the muscle-specific exon cloned in M13, was labeled by primer extension by using a 18-mer complementary to the 3' end of this exon.

Final wash of filters was performed at 65°C in 0.1% sodium dodecyl sulfate containing 0.2× SSC (1× SSC = 0.15 M NaCl/0.015 sodium citrate, pH 7) for at least 30 min. Autoradiographic exposure of filters was done for variable times at -70°C with an intensifying screen. The intensities of the bands were measured by scanning with a Shimadzu densitometer.

RESULTS

Detection of Human and Rat Tissue-Specific Transcripts in Nonspecific Cells or Tissues. The cDNA-PCR amplified products from human fibroblasts, lymphoblastoid cell lines, and hepatoma cells were subjected to blot hybridization analysis.

Figs. 2, 3, and 4 show that it was possible to amplify AMH, β -globin, and muscle-specific aldolase A transcript fragments from fibroblasts, lymphoblastoid cell lines, and hepatoma

cells. Factor VIIIc transcripts were also detected in lymphoblastoid cell lines (data not shown). No specific amplification occurred when the first step of cDNA synthesis was omitted or when DNA amplification and cDNA synthesis were performed using sets of primers corresponding to different genes. This showed that specific amplification was due to initial RNA templates and not to contaminating DNA fragments (Fig. 5).

For all three genes studied, the amount of amplified fragment [which is known to be proportional to the abundance of initial template (10)] was 3- to 10-fold higher when starting from lymphoblasts as compared with Hep G2 cells. It was 10- to 40-fold higher when starting from lymphoblasts as compared with fibroblasts. This amount, as expected, was maximal when starting from tissue known to express the corresponding gene (testis for AMH, fetal liver for β -globin, skeletal muscle for aldolase A, and liver for factor VIIIc). We were also able to amplify an important amount of β -globin cDNA fragment starting from fetal lung. This was probably due to the presence of blood reticulocytes and erythroblasts in fetal tissues.

Both liver (L-type)- and erythroid (L'-type)-specific pyruvate kinase transcripts could be amplified from various rat

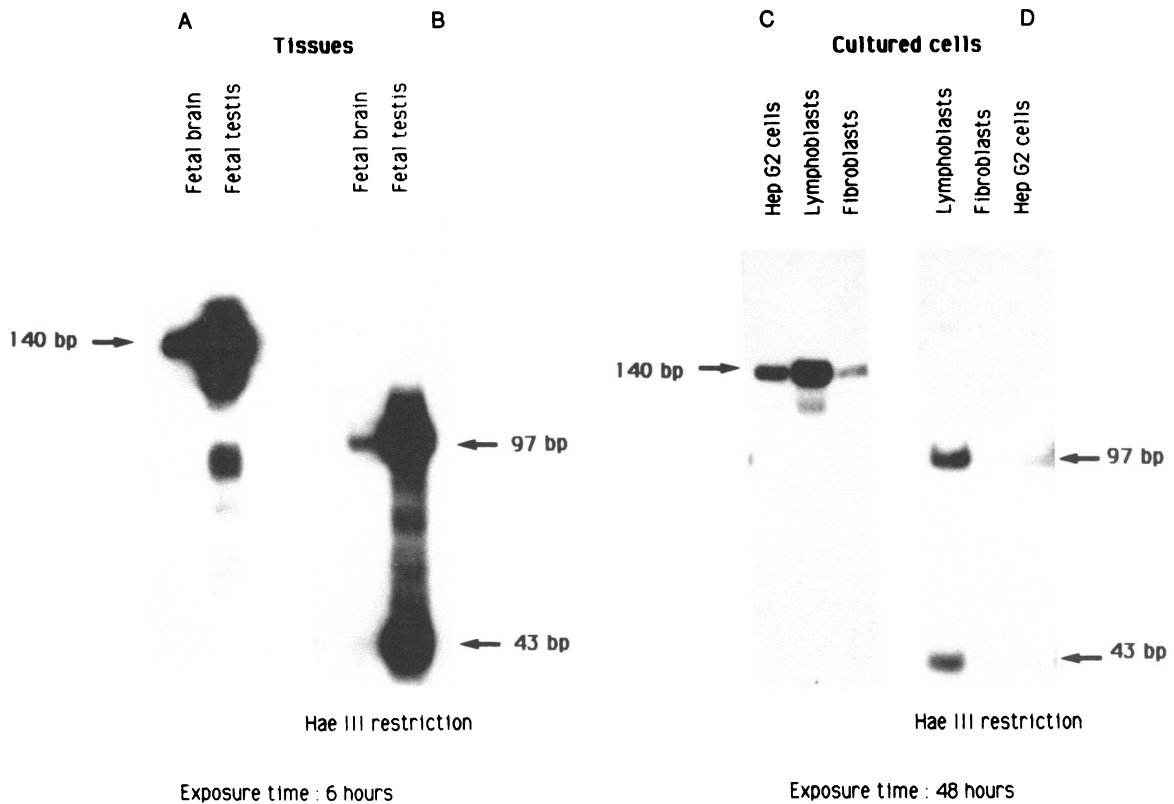


FIG. 2. Detection and characterization by cDNA-PCR of AMH transcripts in different cells. (A and C) Specific hybridization of AMH cDNA fragments amplified from RNAs extracted from various human tissues (A) and cultured cells (C). (B and D) Southern blots of the same cDNA-PCR products as in A and C after restriction by *Hae* III enzyme (the *Hae* III site is shown in Fig. 1). The arrows indicate the size of specific amplified fragments before and after enzymatic restriction. [The 43-base-pair (bp) band is present in fibroblasts and Hep G2 cells but is very faint and hardly visible on the picture.]

tissue RNA preparations. As expected, L-type pyruvate kinase mRNA was especially abundant in liver of carbohydrate-refed rats (11), and L'-type pyruvate kinase mRNA was abundant in fetal liver and spleen of rats treated with acetylphenylhydrazine (a drug that induces hemolytic anemia) (12). Nevertheless, the two types of cDNA fragments, complementary to either L- or L'-type transcripts, could also be amplified from muscle, brain, and lung RNAs (data not shown).

Amplified DNAs Derived from Spliced RNAs. The size of amplified fragments obtained by using primers from different exons corresponded to the expected size of spliced transcripts, indicating that the initial templates were bona fide spliced RNAs and not contaminating genomic DNA or heteronuclear RNA. The identity of amplified RNA-derived DNA fragments was assessed by specific hybridization with appropriate probes (Fig. 2 A and C, Fig. 3 A and C, and Fig. 4). It was further confirmed by restriction analysis. Digestion of amplified products by appropriate restriction enzymes produced fragments of the expected sizes determined by the cDNA sequences (Fig. 2 B and D and Fig. 3 B and D). In addition, the sequence of the amplified AMH cDNA fragment was checked by the Maxam and Gilbert method (13). One of the primers used for the amplification was 5'-end-labeled by polynucleotide kinase reaction. After polymerase chain reaction, the amplified radiolabeled fragment was eluted from the gel, and its sequence was determined. It corresponded exactly to the expected AMH cDNA sequence (2).

DISCUSSION

"Illegitimate Transcription." Tissue-specific and developmentally regulated expression of genes is the basic mechanism of development and differentiation in multicellular

eukaryotes. Therefore, genes can be classified into two groups: tissue-specific genes, which are exclusively expressed at a certain stage of development of certain tissues, and housekeeping genes, which are expressed in essentially all cells (14, 15). The first group encodes proteins involved in functional and phenotypic characteristics of cells; the second group encodes common structural proteins or ubiquitous enzymes. Thus, transcripts of tissue-specific disease genes are considered to exist only in the corresponding tissue—e.g., liver for a liver disease, muscle for a muscle disease, etc.

We show in this paper that, in fact, highly tissue-specific genes, such as those for AMH, factor VIIIc, and β -globin, are expressed as spliced transcripts in nonspecific tissues. The same phenomenon was observed for transcripts of genes that have alternative tissue-specific promoters (genes for aldolase A and L-type pyruvate kinase). These results, following the initial finding of low-level expression of the DMD gene in nonmuscle tissues (1), suggest that any gene may be transcribed at a very low level in any cell type. As a consequence, using the cDNA-PCR procedure, one could amplify DNA fragments complementary to specific gene transcripts by starting from RNAs extracted from any easily accessible cells.

We have previously shown in the case of DMD gene transcripts that fibroblasts and lymphoblasts contained less than one molecule of the specific spliced RNA per 500–1000 cells (1). The same seems to be true in the case of transcripts of the other tissue-specific genes analyzed here in nonspecific cells. It is unlikely that these very rare transcripts play any role. Rather, their presence in nonspecific cells would indicate a basal level of transcription of tissue-specific genes outside of the tissues where they are normally active. We call this low-level ubiquitous transcription of tissue-specific genes "illegitimate transcription." These results are in agree-

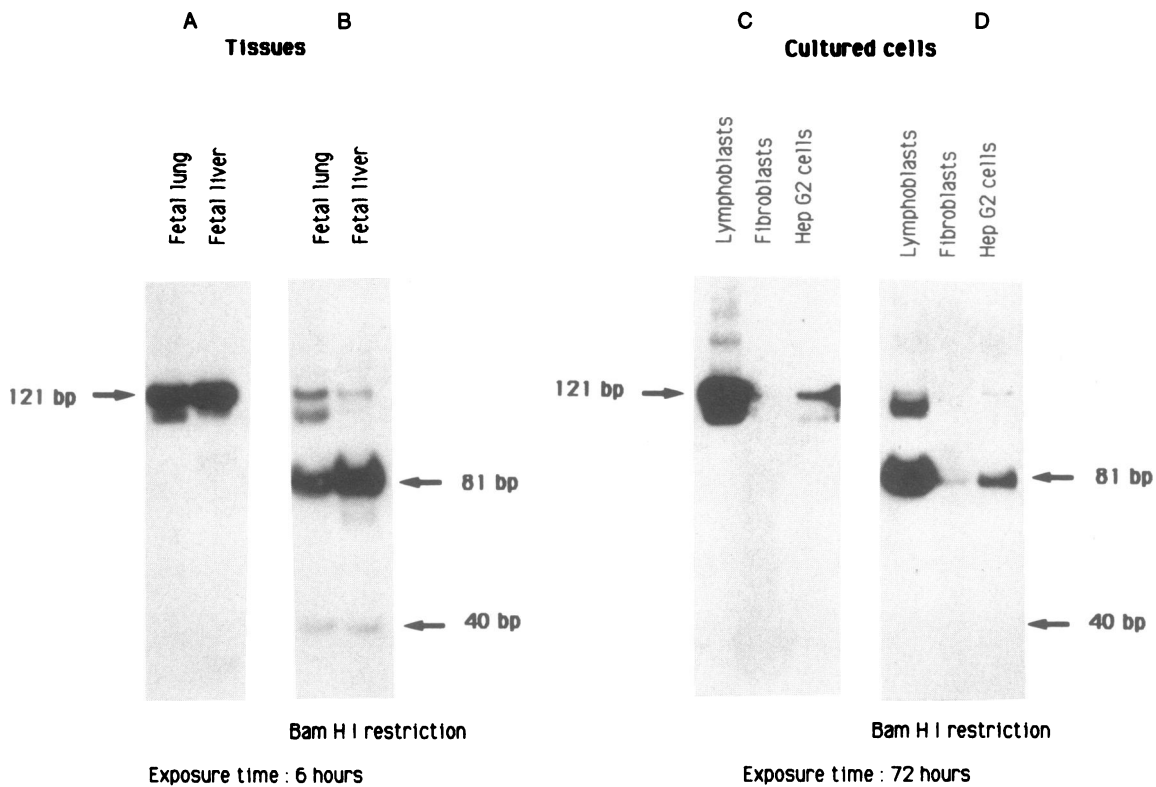


FIG. 3. Detection and characterization by cDNA-PCR of β -globin transcripts in different cells. (A and C) Specific hybridization of β -globin cDNA fragments amplified from RNAs extracted from various human tissues (A) and cultured cells (C). (B and D) Southern blots of the same cDNA-PCR products as in A and C after digestion by *Bam*HI enzyme (the *Bam*HI site is shown in Fig. 1). The 40-bp band is very faint and does not reproduce well on the picture.

ment with the earlier work of Humphries *et al.* (16) who detected mouse β -globin transcripts in normal and transformed fibroblasts using liquid hybridization.

Possible Mechanisms of Illegitimate Transcription. The current view of eukaryotic promoters implies involvement of various transcriptional factors (14, 17), some being ubiquitous [e.g., TATA box factors (14, 18), RNA polymerase, CAAT box binding proteins (14, 19)] and others being tissue-specific (20). Tissue-specific promoters are controlled

by various DNA elements binding both ubiquitous and tissue-specific factors. In the absence of tissue-specific transcriptional factors, gene transcription is probably very low but not null. Activation of some genes in differentiating tissues also involves modifications of chromatin structure and DNA methylation (21–23). Both are transiently changed when double-stranded DNA replicates. We can therefore anticipate that even in the absence of tissue-specific transcriptional factors, all promoters could be minimally active when ubiquitous transcriptional factors reach their cognate DNA elements. This could be made easier by the chromatin

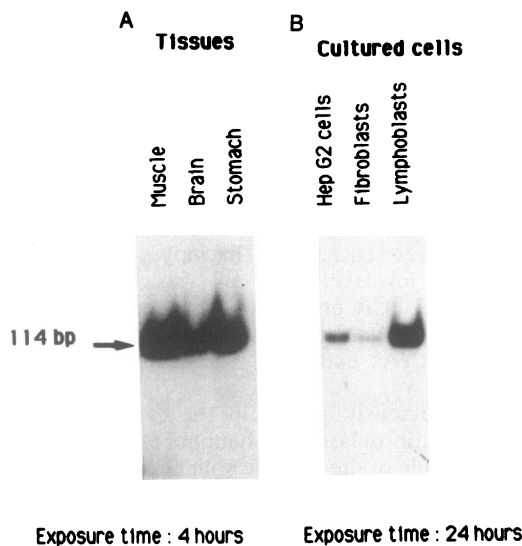


FIG. 4. Detection and characterization by cDNA-PCR of muscle-specific aldolase A transcripts in different cells: specific hybridization of muscle-specific aldolase A cDNA fragments amplified from RNAs extracted from various human tissues (A) and cultured cells (B).

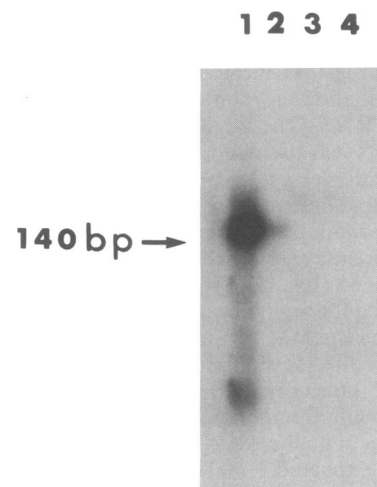


FIG. 5. Specific hybridization of AMH cDNA fragments amplified from RNAs extracted from fetal testis (lanes 1 and 3) and fetal brain (lanes 2 and 4). For fragments in lanes 1 and 2, amplification was performed as described in *Materials and Methods*. For fragments in lanes 3 and 4, the first step of cDNA synthesis was omitted.

disruption which occurs during DNA replication. Such a relationship between illegitimate transcription and DNA replication could explain why we found more illegitimate transcripts in actively proliferating lymphoblasts than in confluent fibroblasts. We are currently testing this hypothesis.

Illegitimate Transcription as a New Tool for the Investigation of Abnormal Highly Tissue-Specific Genes. Whatever the mechanism of illegitimate transcription, this phenomenon provides a powerful tool for investigating pathological transcripts by using easily accessible cells (e.g., fibroblasts, lymphoblasts, and even peripheral blood cells). The coding sequence being more compact in mRNA than in DNA, it is much easier to search for an unknown mutation by studying the messenger rather than the gene itself. This approach is particularly warranted for huge genes, like the DMD and the factor VIIIc genes. However, until now, this approach had been limited by the difficulty of obtaining pertinent mRNAs when they are located in inaccessible cells (e.g., brain, liver, pancreas, and heart), sometimes at a very low level. The finding that such mRNAs are virtually present in all cell types and the possibility of amplifying faithful cDNA fragments by starting from easily accessible cells greatly enlarge the perspective for this approach. Starting from lymphoblast RNAs, one could amplify cDNA fragments encompassing the sequence of any approximately 2 kb-long mRNA by using four or five pairs of oligonucleotide primers.

This method also could be used for quickly cloning a cDNA fragment corresponding to some partially sequenced putative exons, even if the site of maximal transcription of this putative gene is unknown.

We are grateful to Jean-Yves Picard and Daniel Guerrier for giving us specific probes and oligonucleotides and to Mrs. Claudine Brunner for typing the manuscript. This work was supported by grants from the Association Française Contre les Myopathies (AFM), Association de Recherche sur le Cancer, Ligue Française de Recherche Contre le Cancer, and Centre National de la Recherche Scientifique. J.C. is recipient of a fellowship from AFM.

1. Chelly, J., Kaplan, J. C., Maire, P., Gautron, S. & Kahn, A. (1988) *Nature (London)* **333**, 858–860.

2. Cate, R. L., Mattaliano, R. J., Hesscon, C., Tizard, R., Faber, N. M., Cheung, A., Ninfa, E. G., Frey, A. Z., Gash, D. J., Chow, E. P., Fisher, R. A., Bertoni, J. M., Torres, R. C., Manganaro, T. F., MacLaughlin, D. T. & Donahoe, P. K. (1986) *Cell* **45**, 685–698.
3. Jost, A. (1953) *Recent Prog. Horm. Res.* **8**, 379–418.
4. Picard, J. Y. & Josso, N. (1984) *Mol. Cell. Endocrinol.* **34**, 23–29.
5. Maire, P., Gautron, S., Hakim, V., Gregori, C., Mennecier, F. & Kahn, A. (1987) *J. Mol. Biol.* **197**, 425–438.
6. Cagnet, M., Lone, Y., Vaulont, S., Kahn, A. & Marie, J. (1987) *J. Mol. Biol.* **196**, 11–25.
7. Noguchi, T., Yamada, K., Inoue, H., Matsuda, T. & Tanaka, T. (1987) *J. Biol. Chem.* **262**, 14366–14371.
8. Knowles, B. B., Howe, C. C. & Aden, D. P. (1980) *Science* **209**, 497–499.
9. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
10. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239**, 487–491.
11. Weber, A., Marie, J., Cottreau, D., Simon, M. P., Besmond, C., Dreyfus, J. C. & Kahn, A. (1984) *J. Biol. Chem.* **259**, 1798–1802.
12. Marie, J., Simon, M. P., Dreyfus, J. C. & Kahn, A. (1981) *Nature (London)* **292**, 70–72.
13. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
14. Maniatis, T., Goodbourn, S. & Fischer, J. A. (1987) *Science* **236**, 1237–1245.
15. Ryffel, U. G. & McCarthy, B. J. (1975) *Biochemistry* **14**, 1379–1384.
16. Humphries, S., Windass, J. & Williamson, R. (1976) *Cell* **7**, 267–277.
17. Dynan, W. S. & Tjian, R. (1985) *Nature (London)* **316**, 774–777.
18. Chen, W. & Struhl, K. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2691–2695.
19. Santoro, C., Mermod, N., Andrews, P. C. & Tjian, R. (1988) *Nature (London)* **334**, 218–224.
20. Lichsteiner, S., Wuarin, J. & Schibler, U. (1987) *Cell* **51**, 963–973.
21. Charnay, P., Treisman, R., Mellon, P., Moses, C., Axel, R. & Maniatis, T. (1984) *Cell* **38**, 251–263.
22. Doerfler, W. (1983) *Annu. Rev. Biochem.* **52**, 93–124.
23. Weintraub, H. & Groudine, M. (1976) *Science* **193**, 848–856.