A chromosomal region 7p11.2 transcript map: Its development and application to the study of EGFR amplicons in glioblastoma[1]

Greg D. Eley, Jill L. Reiter, Ajay Pandita, Soyeon Park, Robert B. Jenkins, Nita J. Maihle, and C. David James[2]

Department of Laboratory Medicine and Pathology (A.P., R.B.J., C.D.J.) and Tumor Biology Program (G.D.E., J.L.R., S.P., R.B.J., N.J.M., C.D.J.), Mayo Clinic, Rochester, MN 55905

Cumulative information available about the organization of amplified chromosomal regions in human tumors suggests that the amplification repeat units, or amplicons, can be of a simple or complex nature. For the former, amplified regions generally retain their native chromosomal configuration and involve a single amplification target sequence. For complex amplicons, amplified DNAs usually undergo substantial reorganization relative to the normal chromosomal regions from which they evolve, and the regions subject to amplification may contain multiple target sequences. Previous efforts to characterize the 7p11.2 epidermal growth factor receptor (*EGFR*) amplicon in glioblastoma have relied primarily on the use of markers positioned by linkage analysis and/or radiation hybrid mapping, both of which are known to have the potential for being inaccurate when attempting to order loci over relatively short (<1 Mb) chromosomal regions. Due to the limited resolution of genetic maps that have been established through the use of these approaches, we have constructed a 2-Mb bacterial and P1-derived artificial chromosome (BAC-PAC) contig for the *EGFR* region and have applied markers positioned on its associated physical map to the analysis of 7p11.2 amplifications in a series of glioblastomas. Our data indicate that *EGFR* is the sole amplification target within the mapped region, although there are several additional 7p11.2 genes that can be coamplified and overexpressed with *EGFR*. Furthermore, these results are consistent with *EGFR* amplicons retaining the same organization as the native chromosome 7p11.2 region from which they are derived.
Neuro-Oncology 4, 86–94, 2002 (Posted to Neuro-Oncology [serial online], Doc. 01-057, February 21, 2002. URL <neuro-oncology.mc.duke.edu>)

Gene amplification is common in most types of human solid tumors (Knuutila et al., 1998). Amplified DNA is manifested cytogenetically as homogeneously staining regions or double-minutes (Cowell, 1982), each of which contains multiple copies of an amplification repeat unit or amplicon (Amler and Schwab, 1989; Fakharzadeh et al., 1993). The size of the repeat unit for a specific amplicon varies between tumors, but nearly all studies attempting to address this issue report lengths within a range of a few hundred to a few thousand kilobases (Akiyama et al., 1994; Elkahloun et al., 1996; Hiemstra et al., 1994; Mostert et al., 1998; Matsumura et al., 2000; Reiter and Brodeur, 1996).

Amplicon lengths are sufficient to encompass more than 1 expressed sequence, and the common, well-characterized human chromosomal amplification regions have been shown to encompass multiple, distinct genes. However, it is generally considered that most distinct expressed sequences contained in a specific amplicon are not targeted for amplification, but rather are amplified because of their genomic proximity to a single gene

Received 11 October 2001, accepted 2 January 2002.

[1]This study was supported by National Cancer Institute Grants CA85779 (R.B.J., C.D.J.) and CA79808 (N.J.M.).

[2]Address correspondence and reprint requests to C. David James, Department of Laboratory Medicine and Pathology, Mayo Clinic, 200 First St., SW, Hilton Bldg., Rm. 820 D, Rochester, MN 55905.

[3]Abbreviations used are as follows: BAC, bacterial artificial chromosome; EGFR, epidermal growth factor receptor; EST, expressed sequence tag; ORF, open reading frame; PAC, P1-derived artificial chromosome; PCR, polymerase chain reaction; RT, reverse transcriptase; STS, sequence tagged site.

whose amplification-associated overexpression confers tumor growth advantage (Fischer et al., 1996; Reiter and Brodeur 1998). Results associated with recent investigations of 20q13 (Albertson et al., 2000; Collins et al., 1998) and 17q23 (Monni et al., 2001; Orsetti et al., 1999; Wu et al., 2001) amplifications in breast cancer may challenge the validity of this generalization.

Detailed physical and/or transcript maps have been developed for many amplification regions, including those containing *MYCN* (2p24: Amler et al., 1996; Reiter and Brodeur, 1996), *CCND1* (11q13: Cooper et al., 1998; Courseaux et al., 1997), and *CDK4* and *MDM2* (12q13-15: Elkahloun et al., 1996). For the latter of these regions, amplicons have been shown to contain as many as 12 distinct genes (Fischer et al., 1996). There are, however, amplification regions that have yet to be described at a level of detail comparable with those indicated above.

Such is the case for the 7p11.2 localized *EGFR* region, which is amplified in approximately 40% of glioblastomas (Ekstrand et al., 1991; Libermann et al., 1985; Wong et al., 1987). Amplified *EGFRs* have been shown to undergo 5′- as well as 3′-end rearrangement in glioblastoma (Wong et al., 1992; Ekstrand et al., 1992), and although the genomic consequences of these rearrangements have been mapped in detail (Eley et al., 1998; Frederick et al., 2000), little is known about the consequence of *EGFR* amplifications on flanking genomic sequences.

With these issues in mind, the primary goal of the present study was to accurately determine the organization and expressed sequence content for 7p11.2 sequences in instances of 7p11.2 amplification in glioblastoma. Previous attempts toward achieving a similar objective have interpreted 7p11.2 amplifications using maps whose expressed (EST) and nonexpressed (STS) sequence markers were positioned by linkage, radiation hybrid, or YAC analysis (Liu et al., 1998, 2000; Wang et al., 1998a). For the present investigation, we developed a high-resolution BAC-PAC physical map for the 7p11.2 region subject to amplification, and applied information from the map to an analysis of a large series of glioblastoma DNAs.

Materials and Methods

Contig Development

7p11.2 contig construction was initiated by using PCR[3] to screen the RPCI-11 BAC library (Research Genetics, Huntsville, Ala.), as well as RPCI-4/5 PACs localized to 7p11.2 by the Genome Sequencing Center at Washington University (St. Louis, MO), with primers known to amplify *EGFR* genomic DNA from total human DNA (Eley et al., 1998; Frederick et al., 2000). The *EGFR* content of clones identified in this manner was verified by Southern analysis and/or sequencing, and their 7p11.2 localization was demonstrated by fluorescence in situ hybridization. Additional 7p11.2 genomic clones were identified by PCR with primers for EST and STS markers previously shown to be coamplified with *EGFR* (Wang et al., 1998a). The clones obtained through this process were located within the Washington University human BAC library fingerprint database, and predicted overlapping clones were identified by using the FingerPrinted Contig program (Sanger Center, Cambridge, U.K.).

Fingerprint and Restriction Analysis of Genomic Clones

To confirm the FingerPrinted Contig-predicted overlap of BACs and PACs, we analyzed genomic clones by H*ind*III and *Not*I restriction mapping. *Hin*dIII digests for fingerprint analysis were electrophoresed as described by Marra et al. (1997). For the *Not*I restriction digests, bacteria were grown in 1.5 ml Luria-Bertani (LB) broth overnight, and their recombinant DNAs were harvested by alkaline lysis. Harvested DNAs were resuspended in 20 ml water, and 10-µl samples were digested with *Not*I for 2 h at 37°C. Entire digests were run on a 1.0% pulsed field gel for 17 h at 14°C with a 120° field angle, linear ramping factor, and a switch time from 5 to 15 s.

Database-Assisted Identification of 7p11.2-Expressed Sequences

The Human GeneMap ’99 maintained by the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/genemap99/) was accessed to identify ESTs that had been assigned between the DNA markers D7S2422 and D7S494, which are indicated as flanking *EGFR*. Each of the 32 ESTs identified by this approach were analyzed by PCR and/or Southern analysis for possible location within the BAC/PAC contig.

cDNA Selection for Identification of 7p11.2-Expressed Sequences

Equal quantities of mRNAs from fetal brain and placenta (Clontech, Palo Alto, Calif.) were mixed and converted to cDNA (cDNA Synthesis System; Roche Molecular Biochemicals, Indianapolis, Ind.). Linkers were incubated (2 h at 4°C) with 3 mg cDNA in a 10-µl reaction containing 2 mg phosphorylated OLIGO1 and OLIGO2 linkers (Integrated DNA Technologies, Coralville, Iowa), 3 units New England Biolabs T4 DNA ligase, and 3 ml of 10× New England Biolabs ligase buffer. The cDNA with ligated linkers was then denatured with 2 mg of C_ot-1 DNA as well as 500 ng of ribosomal RNA and was allowed to partially reanneal for 4 h at 65°C with gentle shaking in an Eppendorf Thermomixer 5437 (Brinkmann, Westbury, NY).

One hundred nanograms each of BAC and PAC forming an approximate 1.3-Mb core region of the larger contig (Fig. 1B) was biotinylated using the Gibco BRL Bionick Kit (Life Technologies, Inc., Grand Island, N.Y.). The biotinylated DNAs were precipitated with ethanol, resuspended in a volume of 5 ml, and denatured for 5 min at 95°C. The denatured, biotinylated BAC DNAs were immediately added to 5 ml of 2× hybridization buffer (1.5 M NaCl, 40 mM sodium phosphate [pH 7.2], 10 mM EDTA [pH 8.0], 10× Denhardt solution, 1.2% sodium dodecyl sulfate) and 20 ml of partially renatured cDNA, and the resulting mixture was incubated for 54 h at 65°C with shaking in the thermomixer. The biotinylated BAC clones with bound cDNAs were mixed with

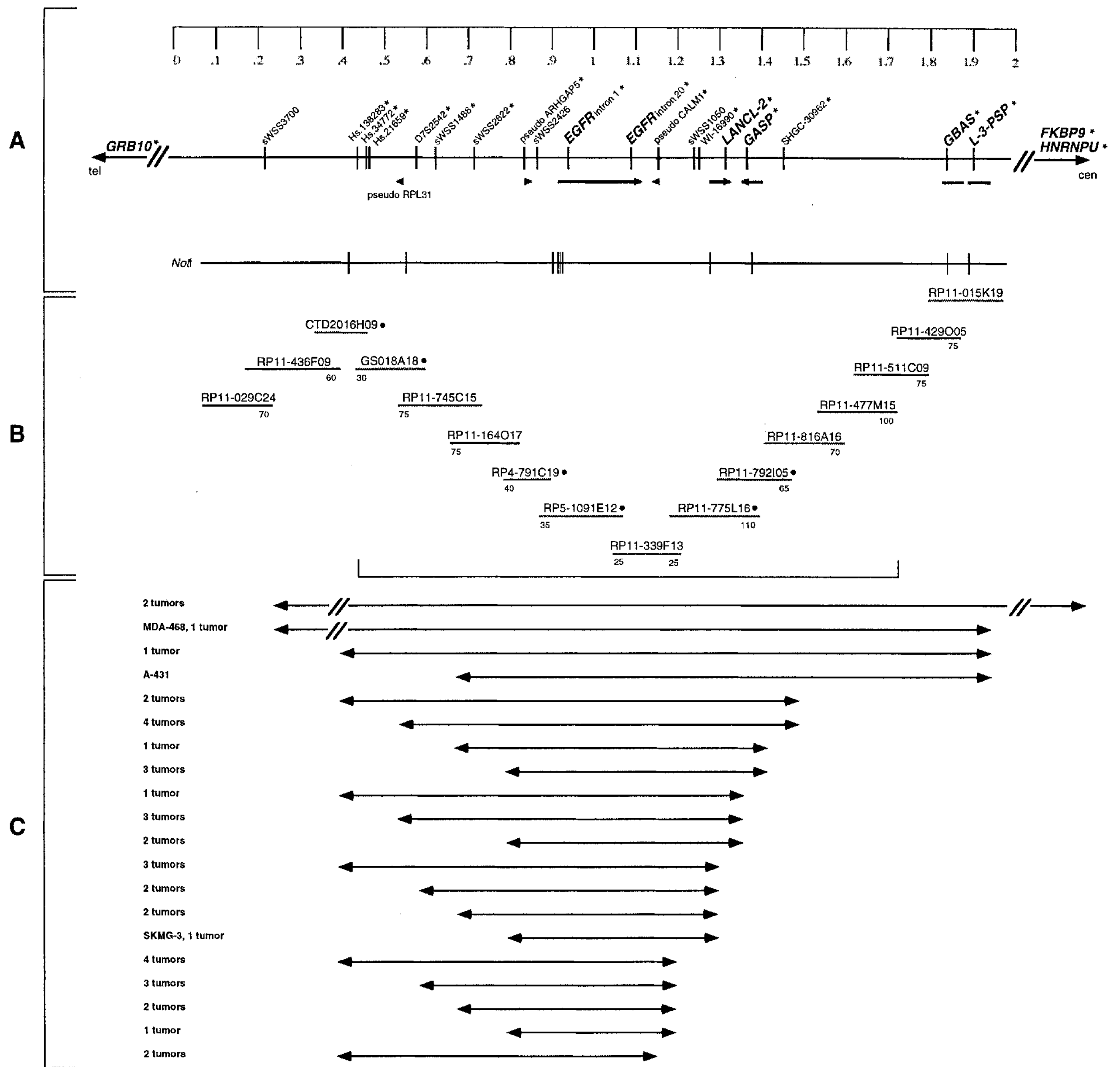

Fig. 1. *EGFR*-anchored 7p11.2 contig and 7p11.2 amplification patterns in glioblastoma. A. Genomic map showing locations of markers used in developing the contig and/or used in the Southern (*) analysis of tumor DNAs. Functional genes are indicated in bold italics, and the portion of the contig spanned by each gene is indicated by a solid line (arrowheads at the ends of solid lines indicate the direction of the associated ORF for 3 of the genes). Isolated arrowheads indicate locations of identified pseudogenes (RF directionality indicated by arrowhead orientation). A length/distance ruler is directly above the genomic map of the contig region (outer boundaries indicated by diagonal lines), and the locations of associated *Not*I restriction sites are indicated directly below. B. BACs and PACs constituting the minimal tile for the contig. Closed circles after library identification numbers indicate clones that have been sequenced. Numbers below each clone indicate extent of overlap (kb) with neighboring clones as determined by fingerprint analysis. Bracketed portion of the contig shows the clones used for cDNA hybrid selection. C. 7p11.2 sequence content determined for 40 tumors and 3 cell lines with 7p11.2 amplification; arrowheads terminate between adjacent markers showing amplicon boundaries. Diagonal lines represent outer boundaries for the region that was examined by Southern analysis. Number of samples showing each type of amplicon pattern is indicated at the left.

1 mg of streptavidin-coated iron beads (Dynabead M280, Dynal, Lake Success, N.Y.) and incubated for 15 min at room temperature. A magnetic particle concentrator was used to separate the beads from the solution, and these were subsequently washed 2 times for 15 min at room temperature in 1× saline–sodium citrate/0.1% sodium dodecyl sulfate, followed by 3 washes for 15 min at 65°C in 0.1× saline–sodium citrate/0.1% sodium dodecyl sulfate. After the final wash, bound cDNAs were eluted by rinsing with 50 ml of 150 mM NaOH for 10 min. The cDNAs were then purified (PCR Purification Kit, Qiagen, Valencia, Calif.) and amplified by PCR using

the Expand HI-FI PCR amplification kit (Boehringer-Mannheim, Berkeley, Calif.) with OLIGO1 and 2 primers. The PCR products were again purified and subjected to a second round of hybrid selection to increase the fractional yield of cDNAs associated with 7p11.2 genes. A library of enriched, selected cDNAs was then generated by cloning the final PCR products into the *Eco*RI site of the pZERO-2 vector (Invitrogen, Calif.). The entire cDNA selection procedure was performed 4 times, and from each cDNA library produced, there were 200 clones selected for fluorescence sequencing of inserts.

Tumor Specimens, Cell Lines, and Nucleic Acid Extraction

All tumors in this study were obtained from patients undergoing surgical treatment at the Mayo Clinic. Three cell lines with *EGFR* amplification were also examined: MDA-468 and A431 cells were purchased from American Type Culture Collection (Rockville, MD), and SKMG-3 cells were obtained from Dr. Christopher Thomas, Mayo Clinic, Jacksonville, FL (Thomas et al., 2001). DNAs were isolated from snap-frozen tumor tissues or cell culture pellets and purified as described previously (James et al., 1988). RNAs were extracted from ~2 mm–square sections of frozen tumor tissue with Trizol (Life Technologies, Inc.).

Southern and Northern Analysis

Southern analysis of tumor DNAs and large insert genomic clones was performed as previously described (Eley et al., 1998). Briefly, *Hin*dIII- or *Not*I-treated DNAs were electrophoresed through agarose gels, transferred to nylon-reinforced nitrocellulose membranes (Nytran, Schleicher and Schuell, Keene, N.H.), and fixed to the membrane by baking in a vacuum oven at 80°C for 2 h. Membranes were hybridized overnight with probes labeled by random priming and were washed the next day before being exposed to X-ray film. Ambion Human Multiple Tissue Northern Blot II filters (Ambion, Austin, Tex.) were used to assess normal tissue expression of candidate expressed sequences. Conditions used for filter hybridization and rinsing have been previously described (Reiter et al., 2001).

RT-PCR Analysis of Gene Expression

Expression of *LANCL2* and *GASP* in glioblastomas was assessed using multiplexed, competitive RT-PCR in which *HPRT* (hypoxanthine phosphoribosyltransferase) served as the reference transcript. Reaction mixtures consisted of 0.1 mg tumor RNA, 3 units RT, and 0.15 units AmpliTaq Gold (Perkin-Elmer, Foster City, Calif.) and were performed in solutions containing 200 mM of each dNTP, 1.5 mM $MgCl_2$, 50 mM KCl, and 10 mM Tris-HCl (pH 8.3) by using a reaction profile of 94°C for 20 s, 55°C for 30 s, and 72°C for 60 s. Reactions were for 30 cycles. Primer sequences used for RT-PCR were as follows: forward primer 5′-TGACCAAACCTACCTGCT CC-3′ with reverse primer 5′-TACAGATAACCTGCCC GTCC-3′ for *LANCL2* (cDNA product = 252 bp), forward primer 5′-GAGCCAGCCTTCAATGTGTC-3′ with reverse primer 5′-AAGGCCCTCTCCTGTCTCTC-3′ for *GASP* (cDNA product = 283 bp), and forward primer 5′-TGTGATGAAGGAGATGGGAGGC-3′ with reverse primer 5′-TTTTATGTCCCCTGTTGACTGC-3′ for *HPRT* (cDNA product = 190 bp). Upstream and downstream primers for each target gene were known to be separated by noncoding intronic sequence, as indicated by the sequences of BAC clones RP11-775L16 and RP11-792I05. Multiplexed PCR products were electrophoresed on a 1.5% agarose gel, and images of ethidium bromide–stained gels were evaluated for target:reference cDNA signal intensity ratios, as determined using Molecular Analyst software (Bio-Rad, Hercules, Calif.). The intensity ratio values were sorted into 2 groups (tumors with or without amplification of the corresponding gene) and analyzed using the Mann-Whitney *U* test.

Generation of Full-Length cDNAs for Coamplified Genes

Expressed sequences associated with Unigenes Hs.134342 and Hs.4750 that showed amplification, amplification-associated overexpression, and normal tissue expression were expanded to full-length cDNAs by electronic walking based on sequences in dbEST and by 5′ and 3′ RACE (SMART RACE cDNA Amplification Kit, Clontech, Palo Alto, Calif.) as previously described (Reiter et al., 2001; Wang et al., 1998b). Resulting full-length sequences for *LANCL2* and *GASP* have been deposited in Genbank (Accession Nos. AF353942 and AF395824, respectively).

Results

The development of a contig for analyzing 7p11.2 amplicons was initiated by using PCR to screen an arrayed BAC library and selected PAC clones with primers specific for the synthesis of *EGFR* genomic fragments. Positive clones were further examined by PCR for the extent of their *EGFR* coding sequence content, and this analysis revealed that the entire *EGFR* coding sequence was contained within the combined genomic sequences of PAC RP5-1091E12 and BAC RP11-339F13. The genomic sequence from these clones showed that the *EGFR* ORF spans 193 kb (Accession Nos. AC006977 and AF288738) and lies in a telomere-to-centromere orientation (Reiter et al., 2001).

To build the contig outward from the *EGFR*-containing clones, primers for STS and EST markers previously determined to be coamplified with *EGFR* (Wang et al., 1998a), were used to screen the same BAC library. This approach identified several BACs, although none of these overlapped with the 2 *EGFR* clones. To fill in the gaps between *EGFR*-containing and *EGFR*-flanking clones, additional, overlapping BACs were identified by FingerPrinted Contig analysis of the Washington University fingerprint database and by again screening the BAC library using primers, the sequences of which were derived from end-sequencing of the 7p11.2 clones that had already been identified. The repeated application of these approaches led to the assembly of an ~2-Mb contig,

for which the least-tiling path consists of 16 BACs and 2 PACs (Fig. 1A and 1B).

To identify functional genes in the 2-Mb contig, we used the following approaches. First, a 40-cR (centiRay) interval (GB4 map), defined by the *EGFR*-flanking markers D7S2422 and D7S494, was examined for its EST content and determined to contain 32 nonredundant expressed sequences (GeneMap '99: http://www.ncbi.nlm.nih.gov/genemap99/). A probe for each of these was synthesized and used for the Southern analysis of tumor DNAs as discussed below.

To ensure that EST databases included all expressed sequences contained within the 7p11.2 contig region, we used cDNA hybrid selection (Parimoo et al., 1991) as a second and laboratory-based approach for identifying functional genes. The selection procedure was carried out 4 times with BAC and PAC clones constituting a 1.3-Mb core of the 2-Mb contig (Fig. 1B). Two hundred cDNAs were sequenced from each of the 4 selections (800 cDNAs total). After discarding clones containing repetitive elements, sequences for the remaining clones were determined to contain expressed sequence for *EGFR* (40.5%), Unigene Hs.4750 (10.7%), Unigene Hs.134342 (8.9%), *ARHGAP5* (Rho GTPase-activating protein 5: 4.8%), or *CALM1* (Calmodulin 1: 1.7%). In addition to the trapped cDNAs for the 3 named genes and the 2 unnamed Unigenes, the genomic clones used as bait for the cDNA selection also trapped sequences from *EGFR* intron 1, which spans more than 120 kb, as well as from *EGFR* intron 20.

ESTs that had been identified through database analysis, as well as cDNAs identified through hybrid selection, were used as Southern probes to examine a panel of 118 glioblastomas, 40 of which had been previously determined to have amplified *EGFR*. This procedure revealed that probes for 14 putative expressed sequences, in addition to probes for trapped *EGFR* intron sequences, were amplified in multiple tumors (Table 1). All amplifications were observed among the 40 tumors previously determined to have amplified *EGFR*: There were no EST amplifications independent of *EGFR*.

A diagrammatic summary of the Southern data for the 40 tumors and 3 cell lines with *EGFR* amplification (SK-MG-3, MDA-468, and A-431) shows a significant variation of 7p11.2 sequence content between tumor amplicons (Fig. 1C), with one extreme represented by tumors having amplicons that appear to minimally encompass a 1.5-Mb region, and the other extreme represented by tumors with amplicons containing little more than the *EGFR* gene. The cumulative results from Southern analysis indicated that the frequency of

Table 1. Locations and amplification frequencies of 7p11.2 markers

Locus	Distance from *EGFR*	Amplification frequency[a] (%)	Contig clone(s)
GRB10	> 1 Mb	7.5	Outside of contig
Hs.138283[b]	370 kb	40	CTD-2016H9
			RP11-436F9
Hs.34772[b]	350 kb	40	CTD-2016H9
Hs.21659[b]	350 kb	40	GS-018A18
			CTD-2016H9
D7S2542[c]	235 kb	57.5	GS-018A18
			RP11-745C15
sWSS1488[c]	175 kb	70	RP11-164O17
			RP11-745C15
sWSS2822[c]	130 kb	82.5	RP11-164O17
Pseudo-*ARHGAP5*	70 kb	100	RP4-791C19
EGFR intron sequences	—	100	RP5-1091E12
			RP11-339F13
Pseudo-*CALM1*	50 kb	95	RP11-339F13
WI-16990[b]	170 kb	70	RP11-775L16
			RP11-792I05
LANCL2	260 kb	50	RP11-775L16
			RP11-792I05
GASP	340 kb	35	RP11-792I05
SHGC-30962[b]	420 kb	25	RP11-792I05
GBAS	700 kb	10	RP11-015K19
L-3-PSP	700 kb	10	RP11-015K19
FKBP9	> 1 Mb	5	Outside of contig
HNRNPU	> 1 Mb	5	Outside of contig

[a]Based on results from 40 glioblastomas with *EGFR* amplification. None of these markers was amplified in any of 78 tumors without amplified *EGFR*.

[b]Denotes unnamed Unigene or EST.

[c]Denotes STS probe for which there is no known expressed sequence association.

marker amplification is inversely proportional to its distance from the amplification target sequence, *EGFR*. In all instances, the results for individual tumors suggested that the 7p11.2 DNA contained within an amplicon is contiguous, or uninterrupted relative to its normal chromosomal configuration. The only indication of reorganization of 7p11.2 sequences was made evident in tumor DNAs with intragenic rearrangements of *EGFR* that have been previously described (Eley et al., 1998; Frederick et al., 2000).

Probes representing each of the 14 putative expressed sequences outside of the *EGFR* coding region, as well as additional probes for trapped *EGFR* intron sequences, were used for the Northern analysis of normal tissue RNAs and for cell line RNAs established from tumors with *EGFR* amplification. Each of the *EGFR* intron probes as well as 5 of the 12 probes for sequences outside of *EGFR* (Hs.138283, Hs.34772, Hs.21659, WI-16990, and SHGC-30962) failed to show expression by this analysis: Similar results have been noted by others who have examined catalogued EST sequences for evidence of expression (Ewing and Green, 2000; Hillier et al., 1996). In addition, Northern results associated with the 3 cell lines having 7p11.2 amplification, combined with sequence data for genomic clones near the *EGFR* core region (Fig. 1), indicated that the *ARHGAP5* and *CALM1* cDNAs were trapped by pseudogenes flanking *EGFR* telomerically and centromerically, respectively (functional *ARHGAP5* and *CALM1* genes reside on 14q). An additional pseudogene for *RPL31* (ribosomal protein L31) was localized to the contig region through sequence analysis of BAC GS018A18.

Seven of the expressed sequence probes were amplified in tumor DNAs, expressed in normal tissue RNAs, and overexpressed in cell line RNAs in instances where their corresponding gene was amplified. Five of the 7 probes were associated with known genes: *GRB10*, *GBAS*, *L-3-PSP*, *FKBP9*, and *HNRNPU*. For the 2 probes for which corresponding sequences were associated with unnamed Unigene entries, Hs.134342 and Hs.4750, the Northern results showed distinct transcript sizes for each (1.8 and 3.8 kb for Hs.134342, and 3.6 and 4.2 kb for Hs.4750: Fig. 2), suggesting that each was associated with a unique gene. Furthermore, the probes for these Unigenes showed distinct normal tissue expression patterns relative to each other, as well as to *EGFR*: Hs.134342 was most abundant in testis, Hs.4750 was most abundant in thymus and ovary, and *EGFR* was most abundant in placenta (Fig. 2).

Electronic walking, performed by aligning ESTs associated with each Unigene, was used to generate 1.5- and 1.2-kb cDNAs for the Hs.134342 and Hs.4750 transcripts, respectively. SMART RACE was then employed to extend the 1.5-kb sequence for Hs.134342 such that a longest ORF of 1353 nucleotides was identified and to extend the 1.2-kb transcript of Hs.4750 in order to reveal its longest ORF of 507 nucleotides. The gene with which Hs.134342 is associated has been named Lantibiotic Synthetase C–like-2 (*LANCL2*, Accession No. AF353942) because of its homology with the bacterial Lantibiotic Synthetase C family of genes, and it represents the second such gene cloned from humans (Bauer et al., 2000). Extensive homology and pattern searches of the 507-bp ORF generated for Hs.4750 have yet to reveal any significant similarities with known genes. However, the cDNA encodes a protein that is generally hydrophilic and contains a leader peptide, suggesting that it is secreted. Due to the frequency of its amplification in glioblastoma, as well as the putative leader peptide and hydrophilic nature of its encoded protein, we have named the gene for Hs.4750 *GASP* (Glioblastoma Amplified Secreted Protein, Accession No. AF395824).

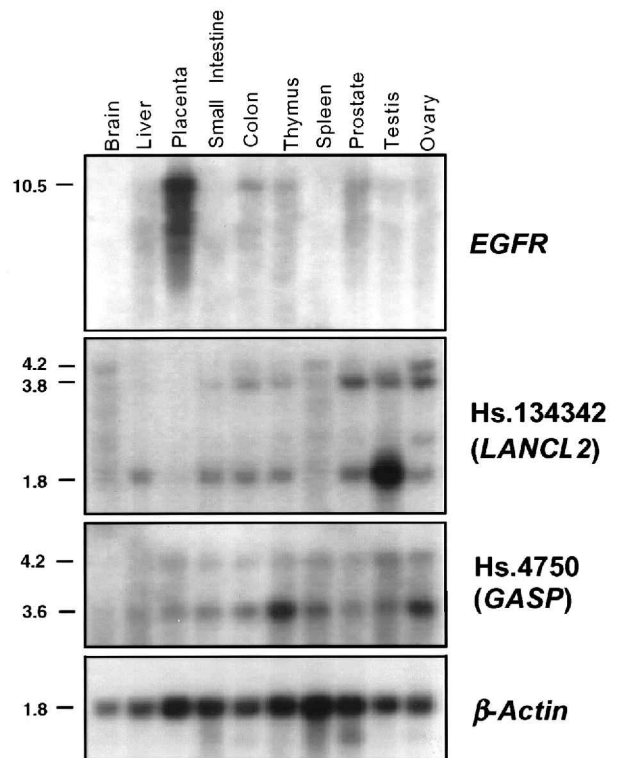

Fig. 2. Normal tissue expression patterns for *EGFR* and 2 novel genes identified in association with this study, as indicated by Northern analysis. Probes used for *LANCL2* and *GASP* were full-length cDNAs. The sizes of transcripts (kb) revealed by each probe are indicated to the left of their corresponding image.

To determine whether *LANCL2* and *GASP* show transcript overexpression in association with gene amplification, competitive RT-PCR analyses were performed using a subset of glioblastoma RNAs for which the amplification status of each gene was known in the corresponding tumor (Fig. 3). The results from these analyses showed a significant association between amplification and increased transcript level for both *LANCL2* (Fig. 3) and *GASP*: P = 0.024 and 0.016, respectively. Gene dosage is, therefore, a major determinant of *LANCL2* and *GASP* expression in glioblastoma, although the variable levels of *LANCL2* (Fig. 3) and *GASP* transcript that were indicated among tumors without amplification suggest the existence of additional factors that influence the expression of these genes.

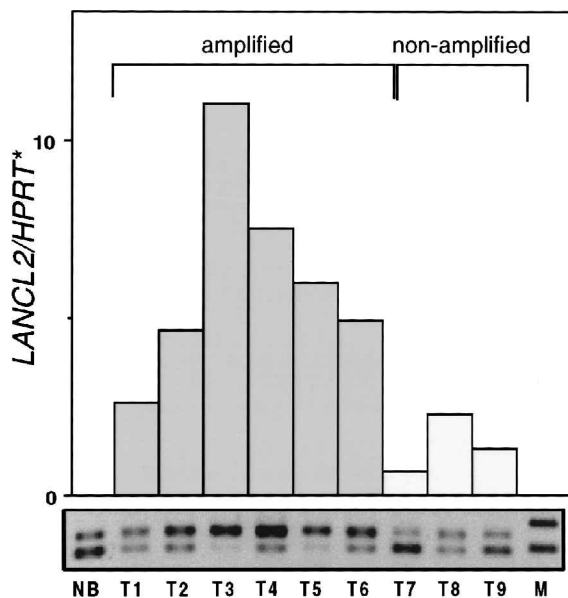

Fig. 3. Comparison of competitive RT-PCR results for *LANCL2* expression in glioblastomas with (T1-T6) or without (T7-T9) *LANCL2* amplification. RT-PCR reaction products were resolved in a 1.5% agarose gel, and a negative image of the ethidium bromide–stained cDNAs (upper band, *LANCL2*; lower band, *HPRT*) were examined for intensity using Molecular Analyst software (Bio-Rad). The results for the normalized tumor *LANCL2*:*HPRT* intensity ratios are indicated by the heights of the histogram bars, the unit definition for which is based on the signal ratio determined for normal brain (NB). Histogram bars are shown directly above the corresponding reaction products from which their values (heights) were determined. The association between elevated *LANCL2* expression and *LANCL2* amplification was determined to be significant ($P = 0.024$).

Discussion

The primary objective of this study was to develop a high resolution physical and transcript map of the genomic region surrounding *EGFR* to more precisely characterize the content of amplicons associated with chromosomal region 7p11.2. Our primary interest in developing such a map is related to the results of recent investigations that suggest the existence of multiple independent amplification targets in at least some chromosomal regions (Albertson et al., 2000; Orsetti et al., 1999). In the event that *EGFR* proved to be the sole amplification target in the 7p11.2 region, the application of a physical-transcript map to the analysis of *EGFR* amplicons was still expected to be informative about the identities of bystander genes that can be coamplified with *EGFR* and the frequencies of their coamplification.

The contig that was developed to address these issues spans approximately 2 Mb of genomic sequence. Application of the physical map derived from the contig, in combination with the Southern analysis of tumor DNAs, sequence analysis of contig clones, and the Northern and RT-PCR analysis of normal as well as tumor RNAs, indicates there are 5 functional genes and 3 pseudogenes within the contig region (Fig. 1); 3 additional genes, *GRB10*, *HNRNPU*, and *FKBP9*, lie outside of the contig region, but are nonetheless occasionally coamplified with *EGFR*.

L-3-PSP and *GBAS*, as well as the 2 novel expressed sequences identified in conjunction with this project, *LANCL2* and *GASP*, are all proximal (centromeric) to *EGFR*. Of these, *LANCL2* and *GASP* are nearest to *EGFR* and are within 250 kb from its centromeric, 3′ end, whereas *GBAS* and *L-3-PSP* are approximately 700 kb centromeric of *EGFR*. Due to their genomic proximity, *LANCL2* and *GASP*, as well as *GBAS* and *L-3-PSP*, are contained on single BAC clones: RP11-775L16 and RP11-792I05 for the former pair, and RP11-015K19 for the latter.

Unexpectedly, our results suggest that there are no functional genes in a 900-kb region that is telomeric of *EGFR*. This finding is especially surprising given the localization of 3 independent Unigenes, and a *Not*I restriction site, within a region 450 to 480 kb from the 5′ end of *EGFR* (Fig. 1A). However, neither our multiple tissue Northern analysis, performed with portions of the Unigene sequences in question, nor the cDNA hybrid selection procedure yielded results that support the assignment of functional genes to this region.

The results of Southern analysis of 7p11.2 sequences in 118 tumors and 3 cell lines support the conclusion that *EGFR* is the only amplification target of the 7p11.2 region. The lack of *EGFR*-independent amplification for any of the contig-localized probes, as well as for probes associated with the 3 genes outside of the contig region, is consistent with such a conclusion. We had previously established this point based on the interpretation of *EGFR*-7p11.2 amplifications with a map, the markers of which were positioned by linkage and radiation hybrid analysis (Wang et al., 1998a). Although the change from radiation-hybrid to physical map has not affected our conclusion regarding the singular nature of the 7p11.2 amplification target, our current interpretation of *EGFR*-7p11.2 amplicon structure, using a 2-Mb physical map with correctly ordered markers, shows that 7p11.2 amplicons are continuous and not structurally complex as we and others have previously suggested (Liu et al., 1998, 2000; Wang et al., 1998a). The simple structure of 7p11.2 amplicons appears analogous to that for the *MYCN*-2p24 amplification region in neuroblastoma (Noguchi et al., 1996; Reiter and Brodeur, 1998) and contrasts with structurally complex amplicons such as those arising from 20q13 (Albertson et al., 2000; Collins et al., 1998) and 17q23 (Monni et al., 2001; Orsetti et al., 1997; Wu et al., 2001) in breast cancer.

As shown for *MYCN* amplicons (Manohar et al., 1995; Noguchi et al., 1996), *EGFR* amplicons frequently contain bystander genes, and the 7p11.2 bystanders can be coamplified and overexpressed with *EGFR*. Approximately one half of the tumors and cell lines that were examined for 7p11.2 amplicon content showed amplification of multiple 7p11.2 genes. However, only 2 of these genes were coamplified at significant frequencies: *LANCL2* (50%) and *GASP* (35%). *LANCL2* amplification was observed in 17% of all glioblastomas examined here, making it the second most commonly amplified gene in these tumors.

Amplification of *EGFR* invariably results in increased expression of its corresponding mRNA and protein (Wong et al., 1987), and the results presented here sug-

gest this to be the case for *GASP* and *LANCL2* as well. In instances where *GASP* and/or *LANCL2* were coamplified with *EGFR*, competitive RT-PCR analysis showed overexpression of their corresponding transcripts. Whether there are biologic consequences associated with the coamplification and overexpression of these bystander genes is an issue that has yet to be addressed and is one of considerable interest because of the prevalence of bystander gene amplification in cancer. Based on the data shown here, however, it seems likely that *LANCL2* and *GASP* amplification minimally result in corresponding protein overexpression, and knowledge of the properties of these proteins could provide useful information for the development of innovative therapies.

References

Akiyama, K., Kanda, N., Yamada, M., Tadokoro, K., Matsunaga, T., and Nishi, Y. (1994) Megabase-scale analysis of the origin of N-myc amplicons in human neuroblastomas. *Nucleic Acids Res.* **22**, 187-193.

Albertson, D.G., Ylstra, B., Segraves, R., Collins, C., Dairkee, S.H., Kowbel, D., Kuo, W.L., Gray, J.W., and Pinkel, D. (2000) Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. *Nat. Genet.* **25**, 144-146.

Amler, L.C., and Schwab, M. (1989) Amplified *MYCN* in human neuroblastoma cells is often arranged as clustered tandem repeats of differently recombined DNA. *Mol. Cell. Biol.* **9**, 4903-4913.

Amler, L.C., Schurmann, J., and Schwab, M. (1996) The DDX1 gene maps within 400 kbp 5′ to *MYCN* and is frequently coamplified in human neuroblastoma. *Genes Chromosomes Cancer* **15**, 134-137.

Bauer, H., Mayer, H., Marchler-Bauer, A., Salzer, U., and Prohaska, R. (2000) Characterization of p40/GPR69A as a peripheral membrane protein related to the lantibiotic synthetase component C. *Biochem. Biophys. Res. Commun.* **275**, 69-74.

Collins, C., Rommens, J.M., Kowbel, D., Godfrey, T., Tanner, M., Hwang, S.I., Polikoff, D., Nonet, G., Cochran, J., Myambo, K., Jay, K.E., Froula, J., Cloutier, T., Kuo, W.L., Yaswen, P., Dairkee, S., Giovanola, J., Hutchinson, G.B., Isola, J., Kallioniemi, O.P., Palazzolo, M., Martin, C., Ericsson, C., Pinkel, D., and Gray, J.W. (1998) Positional cloning of ZNF217 and NABC1: Genes amplified at 20q13.2 and overexpressed in breast carcinoma. *Proc. Natl. Acad. Sci. USA* **95**, 8703-8708.

Cooper, P.R., Nowak, N.J., Higgins, M.J., Church, D.M., and Shows, T.B. (1998) Transcript mapping of the human chromosome 11q12-q13.1 gene-rich region identifies several newly described conserved genes. *Genomics* **49**, 419-429.

Courseaux, A., Szepetowski, P., Fernandes, M., Serizet, C., Kawaguchi, Y., Grosgeorge, J., Perucca-Lostanlen, D., Shows, T.B., Todd, J.A., Nowak, N.J., and Gaudray, P. (1997) Framework YAC contig anchored into a 3.2-Mb high-resolution physical map in proximal 11q13. *Genomics* **40**, 13-23.

Cowell, J.K. (1982) Double minutes and homogeneously staining regions: gene amplification in mammalian cells. *Annu. Rev. Genet.* **16**, 21-59.

Ekstrand, A.J., James, C.D., Cavenee, W.K., Seliger, B., Pettersson, R.F., and Collins, V.P. (1991) Genes for epidermal growth factor receptor, transforming growth factor α, and epidermal growth factor and their expression in human gliomas in vivo. *Cancer Res.* **51**, 2164-2172.

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.

Eley, G., Frederick, L., Wang, X.Y., Smith, D.I., and James, C.D. (1998) 3′ end structure and rearrangements of *EGFR* in glioblastomas. *Genes Chromosomes Cancer* **23**, 248-254.

Elkahloun, A.G., Bittner, M., Hoskins, K., Gemmill, R., and Meltzer, P.S. (1996) Molecular cytogenetic characterization and physical mapping of 12q13-15 amplification in human cancers. *Genes Chromosomes Cancer* **17**, 205-214.

Ewing, B., and Green, P. (2000) Analysis of expressed sequence tags indicates 35,000 human genes. *Nat. Genet.* **25**, 232-234.

Fakharzadeh, S.S., Rosenblum-Vos, L., Murphy, M., Hoffman, E.K., and George, D.L. (1993) Structure and organization of amplified DNA on double minutes containing the *MDM2* oncogene. *Genomics* **15**, 283-290.

Fischer, U., Meltzer, P., and Meese, E. (1996) Twelve amplified and expressed genes localized in a single domain in glioma. *Hum. Genet.* **98**, 625-628.

Frederick, L., Eley, G., Wang, X.-Y., and James, C.D. (2000) Analysis of genomic rearrangements associated with EGFRvIII expression suggests involvement of Alu repeat elements. *Neuro-Oncology* [serial online], Doc. 00-019, May 23, 2000. URL http://neuro-oncology.mc.duke.edu. *Neuro-Oncology* **2**, 159-163.

Hiemstra, J.L., Schneider, S.S., and Brodeur, G.M. (1994) High-resolution mapping of the N-myc amplicon core domain in neuroblastomas. *Prog. Clin. Biol. Res.* **385**, 51-57.

Hillier, L.D., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chissoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., and Marra, M. (1996) Generation and analysis of 280,000 human expressed sequence tags. *Genome Res.* **6**, 807-828.

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

Knuutila, S., Bjorkqvist, A.M., Autio, K., Tarkkanen, M., Wolf, M., Monni, O., Szymanska, J., Larramendy, M.L., Tapper, J., Pere, H., El-Rifai, W., Hemmer, S., Wasenius, V.M., Vidgren, V., and Zhu, Y. (1998) DNA copy number amplifications in human neoplasms: Review of comparative genomic hybridization studies. *Am. J. Pathol.* **152**, 1107-1123.

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

Liu, L., Ichimura, K., Pettersson, E.H., and Collins, V.P. (1998) Chromosome 7 rearrangements in glioblastomas: Loci adjacent to EGFR are independently amplified. *J. Neuropathol. Exp. Neurol.* **57**, 1138-1145.

Liu, L., Ichimura, K., Pettersson, E.H., Goike, H.M., and Collins, V.P. (2000) The complexity of the 7p12 amplicon in human astrocytic gliomas: Detailed mapping of 246 tumors. *J. Neuropathol. Exp. Neurol.* **59**, 1087-1093.

Manohar, C.F., Salwen, H.R., Brodeur, G.M., and Cohn, S.L. (1995) Coamplification and concomitant high levels of expression of a DEAD box gene with *MYCN* in human neuroblastoma. *Genes Chromosomes Cancer* **14**, 196-203.

Marra, M.A., Kucaba, T.A., Dietrich, N.L., Green, E.D., Brownstein, B., Wilson, R.K., McDonald, K.M., Hillier, L.W., McPherson, J.D., and Waterston, R.H. (1997) High throughput fingerprint analysis of large-insert clones. *Genome Res.* **7**, 1072-1084.

Matsumura, K., Iritani, A., Enomoto, S., Torikata, C., Matsuyama, S., Kurita, A., Kurahashi, H., and Tsuchida, N. (2000) Defining a common region of DNA amplification at 22q11.2-12 in head and neck squamous cell carcinomas by quantitative FISH analysis. *Genes Chromosomes Cancer* **29**, 207-212.

Monni, O., Barlund, M., Mousses, S., Kononen, J., Sauter, G., Heiskanen, M., Paavola, P., Avela, K., Chen, Y., Bittner, M.L., and Kallioniemi, A. (2001) Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5711-5716.

Mostert, M.C., Verkerk, A.J., van de Pol, M., Heighway, J., Marynen, P., Rosenberg, C., van Kessel, A.G., van Echten, J., de Jong, B., Oosterhuis,

J.W., and Looijenga, L.H. (1998) Identification of the critical region of 12p over-representation in testicular germ cell tumors of adolescents and adults. *Oncogene* **16**, 2617-2627.

Noguchi, T., Akiyama, K., Yokoyama, M., Kanda, N., Matsunaga, T., and Nishi, Y. (1996) Amplification of a DEAD box gene (*DDX1*) with the *MYCN* gene in neuroblastomas as a result of cosegregation of sequences flanking the *MYCN* locus. *Genes Chromosomes Cancer* **15**, 129-133.

Orsetti, B., Courjal, F., Cuny, M., Rodriguez, C., and Theillet, C. (1999) 17q21-q25 aberrations in breast cancer: Combined allelotyping and CGH analysis reveals 5 regions of allelic imbalance among which two correspond to DNA amplification. *Oncogene* **18**, 6262-6270.

Parimoo, S., Patanjali, S.R., Shukla, H., Chaplin, D.D., and Weissman, S.M. (1991) cDNA selection: Efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 9623-9627.

Reiter, J.L., and Brodeur, G.M. (1996) High-resolution mapping of a 130-kb core region of the *MYCN* amplicon in neuroblastomas. *Genomics* **32**, 97-103.

Reiter, J.L., and Brodeur, G.M. (1998) *MYCN* is the only highly expressed gene from the core amplified domain in human neuroblastomas. *Genes Chromosomes Cancer* **23**, 134-140.

Reiter, J.L., Threadgill, D.W., Eley, G.D., Strunk, K.E., Danielsen, A.J., Sinclair, C.S., Pearsall, R.S., Green, P.J., Yee, D., Lampland, A.L., Balasubramaniam, S., Crossley, T.D., Magnuson, T.R., James, C.D., and Maihle, N.J. (2001) Comparative genomic sequence analysis and isolation of human and mouse alternative *EGFR* transcripts encoding truncated receptor isoforms. *Genomics* **71**, 1-20.

Thomas, C., Ely, G., James, C.D., Jenkins, R., Kastan, M., Jedlicka, A., Burger, P., and Wharen, R. (2001) Glioblastoma-related gene mutations and over-expression of functional epidermal growth factor receptors in SKMG-3 glioma cells. *Acta Neuropathol.* **101**, 605-615.

Wang, X.Y., Smith, D.I., Frederick, L., and James, C.D. (1998a) Analysis of EGF receptor amplicons reveals amplification of multiple expressed sequences. *Oncogene* **16**, 191-195.

Wang, X.Y., Smith, D.I., Liu, W.G., and James, C.D. (1998b) *GBAS*, a novel gene encoding a protein with tyrosine phosphorylation sites and a transmembrane domain, is coamplified with *EGFR*. *Genomics* **49**, 448-451.

Wong, A.J., Bigner, S.H., Bigner, D.D., Kinzler, K.W., Hamilton, S.R., and Vogelstein, B. (1987) Increased expression of the epidermal growth factor receptor gene in malignant gliomas is invariably associated with gene amplification. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6899-6903.

Wong, A.J., Ruppert, J.M., Bigner, S.H., Grzeschik, C.H., Humphrey, P.A., Bigner, D.D., 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.

Wu, G., Sinclair, C., Hinson, S., Ingle, J.N., Roche, P.C., and Couch, F.J. (2001) Structural analysis of the 17q22-23 amplicon identifies several independent targets of amplification in breast cancer cell lines and tumors. *Cancer Res.* **61**, 4951-4955.