



Published in final edited form as:

Nature. 2008 October 23; 455(7216): 1061–1068. doi:10.1038/nature07385.

Comprehensive genomic characterization defines human glioblastoma genes and core pathways

The Cancer Genome Atlas (TCGA) Research Network

Abstract

Human cancer cells typically harbor multiple chromosomal aberrations, nucleotide substitutions and epigenetic modifications that drive malignant transformation. The Cancer Genome Atlas (TCGA) pilot project aims to assess the value of large-scale multidimensional analysis of these molecular characteristics in human cancer and to provide the data rapidly to the research community. Here, we report the interim integrative analysis of DNA copy number, gene expression and DNA methylation aberrations in 206 glioblastomas (GBM), the most common type of adult brain cancer, and nucleotide sequence aberrations in 91 of the 206 GBMs. This analysis provides new insights into the roles of *ERBB2*, *NF1* and *TP53*, uncovers frequent mutations of the PI3 kinase regulatory subunit gene *PIK3R1*, and provides a network view of the pathways altered in the development of GBM. Furthermore, integration of mutation, DNA methylation and clinical treatment data reveals a link between *MGMT* promoter methylation and a hypermutator phenotype consequent to mismatch repair deficiency in treated glioblastomas, an observation with potential clinical implications. Together, these findings establish the feasibility and power of TCGA, demonstrating that it can rapidly expand knowledge of the molecular basis of cancer.

Cancer is a disease of genome alterations: DNA sequence changes, copy number aberrations, chromosomal rearrangements, and modification in DNA methylation together drive the development and progression of human malignancies. With the complete sequencing of the human genome and continuing improvement of high-throughput genomic technologies, it is now feasible to contemplate comprehensive surveys of human cancer genomes. The Cancer Genome Atlas (TCGA) aims to catalogue and discover major cancer-causing genome alterations in large cohorts of human tumors through integrated multi-dimensional analyses.

Correspondence should be addressed to L.C. (Email: E-mail: lynda_chin@dfci.harvard.edu), or M.M. (Email: E-mail: matthew_meyerson@dfci.harvard.edu).

Author Contributions: The TCGA research network contributed collectively to this study. Biospecimens were provided by the Tissue Source Sites and processed by the Biospecimen Core Resource. Data generation and analyses were performed by the Genome Sequencing Centers and Cancer Genome Characterization Centers. All data were released through the Data Coordinating Center. Project activities were coordinated by the NCI and NHGRI Project Teams. We also acknowledge the following TCGA investigators who contributed substantively to the writing of this manuscript. **Leaders:** Lynda Chin(1,2,3) & Matthew Meyerson(4,5). **Neuropathology:** Ken Aldape(10), Darell Bigner(6), Tom Mikkelsen(8) & Scott VandenBerg(9). **Databases:** Ari Kahn(13). **Biospecimen analysis:** Robert Penny(17), Martin Ferguson(18) & Daniela Gerhard(16). **Copy number:** Gad Getz(4), Cameron Brennan(24), Barry S. Taylor(25,26), Wendy Winckler(4,5), Peter Park(21,22,23), Marc Ladanyi(27). **Gene expression:** Katherine A. Hoadley(38), Roel G.W. Verhaak(4,5), D. Neil Hayes(40) & Paul Spellman(28). **LOH:** Devin Absher(64) & Barbara A. Weir(4,5). **Sequencing:** Gad Getz(4), Li Ding(19), David Wheeler(34), Michael S. Lawrence(4), Kristian Cibulskis(52), Elaine Mardis(19), Jinghui Zhang(68), Rick Wilson(19) **TP53:** Larry Donehower(35), David A. Wheeler(34) **NF1:** Wendy Winckler(4,5), Li Ding(19), Jinghui Zhang(68) **EGFR:** Elizabeth Purdom(29), Wendy Winckler(4,5) **ERBB2:** Wendy Winckler(4,5) **PIK3R1:** Li Ding(19), John Wallis(19), Elaine Mardis(19) **DNA Methylation:** Peter W. Laird(49), James G. Herman(42), Li Ding(19), Daniel J. Weisenberger(49), Stephen B. Baylin(42) **Pathway analysis:** Nikolaus Schultz(25), Larry Donehower(35), David A. Wheeler(34), Jun Yao(2), Ruprecht Wiedemeyer(1), John Weinstein(69), Chris Sander(25)

General: Stephen B. Baylin(42), Richard A. Gibbs(34), Joe Gray(28), Raju Kucherlapati(22), Marc Ladanyi(27), Eric Lander(4,60,61), Richard M. Myers(64), Charles M. Perou(36,37), Richard K. Wilson(19) & John Weinstein(69).

Competing interest statement: The author declares no competing financial interests.

The first cancer studied by TCGA is glioblastoma (GBM), the most common primary brain tumor in adults¹. Primary GBM, which comprises more than 90% of biopsied or resected cases, arises *de novo* without antecedent history of low grade disease, whereas secondary GBM progresses from previously diagnosed low-grade gliomas¹. Patients with newly diagnosed GBM have a median survival of approximately one year with generally poor responses to all therapeutic modalities². Two decades of molecular studies have identified important genetic events in human GBMs, including (i) dysregulation of growth factor signaling via amplification and mutational activation of receptor tyrosine kinase (RTK) genes; (ii) activation of the phosphatidyl inositol 3-kinase (PI3K) pathway; and (iii) inactivation of the p53 and retinoblastoma tumor suppressor pathways¹. Recent genome-wide profiling studies have also shown remarkable genomic heterogeneity among GBM and the existence of molecular subclasses within GBM that may, when fully defined, allow stratification of treatment^{3–8}. Albeit fragmentary, such baseline knowledge of GBM genetics sets the stage to explore whether novel insights can be gained from a more systematic examination of the GBM genome.

Results

As a public resource, all TCGA data are deposited at the Data Coordinating Center (DCC) for public access (<http://cancergenome.nih.gov/>). TCGA data are classified by data type (*e.g.* clinical, mutations, gene expression) and data level to allow structured access to this resource with appropriate patient privacy protection. An overview of the data organization is provided in Methods, and a detailed description is available in the TCGA Data Primer (http://tcga-data.nci.nih.gov/docs/TCGA_Data_Primer.pdf).

Biospecimen collection

Retrospective biospecimen repositories were screened for newly diagnosed GBM based on surgical pathology reports and clinical records (Fig. S1). Samples were further selected for having matched peripheral blood as well as associated demographic, clinical and pathological data (Table S1). Corresponding frozen tissues were reviewed at the Biospecimen Core Resource (BCR) to ensure a minimum of 80% tumor nuclei and a maximum of 50% necrosis (Fig. S1). DNA and RNA extracted from qualified biospecimens were subjected to additional quality control measurements (Methods) prior to distribution to TCGA centers for analyses (Fig. S2).

After exclusion based on insufficient tumor content ($n=234$) and suboptimal nucleic acid quality or quantity ($n=147$), 206 of the 587 biospecimens screened (35%) were qualified for copy number, expression, and DNA methylation analyses. Of these, 143 cases had matched normal peripheral blood DNAs and were therefore appropriate for re-sequencing. This cohort also included 21 post-treatment GBM cases used for exploratory comparisons (Table S1). While it is possible that a small number of progressive secondary GBMs were among the remaining 185 cases of newly diagnosed glioblastomas, this cohort represents predominantly primary GBM. Indeed, when compared with published cohorts, overall survival of the newly diagnosed glioblastoma cases in TCGA is similar to that reported in the literature (Fig. S3, $p=0.2$)^{9–12}.

Genomic and transcriptional aberrations

Genomic copy number alterations (CNAs) were measured on three microarray platforms (Methods) and analyzed with multiple analytical algorithms^{13–15} (Fig. S4; Tables S2–S4). Besides the well-known alterations^{3,13,14}, we detected significantly recurrent focal alterations not previously reported in GBMs, such as homozygous deletions involving *NF1* and *PARK2* and amplifications of *AKT3* (Fig. 1a; Tables S2–S4). Search for informative but infrequent CNAs also uncovered rare focal events, such as amplifications of *FGFR2* and *IRS2* and deletion

of *PTPRD* (Table S4). Abundances of protein-coding genes and non-coding microRNA were also measured by transcript-specific and exon-specific probes on multiple platforms (Methods, and manuscript in preparation). The resulting integrated gene expression data set showed that ~76% of genes within recurrent CNAs have expression patterns that correlate with copy number (Table S2). In addition, SNP-based analyses also catalogued copy-neutral loss of heterozygosity (LOH), with the most significant region being 17p, which contains *TP53* (Methods).

Patterns of somatic nucleotide alterations in GBM

91 matched tumor-normal pairs (72 untreated and 19 treated cases) were selected from the 143 cases for detection of somatic mutations in 601 selected genes (Table S5). The resulting sequences, totaling 97 million base pairs (1.1 ± 0.1 million bases per sample), uncovered 453 validated non-silent somatic mutations (Table S6; http://tcga-data.nci.nih.gov/docs/somatic_mutations/tcga_mutations.htm). The background mutation rates differed drastically between untreated and treated GBMs, averaging 1.4 versus 5.8 somatic silent mutations per sample (98 among 72 untreated vs 111 among 19 treated, $p < 10^{-21}$), respectively. This difference was predominantly driven by seven hypermutated samples, as determined by frequencies of both silent and non-silent mutations (Fig. 1b,c). Four of the 7 hypermutated tumors were from patients previously treated with temozolomide and 3 from patients treated with CCNU alone or in combination (Table S1b). A hypermutator phenotype in GBM has been described in 3 GBM specimens with *MSH6* mutations^{16,17}, prompting us to perform a systematic analysis of the genes involved in mismatch repair (MMR). Indeed, 6 of the 7 hypermutated samples harbored mutations in at least one of the mismatch repair genes *MLH1*, *MSH2*, *MSH6*, or *PMS2*, as compared with only one sample among the 84 non-hypermutated samples ($p = 7 \times 10^{-8}$), suggesting a role of decreased DNA repair competency in these highly mutated samples derived from treated patients.

By applying a statistical analysis of mutation significance¹⁸, we identified eight genes as significantly mutated (false discovery rate (FDR) $< 10^{-3}$) (Fig. 2d, Table S6). Interestingly, 27 *TP53* mutations were detected in the 72 untreated GBMs (37.5%) and 11 mutations in the 19 treated samples (58%). All of those mutations clustered in the DNA binding domain, a well-known hotspot for p53 mutations in human cancers (Fig. S5; Table S6). Given the predominance of primary GBM among this newly diagnosed collection, that result unequivocally proves that p53 mutation is a common event in primary GBM.

NF1 is a human glioblastoma suppressor gene—Although somatic mutations in *NF1* have been reported in a small series of human GBM tumors²¹, their role remains controversial²², despite strong genetic data in mouse model systems^{19,20}. Here, 19 *NF1* somatic mutations were identified in 13 samples (14% of 91), including six nonsense mutations, four splice site mutations, five missense changes, and four frameshift indels (Fig. 2a). Five of these mutations—R1391S (23), R1513* (24), e25 -1 and e29 +1 (25), and Q1966* (26)—have been reported as germline alterations in neurofibromatosis patients, thus are likely inactivating. In addition, 30 heterozygous deletions in *NF1* were observed among the entire interim sample set of 206 cases, 6 of which also harbor point mutation (Tables S8 and S9). Some samples also exhibited loss of expression without evidence of genomic alteration (Fig. 2b). Overall, at least 47 of these 206 patient samples (23%) harbored somatic *NF1* inactivating mutations or deletions, definitively address *NF1*'s relevance to sporadic human GBM.

Prevalence of EGFR family activation—EGFR is frequently activated in primary GBMs. Variant III deletion of the extracellular domain (so-called “vIII mutant”)²⁷ has been the most commonly described event, in addition to extracellular domain point mutations and cytoplasmic domain deletions^{28,29}. Here, high resolution genomic and exon-specific

transcriptomic profiling readily detected vIII and C-terminal deletions with correspondingly altered transcripts (Fig. 2c). Among the 91 GBM cases with somatic mutation data, 22 harbored focal amplification of wild type *EGFR* with no point mutation, 16 had point mutations in addition to focal amplification, and three had *EGFR* point mutations but no amplification (Fig. S6; Table S9). Collectively, *EGFR* alterations were observed in 41 of the 91 sequenced samples.

ERBB2 mutation has previously been reported in only one GBM tumor³⁰. In the TCGA cohort, 11 somatic *ERBB2* mutations in 7 of 91 samples were validated, including 3 in the kinase domain and two involving V777A, a site of recurrent missense and in-frame insertion mutations in lung, gastric, and colon cancers³¹. The remaining eight mutations (including seven missense and one splice-site mutation) occurred in the extracellular domain of the protein, similar to somatic *EGFR* substitutions in GBM (Fig. 2d). Unlike in breast cancers, focal amplifications of *ERBB2* were not observed in GBMs.

Somatic mutations of the PI3K complex in human glioblastoma—The PI3 kinase complex is comprised of a catalytically active protein, p110 α , encoded by *PIK3CA*, and a regulatory protein, p85 α , encoded by *PIK3RI*. Frequent activating missense mutations of *PIK3CA* have been reported in multiple tumor types, including GBM^{32,33}. These mutations occur primarily in the adaptor binding domain (ABD) as well as the C2 helical and kinase domains^{34–36}. Indeed, *PIK3CA* somatic nucleotide substitutions were detected in six of the 91 sequenced samples (Table S6). Besides the 4 matching events already reported in the COSMIC database (<http://www.sanger.ac.uk/genetics/CGP/cosmic/>), two novel in-frame deletions were detected in the ABD of *PIK3CA* (“L10del” and “P17del”). Those deletions may disrupt interactions between p110 α and its regulatory subunit, p85 α ³⁷.

Unlike *PIK3CA*, *PIK3RI* has rarely been reported as mutated in cancers. Among the five reported *PIK3RI* nucleotide substitutions in cancers^{38,39}, one was in a glioblastoma³⁹. In our TCGA cohort, 9 *PIK3RI* somatic mutations were detected among the 91 sequenced GBMs. None of them was in samples with *PIK3CA* mutations. Of the nine mutations, eight lay within the intervening SH2 (or iSH2) domain and four are 3-basepair in-frame deletions (Fig. 3a and Table S6). In accord with the crystal structure of PI3 kinase, which identifies the D560 and N564 amino acid residues in p85 α as contact points with the N345 amino acid residue in the C2 domain of p110 α ³⁷, the mutations detected in GBM cluster around those three amino acid residues (Fig. 3b), including a N345K mutation in *PIK3CA* (previously reported in colon and breast cancers⁴⁰) and two novel D560 mutations in *PIK3RI* (D560Y and N564K). We also identified an 18-basepair deletion spanning residues D560 to S565 (DKRMNS) in *PIK3RI* (Fig. 3b) in addition to three other novel deletions (R574del, T576del, and W583del) in proximity to the 3 key residues. We speculate that spatial constraints due to these deletions might prevent inhibitory contact of p85 α nSH2 with the helical domain of p110 α , causing constitutive PI3K activity. Taken together, the pattern of clustering of the mutations around key residues defined by the crystal structure of PI3K strongly suggest that these novel *PIK3RI* point mutations and insertions/deletions disrupt the important C2-iSH2 interaction, relieving the inhibitory effect of p85 α on p110 α .

MGMT methylation and MMR proficiency in post-treatment GBMs

Cancer-specific DNA methylation of CpG dinucleotides located in CpG islands within the promoters of 2,305 genes were measured relative to normal brain DNA (Table S7; Methods). The promoter methylation status of *MGMT*, a DNA repair enzyme that removes alkyl groups from guanine residues⁴¹, is associated with GBM sensitivity to alkylating agents^{42,43}. Among the 91 sequenced cases, 19 samples were found to contain *MGMT* promoter methylation (including 13 of the 72 untreated and 6 of the 19 treated cases). When juxtaposed

with somatic mutation data, an intriguing relationship between the hypermutator phenotype and *MGMT* methylation status emerged in the treated samples. Specifically, *MGMT* methylation was associated with a profound shift in the nucleotide substitution spectrum of treated GBMs (Fig. 4a). Among the treated samples lacking *MGMT* methylation (n=13), 29% (29/99) of the validated somatic mutations occurred as G:C to A:T transitions in CpG dinucleotides (characteristic of spontaneous deamination of methylated cytosines), and a comparable 23% (23/99) of all mutations occurred as G:C to A:T transitions in non-CpG dinucleotides. In contrast, in the treated samples with *MGMT* methylation (n=6), 81% of all mutations (146/181) turned out to be of the G:C to A:T transition type in non-CpG dinucleotides whereas only 4% (8/181) of all mutations were G:C to A:T transition mutations within CpGs. That pattern is consistent with a failure to repair alkylated guanine residues caused by treatment. In other words, *MGMT* methylation shifted the mutation spectrum of treated samples to a preponderance of G:C to A:T transition at non-CpG sites.

Significantly, the mutational spectra in the mismatch repairs (MMR) genes themselves reflected *MGMT* methylation status and treatment consequences. All seven mutations in MMR genes found in six *MGMT* methylated hypermutated (treated) tumors occurred as G:C to A:T mutations at non-CpG sites (Fig. 4b; Table S6), while neither MMR mutations in non-methylated hypermutated tumors was of this characteristic. Hence, these data show that MMR deficiency and *MGMT* methylation together, in the context of treatment, exert a powerful influence on the overall frequency and pattern of somatic point mutations in GBM tumors, an observation of potential clinical importance.

Integrative analyses define core pathways required for GBM pathogenesis

To begin to construct an integrated view of common genetic alterations in the GBM genome, we mapped the unequivocal genetic alterations—validated somatic nucleotide substitutions, homozygous deletions and focal amplifications—onto major pathways implicated in GBM¹. That analysis identified a highly interconnected network of aberrations (Figs. S7–S8), including three major pathways: receptor tyrosine kinases (RTKs) signaling, and the p53/RB tumor suppressor pathways (Fig. 5).

By copy number data alone, 66%, 70% and 59% of the 206 samples harbored somatic alterations of the RB, TP53 and RTK pathways, respectively (Table S8). In the 91 samples for which there was also sequencing data, the frequencies of somatic alterations increased to 87%, 78% and 88%, respectively (Table S9). There was a statistical tendency toward mutual exclusivity of alterations of components within each pathway (p-values of 9.3×10^{-10} , 2.5×10^{-13} , and 0.022, respectively for the p53, RB, and RTK pathways; Tables S10), consistent with the thesis that deregulation of one component in the pathway relieves the selective pressure for additional ones. However, we observed a greater than random chance (one-tailed p = 0.0018) that a given sample harbors at least one aberrant gene from each of the three pathways (Table S10). In fact, 74% harbored aberrations in all three pathways, a pattern suggesting that deregulation of the three pathways is a core requirement for glioblastoma pathogenesis.

Besides frequent deletions and mutations of the *PTEN* lipid phosphatase tumor suppressor gene, 86% of the GBM samples harbored at least one genetic event in the core RTK/PI3K pathway (Fig. 5a). In addition to *EGFR* and *ERBB*, *PDGFRA* (13%) and *MET* (4%) showed frequent aberrations (Tables S9). 10 of the 91 sequenced samples have amplifications or point mutations in at least two of the four RTKs catalogued (*EGFR*, *ERBB2*, *PDGFRA* and *MET*) (Table S9), suggesting genomic activation can be a mechanism for co-activated RTKs⁴⁴.

Inactivation of the p53 pathway occurred in the form of *ARF* deletions (55%), amplifications of *MDM2* (11%) and *MDM4* (4%), in addition to mutations of p53 itself (Fig. 5b; Table S8). Among 91 sequenced samples (Table S9), genetic lesions in *TP53* were mutually exclusive of

those in *MDM2* or *MDM4* (odds ratios of 0.00 for both; $p = 0.02$ and 0.068 , respectively; Tables S10), but not of those in *ARF*. In fact, 10 of the 32 tumors with *TP53* mutations also deleted *ARF*, suggesting that homozygous deletion of the *CDKN2A* locus (which encodes both p16^{INK4A} and ARF) was at least in part driven by p16^{INK4A}.

Among the 77% samples harboring RB pathway aberrations (Fig. 5c), the most common event was deletion of the *CDKN2A/CDKN2B* locus on chromosome 9p21 (55% and 53%), followed by amplification of the *CDK4* locus (14%) (Fig. 1a; Table S8 and S9). Although copy number alterations in the CDK/RB pathway members can co-occur in the same tumor¹⁴, all nine samples with *RB1* nucleotide substitutions (Table S9) lacked *CDKN2A/B* deletion or other copy number alterations in the pathway, suggesting that inactivation of *RB1* by nucleotide substitution, in contrast to copy number loss, obviates the genetic pressure for activation of upstream cyclin/cyclin-dependent kinases.

Discussion

In establishing this pilot program, TCGA has developed important principles in biospecimen banking and collection (manuscript in preparation), and established the infrastructure that will serve similar efforts in the future. Although it ensured high quality data, the stringent biospecimen selection criteria may have introduced a degree of bias because small samples and samples with high levels of necrosis were excluded. Nonetheless, the clinical parameters of this cohort are similar to other published cohorts (Table S1; Fig. S3).

The integrated analyses of multi-dimensional genomic data from complementary technology platforms have proved informative. In addition to pinpointing deregulation of RB, p53 and RTK/RAS/PI3K pathways as obligatory events in most, and perhaps all, GBM tumors, the patterns of mutations may also inform future therapeutic decisions. It would be reasonable to speculate that patients with deletions or inactivating mutations in *CDKN2A* or *CDKN2C* or patients with amplifications of *CDK4/CDK6* would be candidates for treatment with CDK inhibitors, a strategy not likely to be effective in patients with *RB1* mutation. Similarly, patients with *PTEN* deletions or activating mutations in *PIK3CA* or *PIK3R* might be expected to benefit from a PI3 kinase or PDK1 inhibitor, while tumors in which the PI3 kinase pathway is altered by *AKT3* amplification might prove refractory to those modalities. The presence of genomic co-amplification reinforces the recent report of multiple phosphorylated (activated) RTKs in individual GBM specimens⁴⁴, suggesting a way to tailor anti-RTK therapeutic cocktails to specific patterns of RTK mutation. In addition, combination anti-RTK therapy might synergize with downstream inhibition of PI3K or cell cycle mediators. In contrast, GBMs with *NF1* mutations might benefit from a RAF or MEK inhibitor as part of a combination, as shown for *BRAF* mutant cancers⁴⁵.

One of the most important biomarkers for GBM is the methylation status of *MGMT*, which predicts sensitivity to temozolomide^{42,43}, an alkylating agent that is the current standard of care for GBM patients. Integrative analysis of mutation, DNA methylation and clinical (treatment) data, albeit with small sample numbers, suggests a series of inter-related events that may impact clinical response and outcome. Newly diagnosed glioblastomas with *MGMT* methylation respond well to treatment with alkylating agents, in part as a consequence of unrepaired alkylated guanine residues initiating cycles of futile mismatch repair, which can lead to cell death⁴⁶⁻⁴⁸. Therefore, treatment of *MGMT*-deficient GBMs with alkylating therapy introduces a strong selective pressure to lose mismatch repair function⁴⁹. That conclusion is consistent with our observation that the mismatch repair genes themselves are mutated with characteristic C:G → A:T transitions at non-CpG sites resulting from unrepaired alkylated guanine residues. Thus, initial methylation of *MGMT*, in conjunction with treatment, may lead to both a shift in mutation spectrum affecting mutations at mismatch repair genes

and selective pressure to lose mismatch repair function. In other words, our finding raises the possibility that patients who initially respond to the frontline therapy in use today may evolve not only treatment resistance, but also an MMR-defective hypermutator phenotype. If such a mechanism indeed underlies emergence of MMR-defective resistance, one may speculate that selective strategies targeting mismatch-repair deficiency⁵⁰ would represent a rational upfront combination that may prevent or minimize emergence of such resistance. Validation of this hypothesis will have immediate clinical impact and implication for therapeutic design. For one, it suggests that treatment mediated mutator phenotype may lead to pathway mutations that confer resistance to new targeted therapies thereby raising the concern that combined or serial treatment with alkylating agents and pathway targeted therapies may substantially increase the probability of developing resistance to such targeted drugs.

In conclusion, the power of TCGA to produce unprecedented multi-dimensional data sets employing statistically robust numbers of samples sets the stage for a new era in the discovery of new cancer interventions. The integrative analyses leading to formulation of an unanticipated hypothesis on a potential mechanism of resistance highlights precisely the value and power of such project design, demonstrating how unbiased and systematic cancer genome analyses of large sample cohorts can lead to paradigm-shifting discoveries.

Method Summary

Biospecimens were screened from retrospective banks of Tissue Source Sites under appropriate IRB approvals for newly diagnosed GBM with minimal 80% tumor cell percentage. RNA and DNA extracted from qualified specimens were distributed to TCGA centers for analysis. Whole genome-amplified genomic DNA samples from tumors and normals were sequenced by the Sanger method. Mutations were called, verified using a second genotyping platform, and systematically analyzed to identify significantly mutated genes after correcting for the background mutation rate for nucleotide type and the sequence coverage of each gene. DNA copy number analyses were performed using the Agilent 244K, Affymetrix SNP6.0, and Illumina 550K DNA copy number platforms. Sample-specific and recurrent copy number changes were identified using various algorithms (GISTIC, GTS, RAE). mRNA and miRNA expression profiles were generated using Affymetrix U133A, Affymetrix Exon 1.0 ST, custom Agilent 244K, and Agilent miRNA array platforms. mRNA expression profiles were integrated into a single estimate of relative gene expression for each gene in each sample. Methylation at CpG dinucleotides was measured using the Illumina GoldenGate assay. All data for DNA sequence alterations, copy number, mRNA expression, miRNA expression, and CpG methylation were deposited in standard common formats in the TCGA DCC at <http://cancergenome.nih.gov/dataportal/>. All archives submitted to DCC were validated to ensure a common document structure and to ensure proper use of identifying information.

The Cancer Genome Atlas (TCGA) Research Network

Tissue Source Sites

Duke University Medical School—Roger McLendon(6), Allan Friedman(7), Darrell Bigner(6), **Emory University**: Erwin G Van Meir(45,46,47), Daniel J Brat(47,48), Gena Marie Mastrogiannakis(45), Jeffrey J Olson(45,46,47) **Henry Ford Hospital**: Tom Mikkelsen(8), Norman Lehman(50), **MD Anderson Cancer Center**: Ken Aldape(10), W.K. Alfred Yung(11), Oliver Bogler(12), **University of California San Francisco**: Scott VandenBerg(9), Mitchel Berger(51), Michael Prados(51)

Genome Sequencing Centers

Baylor College of Medicine—Donna Muzny(34), Margaret Morgan(34), Steve Scherer(34), Aniko Sabo(34), Lynn Nazareth(34), Lora Lewis(34), Otis Hall(34), Yiming Zhu(34), Yanru Ren(34), Omar Alvi(34), Jiqiang Yao(34), Alicia Hawes(34), Shalini Jhangiani(34), Gerald Fowler(34), Anthony San Lucas(34), Christie Kovar(34), Andrew Cree(34), Huyen Dinh(34), Jireh Santibanez(34), Vandita Joshi(34), Manuel L. Gonzalez-Garay(34), Christopher A. Miller(34,36), Aleksandar Milosavljevic(34,36,37), Larry Donehower(35), David A. Wheeler(34), Richard A. Gibbs(34), **Broad Institute of MIT and Harvard**: Kristian Cibulskis(52), Carrie Sougnez(53), Tim Fennell(54), Scott Mahan(59), Jane Wilkinson(55), Liuda Ziaugra(56), Robert Onofrio(56), Toby Bloom(57), Rob Nicol(58), Kristin Ardlie(59), Jennifer Baldwin(55), Stacey Gabriel(56), Eric Lander(4,60,61), **Washington University in Saint Louis**: Li Ding(19), Robert S. Fulton(19), Michael D. McLellan(19), John Wallis(19), David E. Larson(19), Xiaoqi Shi(19), Rachel Abbott(19), Lucinda Fulton(19), Ken Chen(19), Daniel C. Koboldt(19), Michael C. Wendl(19), Rick Meyer(19), Yuzhu Tang(19), Ling Lin(19), John R. Osborne(19), Brian H. Dunford-Shore(19), Tracie L. Miner(19), Kim Delehaunty(19), Chris Markovic(19), Gary Swift(19), William Courtney(19), Craig Pohl(19), Scott Abbott(19), Amy Hawkins(19), Shin Leong(19), Carrie Haipek(19), Heather Schmidt(19), Maddy Wiechert(19), Tammi Vickery(19), Sacha Scott(19), David J. Dooling(19), Asif Chinwalla(19), George M. Weinstock(19), Elaine R. Mardis(19), Richard K. Wilson(19)

Cancer Genome Characterization Centers

Broad Institute/Dana-Farber Cancer Institute—Gad Getz(4), Wendy Winckler(4,5), Roel G.W. Verhaak(4,5), Michael S. Lawrence(4), Michael O’Kelly(4), Jim Robinson(4), Gabriele Alexe(4), Rameen Beroukhi(4,5), Scott Carter(4), Derek Chiang(4,5), Josh Gould(4), Supriya Gupta(4), Josh Korn(4), Craig Mermel(4,5), Jill Mesirov(4), Stefano Monti(4), Huy Nguyen(4), Melissa Parkin(4), Michael Reich(4), Nicolas Stransky(4), Barbara A. Weir(4,5), Levi Garraway(4,5), Todd Golub(4,62), Matthew Meyerson(4,5) **Harvard Medical School/Dana-Farber Cancer Institute**: Lynda Chin(1,2,3), Alexei Protopopov(2), Jianhua Zhang(2), Ilana Perna(2), Sandy Aronson(21), Narayan Sathiamoorthy(21), Georgia Ren(2), Jun Yao(2), Hyunsoo Kim(21), Sek Won Kong(23,71) Yonghong Xiao(2), Isaac S. Kohane(21,22,23), Jon Seidman(63), Peter J. Park(21,22,23), Raju Kucherlapati(21) **John Hopkins/University of Southern California**: Peter W. Laird(49), Leslie Cope(43), James G. Herman(42), Daniel J. Weisenberger(49), Fei Pan(49), David Van Den Berg(49), Leander Van Neste(44), Joo Mi Yi(42), Kornel E. Schuebel(42), Stephen B. Baylin(42) **HudsonAlpha Institute/Stanford University**: Devin M. Absher(64), Jun Z. Li(70), Audrey Southwick(32), Shannon Brady(32), Amita Aggarwal(32), Tisha Chung(32), Gavin Sherlock(32), James D. Brooks(33), Richard M. Myers(64) **Lawrence Berkeley National Laboratory**: Paul T. Spellman(28), Elizabeth Purdom(29), Lakshmi R. Jakkula(28), Anna V. Lapuk(28), Henry Marr(28), Shannon Dorton(28), Yoon Gi Choi(30), Ju Han(28), Amrita Ray(28), Victoria Wang(29), Steffen Durinck(28), Mark Robinson(31), Nicholas J. Wang(28), Karen Vranizan(30), Vivian Peng(30), Eric Van Name(30), Gerald V. Fontenay(28), John Ngai(30), John G. Conboy(28), Bahram Parvin(28), Heidi S. Feiler(28), Terence P. Speed(29,31), Joe W. Gray(28) **Memorial Sloan-Kettering Cancer Center**: Cameron Brennan(24), Nicholas D. Socci(25), Adam Olshen(65), Barry S. Taylor(25,26), Alex Lash(25), Nikolaus Schultz(25), Boris Reva(25), Yevgeniy Antipin(25), Alexey Stukalov(25), Benjamin Gross(25), Ethan Cerami(25), Wei Qing Wang(25), Li-Xuan Qin(65), Venkatraman E. Seshan(65), Liliana Villafania(66), Magali Cavatore(66), Laetitia Borsu(27), Agnes Viale(66), William Gerald(27), Chris Sander(25), Marc Ladanyi(27) **University of North Carolina, Chapel Hill**: Charles M. Perou(38,39), D. Neil Hayes(40), Michael D. Topal(39), Katherine A. Hoadley(38), Yuan Qi(40), Sai Balu(41), Yan Shi(41), George Wu(41)

Biospecimen Core Resource: Robert Penny(17), Michael Bittner(67), Troy Shelton(17), Elizabeth Lenkiewicz(17), Scott Morris(17), Debbie Beasley(17), Sheri Sanders(17)

Data Coordinating Center: Ari Kahn(13), Robert Sfeir(13), Jessica Chen(13), David Nassau(13), Larry Feng(13), Erin Hickey(13), Carl Schaefer(68), Subha Madhavan(68), Ken Buetow(68)

Project Teams

National Cancer Institute Anna Barker(16), Daniela Gerhard(16), Joseph Vockley(16), Martin Ferguson(18), Carolyn Compton(16), Jim Vaught(16), Peter Fielding(16) **National Human Genome Research Institute:** Francis Collins(15), Peter Good(15), Mark Guyer(15), Brad Ozenberger(15), Jane Peterson(15) & Elizabeth Thomson(15).

Affiliations for participants:

- 1) Department of Medical Oncology, Dana-Farber Cancer Institute; Boston, Massachusetts 02115, USA
- 2) Center for Applied Cancer Science of the Belfer Institute for Innovative Cancer Science, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA
- 3) Department of Dermatology, Harvard Medical School, Boston, Massachusetts 02115, USA
- 4) The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA
- 5) Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA
- 6) Department of Pathology, Duke University Medical Center, Durham, North Carolina 27710, USA
- 7) Department of Surgery, Duke University Medical Center, Durham, North Carolina 27710, USA
- 8) Departments of Neurological Surgery, Henry Ford Hospital, Detroit, MI 48202, USA
- 9) Department of Pathology, University of California San Francisco, San Francisco, California 94143, USA
- 10) Department of Pathology, University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030, USA
- 11) Department of Neuro-Oncology, University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030, USA
- 12) Department of Neurosurgery, University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030, USA
- 13) SRA International, Fairfax, VA 22033, USA
- 14) Center for Biomedical Informatics and Informational Technology, National Cancer Institute, Rockville, Maryland 20852, USA

- 15) National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA
- 16) National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA
- 17) International Genomics Consortium, Phoenix, AZ 85004 USA
- 18) MLF Consulting, Arlington, MA 02474 USA
- 19) The Genome Center at Washington University, Department of Genetics, Washington University School of Medicine, St. Louis, Missouri 63108, USA
- 20) The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, Maryland 21231, USA
- 21) Harvard Medical School-Partners HealthCare Center for Genetics and Genomics, Boston, MA 02115, USA
- 22) Center for Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA
- 23) Informatics Program, Children's Hospital, Boston, MA 02115 USA
- 24) Department of Neurosurgery, Memorial-Sloan Kettering Cancer Center, New York, NY 10065, USA
- 25) Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, NY 10065 USA
- 26) Department of Physiology and Biophysics, Weill Cornell Graduate School of Medical Sciences, New York, NY 10065 USA
- 27) Department of Pathology, Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065
- 28) Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA
- 29) Department of Statistics, University of California at Berkeley, Berkeley, California 95720, USA
- 30) Department of Molecular and Cellular Biology, University of California at Berkeley, Berkeley, California 95720, USA
- 31) Walter and Eliza Hall Institute, Parkville, Vic 3052, Australia.
- 32) Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA
- 33) Department of Urology, Stanford University School of Medicine, Stanford, California 94305, USA
- 34) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA
- 35) Department of Molecular Virology and Microbiology, Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA

- 36) Graduate Program in Structural and Computational Biology and Molecular Biophysics, Baylor College of Medicine, Houston, TX 77030, USA
- 37) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA
- 38) Department of Genetics, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
- 39) Department of Pathology and Laboratory Medicine, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
- 40) Department of Internal Medicine, Division of Medical Oncology, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
- 41) Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
- 42) Cancer Biology Division, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University, Baltimore, Maryland 21231, USA
- 43) Biometry and Clinical Trials Division, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University, Baltimore, Maryland 21231, USA
- 44) Department of Molecular Biotechnology, Faculty of Bioscience and Engineering, Ghent University, Ghent B-9000, Belgium.
- 45) Department of Neurosurgery, Emory University School of Medicine, Atlanta, Georgia 30322, USA
- 46) Department of Hematology and Medical Oncology, Emory University School of Medicine, Atlanta, Georgia 30322, USA
- 47) Winship Cancer Institute, Emory University School of Medicine, Atlanta, Georgia 30322, USA
- 48) Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, Georgia 30322, USA
- 49) University of Southern California Epigenome Center, University of Southern California, Los Angeles, California 90089, USA
- 50) Department of Pathology, Henry Ford Hospital, Detroit, MI 48202 USA
- 51) Department of Neurosurgery, University of California San Francisco, San Francisco, CA 94143 USA
- 52) Medical Sequencing Analysis and Informatics, The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA, 02142 USA
- 53) Cancer Genome & Medical Resequencing Projects, The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA, 02142 USA

- 54) Directed Sequencing Informatics, The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA, 02142 USA
- 55) Sequencing Platform, The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA, 02142 USA
- 56) Genetic Analysis Platform, The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA, 02142 USA
- 57) Sequencing Platform Informatics, The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA, 02142 USA
- 58) Sequencing Platform Production, The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA, 02142 USA
- 59) Biological Samples Platform, The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA, 02142 USA
- 60) Department of Biology, Institute of Massachusetts Institute of Technology, Cambridge, MA, 02142 USA
- 61) Department of Systems Biology, Harvard University, Boston, MA, 02115 USA
- 62) Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston MA 02115 USA
- 63) Department of Genetics, Harvard Medical School, Boston, MA, 02115 USA
- 64) HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806 USA
- 65) Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY 10065 USA
- 66) Genomics Core Laboratory, Memorial Sloan-Kettering Cancer Center, New York, NY 10065 USA
- 67) Computational Biology Division, Translational Genomics Research Institute, Phoenix, AZ 85004 USA
- 68) Center For Biomedical Informatics and Information Technology, National Cancer Institute, Rockville, MD 20852 USA
- 69) Department of Bioinformatics and Computational Biology, M.D. Anderson Cancer Center, Houston, TX 77030 USA
- 70) Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109 USA
- 71) Department of Cardiology, Children's Hospital, Boston, MA 02115 USA

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank the members of TCGA's External Scientific Committee and the Glioblastoma Disease Working Group (<http://cancergenome.nih.gov/components>) for helpful discussions; Dr. David N. Louis for valuable discussions;

Anika Mirick, Jessica Melone, and Cathy Collins for their crucial administrative coordination of TCGA activities; and Leslie Gaffney for graphic art. This work was supported by the following grants from the United States National Institutes of Health: U54HG003067, U54HG003079, U54HG003273, U24CA126543, U24CA126544, U24CA126546, U24CA126551, U24CA126554, U24CA126561, and U24CA126563.

References

1. Furnari FB, et al. Malignant astrocytic glioma: genetics, biology, and paths to treatment. *Genes Dev* 2007;21:2683–710. [PubMed: 17974913]
2. Mischel PS, Cloughesy TF. Targeted molecular therapy of GBM. *Brain Pathol* 2003;13:52–61. [PubMed: 12580545]
3. Mischel PS, Nelson SF, Cloughesy TF. Molecular analysis of glioblastoma: pathway profiling and its implications for patient therapy. *Cancer Biol Ther* 2003;2:242–7. [PubMed: 12878856]
4. Phillips HS, et al. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* 2006;9:157–73. [PubMed: 16530701]
5. Maher EA, et al. Marked genomic differences characterize primary and secondary glioblastoma subtypes and identify two distinct molecular and clinical secondary glioblastoma entities. *Cancer Res* 2006;66:11502–13. [PubMed: 17114236]
6. Lee J, et al. Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. *Cancer Cell* 2006;9:391–403. [PubMed: 16697959]
7. Diehn M, et al. Identification of noninvasive imaging surrogates for brain tumor gene-expression modules. *Proc Natl Acad Sci U S A* 2008;105:5213–8. [PubMed: 18362333]
8. Liang Y, et al. Gene expression profiling reveals molecularly and clinically distinct subtypes of glioblastoma multiforme. *Proc Natl Acad Sci U S A* 2005;102:5814–9. [PubMed: 15827123]
9. Freije WA, et al. Gene expression profiling of gliomas strongly predicts survival. *Cancer Res* 2004;64:6503–10. [PubMed: 15374961]
10. Murat A, et al. Stem cell-related “self-renewal” signature and high epidermal growth factor receptor expression associated with resistance to concomitant chemoradiotherapy in glioblastoma. *J Clin Oncol* 2008;26:3015–24. [PubMed: 18565887]
11. Nutt CL, et al. Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. *Cancer Res* 2003;63:1602–7. [PubMed: 12670911]
12. Sun L, et al. Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. *Cancer Cell* 2006;9:287–300. [PubMed: 16616334]
13. Beroukhim R, et al. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci U S A* 2007;104:20007–12. [PubMed: 18077431]
14. Wiedemeyer R, et al. Feedback circuit among INK4 tumor suppressors constrains human glioblastoma development. *Cancer Cell* 2008;13:355–64. [PubMed: 18394558]
15. Taylor BS, et al. Functional Copy-number Alterations in Cancer. *PLoS ONE*. 2008in press
16. Hunter C, et al. A hypermutation phenotype and somatic MSH6 mutations in recurrent human malignant gliomas after alkylator chemotherapy. *Cancer Res* 2006;66:3987–91. [PubMed: 16618716]
17. Cahill DP, et al. Loss of the Mismatch Repair Protein MSH6 in Human Glioblastomas Is Associated with Tumor Progression during Temozolomide Treatment. *Clin Cancer Res* 2007;13:2038–45. [PubMed: 17404084]
18. Getz G, et al. Comment on “The consensus coding sequences of human breast and colorectal cancers”. *Science* 2007;317:1500. [PubMed: 17872428]
19. Reilly KM, Loisel DA, Bronson RT, McLaughlin ME, Jacks T. Nf1;Trp53 mutant mice develop glioblastoma with evidence of strain-specific effects. *Nat Genet* 2000;26:109–13. [PubMed: 10973261]
20. Kwon CH, et al. Pten haploinsufficiency accelerates formation of high-grade astrocytomas. *Cancer Res* 2008;68:3286–94. [PubMed: 18451155]

21. Thiel G, et al. Somatic mutations in the neurofibromatosis 1 gene in gliomas and primitive neuroectodermal tumours. *Anticancer Res* 1995;15:2495–9. [PubMed: 8669813]
22. Zhu Y, et al. Early inactivation of p53 tumor suppressor gene cooperating with NF1 loss induces malignant astrocytoma. *Cancer Cell* 2005;8:119–30. [PubMed: 16098465]
23. Upadhyaya M, et al. Mutational and functional analysis of the neurofibromatosis type 1 (NF1) gene. *Hum Genet* 1997;99:88–92. [PubMed: 9003501]
24. Side L, et al. Homozygous inactivation of the NF1 gene in bone marrow cells from children with neurofibromatosis type 1 and malignant myeloid disorders. *N Engl J Med* 1997;336:1713–20. [PubMed: 9180088]
25. Fahsold R, et al. Minor lesion mutational spectrum of the entire NF1 gene does not explain its high mutability but points to a functional domain upstream of the GAP-related domain. *Am J Hum Genet* 2000;66:790–818. [PubMed: 10712197]
26. Messiaen LM, et al. Exhaustive mutation analysis of the NF1 gene allows identification of 95% of mutations and reveals a high frequency of unusual splicing defects. *Hum Mutat* 2000;15:541–55. [PubMed: 10862084]
27. Humphrey PA, et al. 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* 1990;87:4207–11. [PubMed: 1693434]
28. Lee JC, et al. Epidermal Growth Factor Receptor Activation in Glioblastoma through Novel Missense Mutations in the Extracellular Domain. *PLoS Med* 2006;3:e485. [PubMed: 17177598]
29. Ekstrand AJ, Sugawa N, James CD, Collins VP. 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* 1992;89:4309–13. [PubMed: 1584765]
30. Stephens P, et al. Lung cancer: intragenic ERBB2 kinase mutations in tumours. *Nature* 2004;431:525–6. [PubMed: 15457249]
31. Bamford S, et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer* 2004;91:355–8. [PubMed: 15188009]
32. Samuels Y, et al. High frequency of mutations of the PIK3CA gene in human cancers. *Science* 2004;304:554. [PubMed: 15016963]
33. Gallia GL, et al. PIK3CA gene mutations in pediatric and adult glioblastoma multiforme. *Mol Cancer Res* 2006;4:709–14. [PubMed: 17050665]
34. Bader AG, Kang S, Zhao L, Vogt PK. Oncogenic PI3K deregulates transcription and translation. *Nat Rev Cancer* 2005;5:921–9. [PubMed: 16341083]
35. Bader AG, Kang S, Vogt PK. Cancer-specific mutations in PIK3CA are oncogenic in vivo. *Proc Natl Acad Sci U S A* 2006;103:1475–9. [PubMed: 16432179]
36. Liu Z, Roberts TM. Human tumor mutants in the p110alpha subunit of PI3K. *Cell Cycle* 2006;5:675–7. [PubMed: 16627990]
37. Huang CH, et al. The structure of a human p110alpha/p85alpha complex elucidates the effects of oncogenic PI3Kalpha mutations. *Science* 2007;318:1744–8. [PubMed: 18079394]
38. Philp AJ, et al. The phosphatidylinositol 3'-kinase p85alpha gene is an oncogene in human ovarian and colon tumors. *Cancer Res* 2001;61:7426–9. [PubMed: 11606375]
39. Mizoguchi M, Nutt CL, Mohapatra G, Louis DN. Genetic alterations of phosphoinositide 3-kinase subunit genes in human glioblastomas. *Brain Pathol* 2004;14:372–7. [PubMed: 15605984]
40. Zhang H, et al. Comprehensive analysis of oncogenic effects of PIK3CA mutations in human mammary epithelial cells. *Breast Cancer Res Treat.* 2007
41. Pegg AE, Dolan ME, Moschel RC. Structure, function, and inhibition of O6-alkylguanine-DNA alkyltransferase. *Prog Nucleic Acid Res Mol Biol* 1995;51:167–223. [PubMed: 7659775]
42. Esteller M, et al. Inactivation of the DNA-repair gene MGMT and the clinical response of gliomas to alkylating agents. *N Engl J Med* 2000;343:1350–4. [PubMed: 11070098]
43. Hegi ME, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. *N Engl J Med* 2005;352:997–1003. [PubMed: 15758010]
44. Stommel JM, et al. Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies. *Science* 2007;318:287–90. [PubMed: 17872411]

45. Solit DB, et al. BRAF mutation predicts sensitivity to MEK inhibition. *Nature* 2006;439:358–62. [PubMed: 16273091]
46. Drablos F, et al. Alkylation damage in DNA and RNA--repair mechanisms and medical significance. *DNA Repair (Amst)* 2004;3:1389–407. [PubMed: 15380096]
47. Hirose Y, Kreklau EL, Erickson LC, Berger MS, Pieper RO. Delayed repletion of O6-methylguanine-DNA methyltransferase resulting in failure to protect the human glioblastoma cell line SF767 from temozolomide-induced cytotoxicity. *J Neurosurg* 2003;98:591–8. [PubMed: 12650433]
48. Kaina B, Christmann M, Naumann S, Roos WP. MGMT: key node in the battle against genotoxicity, carcinogenicity and apoptosis induced by alkylating agents. *DNA Repair (Amst)* 2007;6:1079–99. [PubMed: 17485253]
49. Casorelli I, Russo MT, Bignami M. Role of mismatch repair and MGMT in response to anticancer therapies. *Anticancer Agents Med Chem* 2008;8:368–80. [PubMed: 18473722]
50. Yang JL, Qu XJ, Yu Y, Kohn EC, Friedlander ML. Selective sensitivity to carboxyamidotriazole by human tumor cell lines with DNA mismatch repair deficiency. *Int J Cancer* 2008;123:258–63. [PubMed: 18464258]

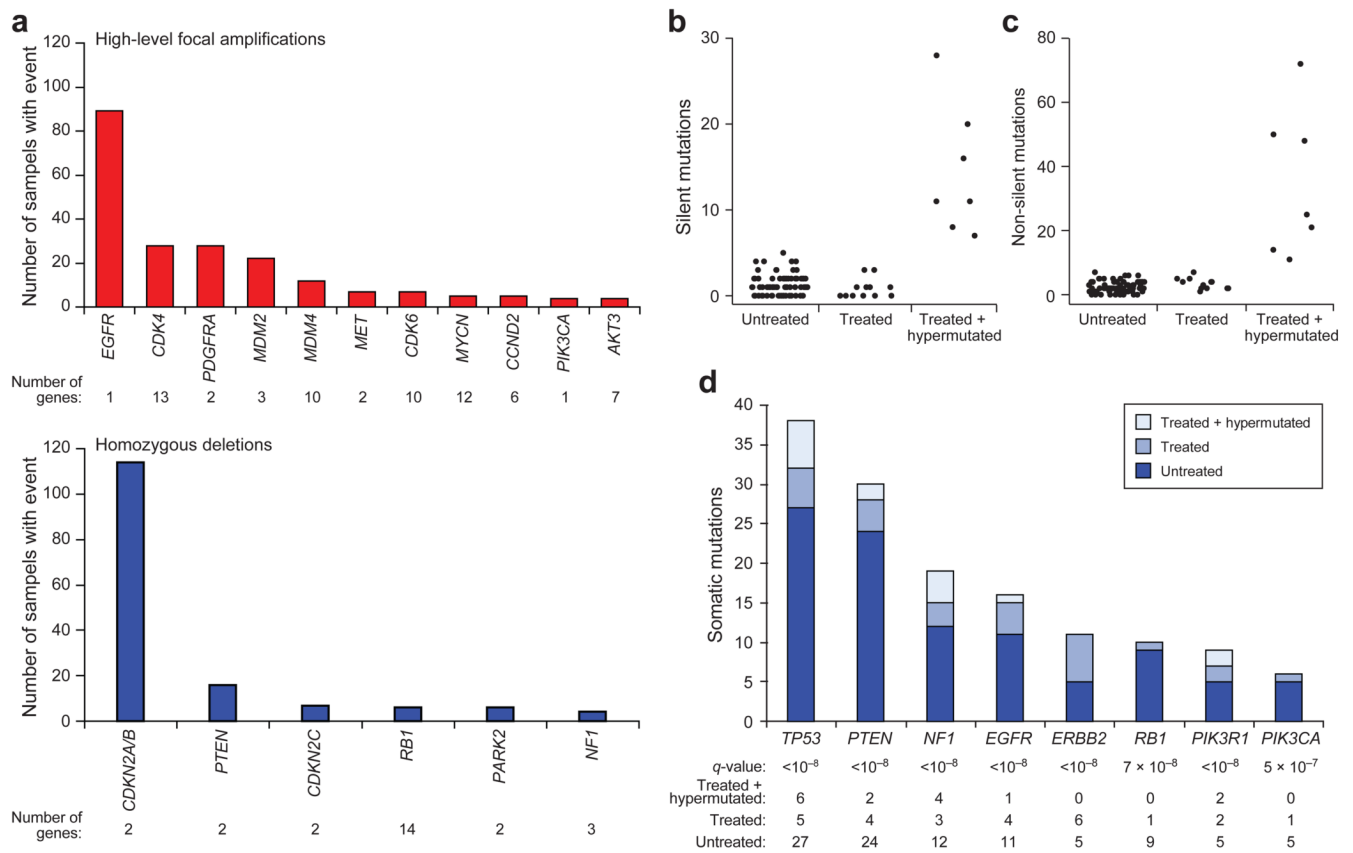


Figure 1. Significant copy number aberrations and pattern of somatic mutations

(a) Frequency and significance of focal high-level copy-number alterations. Known and putative target genes are listed for each significant CNA, with “Number of Genes” denoting the total number of genes within each focal CNA boundary.

(b–c) Distribution of the number of (b) silent and (c) non-silent mutations across the 91 GBM samples separated according to their treatment status, showing hypermutation in 7 out of the 19 treated samples.

(d) Significantly mutated genes in 91 glioblastomas. The eight genes attaining a false discovery rate <0.1 are displayed here. Somatic mutations occurring in untreated samples are in dark blue; those found in statistically non-hypermutated and hypermutated samples among the treated cohort are in respectively lighter shades of blue.

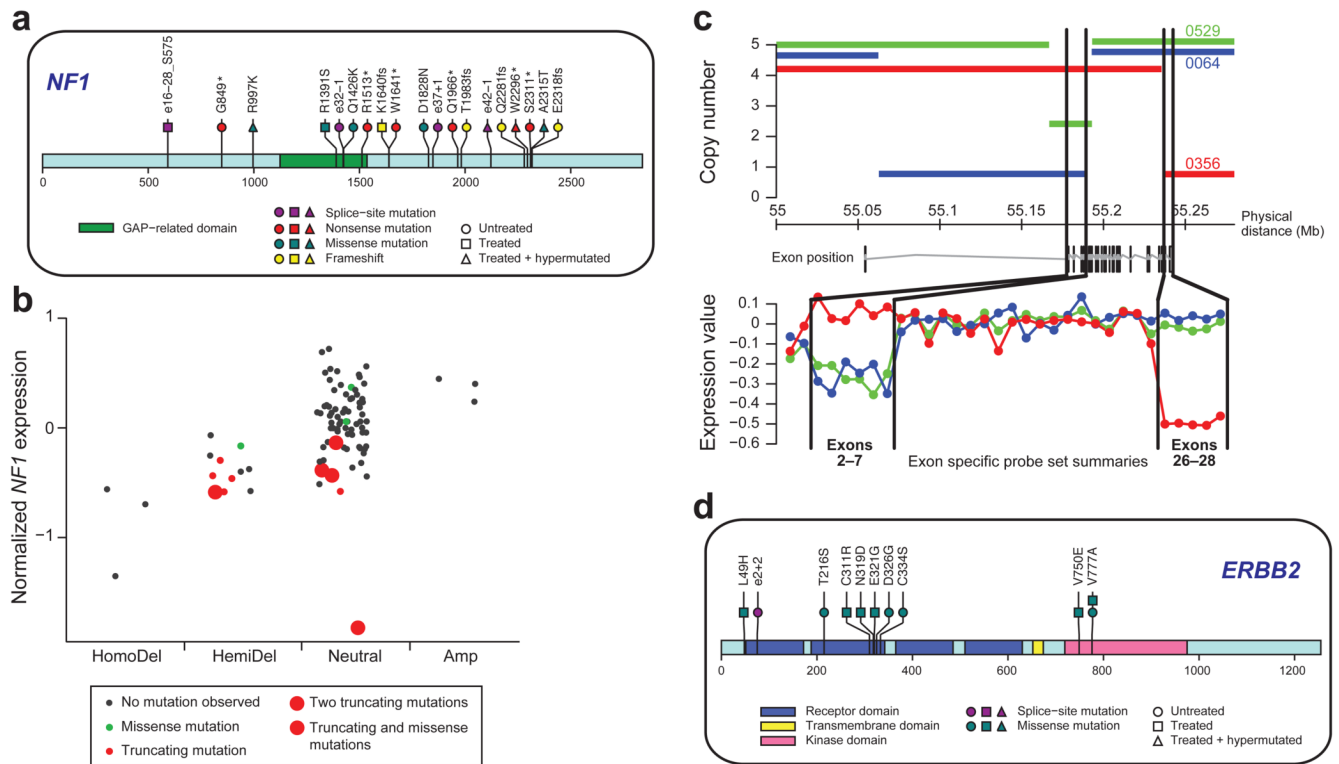


Figure 2. Mutations in *NF1* tumor suppressor gene and EGFR family members

(a) *NF1* somatic mutations in 91 glioblastoma tumors. Both missense mutations and truncating nonsense, frameshift, and splice site mutations were observed. Splice positions are given in number of bases to the closest exon (e#) numbered according to the *NF1* reference transcript in the Human Gene Mutation Database; positive = 3' of exon, negative = 5' of exon. *: stop codon. fs: frameshift.

(b) Correlation of copy number and mutation status at the *NF1* locus with level of expression (Y axis). Mutation events predicted to result in fewer expressed copies (including deletion, nonsense, splice site, and frameshift mutations) generally have lower observed expression. HomoDel = homozygous deletion; HemiDel = single-copy loss; Neutral = no change in copy number (presumed diploid); Amp = increased copy number. Copy number status of the *NF1* locus in each sample was determined as described in the Supplementary Information.

(c) DNA Copy number and mRNA expression profiles for TCGA samples TCGA-08-0356 (red), TCGA-02-0064 (blue), and TCGA-02-0529 (green) at the *EGFR* locus. The upper panel shows the segmented DNA copy number (based on Affymetrix SNP6.0 data) versus genomic coordinates on chromosome 7. The lower panel shows relative exon expression levels across the known *EGFR* exons from the Affymetrix Exon array ordered by genomic position, where relative expression is the median-centered difference in exon intensity and gene intensity. The *EGFR* gene model lies between the two plots. Black lines map the genomic positions of exons 2 through 7 and 26 through 28. Note that structural deletions cause the relatively lower expression of exons 2–7 in the green and blue samples and exons 26–28 in the red sample.

(d) *ERBB2* somatic mutations in 91 glioblastoma tumors. Mutations cluster in the extracellular domain in both genes. Splice site mutation position is given in number of bases to the closest exon (e#); positive = 3' of exon.

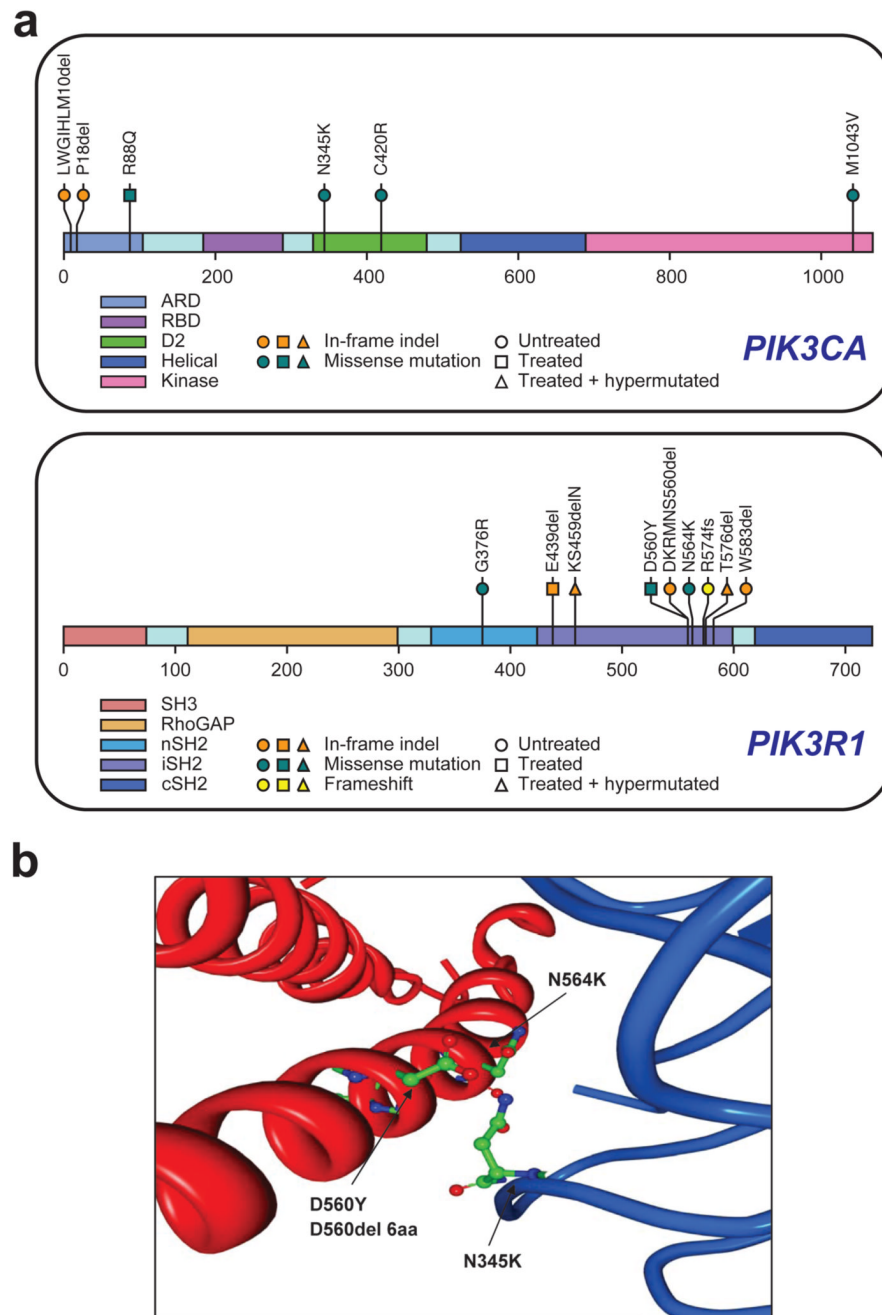


Figure 3. *PIK3R1* and *PIK3CA* mutations in GBM

(a). Diamonds above the backbone indicate the locations of mutations found in TCGA tumors. ABD: adaptor binding domain; RBD: Ras binding domain; C2: membrane-binding domain; iSH2: intervening domain.

(b). Four mutations found in the interaction interface of the p110 α ; C2 domain with iSH2 of p85 α . Two residues of p85 α , D560 and N564, are within hydrogen-bonding distance of the C2 residue of p110 α , N345.

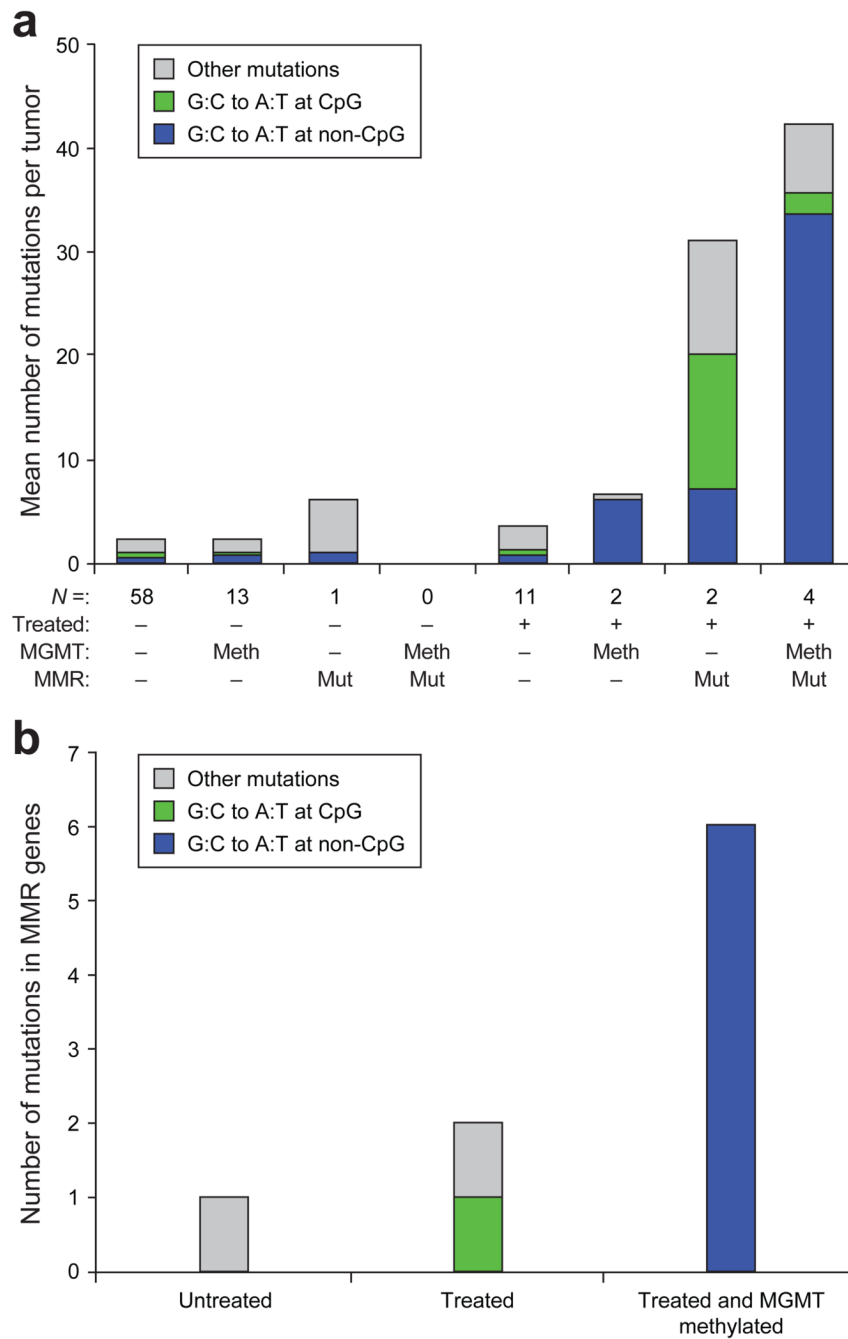


Figure 4. Pattern of somatic mutations, *MGMT* DNA methylation, and MMR gene mutations in treated GBMs

(a). The mean number of validated somatic nucleotide substitutions per tumor for key sample groups is indicated on the Y-axis and denoted by the height of the bar histograms. Samples are grouped along the X-axis according to treatment status of the patient (- = untreated; + = treated), DNA methylation status of *MGMT* (meth = DNA methylated; - = not methylated), and genetic status of MMR genes - = no genes mutated and mut = one or more of the *MLH1*, *MSH2*, *MSH6*, or *PMS2* genes mutated); the number below each bar indicates the number of samples in the group. Bars are color-coded for types of nucleotide substitutions including G-

to-A transitions at non-CpG sites (orange), G-to-A transitions at CpG sites (blue), and other mutation types (green).

(b). Bar histogram for mutation spectrum in the MMR genes as a function of treatment status, and methylation status of *MGMT*. The color code for substitution types is the same as in (a).

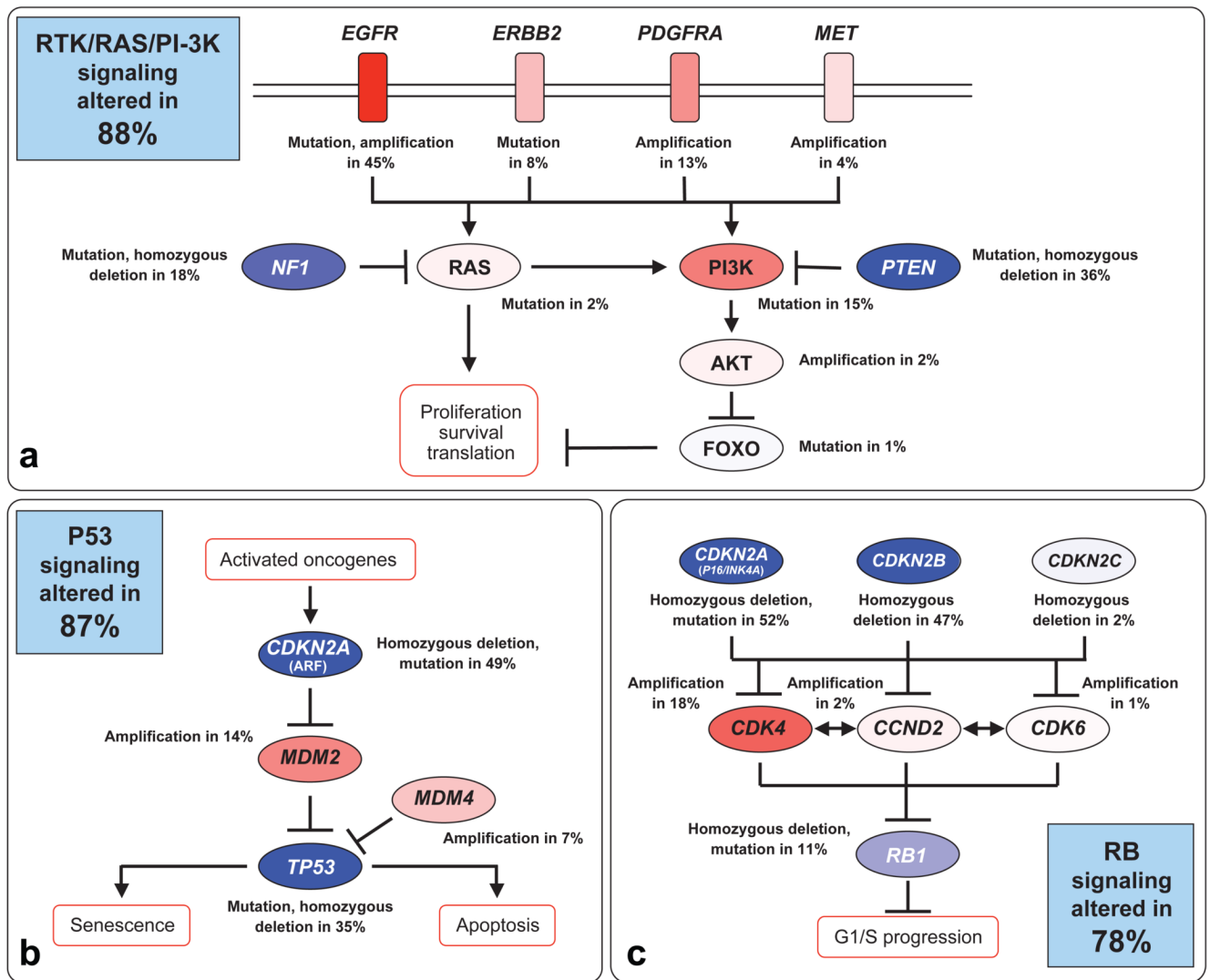


Figure 5. Frequent genetic alterations in three critical signaling pathways

Primary sequence alterations and significant copy number changes for components of the (a) RTK/RAS/PI-3K, (b) p53, and (c) RB signaling pathways are shown. Red indicates activating genetic alterations, with frequently altered genes showing deeper shades of red. Conversely, blue indicates inactivating alterations, with darker shades corresponding to a higher percentage of alteration. For each altered component of a particular pathway, the nature of the alteration and the percentage of affected tumors affected are indicated. Blue boxes contain the final percentages of glioblastomas with alterations in at least one known component gene of the designated pathway.