Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 4490–4494, May 1993
Genetics

The retroviral oncogene *qin* belongs to the transcription factor family that includes the homeotic gene fork head

JIAN LI* AND PETER K. VOGT*

Department of Microbiology, University of Southern California School of Medicine and Norris Cancer Center, 2011 Zonal Avenue, Los Angeles, CA 90033-1034

Contributed by Peter K. Vogt, February 16, 1993

ABSTRACT **Avian sarcoma virus 31 contains an oncogene that we have named *qin*. *qin* codes for a nuclear protein, Qin, that is a member of the HNF-3/fork head family of transcriptional regulators. Within this family Qin is particularly closely related to rat brain factor 1 (BF-1), a telencephalon-specific gene presumed to play an important role in the development of the mammalian brain.**

Retroviral oncogenes are cellular growth-regulatory genes mutated and inserted into the viral genome and expressed under control of viral regulatory sequences. Incorporation of a mutated cellular oncogene makes a retrovirus highly tumorigenic and capable of transforming cells in culture. Oncogenic sequences found in retroviral genomes can be used to identify cellular homologs which are important growth-regulatory genes.

We have used retroviruses isolated from spontaneous tumors of chickens to search for cellular growth-regulatory genes. Here we describe the isolation of *qin*,† a gene that codes for a protein that belongs to the family of transcription factors typified by hepatocyte nuclear factor 3 (HNF-3) and the *Drosophila* homeotic regulator fork head.

MATERIALS AND METHODS

Virus and Cell Culture. Avian sarcoma virus (ASV) 31 was isolated from a spontaneous sarcoma in an adult chicken. Chicken embryo fibroblasts (CEFs) were cultured and viral focus assays were performed according to standard techniques (1).

Northern Blot Hybridization. Poly(A)$^+$ RNAs were isolated from avian leukosis virus- and from ASV 31-infected CEFs with the FastTrack kit (Invitrogen, San Diego). The RNA preparations were fractionated by electrophoresis through a 0.8% agarose gel containing 3.07% formaldehyde and were hybridized against probes derived from avian retroviral long terminal repeat (LTR), *gag*, *pol*, and *env* sequences (2).

Construction of a cDNA Library of ASV 31-Transformed CEFs and Screening for ASV 31. By using the linker–primer GGACAGGCCGAGGCGGCC(T)$_{40}$, cDNA was synthesized according to the Copy Kit protocol (Invitrogen), cloned into λpCEV27 arms, and packaged as described (3–5). The complete cDNA library (titer, 5×10^5 plaque-forming units/ml) was screened by plaque hybridization using ^{32}P-labeled *gag*, *pol*, and *env* probes derived from avian retroviral genomes.

Molecular Cloning. Phage DNA from selected *gag*$^+$, *pol*$^-$, *env*$^+$ λ clones was prepared and digested with restriction endonuclease *Sal* I. Inserts of 4–5 kb were cloned into the *Sal* I site of Bluescript II SK(−) (Stratagene), and clone c714 was selected for study. The full 4.5-kb c714 insert was further cloned into the *Sal* I site of CLA12NCO (6). A *Cla* I fragment of about 4610 bp was excised from this adaptor and religated to *Cla* I-cleaved RCAS avian retroviral expression plasmid, generating RCAS-c714 (6).

DNA Sequencing. Nested deletions of clone c714 were constructed with exonuclease III (5). Single-stranded DNAs were rescued (Stratagene) from the c714 deletion clones and sequenced by the dideoxy chain-termination method (7) with the Sequenase kit (United States Biochemical). The opposite strand sequence of clone c714 was determined by using a series of oligonucleotides from the established complementary sequence as primers.

DNA Transfection. CEFs were prepared according to standard techniques (1) and cultured in Ham's F10 medium with 10% fetal bovine serum. They were seeded at 10^6 per 35-mm well and grown at 37°C overnight. The cultures were transfected with 0.5–2.0 μg of RCAS-c714 plasmid DNA by the dimethyl sulfoxide Polybrene method (8). Transfected cells were overlayed with nutrient agar the day after transfection and incubated at 37°C until foci of transformed cells developed (14–18 days).

Immunofluorescence. Immunofluorescent staining was performed as described (9) with a 1:40 dilution of monoclonal antibody 1A1, which is directed against the avian retroviral Gag protein p19.

RESULTS

ASV 31 Codes for a Nuclear Oncogene Product. ASV 31 is one of several viruses isolated from spontaneous sarcomas in adult chickens. It was found to induce foci of transformed cells in CEFs and fibrosarcomas at the site of inoculation in young chickens (T. Ruscio, S. Benedict, and P.K.V., 1983, unpublished observations). ASV 31 is a retrovirus; initial information on its genome was obtained by Northern blotting (Fig. 1). ASV 31-infected CEFs contained three major mRNA species (7.8, 4.9, and 2.7 kb) that hybridized to an avian retroviral LTR probe. mRNAs of the same size were also detected with a retroviral *env*-specific probe. An avian *gag*-specific probe detected the two larger mRNAs, and a *pol* probe only the largest one. A nontransforming avian leukosis helper virus produced the expected two sizes of mRNA detectable with an LTR probe. The larger of the two represented the full-length genome and was also detectable with *gag*, *pol*, and *env* probes. The smaller represented the subgenomic *env* message. It reacted only with the LTR and *env* probes. A comparison of the ASV 31 and helper virus RNA species suggests that the 4.9-kb band obtained from ASV 31-infected cells is the genome of a defective transforming retrovirus which still carries at least partial *gag* and *env*

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: ASV, avian sarcoma virus; CEF, chicken embryo fibroblast; BF-1, brain factor 1; LTR, long terminal repeat; HNF, hepatocyte nuclear factor; ILF, interleukin binding factor.

*Current address: Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, CA 92037.

†The sequence reported in this paper has been deposited in the GenBank data base (accession no. L10719).

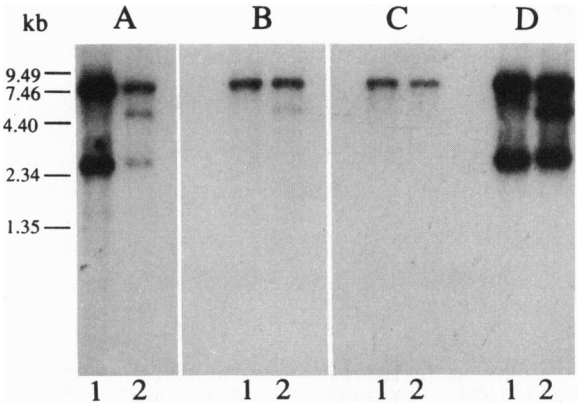

FIG. 1. Northern blot hybridization of poly(A)$^+$ RNA from CEFs infected with an avian leukosis retrovirus (lanes 1) or ASV 31 (lanes 2). The radioactive avian retroviral probes were from the long terminal repeat (LTR) (*A*), *gag* (*B*), *pol* (*C*) and *env* (*D*).

sequences but which has lost the *pol* gene, presumably in exchange for a cell-derived oncogene.

CEFs transformed by ASV 31 were stained in immunofluorescence assays with a monoclonal antibody specific for the avian retroviral Gag protein p19 (Fig. 2*A*). Such cells showed cytoplasmic as well as nuclear fluorescence, in contrast to the p19 protein of replication-competent retroviruses, which is located exclusively in the cytoplasm and at the cell surface. The Gag-specific nuclear staining seen in ASV 31-transformed cells may therefore be indicative of a fusion protein in which Gag epitopes have become linked to a cellular protein that is translocated to the nucleus. ASV 31 may contain a "nuclear" oncogene. For this reason, ASV 31 RNA was tested by Northern blot hybridization for possible homology to the known nuclear oncogenes *myc, myb, jun, fos, ski, maf,* and *erbA*, but no such homology was found (data not shown).

The Oncogene *qin* of ASV 31 Is a Member of the HNF-3/fork head Family of Transcription Factors Closely Related to Brain Factor 1 (BF-1). The genome of ASV 31 was cloned from a cDNA library in the λpCEV27 vector (4). Clones were selected with the avian retroviral *gag* and *env* probes and counterselected with the avian retroviral *pol* probe. Of 23 *gag*-positive clones, 5 clones which reacted with the *env* probe but not with the *pol* probe were picked for DNA analysis. The clones contained inserts of about 4–5 kb, the expected size of the ASV 31 genome (Fig. 1). They were then subcloned in pBluescript II SK(−) (Stratagene). Nested

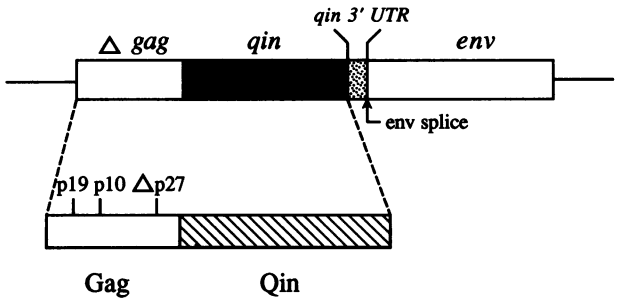

FIG. 3. Map of ASV 31 clone c714 and its protein product. UTR, untranslated region.

deletions of the insert of one of the clones, c714, were sequenced. The nucleotide sequence of the cloned viral DNA revealed the existence of a nonviral insertion, a putative cell-derived oncogene. This oncogene was named *qin*, an allusion to the Chinese word 禽, meaning "avian," to signify the origin of the gene. The length of *qin* is 1397 bp, and it contains an open reading frame for 395 amino acids. As expected, the 5′-terminal portion of the gene is fused with viral *gag*. The 3′ recombination junction of *qin* is the splice site of *env* (Fig. 3). Computer-assisted sequence comparison revealed that *qin* belongs to the HNF-3/fork head family of transcription factors. Within that family it is particularly closely related to the gene encoding BF-1, a telencephalon-specific transcription factor gene isolated from a rat brain cDNA library (Fig. 4) (10). Within the highly conserved fork head domain, the *Drosophila* segmentation gene sloppy pair (*slp-2*) is closest to BF-1 and *qin* (11). More distant but still highly significant relationships can be seen to interleukin binding factor (ILF) (12), HNF-3 (13, 14), and *Drosophila* and *Xenopus* fork head proteins (15, 16).

A *qin*-Containing Construct Induces Oncogenic Transformation in Cell Culture and Tumors in the Animal. The genome of ASV 31, excised from the pBluescript clone c714 and encompassing a fragment of about 4500 bp, was subcloned in the avian retroviral expression vector RCAS (6), and the construct was transfected into CEFs. Foci of transformed cells similar to the ones seen with ASV 31 appeared about 15 days after transfection, and the cultures became completely transformed after passage (Fig. 5). Immunofluorescent staining with avian retroviral Gag-specific antibody showed cell surface and nuclear localization of the protein in the transformed cultures, while controls transfected with the vector alone showed only the cell surface fluorescence (Fig. 2 *B* and

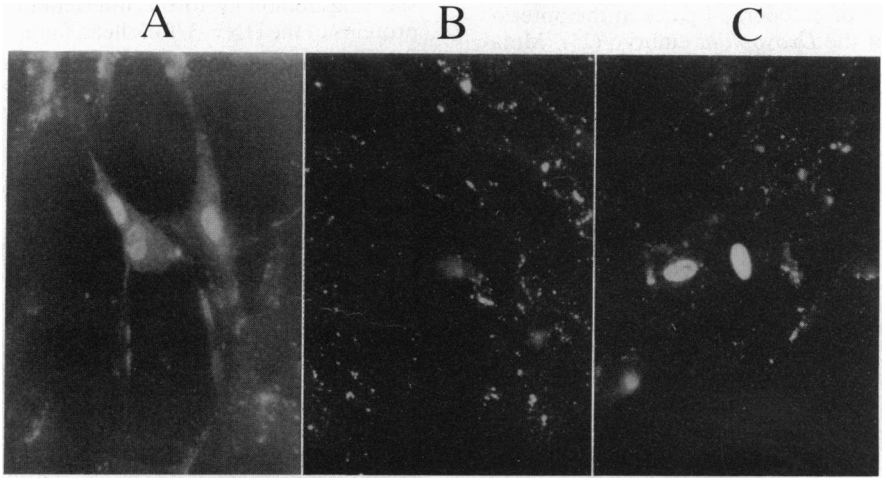

FIG. 2. Immunofluorescent staining of CEFs infected with ASV 31 (*A*), RCAS vector alone (*B*), and RCAS containing clone c714 as insert (*C*). The antibody is directed against avian retroviral Gag protein. (×250.)

```
BF1          1 MLDMGDRKEVKMiPKSSFSINsLVPEAVQnDNHhasHgHHNSHHPqHHHHHHHHHHPPPPa
               |||||||||||| |||||||| |||||| |||    | ||||||| ||||||||||||||
QIN  1 aapsaqrvMLDMGDRKEVKMlPKSSFSINnLVPEAVQsDNH sgHsHHNSHHP HHHHHHHHHHPPPP

BF1         62 pqpppppppqqqQQPppapqppqArgApaaDddKgPqpLllPpsaALdgAKAdALgaKGEpG
                        |||         |  |   |  | |  |  |   ||  ||| ||  ||| |
QIN         67          QQP       qrAaaAeeeDeeKaPllLppPaagALeaAKAeALagKGEaG

BF1        123 ggpAELapvgpdEKEKgagaggEEKKGAgEGGKDGEgGKEGdKKNGKYEKPPFSYNALIMM
                  |||       ||||       |||||| ||||||| |||| ||||||||||||||||||
QIN        111 aaaAEL     eEKEK    aaEEKKGAaEGGKDGEsGKEGeKKNGKYEKPPFSYNALIMM

BF1        184 AIRQSPEKRLTLNGIYEFIMKNFPYYRENKQGWQNSIRHNLSLNKCFVKVPRHYDDPGKGN
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QIN        163 AIRQSPEKRLTLNGIYEFIMKNFPYYRENKQGWQNSIRHNLSLNKCFVKVPRHYDDPGKGN

BF1        245 YWMLDPSSDDVFIGGTTgKLRRRSTTSRAKLAFKRGARLTSTGLTFMDRAGSLYWPMSPFL
               |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
QIN        224 YWMLDPSSDDVFIGGTTdKLRRRSTTSRAKLAFKRGARLTSTGLTFMDRAGSLYWPMSPFL

BF1        306 SLHHPRASSTLSYNGTtSAYPSHPMPYSSVLTQNSLGNNHSFSTaNGLSVDRLVNGEIPYA
               |||||||||||||||| |||||||||||||||||||||||||| |||||||||||||||||
QIN        285 SLHHPRASSTLSYNGTaSAYPSHPMPYSSVLTQNSLGNNHSFSTsNGLSVDRLVNGEIPYA

BF1        367 THHLTAAALAASVPCGLSVPCSGTYSLNPCSVNLLAGQTSYFfphvphpsmtsqtstsmsa
               |||||||||||||||||||||||||||||||||||||||||||
QIN        346 THHLTAAALAASVPCGLSVPCSGTYSLNPCSVNLLAGQTSYF

BF1        428 raassstSPqapstlpceSLrPslpsfttglsgglsdyfthqnqgsssnplih.
                       ||          || |
QIN        388         SP         tSLtPq.
```

FIG. 4. Amino acid sequence comparison of the *qin* protein with rat BF-1.

C). The RCAS vector produced infectious retroviral progeny, and eventually all cells in transfected cultures became transformed (Fig. 5). These virus-producing cells were injected into the wing web of 1- to 4-day-old chickens. They induced slowly progressive fibrosarcomas at the site of injection after a relatively long latent period of 3 weeks (compared with 1 week for the most highly transforming constructs containing the v-*jun* oncogene). These results demonstrated that clone c714, containing the *qin* insert, codes for a transforming and oncogenic retrovirus.

DISCUSSION

Proteins of the HNF-3/fork head family of transcriptional regulators occur in vertebrates and in invertebrate animals (10–18). The expression of most of these proteins is developmentally regulated, and they are themselves important determinants of development and differentiation. A prototype member of the family, the *Drosophila* fork head protein, is the product of a region-specific homeotic gene, *fkh*, that directs the development of ectodermal parts in the anterior and posterior termini of the *Drosophila* embryo (15). Mutations in the *fkh* gene lead to the replacement of foregut and hindgut by ectopic head structures. In later stages of development *fkh* is also expressed in the gut, the yolk nuclei, the salivary glands, and some cells of the nervous system (15, 19). The *fkh*-encoded protein does not contain a homeodomain and, unlike the homeotic genes antennapedia or bithorax, *fkh* is not regulated by the polycomb group of genes; rather, *fkh* belongs to a distinct group of homeotic genes. Seven *fkh*-related genes have been identified in *Drosophila* (*FD1–5* and *slp-1* and -2) (11, 17). The family extends to vertebrates, and there it includes two developmentally regulated genes, *Xenopus* fork head (XFKH-1) and the rat BF-1 gene. XFKH-1 is expressed in the blastopore lip of *Xenopus* gastrulas and in the notochord and neural floor plate (16). BF-1 gene expression is restricted to the prospective telencephalon region of the neural tube (10). Important members of the family are the HNF-3 proteins that mediate liver differentiation and function in the hierarchical control of liver-specific transcription (13, 14, 20). Another interesting member of the HNF-3/fork head family is ILF, which binds to a region of the human immunodeficiency virus LTR showing homology to the interleukin 2 promoter (12). The proteins of the HNF-3/fork head family are related by several domains of homologous sequences (Fig. 6). These consist of two putative α-helices, a possible hinge region, and a basic domain, all of them located in the center of the molecule. They are collectively referred to as the fork head box (16, 18). The region of homology includes the DNA-binding domain of these proteins. Some members of the family also contain short carboxyl-terminal and amino-terminal regions of homology (11, 14). HNF-3/fork head proteins are located in the cell nucleus and bind, probably as monomers, to specific DNA sequences (14). The consensus sequences for optimal binding remain to be determined, but regions of the transthyretin and HNF-1α promoters and of the human immunodeficiency virus LTR have been shown to bind various HNF-3/fork head proteins (12, 14, 20). On the basis of such binding specificity and of amino acid sequence homology in the fork head box, one could distinguish two subgroups of the family, one encompassing *Drosophila* and *Xenopus* fork head proper and the HNF-3 proteins, the other consisting of *Drosophila* sloppy paired (slp) proteins and rat BF-1. The Qin

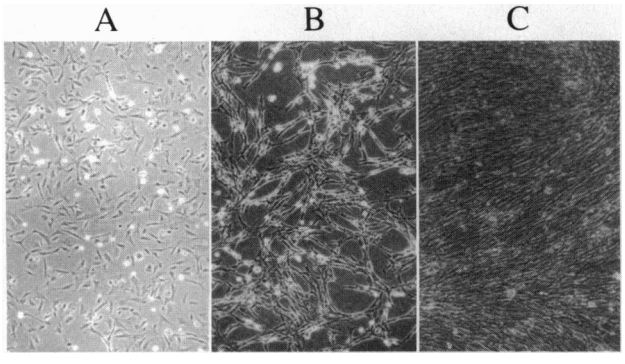

FIG. 5. Transformation of CEFs with RCAS-c714. (*A*) Uninfected CEFs. (*B*) CEFs transfected with RCAS-c714. (*C*) Confluent culture of CEFs transformed by RCAS-c714. (×60.)

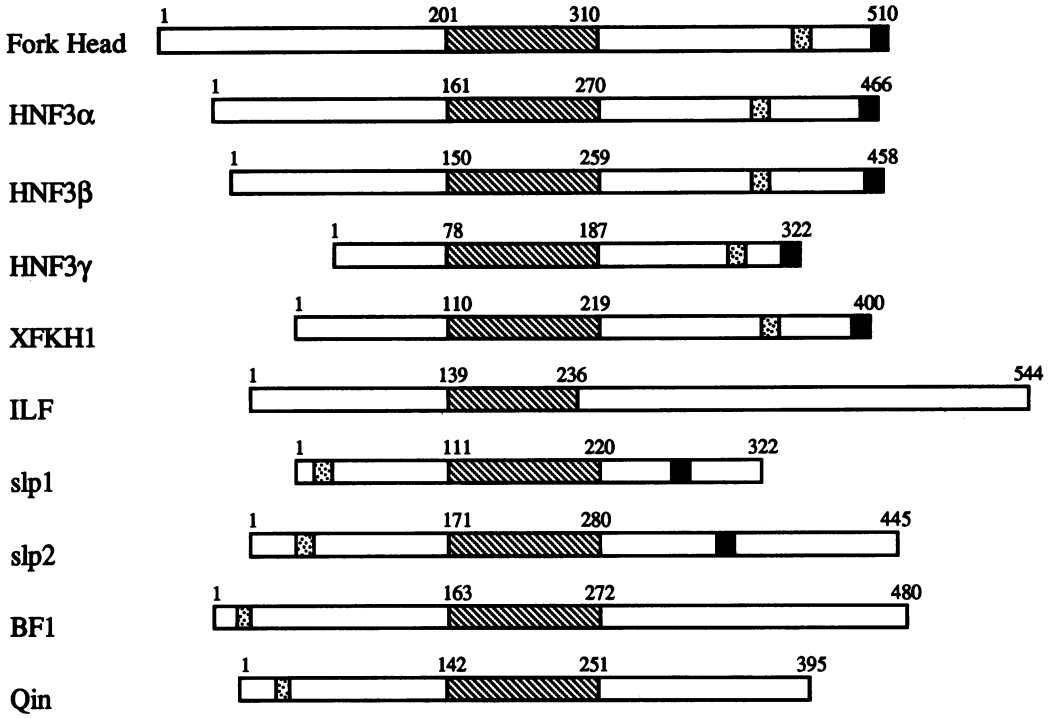

FIG. 6. The HNF-3/fork head family of transcriptional regulators. The shaded region represents the fork head domain common to all members of the family. Stippled and black regions represent smaller areas of homology not shared by all HNF-3/fork head proteins. Only partial sequences are available for the FD-1 to FD-5 members of the family; these are not shown in Fig. 6.

protein belongs to this latter group. Its amino acid sequence within the fork head box is identical to that of BF-1, with homologies extending beyond the fork head box (Fig. 4). Thus the region between amino acids 122 and 387 shows 96% identity to BF-1. Qin does, however, diverge from BF-1 significantly in the amino-terminal portion, amino acids 9–92 are only 54% identical to BF-1, showing both differences in amino acid sequence and several small deletions. Whether the latter have resulted from retroviral transduction or are present in the cellular version of *qin* is not known. The viral Qin protein extends 8 amino acids amino terminally beyond BF-1, and it appears to be truncated by a premature stop codon due to a single base deletion. Qin shares with BF-1 the preferential binding to the B2 DNA site of the HNF-1α promoter (ref. 10; J.L., E.J. Parker and P.K.V., unpublished observations).

qin has been isolated from a fibrosarcoma-inducing virus and induces the same type of tumor upon subcutaneous injection. The full tumor spectrum of *qin* remains to be determined, however. The possibility that the *qin* oncogene also transforms other cell types and that it is involved in the genesis of human tumors needs to be examined. The normal functions of the cellular *qin* gene are still unknown. Patterns of expression in various tissues and during embryonal development, as well as "knock-out" experiments, will provide initial answers to this question.

Recently, homeobox-containing genes have been implicated in childhood leukemias. The t(1;19) translocation seen in pre-B acute lymphoblastic leukemia fuses a truncated *E2A* transcription factor gene to a homeobox-encoding gene, *pbx-1*. The E2A–Pbx-1 chimeric protein produced in leukemic cells is transforming for NIH 3T3 mouse cells (21–23). In the t(10;14) translocation of childhood T-cell acute leukemia, the homeobox gene *HOX11* becomes transcriptionally activated and may play an important causative role in leukemogenesis (24). Similarly, in a mouse myeloid leukemia (WEHI-3B), the *Hox-2.4* gene is rearranged and transcriptionally activated by the insertion of an intracisternal A particle (IAP) provirus in its vicinity (25, 26). The overexpressed *Hox-2.4A* gene is able to transform NIH 3T3 mouse fibroblasts (27). Our results with *qin* show that the HNF-3/fork head family of developmentally controlled transcription factors also has oncogenic potential.

We thank Felipe S. Monteclaro for assistance in nucleic acid and protein sequence comparisons. Leanne Rochanda provided valuable technical support. Sarah Olivo, Esther Olivo, and Arianne Helenkamp produced the manuscript. Volker Vogt kindly donated the 1A1 anti-Gag monoclonal antibody and hybridoma. This work was supported by Public Health Service Research Grant CA42564 and by Grant 1951 from the Council for Tobacco Research.

1. Vogt, P. K. (1969) in *Fundamental Techniques in Virology*, eds. Habel, K. & Salzman, N. P. (Academic, New York), pp. 198–211.
2. Tsuchie, H., Chang, C. H. W., Yoshida, M. & Vogt, P. K. (1989) *Oncogene* **4**, 1281–1284.
3. Miki, T., Matsui, T., Heideran, M. A. & Aaronson, S. A. (1989) *Gene* **83**, 137–146.
4. Miki, T., Fleming, T. P., Crescenzi, M., Molly, C. J., Blam, S. B., Reynolds, S. H. & Aaronson, S. A. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 5167–5171.
5. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Plainview, NY).
6. Hughes, S. H., Greenhouse, J. J., Petropoulos, C. J. & Sutrav, P. (1987) *J. Virol.* **61**, 3004–3012.
7. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
8. Kawai, S. & Nishizawa, M. (1984) *Mol. Cell. Biol.* **4**, 1172–1174.
9. Bos, T. J., Monteclaro, F. S., Mitsunobu, F., Ball, A. R., Chang, C. H. W., Nishimura, T. & Vogt, P. K. (1990) *Genes Dev.* **4**, 1677–1687.
10. Tao, W. & Lai, E. (1992) *Neuron* **8**, 957–966.
11. Grossniklaus, U., Pearson, R. K. & Gehring, W. J. (1992) *Genes Dev.* **6**, 1030–1051.
12. Li, C., Lai, C., Sigman, D. S. & Gaynor, R. B. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 7739–7743.
13. Lai, E., Prezioso, V. R., Smith, E., Litvin, O., Costa, R. H. & Darnell, J. E., Jr. (1990) *Genes Dev.* **4**, 1427–1436.

14. Lai, E., Prezioso, V. R., Tao, W., Chen, W. S. & Darnell, J. E., Jr. (1991) *Genes Dev.* **5**, 416–427.
15. Weigel, D., Jürgens, G., Küttner, F., Seifert, E. & Jäckle, H. (1989) *Cell* **57**, 645–658.
16. Dirksen, M. L. & Jamrich, M. (1992) *Genes Dev.* **6**, 599–608.
17. Häcker, U., Grossniklaus, U., Gehring, W. J. & Jäckle, H. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 8754–8758.
18. Weigel, D. & Jäckle, H. (1990) *Cell* **63**, 455–456.
19. Weigel, D., Beller, H. J., Jürgens, G. & Jäckle, H. (1989) *Roux's Arch. Dev. Biol.* **198**, 201–210.
20. Kuo, C. J., Conley, P. B., Chen, L., Sladek, F. M., Darnell, J. E., Jr., & Grabtree, G. R. (1992) *Nature (London)* **355**, 457–461.
21. Kamps, M. P., Murre, C., Sun, X. H. & Baltimore, D. (1990) *Cell* **60**, 547–555.
22. Nourse, J., Mellentin, J. D., Galili, N., Wilkinson, J., Stanbridge, E., Smith, S. D. & Cleary, M. L. (1990) *Cell* **60**, 535–545.
23. Kamps, M. P., Look, A. T. & Baltimore, D. (1991) *Genes Dev.* **5**, 358–368.
24. Kennedy, M. A., Gonzalez-Sarmiento, R., Kees, U. R., Lampert, F., Dear, N., Boen, T. & Rabbitts, T. H. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 8900–8904.
25. Blatt, C. D., Aberdam, D., Schwartz, R. & Sachs, L. (1988) *EMBO J.* **7**, 4283–4290.
26. Kongsuwan, K., Allen, J. & Adams, J. M. (1989) *Nucleic Acids Res.* **17**, 1881–1892.
27. Aberdam, D., Negreanu, V., Sachs, L. & Blatt, C. (1991) *Mol. Cell. Biol.* **11**, 554–557.