Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 4629–4632, May 1996
Genetics

Genomic structure of the human retinoblastoma-related Rb2/p130 gene

(promoter/exon/intron)

ALFONSO BALDI*, VINCENZO BOCCIA, PIER PAOLO CLAUDIO, ANTONIO DE LUCA*, AND ANTONIO GIORDANO†

Departments of Microbiology/Immunology and Pathology, Jefferson Cancer Institute, Thomas Jefferson University, Philadelphia, PA 19107

Communicated by Sidney Weinhouse, Thomas Jefferson University, Philadelphia, PA, January 16, 1996 (received for review October 20, 1995)

ABSTRACT **The human Rb2/p130 gene shares many structural and functional features with the retinoblastoma gene and the retinoblastoma-related p107 gene. In the present study, we have cloned and partially sequenced the gene coding for the Rb2/p130 protein from human genomic libraries. The complete intron–exon organization of this gene has been elucidated. The gene contains 22 exons spanning over 50 kb of genomic DNA. The length of individual exons ranges from 65 to 1517 bp. The largest intron spans over 9 kb, and the smallest has only 82 bp. The 5′ flanking region revealed a structural organization characteristic of promoters of "housekeeping" and growth control-related genes. A typical TATA or CAAT box is not present, but there are several GC boxes and potential binding sites for numerous transcription factors. This study provides the molecular basis for understanding the transcriptional control of the Rb2/p130 gene and for implementing a comprehensive Rb2/p130 mutation screen using genomic DNA as a template.**

The retinoblastoma protein family includes three members: the retinoblastoma tumor suppressor protein, the related protein p107 (1), and the recently cloned pRb2/p130 (2–4). These proteins display high sequence homology, especially in two discontinuous regions that make up the "pocket domain" (2). The structural identities between the proteins underlie similar functional properties. In fact, the retinoblastoma protein, as well as p107 and pRb2/p130, acts as a negative regulator of cell cycle progression, blocking the cells in the G_1 phase (5–8, 36). Interestingly, these proteins exhibit different growth-suppressive properties in selected cell lines, suggesting that although the different members of the retinoblastoma protein family may complement each other, they are not fully functionally redundant (6–8). The mechanisms by which these three proteins exert their control on cell cycle progression are not fully understood but likely include complex formation and modulation of the activity of several transcription factors (9). The most studied of these complexes is the one with the E2F family of transcription factors. E2Fs are heterodimeric transcription factors composed of E2F-like and DP-like subunits that regulate the expression of genes required for progression through G_0/G_1 S phase of the cell cycle (10). pRb, p107, and pRb2/p130 bind and modulate the activity of distinct E2F/DP1 complexes in different phases of the cell cycle (11–16), thus suggesting distinct roles for these closely related proteins in the regulation of the cell cycle.

Another indication that these proteins play a pivotal role in cell cycle regulation is the interaction of the retinoblastoma family protein with several cyclin/cyclin-dependent kinase (cdk) complexes. pRb can be regulated by different cyclin/cdk complexes, such as cyclin A/cdk2, cyclin E/cdk2, and cyclin D/cdk4, even if stable interaction between pRb and cyclin A/cdk2 or cyclin A/cdk2 has not been found *in vivo*. (17). On the other hand, both p107 and pRb2/p130 stably interact *in vivo* with cyclin E/cdk2 and cyclin A/cdk2 complexes (3, 18, 19). These complexes may be responsible for the existence of different phosphorylated forms of pRb, p107, and pRb2/p130 in the various phases of the cell cycle (20–23). As pRb's functional activities are augmented by these phosphorylations, it is likely that p107 and pRb2/p130 are also affected in the same way by these posttranslational modifications. Because p107 and pRB2/p130 show similar but not redundant functional properties to pRb, it is possible to argue that they, like pRb, act as tumor suppressor genes. Our finding that the Rb2/p130 gene maps on the long arm of chromosome 16, which is a region frequently reported to show deletion or constitutional heterozygosity in several human tumors such as breast, ovarian, hepatocellular, and prostatic carcinomas, supports this hypothesis (24).

In this manuscript we have characterized the entire genomic structure of the human Rb2/p130 gene by determining the number and the exact sizes of the exons contained within the currently known cDNA sequence. We have also identified the nucleotide sequence immediately upstream of the starting codon. The genomic sequence of Rb2/p130 should be valuable for future studies aimed at understanding the molecular mechanisms that regulate its transcription and at delineating potential Rb2/p130 mutations in human tumors.

MATERIALS AND METHODS

Isolation and Characterization of Genomic Clones. To isolate the entire human Rb2/p130 gene, a human P1 genomic library was screened by Genome System, St. Louis. One of the clones provided by Genome System (no. 1437) contained a part of the Rb2/p130 gene, as determined by Southern blot analysis. To gain additional 5′ sequence of the Rb2/p130 gene, we screened a human placenta genomic DNA phage library (EMBL3 SP6/T7; Clontech) with a cDNA probe corresponding to the first 430 bp after the start codon of the published sequence (2–4). Screening was carried out following standard protocols. Of the two positive clones obtained, one (SCR3) contained the 5′-flanking region of the Rb2/p130 gene. Sequencing of the recombinant clones was carried out in part by automated DNA sequencing using the dideoxy terminator reaction chemistry for sequence analysis on the Applied Biosystems model 373A DNA sequencer and in part by using the dsDNA Cycle Sequencing System kit purchased from GIBCO/BRL, Gaithersburg, MD.

Cell Culture and RNA Extraction. The human HeLa (cervix epithelioid carcinoma) cell line was obtained from the American Type Culture Collection and maintained in culture in DMEM with 10% fetal calf serum at 37°C in a 10% CO_2-

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.

Abbreviation: cdk, cyclin-dependent kinase.

Data deposition: The sequence reported in this paper has been deposited in the GenBank data base (accession no. U53220).

*On leave of absence from the Department of Anatomic Pathology, School of Medicine, University of Naples "Federico II," Naples, Italy.
†To whom reprint requests should be addressed.

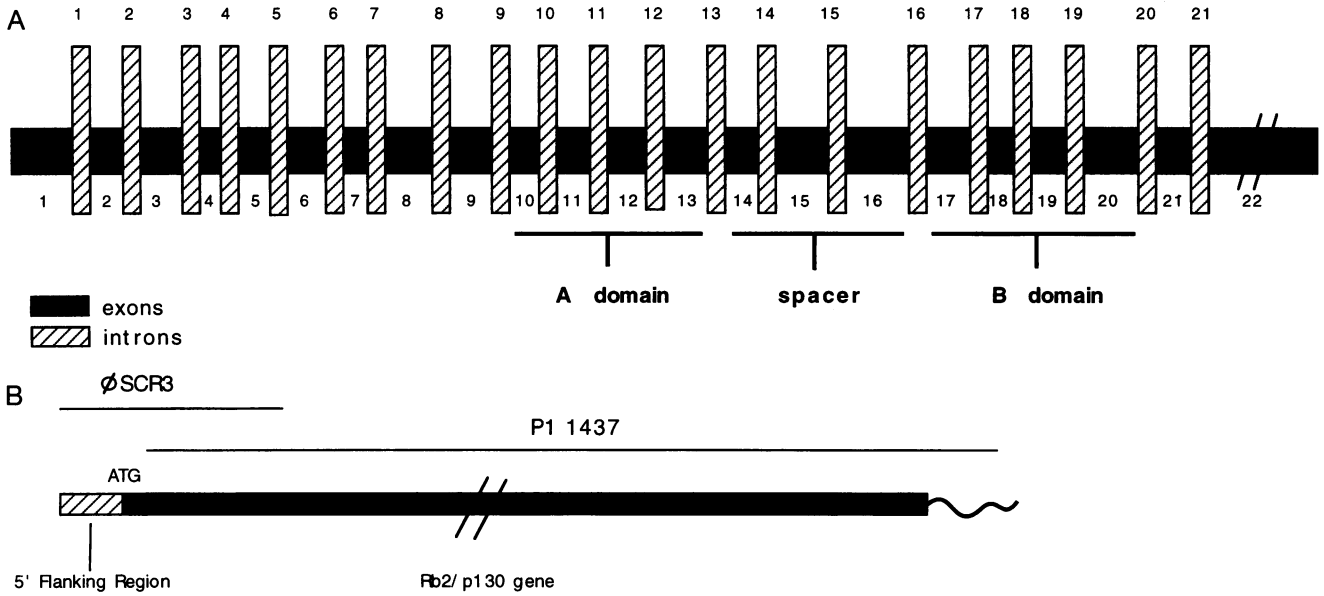

FIG. 1. (*A*) Structure of the human Rb2/p130 gene. Exons are represented by solid rectangles, drawn approximately to scale, and introns by hatched vertical bars. Also indicated are the positions of the A and B domains, which contain the pocket domain, in the Rb2/p130 gene. (*B*) Schematic representation of the human Rb2/p130 genomic clones derived from P1 or λ phage (ϕ) libraries.

containing atmosphere. Cytoplasmic RNA was extracted using the RNAzol B method (Cinna/Biotecx, Friendswood, TX).

Primer Extension Analysis. For the primer extension experiment, an avian myeloblastosis virus reverse transcriptase primer extension system (Promega) was used. Briefly, a 24-mer oligonucleotide 5′-ACCTCAGGTGAGGTGAGGGCCCGG-3′ complementary to the sequence starting at position −22 (see Fig. 3) was end-labeled with [γ-^{32}P]ATP and hybridized overnight with 20 μg of cytoplasmic RNA at 42°C. The primer-annealed RNA was converted into cDNA by avian myeloblastosis virus reverse transcriptase in the presence of 2 mM deoxynucleotides

Table 1. Intron–Exon boundaries of the human Rb2/p130 gene

Exon No. (bp)	5′ Donor sequence	3′ Acceptor sequence	Intron	No. (bp)
1 (<300)	ACGCTGGAG^{309}gtgcgctcgc	tcttttacag^{310}GGAAATGAT	1	(~4500)
2 (131)	AGAGCAGAG^{440}gtaactatgt	ttaataccag^{441}CTTAATCGA	2	(~3800)
3 (201)	GAAACAGCG^{641}gtaggttttc	tcccccaaag^{642}GCGACAGCC	3	(~3865)
4 (65)	ATGCAAAAG^{706}gtaagaaaat	aatcctgcag^{707}GTAATTTCC	4	(4576)
5 (129)	ATTTTAAAG^{835}gtaggtttgt	acaccatag^{836}GCTTATCTG	5	(1618)
6 (161)	GAAAAAAAG^{996}gtttgtaagt	ttcatcatag^{997}CTCCTTAAG	6	(92)
7 (65)	AGAGAGTTT^{1061}gtgagtactt	ttcctatag^{1062}TAAAGCCAT	7	(1046)
8 (187)	TTTGACAAG^{1248}gtgagtttag	ttttctttag^{1249}TCCAAAGCA	8	(~4500)
9 (167)	GATTCTCAG^{1415}gttagtttga	ccttttttag^{1416}GACATGTTC	9	(2127)
10 (90)	GTGCTAAAG^{1525}gtaattgtgc	atttctacag^{1526}AAATTGCCA	10	(716)
11 (104)	GATTTATCT^{1629}gtgagtaaaa	attttatag^{1630}GGTATTCTG	11	(~1000)
12 (138)	TTTTATAAG^{1767}gtatttccca	tttatttcag^{1768}GTGATAGAA	12	(1081)
13 (165)	TGTGAAGAG^{1932}gtgaaaatca	tcttcatag^{1933}GTCATGCCA	13	(1455)
14 (112)	TTGGAAGGA^{2044}gtaagtttaa	ttgacccctag^{2045}GCATAACAT	14	(2741)
15 (270)	CTGTGCAAG^{2314}gtaaggaagg	ctgtcactag^{2315}GTATTGCCA	15	(197)
16 (281)	TTTAGAAAG^{2595}gtaatttttc	tatctcctag^{2596}GTATACCAT	16	(82)
17 (177)	ATGGCAAAG^{2772}gtgagtacca	gtttgccag^{2773}GTCACAAAA	17	(1079)
18 (72)	CGGAGCCAG^{2844}gtaactacat	ttctctaaag^{2845}GTGTATAGA	18	(659)
19 (107)	AAGATAGAA^{2950}gtgggatctt	ctggctgcag^{2951}CCAGTAGAG	19	(572)
20 (202)	CAGGCAAAT^{3153}gtaagtatga	tttttaaacag^{3154}ATGGGATGC	20	(902)
21 (165)	CCTTCAAAG^{3318}gtgagcctaa	cccaccatag^{3319}AGACTGAGA	21	(~9000)
22 (1517)	to the polyadenylation signal			

Intron–exon organization of the human Rb2/p130 gene. Exon sequences are in capital letters; intron sequences are in lower case letters. Also shown are nucleotide positions (superscript numbers) of intron–exon boundaries according to the published sequence (3). Intron sizes were determined by either direct sequencing or amplification of intronic DNA within genomic clones by PCR with cDNA primers.

at 42°C for 45 min. The cDNA product was then analyzed on 7% sequencing gel containing 8 M urea.

RESULTS

Intron–Exon Organization of the Human Rb2/p130 Gene. The human Rb2/p130 gene consisted of 22 exons and spanned >50 kb of genomic DNA. The organization of the 22 exons along with the position of the two genomic clones encompassing all the exons and the 5′-flanking region of the Rb2/p130 gene is illustrated in Fig. 1. To precisely characterize the position of the exons and intron–exon boundaries in the genomic DNA, we used a set of oligonucleotides as primers for sequencing genomic DNA clones. A series of primers were synthesized from both strands at ≈150-bp intervals in the cDNA sequence. All of the intron–exon boundaries were identified in the positions where the genomic DNA sequence diverged from that of the cDNA. The DNA sequence of the introns was determined by using specific primers. Our systematic method of sequencing has enabled us to define the exact number of exons comprising the known cDNA sequence of Rb2/p130. The locations and sizes of each exon and intron, as well as nucleotide sequences at the intron–exon junctions, are illustrated in Table 1. All intron–exon boundaries conform to the GT/AG splice junction rule found in other human genes (25). All the exons were completely sequenced and no discrepancies were found when the genomic sequence of the exons and the previously known cDNA sequence were compared (2–4). With the exception of exon 22, which is 1517 bp long, exon sizes are relatively small; the shortest exons, exons 4 and 7, contain only 65 nt each. Some of the introns were completely sequenced, whereas the size of the others was determined by amplification of intron sequence within genomic clones by PCR with cDNA primers. The shortest intron is only 82 bp long and is located between exons 16 and 17, whereas the largest spans ≈9000 bp between exons 21 and 22. The pocket domain of Rb2/p130 protein, which is the region of highest homology with the retinoblastoma protein and is responsible for binding to viral oncoprotein or transcription factors (2), is contained in exons 10–13 (domain A) and exons 17–20 (domain B) of the genomic structure, whereas the spacer (the region between these two domains) is contained in exons 14, 15, and 16 (Fig. 2).

Analysis of the 5′-Flanking Promoter Sequence of the Rb2/p130 Gene. As described in *Materials and Methods*, a genomic DNA fragment containing ≈2 kb upstream of the start codon of the published sequence of the Rb2/p130 gene, was cloned by screening a human genomic DNA library with a probe spanning the first 430 bp downstream of the start codon of the cDNA sequence. Part of the sequence of the Rb2/p130 gene, along with the entire sequence of the first exon and beginning of the first intron, is shown in Fig. 2. Of note, the human Rb2/p130 gene promoter lacked canonical TATA or CAAT boxes. The first exon and the region immediately 5′ to this exon are very G+C-rich, with a G+C content of 66%. Also present are multiple binding motifs for the transcription factor Sp1 (26) and multiple consensus sequences for a CAP-site (27). All these features are reminiscent of "housekeeping" genes. The ubiquitous pattern of expression found for the Rb2/p130 mRNA further confirms this idea (7). An additional feature of the Rb2/p130 gene promoter is that the G+C-rich regions, particularly the 5′-CCCC-3′ motifs, may bind to the transcription factor ETF, which induces transcription of promoters lacking TATA boxes but enriched in these polycytosine stretches (26). Several other potential transcription factor-binding sites were identified by their similarity to consensus sequences for known transcription factor-binding sites on the basis of the SIGNAL SCAN program. The most recognizable sequence motifs are the ones showing homology to the transcription factors ker 1, which is involved in keratinocyte-specific transcription and MyoD, which is involved in myogenesis (29, 30).

Determination of the Transcription Start Site. As a first step toward characterizing the Rb2/p130 promoter, we used primer extension analysis to locate the transcription initiation site. A 24-mer oligonucleotide, XR, containing the antisense-strand sequence 26–50 nt upstream from the putative ATG codon (see Fig. 2) was end-labeled and used as a primer for extension reaction on cytoplasmic RNA from HeLa cells. As shown in Fig. 3, a major extended fragment of 78 bp was detected (lane 1); the additional bands detected by primer extension could represent additional initiation sites. This is consistent with a transcription initiation site 99 nt upstream of the start codon. In contrast, no product was observed when tRNA was used as template (lane 2). The probable position of the major transcription initiation site on the promoter sequence is indicated in Fig. 2. This analysis was repeated three times with identical results. Attempts to perform S1 or RNase protection assays were not successful, perhaps because of the high G+C content of the 5′ end of the gene.

DISCUSSION

In this communication we have analyzed the genomic organization of the human Rb2/p130 gene and characterized its 5′-flanking region. The human Rb2/p130 gene, like the retinoblastoma gene (31–33), has a complex organization. It consists of 22 exons scattered over a region of ≈50 kb. The 21 introns have different sizes, varying from 82 bp (intron 16) to ≈9000 bp (intron 21).

The lack of canonical TATA or CAAT boxes, the presence of a G+C-rich zone in the area immediately surrounding the major transcription initiation site, the existence of multiple consensus sequences for the Sp1 transcription factor and for CAP-site in the Rb2/p130 promoter, and the presence of multiple transcription start sites, are consistent with the characteristics associated with many housekeeping genes, as is the

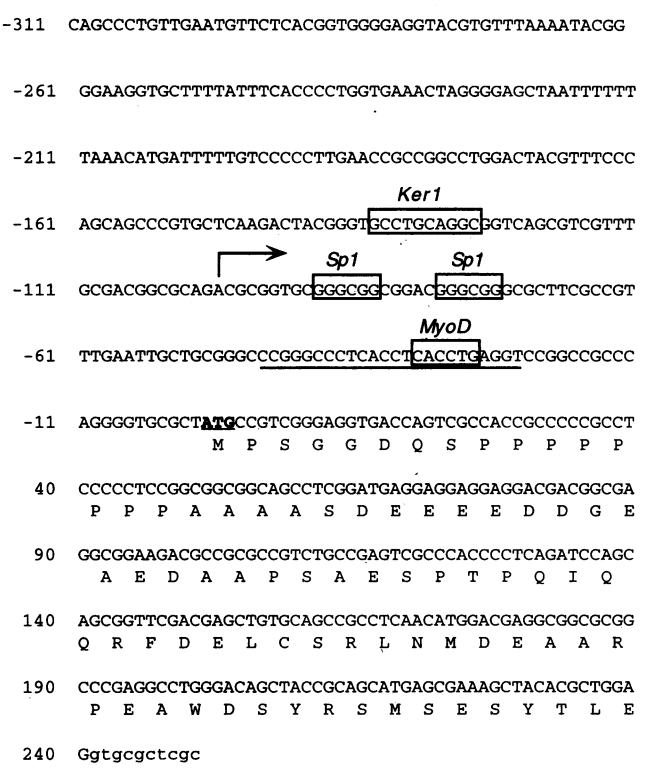

FIG. 2. Nucleotide sequence of the 5′ end and upstream region of the human Rb2/p130 gene. Major transcription start site is designated by an arrow, while position +1 is given to the ATG translation start (underlined and boldface). The sequence corresponding to the primer used for primer extension analysis is underlined. The GGGCGG sequences (at −90 and −79), corresponding to the Sp1 factor recognition motif, are boxed. Also boxed are the sequence motifs corresponding to MyoD and Ker 1 transcription factors. In lower cases are represented the nucleotides corresponding to the beginning of intron 1.

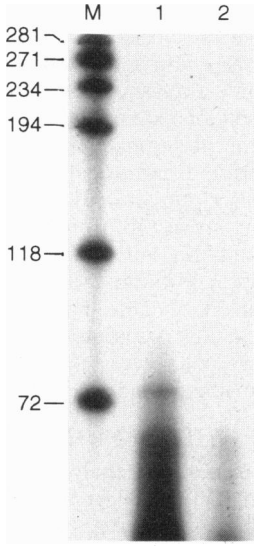

FIG. 3. Identification of the transcription initiation site of the human Rb2/p130 gene by primer extension analysis. Lane 1, primer-extended product using cytoplasmic RNA from HeLa cells (the extra bands detected by primer extension could represent additional initiation sites); and lane 2, primer extension using tRNA as template.

case for the two other members of the retinoblastoma gene family (31–34). The low amount of Rb2/p130 mRNA transcripts, together with the ubiquitous expression of the gene (7), are consistent with this idea. The presence of putative consensus sequences for some transcription factors, such as ker 1 and MyoD, that are involved in keratinocyte-specific transcription and myogenesis, respectively (29, 30), in the Rb2/p130 promoter, supports the hypothesis of an involvement of this gene in the complex pathways regulating differentiation of specific cell systems.

Because the RB2/p130 protein shows structural homology with pRB in the pocket domain and also binds to the transforming domain 2 of the E1A viral oncoprotein (35), one can argue that Rb2/p130 has tumor suppressor activity. Our finding that the Rb2/p130 gene maps on the long arm of chromosome 16, which is a region frequently reported to show deletion or constitutional heterozygosity in several human neoplasias, such as breast, ovarian, hepatocellular, and prostatic carcinomas (21), further supports this hypothesis. The genomic structure of the Rb2/p130 gene presented here will provide a foundation to identify the different mechanisms of its possible inactivation in tumors, including partial deletions, point mutations, and rearrangements, by using PCR directly on genomic DNA. These data will also be useful in identifying naturally occurring polymorphisms at the nucleotide level that may be utilized in diagnostic approaches.

Overall, our results provide the molecular basis for understanding the transcriptional control of the Rb2/p130 gene and for implementing a comprehensive Rb2/p130 mutation screen using genomic DNA as a template.

We acknowledge Dr. H. Alder for automated DNA sequencing and Dr. P. P. Pandolfi for the generous gift of the human placenta genomic library. We sincerely thank M. M. Pisano, C. M. Howard, N. Sang, and T. K. MacLachlan for critical review of the manuscript and P. Pertile for creating the graphs. The technical advice of V. Fidanza is also acknowledged. P.P.C. is supported by a fellowship from the American–Italian Cancer Foundation. A.D.L. is supported by Advanced Fellowship NATO-CNR. This work was supported by the Sbarro Institute for Cancer Research and Molecular Medicine, the Council for Tobacco Research, and National Institutes of Health Grant RO1 CA60999–01A1 to A.G.

1. Ewen, M. E., Xing, Y., Lawrence, J. B. & Livingston, D. M. (1991) *Cell* **66,** 1155–1164.
2. Mayol, X., Grana, X., Baldi, A., Sang, N., Hu, Q. & Giordano A (1993) *Oncogene* **8,** 2561–2566.
3. Li, Y., Graham, C., Lacy, S., Duncan, A. M. & Whyte, P. (1993) *Genes Dev.* **7,** 2366–2377.
4. Hannon, G. J., Demetrick, D. & Beach, D. (1993) *Genes Dev.* **7,** 2378–2391.
5. Goodrich, D. W., Wang, N. P., Qian, Y. W., Lee, E. Y. & Lee, W. H. (1991) *Cell* **67,** 293–302.
6. Zhu, L., Van den Heuvel, S., Helin, K., Fattaey, A., Ewen, M., Livingston, D. M., Dyson, N. & Harlow, E. (1993) *Genes Dev.* **7,** 1111–1125.
7. Claudio, P. P., Howard, C. M., Baldi, A., De Luca, A., Fu, Y., Condorelli, G., Sun, Y., Colburn, N., Calabretta, B. & Giordano, A. (1994) *Cancer Res.* **54,** 5556–5560.
8. Zhu, L., Enders, G., Lees, J. A., Beijersbergen, R. L., Bernards, R. & Harlow, E. (1995) *EMBO J.* **14,** 1904–1913.
9. Sang, N., Baldi, A. & Giordano, A. (1995) *Mol. Cell. Differ.* **3,** 1–29.
10. La Thangue, N. B. (1994) *Trends Biochem. Sci.* **19,** 108–114.
11. Chellapan, S. P., Hiebert, S., Mudryj, M., Horowitz, J. M. & Nevins, J. R. (1991) *Cell* **65,** 1053–1061.
12. Shirodkar, S., Ewen, M., De Caprio, J. A., Morgan, J., Livingston, D. M. & Chitteden, T. (1992) *Cell* **66,** 157–166.
13. Cobrinik, D., Whyte, P., Peeper, D. S., Jacks, T. & Weinberg, R. A. (1993) *Genes Dev.* **7,** 2392–2404.
14. Hijmans, E. M., Voorhoeve, P. M., Beijersbergen, R. L., Van'T Veer, L. J. & Bernards, R. (1995) *Mol. Cell. Biol.* **15,** 3082–3089.
15. Vairo, G., Livingston, D. M. & Ginsberg, D. (1995) *Genes Dev.* **9,** 869–881.
16. Jiang, H., Lin, J., Young, S., Goldstein, N. I., Waxman, S., Davila, V., Chellapan, S. P. & Fisher, P. B. (1995) *Oncogene* **11,** 1179–1189.
17. MacLachlan, T., Sang, N. & Giordano, A. (1995) *Eukaryotic Gene Expression* **5,** 127–156.
18. Ewen, M. E., Faha, B., Harlow, E. & Livingston, D. M. (1992) *Science* **255,** 85–87.
19. Faha, B., Ewen, M. E., Tsai, L. H., Livingston, D. M. & Harlow, E. (1992) *Science* **255,** 87–90.
20. Chen, P. L., Scully, P., Shew, J. Y., Wang, J. Y. J. & Lee, W. H. (1989) *Cell* **58,** 1193–1198.
21. De Caprio, J. A., Furukawa, Y., Ajchenbaum, F., Griffin, J. D. & Livingston, D. M. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 1795–1798.
22. Beijersbergen, R. L., Carlee, L., Kerkhoven, R. M. & Bernards, R. (1993) *Genes Dev.* **9,** 1340–1353.
23. Baldi, A., De Luca, A., Claudio, P. P., Baldi, F., Giordano, G. G., Tommasino, M., Paggi, M. G. & Giordano, A. (1995) *J. Cell. Biochem.* **59,** 402–408.
24. Yeung, R. S., Bell, D. W., Testa, J. R., Mayol, X., Baldi, A., Grana, X., Klinga- Levan, K., Knudson, A. G. & Giordano, A. (1993) *Oncogene* **8,** 3465–3468.
25. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383.
26. Jones, N. C., Rigby, P. W. & Ziff, E. B. (1988) *Genes Dev.* **2,** 267–281.
27. Kozak, M. (1992) *Annu. Rev. Cell Biol.* **8,** 197–225.
28. Kageyama, R., Merlino, G. T. & Pastan, I. (1988) *J. Biol. Chem.* **264,** 15508–15514.
29. Leask, A., Rosenberg, M., Vassar, R. & Fuchs, E. (1990) *Genes Dev.* **4,** 1985–1998.
30. Weintraub, H. (1993) *Cell* **75,** 1241–1244.
31. McGee, T. L., Yandell, D. W. & Dryja, T. P. (1989) *Gene* **80,** 119–128.
32. Hong, F. D., Huang, H. S., To, H., Young, L. J., Oro, A., Bookstein, R., Lee, E. Y.-H. & Lee, W. H. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 5502–5506.
33. T'Ang, A., Wu, K. J., Hashimoto, T., Liu, W. Y., Takahashi, R., Shi, X. H., Mihara, K., Zhang, F. H., Chen, Y. Y., Du, C., Qian, J., Lin, Y. G., Murphree, A. L., Qiu, W. R., Thompson, T., Benedict, W. F. & Fung, Y. K. (1989) *Oncogene* **4,** 401–407.
34. Zhu, L., Zhu, L., Xie, E. & Chang, L. S. (1995) *Mol. Cell. Biol.* **15,** 3552–3562.
35. Giordano, A., McCall, C., Whyte, P. & Franza, B. R., Jr. (1991) *Oncogene* **6,** 481–485.
36. Claudio, P. P., De Luca, A., Howard, C. M., Baldi, A., Firpo, E. J., Koff, A., Paggi, M. G. & Giordano, A. (1996) *Cancer Res.*, in press.