Analysis of genomic rearrangements associated with EGFRvIII expression suggests involvement of Alu repeat elements[1]

Lori Frederick, Greg Eley, Xiao-Yang Wang, and C. David James[2]

Department of Laboratory Medicine and Pathology and Tumor Biology Program, Mayo Clinic and Foundation, Rochester, MN 55905

We have developed a polymerase chain reaction (PCR)-based strategy for the synthesis and analysis of rearranged epidermal growth factor receptor (*EGFR*) fragments associated with the vIII mutant receptor expressed in glioblastomas with *EGFR* amplification. The sequencing of aberrant tumor fragments showed that intragenic deletion rearrangements consistently involve an approximately 600-bp region in intron 7 of *EGFR* and several rearrangement sites interspersed throughout the large (>100 kb) first intron of this gene. Examination of the intron 7 breakpoint region revealed an Alu repeat element, and all intron 7 rearrangement sites were located within or downstream of this repeat sequence. Analysis of intron 1 for similar sequences resulted in the identification of 11 sites containing >80% homology with parts of the Alu element in intron 7. Reverse transcriptase-PCR and/or Western analysis of the tumors showed the presence of EGFRvIII cDNAs and/or proteins, respectively, in all cases for which a rearranged genomic fragment was generated by long-range PCR. Collectively, these data suggest that *EGFR* rearrangements, associated with the synthesis of the most common EGFR mutant, are mediated by a specific sequence element. *Neuro-Oncology 2, 159-163, 2000 (Posted to Neuro-Oncology [serial online], Doc. 00-019, May 23, 2000. URL <neuro-oncology.mc.duke.edu>)*

Received 8 March 2000, accepted 30 March 2000.

[1]Supported by NCI Grant CA55728 (C.D.J.) and a postdoctoral fellowship award from the American Brain Tumor Association (X.-Y.W.).

[2]Address correspondence and reprint requests to C. David James, Mayo Clinic, 200 First St., S.W., Hilton Bldg. Rm. 820-D, Rochester, MN 55905.

[3]Abbreviations used are as follows: EGFR, epidermal growth factor receptor; PCR, polymerase chain reaction; RT, reverse transcriptase.

Internal deletions of *EGFR*[3] genes are known to occur in most glioblastomas with *EGFR* amplification, but until recently it has been difficult to examine these alterations for the potential involvement of specific genomic sequences that could mediate the associated rearrangements and lead to the synthesis of mutant EGF receptors (Bigner et al., 1990; Libermann et al., 1985; Malden et al., 1988; Yamazaki et al., 1988). This has primarily been because of the size of the regions in which the rearrangements take place (Haley et al., 1987), combined with a lack of genomic sequence data for *EGFR*.

Difficulties with the analysis of *EGFR* intragene deletions are particularly evident for rearrangements associated with the synthesis of the type III EGFR variant (EGFRvIII), the most common EGF receptor mutant in glioblastoma (Ekstrand et al., 1992; Frederick et al., 2000). It has been suggested that these rearrangements cause the elimination of exons 2-7 from the *EGFR* gene, and result in the synthesis of transcripts lacking codons for amino acids 6-273 of the protein's extracellular domain (Humphrey et al., 1990; Sugawa et al., 1990).

Ongoing efforts associated with the Human Genome Project have recently provided information that permit a high resolution analysis of the deletions associated with EGFRvIII expression, and perhaps more importantly allow for an examination of rearrangement sites for sequences that might mediate intragene recombination events. Specifically, genomic sequence data for most of the *EGFR* gene are available, and this information, combined with the technique of long distance PCR (Foord and Rose, 1994), has been applied to the analysis of rearranged *EGFR* fragments from tumors known to express the aberrant mRNA and/or protein for EGFRvIII (Frederick et al., 2000). Our results confirm the loss of exons 2-7 in tumors expressing the aberrant transcripts and proteins, and additionally implicate a repetitive sequence element as potentially mediating the associated rearrangements.

Materials and Methods

Tumors and Nucleic Acid Extractions

All tumors used in this study were obtained from patients undergoing surgical treatment at the Mayo Clinic. DNAs from snap-frozen tissues were isolated and purified as described previously (James et al., 1988). RNAs were extracted from approximately 2 mm^2 sections of frozen tumor tissue with Trizol (Gibco BRL, Grand Island, N.Y.).

PCR and RT-PCR

Genomic PCR amplifications were performed using the GeneAmp XL PCR kit (PerkinElmer, Foster City, Calif.). Solutions for these amplifications were 50 µl and contained 1x XL buffer II supplied by the manufacturer, 200 µM dNTPs, 1 mM $Mg(OAc)_2$, 1 U rTth polymerase, 100-200 ng genomic DNA, and either 20 pmol each of intron 1 sense primer mixture and exon 8 antisense primer for the detection of exon 2-7 deletions (Table 1); or 20 pmol each of exon 13 sense primer (bases 1770-1800: GenBank Accession #X00588) and exon 16 antisense primer (bases 2104-2076) for the detection of exon 14-15 deletions. Reaction profiles consisted of a 1-min sample denaturation at 93°C, followed by 35 cycles for 30 s of denaturation at 93°C and 20 min of annealing/extension at 68°C, followed by a final 10-min synthesis step at 72°C. Reaction products were electrophoresed in 0.7% agarose gels and stained with ethidium bromide before their excision and purification using Qiagen's gel extraction kit (Valencia, Calif.).

For cDNA synthesis, 1 µg-samples of total RNA were reverse transcribed at 37°C for 1 h in 20-µl reaction volumes containing random hexamer primers, MMLV RT, and buffer supplied by the manufacturer (Gibco BRL). Following a 2-min, 95°C heat denaturation step, cDNA amplifications were performed in 50 µl-reaction volumes containing 1 µl of product from the RT reaction, 20 pmol of exon 1 sense primer (*EGFR* cDNA bases 172-193: GenBank accession #X00588) and exon 10 antisense primer (bases 1356-1335), 200 µm dNTPs (PerkinElmer), 1.25 U of *Taq* polymerase (AmpliTaq Gold, PerkinElmer), 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, and 0.001% gelatin. After sample denaturation at 95°C for 9 min, PCR amplifications were performed for 43 cycles at 95°C for 30 s, 55°C for 30 s, and 72°C for 1 min (final extension at 72°C for 10 min). RT-PCR products were fractionated in a 1.0% agarose gel, and truncated cDNAs were extracted for sequence analysis.

Sequencing

Sequences were analyzed from the products of 7-µl cycle-sequencing reactions that contained 10-20 ng of *EGFR* genomic DNA PCR product or 10-20 ng of *EGFR* cDNA synthesized by RT-PCR, 50 mM Tris-HCl (pH 8.3), 50 mM KCl, 2.5 mM $MgCl_2$, 10% dimethylsulfoxide, 400 µM ddATP, 600 µM ddTTP, 60 µM ddGTP, 200 µM ddCTP, 10 µM each dATP, dTTP, and dCTP, 20 µM 7-deaza-dGTP (Boehringer Mannheim, Indianapolis, Ind.), 0.05 µM [γ-^{32}P]ATP end-labeled sequencing primer, and 0.05 U of *Taq* polymerase. Reactions were carried out for 30 cycles at 95°C for 20 s, 58°C for 30 s, and 72°C for 1 min using a 1-min ramp time between annealing and elongation phases. Following sample denaturation, reaction products were loaded onto a 6% sequencing gel (19:1 acrylamide, 7 M urea, 0.5X tris-borate-EDTA, 15% formamide). Electrophoresis was at 75 W and room temperature for 1-3 h, after which the gels were dried and exposed to Kodak XAR film. Antisense sequencing primers for identifying intron 7 breakpoints consisted of bases 145610-145589, 144539-144520, 144941-144920, or 145279-145261 (GenBank accession #AC006977), and a sense primer for the analysis of truncated cDNAs contained bases 172-193 (accession #X00588).

Table 1. Sequence identities for sense primer mixtures used in the synthesis of rearranged EGFRvIII fragments

Primer mixture A	Primer mixture B
9361 to 9390	9361 to 9390
12361 to 12390	18640 to 18669
22449 to 22478	28310 to 28339
32450 to 32479	38341 to 38370
42350 to 42379	48301 to 48327
52490 to 52519	58340 to 58369
62491 to 62520	68042 to 68070
72211 to 72240	78391 to 78420
72751 to 72780	88371 to 88399
82384 to 82412	98297 to 98327
92371 to 92400	108369 to 108399
102365 to 102394	118464 to 118493
112442 to 112471	128341 to 128369
122311 to 122340	

Sequence numbering based upon GenBank Accession #AC006977.

Antisense primer identity is 146051 to 146022.

Results

Twenty DNAs, from glioblastomas previously examined for the presence of amplified *EGFR* as well as characterized for specific *EGFR* transcript alterations (Frederick et al., 2000), were subjected to long-range PCR using a single antisense primer located in the eighth exon of the gene, and 1 of 2 sense primer mixtures containing oligonucleotides spaced at approximately 10-kb intervals throughout intron 1, the size of which is greater than 120 kb (Table 1). Reaction products from 10 of the 15 tumor DNAs known to have *EGFR* amplification contained a fragment that was readily visible in an ethidium bromide–stained agarose gel (examples shown in Fig. 1A).

Because of the distance between the upstream and downstream primers, samples without rearrangements in this region of the gene failed to give any PCR product. To show that a negative result was not attributable to the quality of the DNA, tumors were examined with a

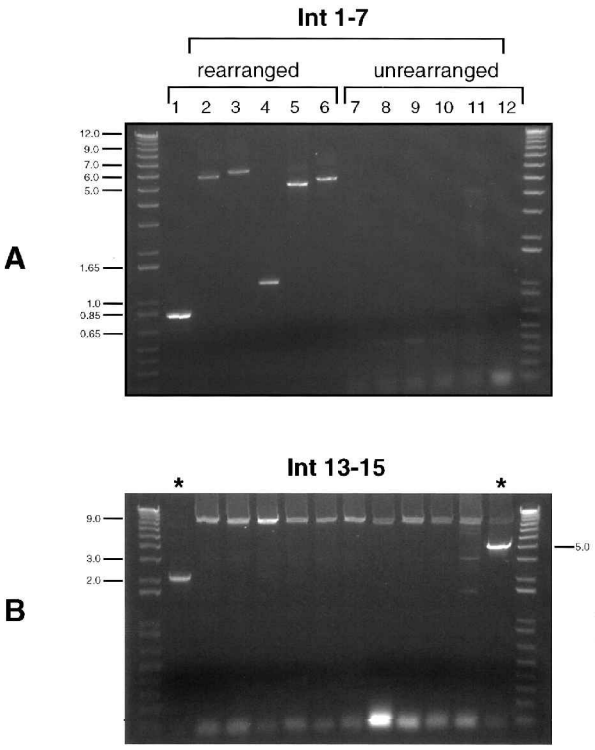

Fig. 1. Long-range PCR amplification of rearranged *EGFR* fragments. DNAs from tumors with *EGFR* amplification, which had been previously characterized for the expression of aberrant RNAs, were subjected to (A) long-range PCR synthesis using a mixture of sense primers for sequences in intron 1 and a single antisense primer in exon 8 or (B) PCR synthesis with a primer pair from exons 13 and 16 of *EGFR*. Reaction products were electrophoresed through 0.7% agarose gels that were stained with ethidium bromide and photographed. The top panel shows samples with or without rearrangement in the intron 1 to intron 7 region, whereas the lower panel shows 2 specimens (asterisks) with rearrangements in the intron 13-15 region of the gene. Faint PCR products, such as those evident in the third lane from the right of the lower panel, were occasionally observed, but were not reproducible between separate amplifications of the same DNA sample, and consequently are presumed to result from nonspecific DNA priming. Marker sizes (in kilobases) are indicated to the left and right of the images.

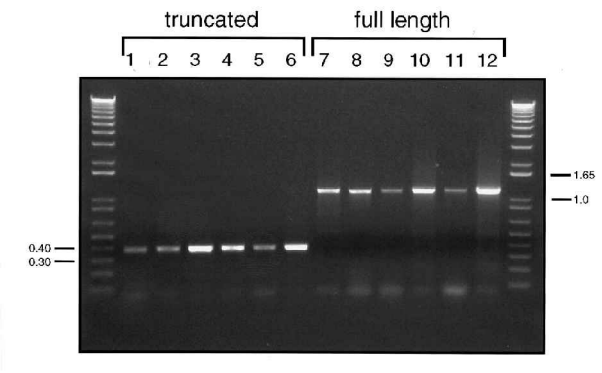

Fig. 2. RT-PCR analysis of *EGFR* transcripts for loss of coding sequence from exons 2-7. RNAs from tumors (the DNAs of which were examined for genomic rearrangements) were converted to cDNAs that were amplified using primers in exons 1 and 10 of *EGFR*. The reaction products were electrophoresed through a 1.0% agarose gel that was stained with ethidium bromide and photographed (sample lane assignments correspond with those in Fig. 1). The results show that tumors with rearrangements in the intron 1-7 region of *EGFR* express truncated transcripts of the same size, whereas those without rearrangements in this region express full length transcripts. Marker sizes (in kilobases) are indicated.

primer pair that flanks an approximately 9-kb genomic region between exons 13 and 16. All tumor DNAs yielded a PCR product from this reaction, and most of these were full length (Fig. 1B). Truncated genomic fragments were generated from 2 of the DNAs with *EGFR* amplification (lanes 1 and 12), and these were from specimens that had been previously determined to have intragene deletion rearrangements eliminating *EGFR* exons 14 and 15 (Frederick et al., 2000).

RT-PCR of tumor RNAs using primers based in exons 1 and 10 of *EGFR* resulted in the synthesis of truncated cDNAs (Fig. 2) in cases for which the corresponding DNA had shown a rearranged genomic fragment in the intron 1-7 region (Fig. 1A). The variability in size of the genomic PCR products contrasted with the uniform size of corresponding truncated cDNAs, suggesting that these 5′ end rearrangements of *EGFR* have a common effect on gene transcription, even though the rearrangement locations are different between tumors. cDNA sequencing revealed that all truncated RT-PCR products were formed as a result of exon 1 to exon 8 splicing, and in instances where there was remaining tissue, Western analysis showed a truncated protein of 140-145 kDa in tumors from which an aberrant cDNA had been synthesized (data not shown).

To determine rearrangement sites in the cases from which a genomic fragment was produced by priming with an intron 1-7 oligonucleotide mixture (Fig. 1A), PCR products were sequenced from an antisense primer located in exon 8 and extended from additional primers if necessary. The results showed that each of the 10 fragments, from tumor DNAs with amplified *EGFR*, contained a unique genomic rearrangement location, both within intron 1 as well as intron 7 (Fig. 3A and B). Inspection of the intron 7 rearrangement locations showed their clustering in an approximately 600-bp region that contained an Alu repeat element (Fig. 3B), and all of the intron 7 breakpoints occurred within or downstream of the Alu repeat. Analysis of intron 1 for similar sequences revealed 11 regions that share significant homology with the intron 7 Alu repeat (Table 2).

Single intron 1 sense primers, which had been used in the sense primer mixtures described previously (Table 1), were evaluated for their ability to amplify rearranged *EGFR* fragments from the tumors that had tested positive for rearrangements in the intron 1-7 region (Fig. 1A). For each of these cases, it was possible to generate a PCR product that was of identical size to that produced when using a primer mixture, but only if the reaction solution contained a sense primer nearest to and upstream of the

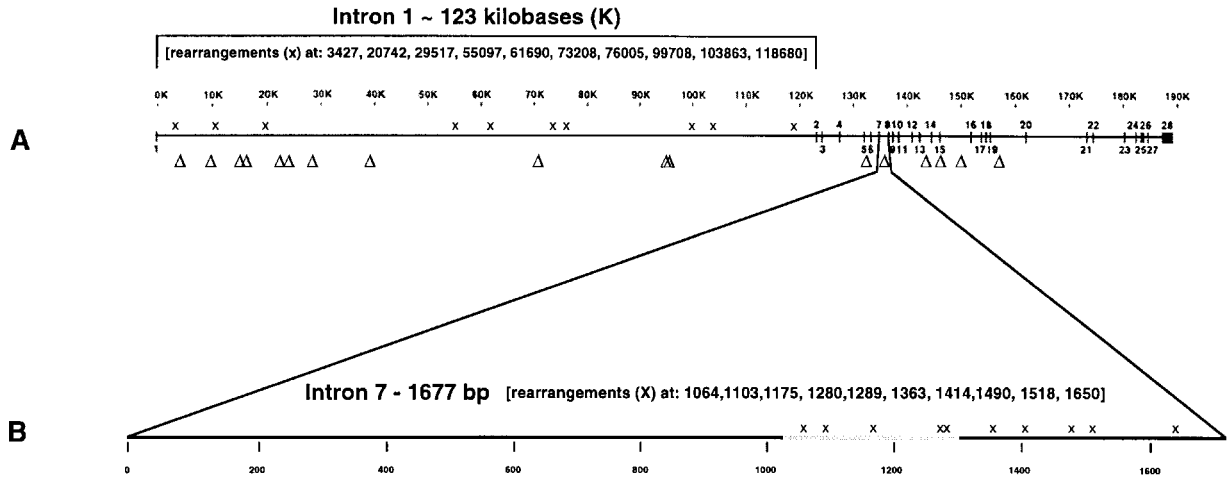

Fig. 3. Location of genomic rearrangements associated with the expression of EGFRvIII. A. Genomic organization of *EGFR* showing locations of intron 1 rearrangement sites (X), and locations of sequences having significant homology with the intron 7 Alu repeat (open triangles). B. Location of intron 7 rearrangements (X) and an Alu repeat sequence (light gray region). Numbering for rearrangement sites is based upon a value of 1 for the first (upstream) nucleotide of the indicated intron.

intron 1 breakpoint for the specific tumor DNA that was being examined (data not shown).

Discussion

Of the common oncogene activation events in human cancer, only *EGFR* is activated through intragenic deletion rearrangements, and to date such alterations have been identified only in glioblastoma. An important aspect of these mutations involves gene copy number, as the rearrangements have been reported only in glioblastomas with *EGFR* amplification. Research aimed at understanding the full range of biologic effects associated with the expression of mutant EGFRs that result from *EGFR* rearrangements in glioblastoma is being conducted by several groups of investigators (Feldkamp et al., 1999; Holland et al., 1998; Moscatello et al., 1998; Nagane et al., 1998). Achieving such an understanding presents a formidable challenge, and represents a goal with no clearly foreseeable end point.

In contrast, the progress and imminent completion of the Human Genome Project (Strohman, 1999) has already allowed for the detailed sequence analysis of several gene alterations in cancer, and sequence data that has recently been made available for *EGFR* is responsible for prompting the study presented here. Previous information concerning the genomic organization of *EGFR* was based upon restriction enzyme mapping of phage clones and RNA-DNA heteroduplex analysis to estimate the number of exons as well as intron lengths for this gene (Haley et al., 1987). Although the overall accuracy of the *EGFR* genomic map that was generated from this study is quite good, the size of intron 1 was substantially underestimated. In addition, the human gene is now known to contain 28 exons rather than the 26 initially proposed (Reiter et al., 2000), with the former value being the same as that determined for the avian *EGFR* gene (Callaghan et al., 1993).

The actual size of the first intron, 123 kb, accounts for prior difficulties in localizing breakpoints in tumors expressing EGFRvIII. The only other successful attempt to achieve this objective involved the construction of phage libraries from human glioblastoma xenografts (Yamazaki et al., 1990). Such an approach, however, is impractical for the rapid analysis of EGFRvIII genomic

Table 2. Alu repeat locations in EGFR

Intron	Sequence location (%)*
1	13638-13771 (86)
1	19156-19259 (89)
1	24381-24468 (85)
1	25686-25805 (84)
1	31412-31515 (84)
1	33483-33589 (90)
1	37755-37835 (86)
1	48381-48513 (81)
1	80093-80213 (83)
1	104483-104575 (80)
1	104993-105116 (87)
5	141871-141981 (85)
7	145429-145562 (100)
13	152935-153021 (89)
15	156148-156233 (91)
	159213-159345 (86)
19	166754-166811 (91)

*Sequence numberings are based on GenBank Accession #AC006977 and represent regions showing the highest percent homology (indicated in parentheses) with the intron 7 Alu repeat.

breakpoint sites, whereas the PCR-based approach we have used here is amenable to this purpose.

The major finding of this study is the possible involvement of a highly repetitive sequence element in *EGFR* rearrangements that causes the expression of EGFRvIII, the most common EGFR mutant in glioblastoma. In addition to the presence of Alu repeats in introns 1 and 7, it is interesting and potentially significant that our analysis of *EGFR* genomic sequence revealed Alu elements in introns 13 and 15, which flank another region of the gene that is occasionally deleted in glioblastomas with *EGFR* amplification (Frederick et al., 2000; Wong et al., 1992). In fact, deletions of this type were evident in 2 of the tumor DNAs examined here (Fig. 1B).

The literature supporting Alu-mediated recombination as a mechanism leading to gene inactivation in cancer is fairly extensive (reviewed by Deininger and Batzer, 1999). To the best of our knowledge, however, this is the first determination of Alu-mediated intragene deletion rearrangement with protein product activation. An additional unique aspect of *EGFR* rearrangements is that they seem to be observed exclusively in tumors with increased gene copy number. The mechanistic basis for this relationship has not been defined and represents a problem whose solution will require further development of model systems for studying the occurrence and evolution of amplified *EGFR* genes (Bigner et al., 1990; Goike et al., 1999; Humphrey et al., 1988).

References

Bigner, S.H., Humphrey, P.A., Wong, A.J., Vogelstein, B., Mark, J., Friedman, H.S., and Bigner, D.D. (1990) Characterization of the epidermal growth factor receptor in human glioma cell lines and xenografts. *Cancer Res.* **50**, 8017-8022.

Callaghan, T., Antczak, M., Flickinger, T., Raines, M., Myers, M., and Kung, H.J. (1993) A complete description of the EGF-receptor exon structure: implication in oncogenic activation and domain evolution. *Oncogene* **8**, 2939-2948.

Deininger, P.L., and Batzer, M.A. (1999) Alu repeats and human disease. *Mol. Genet. Metab.* **67**,183-193.

Ekstrand, A.J., Sugawa, N., James, C.D., and Collins, V.P. (1992) Amplified and rearranged epidermal growth factor receptor genes in human glioblastomas reveal deletions of sequences encoding portions of the N- and/or C-terminal tails. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 4309-4313.

Feldkamp, M.M., Lau, N., Rak, J., Kerbel, R.S., and Guha, A. (1999) Normoxic and hypoxic regulation of vascular endothelial growth factor (VEGF) by astrocytoma cells is mediated by Ras. *Int. J. Cancer* **81**, 118-124.

Foord, O.S., and Rose, E.A. (1994) Long-distance PCR. *PCR Methods Appl.* **3**, S149-S161.

Frederick, L., Wang, X.-Y., Eley, G., and James, C.D. (2000) Diversity and frequency of epidermal growth factor receptor mutations in human glioblastomas. *Cancer Res.* **60**, 1383-1387.

Goike, H.M., Asplund, A.C., Pettersson, E.H., Liu L., Sanoudou, D., and Collins, V.P. (1999) Acquired rearrangement of an amplified epidermal growth factor receptor (EGFR) gene in a human glioblastoma xenograft. *J. Neuropath. Exp. Neurol.* **58**, 697-701.

Haley, J., Whittle, N., Bennet, P., Kinchington, D., Ullrich, A., and Waterfield, M. (1987) The human EGF receptor gene: structure of the 110 kb locus and identification of sequences regulating its transcription. *Oncogene Res.* **1**, 375-396.

Holland, E.C., Hively, W.P., DePinho, R.A., and Varmus, H.E. (1998) A constitutively active epidermal growth factor receptor cooperates with disruption of G1 cell-cycle arrest pathways to induce glioma-like lesions in mice. *Genes Dev.* **12**, 3675-3685.

Humphrey, P.A., Wong, A.J., Vogelstein, B., Friedman, H.S., Werner, M.H., Bigner, D.D., and Bigner, S.H. (1988) Amplification and expression of the epidermal growth factor receptor gene in human glioma xenografts. *Cancer Res.* **48**, 2231-2238.

Humphrey, P.A., Wong, A.J., Vogelstein, B., Zalutsky, M.R., Fuller, G.N., Archer, G.E., Friedman, H.S., Kwatra, M.M., Bigner, S.H., and Bigner, D.D. (1990) Anti-synthetic peptide antibody reacting at the fusion junction of deletion-mutant epidermal growth factor receptors in human glioblastoma. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4207-4211.

James, C.D., Carlbom, E., Dumanski, J.P., Hansen, M., Nordenskjold, M., Collins V.P., and Cavenee, W.K. (1988) Clonal genomic alterations in glioma malignancy stages. *Cancer Res.* **48**, 5546-5551.

Libermann, T.A., Nusbaum, H.R., Razon, N., Kris, R., Lax, I., Soreq, H., Whittle, N., Waterfield, M.D., Ullrich, A., and Schlessinger, J. (1985) Amplification, enhanced expression and possible rearrangement of EGF receptor gene in primary human brain tumours of glial origin. *Nature* **313**, 144-147.

Malden, L.T., Novak, U., Kaye, A.H., and Burgess, A.W. (1988) Selective amplification of the cytoplasmic domain of the epidermal growth factor receptor gene in glioblastoma multiforme. *Cancer Res.* **48**, 2711-2714.

Moscatello, D.K., Holgado-Madruga, M., Emlet, D.R., Montgomery, R.B., and Wong, A.J. (1998) Constitutive activation of phosphatidylinositol 3-kinase by a naturally occurring mutant epidermal growth factor receptor. *J. Biol. Chem.* **273**, 200-206.

Nagane, M., Levitzki, A., Gazit, A., Cavenee, W.K., and Huang, H.J. (1998) Drug resistance of human glioblastoma cells conferred by a tumor-specific mutant epidermal growth factor receptor through modulation of Bcl-XL and caspase-3-like proteases. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5724-5729.

Reiter, J.L., Eley G., Cora, E.M., Baron, A.T., James, C.D., and Maihle, N.J. (2000) Complete human EGFR gene structure and a novel 3 kb alternative transcript encoding a truncated 110 kDa form of the receptor. *Proc. Am. Assoc. Cancer Res.* **41**, 5006.

Strohman, R.C. (1999) Five stages of the Human Genome Project. *Nat. Biotechnol.* **17**, 112.

Sugawa, N., Ekstrand, A.J., James, C.D., and Collins, V.P. (1990) Identical splicing of aberrant epidermal growth factor receptor transcripts from amplified rearranged genes in human glioblastomas. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8602-8606.

Wong, A.J., Ruppert, J.M., Bigner, S.H., Grzeschik, C.H., Humphrey, P.A., Bigner, D.S., and Vogelstein, B. (1992) Structural alterations of the epidermal growth factor receptor gene in human gliomas. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 2965-2969.

Yamazaki, H., Fukui, Y., Ueyama, Y., Tamaoki, N., Kawamoto, T., Taniguchi, S., and Shibuya, M. (1988) Amplification of the structurally and functionally altered epidermal growth factor receptor gene (c-erbB) in human brain tumors. *Mol. Cell. Biol.* **8**, 1816-1820.

Yamazaki, H., Ohba, Y., Tamaoki N., and Shibuya M. (1990) A deletion mutation within the ligand binding domain is responsible for activation of epidermal growth factor receptor gene in human brain tumors. *Jpn. J. Cancer Res.* **81**, 773-779.