The *ets* sequence from the transforming gene of avian erythroblastosis virus, E26, has unique domains on human chromosomes 11 and 21: Both loci are transcriptionally active

(acute leukemia/oncogene expression)

D. K. Watson*, M. J. McWilliams-Smith*, M. F. Nunn†, P. H. Duesberg‡ , S. J. O'Brien§, and T. S. Papas*¶

*Laboratory of Molecular Oncology and §Laboratory of Viral Carcinogenesis, National Cancer Institute, Frederick, MD 21701; †Salk Institute, La Jolla, CA 92122; and ‡Department of Molecular Biology, University of California, Berkeley, CA 94620

ABSTRACT   Human DNA segments homologous to the *ets* region from the transforming gene of avian erythroblastosis virus, E26, were molecularly cloned and shown to be closely related to the viral equivalent by hybridization and partial sequence analysis. The transforming gene of E26 has a tripartite origin with the structure Δgag [1.2 kilobases (kb) from the viral *gag* gene]-*myb* (0.9 kb from the chicken *myb* gene)-*ets* (1.6 kb from the chicken *ets* gene). Human *ets* DNA is located on two distinct human chromosomes. The human *ets-1* locus on chromosome 11 encodes a single mRNA of 6.8 kb; the second locus, *ets-2* on chromosome 21, encodes three mRNAs of 4.7, 3.2, and 2.7 kb. The *ets*-related sequences of human DNA on chromosomes 11 and 21 are discontiguous, except for a small overlap region encoding 14 amino acids, where 12 are conserved between these two loci. By contrast, the chicken homolog has contiguous *ets-1* and *ets-2* sequences and is primarily expressed in normal chicken cells as a single 7.5-kb mRNA. We conclude that the *ets* sequence shared by the virus, the chicken, and humans is likely to contain at least two dissociable functional domains, *ets-1* and *ets-2*. Thus, the tripartite transforming gene of E26 includes four distinct domains that may be functionally relevant for the transforming function of the virus (Δgag, *myb*, *ets-1*, and *ets-2*).

The transforming (*onc*) genes of all oncogenic retroviruses are derivatives of cellular genes, which have been termed proto-*onc* genes. Typically, viral *onc* genes are hybrids of a specific subset of one proto-*onc* gene linked to genetic elements from retroviruses. Avian erythroblastosis virus, E26, is exceptional in that it includes elements from two proto-*onc* genes, chicken proto-*myb* and chicken proto-*ets*, and Δgag from the viral *gag* gene. These elements are linked in the order Δgag [1.2 kilobases (kb)]-*myb* (0.9 kb)-*ets* (1.6 kb) forming a tripartite *onc* gene (1). Previous analyses have indicated that the *myb* domain of the transforming gene of E26 is an internal subset of the proto-*myb* gene of chicken (1–3) and that the *ets* domain of E26 is derived from the chicken proto-*ets* gene (1, 2, 4). These results support the model that viral *onc* genes are genetic hybrids that include subsets of proto-*onc* genes conjugated with retroviral coding and regulatory elements (5–9).

To clarify the structural and the functional relationships between the Δgag-*myb*-*ets* transforming gene of E26 and its cellular homologs, we have analyzed the structures and transcriptional patterns of the *ets*-related genes of man. Human *ets* DNA maps to two different chromosomes, 11 and 21 (10). Utilizing chromosome-specific *ets* probes, we have found that both loci are transcriptionally active. Since the

human *ets* genes are not known progenitors of a viral *onc* gene with an *ets* sequence, we refer to them here as Hu-*ets-1* and -2, rather than proto-*ets-1* and -2. Although human *ets* is not a progenitor of E26, our observation that the human equivalent of the viral *ets* sequence is distributed between two different human genes may provide a basis for a functional analysis of viral and cellular *ets* domains.

MATERIALS AND METHODS

Molecular Cloning, Subcloning, and Nucleotide Sequence of *ets*. Charon 4A recombinant chicken DNA libraries were provided by Bruce Patterson and Terry Robins. A Charon 4A recombinant human DNA library constructed from human placental DNA was provided by Steven Tronick. An EMBL-4 human library was constructed from RD, a human cell line derived from a rhabdomyosarcoma, and provided by James Casey. The libraries were propagated in *Escherichia coli* strain LE392 and were screened by the method of Benton and Davis (11). In addition, a human hepatoma cDNA library was provided by Jeffrey de Wet and was propagated in *E. coli* Y1088.

A subclone of the E26 provirus designated pER4 has been described (2). A 1.28-kb *Bgl* I DNA fragment was further subcloned into the *Eco*RI site of pBR322 and the purified insert was used for screening.

Human *ets*-related *Eco*RI restriction fragments from two nonoverlapping genomic clones were subcloned into pBR322. pRD6K and pRD700 are subclones derived from the λ phage clone λRD3 from the RD library and pH33 was subcloned from λH33 from the amplified placenta library. An *ets*-related cDNA clone, λ cDNA 14, was subcloned in pBR322. pRD700, pH33, and cDNA 14 clones were utilized for sequence analysis, carried out as described (12).

Somatic cell hybrids were derived by PEG-mediated fusion of fresh human lymphocytes to rodent cells (mouse RAG or Chinese hamster E36) lacking the hypoxanthine guanine phosphoribosyltransferase (*HPRT*) gene, thus allowing selection in hypoxanthine/aminopterin/thymidine (HAT) medium. Hybrids were genetically characterized by G-trypsin banding, G-11 staining, and enzyme typing of 20–28 isozyme markers previously assigned to human chromosomes at the same passage from which high molecular weight DNA was extracted (13–15). Genomic DNA was digested with restriction enzymes and subjected to a Southern analysis (16) using the specific probes described above.

Biochemistry: Watson *et al.*

Proc. Natl. Acad. Sci. USA 82 (1985)   7295

RESULTS

Human *ets* Homologs Are Dispersed to Chromosomally Separate Loci. To resolve the chromosomal organization of human *ets* sequences, human DNAs were digested with *Eco*RI and after size-fractionation by agarose gel electrophoresis, the fragments were blotted onto a nitrocellulose sheet and hybridized with the three Hu-*ets*-specific clones. As illustrated in Fig. 1, different Hu-*ets* subclones, specific for different regions of v-*ets* (Fig. 2), recognized different fragments in human genomic DNA. The RD6K probe recognized a 6.2-kb *Eco*RI fragment in human DNA; RD700 recognized a 0.83-kb fragment; and H33 resolved a 3.6-kb fragment. Since each Hu-*ets* subclone recognized a distinct human fragment, it was possible that chromosomally separate DNA segments, each homologous to different portions of the v-*ets* oncogene, existed in the human genome. To explore this possibility, DNAs from a panel of rodent × human somatic cell hybrids, which segregate human chromosomes in different combinations (10, 13, 14), were examined with these Hu-*ets* clones. The results (Fig. 1) indicated that the RD6K-specific 6.2-kb *Eco*RI fragment and the RD700 0.83-kb *Eco*RI fragment were concordantly associated in the hybrids, suggesting they resided on the same human chromosome. The H33-specific 3.6-kb *Eco*RI fragment was highly discordant with the other two fragments, suggesting a minimum of two chromosomally distinct and nonoverlapping *ets* loci present in the human genome. We have subsequently assigned these two loci termed *ets-1* (defined by homology to RD6K and/or RD700) and *ets-2* (defined by homology to H33) to human chromosomes 11 and 21, respectively (10). The concordance of the two loci with their resident human chromosomes is demonstrated here by correlating their appearance with those hybrids of the expected chromosomal phenotype (Fig. 1).

Determination of the Point of Dispersion of v-*ets* Homologous Sequence in Hu-*ets-1* and Hu-*ets-2*. By hybridization analyses, we have shown above that Hu-*ets-1* and Hu-*ets-2* contain unique subsets of v-*ets* homologous sequences. We have determined the partial nucleotide sequence of *ets-1* (RD700) and *ets-2* (cDNA 14 and H33). By alignment of these sequences with those of v-*ets* from E26 (Fig. 3), one can see three regional domains of homology; two related uniquely to Hu-*ets-1* and Hu-*ets-2* and a third that defines the overlap between the v-*ets* related sequences of these loci.

Each human *ets* clone contains a region that is closely related to and colinear with viral *ets* (Figs. 2 and 3). The human *ets* sequences are bordered by regions of nonhomology compared to viral *ets*, which appear to extend the reading frame in the 5' and 3' directions (Fig. 3). However, our data do not distinguish between the possibilities that these sequences of nonhomology are coding regions or are noncoding introns. No splice donor signal can be detected at the 3' border of Hu-*ets-2* and v-*ets*, suggesting that the open reading frame extends beyond this border. Hu-*ets-2* encodes 13 unique amino acids beyond this border; in contrast, the viral gene encodes 16 unique amino acids. Thus, the viral tranforming protein (p135) and the Hu-*ets-2* gene product have different carboxyl termini.

Tentative splice acceptor and donor signals can be identified at the borders of overlap between the Hu-*ets-1* (pRD700) sequence and viral *ets*. Further work analyzing human *ets-1* cDNA clones could help to define the specific coding regions of this gene.

Comparison of the predicted amino acid sequences of Hu-*ets-1* and Hu-*ets-2* with their viral equivalents (Fig. 3) demonstrated 98% and 95% homology, respectively. This strong evolutionary conservation suggests that this gene performs an important cellular function.

To determine whether chicken equivalents of the viral *ets* map in one locus or are scattered over several loci, proto-*ets*
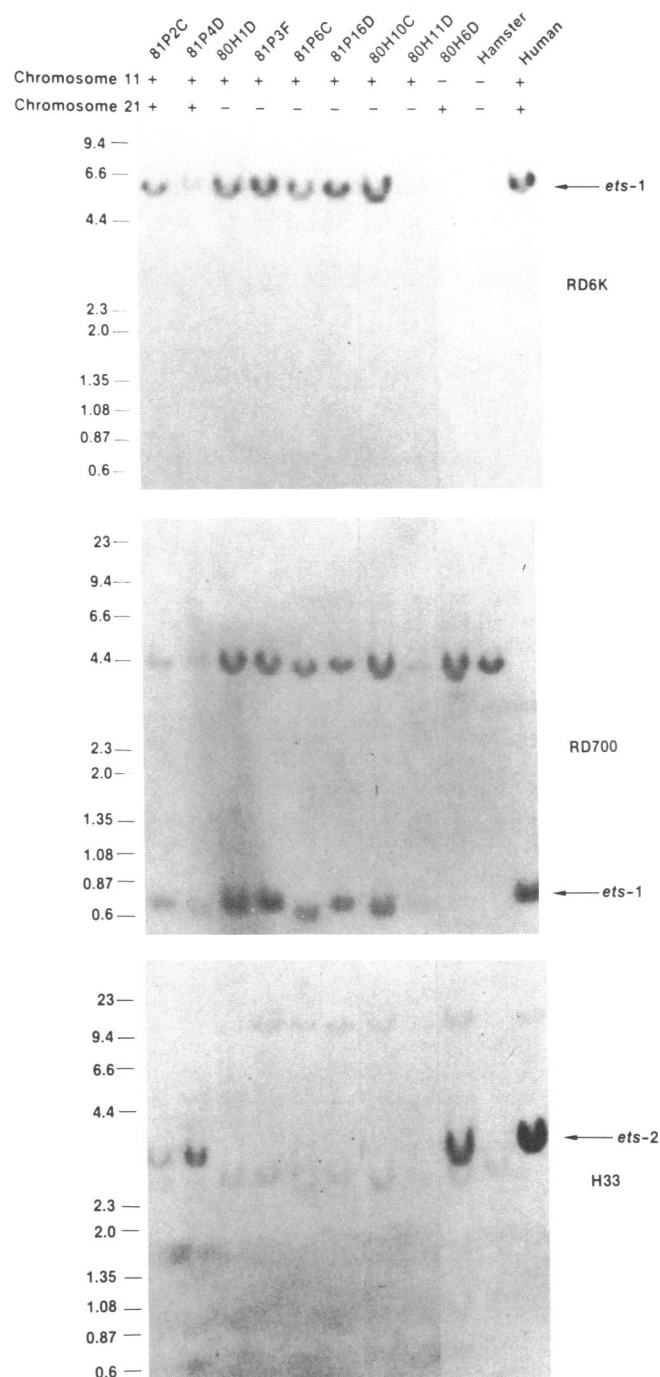
FIG. 1. Hybridization of molecularly cloned Hu-*ets* probes to *Eco*RI digests of normal human lymphocytes × E36 Chinese hamster somatic cell hybrids. λ Phage containing *ets* equivalents of human DNA were selected from two different human libraries using the *Bgl* I v-*ets* clone as a screening probe. One recombinant phage, λRD3, was obtained from the RD library; the other, λH33, was from a human placenta library (see *Materials and Methods*). Two fragments of *ets* DNA sequences from λRD3 were subcloned into pBR322 and were designated pRD6K and pRD700. A subclone of *ets* DNA was derived from λH33 and was termed pH33. The sizes and bordering restriction sites of these clones are diagrammed in Fig. 2. Samples of DNA prepared from somatic cell hybrids between hamster and human cells were digested with *Eco*RI and fragments were resolved on 0.8% agarose gels. Immobilized DNA (16) was hybridized under stringent conditions (50% formamide/0.75 M NaCl/0.075 M Na citrate; 42°C) with human c-*ets* probes RD6K, RD700, and H33. The presence or absence of chromosomes 11 and 21 is indicated by symbols (+, −) under the hybrid cell line name. *ets-1* and *ets-2* denote sequence derived from human chromosome 11 and chromosome 21 loci, respectively. Numbers on left represent kb.

FIG. 2. Comparison of the major features of molecular clones of Hu-*ets-1* and Hu-*ets-2*, and the *ets*-specific region of the E26 virus. The previously sequenced v-*ets* portion of E26 (1) is diagrammed and aligned with the subclone, pRD700, derived from a molecular clone, λRD3, from the Hu-*ets-1* locus, and two subclones cDNA 14 and pH33 derived from the Hu-*ets-2* locus. Alignment was determined by sequence analysis (see Fig. 3). The Hu-*ets-1* subclone is diagrammed above v-*ets* and the Hu-*ets-2* subclones are shown below v-*ets*. Another subclone from λRD3, pRD6K (insert used for probe in data presented in Figs. 1 and 4) has not been localized by sequence analysis but it has homology to the 537-bp *Hph* I/*Pvu* II fragment of v-*ets* (i.e., immediately 5' to the portion of v-*ets* aligned with pRD700). Hatched areas represent homologous domains determined by nucleotide sequence analysis.

FIG. 3. Comparison of nucleotide sequences of Hu-*ets-1*, Hu-*ets-2*, and the viral transforming gene of E26 (v-*ets*). Nucleotide sequence of the E26 *ets*-homologous domains of Hu-*ets* molecular clones are aligned with v-*ets*. The complete line of nucleotides, the numbering, and predicted amino acids are from that previously presented for v-*ets* proceeding in the 5' and 3' direction (1). The nucleotide and amino acid changes found in the human *ets-1* gene (pRD700) are aligned above this partial viral sequence, while those found in the human *ets-2* gene (cDNA 14, pH33) are presented below the viral sequence. Brackets indicate 5' and 3' junctions of viral and cellular flanking DNA. The large arrows define the region of overlap between Hu-*ets-1* and Hu-*ets-2*. Presumptive splice donor ( ) and acceptor ( ) signals are indicated. The symbol (I) defines the junction between the last two Hu-*ets-2* exons.

clones were selected from two chicken genomic libraries and characterized. One recombinant phage, termed λC51, was selected with a viral *ets* probe and was found to contain a chicken DNA insert of ≈15 kb (unpublished observation). To estimate the complexity of the *ets* sequence of λC51, the clone was hybridized with either a 5′ [537-base-pair (bp) *Hph* I/*Pvu* II] or 3′ (169-bp *Hin*dIII/*Bgl* I) viral probe (Fig. 2). Both probes hybridized to the λC51 DNA insert, indicating that sequences complementary to both Hu-*ets-1* and Hu-*ets-2* exist within this 15 kb of chicken DNA.

Hu-*ets-1* and Hu-*ets-2* Are Both Transcriptionally Active. To determine whether the human *ets* loci are transcriptionally active and to estimate the complexity of the human *ets* genes, we have analyzed ets-specific mRNAs. For this purpose poly(A)-selected RNA from human cells was size-fractionated by electrophoresis in an agarose gel, transferred to nitrocellulose paper, and then hybridized with ³²P-labeled DNA specific for Hu-*ets-1* or Hu-*ets-2*. The results show a 6.8-kb ets-1 mRNA and three distinct ets-2 mRNA species of 4.7, 3.2, and 2.7 kb (Fig. 4).

We draw the following conclusions: (*i*) Both human *ets* loci are transcriptionally active. (*ii*) The *ets-1* appears to be a single gene of maximal coding complexity of 6.8 kb and *ets-2* may be a single gene with alternate initiation or splice signals with a complexity of 4.2 kb. (*iii*) There is a small overlap between the *ets*-related sequences of Hu-*ets-1* and Hu-*ets-2*.

Similar studies with RNA from chicken cells, using a v-*ets* probe that contains *ets-1* and *ets-2* sequences, identifies primarily a single major species of ≈7.5 kb. A minor component was also observed that does not correspond to any human mRNA species identified (Fig. 5). Similar results have been reported by others (4).

DISCUSSION

Transduction of a 1.6-kb sequence from chicken proto-*ets* by the retrovirus that generated E26 is as yet the only hint that proto-*ets* has oncogenic potential. The transforming gene of E26 appears to be one of the most complex examples of how proto-*onc* genes and retroviruses generate viral *onc* genes.

Our data suggest that the *ets* region shared by E26, chickens, and humans contains at least two dissociable
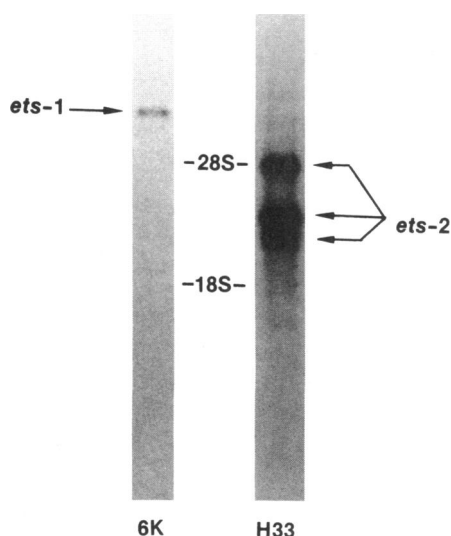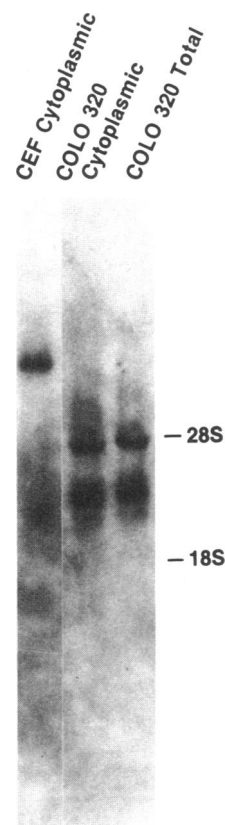

FIG. 5. Chicken *ets* transcription. RNA was prepared from cytoplasmic fractions from chicken embryo fibroblasts (CEF) and COLO 320 cells. Total cell RNA was also prepared from COLO 320 cells. Samples (20 μg) were resolved on 1.2% formaldehyde/agarose gels. CEF RNA was hybridized to a v-*ets* probe (1.28-kb *Bgl* I fragment) and COLO 320 RNA samples were hybridized to H33 under stringent conditions (as described in Fig. 1).

domains. This is deduced from the observation that the human complement of the viral *ets*-related sequence is distributed between two different genes located on two different chromosomes. Furthermore, the high degree of conservation of the *ets*-related genes in chicken and man suggests that these genes encode functions essential to the cellular function. It thus appears that the transforming gene of E26, which is derived from three different progenitor genes, may contain four functional domains—Δ*gag*, *myb*, and two *ets* domains corresponding to human *ets-1* and *ets-2*. We are now constructing E26 deletion mutants to determine functionality of *ets-1* and *ets-2* domains.

Our results delineating the expression of *ets-1* and *ets-2* in human cells give evidence for multiple RNA species (Fig. 4). The Hu-*ets-1* probe reveals one 6.8-kb transcript, while the Hu-*ets-2* probe detected three distinct RNA species of 4.7, 3.2, and 2.7 kb. This suggests functionally distinct proteins potentially translatable from these multiple mRNAs. This result is in marked contrast to that found in chicken cell RNA; hybridization with the v-*ets* probe showed only one major RNA species of 7.5 kb. Taken together, these data indicate that in the chicken only one gene product is functionally transcribed, whereas in man as a result of evolutionary divergence, the *ets-2* domain became part of another gene and became transcriptionally active as a series of mRNAs, perhaps of differing function(s). This hypothesis is further substantiated by our observations that *ets-1* and *ets-2* behave as separate unique loci in the mouse and cat, as well as in man (10). The mammalian homologs of *ets-1* map to human chromosome 11, mouse chromosome 9, and feline chromosome D1. Similarly, the homologs of *ets-2* map to human

FIG. 4. Human *ets-1* and *ets-2* loci are transcriptionally active. Polyadenylylated RNA prepared from HeLa cells was resolved on 1.5% formaldehyde/agarose gels (17) and hybridized to RD6K (purified DNA from pRD6K, *ets-1*) or H33 (purified DNA from pH33, *ets-2*) probes under stringent conditions (as described in Fig. 1). Mobility of the 28S and 18S ribosomal RNAs are as indicated.

chromosome 21, mouse chromosome 16, and feline chromosome C2 (10). These *ets* domains are situated within syntenic groups of homologous, linked loci known to be conserved among the three diverse mammalian species. Thus, in addition to establishing conservation of linkage groups and the occurrence of two distinct *ets*-related loci in three mammalian orders, our work suggests that *ets* encodes separate transcriptionally active and distinct functional products. These genes and their encoded products have presumably been separate and functionally distinct prior to the evolutionary radiation of the Mammalia. We are presently preparing specific antibodies against synthetic oligopeptides derived from the predicted *ets* sequences in an effort to characterize the various gene products. These reagents will also assist in the determination of their respective biological functions, which by analogy with other oncogene products are expected to be involved at some point in control of normal cellular function.

1. Nunn, M. F., Seeburg, P. H., Moscovici, C. & Duesberg, P. H. (1983) *Nature (London)* **306**, 391–395.
2. Nunn, M., Weiher, H., Bullock, P. & Duesberg, P. (1984) *Virology* **139**, 330–339.
3. Klempnauer, K.-H., Gonda, T. J. & Bishop, J. M. (1982) *Cell* **31**, 453–463.
4. Leprine, D., Gegonne, A., Coll, J., deTaisne, C., Schneeberger, A., Lagrou, C. & Stehelin, D. (1983) *Nature (London)* **306**, 395–397.
5. Duesberg, P. (1983) *Nature (London)* **304**, 219–225.
6. Duesberg, P. H., Nunn, M., Kan, N., Watson, D., Seeburg,P. H. & Papas, T. (1985) *Modern Trends in Human Leukemia* (Springer-Verlag, Berlin), Vol. 29, pp. 9–27.
7. Papas, T. S., Kan, N. C., Watson, D. K., Flordellis, C. S., Psallidopoulos, M. C., Lautenberger, J., Samuel, K. P. & Duesberg, P. (1984) *Cancer Cells* (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York), Vol. 2, pp. 153–163.
8. Papas, T. S., Kan, N. C., Watson, D. K., Flordellis, C., Lautenberger, J. A., Psallidopoulos, M. C. Rovigatti, U. G., Samuel, K. P., Ascione, R. & Duesberg, P. H. (1985) *Anticancer Res.* **5**, 73–80.
9. Flordellis, C. S., Kan, N. C., Lautenberger, J. A., Samuel, K. P., Garon, C. F. & Papas, T. S. (1985) *Virology* **141**, 267–274.
10. Watson, D. K., McWilliams-Smith, M. J., Kozak, C., Reeves, R., Gearhart, J., Nash, W., Modi, W., Duesberg, P., Papas, T. S. & O'Brien, S. J. (1985) *Proc. Natl. Acad. Sci. USA*, in press.
11. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
12. Watson, D. K., Reddy, E. P., Duesberg, P. H. & Papas, T. S. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2146–2150.
13. O'Brien, S. J., Bonner, T. I., Cohen, M., O'Connell, C. & Nash, W. G. (1983) *Nature (London)* **303**, 74–77.
14. O'Brien, S. J., Simonson, J. M. & Eichelberger, M. A. (1982) in *Techniques in Somatic Cell Genetics*, ed. Shay, J. W. (Plenum, New York), pp. 342–370.
15. Nash, W. G. & O'Brien, S. J. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6631–6635.
16. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
17. Lehrach, H., Diamond, D., Wozney, J. M. & Boedtker, H. (1977) *Biochemistry* **16**, 4743–4751.