MOLECULAR AND CELLULAR BIOLOGY, Sept. 1986, p. 3109–3116
0270-7306/86/093109-08$02.00/0

Vol. 6, No. 9

Characterization of an Activated Human *ros* Gene

CARMEN BIRCHMEIER, DANIEL BIRNBAUM,† GAYLE WAITCHES, OTTAVIO FASANO,‡ AND MICHAEL WIGLER*

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Received 3 April 1986/Accepted 20 May 1986

A human oncogene, *mcf3*, previously detected by a combination of DNA-mediated gene transfer and a tumorigenicity assay, derives from a human homolog of the avian v-*ros* oncogene. Both v-*ros* and *mcf3* can encode a protein with homology to tyrosine-specific protein kinases, and both *mcf3* and v-*ros* encode a potential transmembrane domain N terminal to the kinase domain. *mcf3* probably arose during gene transfer from a normal human *ros* gene by the loss of a putative extracellular domain. There do not appear to be any other gross rearrangements in the structure of *mcf3*.

Cellular oncogenes, which are genes capable of altering the growth properties of normal cells, were first detected as the homologs of transforming genes of RNA tumor viruses (3). Additional cellular oncogenes have been discovered by their amplification in certain tumor cells (12, 21) and by DNA transfer techniques (1, 7, 14, 24). The most commonly used assay for detecting oncogenes after DNA transfer has been focus formation in NIH 3T3 cells (23). We have been exploring an alternate assay for oncogenes based on tumorigenicity in nude mice (4, 8). The details of the procedure we use have been published previously (8). In brief, DNA isolated from a tumor is cotransfected into NIH 3T3 cells in the presence of a G418 antibiotic resistance gene (25, 31). Cells which have taken up foreign DNA are selected by their resistance to G418, grown into colonies, pooled, and injected into animals.

We previously described our studies of cotransformation with DNA from the human mammary carcinoma cell line MCF-7 (8). After cotransforming NIH 3T3 cells with this DNA and injecting cotransformed cells into nude mice we obtained three primary tumors, called MCF-7-1, MCF-7-2, and MCF-7-3. DNA from each of these was capable of efficiently inducing secondary tumors after another round of cotransformation into NIH 3T3 cells and tumorigenicity assays. The MCF-7-1 tumor, and subsequent tumors derived from MCF-7-1 DNA, were shown to contain the human N-*ras* gene. DNA from all tumors derived from MCF-7-2 DNA contained a common gene which we called *mcf2*; and DNA from MCF-7-3 and its derived tumors all contained a gene which we called *mcf3*. We previously described the isolation of portions of *mcf2* and *mcf3* (8). Here we report the molecular characterization of *mcf3*. *mcf3* derives in part from the closest human homolog of the avian v-*ros* oncogene which we call *ros1*. The human *ros1* gene appears to have been activated during gene transfer. We present the nucleotide sequence of the activated gene. Based on the predicted amino acid sequence, the human *ros1* gene, like the chicken cellular *ros* gene (16), encodes a putative transmembrane protein kinase, possibly a growth factor receptor.

* Corresponding author.
† Present address: U.119 Institute National de la Santé et de la Recherche Médicale, 13009 Marseille, France.
‡ Present address: EMBL, D 6900 Heidelberg, Federal Republic of Germany.

MATERIALS AND METHODS

Cell culture, cotransfection and tumorigenicity assays. NIH 3T3 cells at 8×10^5 cells per plate were cotransformed with 300 ng of pKOneo plasmid DNA (30) and 5 μg of each of the cosmid fragments. The selection for G418 antibiotic resistance and the tumorigenicity assay were performed as previously described (8). Cell cultures were established from excised tumors after surgical removal and mincing and were maintained in Dulbecco medium plus 10% calf serum under our standard culture conditions. Nomenclature for tumors and cell lines derived from them is as follows. Independent secondary tumors and cell lines derived from MCF-7-3 DNA are called MCF-7-3-*n*, where *n* is a number. Independent tertiary tumors and cell lines derived from MCF-7-3-*n* DNA are called MCF-7-3-*n*-*m*, where *m* is a number.

Construction of libraries. DNAs were prepared from tumors or human placentas as previously described (8). Genomic libraries were constructed from placental or MCF-7-3-4 tumor DNAs by partial cleavage with *Eco*RI and cloning into the cosmid vector pHC79 (10). Appropriate fragments from a previously isolated bacteriophage lambda library (8) were used for colony filter hybridization. Additional overlapping cosmid clones were then isolated by hybridization with appropriate probes isolated from cosmid clones. cDNAs were synthesized from poly(A)$^+$ mRNA isolated from the nude mouse tumor-derived cell line MCF-7-3-7 (13). The cDNA library was constructed into λgt10 (11). Phages containing *mcf3* cDNAs were isolated by plaque hybridization, initially with the *Eco*RI fragment 1.4 kilobase pairs (kbp) in length isolated from cosmid clone 101 (see Fig. 1), and later with fragments isolated from cDNA clones.

DNA and RNA analysis. Southern blot analysis under high- and low-stringency conditions was performed as previously described (19, 24). DNA sequences were determined by the dideoxy method of Sanger et al. (20) as modified by Biggin et al. (2). Both strands of the coding sequences of the *mcf3* cDNA were sequenced. To localize the exon sequences, we performed Southern analysis using cDNA fragments as probes. A synthetic oligonucleotide with the sequence 5′-CCAACTATAATAGTAAGTATG-3′, which corresponds to the noncoding strand of the sequences encoding amino acids 10 to 15, was used as probe to localize the exon encoding the transmembrane domain. This oligonucleotide hybridized to the *Eco*RI fragment 4.8 kbp in length and was used as a primer for sequencing double-stranded DNA fragments from

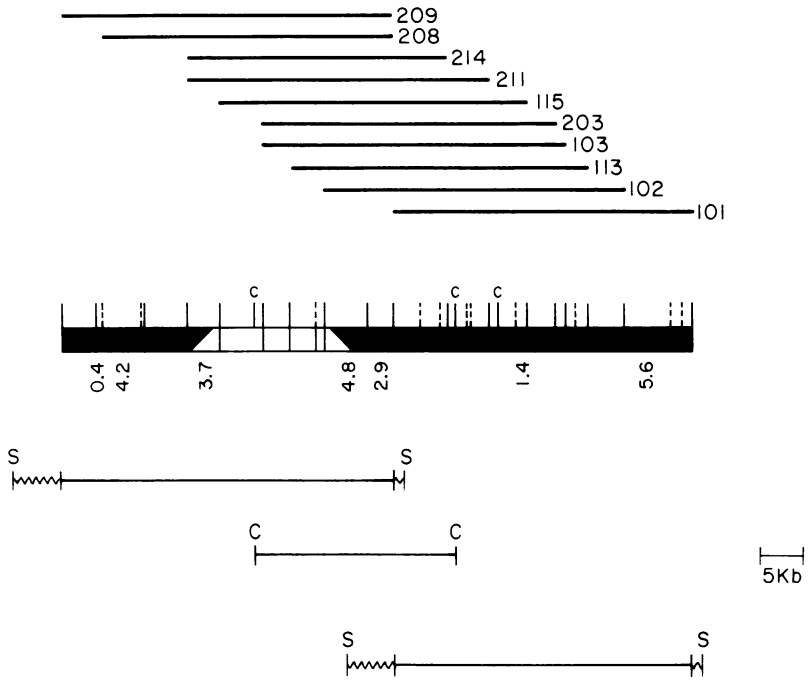


FIG. 1. Structure of *mcf3* locus. The inserts of the cosmid clones isolated from MCF-7-3-4 tumor DNA are shown in the upper panel. Numbers on the right are their identifiers. The *Eco*RI restriction sites deduced from the overlapping cosmid clones are shown by vertical lines, and the length of relevant *Eco*RI fragments (in kilobase pairs) are depicted below. Where there is ambiguity of fragment order, the *Eco*RI sites are designated with dashed vertical lines. The *Cla*I restriction sites are indicated by C. The darkened bars indicate regions of the *mcf3* locus composed of contiguous human DNA. The clear bar indicates the region of the *mcf3* locus of undetermined origin. The cosmids 101 and 209, linearized at their unique *Sal*I sites in the vector DNA, and the *Cla*I-*Cla*I fragment of cosmid 115 which were used for cotransfection and reconstitution of an intact *mcf3* locus are shown in the lower panel. Squiggly lines are cosmid vector sequences.

both placental and *mcf3* cosmid clones by a modification of the dideoxy method (6). The oligonucleotides used in this study were synthesized on an Applied Biosystems DNA synthesizer and purified from polyacrylamide gels as described previously (34). RNA was prepared from cell lines by the guanidinium-hot phenol method (13) and analyzed on Northern blots as described previously (27). RNA protection studies with labeled DNA as hybridization probes were performed as described previously (15).

RESULTS

Structure of the *mcf3* locus. We previously described a set of Charon 4A phage clones containing portions of the *mcf3* locus (8). To obtain a complete set of clones from the *mcf3* locus, we chose to use the cosmid cloning vector pHC79 (10) and constructed a cosmid library from the DNA of a secondary nude mouse tumor MCF-7-3-4. DNA sequences from the Charon phages were used initially as probes to isolate clones from the cosmid library. Overlapping cosmids were cloned until we obtained clones which contained sequences not found in all tumors derived by cotransformation with DNA from MCF-7-3. An *Eco*RI map of this 70-kbp region is depicted in Fig. 1. Only *Eco*RI fragments to the left and including the 5.6-kbp fragment and to the right and including the 0.4-kbp fragment were found in all *mcf3* transformants.

To demonstrate that we had cloned the entire transforming gene and to determine the location of this gene in the 70-kbp-long stretch of cloned DNA, cosmid clones from different regions of the *mcf3* locus were linearized at the unique *Sal*I site in the cosmid vector and tested for their ability to induce the tumorigenic phenotype in cotransfected NIH 3T3 cells. None of the individually tested cosmids scored positive in the assay, indicating that the transforming gene extends over a major part of the *mcf3* locus. Similarly, no pairwise combination of cosmids was able to induce tumorigenicity in NIH 3T3 cells. Therefore, we performed cotransformation experiments with three overlapping pieces of DNA: two linearized cosmids and a third cosmid fragment, which together span the whole locus. The pieces used were cosmid clones 101 and 209, linearized at their unique *Sal*I sites, and the *Cla*I fragment of cosmid clone 115 (Fig. 1, lower panel). The combination of all three fragments scored positive in the cotransfection and tumorigenicity assay, whereas no combination of two fragments did (data not shown). Therefore, to form an intact *mcf3* oncogene, two homologous recombination events had to take place. This shows that we cloned the entire *mcf3* oncogene. It is clearly a very large gene, which may extend over nearly 70 kbp of DNA.

To compare the structure of the *mcf3* locus with normal cellular DNA, we constructed a cosmid library from human placental DNA. Appropriate DNA fragments from the *mcf3* locus were used for screening this library, and five independent cosmid clones were obtained. The DNAs present in these clones could not be aligned over their full length with the *mcf3* locus. Additional cloning and Southern analysis indicated that the different structure of the placental cosmid clones and the *mcf3* locus was not an artifact of cosmid cloning (data not shown). The *mcf3* locus was created by DNA rearrangement involving the fusion of at least three separate fragments of DNA. The first of these is a fragment of human DNA located at what we later show to be the 5′ end of the *mcf3* oncogene, and it extends from the left end of

the map of the *mcf3* locus to the *Eco*RI fragment that is 3.7 kbp in length. This fragment is connected by DNA of unknown origin to another fragment of the human genome which contains the majority of the coding sequences of *mcf3*. The second piece of human DNA extends from the *Eco*RI fragment that is 4.8 kbp in length to the right end of the map of the *mcf3* locus. The lengths and positions of these three fragments are indicated in Fig. 1. The DNA rearrangements which created the *mcf3* locus are probably of functional significance since they span the regions that were required in the cotransfection studies described above.

Analysis and identification of *mcf3* oncogene. Since the *mcf3* gene is large, cloning cDNAs to the transcript of the *mcf3* oncogene was necessary for the further analysis of the gene. To localize the transcription unit in the *mcf3* locus, we analyzed poly(A)$^+$ RNA from a cell line established from the secondary nude mouse tumor MCF-7-3-7. We used various fragments from the *mcf3* locus as probes in Northern blots. The 1.4-kbp *Eco*RI fragment contained in cosmid 101 (Fig. 1) detected several RNA species of 2.8 to 3.3 kb in RNA from MCF-7-3-7 cells (data not shown). These RNA species were not found in RNA from normal NIH 3T3 cells. The *Eco*RI fragment 1.4 kbp in length was then used as probe for screening a cDNA library. The library was constructed by cloning cDNAs synthesized from poly(A)$^+$ RNA of the MCF-7-3-7 cell line into the λgt10 cloning vector (11). The 3′ end of one transcript was localized in cDNA clone M3.9, which contained a poly(A)$^+$ tail. This also determined the direction of transcription in the *mcf3* locus which was subsequently verified by S1 mapping (data not shown).

Heterogeneity in the structure of the isolated cDNA clones was revealed when they were hybridized to *Eco*RI digests of *mcf3* DNA. The heterogeneity was confined to sequences derived from the 5′ end of the gene and indicated that different splicing events occur at the 5′ end of the *mcf3* transcripts (data not shown). All the isolated cDNA clones contained a portion with common structure which hybridized to a common set of *Eco*RI fragments from the *mcf3* locus. These common sequences are transcribed from, and completely contained within, the 3′ portion of the locus which derives from the long contiguous piece of human DNA.

To obtain information about the nature of the *mcf3* transforming protein, we determined the sequence of the open reading frame in the common portion of the *mcf3* cDNAs. These nucleotide sequences and the predicted amino acid sequences are shown in Fig. 2. Initial computer analysis of the amino acid sequence indicated that the *mcf3* protein sequence was similar to that of all oncogenes encoding tyrosine kinases, but closely related to none. When the sequence of v-*ros*, the transforming gene of the avian sarcoma virus UR2, was published (17), it was immediately obvious that the *mcf3* protein was closely related to v-*ros*. For comparison, the amino acid sequence of the v-*ros* protein is shown in Fig. 3. From amino acid 51 to 370 of *mcf3*, 75% homology exists between the two proteins. Only at the C terminus do they differ considerably, where the *mcf3*-encoded protein contains 99 additional amino acids not found in the v-*ros* protein. The close homology between v-*ros* and *mcf3* initially surprised us since in previous studies we found that a fragment of the *mcf3* gene, represented in RNA transcripts, showed no homology to a panel of cloned retroviral oncogenes which included v-*ros* (8). However, the fragment of *mcf3* used in the earlier studies encodes the C-terminal part of the *ros1* protein and has limited homology to v-*ros*.

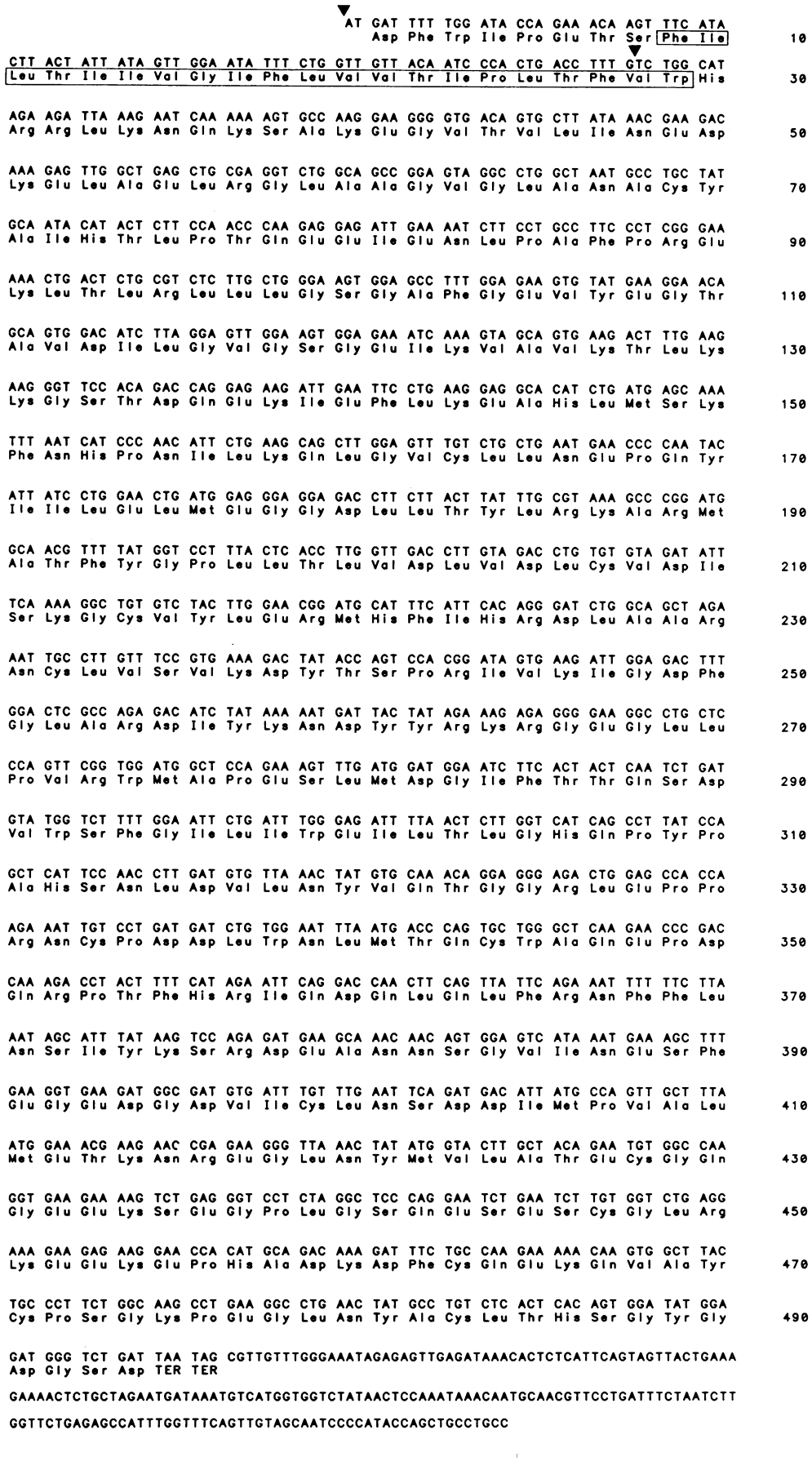
The close homology between the chicken-derived v-*ros* gene and the common part of the *mcf3* cDNAs suggests that the 3′ portion of the *mcf3* locus derives from the human counterpart of the v-*ros* gene. To determine whether this is indeed the case, we performed Southern analyses of total human DNA under conditions of low stringency with two probes, a v-*ros* and an *mcf3* cDNA fragment, which encode roughly analogous sequences of the *ros* proteins. The probe from the *mcf3* cDNA hybridized most strongly to two *Eco*RI fragments, 2.9 and 5.0 kbp, which correspond to two fragments of human DNA known to contain exon sequences (Fig. 4A, lanes 2 and 3). The v-*ros* probe hybridized most strongly to an *Eco*RI fragment in human DNA 2.9 kbp in length, corresponding to the same fragment of the *mcf3* locus (Fig. 4B, lanes 2 and 3). We conclude, therefore, that a major portion of the *mcf3* coding sequence is derived from the gene in humans most closely related to the v-*ros* gene. We call this gene *ros1*, since other human genes related to the v-*ros* or the *mcf3* gene may exist.

Potential membrane-spanning domain and rearrangement of *mcf3* gene. Five of the oncogenes known to encode oncogenic tyrosine kinases, v-*ros* (17), v-*erbB* (32), v-*fms* (9), *neu* (1), and *trk* (14), have hydrophobic potential membrane-spanning domains. The membrane-spanning domains of these proteins are always encoded 5′ to sequences encoding the kinase domain. Inspection of the *mcf3* nucleotide sequence shows that it can encode a highly hydrophobic stretch of 21 amino acids immediately followed by a stretch rich in positively charged amino acids (Fig. 2, boxed sequence). These features are commonly found in membrane-spanning domains. The hydrophobic sequences of v-*ros* are 30 amino acids long, longer than in *mcf3* and longer than is needed for such a domain.

We wanted to know whether the portion of the *mcf3* cDNA encoding the potential transmembrane domain derived from the human *ros1* gene or from other parts of the *mcf3* locus. By combined restriction endonuclease analysis, Southern blotting, and hybridization with synthetic oligonucleotides, two coding exons were localized and sequenced within the 4.8-kbp *Eco*RI fragment which contains the breakpoint in the *ros1*-derived part of the *mcf3* locus (Fig. 1). The positions of the deduced splice junctions are included in Fig. 2. One exon encodes sequences for the putative intracellular domain and for one amino acid of the potential transmembrane domain. The other encodes sequences from the potential transmembrane domain and for eight amino acids of the putative extracellular domain. This last exon is also found in placental DNA clones containing the human *ros1* gene and therefore does not derive from rearranged sequences (data not shown). Homologs to the eight amino acids of the putative extracellular domain are not found in the avian v-*ros* gene.

There was no consensus structure for cDNA derived from parts of the gene 5′ to the breakpoint found in the 4.8-kbp *Eco*RI fragment. The heterogeneity observed in the structure of the cDNAs probably reflects different splicing patterns of transcripts arising from regions of the *mcf3* oncogene which do not derive from *ros1*. Since we do not known which, if any, of our cDNA clones reflect transcripts encoding a transforming protein, we did not analyze these portions of our cDNA clones. From the position of the coding region within the cDNA clones and the length of the longest *mcf3* transcript, we estimate the maximal size of an *mcf3*-encoded protein to be 75 kilodaltons, of which 50 kilodaltons derive from *ros1* sequences.

***ros1* is not rearranged in MCF-7 cells.** We showed in the


AT GAT TTT TGG ATA CCA GAA ACA AGT TTC ATA
Asp Phe Trp Ile Pro Glu Thr Ser Phe Ile 10
CTT ACT ATT ATA GTT GGA ATA TTT CTG GTT GTT ACA ATC CCA CTG ACC TTT GTC TGG CAT
Leu Thr Ile Ile Val Gly Ile Phe Leu Val Val Thr Ile Pro Leu Thr Phe Val Trp His 30
AGA AGA TTA AAG AAT CAA AAA AGT GCC AAG GAA GGG GTG ACA GTG CTT ATA AAC GAA GAC
Arg Arg Leu Lys Asn Gln Lys Ser Ala Lys Glu Gly Val Thr Val Leu Ile Asn Glu Asp 50
AAA GAG TTG GCT GAG CTG CGA GGT CTG GCA GCC GGA GTA GGC CTG GCT AAT GCC TGC TAT
Lys Glu Leu Ala Glu Leu Arg Gly Leu Ala Ala Gly Val Gly Leu Ala Asn Ala Cys Tyr 70
GCA ATA CAT ACT CTT CCA ACC CAA GAG GAG ATT GAA AAT CTT CCT GCC TTC CCT CGG GAA
Ala Ile His Thr Leu Pro Thr Gln Glu Glu Ile Glu Asn Leu Pro Ala Phe Pro Arg Glu 90
AAA CTG ACT CTG CGT CTC TTG CTG GGA AGT GGA GCC TTT GGA GAA GTG TAT GAA GGA ACA
Lys Leu Thr Leu Arg Leu Leu Leu Gly Ser Gly Ala Phe Gly Glu Val Tyr Glu Gly Thr 110
GCA GTG GAC ATC TTA GGA GTT GGA AGT GGA GAA ATC AAA GTA GCA GTG AAG ACT TTG AAG
Ala Val Asp Ile Leu Gly Val Gly Ser Gly Glu Ile Lys Val Ala Val Lys Thr Leu Lys 130
AAG GGT TCC ACA GAC CAG GAG AAG ATT GAA TTC CTG AAG GAG GCA CAT CTG ATG AGC AAA
Lys Gly Ser Thr Asp Gln Glu Lys Ile Glu Phe Leu Lys Glu Ala His Leu Met Ser Lys 150
TTT AAT CAT CCC AAC ATT CTG AAG CAG CTT GGA GTT TGT CTG CTG AAT GAA CCC CAA TAC
Phe Asn His Pro Asn Ile Leu Lys Gln Leu Gly Val Cys Leu Leu Asn Glu Pro Gln Tyr 170
ATT ATC CTG GAA CTG ATG GAG GGA GGA GAC CTT CTT ACT TAT TTG CGT AAA GCC CGG ATG
Ile Ile Leu Glu Leu Met Glu Gly Gly Asp Leu Leu Thr Tyr Leu Arg Lys Ala Arg Met 190
GCA ACG TTT TAT GGT CCT TTA CTC ACC TTG GTT GAC CTT GTA GAC CTG TGT GTA GAT ATT
Ala Thr Phe Tyr Gly Pro Leu Leu Thr Leu Val Asp Leu Val Asp Leu Cys Val Asp Ile 210
TCA AAA GGC TGT GTC TAC TTG GAA CGG ATG CAT TTC ATT CAC AGG GAT CTG GCA GCT AGA
Ser Lys Gly Cys Val Tyr Leu Glu Arg Met His Phe Ile His Arg Asp Leu Ala Ala Arg 230
AAT TGC CTT GTT TCC GTG AAA GAC TAT ACC AGT CCA CGG ATA GTG AAG ATT GGA GAC TTT
Asn Cys Leu Val Ser Val Lys Asp Tyr Thr Ser Pro Arg Ile Val Lys Ile Gly Asp Phe 250
GGA CTC GCC AGA GAC ATC TAT AAA AAT GAT TAC TAT AGA AAG AGA GGG GAA GGC CTG CTC
Gly Leu Ala Arg Asp Ile Tyr Lys Asn Asp Tyr Tyr Arg Lys Arg Gly Glu Gly Leu Leu 270
CCA GTT CGG TGG ATG GCT CCA GAA AGT TTG ATG GAT GGA ATC TTC ACT ACT CAA TCT GAT
Pro Val Arg Trp Met Ala Pro Glu Ser Leu Met Asp Gly Ile Phe Thr Thr Gln Ser Asp 290
GTA TGG TCT TTT GGA ATT CTG ATT TGG GAG ATT TTA ACT CTT GGT CAT CAG CCT TAT CCA
Val Trp Ser Phe Gly Ile Leu Ile Trp Glu Ile Leu Thr Leu Gly His Gln Pro Tyr Pro 310
GCT CAT TCC AAC CTT GAT GTG TTA AAC TAT GTG CAA ACA GGA GGG AGA CTG GAG CCA CCA
Ala His Ser Asn Leu Asp Val Leu Asn Tyr Val Gln Thr Gly Gly Arg Leu Glu Pro Pro 330
AGA AAT TGT CCT GAT GAT CTG TGG AAT TTA ATG ACC CAG TGC TGG GCT CAA GAA CCC GAC
Arg Asn Cys Pro Asp Asp Leu Trp Asn Leu Met Thr Gln Cys Trp Ala Gln Glu Pro Asp 350
CAA AGA CCT ACT TTT CAT AGA ATT CAG GAC CAA CTT CAG TTA TTC AGA AAT TTT TTC TTA
Gln Arg Pro Thr Phe His Arg Ile Gln Asp Gln Leu Gln Leu Phe Arg Asn Phe Phe Leu 370
AAT AGC ATT TAT AAG TCC AGA GAT GAA GCA AAC AAC AGT GGA GTC ATA AAT GAA AGC TTT
Asn Ser Ile Tyr Lys Ser Arg Asp Glu Ala Asn Asn Ser Gly Val Ile Asn Glu Ser Phe 390
GAA GGT GAA GAT GGC GAT GTG ATT TGT TTG AAT TCA GAT GAC ATT ATG CCA GTT GCT TTA
Glu Gly Glu Asp Gly Asp Val Ile Cys Leu Asn Ser Asp Asp Ile Met Pro Val Ala Leu 410
ATG GAA ACG AAG AAC CGA GAA GGG TTA AAC TAT ATG GTA CTT GCT ACA GAA TGT GGC CAA
Met Glu Thr Lys Asn Arg Glu Gly Leu Asn Tyr Met Val Leu Ala Thr Glu Cys Gly Gln 430
GGT GAA GAA AAG TCT GAG GGT CCT CTA GGC TCC CAG GAA TCT GAA TCT TGT GGT CTG AGG
Gly Glu Glu Lys Ser Glu Gly Pro Leu Gly Ser Gln Glu Ser Glu Ser Cys Gly Leu Arg 450
AAA GAA GAG AAG GAA CCA CAT GCA GAC AAA GAT TTC TGC CAA GAA AAA CAA GTG GCT TAC
Lys Glu Glu Lys Glu Pro His Ala Asp Lys Asp Phe Cys Gln Glu Lys Gln Val Ala Tyr 470
TGC CCT TCT GGC AAG CCT GAA GGC CTG AAC TAT GCC TGT CTC ACT CAC AGT GGA TAT GGA
Cys Pro Ser Gly Lys Pro Glu Gly Leu Asn Tyr Ala Cys Leu Thr His Ser Gly Tyr Gly 490
GAT GGG TCT GAT TAA TAG CGTTGTTTGGGAAATAGAGAGTTGAGATAAACACTCTCATTCAGTAGTTACTGAAA
Asp Gly Ser Asp TER TER
GAAAACTCTGCTAGAATGATAAATGTCATGGTGGTCTATAACTCCAAATAAACAATGCAACGTTCCTGATTTCTAATCTT
GGTTCTGAGAGCCATTTGGTTTCAGTTGTAGCAATCCCCATACCAGCTGCCTGCC

```
mcf3    (1)    NFWIPETSFILTIIVGIFLVVT---IPLT-FVWHRRLKNQKSAKEGVTVLINEDKELAEL
                  *   *   *   *   *  * ** **** * *  *  *  *  *  **  ****** *
        (153)  TVTSPDITAIVAVIGAVVLGLTSLTIILTGFVWHQRWKSRKPASTGQIVLVKEDKELAQL

mcf3    (57)   RGLAAGVGLANACYAIHTLPTQEEIENLPAFPREKLTLRLLLGSGAFGEVYEGTAVDILG
               ** *  *********  *** * *** ****** ** *  ************** ***
v-ros   (213)  RGMAETVGLANACYAVSTLPSQAEIESLPAFPRDKLNLHKLLGSGAFGEVYEGTALDILA

mcf3    (117)  VGSGEIKVAVKTLKKGSTDQEKIEFLKEAHLMSKFNHPNILKQLGVCLLNEPQYIILELM
                ****  ******* * ***** ************ ** *** ********** *****
v-ros   (273)  DGSGESRVAVKTLKRGATDQEKSEFLKEAHLMSKFDHPHILKLLGVCLLNEPQYLILELM

mcf3    (177)  EGGDLLTYLRKARMATFYGPLLTLVDLVDLCVDISKGCVYLERMHFIHRDLAARNCLVSV
               ****** *** **   *  ***** ** * * ** ******* * **************
v-ros   (333)  EGGDLLSYLRGARKQKFQSPLLTLTDLLDICLDICKGCVYLEKMRFIHRDLAARNCLVSE

mcf3    (237)  KDYTS-PRIVKIGDFGLARDIYKNDYYRKRGEGLLPVRWMAPESLMDGIFTTQSDVWSFG
               * * *  * ************************************ ** **  **** **
v-ros   (393)  KQYGSCSRVVKIGDFGLARDIYKNDYYRKRGEGLLPVRWMAPESLIDGVFTNHSDVWAFG

mcf3    (296)  ILIWEILTLGHQPYPAHSNLDVLNYVQTGGRLEPPRNCPDDLWNLMTQCWAQEPDQRPTF
                * ** **** ****  **  **  *  ***** * *****  *** **** *  ****
v-ros   (453)  VLVWETLTLGQQPYPGLSNIEVLHHVRSGGRLESPNNCPDDIRDLMTRCWAQDPHNRPTF

mcf3    (356)  HRIQDQLQLFRNFFLNSIYKSRDEANNSGVINESFEGEDG...
                  **  **  *   *   *   *
v-ros   (513)  FYIQHKLQEIRHSPLCFSYFLGDKESVAPLRIQTAFFQPL
```

FIG. 3. Comparison of deduced amino acid sequences of *mcf3* and v-*ros* genes. The predicted amino acid sequences of the gene products of *mcf3* and v-*ros* (17) are shown. The residues shared between *mcf3* and v-*ros* are indicated by asterisks. The numbers to the left refer to the first amino acid shown in the corresponding line. The end of the v-*ros* sequence is shown, whereas the *mcf3* sequence continues for another 99 amino acids.

above studies that the *mcf3* gene derives from a rearrangement involving the human *ros1* gene and that putative extracellular sequences of this gene have been lost. Since similar rearrangements have occurred during the biogenesis of the v-*ros* (17), v-*erbB* (32), and *trk* (14) genes, they are probably of functional significance. It was thus of interest to determine whether DNA from the MCF-7 cell line used in the original cotransformation studies contained the rearranged *mcf3* locus or whether this rearrangement was introduced during DNA transfer. The structure of the locus in the MCF-7 cell line was analyzed by Southern blotting and compared with either the normal configuration of the locus in placental DNA or the rearranged configuration in MCF-7-3 and its derived tumors. The probe used for these experiments contained sequences from the 4.8-kbp *Eco*RI fragment which contains one of the breakpoints introduced during rearrangement. This probe hybridized to a *Bam*HI fragment approximately 12 kbp in length in tumor DNAs containing the rearranged *mcf3* oncogene (Fig. 5, lanes 4 to 6), but hybridized to a band approximately 10 kbp in length in DNA from MCF-7 cells and from placental DNA (Fig. 5, lanes 1 and 2). Therefore, we can conclude that the rearrangement responsible for the creation of the *mcf3* oncogene had not occurred in MCF-7 DNA. The alteration is only found in DNA from MCF-7-3 and its derived tumors and must therefore have occurred during or after DNA transfer into NIH 3T3 cells. The blot hybridization studies also indicate that the *mcf3* oncogene is very highly amplified in all transformed NIH 3T3 cells, a point which we made previously (8). The activation of the *ros1* gene was therefore probably an artifact of DNA transfer. To test this hypothesis further, we analyzed RNA from the MCF-7 cell line by Northern blotting and by RNA protection studies (see Materials and Methods). We could not detect expression of the *ros1* gene in MCF-7 cells. If expressed at all, levels of the *ros1* transcript in MCF-7 cells must be 50-fold lower than levels found in cells transformed with the *mcf3* locus. We conclude that the oncogenic potential of the human *ros1* gene may have been activated by rearrangement and gene amplification occurring during or after gene transfer.

DISCUSSION

We presented data describing the structure of an oncogene we have called *mcf3* (8). The gene was detected by a combination of DNA-mediated gene transfer and a tumorigenicity assay in nude mice. The DNA originally used in the first transfections came from MCF-7 cells, a human mammary carcinoma cell line (8). The *mcf3* oncogene was isolated by molecular cloning, and its structure was compared with the structure of its normal counterpart in human placental DNA. We found that the *mcf3* oncogene was a product of a major DNA rearrangement. DNA cotransfection studies indicated that this rearrangement spanned functionally important domains of *mcf3*. The peculiar rearrangement associated with the *mcf3* oncogene was shown to be present only in DNA isolated from transfected and

FIG. 2. Sequence of common region of *mcf3* cDNA clones. The nucleotide sequence of the common region of the *mcf3* cDNA clones is depicted. Below the nucleotide sequence, the predicted amino acid sequence is shown. The numbers at the end of each line refer to the position of the predicted amino acid residues, with position +1 defining the first amino acid of the *mcf3* cDNA which is encoded by the *ros1*-derived part of the locus. The potential membrane-spanning domain of 21 hydrophobic amino acids is boxed. The positions of two splice junctions close to the point of rearrangement are indicated by arrowheads.

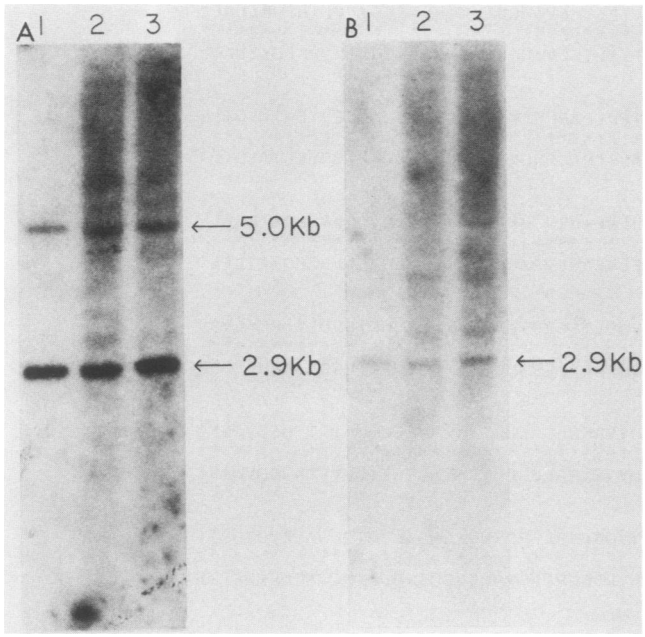


FIG. 4. Southern blot analysis of the human genome with *mcf3* cDNA and v-*ros* probes. Genomic DNA (10 μg) and DNA (0.1 ng) from a cosmid clone (3P7) derived from human placental DNA were digested with *Eco*RI and analyzed by Southern hybridization under low-stringency conditions. A fragment from an *mcf3* cDNA clone encoding amino acids 3 to 150 and a v-*ros* fragment encoding amino acids 1 to 305 of the *gag-ros* fusion protein were used as probes in panels A and B, respectively. The numbers on the right indicate the sizes of the DNA fragments of the *ros1* gene hybridizing to the *mcf3* or v-*ros* probe. Lanes: 1, DNA from cosmid clone 3P7; 2, DNA from human placenta; 3, DNA from MCF-7 cells.

tumorigenic cells, but not in the original donor DNA from the MCF-7 cell line. Since we cannot detect any expression of the *mcf3* gene in the MCF-7 cells, it seems unlikely that this gene contributes to the transformed phenotype of MCF-7. Thus, a functional *mcf3* gene was created by a rearrangement introduced during or after gene transfer. We have found a second instance of this in an independent line of experiments (33). Rearrangements and amplifications also occur in the standard NIH 3T3 focus assay and have been reported to lead to proto-oncogene activation (26). In our experience, such events are rare in the focus assay but common with the cotransformation and tumorigenicity assay. The cotransformation and tumorigenicity assay, therefore, may be unreliable for the detection of oncogenes in tumor DNAs. However, it may be a good method for searching for proto-oncogenes which can be activated by rearrangement or amplification.

Sequence analysis indicated that *mcf3* was related to the v-*ros* gene of the avian retrovirus UR2 (17). Both *mcf3* and v-*ros* probes hybridize most strongly to the same DNA fragment in human DNA, and we therefore designated the gene from which *mcf3* was derived as *ros1*. We use this nomenclature since there may be other *ros*-related genes in the human genome. In the human genome, the *ros1* gene is localized on chromosome 6, bands q11 to q22 (M. Rabin, D. Birnbaum, M. Wigler, and F. Ruddle, Am. J. Hum. Genet. **37**:A36). Translocations of chromosome 6 in this region have been previously observed in human tumors. Thus, although *ros1* is probably not activated in MCF-7 tumor cells, it is possible that chromosomal alterations affect the *ros1* locus and lead to an activation of its oncogenic potential in other tumor cells. Although most human tumor cell lines, like MCF-7, do not express detectable levels of *ros1*, we have found several human tumor cell lines that do contain significant levels of *ros1* transcripts (preliminary data).

The majority of known oncogenes encode proteins with tyrosine-specific protein kinase activity. However, potential membrane-spanning domains N terminal to kinase domains are found only in v-*erbB* (32), v-*fms* (9), *neu* (1), *trk* (14), and the v-*ros* and cellular *ros* genes (16, 17). The cellular analogs of the v-*erbB* and the v-*fms* genes probably encode the receptors for the epidermal growth factor and macrophage colony-stimulating factor, respectively (22, 29). Thus, the cellular *ros* gene very likely encodes a hormone receptor as well. As has been noted before, the tyrosine kinase most closely related to *ros* is the insulin receptor (28). In particular, a stretch of very high (75%) homology to the insulin receptor exists between positions 245 and 288 in *ros1*, which can be aligned with positions 178 to 215 in the cyclic AMP-dependent protein kinase. Since Cys-198 of the cyclic AMP-dependent protein kinase is protected from chemical modification by peptide substrates, this region has been implicated in substrate binding (5). The high degree of homology between the amino acid sequences of *ros1* and the insulin receptor in this putative substrate-binding domain might indicate a similar substrate specificity for the tyrosine kinase activities of these two proteins.

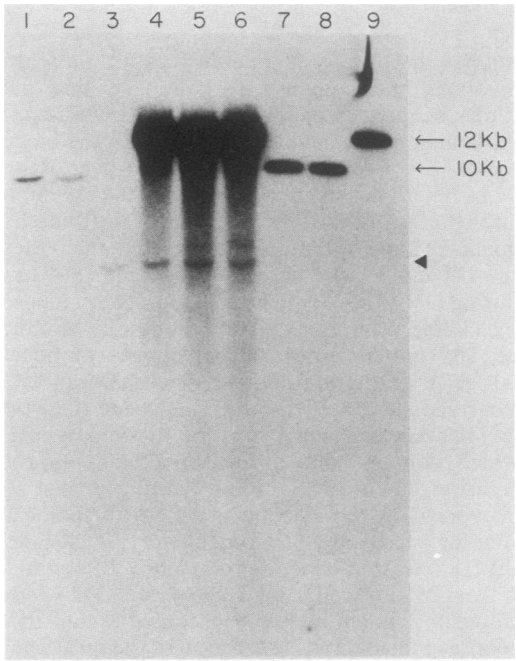


FIG. 5. Rearrangement of *mcf3* locus in MCF-7-3 tumor DNAs. Genomic DNA (10 μg) and cosmid DNA (1 ng) were digested with *Bam*HI and analyzed by Southern hybridization. A cDNA fragment encoding amino acids 3 to 150 was used as a probe (see the legend to Fig. 4). The numbers on the right indicate the sizes of the fragments in human DNA hybridizing to the cDNA fragment. The arrowhead indicates the fragment in mouse DNA which hybridized to the human cDNA fragment. Lanes: 1, DNA from human placenta; 2, DNA from MCF-7 cells; 3, DNA from NIH 3T3 cells; 4, DNA from the nude mouse tumor MCF-7-3; 5, DNA from the nude mouse tumor MCF-7-3-1; 6, DNA from the nude mouse tumor MCF-7-3-1-1; 7 and 8, DNA from cosmid clone 3P7 (see the legend to Fig. 4); 9, DNA from cosmid clone 115.

One effect of the rearrangement which created the oncogenic *mcf3* gene is a deletion of all but eight amino acids of the putative extracellular domain of *ros1*. DNA pieces of unknown origin replaced this part of the *ros1* gene in the *mcf3* locus. In the v-*ros* gene the point of fusion between cellular and viral sequences is located in an analogous position leading to the loss of most of the putative extracellular domain (16). The deletion of the extracellular domain may be an important event in the activation of the oncogenic potential of the *ros1* gene. Similar events have been observed previously for the v-*erbB* (32) and the c-*erbB* (18) gene in avian leukosis virus-induced erythroblastosis. Insertion of the provirus into the middle of the c-*erbB* gene leads to the production of a truncated *erbB* transcript which encodes only 64 amino acids of the extracellular domain but an intact membrane-spanning and intracellular domain. Similarily, *trk* seems to have been formed by a somatic rearrangement that replaced the extracellular domain of a putative transmembrane receptor with the first 221 amino acids of a nonmuscle tropomyosin protein (14). We have not analyzed the nature of the sequences in *mcf3* which have replaced the extracellular domain of the *ros1* gene. However, we should be cautious in concluding that the loss of these extracellular sequences has led to the activation of *ros1*. The amino acid sequence of *mcf3* from position 1 to 392 is identical to the deduced coding sequence of *ros1* as determined from human placental DNA (H. Matsushime, L.-H. Wang, and M. Shibuya, personal communication), but the corresponding coding sequences C terminal to amino acid 392 have not yet been determined for *ros1*. Although we could not detect any gross structural differences between the DNA of *mcf3* and *ros1* coding for the C-terminal part of the protein, we would not have detected any subtle changes such as point mutations or small deletions by our methods of analysis. Moreover, the rearranged *ros1* gene is very highly amplified in all *mcf3*-transformed NIH 3T3 cells we examined. We cannot presently assess the relative contributions of rearrangement and amplification of this gene to its oncogenicity.

ACKNOWLEDGMENTS

We thank J. Loveless for performing nude mice tumorigenicity assays, J. Lewis for advice concerning the λgt10 cDNA cloning system, M. Zoller for the synthesis of oligonucleotides, and P. Bird for her help in preparing this manuscript.

This work was supported by grants from the American Cancer Society, American Business for Cancer Research Foundation, and the National Institutes of Health. C.B. was supported by the Swiss National Science Foundation. D.B. was supported by Institute National de la Santé et de la Recherche Médicale and Philippe Foundation fellowships. M.W. is an American Cancer Society Research Professor.

LITERATURE CITED

1. **Bargmann, C. I., M.-C. Hung, and R. A. Weinberg.** 1986. The *neu* oncogene encodes an epidermal growth factor receptor-related protein. Nature (London) **319**:226–230.
2. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80**:3963–3965.
3. **Bishop, J. M.** 1985. Viral oncogenes. Cell **42**:23–38.
4. **Blair, D. G., C. S. Cooper, M. K. Oskarsson, L. A. Eader, and G. F. Vande Woude.** 1982. New method for detecting cellular transforming genes. Science **281**:1122–1125.
5. **Branson, H. N., N. Thomas, R. Matsueda, N. C. Nelson, S. S. Taylor, and E. T. Kaiser.** 1982. Modification of the catalytic subunit of bovine heart cAMP-dependent protein kinase with affinity labels related to peptide substrates. J. Biol. Chem. **257**:10575–10582.
6. **Chen, E. Y., and P. H. Seeburg.** 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA **4**:165–170.
7. **Cooper, C. S., M. Park, D. G. Blair, M. A. Tainsky, K. Huebner, C. M. Croce, and G. F. Vande Woude.** 1984. Molecular cloning of a new transforming gene from a chemically transformed human cell line. Nature (London) **311**:29–33.
8. **Fasano, O., D. Birnbaum, L. Edlund, J. Fogh, and M. Wigler.** 1984. New human transforming genes detected by a tumorigenicity assay. Mol. Cell. Biol. **4**:1695–1705.
9. **Hampe, A., M. Gobet, C. Sherr, and F. Galibert.** 1984. Nucleotide sequence of the feline retroviral oncogene v-*fms* shows unexpected homology with oncogenes encoding tyrosine-specific protein kinases. Proc. Natl. Acad. Sci. USA **81**:85–89.
10. **Hohn, B., and J. Collins.** 1980. A small cosmid for efficient cloning of large DNA fragments. Gene **11**:291–298.
11. **Huynh, T. V., R. A. Young, and R. W. Davis.** 1984. Constructing and screening cDNA libraries in λgt10 and λgt11, p. 49–78. *In* D. Glover (ed.), DNA cloning techniques: a practical approach. IRL Press, Oxford.
12. **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.
13. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
14. **Martin-Zanca, D., S. Hughes, and M. Barbacid.** 1986. A human oncogene formed by the fusion of a truncated tropomyosin and protein tyrosine kinase sequences. Nature (London) **319**:743–748.
15. **Melton, D. A., P. A. Krieg, M. R. Rebugliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12**:7035–7056.
16. **Neckameyer, W. S., M. Shibuya, M.-T. Hsu, and L. H. Wang.** 1986. Proto-oncogene c-*ros* codes for a molecule with structural features common to those of growth factor receptors and displays tissue-specific and developmentally regulated expression. Mol. Cell. Biol. **6**:1478–1486.
17. **Neckameyer, W. S., and L. H. Wang.** 1985. Nucleotide sequence of avian sarcoma virus UR2 and comparison of its transforming gene with orther members of the tyrosine protein kinase oncogene family. J. Virol. **53**:879–884.
18. **Nilsen, T., P. Maroney, R. Goodwin, F. Rottman, L. Crittenden, M. Raines, and H.-J. Kung.** 1985. c-*erbB* activation in ALV-induced erythroblastosis: novel RNA processing and promotor insertion result in expression of an amino-truncated EGF receptor. Cell **41**:719–726.
19. **Powers, S., T. Kataoka, O. Fasano, M. Goldfarb, J. Strathern, J. Broach, and M. Wigler.** 1984. Genes in *S. cerevisiae* encoding proteins with domains homologous to the mammalian *ras* proteins. Cell **36**:607–612.
20. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
21. **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 tumor. Nature (London) **305**:245–248.
22. **Sherr, C., C. Rettenmier, R. Sacca, M. Roussel, A. Look, and E. Stanley.** 1985. The c-*fms* proto-oncogene product is related to the receptor for the mononuclear phagocyte growth factor, CSF-1. Cell **41**:665–676.
23. **Shih, C., B. Shilo, M. Goldfarb, A. Dannenberg, and R. Weinberg.** 1979. Passage of phenotypes of chemically transformed cells via transfection of DNA and chromatin. Proc. Natl. Acad. Sci. USA **76**:5714–1518.
24. **Shimizu, K., M. Goldfarb, Y. Suard, M. Perucho, Y. Li, T. Kamata, J. Ferimisco, E. Stavnezer, J. Fogh, and M. Wigler.**

1983. Three human transforming genes are related to the viral *ras* oncogenes. Proc. Natl. Acad. Sci. USA **80**:2112–2116.

25. **Southern, R. J., and P. Berg.** 1982. Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. J. Mol. Appl. Genet. **1**:327–341.
26. **Takahashi, M., J. Ritz, and G. M. Cooper.** 1985. Activation of a novel human transforming gene, *ret*, by DNA rearrangement. Cell **42**:581–588.
27. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small DNA fragment transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **77**:5201–5206.
28. **Ullrich, A., J. R. Bell, E. Y. Chen, R. Herrera, L. M. Petruzelli, T. J. Dull, A. Gray, L. Coussens, Y.-C. Liao, M. Tsubokava, A. Mason, P. H. Seeberg, C. Grunfeld, O. M. Rosen, and J. Ramachandran.** 1985. Human insulin receptor and its relationship to the tyrosine kinase family of oncogenes. Nature (London) **313**:756–763.
29. **Ullrich, A., L. Coussens, J. Hayflick, T. Dull, A. Gray, A. Tam, J. Lee, Y. Yarden, T. Libermann, J. Schlesinger, J. Downward, E. Mayes, N. Whittle, M. Waterfield, and P. Seeburg.** 1984. Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. Nature (London) **309**:418–425.
30. **Van Daren, K., D. Hanahan, and Y. Gluzman.** 1984. Infection of eucaryotic cells by helper-independent recombinant adenovirus: early region I not obligatory for integration of viral DNA. J. Virol. **50**:606–614.
31. **Wigler, M., R. Sweet, G. K. Sim, B. Wold, A. Pellicer, E. Lacy, T. Maniatis, S. Silverstein, and R. Axel.** 1979. Transformation of mammalian cells with genes from procaryotes and eucaryotes. Cell **16**:777–785.
32. **Yamamoto, T., T. Nishida, N. Miyajima, S. Kawai, T. Ooi, and K. Toyoshima.** 1983. The *erbB* gene of avian erythroblastosis virus is a member of the *src* gene family. Cell **35**:71–78.
33. **Young, D., G. Waitches, C. Birchmeier, O. Fasano, and M. Wigler.** 1986. Isolation and characterization of a new cellular oncogene encoding a protein with multiple transmembrane domains. Cell **45**:711–719.
34. **Zoller, M., and M. Smith.** 1984. Oligonucleotide directed mutagenesis: a simple method using two oligonucleotide primers and a single stranded DNA template. DNA **3**:479–488.