Comprehensive genomic characterization defines human glioblastoma genes and core pathways

The Cancer Genome Atlas Research Network*

Human cancer cells typically harbour 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 multi-dimensional 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—the most common type of primary adult brain cancer—and nucleotide sequence aberrations in 91 of the 206 glioblastomas. This analysis provides new insights into the roles of ERBB2, NF1 and TPS3, uncovers frequent mutations of the phosphatidylinositol-3-OH kinase regulatory subunit gene PIK3R1, and provides a network view of the pathways altered in the development of glioblastoma. 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 aims to catalogue and discover major cancer-causing genome alterations in large cohorts of human tumours through integrated multi-dimensional analyses.

The first cancer studied by TCGA is glioblastoma (World Health Organization grade IV), the most common primary brain tumour in adults. Primary glioblastoma, which comprises more than 90% of biopsied or resected cases, arises de novo without antecedent history of low-grade disease, whereas secondary glioblastoma progresses from previously diagnosed low-grade gliomas. Patients with newly diagnosed glioblastoma have a median survival of approximately 1 year with generally poor responses to all therapeutic modalities. Two decades of molecular studies have identified important genetic events in human glioblastomas, including the following: (1) dysregulation of growth factor signalling via amplification and mutational activation of receptor tyrosine kinase (RTK) genes; (2) activation of the phosphatidylinositol-3-OH kinase (PI3K) pathway; and (3) inactivation of the p53 and retinoblastoma tumour suppressor pathways. Recent genome-wide profiling studies have also shown remarkable genomic heterogeneity among glioblastoma and the existence of molecular subclasses within glioblastoma that may, when fully defined, allow stratification of treatment.

Albeit fragmentary, such baseline knowledge of glioblastoma genetics sets the stage to explore whether novel insights can be gained from a more systematic examination of the glioblastoma genome.

Results

Data release. 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 (for example, 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 the Supplementary 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 glioblastoma based on surgical pathology reports and clinical records (Supplementary Fig. 1). Samples were further selected for having matched normal tissues as well as associated demographic, clinical and pathological data (Supplementary Table 1). Corresponding frozen tissues were reviewed at the Biospecimen Core Resource (BCR) to ensure a minimum of 80% tumour nuclei and a maximum of 50% necrosis (Supplementary Fig. 1). DNA and RNA extracted from qualified biospecimens were subjected to additional quality control measurements (Supplementary Methods) before distribution to TCGA centres for analyses (Supplementary Fig. 2).

After exclusion based on insufficient tumour 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 or normal tissue DNAs and were therefore appropriate for re-sequencing. This cohort also included 21 post-treatment glioblastoma cases used for exploratory comparisons.

*Lists of participants and their affiliations appear at the end of the paper.
Genomic and transcriptional aberrations

Genomic copy number alterations (CNAs) were measured on three microarray platforms (Supplementary Methods) and analysed with multiple analytical algorithms\(^3,13\) (Supplementary Fig. 4 and Supplementary Tables 2–4). In addition to the well-known alterations\(^3,13\), we detected significantly recurrent focal alterations not previously reported in glioblastomas, such as homozygous deletions involving NF1 and PARK2, and amplifications of AKT3 (Fig. 1a and Supplementary Tables 2–4). Search for informative but infrequent CNAs also uncovered rare focal events, such as amplifications of FGFR2 and IRS2, and deletion of PTPRD (Supplementary Table 4).

Abundance of protein-coding genes and non-coding microRNA was also measured by transcript-specific and exon-specific probes on multiple platforms (Supplementary Methods). The resulting integrated gene expression data set showed that ~76% of genes within recurrent CNAs have expression patterns that correlate with copy number (Supplementary Table 2). In addition, single-nucleotide polymorphism (SNP)-based analyses also catalogued copy-neutral loss of heterozygosity (LOH), with the most significant region being 17p, which contains TP53 (Supplementary Methods).

Patterns of somatic nucleotide alterations in glioblastoma

A total of 91 matched tumour–normal pairs (72 untreated and 19 treated cases) were selected from the 143 cases for detection of somatic mutations in 601 selected genes (Supplementary Table 5). The resulting sequences, totalling 97 million base pairs (1.1 ± 0.1 million bases per sample), uncovered 453 validated non-silent somatic mutations in 223 unique genes, 79 of which contained two or more events (Supplementary Table 6; see also http://tcga-data.nci.nih.gov/docs/somatic_mutations/tcga_mutations.htm). The background mutation rates differed markedly between untreated and treated glioblastomas, averaging 1.4 versus 5.8 somatic silent mutations per sample (98 events among 72 untreated cases versus 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 seven hypermutated tumours were from patients previously treated with temozolomide and three were from patients treated with CCNU (lomustine) alone or in combination (Supplementary Table 1b). A hypermutator phenotype in glioblastoma has been described in three glioblastoma specimens with MSH6 mutations\(^16,17\), prompting us to perform a systematic analysis of the genes involved in mismatch repair (MMR). Indeed, six of the seven hypermutated samples harboured mutations in at least one of the MMR genes MLH1, MSH2, MSH6 or PMS2, as compared with only one sample among the eighty-four 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 significance\(^18\), we identified eight genes as significantly mutated (false discovery rate < 10\(^{-8}\)) (Fig. 2d and Supplementary Table 6). Interestingly, 27 TP53 mutations were detected in the 72 untreated glioblastomas (37.3%) 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 (Supplementary Fig. 5 and Supplementary Table 6). Given the predominance of primary status, showing hypermutation in 7 out of the 19 treated samples.

### Figure 1 | Significant copy number aberrations and pattern of somatic mutations

**a.** Frequency and significance of focal high-level CNAs. 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 silent (b) and non-silent (c) mutations across the 91 glioblastoma 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. Numbers of events in each group are noted.
NF1 is a human glioblastoma suppressor gene. Although somatic mutations in NF1 have been reported in a small series of human glioblastoma tumours\(^6\), their role remains controversial\(^2\), despite strong genetic data in mouse model systems\(^20\). Here, 19 NF1 somatic mutations were identified in 13 samples (14% of 91), including 6 nonsense mutations, 4 splice site mutations, 5 missense changes and 4 frameshift insertions/deletions (indels) (Fig. 2a). Five of these mutations—R1391S (ref. 23), R1513* (ref. 24), E25 -1 and E29 +1 (ref. 25) and Q1966* (ref. 26)—have been reported as germline alterations in neurofibromatosis patients, and thus are probably inactivating. In addition, 30 heterozygous deletions in NF1 were observed among the entire interim sample set of 206 cases, 6 of which also harbour point mutation (Supplementary Tables 8 and 9). Some samples also exhibited loss of expression without evidence of genomic alteration (Fig. 2b). Overall, at least 47 of these 206 patient samples (23%) harboured somatic NF1 inactivating mutations or deletions, definitively addressing NF1’s relevance to sporadic human glioblastoma.

Prevalence of EGFR family activation. EGFR is frequently activated in primary glioblastomas. Variant III deletion of the extracellular domain (‘vIII mutant’)\(^{37}\) has been the most commonly described event, in addition to extracellular domain point mutations and cytoplasmic domain deletions\(^{38,39}\). Here, high-resolution genomic and exon-specific transcriptomic profiling readily detected vIII and carboxy-terminal deletions with correspondingly altered transcripts (Fig. 2c). Among the 91 glioblastoma cases with somatic mutation data, 22 harboured focal amplification of wild-type \textit{EGFR} with no point mutation, 16 had point mutations in addition to focal amplification, and 3 had \textit{EGFR} point mutations but no amplification (Supplementary Fig. 6 and Supplementary Table 9). Collectively, \textit{EGFR} alterations were observed in 41 of the 91 sequenced samples.

\textit{ERBB2} mutation has previously been reported in only one glioblastoma tumour\(^9\). In the TCGA cohort, 11 somatic \textit{ERBB2} mutations in 7 of 91 samples were validated, including 3 in the kinase domain and 2 involving V777A, a site of recurrent missense and in-frame insertion mutations in lung, gastric and colon cancers\(^9\). The remaining eight mutations (including seven missense and one splice-site mutation) occurred in the extracellular domain of the protein, similar to somatic \textit{EGFR} substitutions in glioblastoma (Fig. 2d). Unlike in breast cancers, focal amplifications of \textit{ERBB2} were not observed in glioblastomas.

Somatic mutations of the PI(3)K complex in human glioblastoma. The PI(3)K complex consists of a catalytically active protein, p110\(\alpha\), and a regulatory protein, p85\(\alpha\), encoded by \textit{PIK3CA} and \textit{PIK3R1}. Frequent activating missense mutations of \textit{PIK3CA} have been reported in multiple tumour types, including glioblastoma\(^{40,41}\). These mutations occur primarily in the adaptor binding domain (ABD) as well as the C2 helical and kinase domains\(^{44,45}\). Indeed, \textit{PIK3CA} somatic nucleotide substitutions were detected in 6 of the 91 sequenced samples (Supplementary Table 6). Apart from the four mutations already reported in the COSMIC database (http://www.sanger.ac.uk/genetics/CGP/cosmic/), two novel in-frame deletions were detected in the adaptor binding domain of \textit{PIK3CA} (‘L10del’ and ‘P17del’). Those deletions may disrupt interactions between p110\(\alpha\) and its regulatory subunit, p85\(\alpha\) (ref. 37).

![Figure 2](link)

**Figure 2** | Mutations in NF1 tumour suppressor gene and EGFR family members. a, NF1 somatic mutations in 91 glioblastoma tumours. 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 \textit{Human Gene Mutation Database}; positive indicates 3’ of exon, negative indicates 5’ of exon. Asterisk indicates a 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-02-0329 (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 \textit{EGFR} exons from the Affymetrix exon array ordered by genomic position, where relative expression is the median-centred difference in exon intensity and gene intensity. The \textit{EGFR} gene model lies between the two plots. Black lines map the genomic positions of exons 2 through to 7 and 26 through to 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, \textit{ERBB2} somatic mutations in 91 glioblastoma tumours. Mutations cluster in the extracellular domain as seen in \textit{EGFR} in glioblastoma. Splice site mutation position is given in number of bases to the closest exon (e#); positive indicates 3’ of exon.
Unlike PIK3CA, PIK3R1 has rarely been reported as mutated in cancers. Among the five reported PIK3R1 nucleotide substitutions in cancers\(^{36,39}\), one was in a glioblastoma\(^{39}\). In our TCGA cohort, 9 PIK3R1 somatic mutations were detected among the 91 sequenced glioblastomas. 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-bp in-frame deletions (Fig. 3a and Supplementary Table 6). In accord with the crystal structure of PI(3)K, which identifies the D560 and N564 amino acid residues in proximity to the two key residues. We speculate that spatial constraints due to these deletions might prevent inhibitory contact of the p85:x N-terminal SH2 (nSH2) domain with the helical domain of PIK3CA (previously reported in colon and breast cancers\(^{36}\)) and D560Y and N564K mutations in PIK3R1. We also identified an 18-bp deletion spanning residues D560 to S565 (Fig. 3b) in addition to three other novel deletions (R574del, T576del and W583del) in proximity to the two key residues. Among the nine mutations, eight lay within the intervening SH2 (or iSH2) domain of p85et (ref. 37), the mutations detected in glioblastoma cluster around those three amino acid residues (Fig. 3b), including a N345K mutation in PIK3CA (previously reported in colon and breast cancers\(^{36}\)) and D560Y and N564K mutations in PIK3R1. PI(3)K strongly suggests that these novel PIK3R1 point mutations and indels disrupt the important C2-iSH2 interaction, relieving the inhibitory effect of p85\(\alpha\) on p110\(\alpha\).

**MGMT methylation and MMR in treated glioblastomas**

Cancer-specific DNA methylation of CpG dinucleotides located in CpG islands within the promoters of 2,305 genes was measured relative to normal brain DNA (Supplementary Table 7 and Supplementary Methods). The promoter methylation status of MGMT, a DNA repair enzyme that removes alkyl groups from guanine residues\(^{40}\), is associated with glioblastoma sensitivity to alkylating agents\(^{31,40}\). Among the 91 sequenced cases, 19 samples were found to contain MGMT promoter methylation (including 13 of the 72 untreated cases and 6 of the 19 treated cases). When juxtaposed with somatic mutation data, an

![Figure 3](https://example.com/figure3.png)

**Figure 3 | PIK3R1 and PIK3CA mutations in glioblastoma.** a, The locations of mutations found in TCGA tumours are indicated above the backbone. ABD, adaptor binding domain; RBD, Ras binding domain; C2, membrane-binding domain; nSH2, N-terminal SH2 domain; iSH2, inter-SH2 domain; cSH2, C-terminal SH2 domain. b, Three mutations found in the interaction interface of the C2 domain of p110\(\alpha\) with iSH2 of p85\(\alpha\). Two residues of p85\(\alpha\), D560 and N564, are within hydrogen-bonding distance of the C2 residue of p110\(\alpha\), N345.
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 glioblastomas (Fig. 4a). Among the 13 treated samples without MGMT methylation, 29% (29 out of 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 out of 99) of all mutations occurred as G•C to A•T transitions in non-CpG dinucleotides. In contrast, in the six treated samples with MGMT methylation, 81% of all mutations (146 out of 181) turned out to be of the G•C to A•T transition type in non-CpG dinucleotides, whereas only 4% (8 out of 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.

Notably, the mutational spectra in the MMR genes themselves reflected MGMT methylation status and treatment consequences. All seven mutations in MMR genes found in six MGMT methylated, hypermutated (treated) tumours occurred as G•C to A•T mutations at non-CpG sites (Fig. 4b and Supplementary Table 6), whereas neither MMR mutation in non-methylated, hypermutated tumours 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 glioblastoma tumours, an observation of potential clinical importance.

**Integrative analyses define glioblastoma core pathways**

To begin to construct an integrated view of common genetic alterations in the glioblastoma genome, we mapped the unequivocal genetic alterations—validated somatic nucleotide substitutions, homozygous deletions and focal amplifications—onto major pathways implicated in glioblastoma. That analysis identified a highly interconnected network of aberrations (Supplementary Figs 7 and 8), including three major pathways: RTK signalling, and the p53 and RB tumour suppressor pathways (Fig. 5). By copy number data alone, 66%, 70% and 59% of the 206 samples harboured somatic alterations in core components of the RB, TP53 and RTK pathways, respectively (Supplementary Table 8). In the 91 samples for which there was also sequencing data, the frequencies of somatic alterations increased to 87%, 78% and 88%, respectively (Supplementary Table 9). There was a statistical tendency towards mutual exclusivity of alterations of components within each pathway (P-values of $9.3 \times 10^{-10}$, $2.5 \times 10^{-13}$ and 0.022, respectively, for the p53, RB and RTK pathways; Supplementary Table 10), 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 harbours at least one aberrant gene from each of the three pathways (Supplementary Table 10). In fact, 74% harboured aberrations in all three pathways, a pattern suggesting that deregulation of the three pathways is a core requirement for glioblastoma pathogenesis.

As well as frequent deletions and mutations of the PTEN lipid phosphatase tumour suppressor gene, 86% of the glioblastoma samples harboured at least one genetic event in the core RTK/PI3K pathway (Fig. 5a). In addition to EGFR and ERBB2, PDGFRA (13%) and MET (4%) showed frequent aberrations (Supplementary Table 9). A total of 10 of the 91 sequenced samples have amplifications or point mutations in at least 2 of the 4 RTKs catalogued (EGFR, ERBB2, PDGFRA and MET; Supplementary Table 9), suggesting that genomic activation can be a mechanism for co-activated RTKs.

**Figure 5 | Frequent genetic alterations in three critical signalling pathways.** a–c. Primary sequence alterations and significant copy number changes for components of the RTK/RAS/PI(3)K (a), p53 (b) and RB (c) signalling 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 tumours affected are indicated. Boxes contain the final percentages of glioblastomas with alterations in at least one known component gene of the designated pathway.
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 and Supplementary Table 8). Among 91 sequenced samples (Supplementary Table 9), 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; Supplementary Table 10), but not of those in ARF. In fact, 10 of the 32 tumours with TP53 mutations also had deleted ARF, suggesting that homozygous deletion of the CDKN2A locus (which encodes both p16INK4A and ARF) was at least in part driven by p14ARF.

Among the 77% samples harbouring RB pathway aberrations (Fig. 5c), the most common event was deletion of the CDKN2A/CDKN2B locus on chromosome 13q21 (55% and 53%), followed by amplification of the CDK4 locus (14%) (Fig. 5c and Supplementary Tables 8 and 9). Although CNAs in the CDK/RB pathway members can co-occur in the same tumour, all nine samples with RB1 nucleotide substitutions (Supplementary Table 9) lacked CDKN2A/CDKN2B deletion or other CNAs 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 programme, TCGA has developed important principles in biospecimen banking and collection, 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 (Supplementary Fig. 3 and Supplementary Table 1).

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, glioblastoma tumours, 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 PIK3R1 might be expected to benefit from a PI(3)K or PDK1 inhibitor, whereas tumours in which the PI(3)K 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 glioblastoma specimens, 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 PI(3)K or cell cycle mediators. In contrast, glioblastomas with NF1 mutations might benefit from a RAF or MEK inhibitor as part of a combination, as shown for BRAF mutant cancers.

One of the most important biomarkers for glioblastomas is the methylation status of MGMT, which predicts sensitivity to temozolomide, an alkylating agent that is the current standard of care for glioblastoma 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 have an impact on clinical response and outcome. Newly diagnosed glioblastomas with MGMT methylation respond well to treatment with alkylating agents, in part as a consequence of unreacted alkylated guanine residues initiating cycles of futile mismatch repair, which can lead to cell death. Therefore, treatment of MGMT-deficient glioblastomas with alkylating therapy introduces a strong selective pressure to lose mismatch repair function. That conclusion is consistent with our observation that the mismatch repair genes themselves are mutated with characteristic C·G to A·T transitions at non-CpG sites resulting from unrepaird 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 hypothesis is validated, one may speculate that selective strategies designed to target mismatch-repair-deficient cells would represent a rational upfront combination with alkylating agent that together may prevent or minimize emergence of such resistance. Conversely, such a treatment-mediated mutator phenotype may enhance pathway mutations that can confer resistance to targeted therapies, thereby cautioning the combination of alkylating agents with targeted agents, as this may substantially increase the probability of developing resistance to such targeted drugs.

The power of TCGA to produce unprecedented multi-dimensional data sets using statistically robust numbers of samples sets the stage for a new era in the discovery of new cancer interventions. The integrative analyses leading to the 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 collections can lead to important discoveries.

**METHODS SUMMARY**

Biospecimens were screened from retrospective banks of tissue source sites under appropriate Institutional Review Board approvals for newly diagnosed glioblastoma with minimal 80% tumour cell percentage. RNA and DNA extracted from qualified specimens were distributed to TCGA centres for analysis. Whole-genome-amplified genomic DNA samples from tumours and normal samples were sequenced by the Sanger method. Mutations were called, verified using a second genotyping platform, and systematically analysed 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). Messenger RNA and microRNA (miRNA) expression profiles were generated using Affymetrix U133A, AffymetrixExon 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.

Received 28 July; accepted 1 September 2008.

Published online 4 September 2008, corrected online 17 September 2008.


The Cancer Genome Atlas Research Network

Tissue source sites: Duke University Medical School Roger McLendon; Allan Friedman; Darrell Bigner; Emory University Erwin G. Van Meir; S. R. Daniel; J. Brat; Gena M. Mastrogiannakis; Jeffrey J. Olson; Henry Ford Hospital Tom Mikkelson; Norman Leaman; MD Anderson Cancer Center Ken Aldape; W. K. Alfred Yung; Roger Bogler; John N. Weinstein; University of California San Francisco Scott Vanderberg; Mitchell Berger; Richard Prados

Genome sequencing centres: Baylor College of Medicine Donna Muzny; Margaret Morgan; Steve Scherer; Aniko Sabo; Lynn Nazareth; Lora Lewis; Otis Hall; Yiming Zhu; Yang Ren; Omar Alvi; Jiagong Yao; Alicia Hawes; Shalini Jhangiani; Gerald Fowler; Anthony San Lucas; Christie Kovar; Andrew Cree; Huyen Dinh; Jirhe Santibanez; Vandita Joshi; Manuel Gonzalez-Garay; Christopher A. Miller; Alejandro Milosavljevic; Larry Donehower; David A. Wheeler; Richard A. Gibbs; Broad Institute of MIT and Harvard Kristian

Acknowledgements We thank the members of TCGA’s External Scientific Committee, the Glioblastoma Disease Working Group (http://cancergenome.nih.gov/components) and D. N. Louis for discussions; A. Mirick, J. Melone and C. Collins for administrative coordination of TCGA activities; and L. G. Foster for graphic art. This work was supported in part by the United States National Institutes of Health: U54HG030367, U54HG030379, U54HG030327, U24CA126543, U24CA126544, U24CA126546, U24CA126551, U24CA126554, U24CA126561 and U24CA126563.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.