MOLECULAR AND CELLULAR BIOLOGY, July 1987, p. 2597–2601
0270-7306/87/072597-05$02.00/0

Vol. 7, No. 7

Human HER2 (*neu*) Promoter: Evidence for Multiple Mechanisms for Transcriptional Initiation

MICHAEL TAL,[1] C. RICHTER KING,[2] MATTHIAS H. KRAUS,[2] AXEL ULLRICH,[3] JOSEPH SCHLESSINGER,[1] AND DAVID GIVOL[1]†*

Department of Chemical Immunology, The Weizmann Institute of Science, Rehovot 76100, Israel[1]; Laboratory of Cellular and Molecular Biology, National Cancer Institute, Bethesda, Maryland 20892[2]; and Department of Developmental Biology, Genentech, Inc., South San Francisco, California 94080[3]

Received 9 February 1987/Accepted 7 April 1987

We localized the 5′ region of the human gene HER2 in a cloned fragment of genomic DNA. This clone contained exons 1 to 4 of HER2, spanning the coding sequence for the first 191 amino acids. The promoter region of HER2 was identified upstream to exon 1 by nuclease S1 mapping and by a functional assay in which the promoter region drives the expression of a chloramphenicol acetyltransferase gene. The HER2 promoter is different from the promoter of the epidermal growth factor receptor gene (HER1), and the GC boxes which are typical of the promoter of the epidermal growth factor receptor gene are absent from the HER2 promoter. One major and two minor RNA start sites located at nucleotides 178, 244, and 257 upstream to the initiator ATG were identified. The first one is 21 and 70 base pairs downstream from typical TATAA and CAAT boxes, respectively. This indicates that transcription of HER2/*neu* can be regulated by a mechanism involving a TATA box, as well as by other unidentified regulatory elements.

The particularly close relationship between the human epidermal growth factor (EGF) receptor gene (HER1) and a recently cloned gene designated HER2 or c-*erbB2* (2, 22) suggests that they have evolved from a common ancestral gene (2). Both genes have been linked to neoplastic transformation. HER1 is the cellular homolog of the oncogene v-*erbB* (21), and HER2 appears to be the normal cellular homolog of the oncogene *neu*, identified as a dominant transforming gene present in ethyl-nitrosourea-induced rat neuroblastomas (16, 19). HER2 has also been found to be amplified in some cases of human gastric, breast, or salivary adenocarcinoma (3, 9, 18, 24). Recently, the amplification of HER2 in human mammary cancer has been linked to increased malignancy (20). Overexpression of growth factor receptors may affect the cell phenotype and contribute to transformation (12). In fact, overexpression of HER2 occurs in some mammary tumor cell lines without accompanying gene amplification (10). The analysis of mechanisms controlling the regulation of expression of genes for growth factor receptors is therefore important for our understanding of the malignant state. To investigate the regulation of expression of HER2 at the molecular level, we identified and characterized the promoter region of this gene. We verified the location of the promoter region by sequence analysis of genomic and cDNA clones, nuclease S1 protection analysis, and its ability to cause transcription of a linked chloramphenicol acetyltransferase (CAT) gene. Our analysis indicated that, despite their strong similarity in the coding region, the promoters of HER2 and HER1 diverge considerably both in sequence and in the features they contain.

Isolation and sequence of 5′-specific human HER2 clone. As a probe for the 5′ region of the HER2 gene, we prepared the 5′ *Eco*RI fragments from HER2 cDNA (2), corresponding to nucleotides 1 to 1451, and used them to screen a human gene library (13, 14). Several strongly hybridizing recombinant bacteriophages were isolated, and one of them, phage λn2, contained two *Eco*RI fragments of 9 and 4 kilobases (kb) (Fig. 1). By digesting the 5′ region of HER2 cDNA with *Nco*I, we generated a 123-base-pair (bp) fragment (2) which represents the extreme 5′ probe. The 9-kb *Eco*RI fragment of λn2 hybridized to this probe, and the hybridization was confined to a 2.5-kb *Hin*III fragment (Fig. 1), whereas the 4-kb *Eco*RI fragment of λn2 hybridized strongly to the rest of the 5′ probe but not to the extreme 5′ *Eco*RI-*Nco*I 123-bp fragment of the cDNA. To determine the exon-intron relationship in λn2, as well as the structure of the HER2 promoter, we subcloned the 2.5-kb *Hin*III fragment and the 4-kb *Eco*RI fragment (Fig. 1) in plasmids pTZ18 and pTZ19. Sequence analysis of the relevant regions were performed either by the dideoxy method (17) on single-stranded phages generated from pTZ subclones or by the chemical degradation method (15) on the plasmid DNA.

The sequences of the four 5′ exons of HER2, which encompass the coding region for the N-terminal 191 amino acids, is shown in Fig. 2. The sequence of exon 1 also includes 677 bp upstream to the first ATG. Exon 1 encodes the signal peptide of the pro-HER2 polypeptide and is followed by an approximately 7-kb intron. Sequences begin-

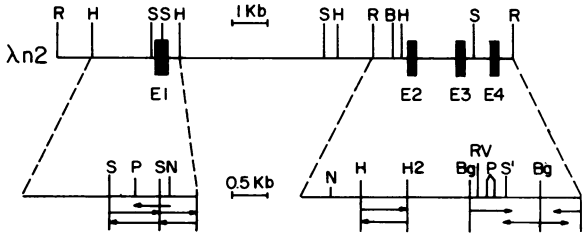

FIG. 1. Physical map of cloned phage λn2 containing the four 5′ exons of HER2. The sequencing strategy for the 2.5-kb *Hin*III and 4-kb *Eco*RI subclones is also shown. DNA sequencing was performed on PTZ plasmids and PTZ single-stranded phages. E, Exon; B, *Bam*HI; Bg, *Bgl*II; H, *Hin*III; H2, *Hin*cII; N, *Nco*I; P, *Pst*I; S, *Sma*I; R, *Eco*RI; RV, *Eco*RV.

* Corresponding author.
† Present address: Biotechnology Research Center, Meloy Laboratories, Inc., Rockville, MD 20850.

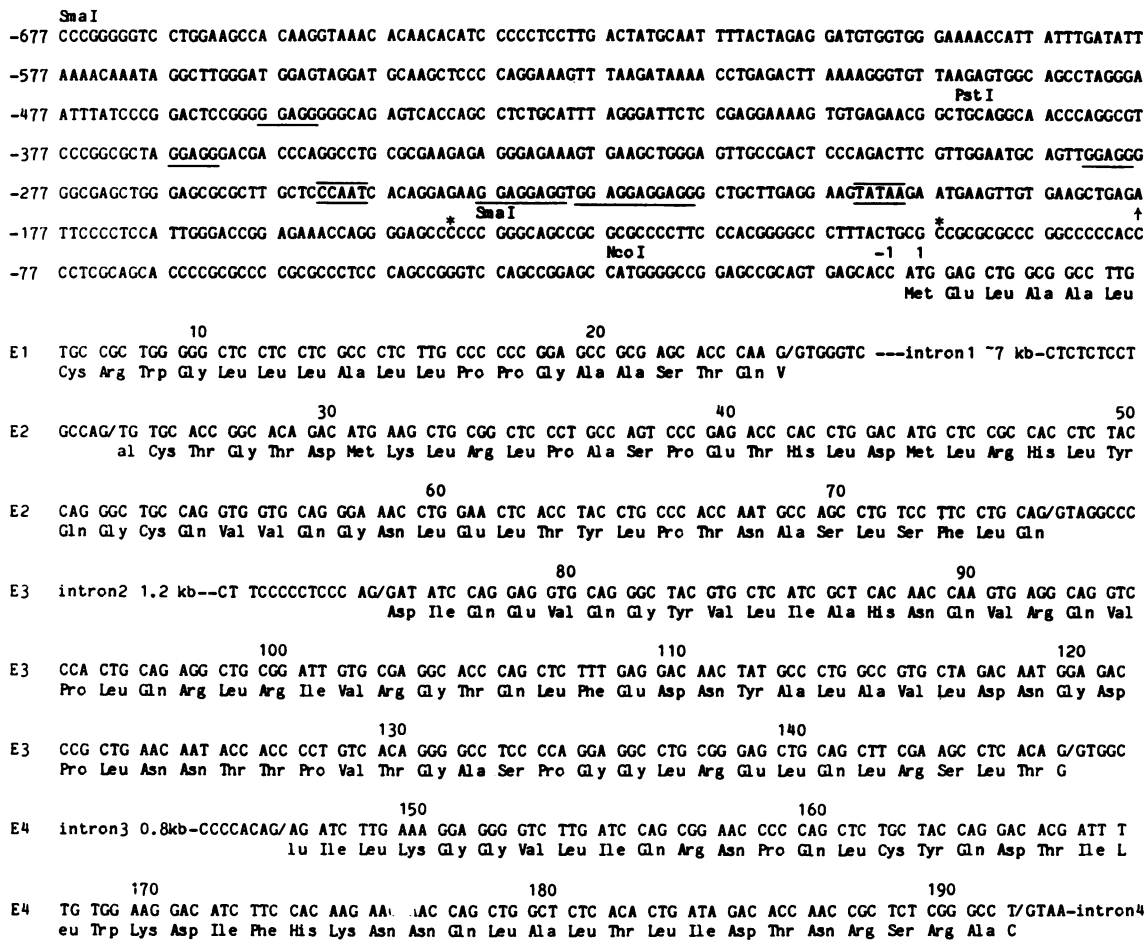

FIG. 2. DNA sequence of HER2 promoter and the four 5′ exons. The sequence shows the region upstream to exon 1, as well as exons 1 to 4 (E1 to E4) with the intron boundaries and estimated intron sizes (Fig. 1). Amino acid numbers are shown above the sequence in accordance with Coussens et al. (2). The sequence upstream to the ATG in exon 1 was given negative numbers, −1 being the first nucleotide preceding the ATG. The beginnings of the corresponding cDNA sequences described by Coussens et al. (2) and Yamamoto et al. (23) are indicated by asterisks. The CCAAT and TATA signals are boxed. The repeating sequence GGAGG is underlined. The position of the major mRNA start site as determined by nuclease S1 analysis (Fig. 3) is indicated by a vertical arrow.

ning at −97 and −141 upstream to the ATG are identical to those of the cDNA clones described by Coussens et al. (2) and Yamamoto et al. (23), respectively (Fig. 2). In addition, the sequence upstream to ATG contains the classical promoter motifs TATAA (−204 to −200) and CCAAT (−253 to −249) separated by 44 bp. This suggested that the region cloned in λn2, which is adjacent to the initiator ATG, may contain the transcriptional initiation site as well.

Nuclease S1 mapping. To locate mRNA start sites, we used nuclease S1 protection analysis (1, 22). To generate a probe for the S1 protection experiment, the 2.5-kb *Hin*III fragment containing exon 1 (Fig. 1) in pTZ19 was digested with *Nco*I (Fig. 2, nucleotide 27) and *Sph*I in the polylinker of the plasmid, blunt-ended with Klenow enzyme, and religated to generate clone pn1. This clone in its single-stranded form was used to generate the probe for nuclease S1 mapping by extension of the T7 promoter primer and digestion with *Pst*I (Fig. 2, nucleotide −396). When this labeled DNA probe was hybridized to 50 μg of total placenta RNA and subjected to digestion with nuclease S1 and electrophoresis on a 6% acrylamide-urea gel, the protected band had a length of approximately 150 nucleotides (Fig. 3), corresponding to the mRNA start site at nucleotide −178 (Fig. 2). This initiation site is 21 bp downstream from the last nucleotide of the TATAA box, suggesting that the HER2 promoter, unlike the EGF receptor promoter (8), contains functional TATA and CAAT boxes. Two additional S1 protected bands were detected (Fig. 3). These sites were assigned to nucleotides −244 and −257, and they appear to be located upstream to the TATA box.

Additional evidence for the mRNA start site was obtained from a primer extension experiment with a 25-mer synthetic oligonucleotide corresponding to nucleotides −87 to −111 (Fig. 2). The primer was annealed to placenta mRNA and extended by reverse transcriptase to yield a 90-nucleotide band (data not shown). This is consistent with the nuclease S1 major protected band and also indicates the existence of an mRNA start site at nucleotide −178 (Fig. 2).

To analyze the utilization of the different mRNA start sites in various tumors, we performed nuclease S1 analysis on RNA derived from 23 different tumor cell lines. These included three sarcomas, three glioblastomas, three melanomas, five breast carcinomas, three lung carcinomas, and four carcinomas of other origin. The results showed that in all instances but one, the three nuclease S1 protected bands described above were present in similar proportions (data

not shown). Of particular interest are the HER2 mRNA initiation sites in the mammary tumor cell line SK-BR-3. This cell line contains amplification of the HER2 locus and a 130-fold increase in mRNA. The mRNA start sites at −178, −244, and −257 agree well with those found for the placenta mRNA (Fig. 4). This indicates that despite gene amplification, the fidelity of mRNA transcript initiation is retained by the HER2 promoter. In the breast carcinoma line MCF-7, different S1 protected bands were present, suggesting mRNA start sites upstream to the RNA start sites present in all other instances (Fig. 4). This suggests that the HER2 mRNA in MCF-7 is larger than that of the placenta and other cell lines.

Functional analysis of the HER2 promoter. To further demonstrate the promoter activity of the sequence upstream to exon 1 of HER2, the plasmid pHER2CAT was constructed. This plasmid contains the 537-bp *Sma*I fragment from λn2 (nucleotides −674 to −137 in Fig. 2) subcloned into pCAT3M, which has the CAT gene but lacks enhancer or promoter elements (11). The orientation of the insert in relation to the CAT gene was determined by sequencing it from the *Xba*I site of pCAT3M. pHER2CAT was transfected into Fisher rat embryo fibroblasts in parallel with pSV2CAT and pCAT3M as controls. After 48 h the cells were harvested and lysed, and the lysate was tested for CAT activity. The *Sma*I fragment of HER2 (Fig. 2) was able to drive expression of CAT in transiently transfected rat embryo fibroblast cells (Fig. 5). This indicates a functional promoter activity in this fragment.

The sequence upstream to the initiator ATG contains

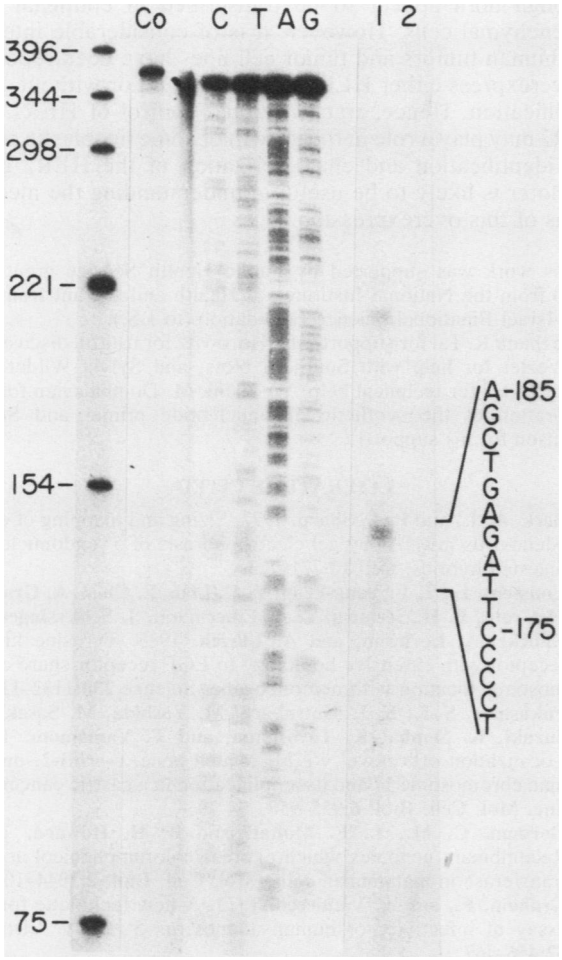

FIG. 3. Nuclease S1 analysis of human placenta RNA with the *Pst*I fragment of HER2. The homogeneously labeled *Pst*I fragment obtained from clone pn1 (see the text) was hybridized to 50 μg of placenta RNA and treated with nuclease S1 as described previously (1, 22). The S1 reaction products were electrophoresed through a 6% acrylamide-urea gel. The sequence of the corresponding *Nco*I-*Pst*I fragment labeled at the 3′ end of the *Nco*I site (Fig. 2, nucleotide −27) was run in parallel (CTAG). Because of the preparation of pn1 by filling in the *Nco*I site with Klenow enzyme, the S1 protected probe was 3 bp longer than the labeled *Nco*I fragment used for sequencing. This was confirmed by sequencing. The numbers next to the sequence refer to the DNA sequence in Fig. 2. Lanes: Co, control without nuclease S1; 1, placenta RNA; 2, tRNA. The numbers on the left are sizes in nucleotides of markers derived from *Hin*fI-digested pBR322.

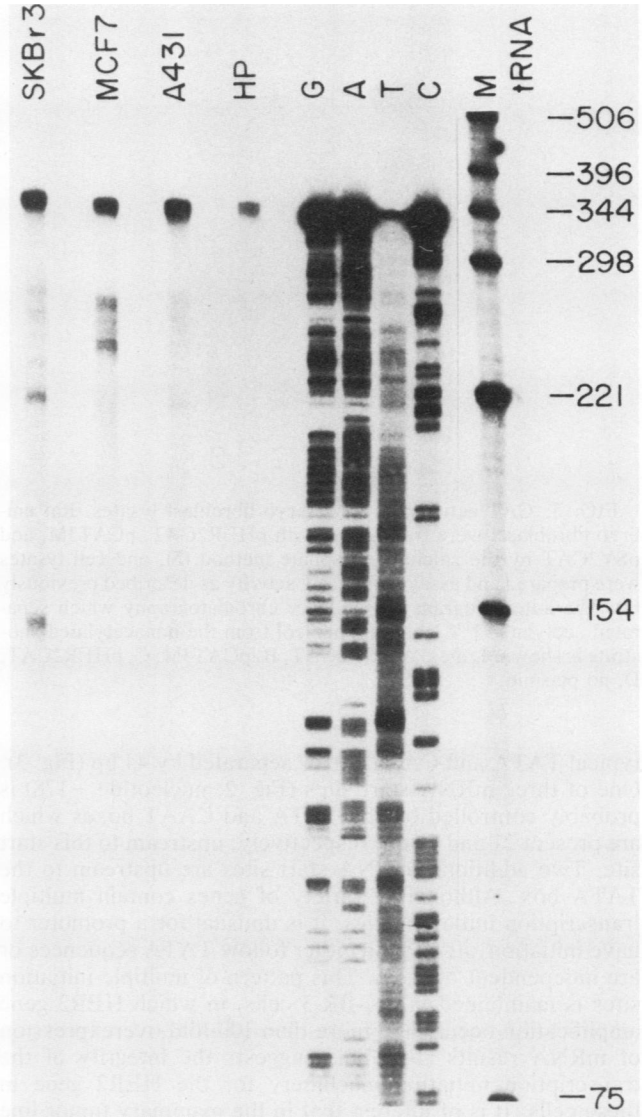

FIG. 4. Nuclease S1 analysis of RNA of tumor cell lines. Analysis was performed as described in the legend to Fig. 3 except that 12 μg of RNA was used. Cell lines: SK-BR-3 and MCF-7, mammary carcinomas; A431, squamous carcinoma; HP, human placenta. Lane M, Size markers. Other details are as described in the legend to Fig. 3.

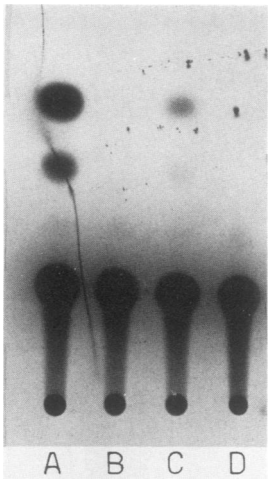

FIG. 5. CAT activity in rat embryo fibroblast lysates. Rat embryo fibroblasts were transfected with pHER2CAT, pCAT3M, and pSV2CAT by the calcium phosphate method (5), and cell lysates were prepared and assayed for CAT activity as described previously (4). The autoradiogram of thin-layer chromatography which separated acetylated [^{14}C]chloramphenicol from the nonacetylated substrate is shown. Lanes: A, pSV2CAT; B, pCAT3M; C, pHER2CAT; D, no plasmid.

typical TATA and CAAT boxes separated by 44 bp (Fig. 3). One of three mRNA start sites (Fig. 2, nucleotide −178) is probably controlled by the TATA and CAAT boxes which are present 21 and 70 bp, respectively, upstream to this start site. Two additional mRNA start sites are upstream to the TATA box. Although a variety of genes contain multiple transcription initiation sites, it is unusual for a promoter to have initiation sites which either follow TATA sequences or are independent of them. This pattern of multiple initiation sites is maintained in SK-BR-3 cells, in which HER2 gene amplification occurs and more than 100-fold overexpression of mRNA results (10). This suggests the integrity of the transcription initiation machinery for the HER2 gene in these cells. It is of interest that in the mammary tumor line MCF-7, different mRNA start sites were identified. These are located approximately 270 and 315 bp upstream to the ATG, and therefore the HER2/*neu* mRNA in this cell line should be longer. Indeed, a cDNA clone isolated from an MCF-7 library extends the cDNA sequence upstream to previously published cDNA sequences, and its 5′ end corresponds to nucleotide −238 in Fig. 3 (unpublished observations). Since this cDNA clone shows uninterrupted homology across the TATA region, it demonstrates that initiation upstream of TATA may indeed occur.

Recently, the promoters of several cellular oncogenes like c-Ha-*ras* and the EGF receptor gene were characterized and found to lack the typical TATA and CAAT boxes but to contain the GC box motif. For example, the EGF receptor gene promoter contains 6 copies of the sequence CCGCCC (8), and the *ras* promoter contains 10 copies of CCGCCC and its inverted complement GGGCGG (7). Based on this resemblance, it has been proposed that growth control genes may be characterized by promoters rich in GC boxes, which may be regulated by the Sp1 protein (6). The HER2 proto-oncogene is also a growth control gene, since its encoded protein is highly homologous to the EGF receptor. Our results show that the HER2 promoter diverges markedly both in sequence and in repeating motifs from that of the EGF receptor gene. It contains the CAAT and TATA boxes at the typical positions upstream to a major mRNA start site, and it lacks GC boxes. One similarity between the two promoters is the presence of another repeating motif, the pentanucleotide GGAGG (Fig. 2), which appears eight times (some of them overlap) in the region upstream to the mRNA start site in HER2. In the HER1 promoter, the complement of this sequence, CCTCCTCC, appears four times between 360 and 290 bp upstream to the first ATG (8). The HER2 promoter sequence is also G+C rich. The 300 bp upstream to the mRNA start site has a G+C content of 59%. This value is smaller than that reported for the promoter of the EGF receptor gene (8). The different structural features of the EGF receptor gene and the HER2 gene suggest that transcription may be regulated by different signals. Little is known about the exact tissue specificity of either gene, although both appear to be transcribed in epithelial and mesenchymal cells. However, it is of considerable interest that human tumors and tumor cell lines have been reported to overexpress either HER1 or HER2, with or without gene amplification. Hence, transcriptional control of HER1 and HER2 may play a role in the growth of some neoplastic cells. The identification and characterization of the HER2 gene promoter is likely to be useful in understanding the mechanisms of this overexpression.

This work was supported by Public Health Service grant CA 25820 from the National Institutes of Health and a grant from the U.S.-Israel Binational Science Foundation (to J.S.).

We thank R. Tal for support, Mia Horowitz for fruitful discussion, M. Wezler for help with Southern blots, and Sylvia Wilder and Laura Porter for technical help. We thank M. Dombalagian for the preparation of the synthetic oligonucleotide primer and Stuart Aaronson for his support.

LITERATURE CITED

1. **Berk, A. J., and P. A. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of 5′ endonuclease-digested hybrids. Cell **12**:721–723.
2. **Coussens, L., T. L. Yang-Feng, Y. C. Liao, E. Chen, A. Gray, J. McGrath, P. H. Seeburg, T. A. Libermann, J. Schlessinger, U. Hancke, A. Levinson, and A. Ullrich.** 1985. Tyrosine kinase receptor with extensive homology to EGF receptor share chromosomal location with neu oncogene. Science **230**:1132–1139.
3. **Fukushige, S.-I., K.-I. Matsubara, M. Yoshida, M. Sasaki, T. Suzuki, K. Semba, K. Toyoshima, and T. Yamamoto.** 1986. Localization of a novel v-*erb*B-related gene, c-*erb*B-2, on human chromosome 17 and its amplification in a gastric cancer cell line. Mol. Cell. Biol. **6**:955–958.
4. **Gorman, C. M., L. F. Moffat, and B. H. Howard.** 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. **2**:1044–1051.
5. **Graham, F., and A. Vandereb.** 1973. A new technique for the assay of infectivity of human adenovirus 5′ DNA. Virology **52**:456–467.
6. **Ishii, S., J. T. Kadonaga, R. Tjian, J. N. Brady, G. T. Merlino, and I. Pastan.** 1986. Binding of the Sp1 transcription factor by the human Harvey ras-1 proto-oncogen promoter. Science **232**: 1410–1413.
7. **Ishii, S., G. T. Merlino, and I. Pastan.** 1985. Promoter region of the human Harvey ras proto-oncogene. Similarity to the EGF receptor proto-oncogene promoter. Science **230**:1378–1380.
8. **Ishii, S., Y.-H. Xu, R. H. Stratton, B. A. Roe, G. T. Merlino, and I. Pastan.** 1985. Characterization and sequence of the promoter region of the human epidermal growth factor receptor gene. Proc. Natl. Acad. Sci. USA **82**:4920–4924.
9. **King, C. R., M. H. Kraus, and S. A. Aaronson.** 1985. Amplification of a novel v-erbB-related gene in a human mammary carcinoma. Science **229**:974–976.
10. **Kraus, M. H., N. C. Popescu, S. C. Amsbaugh, and C. R. King.**

1987. Overexpression of the EGF receptor-related proto-oncogene erbB-2 in human mammary tumor cell lines by different molecular mechanisms. EMBO J. **6**:605–610.

11. **Laimins, L. A., P. Gruss, R. Pozzatti, and G. Khoury.** 1984. Characterization of enhancer elements in the long terminal repeat of Moloney murine sarcoma virus. J. Virol. **49**:183–189.
12. **Libermann, T. A., H. R. Nusbaum, N. Razon, R. Kris, I. Lax, H. Soreq, N. Whittle, M. D. Waterfield, A. Ullrich, and J. Schlessinger.** 1985. Amplification, enhanced expression and possible rearrangement of the EGF-receptor gene in primary brain tumors of glial origin. Nature (London) **313**:144–147.
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. **Maniatis, T., R. C. Hardison, E. Lacy, J. Lauer, C. O'Connell, D. Quon, D. K. Sim, A. Efstratiadis.** 1978. The isolation of structural genes from libraries of eukaryotic DNA. Cell **15**: 687–695.
15. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74**:560–564.
16. **Padhy, L. C., C. Shih, D. Cowing, R. Finkelstein, and R. A. Weinberg.** 1982. Identification of a phosphoprotein specifically induced by the transforming DNA of rat neuroblastomas. Cell **28**:865–871.
17. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
18. **Semba, K., N. Kamata, K. Toyoshima, and T. Yamamoto.** 1985. A v-erbB-related proto-oncogene, c-erbB-2, is distinct from the c-erbB-1/epidermal growth factor-receptor gene and is amplified in human salivary gland adenocarcinoma. Proc. Natl. Acad. Sci. USA **82**:6497–6501.
19. **Shih, C., L. C. Padhy, M. Murray, and R. A. Weinberg.** 1981. Transforming genes of carcinomas and neuroblastomas introduced into mouse fibroblast. Nature (London) **290**:261–264.
20. **Slamon, D. J., G. M. Clark, S. G. Wong, W. J. Levin, A. Ullrich, and W. L. McGuire.** 1986. Human breast cancer: correlation of relapse and survival with amplification of the HER2/neu oncogene. Science **235**:177–182.
21. **Ullrich, A., L. Coussens, J. S. Hayflick, T. J. Dull, A. Gray, A. W. Tam, J. Lee, Y. Yarden, T. A. Libermann, J. Schlessinger, J. Downward, E. L. V. Mayes, N. Whittle, M. D. Waterfield, and P. H. 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.
22. **Weaver, R., and S. Weissmann.** 1979. Mapping of DNA by a modification of the Berk-Sharp procedure: the 5′ termini of 15S β-globin mRNA precursor and mature 10S β-globin mRNA have identical map coordinate. Nucleic Acids Res. **7**:1175–1193.
23. **Yamamoto, T., S. Ikawa, T. Akiyama, K. Semba, N. Nomura, N. Miyajima, T. Saito, and K. Toyoshima.** 1986. Similarity of protein encoded by human c-erbB-2 gene to epidermal growth factor receptor. Nature (London) **319**:230–234.
24. **Yokota, J., K. Toyoshima, T. Sugimura, T. Yamamoto, M. Terada, H. Battifora, and M. J. Cline.** 1986. Amplification of c-erbB-2 oncogene in human adenocarcinomas in vivo. Lancet **i**(8484):765–766.