GENOMICS 7, 555–565 (1990)

Identification and Characterization of Transcripts from the Neurofibromatosis 1 Region: The Sequence and Genomic Structure of *EVI2* and Mapping of Other Transcripts

RICHARD M. CAWTHON,* PETER O'CONNELL,* ARTHUR M. BUCHBERG,† DAVID VISKOCHIL,* ROBERT B. WEISS,* MELANIE CULVER,* JEFFREY STEVENS,* NANCY A. JENKINS,† NEAL G. COPELAND,† AND RAY WHITE*

*Howard Hughes Medical Institute, University of Utah, Salt Lake City, Utah 84132; and †BRI–Basic Research Program, NCI–Frederick Cancer Research Facility, Frederick, Maryland 21701

Received January 31, 1990; revised March 21, 1990

Mapping of the *EVI2* gene between the translocation breakpoints of two patients with neurofibromatosis type 1 (NF1), combined with the likely role of its murine homolog in neoplastic disease, implicates *EVI2* as a possible candidate for the *NF1* gene. We report here the expression of a 1.6-kb *EVI2* transcript in normal human brain and peripheral blood mononuclear cells. Sequencing studies predict an *EVI2* protein of 232 amino acids that contains an N-terminal signal peptide, an extracellular domain with five potential glycosylation sites, a single hydrophobic transmembrane domain with a leucine zipper, and a hydrophilic cytoplasmic domain. These features are all well-conserved with respect to the mouse *Evi-2* protein and are consistent with the hypothesis that *EVI2* is a membrane protein that may complex with itself and/or other proteins within the membrane, perhaps to function as part of a cell-surface receptor. In the course of these studies we have also identified three other transcripts (classes of cDNAs) from the NF1 region. Two of these transcripts map between the NF1 translocation breakpoints; the remaining transcript maps just outside this region.

INTRODUCTION

Balanced chromosomal translocations found in two neurofibromatosis 1 patients (Ledbetter *et al.*, 1989; Menon *et al.*, 1989) provide important clues in the search for the von Recklinghausen neurofibromatosis (*NF1*) gene. Because the chromosomal rearrangements may have caused the disease by interrupting the *NF1* gene, at least a portion of the gene might be expected to lie within the region defined by the translocations. Any gene that maps into this region, therefore, becomes a candidate in the search for the *NF1* gene. We have previously reported the mapping of the human homolog of a mouse gene implicated in leukemogenesis, *Evi-2* (Buchberg *et al.*, 1988, 1989, and submitted for publication), into the 60-kb region between the NF1 translocation breakpoints (O'Connell *et al.*, 1990, accompanying paper). Here we describe the pattern of *EVI2* gene expression in several tissues, including tissues affected by NF1; report the sequences of human *EVI2* cDNA and genomic clones; predict features of the encoded protein product; compare the predicted amino acid sequence to that of the mouse gene; and show the location and orientation of the gene on the genomic map of the breakpoint region. We also describe the initial identification and mapping of several other transcripts from the NF1 region.

MATERIALS AND METHODS

cDNA Libraries and Plaque Screening

A normal adult human bone marrow cDNA library in λgt10, catalogue number HL1058a, was purchased from Clontech Laboratories, Inc. (Palo Alto, CA). A cDNA library in λ Zap, catalogue number 935203, derived from frontal cortex of a child without NF1, was purchased from Stratagene (San Diego, CA). Phage from the libraries were incubated with *Escherichia coli* strain LE392 and plated according to the manufacturer's instructions. Duplicate plaque lifts were made with 0.2-μm pore size Biotrans filters (Pall Biodyne) by the method of Benton and Davis (1977).

Radiolabeling of DNA probes was performed according to Feinberg and Vogelstein (1984). First, cosmid and cDNA inserts were separated from vector by restriction enzyme digestion and gel electrophoresis in low-melting agarose (NuSieve GTG agarose, FMC BioProducts, Rockland, ME). DNA in gel slices was

0888-7543/90 $3.00

used directly for radiolabeling. Prehybridization of the cDNA library filters was in 50% formamide, 5× SSC, 50 m*M* $NaPO_4$ (pH 6.5), 2× Denhardt's solution, and 500 μg/ml sheared, denatured human DNA at 42°C for 2 h. Human DNA was included to suppress hybridization of the probes (radiolabeled genomic or cDNA fragments) to repetitive sequences. Separate prehybridization of the radiolabeled probes was simultaneously carried out in a solution of the same composition except that 50 μg/ml of DNA from the vector used to clone the probe was included. After prehybridization, the probe and filters were combined and hybridization was carried out for 4–16 h at 42°C. When genomic cosmids were used as probes, washes were in 2× SSC, 0.1% SDS at room temperature. When cDNAs were used as probes, washes were in 0.1× SSC, 0.1% SDS at 55–65°C. After washing, filters were placed on X-Omat AR film with an intensifying screen at −70°C.

Anchored PCR cDNA Library

Anchored PCR, to obtain cDNA clones from the 5′ end of *EVI2* mRNA, was performed as described by Loh *et al.* (1989). For this purpose two *EVI2* oligonucleotide primers were constructed based on the furthest 5′ sequence shared by cDNA clones E-9 and E-34, and RNA was prepared from human brain frontal cortex. Following first-strand cDNA synthesis with the first *EVI2*-specific primer and addition of poly(dG) tails with terminal transferase, the second, more 5′ *EVI2* primer was used for PCR together with an "anchored primer" containing poly(dC) at its 3′ end. The amplified product was electrophoresed on a 2% low-melting agarose gel and visualized with ethidium bromide, revealing a heterogeneous set of products ranging from approximately 50 to 300 bp. The gel lane was cut into slices 2 mm thick starting about 1 cm above the region of highest molecular size that still contained visible product. Ten microliters of each gel slice was subjected to PCR using the same pair of primers and analyzed by gel electrophoresis. The mobility of the products matched the order of the gel slices in the lane, and the first fraction showing visible product was cloned into Bluescript SK(+) (Stratagene) and transformed into DH5αF′ cells. Colonies were screened with an *EVI2* probe still further 5′ in the known sequence; approximately 150 positive clones were identified in this way.

Cell Lines and Tissues

Lymphoblastoid cell lines from normal and NF1 individuals, established in our laboratory by transformation of peripheral blood mononuclear cells with Epstein–Barr virus, and somatic cell hybrid lines containing various portions of chromosome 17 were maintained in culture as described by O'Connell *et al.* (1990) in the accompanying manuscript. Skin fibroblasts were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and passaged every 5–7 days with a trypsin/EDTA solution. Mononuclear cells were prepared from whole peripheral blood by the LeucoPREP procedure according to the manufacturer's instructions (Becton Dickinson and Co., Lincoln Park, NJ). Bone marrow aspirates (performed at University Hospital, University of Utah Medical Center, with informed consent of the patients and Institutional Review Board approval) drawn into tubes with heparin were spun at 2000 rpm for 10 min in a tabletop centrifuge; the pellets were then used for preparation of total RNA (see below). Other tissues removed from patients for therapeutic reasons or at autopsy were immediately placed in liquid nitrogen and stored at −135°C until used for RNA preparation. Some tissue specimens were obtained from the National Neurological Research Bank (Veterans Administration Medical Center Wadsworth, Los Angeles, CA), which is sponsored by NINCDS/NIMH, NMSS, HD Foundation, TS Association, and the Veterans Administration.

Sequencing

All DNA sequencing was based on the dideoxy-termination method of Sanger (Sanger *et al.*, 1977). *EVI2* cDNA clone E-9.4 was digested with *Sau*3AI and subcloned into M13mp18; single-strand preparations were sequenced with fluorescently tagged M13 primers on an Applied Biosystems, Inc., Model 370A DNA sequencer (ABI, Foster City, CA), using T7 DNA polymerase or *Taq*I DNA polymerase. Double-stranded plasmid preparations of clones E-32, E-34, and E-37 in Bluescript SK(−) were made by the cesium chloride/ethidium bromide centrifugation method (Maniatis *et al.*, 1982a), denatured by treatment with 0.2 *N* NaOH, neutralized, and sequenced from M13 primers on the ABI machine. Ten anchored PCR cDNA clones including E-An3, E-An4, and E-An7 were symmetrically amplified from picked colonies using M13 primers and sized by agarose gel electrophoresis. The largest three inserts were then asymmetrically amplified from the initial amplified product; the single strands were sequenced on the ABI machine, and also sequenced manually using the Sequenase kit (U.S. Biochemical Corp., Cleveland, OH). All segments of the *EVI2* cDNA consensus sequence were sequenced at least three times.

DNA sequence from cosmid clone cEVI20 was obtained by multiplexed dideoxy sequencing of random shotgun clones. cEVI20 DNA was sonicated, size fractionated (500–1000 bp) on an agarose gel, and blunt-end ligated to a set of 10 vectors carrying multiplex identifier tag sequences (pKZ vectors; R. Weiss, manuscript in preparation). Primary clones (720) were iso-

lated (72/vector); these clones were then distributed into 72 mixtures, each containing 10 clones (one clone from each of the 10 vectors), and plasmid DNA was prepared and sequenced from these mixtures. The mixed sequences were electroblotted from standard sequencing gels onto GeneScreen nylon membranes and probed sequentially for the individual sequence ladders. From 20 probings, 690 readable sequence ladders were recovered, and these were melded into 61 contigs (approximate total melded sequence is 35,000 nucleotides).

DNA and Peptide Sequence Analysis

DNA sequences were aligned, examined for open reading frames, and compared to DNA sequences in the EMBL and GenBank databases using the IntelliGenetics Suite of programs (IntelliGenetics, Inc., Mountain View, CA) running on a VAX computer. The predicted *EVI2* peptide was aligned with the murine sequence, analyzed for hydropathy by the Kyte and Doolittle (1982) method, and compared to peptide sequences in the Swiss-Prot and PIR protein databases, again using the IntelliGenetics Suite. The most likely site for cleavage of the putative signal peptide was determined using the SIGSEQ1 and SIGSEQ2 programs of Folz and Gordon (1987), based on a weight-matrix method developed by von Heijne (1986).

RNA Preparation and Analysis

Total RNA from peripheral blood mononuclear cells, bone marrow aspirates, and cultured cells was prepared by the guanidinium–acid–phenol method as described by Chomczynski and Sacchi (1987). RNA from tissues that had been frozen in liquid nitrogen was similarly prepared, except that 1-g portions of tissue were transferred directly from liquid nitrogen into a tube containing 10 ml of guanidinium solution and immediately homogenized for 1 min at top speed with a Brinkmann Polytron (Brinkmann Instruments, Inc., Westbury, NY).

Total RNA was electrophoresed through 1% agarose gels containing formaldehyde and blotted overnight onto Duralon uncharged nylon membranes (Stratagene) essentially as described by Maniatis *et al.* (1982b), except that 0.37 *M* formaldehyde was used in the gels as described by Selden (1989) and in the electrophoresis buffer as well instead of 2.2 *M* formaldehyde. Filters were crosslinked with uv irradiation in a Stratalinker (Stratagene). Radiolabeling of probes for RNA analysis and hybridization, washing, and X-ray film exposures of RNA blots were performed as per the methods described above for screening cDNA libraries. RNA blots were stripped by swirling the blot in a container with near to boiling 0.1× SSC, 0.1% SDS; the stripping solution was changed three times over a 5-min period.

RESULTS

Isolation of cDNA Clones Encoded by the Human EVI2 Locus

The human genomic DNA segment contained in a cosmid, cEVI36, has been shown to include a region homologous to a portion of murine *Evi-2* (O'Connell *et al.*, 1990). Because in mouse the greatest abundance of *Evi-2* transcript is in macrophages (Buchberg *et al.*, submitted for publication), the insert from cEVI36 was used to screen an oligo(dT)-primed cDNA library from normal human bone marrow in λgt10. Thirteen positive clones (from 1 million plated plaques) were picked and plaque-purified; rescreening these 13 with the murine *Evi-2* cDNA BK3 (Buchberg *et al.*, submitted for publication) as probe identified 4 as being related to *Evi-2*. The largest of these four clones, cDNA E-9, contained a 5.1-kb insert (Fig. 1). Hybridization of the E-9 insert to genomic DNA detected a single, 15-kb *Bgl*II fragment, which maps between the translocation breakpoints of the two NF1 patients (O'Connell *et al.*, 1990).

Sequencing revealed no poly(A) tract in this cDNA clone; however, hybridization of RNA blots with single-stranded probes identified the sense strand. Further analysis of cDNA E-9 showed that only the 1.3-kb segment at the 3′ end detected a transcript on RNA blots. Because the remaining 5′ segment did not hybridize to any transcript from bone marrow or from several other human tissues, to murine *Evi-2* cDNA clones, or to Southern transfers of human genomic DNA, it is likely to represent a cloning artifact of non-human origin.

Analysis of EVI2 mRNA Expression in NF1 Tissues

CDNA E-9 identified a single class of transcript of 1.6 kb on RNA blots of bone marrow and other tissues. Figure 2 shows an autoradiograph of a blot of samples of total RNA probed with cDNA E-9.4, a 1.8-kb subclone from the 3′ end of clone E-9. The *EVI2* transcript is most abundant in peripheral blood mononuclear cells and brain, less abundant in bone marrow, detectable in Epstein–Barr virus (EBV)-transformed lymphoblastoid cell lines and fibroblasts, but not detected in choriocarcinoma. RNA samples from a peripheral nerve, a neurofibroma, and a malignant schwannoma, all from the same NF1 patient, also contained *EVI2* transcript, perhaps consistent with a role for *EVI2* in NF1. However, *EVI2* transcript could not be detected in neurofibromas from an additional two NF1 patients. It is possible that the relatively low abundance of *EVI2* transcripts found here in some of the pathologic specimens and autopsied tissues actually derives from peripheral blood cells in the vasculature of these tissues. The weak signals in some samples just above the 4.4-kb marker appear to represent nonspecific binding to

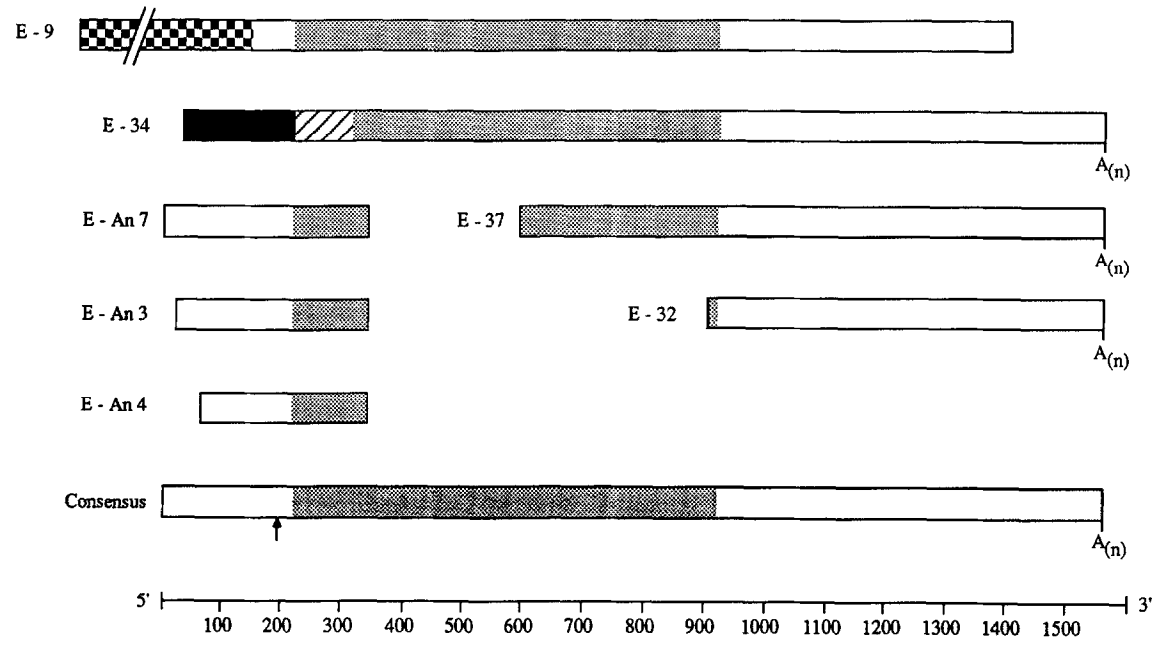

FIG. 1. Human *EVI2* cDNA clones. The scale is in nucleotide base pairs. E-9 is a 5.1-kb bone marrow cDNA. E-34, E-37, and E-32 are brain frontal-cortex cDNAs. E-An7, E-An3, and E-An4 are brain frontal-cortex cDNAs prepared by the anchored polymerase chain reaction, as described under Materials and Methods. The stippled segments contain the *EVI2* open reading frame. The solid white segments are 5′ and 3′ untranslated *EVI2* sequences. The checkered segment in E-9 and the solid black segment in E-34 contain sequences of unknown origin, unrelated to the 1.6-kb *EVI2* transcript. The cross-hatched segment in E-34 appears to be the inverse complement of the corresponding region in E-9, E-An7, E-An3, and E-An4. $A_{(n)}$ designates the presence of a poly(A) tail. The human brain *EVI2* consensus cDNA is shown at the bottom. In the genome a single 2.3-kb intron interrupts the sequence derived from the cDNA at the position indicated by the arrow.

28 S ribosomal RNA, since (1) by ethidium bromide staining the 28 S ribosomal RNA had the same mobility as this signal, (2) the intensity of this autoradiographic signal in the various samples paralleled the intensity of the ethidium bromide-stained 28 S RNA across the samples, and (3) blots of oligo(dT)-selected poly(A) RNA from equivalent samples had no detectable signal at this level (data not shown).

The blot shown in Fig. 2 includes RNA samples from the two NF1 lymphoblastoid cell lines containing, respectively, the t(17;22) and t(1;17) balanced translocations. RNA samples from a total of six unrelated NF1 patients are included on the blot. *EVI2* transcripts of abnormal mobility were not detected in any of these samples.

cDNA and Genomic Nucleotide Sequences and Predicted Amino Acid Sequence of EVI2

To identify additional *EVI2* cDNA clones, fragment E-9.4 was used to screen an oligo(dT)-primed cDNA library from normal human frontal cortex in λ Zap. Three independent positive clones were isolated, plaque-purified, and sequenced (Fig. 1). The largest of these, E-34, is 1522 bp long and contains within it the 1110 bp of sequence at the 3′ end of the E-9 clone. The E-34 sequence continues 3′ of this shared region for another 146 bp, then terminates with a run of 12 A residues. The 88-bp segment in E-34 just 5′ of the 1110-bp shared region, however, appears to be an inverse complement of the corresponding 88 bp in E-9 just 5′ of the shared region. The remaining sequence of 178 bp at the 5′ end of E-34 detects no transcript on RNA blots, shows no homology to any part of E-9, and does not hybridize to Southern transfers of chromosome 17 DNA.

To identify additional sequences from the 5′ end of the *EVI2* mRNA, we used the anchored PCR (Frohman *et al.*, 1988; Loh *et al.*, 1989) to construct a set of frontal-cortex cDNA clones enriched for the 5′ end of *EVI2*. We initiated first-strand cDNA synthesis on frontal-cortex RNA with a primer complementary to sense-strand sequence shared by E-34 and E-9. Three of the largest resulting cDNA clones were sequenced and found to be a nested overlapping set (Fig. 1). The largest of these, 341 bp, is E-An7. The last 138 bp at its 5′ end is sequence not found in either E-9 or E-34; this sequence used as a probe detects the same 1.6-kb transcript that is detected by E-9.

Assuming that only sequences shared by at least two independent clones represent true *EVI2* mRNA sequences, we arrived at the consensus *EVI2* cDNA shown in Fig. 1. It is 1563 bp long; the mRNA it detects is approximately 1.6 kb. The cDNA sequence, flanking

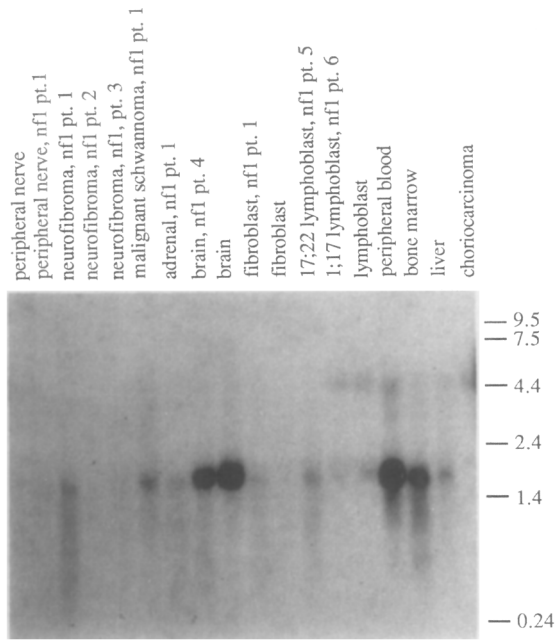

FIG. 2. RNA blot of human tissues, probed with *EVI2* cDNA E-9.4. Five micrograms of total RNA was loaded in each lane. An RNA ladder containing 9.5-, 7.5-, 4.4-, 2.4-, 1.4-, and 0.24-kb standards was included in the gel run to estimate the size of the detected transcripts. The six different NF1 patients from whom the samples were derived are labeled "nf1 pt. 1–6." One peripheral nerve sample, one neurofibroma sample, one fibroblast sample, and the samples labeled malignant schwannoma and adrenal are from a single NF1 patient (patient 1). The remaining five NF1 samples are all from different patients. These include the t(17;22) lymphoblastoid line from the NF1 patient with the t(17;22) balanced translocation and the t(1;17) lymphoblastoid line from the NF1 patient with the t(1;17) translocation. All other samples are from non-NF1 subjects. The peripheral blood sample contained mononuclear cells only.

genomic sequences, predicted amino acid sequence, and a hydropathy plot of human *EVI2* are shown in Fig. 3. A single intervening sequence has been found; it lies between nucleotide positions −23 and −22 of the cDNA sequence. A single, large open reading frame (ORF) extends from the A at the −12 position to the T at +696. Although the first ATG in this reading frame does begin at the −12 position, it does not match the consensus initiation site of eukaryotic mRNAs (Kozak, 1984), and an ATG does not occur at this position in the murine cDNA clones (Buchberg *et al.*, submitted for publication). The next ATG in the human *EVI2* sequence lies 13 bases downstream of the first ATG; this site matches the Kozak consensus initiation sequence 5′-RNNATGG-3′ and also appears in the murine cDNAs. Therefore, we have labeled this the +1 position for both the nucleotide and the amino acid sequence.

The *EVI2* ORF predicts a 232-amino-acid protein. The N-terminal segment has features characteristic of a signal peptide (von Heijne, 1985; Gierasch, 1989), with the most likely cleavage point (von Heijne, 1986; Folz and Gordon, 1987) lying between the alanine at position 26 and the asparagine at position 27. Next is a 103-residue hydrophilic, presumably extracellular domain, with five potential glycosylation sites. Then follows a segment of 28 hydrophobic amino acids with an average hydropathy score of 2.7 by Kyte and Doolittle (1982) analysis, which indicates that it is a transmembrane domain. The last 17 amino acids of this hydrophobic stretch are part of a 22-residue leucine zipper (Landschulz *et al.*, 1988); this feature suggests that *EVI2* forms homodimers or is able to bind to other proteins in the membrane. The region N-terminal of the zipper motif lacks the pattern of basic residues found in leucine zipper proteins that bind to DNA, suggesting that *EVI2* is not a DNA-binding protein (Vinson *et al.*, 1989). The remaining hydrophilic domain of 70 amino acid residues is presumably cytoplasmic. We have sought structural motifs characteristic of known oncogenes, but without success. Comparison of the DNA sequence to sequences in the EMBL and GenBank databases, and of the predicted peptide sequence to those in the PIR and Swiss-Prot databases, has shown that *EVI2* is a novel gene with no significant specific homologies to previously reported DNA or protein sequences.

The 3′ untranslated sequence of the *EVI2* mRNA contains four instances of the AUUUA pentamer. This motif is common to the 3′ untranslated regions of many oncogene and cytokine mRNAs, and it appears to promote rapid cytoplasmic degradation of mRNAs (Shaw and Kamen, 1986; Greenberg *et al.*, 1986; Wilson and Treisman, 1988; Brawerman, 1989; Malter, 1989).

Comparison of cDNA and Amino Acid Sequences of Murine Evi-2 and Human EVI2

The most striking difference between human *EVI2* and murine *Evi-2* is that in RNA blots of human tissues only a 1.6-kb *EVI2* transcript is detected, whereas in mouse tissues two classes of transcripts are seen, one 1.8 kb in length and the other 2.2 kb in length (Buchberg *et al.*, submitted for publication). In mouse, two classes of *Evi-2* cDNA clones from brain have been isolated; they have the same 3′ end (which contains the entire ORF of both cDNA classes and hybridizes to both sizes of RNA transcript), but 5′ of the ORF one cDNA class has been spliced at the same point where the human *EVI2* cDNA has been spliced, while the other cDNA has not been spliced at this point but continues to match mouse genome sequence. DNA probes based on the 5′ sequence of this second cDNA class hybridize exclusively to the 2.2-kb transcript in mouse RNA blots. This transcript is too small to be simply an immature form of the 1.6-kb transcript that has retained its intron (Buchberg *et al.*, submitted for pub-

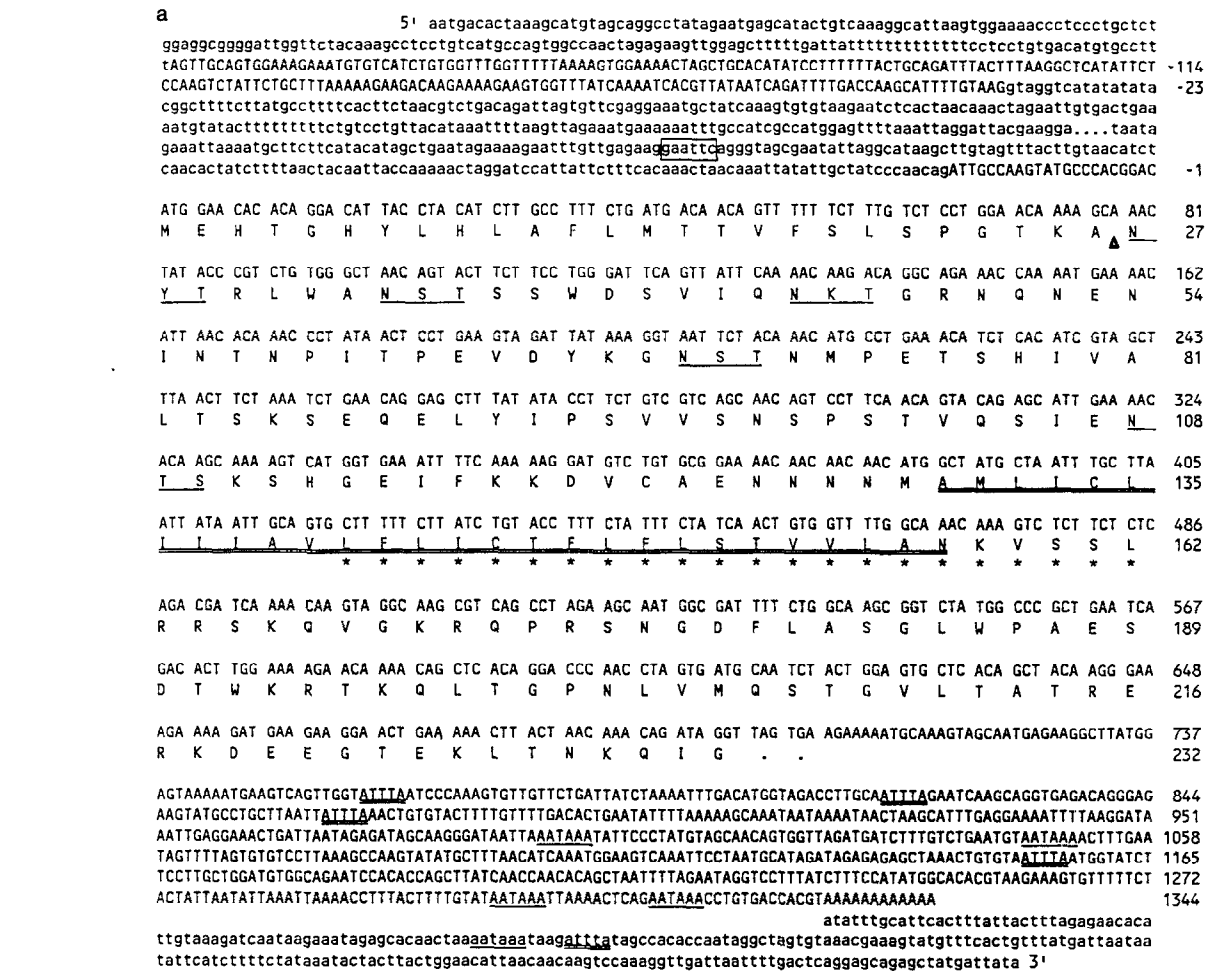

FIG. 3. cDNA and genomic nucleotide sequences and predicted amino acid sequence of *EVI2* (**a**) and hydropathy profile of predicted *EVI2* protein (**b**). (**a**) The cDNA nucleotide sequence is capitalized; the flanking genomic sequences are in lowercase. The cDNA sequence is numbered beginning with the putative initiating ATG codon and ending with the T nucleotide at position 1344; in the cDNA this T is followed by a series of 12 A's that end at the *Eco*RI cloning site of the vector. The genomic sequence corresponding to the position of the poly(A) tract is aligned beneath it. The flanking genomic sequences are not numbered. The single 2.3-kb intron of *EVI2* is not shown in its entirety; ellipses . . . represent the omitted portion of the intron. The *Eco*RI site that helped to position this sequence on the restriction map of the region (Fig. 5) is boxed. In the open reading frame (ORF) the triangle lying between predicted amino acids 26 and 27 indicates the most likely cleavage site for the putative signal peptide; single underlining denotes potential sites for N-linked glycosylation; double underlining indicates the strongly hydrophobic, presumably transmembrane, segment of the predicted protein; and asterisks mark the amino acids contributing to the leucine zipper. In the 3′ noncoding sequence single underlining identifies potential polyadenylation signals, and double underlining marks a reiterated pentamer shared by rapidly degraded mRNAs. (**b**) Hydropathy profile. Deflections above zero indicate hydrophobic regions.

lication). Another difference between mouse *Evi-2* and human *EVI2* transcripts is that in the 3′ untranslated region the mouse transcript has no AUUUA pentamers, although four are present in the human transcript.

Figure 4 compares the predicted peptide sequences of *EVI2* from human brain and *Evi-2* from murine brain. The overall identity of amino acid residues is approximately 65%, with greater homology within two well-conserved regions. First, 14 of 17 residues (82%) match (with no gaps) in the signal peptide. Second, 86 of 99 amino acids (87%) match (with one gap of a single residue) in a segment that begins with the transmembrane domain and continues through the cytoplasmic domain. The longest perfectly conserved segment (100% identity, no gaps) is a 25-residue segment that contains the leucine zipper. Three cysteine residues are found in both proteins; since these are conserved and lie near or within the leucine zipper, they may contribute stabilizing disulfide bridges within the *EVI2* protein or between it and the protein to which it binds. In the less conserved extracellular domain, 47 of 112 residues (42%) match and there are three single-residue gaps and one four-residue gap. This region contains five potential glycosylation sites in both mouse and human; two of the five sites are conserved.

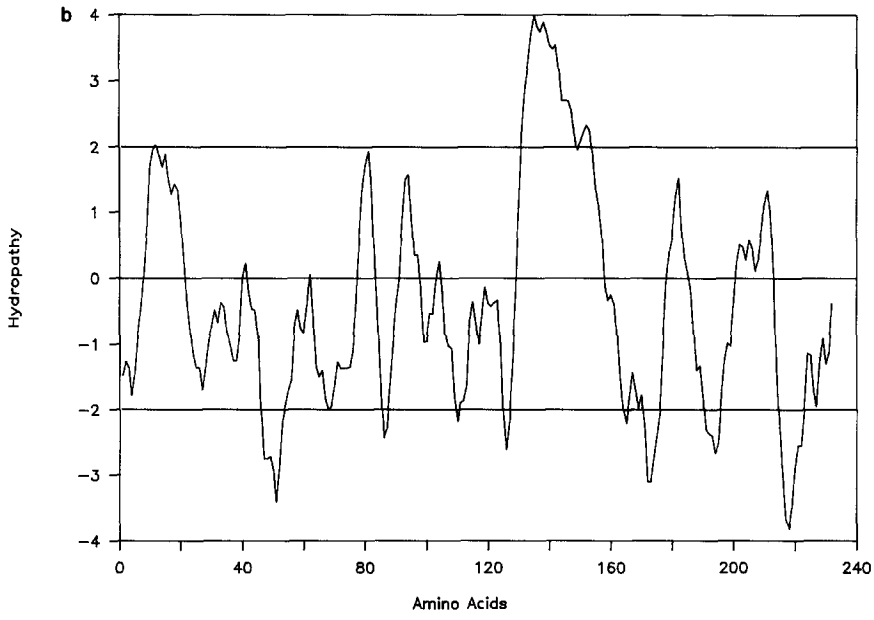

FIG. 3—*Continued*

Location of EVI2 with Respect to the NF1 Translocation Breakpoints

As shown in Fig. 5, the entire *EVI2* consensus cDNA sequence that has been obtained thus far lies within the interval between the translocation breakpoints and consists of two exons separated in genomic DNA by a 2.3-kb intron. Transcription is in the direction indicated by the arrow; i.e., the 5′ end of the sense strand is toward the telomere of 17q. This genomic organization was deduced from the following experiments.

First, comparison of the consensus cDNA sequence to the sequenced portions of cosmid cEVI20 (R. Weiss, manuscript in preparation) showed that the 197 bases at the 5′ end of the cDNA were identical to a 197-base sequence contained within one cosmid segment, and the 1366 bases at the 3′ end of the cDNA were identical to a 1366-base sequence contained within a different cosmid segment (Fig. 3a). In each of these two cosmid segments, sequences that continued beyond both boundaries of the sequence shared with the cDNA were found. The genomic sequences lying 3′ of the 197-base cDNA sequence and 5′ of the 1366-base cDNA sequence appear to be splice junctions.

Probes from the 3′ exon (exon 2 of Fig. 5) hybridized only to the 9.5-kb *Eco*RI fragment from cEVI20 and cEVI36, as indicated in Fig. 5. A probe from the 5′ exon (exon 1 of Fig. 5) hybridized only to the adjacent 10-

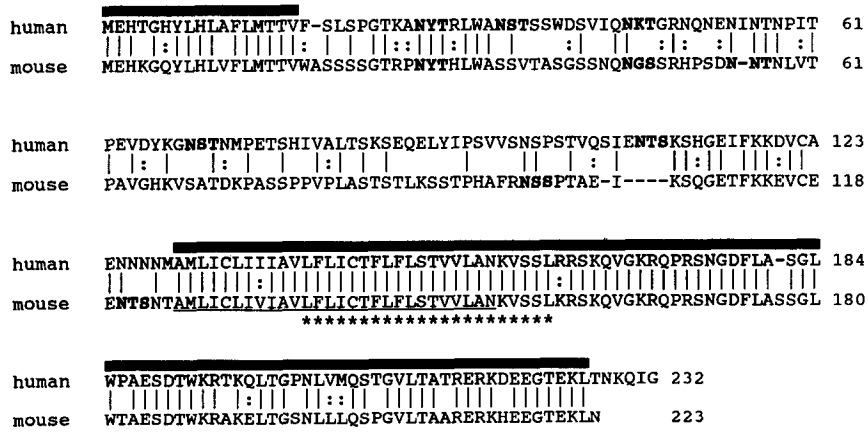

FIG. 4. Amino acid homologies of predicted *EVI2* proteins from human and mouse. Identical residues are joined by a vertical bar. Conservative substitutions are joined by a colon. Gaps introduced to optimize the alignment are shown as hyphens. Heavy overlining identifies segments highly conserved between mouse and human. Potential N-glycosylation signals (with N as the glycosylation site) are in boldface. The single long hydrophobic segment is underlined. Asterisks indicate the leucine zipper.

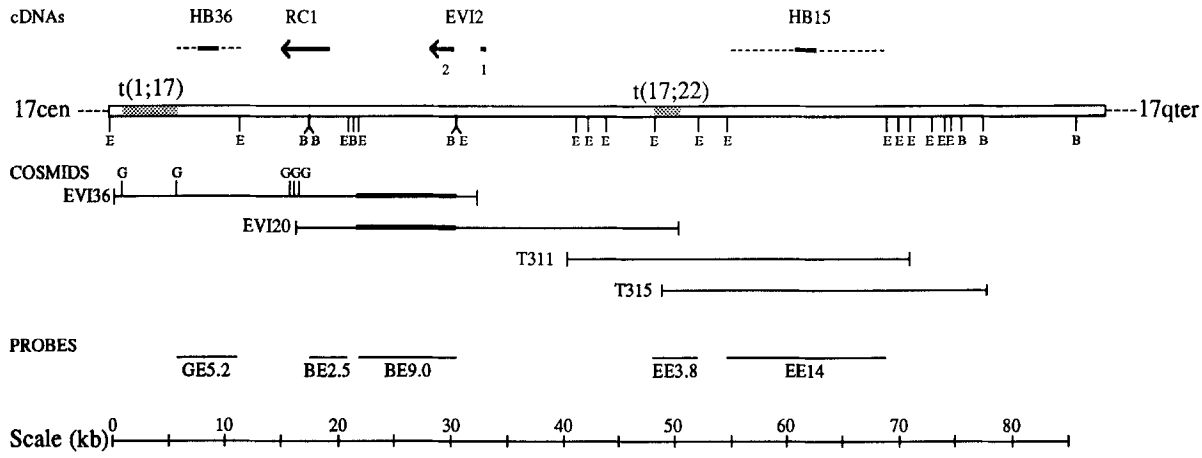

FIG. 5. Location of transcripts and conserved genomic segments in the translocation–breakpoint region. The long open bar is a *Bam*HI (B) and *Eco*RI (E) restriction map of the region of chromosome 17q containing the two NF1 translocation breakpoints. The centromere is toward the left; the telomere is toward the right. The t(1;17) and t(17;22) breakpoints lie somewhere within the respective stippled sections. Below the restriction map are shown four genomic cosmids that span the region and five shorter restriction fragments. The *Bgl*II (G) restriction map of cosmid cEVI36 is included. Above the *Eco*RI/*Bam*HI restriction map cDNA clones are represented as arrows and heavy bars overlying the restriction fragments to which they hybridize. The length of an arrow or heavy bar is scaled to the length of the isolated cDNA. The two exons of the *EVI2* consensus cDNA, numbered "1" and "2", have been precisely mapped and are shown as an interrupted heavy arrow; the direction of the arrow indicates the direction of transcription. The 9.5-kb restriction fragment of cosmids cEVI20 and cEVI36 to which *EVI2* exon 2 hybridizes is emphasized by doubly thick line segments within the two cosmids. The arrow centromeric of *EVI2* represents the largest cDNA clone isolated thus far from the *RC1* locus. This cDNA also has been precisely mapped by PCR; again, the direction of transcription is indicated by the arrow. cDNAs from the *HB36* and *HB15* loci are represented by bars instead of arrows because the direction of transcription is not yet known. Since precise mapping of these clones has not yet been accomplished, they are shown centered over the restriction fragments to which they hybridize, with dashed lines indicating that their sequences could lie anywhere within these restriction fragments. The five restriction fragments at the bottom of the figure identify genomic segments apparently evolutionarily conserved between mouse and human.

kb *Eco*RI fragment from cEVI20 (and not at all to cEVI36). Furthermore, as neither exon contains an *Eco*RI site, each exon is contained entirely within the genomic *Eco*RI fragment it detects. We were able to determine the exact position of the 3′ exon within the 9.5-kb *Eco*RI fragment because the sequenced genomic segment containing the 3′ exon also includes the *Eco*RI and *Bam*HI sites that lie between the exons. This *Eco*RI site lies 134 bp telomeric of the 5′ end of the 3′ *EVI2* exon (see box in Fig. 3a).

The size of the intron was determined by PCR. A primer based on the sequenced genomic segment containing the 5′ *EVI2* exon and another primer based on the genomic segment containing the 3′ exon were used for PCR with cEVI20 as template. After agarose gel electrophoresis, ethidium bromide staining revealed a single band. The size of this band and the known positions of the primer sequences relative to the exons indicate that the intron is 2.3 kb long.

Search for EVI2 Mutations Specific to NF1 Patients

As yet, no mutations in *EVI2* specific to NF1 patients have been found, although several techniques capable of detecting different types of mutation have been applied in the course of this search. All NF1 patients were diagnosed by the criteria of Stumpf *et al.* (1987). To screen for large-scale deletions or insertions in the region of *EVI2, EVI2* cDNA clones were used to probe genomic DNA from 38 unrelated NF1 patients and several non-NF1 controls, after digestion of the DNA with rare-cutting restriction endonucleases followed by pulsed-field gel electrophoresis. To screen for medium-scale deletions or insertions, these same probes were used to screen genomic DNA from 87 unrelated NF1 patients after digestion with restriction enzymes with 6-base recognition sites and standard agarose gel electrophoresis (O'Connell *et al.*, 1990). To screen for small deletions or insertions in the ORF of *EVI2,* three pairs of oligonucleotide primers were used to amplify by PCR the ORF from genomic DNA of 97 NF1 patients, in overlapping segments from 230 to 380 bp long, and the amplified products were electrophoresed through 4% sieving agarose gels. To screen for point mutations in *EVI2* we analyzed genomic DNA from four sporadic NF1 patients and their (unaffected) parents, by the PCR-single strand conformation polymorphism (PCR-SSCP) method of Orita *et al.* (1989); three overlapping PCR products from 230 to 380 bp long encompassing the *EVI2* ORF and one PCR product 305 bp long containing all of the 5′ exon sequence represented in *EVI2* cDNA clone E-An7 (see Fig. 1) were prepared from each individual, heat-denatured, and subjected to electrophoresis through 4–5% nondenaturing polyacrylamide gels. None of the above

procedures identified a restriction fragment or PCR product specific to NF1 patients. The *EVI2* ORFs of two of these sporadic patients have been sequenced in their entirety and found to match non-NF1 sequence. To look for abnormalities at the RNA level, EBV-transformed lymphoblast RNAs from 12 unrelated NF1 patients were screened with cDNA clone E-9.4; no aberrant *EVI2* transcripts were found.

Detection of Additional cDNA Clones and Evolutionarily Conserved Genomic Regions: Location with Respect to the NF1 Translocation Breakpoints

Using genomic DNA segments from the NF1 translocation breakpoint region to probe cDNA libraries, we have detected three additional classes of cDNAs apparently unrelated to each other or to the *EVI2* cDNAs described above; cDNAs in a given class hybridize to other cDNAs in the same class but not to cDNAs of other classes. Figure 5 shows the location of the largest cDNA clone from each of these loci (*HB36, RC1,* and *HB15*) on the map of the translocation-breakpoint region. Preliminary mapping of these cDNA clones onto the genome indicates that none of them spans either of the two NF1 translocation breakpoints. The cosmid restriction fragments GE5.2, BE2.5, BE9.0, and EE14 (Fig. 5), which hybridize to cDNAs from the *HB36, RC1, EVI2,* and *HB15* loci, respectively, also cross-hybridize to restriction fragments of rodent DNA, suggesting the presence of evolutionarily conserved regions likely to contain genes. One additional evolutionarily conserved region has been found with probe EE3.8, a 3.8-kb *Eco*RI fragment from cT311 that does span the t(17;22) translocation breakpoint; however, thus far this probe has not detected any positive clones in screens of several cDNA libraries, or any transcripts on RNA blots.

DISCUSSION

The location of the *EVI2* gene between NF1 translocation breakpoints and the association of its murine homolog with neoplasia make it a strong candidate for the *NF1* gene. Although the predicted protein product of *EVI2* has no significant homology to the products of known oncogenes, its structure suggests that it may be a growth-factor receptor. It appears to be a glycosylated protein with a single membrane-spanning domain that, by virtue of its leucine zipper, may interact with another membrane-bound protein. Such a structure is not inconsistent with our expectation for a gene product involved in neoplasia.

The high degree of conservation of membrane-spanning and cytoplasmic domains between mouse *Evi-2* and human *EVI2* suggests that any mutation in these regions may disrupt function. Even in the less well-conserved extracellular domain, two of the potential glycosylation sites are identically located. However, the presence in mouse of a second, 2.2-kb transcript with a different 5′ sequence that varies in proportion in different tissues, the absence of the AUUUA motif in the 3′ noncoding region of the murine *Evi-2* transcript, and the divergence between mouse and human amino acid sequences in the extracellular domain of the predicted protein product suggest that there may be significant differences between mouse and human in the regulation and function of this gene. If *EVI2* is the *NF1* gene, such differences may account for the fact that no mouse with a clinical phenotype like NF1 has ever been found. On the other hand, in light of the apparent role of *Evi-2* in murine leukemogenesis, it is worth noting that juvenile chronic myelogenous leukemias have been reported to occur more often in NF1 patients than in the general population (Bader and Miller, 1978; Mays *et al.,* 1980; Clark and Hutter, 1982). It is also possible that *EVI2* is not the *NF1* gene but when disrupted in the human (as in the mouse) predisposes the individual to particular leukemias; thus, NF1 patients with leukemias may have mutations (e.g., large deletions) that affect the *NF1* gene and the *EVI2* gene simultaneously.

An extensive, but not exhaustive, search for mutations in the *EVI2* gene specific to NF1 patients has been unsuccessful. If deletions removing part or all of the *NF1* gene or point mutations in the ORF of the gene are common lesions causing the disease, then *EVI2* is probably not the *NF1* gene, since our screening procedures would have been likely to detect such changes. However, we have not yet found the 5′ end of the *EVI2* mRNA, and we have not yet carefully examined the 2.3-kb intron, the 3′ noncoding sequences of the mRNA, or the 5′ and 3′ genomic sequences flanking the transcribed region of *EVI2.* These studies are under way and may yet reveal mutations specific to NF1 patients.

Could the DNA rearrangements in the two NF1 patients with chromosomal translocations be exerting their influence in some way other than by interruption of the *NF1* gene? One possibility is that the translocation breakpoints act over a distance to interfere with the normal expression of *EVI2* or another transcription unit in the vicinity. Consistent with this idea are studies in the mouse that suggest that viral insertions downstream of *Evi-2* can activate the oncogenic potential of this locus (Buchberg *et al.,* submitted for publication). Furthermore, many of the activations of c-*myc* by chromosomal translocation in Burkitt lymphomas do not appear to involve rearrangement of the c-*myc* gene itself; some breakpoints are as far as 260 kb away (Haluska *et al.,* 1987a,b; Henglein *et al.,* 1989). Recently, it was suggested that a t(Y;22) translocation described by Page *et al.* (1987) may have inactivated

the testis-determining gene, located at least 80 kb away from the translocation breakpoint (Palmer *et al.*, 1989). An additional precedent for translocational inactivation or alteration of the regulation of genes over a local region of a chromosome is the "position-effect variegation" observed among translocations and chromosomal rearrangements in *Drosophila* (e.g., see Spoffard, 1976). Thus, if the t(1;17) and t(17;22) translocation breakpoints are affecting the *NF1* gene without interrupting it, any of the genes described here could be the *NF1* gene; but it is also possible that none of them is the *NF1* gene and that the *NF1* gene lies either centromeric or telomeric of the cloned region analyzed here, perhaps many kilobases away.

Determination of whether any of the transcripts reported here has an etiogical role in NF1 will require the identification in these transcripts of mutations that are specific to NF1 patients. Despite our initial expectation that changes in a gene with such a high mutation frequency (1 in 10,000 gametes per generation; Riccardi and Eichner, 1986) often would be deletions, deletions have not yet been found in the NF1 region. Therefore, candidate genes must be screened in NF1 patients by techniques capable of detecting point mutations. This task will be made somewhat easier because approximately half of all NF1 patients carry a new (not inherited) mutation; in these people the NF1 mutation should be distinguishable as a DNA sequence found in neither parent. With the sequence data presented here, investigators can rapidly test whether *EVI2* plays a causative role in NF1.

ACKNOWLEDGMENTS

We thank V. M. Riccardi and D. Fults for providing NF1 tissues; M. Robertson for the fluorescent sequencing of *EVI2* cDNAs on the Applied Biosystems, Inc., Model 370A DNA sequencer; S. Nothwehr for analyzing *EVI2* with the SIGSEQ programs; and M. Emi, J. Groden, G. Joslin, L. Gelbert, and J. Swent for helpful discussions. R. Foltz edited the manuscript and prepared the figures. This work was supported in part by the National Cancer Institute, Department of Health and Human Services, under Contract N01-CO-74101 with BRI. R. White is an Investigator at the Howard Hughes Medical Institute.

Note added in proof. PFGE analysis of 40 additional NF1 patient DNAs reveals 2 samples, each with a smaller fragment in addition to the normal *Not*I fragment that encompasses the two translocation breakpoints. This provides additional support for the localization of *NF1* to this region.

REFERENCES

1. BADER, J. L., AND MILLER, R. W. (1978). Neurofibromatosis and childhood leukemia. *J. Pediatr.* **92:** 925–929.
2. BENTON, W. D., AND DAVIS, R. W. (1977). Screening lambda-gt recombinant clones by hybridization to single plaques in situ. *Science* **196:** 180–182.
3. BRAWERMAN, G. (1989). mRNA decay: Finding the right targets. *Cell* **57:** 9–10.
4. BUCHBERG, A. M., BEDIGIAN, H. G., TAYLOR, B. A., BROWNELL, E., IHLE, J. N., NAGATA, S., JENKINS, N. A., AND COPELAND, N. (1988). Localization of *Evi-2* to chromosome 11: Linkage to other protooncogene and growth factor loci using interspecific backcross mice. *Oncogene Res.* **2:** 149–165.
5. BUCHBERG, A. M., BROWNELL, E., NAGATA, S., JENKINS, N., AND COPELAND, N. (1989). A comprehensive genetic map of murine chromosome 11 reveals extensive linkage conservation between mouse and human. *Genetics* **122:** 153–161.
6. BUCHBERG, A. M., BEDIGIAN, H. G., JENKINS, N. A., AND COPELAND, N. G. Identification of a putative protooncogene, *Evi-2*, in murine myeloid leukemias that is closely linked to the von Recklinghausen neurofibromatosis locus in humans. Submitted for publication.
7. CHOMCZYNSKI, P., AND SACCHI, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Anal. Biochem.* **162:** 156–159.
8. CLARK, R. D., AND HUTTER, J. J., JR. (1982). Familial neurofibromatosis and juvenile chronic myelogenous leukemia. *Hum. Genet.* **60:** 230–232.
9. FEINBERG, A., AND VOGELSTEIN, B. (1984). Addendum: A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **137:** 266–267.
10. FOLZ, R. J., AND GORDON, J. I. (1987). Computer-assisted predictions of signal peptidase processing sites. *Biochem. Biophys. Res. Commun.* **146:** 870–877.
11. FROHMAN, M. A., DUSH, M. K., AND MARTIN, G. R. (1988). Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85:** 8998–9002.
12. GIERASCH, L. M. (1989). Signal sequences. *Biochemistry* **28:** 923–930.
13. GREENBERG, M. E., HERMANOWSKI, A. L., AND ZIFF, E. B. (1986). Effect of protein synthesis inhibitors on growth factor activation of c-fos, c-myc, and actin gene transcription. *Mol. Cell. Biol.* **6:** 1050–1057.
14. HALUSKA, F. G., TSUJIMOTO, Y., AND CROCE, C. M. (1987a). The t(8;14) chromosome translocation of the Burkitt lymphoma cell line Daudi occurred during immunoglobulin gene rearrangement and involved the heavy chain diversity region. *Proc. Natl. Acad. Sci. USA* **84:** 6835–6839.
15. HALUSKA, F. G., TSUJIMOTO, Y., AND CROCE, C. M. (1987b). Oncogene activation by chromosome translocation in human malignancy. *Annu. Rev. Genet.* **21:** 321–345.
16. HENGLEIN, B., SYNOVZIK, H., GROITL, P., BORNKAMM, G., HARTL, P., AND LIPP, M. (1989). Three breakpoints of variant t(2;8) translocations in Burkitt's lymphoma cell fall within a region 140 kilobases distant from c-*myc*. *Mol. Cell. Biol.* **9:** 2105–2113.
17. KOZAK, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* **12:** 857–872.
18. KYTE, J., AND DOOLITTLE, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157:** 105–132.
19. LANDSCHULZ, W. H., JOHNSON, P. F., AND MCKNIGHT, S. L. (1988). The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science* **240:** 1759–1764.
20. LEDBETTER, D. H., RICH, D. C., O'CONNELL, P., LEPPERT, M., AND CAREY, J. C. (1989). Precise localization of NF1 to 17q11.2 by balanced translocation. *Amer. J. Hum. Genet.* **44:** 20–24.
21. LOH, E. Y., ELLIOTT, J. F., CWIRLA, S., LANIER, L. L., AND

DAVIS, M. M. (1989). Polymerase chain reaction with single-sided specificity: Analysis of T cell receptor delta chain. *Science* **243:** 217–220.

22. MALTER, J. S. (1989). Identification of an AUUUA-specific messenger RNA binding protein. *Science* **246:** 664–666.
23. MANIATIS, T., FRITSCH, E. F., AND SAMBROOK, J. (1982a). Purification of closed circular DNA by centrifugation to equilibrium in cesium chloride–ethidium bromide gradients. "Molecular Cloning: A Laboratory Manual," pp. 93–94, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
24. MANIATIS, T., FRITSCH, E. F., AND SAMBROOK, J. (1982b). Electrophoresis of RNA through gels containing formaldehyde. "Molecular Cloning: A Laboratory Manual," pp. 202–203, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
25. MAYS, J. A., NEERHOUT, R. C., BAGBY, G. C., AND KOLER, R. D. (1980). Juvenile chronic granulocytic leukemia: Emphasis on cutaneous manifestations and underlying neurofibromatosis. *Amer. J. Dis. Child.* **134:** 654–658.
26. MENON, A. G., LEDBETTER, D. H., RICH, D. C., SEIZINGER, B. R., ROULEAU, G. A., MICHAELS, V. F., SCHMIDT, M. S., DEWALD, G., DELLA-TORRE, C. M., HAINES, J. L., AND GUSELLA, J. F. (1989). Characterization of a translocation within the von Recklinghausen neurofibromatosis region of chromosome 17. *Genomics* **5:** 245–249.
27. O'CONNELL, P., VISKOCHIL, D., BUCHBERG, A. M., FOUNTAIN, J., CAWTHON, R., CULVER, M., STEVENS, J., RICH, D. C., LEDBETTER, D. H., WALLACE, M., CAREY, J. C., JENKINS, N. A., COPELAND, N. G., COLLINS, F. S., AND WHITE, R. (1990). The human homolog of murine *Evi-2* lies between two von Recklinghausen neurofibromatosis translocations. *Genomics* **7:** 547–554.
28. ORITA, M., SUZUKI, Y., SEKIYA, T., AND HAYASHI, K. (1989). Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5:** 874–879.
29. PAGE, D. C., MOSHER, R., SIMPSON, E. M., FISHER, E. M. C., MARDON, G., POLLACK, J., MCGILLIVRAY, B., DE LA CHAPELLE, A., AND BROWN, L. G. (1987). The sex-determining region of the human Y chromosome encodes a finger protein. *Cell* **51:** 1091–1104.
30. PALMER, M. S., SINCLAIR, A. H., BERTA, P., ELLIS, N. A., GOODFELLOW, P. N., ABBAS, N. E., AND FELLOUS, M. (1989). Genetic evidence that ZFY is not the testis-determining factor. *Nature (London)* **342:** 937–939.
31. RICCARDI, V. M., AND EICHER, J. E. (1986). "Neurofibromatosis: Phenotype, Natural History, and Pathogenesis," Johns Hopkins Univ. Press, Baltimore, MD.
32. SANGER, F., NICKLEN, S., AND COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74:** 5463–5467.
33. SELDEN, R. F. (1989). Analysis of RNA by Northern hybridization. *In* "Current Protocols in Molecular Biology," pp. 4.9.1–4.9.8, Wiley, New York.
34. SHAW, G., AND KAMEN, R. (1986). A conserved AU sequence from the 3′ untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* **46:** 659–667.
35. SPOFFARD, A. (1976). Position effect variegation in Drosophila. *In* "Genetics and Biology of Drosophila" (M. Ashburner and E. Novitski, Eds.), Vol. 1c, pp. 955–1018, Academic Press, New York.
36. STUMPF, D. A., ALKSNE, J. F., ANNEGERS, J. F., BROWN, S. S., CONNEALLY, P. M., HOUSMAN, D., LEPPERT, M., MILLER, J. P., MOSS, M. L., PILEGGI, A. J., RAPIN, I., STROHMAN, R. C., SWANSON, L. W., AND ZIMMERMAN, A. (1987). Neurofibromatosis. "NIH Consensus Development Conference Statement": **6,** No. 12.
37. VINSON, C. R., SIGLER, P. B., AND MCKNIGHT, S. L. (1989). Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science* **246:** 911–916.
38. VON HEIJNE, G. (1985). Signal sequences: The limits of variation. *J. Mol. Biol.* **184:** 99–105.
39. VON HEIJNE, G. (1986). A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* **14:** 4683–4690.
40. WILSON, T., AND TREISMAN, R. (1988). Removal of poly(A) and consequent degradation of c-fos mRNA facilitated by 3′ AU-rich sequences. *Nature (London)* **336:** 396–399.