The *GLI*-Kruppel Family of Human Genes

JOHN M. RUPPERT,[1] KENNETH W. KINZLER,[1,2] ALBERT J. WONG,[1] SANDRA H. BIGNER,[3] FA-TEN KAO,[4]
MARTHA L. LAW,[4] HECTOR N. SEUANEZ,[5] STEPHEN J. O'BRIEN,[5] AND BERT VOGELSTEIN[1]*

*The Oncology Center[1] and Department of Pharmacology and Molecular Sciences,[2] Johns Hopkins University School of
Medicine, Baltimore, Maryland 21231; Department of Pathology, Duke University Medical Center, Durham, North
Carolina 27710[3]; Eleanor Roosevelt Institute for Cancer Research and School of Medicine, University of Colorado Health
Sciences Center, Denver, Colorado 80262[4]; and Laboratory of Viral Carcinogenesis, National Cancer Institute,
Frederick, Maryland 21701[5]*

Previous characterization of *GLI*, a gene found to be amplified and expressed in a subset of human brain tumors, revealed the presence of five tandem zinc fingers related to those of Krüppel (*Kr*), a *Drosophila* segmentation gene of the gap class. We have used the *GLI* cDNA as a molecular probe to isolate related sequences from the human genome. Partial characterization of six related loci, including sequence determination, expression studies, and chromosome localization, revealed that each locus could encode a separate finger protein. The predicted proteins all had similar H-C links, i.e., a conserved stretch of 9 amino acids connecting the C-terminal histidine of one finger to the N-terminal cysteine of the next. On the basis of amino acid sequence and intron-exon organization, the genes could be placed into one of two subgroups: the *GLI* subgroup (with the consensus finger amino acid sequence [Y/F]XCX$_3$GCX$_3$[F/Y]X$_5$LX$_2$HX$_{3-4}$H[T/S]GEKP) or the *Kr* subgroup (with the consensus finger amino acid sequence [Y/F]XCX$_2$CX$_3$FX$_5$LX$_2$HXRXHTGEKP). Unlike *GLI* or *Kr*, most of the newly isolated genes were expressed in many adult tissues. The predicted proteins probably control the expression of other genes and, by analogy with *Kr* and *GLI*, may be important in human development, tissue-specific differentiation, or neoplasia.

One of the principal objectives of molecular biology is to gain an understanding of mechanisms resulting in specific temporal and spacial patterns of gene expression. Among the proteins important in this regard are those containing distinct DNA-binding regions. In *Drosophila melanogaster*, two classes of such proteins have been identified, containing either homeo box regions or zinc fingers in their putative DNA-binding domains. Homeo box regions were first detected in the antennapedia and fushi tarazu genes and have subsequently been identified in other *Drosophila* developmental genes (19). The homeo box elements are conserved in genes of a diverse array of species, and recent experiments have indicated a role for such proteins in mammalian embryonic development (for a review, see reference 20).

Zinc fingers were first identified in the *Xenopus* transcription factor TFIIIA (6, 37). Such fingers have been proposed to bind specific nucleic acid sequences while tetrahedrally coordinating a metal ion (zinc) via conserved cysteine and histidine residues (23, 37). Zinc fingers have been found in many regulatory genes (2, 16, 32, 58, 62). A family of zinc finger genes related to Krüppel (*Kr*) is particularly relevant to the present work. The *Drosophila* gene *Kr* is a member of the gap class of segmentation genes, and thoracic and anterior abdominal segments fail to form in *Kr* mutant embryos (44). *Kr* encodes a chromatin-associated phosphoprotein (41) which contains five zinc fingers with the consensus sequence (Y/F)XCX$_2$CX$_3$FX$_5$LX$_2$HX$_3$HTGEKP, in which X can be any amino acid (46). These consensus features, including the H-C link (the amino acid sequence HTGEKP(Y/F)XC connecting the histidine of one finger to the cysteine of the next), define the Krüppel family of zinc finger genes (49). Conservation of the contiguous stretch of

nucleotides encoding the H-C link has allowed cloning of Krüppel-related genes from *D. melanogaster*, mouse, and frog by hybridization with Krüppel cDNA at low stringency (10, 47, 49). Each of these Krüppel family members has been shown to be expressed in embryonic cells, suggesting a role for them in normal development.

Our interest in this family of genes was sparked by the identification of the *GLI* gene as a member of the Krüppel family (31). The *GLI* gene was discovered by virtue of its amplification in a subset of human brain tumors (30). Sequencing of *GLI* cDNA clones revealed the presence of five tandem fingers connected by H-C links similar to those of Krüppel. The fact that a structural motif proposed to mediate sequence-specific nucleic acid binding is found both in *Drosophila* developmental genes and in a gene implicated in human neoplasia suggested that other genes of this class might prove important in normal or disease states. Indeed, other genes important in neoplasia, such as *N-NYC* (33, 50), *L-MYC* (38), *HER-2* (12, 29), and *N-RAS* (54), were identified partly through their homology to previously identified oncogenes. Similarly, genes of potential importance in development have been identified through sequence similarity to known developmental genes (e.g., the *HOX* family of genes in mammals; 11, 34, 35). We have therefore used a *GLI* cDNA fragment which encodes the finger region to isolate related human sequences. Six distinct loci were identified in this manner and shown to be present on five different chromosomes. Partial sequencing revealed that each had open reading frames capable of encoding fingers with H-C links. Unlike Krüppel family genes previously identified in other species, most of these sequences were expressed in several adult tissues. Further study of these genes may provide insights into transcriptional mechanisms, normal human development, and/or neoplasia.

---

* Corresponding author.

MATERIALS AND METHODS

Library construction. Genomic DNA was purified from mouse xenografts of the glioblastoma multiforme cell line D320MG (25a) as described previously (61). After partial *Mbo*I digestion, DNA was size fractionated by sucrose density gradient ultracentrifugation. The fractions containing 17- to 24-kilobase fragments were cloned into the *Bam*HI site of Lambda Fix (Stratagene) after partial fill in of *Mbo*I ends according to the instructions of the manufacturer. The ligation product was packaged with lambda phage extracts (Stratagene) and used to infect *Escherichia coli* C600 cells. DNA from the resulting plaques was lifted with Colony Plaque Screen nylon membrane (Dupont, NEN Research Products) and bound by treatment with 0.4 M NaOH for 15 min (45). The filters were hybridized to a *GLI* finger probe (pGLIMBD) containing a 395-base-pair fragment of *GLI* cDNA spanning fingers 2 to 5 and containing nucleotides 873 to 1267 (31). Three rounds of plaque selection via hybridization and subsequent replating were used before purified DNA from individual plaques was analyzed.

DNA hybridization. After digestion with *Eco*RI or *Hind*III, genomic or phage DNA fragments were separated by electrophoresis through 1% agarose gels and transferred to nylon as described (45). DNA was labeled with ^{32}P by oligo labeling (17) and hybridized as described (61), except that nonfat dried milk was used (0.5%) (26), and 10% polyethylene glycol was included. For reduced-stringency experiments, Southern blots or plaque lifts were hybridized at 55°C and washed at 55 to 60°C in 0.3× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) and 0.3% sodium dodecyl sulfate (SDS) for 1 h. For normal stringency, the same wash solution was used but at 65°C for 1 h.

Plasmid subcloning. For isolation of subclones from recombinant bacteriophage, *Eco*RI, *Hind*III, or *Eco*RI-*Sal*I fragments which hybridized to pGLIMBD were separated from other fragments by electrophoresis and eluted from the gel (60). (*Sal*I sites are adjacent to the *Bam*HI cloning sites in Lambda Fix). Fragments were subcloned in alkaline phosphatase-treated Bluescript M13 + KS vector (Stratagene). For isolation of finger-containing sequences in smaller fragments for sequencing, these plasmid subclones were digested with several restriction endonucleases which had 4-bp recognition sites. The fragments were separated by electrophoresis in 1.5% agarose gels and studied by Southern blot analysis by using the pGLIMBD probe. Hybridizing fragments were eluted from agarose gels, and the ends were filled in with the Klenow fragment of *Pol*I. These blunt-ended fragments were ligated into the *Eco*RV site of Bluescript KS. Resulting clones were digested with *Eco*RI and *Hind*III, and the inserts were recloned into the same sites of Bluescript SK for sequencing in the opposite direction.

Thus, for each of the six loci indicated in Fig. 1, three sets of clones were generated: the original phage clone, a plasmid subclone with a large insert containing the finger region (Fig. 1), and small subclones of these plasmids which contained the pGLIMBD hybridizing regions and which were used for sequencing (Table 1). Two phages contained sequences that hybridized to non-zinc finger regions of the *GLI* cDNA probe (see text). Subcloning and sequencing of these regions were done in the same way as described above for the finger regions. The large subclone of GLI2 which hybridized to the upstream probe of *GLI* contained the 1.0-kilobase *Eco*RI fragment located 7.6 kilobases to the left of pGLI2RR (Fig. 1).

Sequencing. Sequencing was performed by the chain ter-

mination method with modified T7 polymerase (53). Single-strand templates were obtained from subclones in Bluescript by using f1 helper phage R408 (48). Most sequences were obtained by using data from both strands. The labeling and electrophoresis conditions we employed precluded sequencing of the terminal 10 to 15 bases at the end of each plasmid insert.

Ribonuclease protection. Total RNA was isolated by the acid-guanidium extraction method described by Chomczynski and Sacchi (9). ^{32}P-labeled RNA transcripts were generated in vitro from the sequencing subclones described above by using T3 or T7 RNA polymerase. Ribonuclease protection was performed as described (63) with the following modifications: hybridizations were performed in a final volume of 10 μl; only RNase A at 12.5 μg per ml was used; and the RNase A and proteinase K digestions were performed at room temperature for 30 min. Tera-1 cells (18) were obtained from the American Type Culture Collection. NTera-2 was generously provided by P. Andrews (1).

Chromosome localization. Southern blot analysis of human-rodent somatic cell hybrid DNA (8, 28, 39, 40) was performed by using inserts from the sequencing subclones described above.

RESULTS

Isolation of *GLI*-related sequences. To identify *GLI*-related sequences, a fragment of *GLI* cDNA encoding fingers two to five was subcloned (pGLIMBD) as described in Materials and Methods. Reduced stringency hybridization with this probe to Southern blots of human DNA revealed several bands not seen at normal stringency (not shown). To clone these sequences, 800,000 recombinant bacteriophage (4.5 genome equivalents) from a human genomic library were screened by using the same reduced stringency conditions. Fourteen phages were identified through this screen. Analysis of phage DNA revealed seven unique restriction maps. One set of clones represented the *GLI* locus, and the others represented six new loci. Comparison of the sizes of hybridizing fragments in the phage clones with those seen in genomic Southern blots revealed that most of the sequences detected in the Southern blots had been cloned.

***GLI*-related sequences encoded fingers similar to *GLI* or *Kr*.** Representative phages from each of the six loci defined by restriction mapping were chosen for further study (Fig. 1). For each of the loci, regions of the phage which hybridized to pGLIMBD were subcloned in plasmid vectors. Sequencing of these subclones revealed open reading frames encoding fingers with H-C links for each of the six loci (Fig. 2). No in-frame stop codons were found within the fingers. Intron-exon junctions were predicted adjacent to the finger regions on the basis of consensus splice sequences (51).

Analysis of the predicted amino acid sequences showed that the clones could be placed into one of two subgroups. Two of the six loci (*GLI2* and *GLI3*) encoded fingers that were very similar to those of *GLI* and *MGLI* (the mouse homolog of *GLI*). *GLI2* and *GLI3* had 89% and 84% of their residues in common with *GLI* in the finger regions, respectively (Fig. 3A). Over the 50-amino-acid region for which *GLI*, *GLI2*, and *GLI3* sequences were all available (i.e., from the middle of finger 1 to the amino terminus of finger 3), *GLI2* and *GLI3* were more similar to each other (92% amino acid identity) than either was to *GLI* (84% amino acid identity for each).

Alignment of *GLI2*, *GLI3*, and *MGLI* genomic sequences with the *GLI* cDNA sequence and identification of consen-
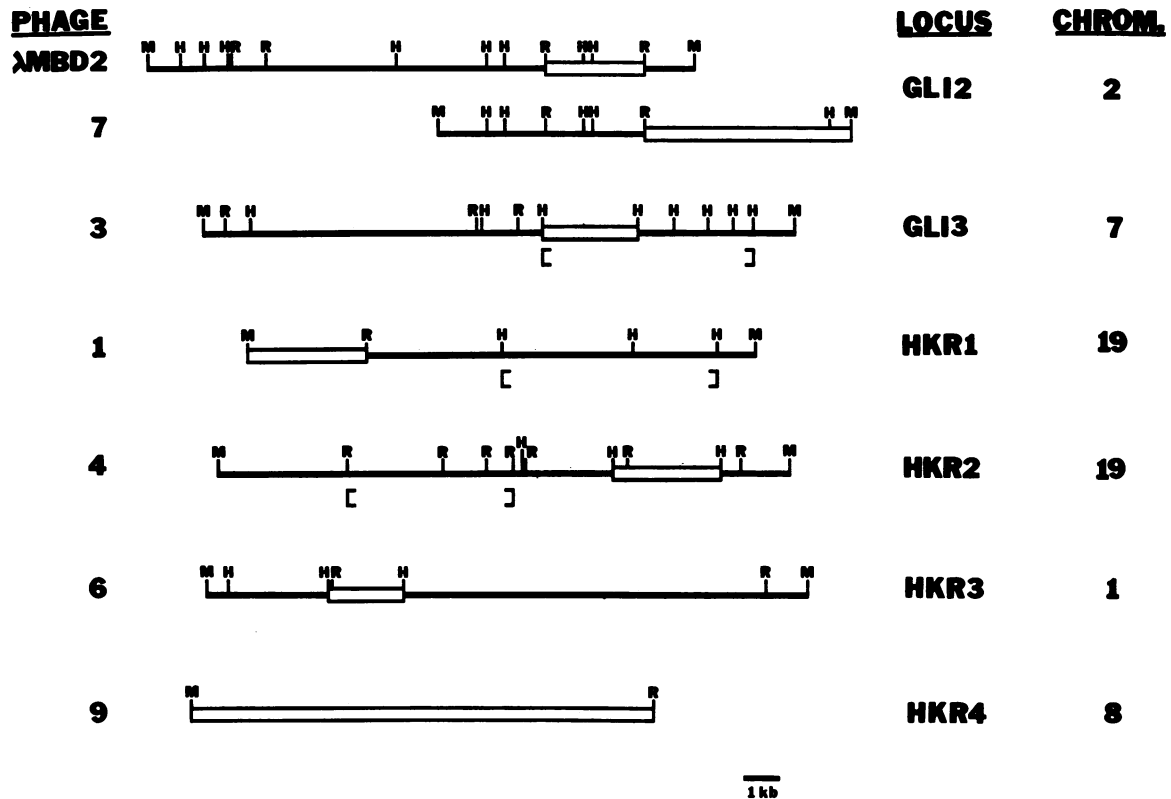
FIG. 1. *GLI*-related loci. Reduced-stringency hybridization of pGLIMBD to a λ phage library allowed isolation of clones containing related sequences. Representative phage for each of the six new loci identified are shown. *Eco*RI (R) and *Hin*dIII (H) recognition sites within the phages are illustrated. M indicates the *Mbo*I site at the junction between human and phage sequences. Brackets below the maps indicate restriction fragments not ordered with respect to one another. Subcloned regions which hybridize to pGLIMBD are shown as open boxes (for plasmid names see Table 1). Hybridization of plasmid subclones to DNA from human-rodent hybrid panels was used to determine chromosome localizations of the six different loci (shown on the right).

sus splice donor and acceptor sequences allowed prediction of intron-exon junctions. Each of the nine predicted intron-exon junction sequences examined showed that the positions of predicted splice junctions within the finger region were exactly conserved (Fig. 3A). For both *GLI2* and *GLI3*, the sequences predicted an exon extending from the middle of finger 1 to near the beginning of finger 3. For *GLI2*, the sequence predicted another exon extending from upstream of the middle of finger 3 (where the sequence was truncated

during cloning) to the C-terminus of finger 4. The conservation of intron-exon junctions and the conservation of amino acid sequences within individual fingers supported the hypothesis of a common evolutionary origin for this subgroup of genes (the *GLI* subgroup).

The four other finger-related clones demonstrated amino acid and nucleotide similarity to *GLI* only in the H-C link and in the other consensus residues (i.e., C-C-F-L-H-H) found commonly in many finger proteins. The H-C link

TABLE 1. Loci, phages, and clones used in this study

Locus	Phage	Large plasmid subclone[a]	Sequencing plasmid subclone
GLI2	λMBD2	pGLI2RR	pGLI2RR-*Bst*NI-600
GLI2	λMBD7	pGLI2RS	pGLI2RS-*Taq*I-480[b]
GLI3	λMBD3	pGLI3HH	pGLI3HH-*Hae*III-640[b]
HKR1	λMBD1	pHKR1RS	pHKR1RS-*Rsa*I-680[b]
HKR2	λMBD4	pHKR2HH	pHKR2HH-*Hinf*I-325[b]
HKR3	λMBD6	pHKR3HH	pHKR3HH-*Msp*I-490[b]
HKR4	λMBD9	pHKR4RS	pHKR4RS-*Bst*NI-400
HKR4	λMBD9	pHKR4RS	pHKR4RS-*Bst*NI-450[b]
GLI2	λMBD2	pGLI2RR-2	pGLI2RR-2-*Hae*III-440[c]
HKR3	λMBD6	pHKR3HH	pHKR3HH-HPX-550[c]

[a] The plasmid subclone names include the genomic restriction fragment used as insert (H, *Hin*dIII; R, *Eco*RI; S, *Sal*I; P, *Pst*I; X, *Xba*I) and, in the case of the small subclones used for sequencing, the size of the insert. Thus, the plasmid pGLI2RR-*Bst*NI-600 refers to a plasmid subclone of pGLI2 RR (boxed in Fig. 1) containing a 600-bp *Bst*NI insert.

[b] Subclone used for RNase protection, chromosome localization, and cross-species hybridization studies.

[c] Subclone isolated by hybridization with non-zinc finger regions of *GLI* cDNA (see Materials and Methods).

```
GLI2    TGCCAATATAGACGGTCACTTTGAGGACTGATTGGAAAATGCAGGGCAGATAAGCGGCACGCAGTAGGCCTTCATTTAAGGGCAGCTAGTGATGATTCATGAGGAGGCGTGGCCCGGCTC  120

        GTAGGGGTGGTCTTGTTGGGTCAGTGGTCTGCGTATAGTAGGTGGGATGGGCAATGAGGAGCACGGTCTGAACATAATAGGTGTGGTCAGCAGGGTGGATGGTCTGCACACAGGTCGGGG  240

        AGGGCTGGGGGGGTGCCGGTGCAGGTGGTCAGCTGACAGCAGGGGGTGGTCTGTGGGCCTCTGGCTGCTTACCCTCTTCTGGGCATGTCCCTCCGGCAGCACATCAACAACGAGCACATC  360
                                                                                                              H  I  N  N  E  H  I

        CACGGGGAGAAGAAGGAGTTTGTGTGCCGCTGGCAGGCCTGCACGCGGGAGCAGAAGCCCTTCAAGGCGCAGTACATGCTGGTGGTGCACATGCGGCGACACACGGGCGAGAAGCCCCAC  480
        H  G  E  K  K  E  F  V  C  R  W  Q  A  C  T  R  E  Q  K  P  F  K  A  Q  Y  M  L  V  V  H  M  R  R  H  T  G  E  K  P  H

        AAGTGCACGGTGAGTGGCCTTCTCCCCACCCCCGCCGCAGCATCAAGACTGGCCTGTC  538
        K  C  T

        AAGGCCTACTCCCGCCTGGAGAACCTGAAGACACACCTGCGGTCCCACACCGGGGAGAAGCCCATATGTGTGTGAGCACGAGGGCTGCAACAAAGCCTTCTCCAACGCCTCGGACCGCGCC  120
        K  A  Y  S  R  L  E  N  L  K  T  H  L  R  S  H  T  G  E  K  P  Y  V  C  E  H  E  G  C  N  K  A  F  S  N  A  S  D  R  A

        AAGCACCAGAATCGCACCCACTCCAACGAGGTACCTCTGCGGGGCATGCACTGGGCATGCACACTGGGGCCCCACTGATGCCCCATGGCTTCCAGGCATCTGTAGTCCAGTAGATAGCCA  240
        K  H  Q  N  R  T  H  S  N  E

        GGTGCCCCATGTCCCGGCCCCCGCCACTGACCAGCTATACATCCTCAGATACAGACCAGGACTCTTTGAGACCTAGACAGACCGGTCTCTTTCTCTATGCCACATCCTTTCCTCTAAAAG  360

        ATGCCCCGTGAGGTCCCATGGCATACTGACCCTCTCTGTGCCTCAATTTCCTTATCTCCATAATAAATATAATAGTGCCCTCCCCAAAGGCCATCATGAGGTCC  464
```

```
GLI3    GTGTACTGTAGTATATTTCCTGGAATGCCAGGCTCCATCCCGGTCTTGCCACTCCCACTGGTTAGGAAGCATGCATACACAGTTAGCAAAAGCAAATTTACTTGAGAAGTGTACTAACTG  120

        GAATTTTTATAATCTGTAATATGGTAAAGTTATTCTTGTGGTCTTTAGAATTTTTAAGATTGGGGTATTTTCTGCATTTTCTTCTGTCGCCTTCAGAAATGATGAATACGTTTCCATTTG  240

        ATTTCTCTCTCTTTTTCCAAACCCCAGCATATAAAATAACGACCATATTCATGGAGAGAAGAAGGAGTTCGTGTGCAGGTGGCTGGACTGCTCAAGAGAGCAGAAACCCTTCAAAGCCCAG  360
                                    H  I  N  N  D  H  I  H  G  E  K  K  E  F  V  C  R  W  L  D  C  S  R  E  Q  K  P  F  K  A  Q

        TATATGTTGGTAGTGCATATGAGAAGACACACGGGCGAGAAGCCTCACAAATGCACTGTGAGTACGAAGTGGGTGGCAGGCACACTTGGGTCCTTCCCCATACTCATAAGAGCTGTGCA  480
        Y  M  L  V  V  H  M  R  R  H  T  G  E  K  P  H  K  C  T

        AACTGATGACACCAGCTCAGGGGAAGCGTGGTTGGGTGGCAAGAGAAGAGGACTAGAGTATTGTCTGGTGATAAATGCCTCGGTGTTTGAGTGACTGAGTTGGACAGCGCAGCTCTGCAC  600

        AGTGGCAAGGGGGCAGTGGTGGCTGGCACTGG  632
```

```
HKR1    ACTGAGTGGGGAGACAGCTTTGGCAGTATGTCAGTCCTCATCAAAAACCCAAGGACACACTCTGGGGGAAAGCCTTATGTGTGCAGGGAATGTGGGCGAGGCTTTACGTGGAAGTCAAAC  120
        T  E  W  G  D  S  F  G  S  M  S  V  L  I  K  N  P  R  T  H  S  G  G  K  P  Y  V  C  R  E  C  G  R  G  F  T  W  K  S  N

        CTGATCACACATCAGAGGACACACTCAGGGGAGAAACCTTATGTGTGCAAGGATTGTGGACGAGGCTTTACTTGGAAGTCGAACCTCTTTACACATCAGCGGACACACTCAGGGCTCAAG  240
        L  I  T  H  Q  R  T  H  S  G  E  K  P  Y  V  C  K  D  C  G  R  G  F  T  W  K  S  N  L  F  T  H  Q  R  T  H  S  G  L  K

        CCTTATGTGTGCAAGGAATGTGGGCAGAGCTTTAGCCTGAAGTCAAACCTCATTACCCACCAGAGGGCGCACACTGGGGAGAAGCCTTATGTTTGCAGGGAATGTGGGCGTGGCTTTCGC  360
        P  Y  V  C  K  E  C  G  Q  S  F  S  L  K  S  N  L  I  T  H  Q  R  A  H  T  G  E  K  P  Y  V  C  R  E  C  G  R  G  F  R

        CAGCATTCACACCTGGTCAGACACAAGAGGACACATTCAGGAGGAGAAGCCTTACATTTGCAGGGAGTGTGAGCAAGGCTTTAGCCAGAAGTCACACCTCATCAGACACTTAAGGACACAC  480
        Q  H  S  H  L  V  R  H  K  R  T  H  S  G  E  K  P  Y  I  C  R  E  C  E  Q  G  F  S  Q  K  S  H  L  I  R  H  L  R  T  H

        ACAGGAGAGAAGCCTTATGTATGCACAGAATGTGGGCGTCACTTTAGCTGGAAATCAAACCTCAAAACACACCAGAGGACACACTCAGGGGTTAAACCTTATGTCTGCCTGGAGTGCGGG  600
        T  G  E  K  P  Y  V  C  T  E  C  G  R  H  F  S  W  K  S  N  L  K  T  H  Q  R  T  H  S  G  V  K  P  Y  V  C  L  E  C  G

        CAGTGCTTTAGCCTGAAGTCAAACCTTAACAAACACCAGAGGTCACACACGGGGGAGAAGCCATTTGTATGT  672
        Q  C  F  S  L  K  S  N  L  N  K  H  Q  R  S  H  T  G  E  K  P  F  V  C
```

```
HKR2    CCCAACAAAAGGAGACCCTGTACGTGCGAAGAGTGTGGCAAGGCCTTTGGACAGAGAAGTCACCTTGTTCAGCACACCAGTGAGAAGCTGTATGCATGCCAGGAATGTGGGTGTACCTTC  120
        P  N  K  R  R  P  C  T  C  E  E  C  G  K  A  F  G  Q  R  S  H  L  V  Q  H  T  S  E  K  L  Y  A  C  Q  E  C  G  C  T  F

        AGCAACAATTCATCTCTAGTCAAGCACTGGCACGTCCACACAGGCGAGAAGCCCTACATGTGTGGCCACTGTGGCAAGTGCTTCCGAGAGAGCTCATCCCTTGCCAAGCACCAGCGTGTG  240
        S  N  N  S  S  L  V  K  H  W  H  V  H  T  G  E  K  P  Y  M  C  G  E  C  G  K  C  F  R  E  S  S  S  L  A  K  H  Q  R  V

        CACACAGGTGAGAAGCCATACGTGTGTGGTGAGTGTGGACGCACCTTCAGCGAGAGCACACACCTTGTACAGCACTGGTG  320
        H  T  G  E  K  P  Y  V  C  G  E  C  G  R  T  F  S  E  S  T  H  L  V  Q  H  W
```

```
HKR3    TTCCTGTGACTCCTGCTCATAGATTGTCCTTCTGCTCTCGGGGTGGAGGTGGTGCCCCTTTCCTAGCACTGCCCAAGCCCTCTTTCCACCAGGCATGCTCCTAGCTGTAGCAGAGCAAGG  120

        GGGTCACTTCCCTTGGTGATGGCCTCTGCCCCATGTCCCCACCTTAACAGGAATGAGGAGGCCACACGTATGTGAGTTCTGCAGCCACGCCTTCACCCAAAAGGCCAATCTCAACATGCAC  240
                                                 N  E  R  P  H  V  C  E  F  C  S  H  A  F  T  Q  K  A  N  L  N  M  H

        CTGCGCACACACACGGGTGAGAAGCCCTTCCAGTGCCACCTCTGTGGCAAGACCTTCCGAACCCAAGGTGAGGTACGCCCTGCCCCTCCCCTCGCCTCCCCACCTGAGGCCAAGGCCACA  360
        L  R  T  H  T  G  E  K  P  F  Q  C  H  L  C  G  K  T  F  R  T  Q

        GGCTGAGCTCTTGCCTTGTGCCTGCAGCCAGCCTGGACAAGCACAACCGCACCCACACCG  420
        A  S  L  D  K  H  N  R  T  H  T
```

```
HKR4    GCGCGTGACCTCGTGCAGCAAGCAGCGGCCGGGCCCGAGGGTGCGCCCGAGCGGGCTGCCGAGCTGGGAGTCAACTTCGGTCGGAGCCGGCAGGGCAGCGCGCGGGGGACCAAGCCGCAC  120
        A  R  D  L  V  Q  Q  A  A  A  G  P  E  G  A  P  E  R  A  A  E  L  G  V  N  F  G  R  S  R  Q  G  S  A  R  G  T  K  P  H

        AGGTGCGAGGCCTGCGGCAAGAGTTTCAAGTATAACTCGCTGCTCCTGAAGCACCAGCGCATCCACACGGGCGAGAAGCCCTACGCCTGCCACGAGTGCGGCAAGCGCTTCCGCGGCTGG  240
        R  C  E  A  C  G  K  S  F  K  Y  N  S  L  L  L  K  H  Q  R  I  H  T  G  E  K  P  Y  A  C  H  E  C  G  K  R  F  R  G  W

        TCGGGCTTCATCCAGCACCACCGCATCCACACGGGCGAGAAGCCCTACGAGTGCGGCCAGTGCGGCCGCGCCTTCAGCCACAGCTCGCACTTCACGCAGCACCTGCGCATCCACAACGGC  360
        S  G  F  I  Q  H  H  R  I  H  T  G  E  K  P  Y  E  C  G  Q  C  G  R  A  F  S  H  S  S  H  F  T  Q  H  L  R  I  H  N  G

        GAGAAGCCCTACAAGTGCGGCGAGTGC  387
        E  K  P  Y  K  C  G  E  C

        GTGCGCCACCAGCGGCTGCACACGGGTGAGAAGCCCTACGCCTGCAGCCAGTGCGGCAAGGCCTTCATCTGGAGCTCCGTGCTCATCGAGCACCAGCGCATCCACACTGGCGAGAAGCCC  120
        V  R  H  Q  R  L  H  T  G  E  K  P  Y  A  C  S  Q  C  G  K  A  F  I  W  S  S  V  L  I  E  H  Q  R  I  H  T  G  E  K  P

        TACGAGTGCTCCGACTGCGGCAAAGCCTTCCGCGGCCGCTCGCACTTCTTCCGGCACCTGCGGACCCACACGGGCGAGAAGCCCTTCGCGTGTGGCGCCTGCGGCAAGGCCTTCGGCCAG  240
        Y  E  C  S  D  C  G  K  A  F  R  G  R  S  H  F  F  R  H  L  R  T  H  T  G  E  K  P  F  A  C  G  A  C  G  K  A  F  G  Q

        AGCTCCCAGCTCATCCAGCACCAGCGGGTGCACTACCGCGAGTAGCCGGGCGGGGGCTCGGGGCTCGGCCTCCTACCTGCCCCCAACCCACCCTCCACCCCGTCCCCCACGGTGGGCACT  360
        S  S  Q  L  I  Q  H  Q  R  V  H  Y  R  E  *

        GCCCAGCACCGCATGCCACGTGTCCGGAATAAATTCTTTTTGATTGTTGGAAGTGGGAGCC  421
```

FIG. 2. Sequence analysis of GLI-related loci. Sequences of GLI2, GLI3, HKR1, HKR2, HKR3, and HKR4 are shown. Sequencing was performed as described in Materials and Methods. Splice consensus sequences (51) were present adjacent to regions of GLI cDNA similarity. Closed arrows represent predicted splice donor sites, and open arrows show predicted splice acceptor sites. The sequences obtained for HKR1 and HKR2 each had one continuous open reading frame encoding fingers. The GLI2 and HKR3 sequences each predicted two different exons. Two BstNI fragments from HKR4 contained finger-hybridizing sequences; it is not known whether these two BstNI fragments were contiguous in the genome. The presence of an in-frame stop codon (asterisk) in the larger fragment of HKR4, the results of ribonuclease protection (Fig. 4), and the presence of a possible polyadenylation signal (underlined) downstream of the stop codon suggested that these sequences lie at the C-terminus of the HKR4 protein.

A)

```
FINGER                                              **     *
            GLI3                              }INNND HIHGEKKE
            GLI2                              }INNNE HIHGEKKE
    1       GLI     TD CRWDG CSOE FDSOEO LVH }INSE HIHGERKE
            MGLI                             }INSE HIHGERKE


                    *  **   *   **
            GLI3    FV CRWLD CSREOKP FKAQYH LVV HHRR HTGEKP
            GLI2    FV CRWOA CTREOKP FKAQYH LVV HHRR HTGEKP
    2       GLI     FV CHWGG CSRELRP FKAQYH LVV HHRR HTGEKP
            MGLI    F


            GLI3    HK CT     *
            GLI2    HK CT     KA YSRLEN LKT HLRS HTGEKP
    3       GLI     HK CTFEG CRKS YSRLEN LKT HLRS HTGEKP
            MGLI       }EG CRKS YSRLEN LKT HLRS HTGEKP


                    *    *    *
            GLI2    YV CEHEG CNKA FSNASD RAK HQNRT HSNE
    4       GLI     YM CEHEG CSKA FSNASD RAK HQNRT HSNEKP
            MGLI    YM CEQEG CSKA FSNASD RAK HQNRT HSNEKP


    5       GLI     YV CKLPG CTKR YTDPSS LRK HVKTV HGPDAHVTKRH
            MGLI    YV CKLPG CTKR YTDPSS LRK HVKTV HGPDAHVTKRH


                    Y        S    F              R    T
CONSENSUS           -X CXXXG C-KX -XXXXX LXX HX-XX H-GEKP
                    F        T    Y              N    S
```

B)

```
HKR1        TE WGDS FGSHSV LIK NPRT HSGGKP
         YV CRE CGRG FTWKSN LIT HQRT HSGEKP
         YV CKD CGRG FTWKSN LFT HQRT HSGLKP
         YV CKE CGQS FSLKSN LIT HQRA HTGEKP
         YV CRE CGRG FROHSH LVR HKRT HSGEKP
         YI CRE CEQG FSQKSH LIR HLRT HTGEKP
         YV CTE CGRH FSWKSN LKT HQRT HSGVKP
         YV CLE CGQC FSLKSN LNK HQRS HTGEKP
         FV C

HKR2     CT CEE CGKA FGQRSH LVQ H    TSEKL
         YA CQE CGCT FSNNSS LVK HWHV HTGEKP
         YM CGH CGKC FRESSS LAK HQRV HTGEKP
         YV CGE CGRT FSESTH LVQ HW

HKR3                                      NERP
         HV CEF CSHA FTQKAN LHM HLRT HTGEKP
         FQ CHL CGKT FRTQAS LDK HNRT HT

HKR4                                      RGTKP
         HR CEA CGKS FKYNSL LLK HQRI HTGEKP
         YA CHE CGKR FRGWSG FIQ HHRI HTGEKP
         YE CGQ CGRA FSHSSH FTQ HLRI HNGEKP
         YK CGE C
                                    VR HQRL HTGEKP
         YA CSQ CGKA FIWSSV LIE HQRI HTGEKP
         YE CSD CGKA FRGRSH FFR HLRT HTGEKP
         FA CGA CGKA FGQSSQ LIQ HQRV HYRE*

         Y
KR       -X CXX CDXX FXXXHX LXX HXRX HTGEKP
         F

MKR      YE CXE CGKX FXXXSX LXX HQRI HTGEKP

                 K                    T
HKR      YX CXE CG-X FXXXSX LXX HXRX H-GEKP
                 R                    S
```

C)
GLI AA74-107

```
GLI2(DNA) ACGCCCCGCCTGAGCCGCAAGCGGGCGCTGTCCATCTCCCCACTCTCAGAC
GLI2(AA)    T  P  R  L  S  R  K  R  A  L  S  I  S  P  L  S  D
GLI         A  V  K  L  T  K  K  R  A  L  S  I  S  P  L  S  D
MGLI        S  V  K  L  T  K  K  R  A  L  S  I  S  P  L  S  D
            *  *  *     *  *

GLI2      GCCAGCCTGGACCTGCAGCGGATGATCCGCACCTCACCCAACTCGCTAGTG
GLI2        A  S  L  D  L  Q  R  M  I  R  T  S  P  N  S  L  V
GLI         A  S  L  D  L  Q  T  V  I  R  T  S  P  S  S  L  V
MGLI        A  S  L  D  L  Q  T  V  I  R  T  S  P  S  S  L  V
                              *  *              *
```

GLI AA441-456

```
HKR3(DNA) TCTCTTGGGGCCCAGTCCTGCTGCCAGCCCAGGCTTGCACCGGCAGGG
HKR3(AA)    S  L  G  A  Q  S  C  C  Q  P  R  L  A  P  A  G
GLI         S  P  G  A  Q  S  S  C  S  S  D  H  S  P  A  G
            *              *     *  *  *  *  *
```
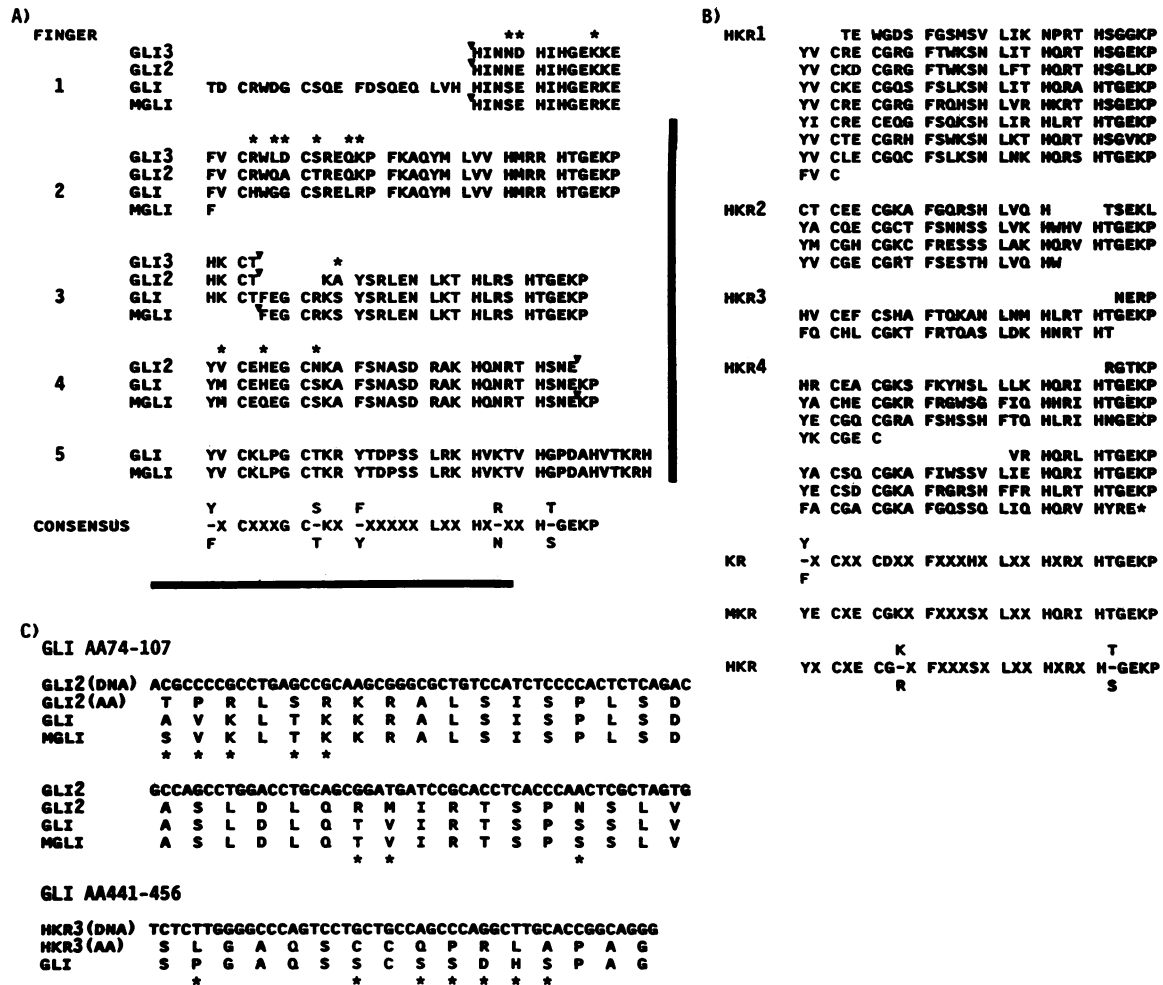
FIG. 3. (A) *GLI* family finger structure. Predicted amino acid sequences obtained from *GLI2*, *GLI3*, and *MGLI* genomic DNA sequences were aligned with *GLI* amino acids 235 to 397. The asterisks mark positions where amino acids varied. Arrowheads show positions of predicted intron-exon junctions (see text and Fig. 2). A consensus sequence common to the majority of *GLI* fingers is shown at the bottom. X indicates any amino acid. (B) Krüppel family structure. Predicted amino acid sequences obtained from *HKR1* to 4 genomic DNA sequences are illustrated, with alignment of fingers to demonstrate the consensus features typical of Krüppel-related genes. The consensus sequences of the fingers for the *HKR*, mouse Krüppel (*MKR*, [10]) and Krüppel (46) genes are shown at the bottom; the asterisk indicates a translation stop codon. (C) Similarity to *GLI* in nonfinger regions. Phage from each of the six loci depicted in Fig. 1 were hybridized with *GLI* cDNA probes corresponding to regions upstream or downstream of the fingers (see text). Subclones containing the hybridizing regions from *GLI2* and *HKR3* were sequenced. These sequences were aligned with the GLI cDNA and the predicted amino acids were compared. The asterisks indicate positions of amino acid variation.

accounts for the only contiguous stretch of *GLI* similarity and thus was responsible for the hybridization to these genes. Inspection of the predicted amino acid sequences of the four clones (Fig. 3B) showed that they were more similar to previously described *Kr* family genes than to the *GLI* genes described above. The four loci were therefore named *HKR1* to 4 (for human Krüppel-related genes). Some *HKR* fingers had up to 75% identity with those of mouse (*MKR1*, *MKR2*, [10]) or *Xenopus* (*Xfin* [47]) Krüppel-related genes. For example, one of the fingers of *HKR4* shared 78% identity with one finger of *MKR2* and 75% identity with the finger 34 of *Xfin*. However, this strong similarity was only observed in isolated fingers, and none of the four *HKR* genes appeared to represent the human homolog of a previously described zinc finger gene.

Spacing of the first two cysteines within *HKR* gene fingers (CX$_2$C) differed from the *GLI* gene fingers (CX$_4$C). In addition, inspection of the finger sequences revealed specific

conserved amino acids that distinguished the *HKR* subgroup from the *GLI* subgroup (Fig. 3A and B). Those amino acids that were common to *HKR* fingers were also common to those of the mouse Krüppel genes (Fig. 3B). Another difference between the *GLI* and *HKR* genes was that *HKR* had several fingers encoded by one exon, while in *GLI* exons generally encoded only one complete finger (Fig. 2). The mouse Krüppel family genes also encode multiple fingers in individual exons (10).

For *HKR4*, an in-frame stop codon occurred at the carboxy terminus of a predicted finger, followed by a possible polyadenylation signal (AATAAA) 103 base pairs downstream (Fig. 3B). Location of a finger region at the C-terminus has also been observed in the *Drosophila* finger protein gene *snail* (5).

Similarity in nonfinger regions. It was of interest to determine whether the six loci described above contained regions of similarity to *GLI* other than those in the zinc finger

domain. Towards this end, phage clones representing each of the six loci were hybridized to *GLI* cDNA probes containing coding sequences upstream (nucleotides 118 to 873) or downstream (nucleotides 1255 to 2426) of the fingers. *GLI2* was found to hybridize to the upstream probe, while *HKR3* hybridized to the downstream probe. Genomic sequences from the hybridizing regions of *GLI2* and *HKR3* were subcloned and sequenced, and the predicted amino acid sequences were compared to *GLI* (Fig. 3C). *GLI2* and *GLI* showed 76% identity over a 34-amino-acid region upstream of the *GLI* fingers (amino acids 74 to 107 of *GLI*). Over the same region, *MGLI* and *GLI* were identical at all but the first position. *HKR3* and *GLI* shared 56% of residues over a 16-amino-acid region downstream of the *GLI* zinc fingers (amino acids 441 to 456 of *GLI*).

Expression of finger-encoding genes. To assess expression, subclones containing finger regions from each of the six loci described in Fig. 1 were used to generate synthetic antisense RNA transcripts. RNA from normal human tissues, a glioblastoma tumor (D245MG [4]), and two embryonal carcinoma cell lines (Tera-1 [18] and NTera-2 [1]) were then tested for their abilities to protect the in vitro transcripts from RNase A digestion. The glioblastoma and embryonal carcinoma cell lines were chosen because previous studies had indicated that *GLI* and other Krüppel-related genes were expressed in analogous cells (10, 31).

Expression was demonstrated for all loci except *HKR2* (Fig. 4). In each case, the length of the protected RNA transcript corresponded to that expected from sequence data (Fig. 2). The levels of expression in different cell types varied widely among the six loci. *GLI2* was found to be expressed in all normal tissues except for placenta, with highest levels expressed in testes, myometrium, and kidney. In contrast, *GLI3* was found to be expressed in all normal tissues except for kidney and brain, with highest levels in myometrium and lung. *HKR1* and *HKR4* were found to be expressed at detectable levels in all normal tissues studied, while *HKR3* was found in all except placenta. For each of the expressed loci, levels of expression were considerably higher in the glioblastoma multiforme or embryonal carcinoma cell lines than in any normal tissue.

Similar sequences in other species. The data on RNA expression showed that five of the six loci contained genes which were expressed in normal adult tissues. Further evidence for the potential importance of these genes was provided by the observation that sequences related to each of the genes were evolutionarily conserved. Human, mouse, rat, chicken, frog, fly, and yeast DNAs were digested with *Eco*RI or *Hind*III, and Southern blots were prepared. At reduced stringencies (see Materials and Methods), probes containing finger-encoding regions from each of the six loci identified one or a few prominent restriction fragments in various other species (Fig. 5 and data not shown). For example, *HKR1* detected related fragments in mouse and rat (Fig. 5A), whereas *GLI3* detected fragments in every species tested (Fig. 5B). *HKR3* identified several fragments in every species except for yeast and frog. Comparison of the sizes of the *Eco*RI and *Hind*III fragments identified by each of the *HKR* loci revealed that each hybridized to a different set of restriction fragments. Therefore, the H-C link, which was present in all six loci, was not responsible for the cross-species hybridization patterns. Rather, the patterns indicated that sequences within each of the identified loci were evolutionarily conserved separately. For two of the six probes (*HKR3* and *HKR4*), further reduction of stringency
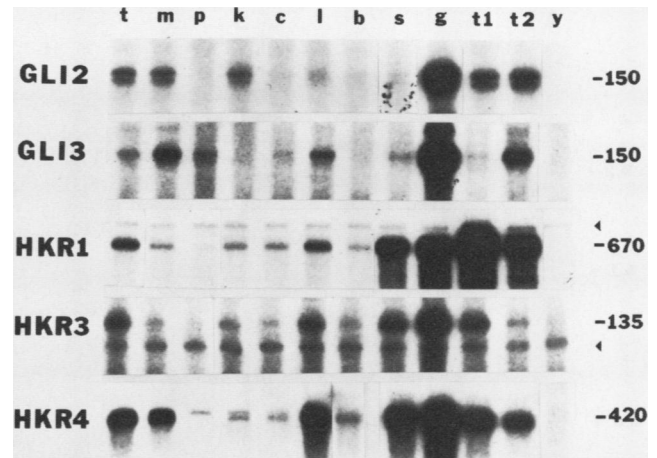

FIG. 4. Expression of *GLI-Kr*. Expression was assessed by protection of labeled antisense RNA from ribonuclease digestion. Labeled transcripts of the plasmid subclones corresponding to each of the six loci depicted in Fig. 1 were hybridized with 20 μg of total RNA from the sources listed below. After digestion with ribonuclease A, RNA was separated by electrophoresis on denaturing polyacrylamide gels and exposed to autoradiographic film. The plasmid subclones used for protection are described in Materials and Methods. None of the RNA samples analyzed protected *HKR2*. The sizes of the protected transcripts, shown on the right in bases, correspond to those expected from sequence data and support the intron-exon junction predictions (Fig. 2). Doublets can be seen in some lanes with the *GLI2* and *HKR4* probes. It is not known whether these resulted from incomplete RNase protection in vitro or alternative RNA processing in vivo. Arrowheads indicate nonspecific signals present in all lanes, including yeast tRNA lanes. Lanes: t, testes; m, myometrium; p. placenta; k, kidney; c, colon; l, lung; b, brain; s, spleen; g, glioblastoma multiforme D245MG; t1, embryonal carcinoma cell line Tera-1; t2, embryonal carcinoma cell line NTera-2; y, yeast tRNA.

allowed visualization of many additional fragments in all species tested except for yeast (cf. Fig. 5C and 5D).

Chromosome localization of the six loci. The data on RNA expression (Fig. 4) and species conservation (Fig. 5) indicated significant differences among the six loci, and were consistent with the hypothesis that each encoded a different gene. Further evidence for this hypothesis was provided by chromosome localization studies. DNA from human-rodent hybrid panels (8, 28, 40) was used in hybridization studies with probes representing each of the six loci. The six loci were found to be localized to five different chromosomes (Fig. 1). These chromosome localizations, based on karyotypic and enzymatic analysis of hybrids, were confirmed by syntenic mapping, wherein probes from each of the chromosomes predicted to contain one of the six genes were rehybridized to the same Southern blots, duplicating the initial patterns. Thus, the segregation of *GLI2* was concordant with *N-MYC* on chromosome 2, *GLI3* with the epidermal growth factor receptor gene on chromosome 7, *HKR1* and 2 with *APO-E* on chromosome 19, *HKR3* with *N-RAS* on chromosome 1, and *HKR4* with *C-MYC* on chromosome 8.

DISCUSSION

We used a *GLI* finger probe to isolate sequences from six loci, each of which encoded zinc fingers with conserved H-C links. By analogy with previously characterized zinc finger
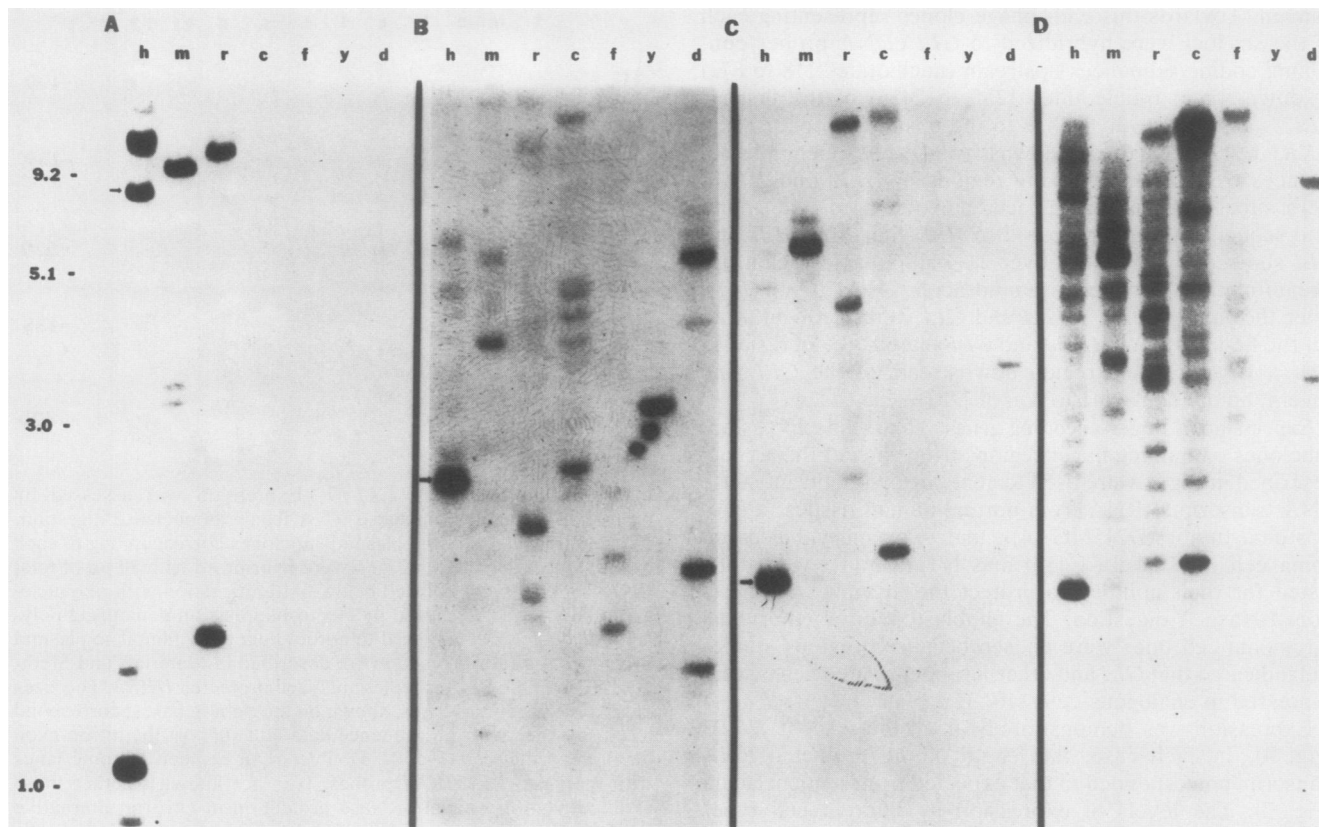
FIG. 5. Evolutionary conservation of finger-encoding sequences. Southern blots containing DNA from several eucaryotes (listed below) were hybridized at reduced stringency to radiolabeled DNA probes prepared from the subclones described in Materials and Methods. Panels A, B, and C show *HKR1*, *GLI3*, and *HKR3* probes hybridized at moderate stringency. Panel D shows hybridization to the *HKR3* probe at lower stringency. Arrows indicate the unique human *Hin*dIII fragments which hybridized at normal stringency (data not shown). Numbers on the left are kilobases. Lanes: h, human (5 μg); m, mouse (5 μg); r, rat (5 μg); c, chicken (5 μg); f, frog (5 μg); y, yeast (1 μg); d, *D. melanogaster* (1 μg).

genes (6, 27, 37), these probably encode DNA (or RNA) binding proteins which may regulate transcription. At least two of the loci (*HKR3* and *GLI2*) contained sequences related to other parts of the *GLI* cDNA as well, suggesting that the putative proteins encoded by these loci have *GLI* similarity aside from that in the finger region. Five of the six loci were found to be variably expressed in several normal adult tissues as well as in embryonal carcinoma cell lines. The genes were mapped to five different chromosomes, and each was evolutionarily conserved. These studies, therefore, document the basic features of a family of human genes which have in common an H-C link connecting zinc fingers. There are probably additional members of this family besides the seven genes discussed here (*GLI*, *GLI2*, *GLI3*, and *HKR1* -2, -3, and -4), since low-stringency hybridization with probes from some of these loci revealed the presence of many restriction fragments (Fig. 5D).

Conservation of DNA-binding protein motifs has allowed isolation of related genes within families in the case of homeo box-containing genes (11, 34, 35) and for two classes of finger proteins, the nuclear receptor family (e.g., estrogen [22] or retinoic acid [21, 43] receptors) and the Krüppel family. The Krüppel family genes have been isolated from several species; some have been isolated by hybridization with Krüppel cDNA (7, 10, 47, 49), and others have been cloned independently (see Table 2). In general, similarity outside of the finger region has not been demonstrated between Krüppel family members (other than the small

regions of similarity described in Fig. 3C and in Chavrier et al. [7]), suggesting a significant diversity of function.

Because of the large number of finger-containing genes, it is useful to subclassify them on the basis of common conserved elements (2, 16, 32). One major class of finger proteins, termed C_2H_2 (16), contains several adjacent fingers of the form $(Y/F)XCX_{2,4}CX_3$ $FX_5LX_{2-3}HX_{3-4}HX_5$. This class is distinguished from other classes of finger proteins by virtue of the fact that the putative metal-binding amino acids are always C-C- - -H-H, and phenylalanine and leucine residues are present near the finger midpoints. Three subclasses with these features can be defined (Table 2). The $C_2H_2-X_5$ subclass genes lack any consensus sequence connecting the zinc fingers. In contrast, the *GLI*-Krüppel gene family members are notable for having a conserved H-C link (HTGEKP(Y/F)XC) present between most fingers (10, 49). Identical H-C links are only occasionally found in $C_2H_2-X_5$ proteins (e.g., between one pair of the nine fingers of *TFIIIA*). The *GLI*-Krüppel genes can be divided into Krüppel (C_2H_2-Kr) or *GLI* (C_2H_2-GLI) subclasses, depending on spacing within fingers and other conserved primary sequence features (Fig. 3A and B; Table 2). Interestingly, the *Sp1* gene was recently sequenced and found to have three zinc fingers with connecting H-C links, which is typical of the *GLI*-Krüppel family (27). One finger of *Sp1* fits the *Kr* subgroup consensus, one finger fits the *GLI* subgroup consensus, and one finger fits neither. In contrast, the six genes described in this report were clearly divisible into *GLI* or

TABLE 2. C_2H_2 finger proteins

Subclass	H-C links	Consensus finger sequence	Gene	Source	Reference	Method of isolation
C_2H_2-X_5	Minority of fingers	(Y/F)XCX$_{2-4}$CX$_3$FX$_5$LX$_{2-3}$HX$_{3-4}$HX$_5$	*TFIIIA*	*Xenopus*	6, 37	Independent
			serendipity	*Drosophila*	59	Independent
			hunchback	*Drosophila*	55	independent
			snail	*Drosophila*	5	Independent
			pDP1007	Human	42	Independent
C_2H_2-*KR*	Most fingers	(Y/F)XCX$_2$CX$_3$FX$_5$LX$_2$HXRXHTGEKP	Krüppel	*Drosophila*	46	Independent
			ADR1	Yeast	24	Independent
			KR-H	*Drosophila*	49	*KR*-related
			Xfin	*Xenopus*	47	*KR*-related
			NGFI-A	Rat	36	Independent
			MKR1,2	Mouse	10	*KR*-related
			KROX4,6,8,9,20	Mouse	7	*KR*-related
			EGR-1	Mouse	52	Independent
			Sp1	Human	27	Independent
			HKR1-4	Human	This work	*GLI*-related
C_2H_2-*GLI*	Most fingers	(Y/F)XCX$_3$GCX$_3$(F/Y)X$_5$LX$_2$HX$_{3-4}$H(T/S)GEKP	*GLI*	Human	31	Independent
			MGLI	Mouse	31	*GLI*-related
			GLI2,3	Human	This work	*GLI*-related

Krüppel subgroups on the basis of the features described in Fig. 3 and Table 2. Perhaps reflecting the large size of the zinc finger gene family, characterization of genes from various species has resulted in the isolation of homologs only rarely. Of the genes listed in Table 2, apparent homologs include only (i) *NGFI-A* (rat) and *EGR-1* (mouse) (36, 52) and (ii) *GLI* (human) and *MGLI* (mouse) (31).

There were several distinct differences between the C_2H_2-*GLI* and C_2H_2-*Kr* subclasses. First, the spacing of amino acids between the invariant cysteine and histidine residues for C_2H_2-*GLI* (CX$_3$GC. . .HX$_{3-4}$H) differed from C_2H_2-*Kr* (CX$_2$C. . .HXRXH). Second, exons of the *GLI* subgroup contained only one complete finger, while exons of the Krüppel subgroup contained several fingers. It has previously been noted that most of the nine fingers of *TFIIIA* are contained within individual exons (57). This intron-exon organization of the *TFIIIA* and *GLI* subgroup genes gives support to the hypothesis that finger proteins have evolved by gene duplication of a small subunit encoding one finger of approximately 30 amino acids (37). Another difference was the conservation of amino acid sequence in the finger regions. Although the fingers of *HKR1* to *4* were almost invariant with respect to certain amino acids common to all Krüppel family fingers (Fig. 3B), there were variable amino acids at other positions, so that each of the finger domains within *HKR1* to *4* was unique. This suggests that the DNA-binding specificities or affinities or both would vary considerably among the different proteins encoded by these loci. In contrast, sequences within the finger regions were highly conserved among the *GLI* genes (Fig. 3A). If *GLI* is a sequence-specific DNA-binding transcription factor, as has been demonstrated for a growing list of finger proteins (e.g., *TFIIIA* [15], *SP1* [14], and human glucocorticoid receptor [25]), then the remarkable similarity of finger sequences among *GLI*, *GLI2*, and *GLI3* suggest that each may bind to the same or similar sequences. The study of amino acid variation in fingers binding similar sequences may further the understanding of sequence-specific binding. Notably, sequence variations between *GLI*, *MGLI*, *GLI2*, and *GLI3* are not found at those finger positions proposed to make sequence-specific contacts with DNA (i.e., those amino acids near the conserved leucine; 3).

Consistent with their putative role in nucleic acid binding and transcriptional regulation, various finger proteins have been shown to be involved in important biological processes. Hunchback (55) and snail (5) play key roles in *Drosophila* development, and the retinoic acid receptor may be important in morphogenesis of the chick limb (21, 43, 56). Some members of the *GLI*-Krüppel family have been implicated in the processes of embryonic development and neoplasia (31, 44, 47, 49); the transcription activator *Sp1* acts on a variety of gene promoters in normal tissues (13, 27). Further study of the members of the *GLI*-Krüppel gene family described here may prove important to the understanding of transcriptional mechanisms or differential gene expression or both and may be relevant to both normal and pathological states in humans.

LITERATURE CITED

1. **Andrews, P. W., I. Damjanov, D. Simon, G. S. Banting, C. Carlin, N. C. Dracopoli, and J. Fogh.** 1984. Pluripotent embryonal carcinoma clones derived from the human teratocarcinoma cell line tera-2. J. Lab. Invest. **50**:147–162.
2. **Berg, J. M.** 1986. Potential metal-binding domains in nucleic acid binding proteins. Science **232**:485–487.
3. **Berg, J. M.** 1988. A proposed structure for the zinc binding domains from transcription factor IIIA and related proteins. Proc. Natl. Acad. Sci. USA **85**:99–102.
4. **Bigner, S. H., J. Mark, and D. D. Bigner.** 1987. Chromosomal progression of malignant human gliomas from biopsy to establishment as permanent lines in vitro. Cancer Genet. Cytogenet. **24**:163–176.
5. **Boulay, J. L., C. Dennefeld, and A. Alberga.** 1987. The *Drosophila* developmental gene *snail* encodes a protein with nucleic acid binding fingers. Nature (London) **330**:395–398.
6. **Brown, R. S., C. Sander, and P. Argos.** 1985. The primary

structure of transcription factor TFIIIA has 12 consecutive repeats. FEBS Lett. **186**:271–274.

7. **Chavrier, P., P. Lemaire, O. Revelant, R. Bravo, and P. Charnay.** 1988. Characterization of a mouse multigene family that encodes zinc finger structures. Mol. Cell. Biol. **8**:1319–1326.

8. **Cheung, P., F. Kao, M. L. Law, C. Jones, T. T. Puck, and L. Chan.** 1984. Localization of the structural gene for human apolipoprotein A on the long arm of human chromosome 11. Proc. Natl. Acad. Sci. USA **81**:508–511.

9. **Chomczynski, P., and N. Sacchi.** 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162**:156–159.

10. **Chowdhury, K., U. Deutsch, and P. Gruss.** 1987. A multigene family encoding several "finger" structures is present and differentially active in mammalian genomes. Cell **48**:771–778.

11. **Colberg-Poley, A. M., S. D. Voss, K. Chowdhury, and P. Gruss.** 1985. Structural analysis of murine genes containing homeo box sequences and their expression in embryonal carcinoma cells. Nature (London) **314**:713–718.

12. **Coussens, L., T. L. Yang-Feng, Y.-C. Liao, E. Chen, A. Gray, J. McGrath, P. H. Seeburg, T. A. Libermann, J. Schlessinger, U. Francke, A. Levinson, and A. Ullrich.** 1985. Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with *neu* oncogene. Science **230**:1132–1139.

13. **Dynan, W., and R. Tjian.** 1985. Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. Nature (London) **316**:774–778.

14. **Dynan, W. S., and R. Tjian.** 1983. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell **35**:79–87.

15. **Engelke, D. R., S.-Y. Ng, B. S. Shastry, and R. G. Roeder.** 1980. Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. Cell **19**:717–728.

16. **Evans, R. M., and S. M. Hollenberg.** 1988. Zinc fingers: gilt by association. Cell **52**:1–3.

17. **Feinberg, A. P., and B. Vogelstein.** 1984. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **137**:266–267.

18. **Fogh, J.** 1978. Cultivation, characterization, and identification of human tumor cells with emphasis on kidney, testis, and bladder tumors. Natl. Cancer Inst. Monogr. **49**:5–9.

19. **Gehring, W. J.** 1985. Homeotic genes, the homeo box, and the genetic control of development. Cold Spring Harbor Symp. Quant. Biol. **50**:243–252.

20. **Gehring, W. J.** 1987. Homeo boxes in the study of development. Science **236**:1245–1252.

21. **Giguere, V., E. S. Ong, P. Segui, and R. M. Evans.** 1987. Identification of a receptor for the morphogen retinoic acid. Nature (London) **330**:624–629.

22. **Giguere, V., N. Yang, P. Segui, and R. M. Evans.** 1988. Identification of a new class of steroid hormone receptors. Nature (London) **331**:91–94.

23. **Hanas, J. S., D. J. Hazuda, D. F. Bogenhagen, F. Y.-H. Wu, and C.-W. Wu.** 1983. *Xenopus* transcription factor A requires zinc for binding to the 5S RNA gene. J. Biol. Chem. **258**:14120–14125.

24. **Hartshorne, T. A., H. Blumberg, and E. T. Young.** 1986. Sequence homology of the yeast regulatory protein ADR1 with Xenopus transcription factor TFIIIA. Nature (London) **320**:283–287.

25. **Hollenberg, S. M., V. Giguere, P. Segui, and R. M. Evans.** 1987. Colocalization of DNA-binding and transcriptional activation functions in the human glucocorticoid receptor. Cell **49**:39–46.

25a.**Humphrey, P. A., A. J. Wong, B. Vogelstein, H. S. Friedman, M. H. Werner, D. D. Bigner, and S. H. Bigner.** 1988. Amplification and expression of the epidermal growth factor receptor gene in human glioma xeno grafts. Cancer Res. **48**:2231–2238.

26. **Johnson, D. A., J. W. Gautsch, J. R. Sportsman, and J. H. Elder.** 1984. Gene analysis techniques. Gene Anal. Tech. **1**:308.

27. **Kadonaga, J. T., K. R. Carner, F. R. Masiarz, and R. Tijian.** 1987. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell **51**:1079–

1090.

28. **Kao, F., J. A. Hartz, M. L. Law, and J. N. Davidson.** 1982. Isolation and chromosomal localization of unique DNA sequences from a human genomic library. Proc. Natl. Acad. Sci. USA **79**:865–869.

29. **King, C. R., M. H. Kraus, and S. A. Aaronson.** 1985. Amplification of a novel v-*erb*B-related gene in a human mammary carcinoma. Science **229**:974–976.

30. **Kinzler, K. W., S. H. Bigner, D. D. Bigner, J. M. Trent, M. L. Law, S. J. O'Brien, A. J. Wong, and B. Vogelstein.** 1987. Identification of an amplified, highly expressed gene in a human glioma. Science **236**:70–73.

31. **Kinzler, K. W., J. M. Ruppert, S. H. Bigner, and B. Vogelstein.** 1988. The Gli gene is a member of the Kruppel family of zinc finger proteins. Nature (London) **332**:371–374.

32. **Klug, A., and D. Rhodes.** 1987. 'Zinc fingers': a novel protein motif for nucleic acid recognition. Trends Biochem. Sci. **12**:464–469.

33. **Kohl, N. E., N. Kanda, R. R. Schreck, G. Bruns, S. A. Latt, F. Gilbert, and F. W. Alt.** 1983. Transposition and amplification of oncogene-related sequences in human neuroblastomas. Cell **35**:359–367.

34. **Levine, M., G. M. Rubin, and R. Tjian.** 1984. Human DNA sequences homologous to a protein coding region conserved between homeotic genes of Drosophila. Cell **38**:667–673.

35. **McGinnis, W., C. P. Hart, W. J. Gehring, and F. H. Ruddle.** 1984. Molecular cloning and chromosome mapping of a mouse DNA sequence homologous to homeotic genes of Drosophila. Cell **38**:675–680.

36. **Milbrandt, J.** 1987. A nerve growth factor-induced gene encodes a possible transcriptional regulatory factor. Science **238**:797–799.

37. **Miller, J., A. D. McLachlan, and A. Klug.** 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus ocytes. EMBO J. **4**:1609–1614.

38. **Nau, M. M., B. J. Brooks, J. Battey, E. Sausville, A. F. Gazdar, I. R. Kirsch, O.W. McBride, V. Bertness, G. F. Hollis, and J. D. Minna.** 1985. L-*myc*, a new *myc*-related gene amplified and expressed in human small cell lung cancer. Nature (London) **318**:69–73.

39. **O'Brien, S. J., W. G. Nash, J. L. Goodwin, D. R. Lowry, and E. H. Chang.** 1983. Dispersion of the *ras* family of transforming genes to four different chromosomes in man. Nature (London) **302**:839–842.

40. **S. J. O'Brien, T. I. Bonner, M. Cohen, C. O'Connell, and W. G. Nash.** 1983. Mapping of an endogenous retroviral sequence to human chromosome 18. Nature (London) **303**:74–77.

41. **Ollo, R., and T. Maniatis.** 1987. *Drosophila* Kruppel gene product produced in a baculovirus expression system is a nuclear phosphoprotein that binds to DNA. Proc. Natl. Acad. Sci. USA **84**:5700–5704.

42. **Page, D. C., R. Mosher, E. M. Simpson, E. M. C. Fisher, G. Mardon, J. Pollack, B. McGillivray, A. de la Chapelle, and L. G. Brown.** 1987. The sex-determining region of the human Y chromosome encodes a finger protein. Cell **51**:1091–1104.

43. **Petkovich, M., N. J. Brand, A. Krust, and P. Chambon.** 1987. A human retinoic acid receptor which belongs to the family of nuclear receptors. Nature (London) **330**:444–450.

44. **Preiss, A., U. B. Rosenberg, A. Kienlin, E. Seifert, and H. Jackle.** 1985. Molecular genetics of *Kruppel*, a gene required for segmentation of the *Drosophila* embryo. Nature (London) **313**:27–32.

45. **Reed, K. C., and D. A. Mann.** 1985. Rapid transfer of DNA from agarose gels to nylon membranes. Nucleic Acids Res. **13**:7207–7221.

46. **Rosenberg, U. B., C. Schroder, A. Preiss, A. Kienlin, S. Cote, I. Riede, and H. Jackle.** 1986. Structural homology of the product of the Drosophila *Kruppel* gene with Xenopus transcription factor IIIA. Nature (London) **319**:336–339.

47. **Ruiz i Altaba, A., H. Perry-O'Keefe, and D. A. Melton.** 1987. *Xfin*: an embryonic gene encoding a multifingered protein in *Xenopus*. EMBO. J. **6**:3065–3070.

48. **Russel, M., S. Kidd, and M. R. Kelley.** 1986. An improved

filamentous helper phage for generating single-stranded plasmid DNA. Gene 45:333–338.

49. Schuh, R., W. Aicher, U. Gaul, S. Cote, A. Preiss, D. Maier, E. Siefert, U. Nauber, C. Schroder, R. Kemler, and H. Jackle. 1986. A conserved family of nuclear proteins containing structural elements of the finger protein encoded by *Kruppel*, a Drosophila segmentation gene. Cell 47:1025–1032.

50. Schwab, M., K. Alitalo, K. H. Klempnauer, H. E. Varmus, J. M. Bishop, F. Gilbert, G. Brodeur, M. Goldstein, and J. Trent. 1983. Amplified DNA with limited homology to *myc* cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. Nature (London) 305:245–248.

51. Shapiro, M. B., and P. Senapathy. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res. 15:7155–7174.

52. Sukhatme, V. P., X. Cao, L. C. Chang, C.-H. Tsai-Morris, D. Stamenkovich, P. C. P. Ferreira, D. R. Cohen, S. A. Edwards, T. B. Shows, T. Curran, M. M. Le Beau, and E. D. Adamson. 1988. A zinc finger-encoding gene coregulated with c-fos during growth and differentiation, and after cellular depolarization. Cell 53:37–43.

53. Tabor, S., and C. C. Richardson. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA 84:4767–4771.

54. Taparowsky, E., K. Shimizu, M. Goldfarb, and M. Wigler. 1983. Structure and activation of the human N-*ras* gene. Cell 34:581–586.

55. Tautz, D., R. Lehmann, H. Schnurch, R. Schuh, E. Seifert, A. Kienlin, K. Jones, and H. Jackle. 1987. Finger protein of novel

structure encoded by *hunchback*, a second member of the gap class of *Drosophila* segmentation genes. Nature (London) 327:383–389.

56. Thaller, C., and G. Eichele. 1987. Identification and spatial distribution of retinoids in the developing chick limb bud. Nature (London) 327:625–628.

57. Tso, J. Y., D. J. Van Den Berg, and L. J. Korn. 1986. Structure of the gene for *Xenopus* transcription factor TFIIIA. Nucleic Acids Res. 14:2187–2200.

58. Vincent, A. 1986. TFIIIA and homologous genes. The 'finger' proteins. Nucleic Acids Res. 14:4385–4391.

59. Vincent, A., H. V. Colot, and M. Rosbash. 1985. Sequence and structure of the *Serendipity* locus of *Drosophila melanogaster*. A densely transcribed region including a blastoderm-specific gene. J. Mol. Biol. 186:149–166.

60. Vogelstein, B. 1987. Rapid purification of DNA from agarose gels by centrifugation through a disposable plastic column. Anal. Biochem. 160:115–118.

61. Vogelstein, B., E. R. Fearon, S. R. Hamilton, A. C. Preisinger, H. F. Willard, A. M. Michelson, A. D. Riggs, and S. H. Orkin. 1987. Clonal analysis using recombinant DNA probes from the X-chromosome. Cancer Res. 47:4806–4813.

62. Wingender, E. 1988. Compilation of transcription regulating proteins. Nucleic Acids Res. 16:1879–1902.

63. Winter, E., F. Yamamoto, C. Almoguera, and M. Perucho. 1985. A method to detect and characterize point mutations in transcribed genes: amplification and overexpression of the mutant c-Ki-*ras* allele in human tumor cells. Proc. Natl. Acad. Sci. USA 82:7575–7579.