Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 9059–9063, December 1987
Genetics

Deletions of a DNA sequence in retinoblastomas and mesenchymal tumors: Organization of the sequence and its encoded protein

(recessive oncogenes/cancer/soft tissue sarcomas)

STEPHEN H. FRIEND*†‡, JONATHAN M. HOROWITZ*, MONICA R. GERBER*, XIAO-FAN WANG*, EMIL BOGENMANN§, FREDERICK P. LI¶, AND ROBERT A. WEINBERG*‖

*Whitehead Institute for Biomedical Research, Cambridge, MA 02142; †Division of Hematology–Oncology, The Children's Hospital, Dana–Farber Cancer Institute, Department of Pediatrics, Harvard Medical School, Boston, MA 02115; §Division of Hematology–Oncology, Children's Hospital of Los Angeles, Los Angeles, CA 90027; ¶Clinical Epidemiology Branch, National Cancer Institute, Boston, MA 02115; ‖Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

Contributed by Robert A. Weinberg, July 20, 1987

ABSTRACT **Retinoblastoma is a childhood tumor that can arise because of mutant alleles acquired as somatic or germinal mutations. The mutant allele can be carried in the germ line. The mutations creating these alleles act by inactivating copies of a recessive oncogene located within band q14 of chromosome 13 and termed the *RB1* locus. We have reported isolation of a cDNA fragment that recognizes chromosomal sequences possessing many of the attributes of the retinoblastoma gene associated with the *RB1* locus. We now report that this segment is additionally the target of somatic mutations in mesenchymal tumors among patients having no apparent predisposition to retinoblastoma and no previous evidence of retinoblastoma. These tumors provide additional evidence that the cloned sequences are representative of a gene that is a frequent target of inactivation during tumorigenesis. Sequence analysis of this cDNA provides little insight into its normal functional role.**

Retinoblastoma is a human childhood tumor that may occur because of inborn genetic susceptibility or acquired somatic mutations. The genetic elements involved in both susceptibility and somatic mutations have been mapped to the *RB1* locus found on human chromosome 13q14. Evidence from a variety of sources has suggested that inactivation of both alleles of the retinoblastoma gene is responsible for triggering tumor formation (1–3). Such data further suggest that the *RB1* locus represents an "anti-oncogene" that acts normally to constrain cellular proliferation. In its absence, the growth of retinoblasts is no longer correctly regulated, thereby resulting in the outgrowth of tumor cell clones.

We have reported (4) the isolation of a cDNA clone, termed p4.7R, that is homologous to a chromosomal sequence having many of the properties of the *RB* gene. Others using a probe described in our initial report have extended those initial observations (5). The p4.7R clone (4) was generated by reverse transcription of a 4.7-kilobase (kb) mRNA that is expressed in virus-immortalized human retinoblasts but was not detected in several retinoblastoma tumor cell lines. When used as a probe in Southern blot hybridization, the clone reacts with DNA sequences that are distributed over at least 180 kb of human chromosome 13 (J. Wiggs and T. Dryja, personal communication). Use of the p4.7R probe in analysis of 40 retinoblastomas revealed gross structural changes in chromosomal sequences in ≈30% of these tumors (4). In five of the retinoblastomas, the observed changes were present homozygously, thus affecting both homologous copies of this gene. In addition, a homozygous deletion in one osteosarcoma DNA provided direct evidence of a link between the inactivation of the p4.7R-homologous gene and tumor formation in an organ entirely separate from the retina. This connection with osteosarcomas had been considered because patients with retinoblastoma have a several hundred-fold increased risk of contracting osteosarcomas (6). Indeed, indirect genetic evidence regarding closely linked markers on chromosome 13q14 had led others to postulate the involvement of a common genetic element in the etiology of both types of tumor (7, 8). We now provide direct evidence of the frequent change of this gene in the genomes of nonretinoblastoma tumors as well as more detailed characterization of the cloned gene.**

MATERIALS AND METHODS

Southern Hybridization. DNA was isolated from human tumors and retinal cells as described (9). Restriction endonuclease digestion of DNA samples, agarose gel electrophoresis, Southern blotting, and hybridization were performed according to standard methods (33). The p4.7R plasmid was subcloned using an internal *Eco*RI site to yield 0.9-kb and 3.8-kb fragments that were isolated and used as probes. Radiolabeling of the p3.8R and p0.9R probes with [^{32}P]dCTP was performed using an oligo-labeling technique (10).

RNA Gel Blot Hybridization. Total cytoplasmic RNA was isolated using the Nonidet P-40 extraction technique, resolved by electrophoresis through a 1% agarose/formaldehyde gel, and transferred onto nitrocellulose as described by Schrier *et al.* (11). Radiolabeling of the probes with [^{32}P]dCTP was performed using an oligo-labeling technique (10).

Sequencing. Both strands of the entire p4.7R clone were sequenced using the dideoxy technique of Sanger *et al.* (12). A sequential series of overlapping clones was prepared in M13 bacteriophage (13). Additional sequencing of genomic DNA was used for confirmation of some regions. 7-Deaza-2′-deoxyguanosine 5′-triphosphate was used during sequence analysis for enhanced resolution of G+C-rich regions (14).

Sequence Management and Searching Programs. Overlapping DNA segments were aligned using the programs of Staden (15, 16). Homology searches (17) accessed the National Biomedical Research Foundation Data Base.††

RESULTS

Deletions in Mesenchymal Tumors and Retinoblastoma. Patients with an inherited susceptibility to retinoblastoma have a high incidence of second malignant tumors occurring primarily in mesenchymal tissues (6, 18–22). Almost two-thirds of these second site tumors are osteosarcomas or soft

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.

‡To whom reprint requests should be addressed.

**This sequence is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J02994).

††Protein Identification Resource (1987) Protein Sequence Database (Natl'l. Biomed. Res. Found., Washington, DC), Release 12.

tissue sarcomas. This suggests that inactivation of the *RB* locus affects tumor predisposition in tissues of mesenchymal origin. Accordingly, we have now surveyed the DNA from a number of tumor types, including those histologic types that are seen at second sites in patients initially treated for retinoblastoma. A listing of the tumor types is provided in Table 1. The particular tumors represented in our own sampling arose in patients having no previous retinoblastomas and no familial predisposition to such tumors.

As in the previous analysis of retinoblastoma DNAs (4), we detect both heterozygous and homozygous changes in the chromosomal segments reactive with the p4.7R probe. We confine our present analysis to the homozygous changes; heterozygous deletions of segments are observed with an ≈2-fold greater frequency, yet they often yield autoradiographic patterns that are less clearly interpreted.

Table 1 and Fig. 1 indicate that the gene recognized by the p4.7R probe is often altered in a homozygous fashion in osteosarcomas (3 out of 13 tumors) and several types of soft tissue sarcomas (3 out of 16 tumors). In contrast, DNA samples from other, nonmesenchymal tumor types surveyed here provide no evidence of gross homozygous structural changes in the sequences reactive with the p4.7R probe. Because the particular tumors studied here are not from patients with a known inborn predisposition to retinoblastoma, we conclude that the observed lesions affecting this gene occurred as a consequence of somatic mutation.

Only a portion of the genetic events that might inactivate the *RB* gene can be seen as alterations detectable by Southern blot analysis. The remainder, which may involve minor changes in DNA structure, represent inactivating events that escape detection by this method. Consequently, it is likely that many further, as yet undetected, inactivations of this gene are represented among the mesenchymal tumors listed in Table 1.

We note in passing the absence of obvious structural changes of this chromosomal segment in the DNAs of 14 breast carcinomas including 10 ductal carcinomas of the breast (studied in collaboration with Magnus Nordenskjold). The ductal breast carcinomas were of special interest because karyotypic analysis of tumor cells indicates specific deletion of a chromosomal segment mapping to 13q14.1, very close to the *RB1* locus (24). In the 3 ductal breast carcinomas carrying karyotypic changes in one of their chromosomes 13, we observe a heterozygous loss of the DNA within the entire region detected by the p4.7R clone. This apparently stems from loss of a chromosomal region encompassing both the p4.7R region and a large array of other genes. Such results do not necessarily indicate that the gene that we have cloned is involved in the formation of these breast carcinomas. It is equally plausible that these tumors derive from the inactivation of a locus that is closely linked to but distinct from the *RB* gene itself.

Table 1. Occurrence of various tumor types as second tumors in retinoblastoma patients and the frequency of homozygous deletions spanned by the p4.7R clone

Tumor type	% second tumors	Fraction of homozygous deletions
Retinoblastoma	—	7/49
Osteosarcoma	41	3/13
Soft tissue sarcomas	25	3/16
Fibrosarcoma		0/1
Malignant fibrous histiocytoma		1/2
Leiomyosarcoma		1/2
Liposarcoma		0/2
Rhabdomyosarcoma		0/3
Synovial cell sarcoma		0/2
Undifferentiated sarcoma		1/4
Melanoma	6	0/17
Brain tumors	2	—
Epithelial tumors and carcinomas	<2	0/17
Bladder carcinoma		0/1
Ductal breast carcinoma		0/14
Renal cell adenocarcinoma		0/2
Peripheral nerve tumors	<2	0/2
Neurosarcoma		0/1
Schwannoma		0/1
All other tumors	<2	—

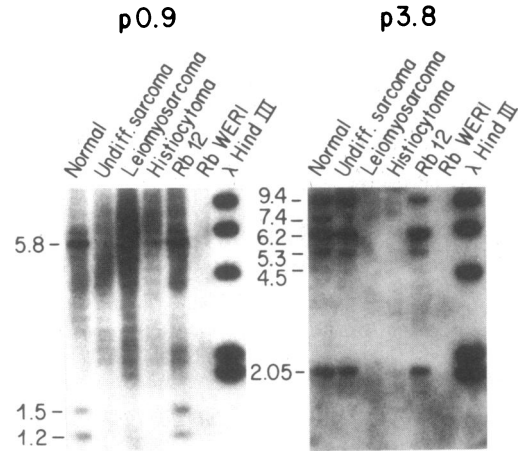

FIG. 1. Southern blot analysis of soft tissue sarcoma, retinoblastoma, and normal DNAs electrophoresed on a 1% agarose gel and probed with subclones of p4.7R. Lane Normal contains lymphocyte DNA from a normal healthy human. The adjacent lanes have DNA from the following tumors. Lanes: Undiff. sarcoma, highly undifferentiated sarcoma NCI 2596-01-006; Leiomyosarcoma, NCI 2596-01-008; Histiocytoma, NCI 2596-01-004; Rb12, retinoblastoma RBLA-12 established by E.B.; Rb WERI, retinoblastoma WERI, ref. 23; DNA, 4 μg per lane was digested with *Hin*dIII. *Hin*dIII-digested λ DNA is included as molecular weight markers. (*Left*) Subcloned 0.9-kb *Eco*RI fragment of p4.7R (ref. 1) was used as a probe. (*Right*) Subcloned 3.8-kb *Eco*RI fragment of p4.7R (ref. 1) was used as a probe. Molecular sizes are in kb.

Further Evidence for Involvement of the Cloned Sequence in Tumorigenesis. The analysis of soft tissue sarcoma DNA has also strengthened the association between sequences reactive with the p4.7R probe and the chromosomal segment that is a specific target of genetic inactivation during tumorigenesis. Our work has described (4) a number of deletions that affected either the entirety of the chromosomal domain detected by this probe or the right half of this domain as plotted in Fig. 2. Such deletions, each arising as a consequence of an independent mutational event, indicated either (*i*) that the cloned gene was the critical target of inactivation during retinoblastoma formation or (*ii*) that the target of inactivation is a separate gene that lies somewhere to the right on the chromosome. Two internal deletions, each beginning and ending within the confines of the cloned gene, provided some assurance that the cloned gene indeed encodes the function that must be inactivated to trigger tumorigenesis. These deletions were present in homozygous form in an osteosarcoma and in heterozygous form in a retinoblastoma (4).

Our data strengthen the earlier assignment. In an undifferentiated soft tissue sarcoma (Fig. 2), we observed a deletion that begins in the middle of the gene and extends beyond the end of the gene in a leftward direction. This militates against the presence of a target gene mapping to the right of the cloned segment (see above). Equally persuasive is a retinoblastoma DNA (Fig. 2), in which a totally internal deletion is present homozygously. The tumor cell line Rb-12 was established by E.B. from the same tumor as reported by Fung *et al.* (34). These results make it increasingly unlikely that the gene targeted for inactivation in these various tumors

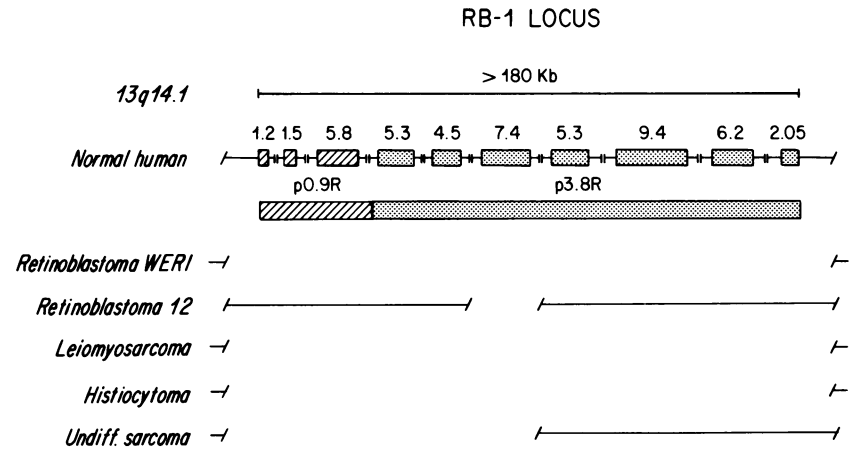

FIG. 2. Schematic ordering of deletions in the genomic DNA from two retinoblastomas and three soft tissue sarcomas. The genomic *Hin*dIII fragments detected by the 0.9- and 3.8-kb *Eco*RI fragments, denoted by hatched boxes and stippled boxes, respectively, are shown at the top according to their relative positions in the genome and labeled with sizes in kilobases. Normal and tumor DNAs, labeled as in Fig. 1, have solid lines where both copies are present. The gaps denote homozygously deleted segments.

lies anywhere except within the limits of the chromosomal region surveyed by the p4.7R cDNA clone.

RNA Gel Blot Analysis of Normal Retinal Cells and Retinoblastoma Cells. Supporting evidence that associates the p4.7R clone with the *RB* locus also came from study of RNA expression in retinal and retinoblastoma cells. Previously, we reported an inability to detect expression of p4.7R-reactive sequences in RNAs prepared from four retinoblastomas and an osteosarcoma (4). We speculated that this apparent lack of expression reflected the damaged state of the *RB* gene in these tumors. However, such absence of expression might alternatively reflect the derivation of retinoblastomas from an embryonic cell type that does not normally express this gene.

To address these possibilities, we have studied the RNAs prepared from normal retinal cells and from a group of retinoblastomas. We have performed this analysis using probes labeled to a higher specific activity than in our previous survey (10). As shown in Fig. 3, embryonic retinal cells at 19 weeks gestation synthesize significant amounts of the p4.7R transcript. These embryonic retinal cells are mitotically active *in vivo* and are capable of growth *in vitro* as well. They have marginally lower amounts of reactive RNA than are found in cells of the adult retina shown in the next lane. We conclude that retinoblastoma cells derive from a cell lineage that normally expresses detectable amounts of p4.7R-reactive RNA.

Our present analysis reveals that the expression of this RNA in 11 retinoblastomas is, in contrast, highly variable. The retinoblastomas have several different patterns of expression that can be summarized by the four examples shown in Fig. 3. Some retinoblastomas such as "WERI" cells (23) have no detectable transcript. Some tumors such as the Y79 cells, in which we previously detected no message, have a transcript of altered size, migrating as a 4.3-kb RNA. Other tumors show normal-sized transcripts expressed either at levels close to embryonic retinal cells (Rb-12) or at very low levels (Rb-22). Thus, our present analysis of tumor RNAs reveals that probe-reactive RNAs can indeed be found in some retinoblastoma samples. The patterns of expression and transcript size may well reflect the variety of molecular mechanisms through which the *RB* gene is inactivated.

Sequence Analysis. The p4.7R cDNA used in these hybridizations has been subjected to sequence analysis to determine the structure of its encoded protein. Both of the strands were subjected to complete sequencing by the Sanger dideoxy method (12), and the resulting sequence is shown in Fig. 4. Choice of the long open reading frame that specifies a protein of a maximum of 927 amino acids allowed us to establish a 5′→3′ orientation of the chromosomal DNA map. Either of two methionines at nucleotide positions 4 and 340 may initiate translation of the long open reading frame. The first of these conforms more closely to the initiator consensus sequence rules of Kozak (26).

Others using a probe described in our previous report (5) have isolated a similar cDNA and presented sequence data that is substantially similar to that presented here (5). We find 12 differences between the two sequences, the majority of these affecting untranslated regions. A major discrepancy is found between nucleotide residues 346 and 376 at the beginning of the open reading frame. This affects the coding of amino acid residues 116 through 123. To confirm our own sequence, we isolated and sequenced the chromosomal segment that is homologous to this portion of the cDNA. All analyses suggest that in this region the reported sequence (5) is shifted out of the correct reading frame and returns shortly thereafter to the correct frame. All other deviations involve single base-pair changes at the following nucleotide posi-

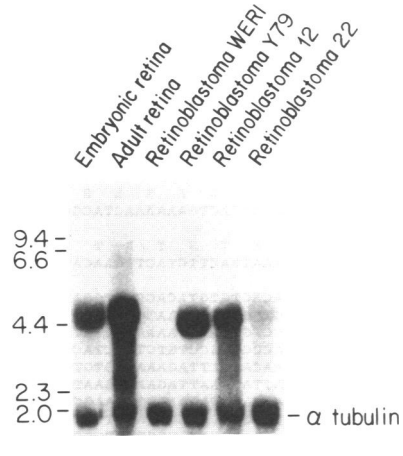

FIG. 3. RNA gel blot analysis of RNA isolated from retinal and retinoblastoma cells. Total cytoplasmic RNA (20 μg) was loaded from embryonic retinal cells (19 weeks of gestation), adult retinal cells, retinoblastoma WERI (10), retinoblastoma Y79 (13), retinoblastoma 12, and retinoblastoma 22. ^{32}P-labeled p3.8R insert was used as a probe. The filter was also hybridized with a probe derived from the rat α-tubulin locus that detects a homologous 1.7-kb message (25). Positions of molecular mass markers in kb are indicated on the left.

```
      M  P  P  K  T  P  R  K  T  A  A  T  A  A  A  A  A  A  E  P  P  A  P  P  P  P  P  P  P  E  E  D  P
    1 GTCATGCCGCCCAAAACCCCCCGAAAAACGGCCGCCACCGCCGCCGCTGCCGCCGCGGAACCCCCGGCACCGCCGCCGCCGCCCCCTCCTGAGGAGGACC  100

      E  Q  D  S  G  P  E  D  L  P  L  V  R  L  E  F  E  E  T  E  E  P  D  F  T  A  L  C  Q  K  L  K  I
  101 CAGAGCAGGACAGCGGCCCGGAGGACCTGCCTCTCGTCAGGCTTGAGTTTGAAGAAACAGAAGAACCTGATTTTACTGCATTATGTCAGAAATTAAAGAT  200

      P  D  H  V  R  E  R  A  W  L  T  W  E  K  V  S  S  V  D  G  V  L  G  G  Y  I  Q  K  K  K  E  L  W
  201 ACCAGATCATGTCAGAGAGAGAGCTTGGTTAACTTGGGAGAAAGTTTCATCTGTGGATGGAGTATTGGGAGGTTATATTCAAAAGAAAAAGGAACTGTGG  300

      G  I  C  I  F  I  A  A  V  D  L  D  E  M  S  F  T  F  T  E  L  Q  K  N  I  E  I  S  V  H  K  F  F  N
  301 GGAATCTGTATCTTTATTGCAGCAGTTGACCTAGATGAGATGTCGTTCACTTTTACTGAGCTACAGAAAAACATAGAAATCAGTGTCCATAAATTCTTTA  400

      L  L  K  E  I  D  T  S  T  K  V  D  N  A  M  S  R  L  L  K  K  Y  D  V  L  F  A  L  F  S  K  L  E
  401 ACTTACTAAAAGAAATTGATACCAGTACCAAAGTTGATAATGCTATGTCAAGACTGTTGAAGAAGTATGATGTATTGTTTGCACTCTTCAGCAAATTGGA  500

      R  T  C  E  L  I  Y  L  T  Q  P  S  S  S  I  S  T  E  I  N  S  A  L  V  L  K  V  S  W  I  T  F  L
  501 AAGGACATGTGAACTTATATATTTGACACAACCCAGCAGTTCGATATCTACTGAAATAAATTCTGCATTGGTGCTAAAAGTTTCTTGGATCACATTTTTA  600

      L  A  K  G  E  V  L  Q  M  E  D  D  L  V  I  S  F  Q  L  M  L  C  V  L  D  Y  F  I  K  L  S  P  P  M
  601 TTAGCTAAAGGGGAAGTATTACAAATGGAAGATGATCTGGTGATTTCATTTCAGTTAATGCTATGTGTCCTTGACTATTTTATTAAACTCTCACCTCCCA  700

      L  L  K  E  P  Y  K  T  A  V  I  P  I  N  G  S  P  R  T  P  R  R  G  Q  N  R  S  A  R  I  A  K  Q
  701 TGTTGCTCAAAGAACCATATAAAACAGCTGTTATACCCATTAATGGTTCACCTCGAACACCCAGGCGAGGTCAGAACAGGAGTGCACGGATAGCAAAACA  800

      L  E  N  D  T  R  I  I  E  V  L  C  K  E  H  E  C  N  I  D  E  V  K  N  V  Y  F  K  N  F  I  P  F
  801 ACTAGAAAATGATACAAGAATTATTGAAGTTCTCTGTAAAGAACATGAATGTAATATAGATGAGGTGAAAAATGTTTATTTCAAAAATTTTATACCTTTT  900

      M  N  S  L  G  L  V  T  S  N  G  L  P  E  V  E  N  L  S  K  R  Y  E  E  I  Y  L  K  N  K  D  L  D  A
  901 ATGAATTCTCTTGGACTTGTAACATCTAATGGACTTCCAGAGGTTGAAAATCTTTCTAAACGATACGAAGAAATTTATCTTAAAAATAAAGATCTAGATG 1000

      R  L  F  L  D  H  D  K  T  L  Q  T  D  S  I  D  S  F  E  T  Q  R  T  P  R  K  S  N  L  D  E  E  V
 1001 CAAGATTATTTTTGGATCATGATAAAACTCTTCAGACTGATTCTATAGACAGTTTTGAAACACAGAGAACACCACGAAAAAGTAACCTTGATGAAGAGGT 1100

      N  V  I  P  P  H  T  P  V  R  T  V  M  N  T  I  Q  Q  L  M  M  I  L  N  S  A  S  D  Q  P  S  E  N
 1101 GAATGTAATTCCTCCACACACTCCAGTTAGGACTGTTATGAACACTATCCAACAATTAATGATGATTTTAAATTCAGCAAGTGATCAACCTTCAGAAAAT 1200

      L  I  S  Y  F  N  N  C  T  V  N  P  K  E  S  I  L  K  R  V  K  D  I  G  Y  I  F  K  E  K  F  A  K  A
 1201 CTGATTTCCTATTTTAACAACTGCACAGTGAATCCAAAAGAAAGTATACTGAAAAGAGTGAAGGATATAGGATACATCTTTAAAGAGAAATTTGCTAAAG 1300

      V  G  Q  G  C  V  E  I  G  S  Q  R  Y  K  L  G  V  R  L  Y  Y  R  V  M  E  S  M  L  K  S  E  E  E
 1301 CTGTGGGACAGGGTTGTGTCGAAATTGGATCACAGCGATACAAACTTGGAGTTCGCTTGTATTACCGAGTAATGGAATCCATGCTTAAATCAGAAGAAGA 1400

      R  L  S  I  Q  N  F  S  K  L  L  N  D  N  I  F  H  M  S  L  L  A  C  A  L  E  V  V  M  A  T  Y  S
 1401 ACGATTATCCATTCAAAATTTTAGCAAACTTCTGAATGACAACATTTTTCATATGTCTTTATTGGCGTGCGCTCTTGAGGTTGTAATGGCCACATATAGC 1500

      R  S  T  S  Q  N  L  D  S  G  T  D  L  S  F  P  W  I  L  N  V  L  N  L  K  A  F  D  F  Y  K  V  I  E
 1501 AGAAGTACATCTCAGAATCTTGATTCTGGAACAGATTTGTCTTTCCCATGGATTCTGAATGTGCTTAATTTAAAAGCCTTTGATTTTTACAAAGTGATCG 1600

      S  F  I  K  A  E  G  N  L  T  R  E  M  I  K  H  L  E  R  C  E  H  R  I  M  E  S  L  A  W  L  S  D
 1601 AAAGTTTTATCAAAGCAGAAGGCAACTTGACAAGAGAAATGATAAAACATTTAGAACGATGTGAACATCGAATCATGGAATCCCTTGCATGGCTCTCAGA 1700

      S  P  L  F  D  L  I  K  Q  S  K  D  R  E  G  P  T  D  H  L  E  S  A  C  P  L  N  L  P  L  Q  N  N
 1701 TTCACCTTTATTTGATCTTATTAAACAATCAAAGGACCGAGAAGGACCAACTGATCACCTTGAATCTGCTTGTCCTCTTAATCTTCCTCTCCAGAATAAT 1800

      H  T  A  A  D  M  Y  L  S  P  V  R  S  P  K  K  K  G  S  T  T  R  V  N  S  T  A  N  A  E  T  Q  A  T
 1801 CACACTGCAGCAGATATGTATCTTTCTCCTGTAAGATCTCCAAAGAAAAAAGGTTCAACTACGCGTGTAAATTCTACTGCAAATGCAGAGACACAAGCAA 1900

      S  A  F  Q  T  Q  K  P  L  K  S  T  S  L  S  L  F  Y  K  K  V  Y  R  L  A  Y  L  R  L  N  T  L  C
 1901 CCTCAGCCTTCCAGACCCAGAAGCCATTGAAATCTACCTCTCTTTCACTGTTTTATAAAAAAGTGTATCGGCTAGCCTATCTCCGGCTAAATACACTTTG 2000

      E  R  L  L  S  E  H  P  E  L  E  H  I  I  W  T  L  F  Q  H  T  L  Q  N  E  Y  E  L  M  R  D  R  H
 2001 TGAACGCCTTCTGTCTGAGCACCCAGAATTAGAACATATCATCTGGACCCTTTTCCAGCACACCCTGCAGAATGAGTATGAACTCATGAGAGACAGGCAT 2100

      L  D  Q  I  M  M  C  S  M  Y  G  I  C  K  V  K  N  I  D  L  K  F  K  I  I  V  T  A  Y  K  D  L  P  H
 2101 TTGGACCAAATTATGATGTGTTCCATGTATGGCATATGCAAAGTGAAGAATATAGACCTTAAATTCAAAATCATTGTAACAGCATACAAGGATCTTCCTC 2200

      A  V  Q  E  T  F  K  R  V  L  I  K  E  E  E  Y  D  S  I  I  V  F  Y  N  S  V  F  M  Q  R  L  K  T
 2201 ATGCTGTTCAGGAGACATTCAAACGTGTTTTGATCAAAGAAGAGGAGTATGATTCTATTATAGTATTCTATAACTCGGTCTTCATGCAGAGACTGAAAAC 2300

      N  I  L  Q  Y  A  S  T  R  P  P  T  L  S  P  I  P  H  I  P  R  S  P  Y  K  F  P  S  S  P  L  R  I
 2301 AAATATTTTGCAGTATGCTTCCACCAGGCCCCCTACCTTGTCACCAATACCTCACATTCCTCGAAGCCCTTACAAGTTTCCTAGTTCACCCTTACGGATT 2400

      P  G  G  N  I  Y  I  S  P  L  K  S  P  Y  K  I  S  E  G  L  P  T  P  T  K  M  T  P  R  S  R  I  L  V
 2401 CCTGGAGGGAACATCTATATTTCACCCCTGAAGAGTCCATATAAAATTTCAGAAGGTCTGCCAACACCAACAAAAATGACTCCAAGATCAAGAATCTTAG 2500

      S  I  G  E  S  F  G  T  S  E  K  F  Q  K  I  N  Q  M  V  C  N  S  D  R  V  L  K  R  S  A  E  G  S
 2501 TATCAATTGGTGAATCATTCGGGACTTCTGAGAAGTTCCAGAAAATAAATCAGATGGTATGTAACAGCGACCGTGTGCTCAAAAGAAGTGCTGAAGGAAG 2600

      N  P  P  K  P  L  K  K  L  R  F  D  I  E  G  S  D  E  A  D  G  S  K  H  L  P  G  E  S  K  F  Q  Q
 2601 CAACCCTCCTAAACCACTGAAAAAACTACGCTTTGATATTGAAGGATCAGATGAAGCAGATGGAAGTAAACATCTCCCAGGAGAGTCCAAATTTCAGCAG 2700

      K  L  A  E  M  T  S  T  R  T  R  M  Q  K  Q  K  M  N  D  S  M  D  T  S  N  K  E  E  K  *
 2701 AAACTGGCAGAAATGACTTCTACTCGAACACGAATGCAAAAGCAGAAAATGAATGATAGCATGGATACCTCAAACAAGGAAGAGAAATGAGGATCTCAGG 2800

 2801 ACCTTGGTGGACACTGTGTACACCTCTGGATTCATTGTCTCTCACAGATGTGACTGTATAACTTTCCCAGGTTCTGTTTATGGCCACATTTAATATCTTC 2900
 2901 AGCTCTTTTTGTGGATATAAAATGTGCAGATGCAATTGTTTGGGTGATTCCTAAGCCACTTGAAATGTTAGTCATTGTTATTTATACAAGATTGAAAATC 3000
 3001 TTGTGTAAATCCTGCCATTTAAAAAGTTGTAGCAGATTGTTTCCTCTTCCAAAGTAAAATTGCTGTGCTTTATGGATAGTAAGAATGGCCCTAGAGTGGG 3100
 3101 AGTCCTGATAACCCAGGCCTGTCTGACTACTTTGCCTTCTTTTGTAGCATATAGGTGATGTTTGCTCTTGTTTTTATTAATTTATATGTATATTTTTTTA 3200
 3201 ATTTAACATGAACACCCTTAGAAAATGTGTCCTATCTATCTTCCAAATGCAATTTGATTGACTGCCCATTCACCAAAATTATCCTGAACTCTTCTGCAAA 3300
 3301 AATGGATATTATTAGAAATTAGAAAAAAATTACTAATTTTACACATTAGATTTTATTTTACTATTGGAATCTGATATACTGTGTGCTTGTTTTATAAAAT 3400
 3401 TTTGCTTTTAATTAAATAAAAGCTGGAAGCAAAGTATAACCATATGATACTATCATACTACTGAAACAGATTTCATACCTCAGAATGTAAAAGAACTTAC 3500
 3501 TGATTATTTTCTTCATCCAACTTATGTTTTTAAATGAGGATTATTGATAGTACTCTTGGTTTTTATACCATTCAGATCACTGAATTTATAAAGTACCCAT 3600
 3601 CTAGTACTTGAAAAAGTAAAGTGTTCTGCCAGATCTTAGGTATAGAGGACCCTAACACAGTATATCCCAAGTGCACTTTCTAATGTTTCTGGGTCCTGAA 3700
 3701 GAATTAAGATACAAATTAATTTTACTCCATAAACAGACTGTTAATTATAGGAGCCTTAATTTTTTTTTCATAGAGATTTGTCTAATTGCATCTCAAAATT 3800
 3801 ATTCTGCCCTCCTTAATTTGGGAAGGTTTGTGTTTTCTCTGGAATGGTACATGTCTTCCATGTATCTTTTGAACTGGCAATTGTCTATTTATCTTTTATT 3900
 3901 TTTTTAAGTCAGTATGGTCTAACACTGGCATGTTCAAAGCCACATTATTTCTAGTCCAAAATTACAAGTAATCAAGGGTCATTATGGGTTAGGCATTAAT 4000
 4001 GTTTCTATCTGATTTTGTGCAAAAGCTTCAAATTAAAACAGCTGCATTAGAAAAAGAGGCGCTTCTCCCCTCCCCTACACCTAAAGGTGTATTTAAACTA 4100
 4101 TCTTGTGTGATTAACTTATTTAGAGATGCTGTAACTTAAAATAGGGGATATTTAAGGTAGCTTCAGCTAGCTTTTAGGAAAATCACTTTGTCTAACTCAG 4200
 4201 AATTATTTTTAAAAAGAAATCTGGTCTTGTTAGAAAACAAAATTTTATTTTGTGCTCATTTAAGTTTCAAACTTACTATTTTGACAGTTATTTTGATAAC 4300
 4301 AATGACACTAGAAAACTTGACTCCATTTCATCATTGTTTCTGCATGAATATCATACAAATCAGTTAGTTTTTAGGTCAAGGGCTTACTATTTCTGGGTCT 4400
 4401 TTTGCTACTAAGTTCACATTAGAATTAGTGCCAGAATTTTAGGAACTTCAGAGATCGTGTATTGAGATTTCTTAAATAATGCTTCAGATATTATTGCTTT 4500
 4501 ATTGCTTTTTTTGTATTGGTTAAAACTGTACATTTAAAATTGCTATGTTACTATTTTCTACAATTAATAGTTTGTCTATTTTAAAATAAATTAGTTGTTAG 4600
```

FIG. 4. Nucleotide sequence of p4.7R and the deduced amino acid sequence. The 3'-untranslated region extends from position 2788 to position 4600.

tions: 321, 322, 997, 998, 1102, 1146, 2858, 2945, 4036, and 4038. We do not know whether these particular differences represent sequencing artifacts or polymorphisms of this gene that may be present in the human gene pool.

The polypeptide encoded by the open reading frame was analyzed by searching for homologies with other proteins recorded in the Dayhoff data base†† using the Lipman–Pearson protein sequence similarity program (17). We found no closely related sequences.

The p4.7R sequence was also analyzed by searching for regions similar to functional domains of other well-characterized proteins. A possible association of this protein with membranes was tested by searching for a consensus signal domain following the rules of Kyte and Doolittle (27) and by searching for clustered consensus glycosylation sites (28). Neither of these searches found evidence to support a membrane localization of this protein. We also searched for evidence of a consensus ATP-binding domain (29) and Ca^{2+}-binding sites (30) but found none. Although it has been suggested that there are potential metal ion/DNA-binding domains in this protein (5), we find no evidence of such domains using the consensus rules set out in the work of Berg (31). The p4.7R sequence was also compared with several specific groups of proteins having well-studied functions. These groups included the set of known dominant oncogenes, the homeobox genes, and the consensus sequence pattern shared among transforming growth factor β, the inhibins, Mullerian inhibitory substance, and the protein encoded by the *Drosophila* gene associated with decapentaplegia (32). None of these searches indicated any significant homologies. We, therefore, consider p4.7R to represent a gene encoding a protein of an uncharacterized type.

DISCUSSION

Inactivation of the p4.7R-homologous gene, occurring somatically, appears to be involved in the genesis of several mesenchymal tumor types in addition to retinoblastoma. These mesenchymal tumors account for a majority of the tumors that occur at second sites in patients with an inherited predisposition to retinoblastoma. We observed homozygous gross structural changes of this gene present in ≈20% of such DNAs tested (Table 1). Because our analysis excluded heterozygous lesions as well as subtle genetic changes that may escape detection by Southern blotting, we consider it possible that the majority of such mesenchymal tumors may be triggered by somatic inactivation of the gene recognized by this probe. This evokes the question of the apparent connection between the mechanisms of growth regulation of neuronal-type cells represented by retinoblastomas and those of the distantly related mesenchymal cells.

The present results do not address whether these observed gene inactivations are sufficient for tumorigenesis or represent only one of several genetic changes required during tumor progression to achieve full malignant transformation. Moreover, they do not address the reasons why inactivation of a widely expressed gene (4) results in only a narrow spectrum of tumor types; and they provide no clues as to why these mesenchymal-cell tumors appear far later in life than the retinal tumors.

While the data presented here provide additional support for identifying the p4.7R-homologous sequence as the *RB* gene, the ultimate proof of this identity must come from a functional test. Specifically, we would hope that introduction of expression constructs containing the p4.7R clone into retinoblastoma cells will cause those cells to lose at least some of their cancer-specific phenotypes. Success in such an experiment would validate the central role of this cloned gene in retinoblastoma tumorigenesis.

We are grateful to Thaddeus P. Dryja, Sang-Ho Park, John E. Price, Joyce M. Rapaport, and Jennifer M. Joyce for their collaboration and skilled assistance; Daniel M. Albert, Alfred G. Knudson, and Rene Bernards for helpful comments. Magnus Nordenskjold kindly probed the breast carcinomas and Nicholas Dracopoli generously provided DNA from the melanomas. S.H.F. is a Lucille P. Markey Scholar and R.A.W. is an American Cancer Society Research Professor. This work was supported by Grant CA 39826 from the National Institutes of Health, EY 05731 from the National Eye Institute, a grant from the American Business Foundation for Cancer Research, and grants from the Rowland Foundation and Donaldson Trust.

1. Murphree, A. L. & Benedict, W. F. (1984) *Science* **223,** 1028–1033.
2. Cavenee, W. K., Dryja, T. P., Phillips, R. A., Benedict, W. F., Godbout, R., Gallie, B. L., Murphree, A. L., Strong, L. C. & White, R. L. (1983) *Nature (London)* **305,** 779–783.
3. Dryja, T. P., Rapaport, J. M., Joyce, J. M. & Petersen, R. A. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 7391–7394.
4. Friend, S. H., Bernards, R., Rogelj, S., Weinberg, R. A., Rapaport, J. M., Albert, D. M. & Dryja, T. P. (1986) *Nature (London)* **323,** 643–646.
5. Lee, W. H., Bookstein, R., Hong, F., Young, L.-J., Shen, J.-Y. & Lee, E. Y.-H. P. (1987) *Science* **235,** 1394–1399.
6. Abramson, D. H., Ellsworth, R. M., Kitchin, D. & Tung, G. (1984) *Ophthalmology* **91,** 1351–1355.
7. Dryja, T. P., Rapaport, J. M., Epstein, J., Goorin, A. M., Weichselbaum, R., Koufos, A. & Cavenee, W. K. (1986) *Am. J. Genet.* **38,** 59–66.
8. Hansen, M. F., Koufos, A., Gallie, B. L., Phillips, R. A., Fodstad, O., Brogger, A., Gedde-Dahl, T. & Cavenee, W. K. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 6216–6220.
9. Aldridge, J., Kunkel, L., Bruns, G., Tantravahi, U., Lalande, M., Brewster, T., Moreau, E., Wilson, M., Bromley, W., Roderick, T. & Latt, S. A. (1984) *Am. J. Hum. Genet.* **36,** 546–564.
10. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13.
11. Schrier, P. I., Bernards, R., Vaessen, R. T. M. J., Houweling, A. & van der Eb, A. J. (1983) *Nature (London)* **305,** 771–775.
12. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
13. Dale, R. M. K., McClure, B. A. & Houchins, J. P. (1985) *Plasmid* **13,** 31–40.
14. Barr, P. J., Thayer, R. M., Laybourn, P., Najarian, R. C., Seela, F. & Tolan, D. R. (1986) *BioTechniques* **4,** 428–432.
15. Staden, R. (1977) *Nucleic Acids Res.* **4,** 4037–4051.
16. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395.
17. Lipman, D. J. & Pearson, W. R. (1985) *Science* **227,** 1435–1441.
18. Jensen, R. D. & Miller, R. W. (1971) *N. Engl. J. Med.* **285,** 307–311.
19. Francois, J., deSutter, E., Coppieters, R. & deBie, S. (1980) *Ophthalmologica* **181,** 93–99.
20. Meadows, A. T., Baum, E., Fossati-Bellani, F., Green, D., Jenkin, R. D. T., Marsden, B., Nesbit, M., Newton, W., Oberlin, O., Sallan, S. G., Siegel, S., Strong, L. C. & Voute, P. A. (1985) *J. Clin. Oncol.* **3,** 532–538.
21. DerKinderen, D. J., Koten, J. W., DenOtter, W., Tan, K. E. W. P. & Beemer, F. A. (1986) *Lancet* **ii,** 1335.
22. Draper, G. J., Sanders, B. M. & Kingston, J. E. (1986) *Br. J. Cancer* **53,** 661–671.
23. McFall, R. C., Sery, T. W. & Makadon, M. (1977) *Cancer Res.* **37,** 1003–1010.
24. Lundberg, C., Skoog, L., Cavenee, W. J. & Nordenskjold, M. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 2372–2376.
25. Lemischka, I. R., Farmer, S., Racaniello, V. R. & Sharp, P. A. (1981) *J. Mol. Biol.* **150,** 101–120.
26. Kozak, M. (1984) *Nucleic Acids Res.* **12,** 857–872.
27. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
28. Wagh, P. V. & Bahl, O. P. (1981) *CRC Crit. Rev. Biochem.* **10,** 307–377.
29. Hunter, T. & Cooper, J. A. (1985) *Annu. Rev. Biochem.* **54,** 897–930.
30. Szebenyi, D. M. E. & Moffat, K. (1986) *J. Biol. Chem.* **261,** 8761–8777.
31. Berg, J. M. (1986) *Science* **232,** 485–487.
32. Padgett, R. W., St. Johnnston, R. D. & Gelbart, W. M. (1987) *Nature (London)* **325,** 81–84.
33. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
34. Fung, Y.-K. T., Murphree, A. L., T'Ang, A., Qian, J., Hinrichs, S. H. & Benedict, W. F. (1987) *Science* **236,** 1657–1661.