

An Integrated Genomic Analysis of Human Glioblastoma Multiforme

D. Williams Parsons,^{1,2*} Siân Jones,^{1*} Xiaosong Zhang,^{1*} Jimmy Cheng-Ho Lin,^{1*} Rebecca J. Leary,^{1*} Philipp Angenendt,^{1*} Parminder Mankoo,³ Hannah Carter,³ I-Mei Siu,⁴ Gary L. Gallia,⁴ Alessandro Olivi,⁴ Roger McLendon,⁵ B. Ahmed Rasheed,⁵ Stephen Keir,⁵ Tatiana Nikolskaya,⁶ Yuri Nikolsky,⁷ Dana A. Busam,⁸ Hanna Tekleab,⁸ Luis A. Diaz Jr.,¹ James Hartigan,⁹ Doug R. Smith,⁹ Robert L. Strausberg,⁸ Suely Kazue Nagahashi Marie,¹⁰ Sueli Mieko Oba Shinjo,¹⁰ Hai Yan,⁵ Gregory J. Riggins,⁴ Darell D. Bigner,⁵ Rachel Karchin,³ Nick Papadopoulos,¹ Giovanni Parmigiani,¹ Bert Vogelstein,^{1†} Victor E. Velculescu,^{1†} Kenneth W. Kinzler^{1†}

¹Ludwig Center for Cancer Genetics and Therapeutics, and The Howard Hughes Medical Institute at Johns Hopkins Kimmel Cancer Center, Baltimore, MD 21231, USA. ²Department of Pediatrics, Section of Hematology-Oncology, Baylor College of Medicine, Houston TX 77030, USA. ³Department of Biomedical Engineering, Institute of Computational Medicine, Johns Hopkins Medical Institutions, Baltimore, MD 21218, USA. ⁴Department of Neurosurgery, Johns Hopkins Medical Institutions, Baltimore, MD 21231, USA. ⁵Department of Pathology, Pediatric Brain Tumor Foundation, and Preston Robert Tisch Brain Tumor Center at Duke University Medical Center, Durham, NC 27710, USA. ⁶Vavilov Institute for General Genetics, Moscow B333, 117809, Russia. ⁷GeneGo, Inc., St. Joseph, MI 49085, USA. ⁸J. Craig Venter Institute, Rockville, MD 20850, USA. ⁹Agencourt Bioscience Corporation, Beverly, MA 01915, USA. ¹⁰Department of Neurology, School of Medicine, University of São Paulo, São Paulo, Brazil.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: bertvog@gmail.com (B.V.); velculescu@jhmi.edu (V.E.V.); kinzke@welch.jhu.edu (K.W.K.)

Glioblastoma multiforme (GBM) is the most common and lethal type of brain cancer. To identify the genetic alterations in GBMs, we sequenced 20,661 protein coding genes, determined the presence of amplifications and deletions using high-density oligonucleotide arrays, and performed gene expression analyses using next-generation sequencing technologies in 22 human tumor samples. This comprehensive analysis led to the discovery of a variety of genes that were not known to be altered in GBMs. Most notably, we found recurrent mutations in the active site of isocitrate dehydrogenase 1 (*IDH1*) in 12% of GBM patients. Mutations in *IDH1* occurred in a large fraction of young patients and in most patients with secondary GBMs, and were associated with an increase in overall survival. These studies demonstrate the value of unbiased genomic analyses in the characterization of human brain cancer and identify a potentially useful genetic alteration for the classification and targeted therapy of GBMs.

Malignant gliomas are the most frequent and lethal cancers originating in the central nervous system. The most biologically aggressive subtype is glioblastoma multiforme (GBM; WHO grade IV astrocytoma), a tumor associated with a dismal prognosis (1). The current standard of care for GBM patients—surgical resection followed by adjuvant radiation

therapy and chemotherapy with the oral alkylating agent temozolomide—produces a median survival of only 15 months (2). Historically, GBMs have been categorized into two groups (“primary” and “secondary”) on the basis of clinical presentation (3). Secondary GBMs are defined as cancers that have clinical, radiologic, or histopathologic evidence of malignant progression from a preexisting lower-grade tumor, while primary GBMs have no such history and present at diagnosis as advanced cancers (4). Clinical differences have been reported between the two groups, with secondary GBMs occurring less frequently (~5% of GBMs) and predominantly in younger patients (median age ~45 years vs. ~60 years for primary GBM) (5, 6). The histopathologic findings of primary and secondary GBMs are indistinguishable, and the prognosis does not appear to be significantly different after adjustment for age (5, 6).

Substantial research effort has focused on the identification of genetic alterations in GBMs that might help define subclasses of GBM patients with differing prognoses and/or response to specific therapies (7). Distinctions between the genetic lesions found in primary and secondary GBMs have been made, with *TP53* mutations occurring more commonly in secondary GBMs and *EGFR* amplifications and *PTEN* mutations occurring more frequently in primary GBMs

(6, 8, 9); however, none of these alterations are sufficiently specific to distinguish between primary and secondary GBMs. This issue is further confounded by the possibility that a fraction of GBMs designated as primary tumors may follow a sequence of genetic events similar to that of secondary lesions but not come to clinical attention until malignant progression to a GBM has occurred.

The comprehensive elucidation of genetic alterations in GBMs could provide novel targets that might be used for diagnostic, prognostic, or therapeutic purposes as well as identify subgroups of patients that preferentially respond to particular targeted therapies. The determination of the human genome sequence and improvements in sequencing and bioinformatic technologies have recently permitted genome-wide sequence analyses in human cancers. We have previously studied the genomes of 11 breast and 11 colorectal cancers by determining the sequence of the more than 18,000 Consensus Coding Sequence (CCDS) and Reference Sequence (RefSeq) genes (10, 11). Here, we have analyzed 20,661 protein coding genes in 22 human GBM samples. To complement these sequencing data, we have also performed a genome-wide analysis of focal copy number alterations, including amplifications and homozygous deletions, using high-density oligonucleotide microarrays on the same GBM tumors. Finally, we have examined the expression profiles of these same samples using serial analysis of gene expression (SAGE) and next-generation sequencing technologies.

Sequencing strategy. We extended our previous sequencing strategy for identification of somatic mutations to include 23,219 transcripts from 20,661 genes (fig. S1). These included 2783 additional genes from the Ensembl databases that were not present in the CCDS or RefSeq databases analyzed in the previous studies (10, 11). In addition, we redesigned PCR primers for regions of the genome that (i) were difficult to PCR amplify in prior studies; or (ii) were found to share significant identity with other human or mouse sequences. The combination of these new, redesigned, and existing primers sequences resulted in a total of 208,311 primer pairs (table S1) that were successfully used for sequence analysis of the coding exons of these genes.

Twenty-two GBM samples (table S2) were selected for PCR sequence analysis, consisting of 7 samples extracted directly from patient tumors and 15 samples passaged in nude mice as xenografts. In the first stage of this analysis, called the Discovery Screen, the primer pairs were used to amplify and sequence 175,471 coding exons and adjacent intronic splice donor and acceptor sequences in 22 GBM samples and one matched normal sample. The data were assembled for each amplified region and evaluated using stringent quality criteria (12), resulting in successful amplification and sequencing of 95.0% of targeted amplicons and 93.0% of

targeted bases in the 22 tumors (Table 1). A total of 689 Mb of sequence data was generated in this fashion. The amplicon traces were analyzed using automated approaches to identify changes in the tumor sequences that were not present in the reference sequences of each gene. Alterations present in the normal control sample and in single nucleotide polymorphism (SNP) databases were then removed from further analyses. The remaining sequence traces of potential alterations were visually inspected to remove false-positive mutation calls generated by the automated software. All exons containing putative mutations were then reamplified and sequenced in both the affected tumor and the matched normal DNA sample. This process allowed us to confirm the presence of the mutation in the tumor sample and determine whether the alteration was somatic (i.e. tumor-specific) or was present in the germline. All putative somatic mutations were examined computationally and experimentally to confirm that the alterations did not arise through the aberrant coamplification of related gene sequences (12).

Analysis of sequence alterations. Analysis of the identified somatic mutations revealed that one tumor (Br27P), from a patient previously treated with radiation therapy and temozolomide, had 17-fold more alterations than any of the other 21 patients (table S3). The mutation spectrum of this sample was also dramatically different from those of the other GBM patients (12) and was consistent with previous observations of a hypermutation phenotype in glioma samples of patients treated with temozolomide (13, 14). After removing Br27P from consideration, we found that 685 genes (3.3% of the 20,661 genes analyzed) contained at least one nonsilent somatic mutation. The vast majority of these alterations were single-base substitutions (94%), while the others were small insertions, deletions, or duplications (Table 1). The 993 somatic mutations were observed to be distributed relatively evenly among the 21 remaining tumors (table S3), with a mean of 47 mutations per tumor, representing 1.51 mutations per Mb of GBM tumor genome sequenced. The six DNA samples extracted directly from patient tumors had smaller numbers of mutations than those obtained from xenografts, likely because of the masking effect of nonneoplastic cells in the former. It has previously been shown that cell lines and xenografts provide the optimal template DNA for cancer genome sequencing analyses (15) and that they faithfully represent the alterations present in the original tumors (16). Both the total number and frequency of sequence alterations in GBMs were substantially smaller than the number and frequency of such alterations observed in colorectal or breast cancers, and slightly less than in pancreatic cancers (10, 11, 17). The most likely explanation for this difference is the reduced number of cell generations in glial cells prior to the onset of neoplasia (18).

We further evaluated a set of 21 mutated genes identified in the Discovery Screen in a second screen, called a Prevalence Screen, comprising an additional 83 GBMs with well-documented clinical histories (table S2). The 21 genes selected were mutated in at least two Discovery Screen tumors and had mutation frequencies of >10 mutations per Mb of tumor DNA sequenced. Nonsilent somatic mutations were identified in 16 of these 21 genes in the additional tumor samples (table S4). The mutation frequency of all analyzed genes in the Prevalence Screen was 24 mutations per Mb of tumor DNA, markedly increased from the overall mutation frequency in the Discovery Screen of 1.5 mutations per Mb ($P < 0.001$, binomial test). Additionally, the observed ratio of nonsilent to silent mutations among mutations in the Prevalence Screen was 14.8:1, substantially higher than the 3.1:1 ratio that was observed in the Discovery Screen ($P < 0.001$, binomial test). The increased mutation frequency and higher fraction of nonsilent mutations suggested that genes mutated in the Prevalence Screen were enriched for genes that actively contributed to tumorigenesis.

In addition to the frequency of mutations in a gene, the type of mutation can provide information useful for evaluating its potential role in disease (19). The likely effect of missense mutations can be assessed through evaluation of the mutated residue by evolutionary or structural means. To evaluate missense mutations, we developed an algorithm (LS-MUT) that employs machine learning of 58 predictive features based on evolutionary conservation and the physical-chemical properties of amino acids involved in the alteration (12). Approximately 15% of the missense mutations evaluated were predicted to have a statistically significant effect on protein function when assessed by this method (table S3). We also were able to make structural models of 244 of the 870 missense mutations identified in this study (20). In each case, the model was based on x-ray crystallography or nuclear magnetic resonance spectroscopy of the normal protein or a closely related homolog. This analysis showed that 35 of the missense mutations are located close to a domain interface or substrate-binding site and thus are likely to impact protein function [(links to structural models available in (12)].

Analysis of copy number changes. The same tumors were then evaluated for copy number alterations through genomic hybridization of DNA samples to Illumina SNP arrays containing ~1 million probes (21). We have recently developed a sensitive and specific approach for the identification of focal amplifications resulting in 12 or more copies per nucleus (6-fold or greater amplification compared to the diploid genome) as well as deletions of both copies of a gene (homozygous deletions) using such arrays (22). Unlike larger chromosomal aberrations, such focused alterations can

be used to identify underlying candidate genes in these regions.

We identified a total of 147 amplifications (table S5) and 134 homozygous deletions (table S6) in the 22 samples used in the Discovery Screen (Table 1). Although the number of amplifications was similar in samples extracted from patient tumors and those that had been passaged as xenografts, the latter samples allowed detection of a larger number of homozygous deletions (average of 8.0 deletions per sample in the xenografts versus 2.2 per sample in the tumors). These observations are consistent with previous reports documenting the difficulty of identifying homozygous deletions in samples containing contaminating normal DNA (23) and highlight the importance of using purified human tumor cells, such as those present in xenografts or cell lines, for genomic analyses.

Integration of sequencing, copy number and expression analyses. Mutations that arise during tumorigenesis may provide a selective advantage to the tumor cell (driver mutations) or have no net effect on tumor growth (passenger mutations). The mutational data obtained from sequencing and analysis of copy number alterations were integrated in order to identify GBM candidate cancer genes (*CAN*-genes) that are most likely to be drivers and therefore worthy of further investigation. To determine if a gene was likely to harbor driver mutations, we compared the number and type of mutations observed (including sequence changes, amplifications and homozygous deletions) and determined the probability that these alterations would result from passenger mutation rates alone (12) (fig. S1).

The *CAN*-genes, together with their passenger probabilities, are listed in table S7. The *CAN*-genes included several with well-established roles in gliomas, including *TP53*, *PTEN*, *CDKN2A*, *RBI*, *EGFR*, *NF1*, *PIK3CA* and *PIK3R1* (24–34). Of these genes, the most frequently altered were *CDKN2A* (altered in 50% of GBMs), *TP53*, *EGFR*, and *PTEN* (altered in 30 to 40%), *NF1*, *CDK4* and *RBI* (altered in 12 to 15%), and *PIK3CA* and *PIK3R1* (altered in 8 to 10%) (Table 2). Overall, these frequencies, which are similar to or in some cases higher than those previously reported, validate the sensitivity of our approach for detecting somatic alterations.

Through analysis of additional gene members within cell signaling pathways affected by these genes, we identified alterations of critical genes in the *TP53* pathway (*TP53*, *MDM2*, *MDM4*), the *RBI* pathway (*RBI*, *CDK4*, *CDKN2A*), and the *PI3K/PTEN* pathway (*PIK3CA*, *PIK3R1*, *PTEN*, *IRS1*). These alterations affected pathways in a majority of tumors (64%, 68%, and 50%, respectively) and in all cases but one, mutations within each tumor affected only a single

member of each pathway in a mutually exclusive manner ($P < 0.05$) (Table 3).

Systematic analyses of functional gene groups and pathways contained within the well-annotated MetaCore database (35) identified enrichment of alterations in a variety of cellular processes in GBMs, including additional members of the *TP53* and *PI3K/PTEN* pathways. Many of the pathways identified were similar to core signaling pathways found to be altered in pancreas, colorectal, and breast tumors, such as those regulating control of cellular growth, apoptosis and cell adhesion (17, 22, 36). However, several pathways were enriched only in GBMs. These included channels involved in transport of sodium, potassium and calcium ions as well as nervous system-specific cellular pathways such as synaptic transmission, transmission of nerve impulses, and axonal guidance (table S8). Mutations in these latter pathways may represent a subversion of normal glial cell processes to promote dysregulated growth and invasion.

Gene expression patterns can inform the analysis of pathways because they can reflect epigenetic alterations not detectable by sequencing or copy number analyses. To analyze the transcriptome of GBMs, we performed SAGE (serial analysis of gene expression) (37, 38) on all GBM samples for which RNA was available (total of 18 samples) as well as on two independent normal brain RNA controls. When combined with sequencing-by-synthesis methods (39–42), SAGE provides a highly quantitative and sensitive measure of gene expression. We first used the transcript analysis to help identify previously uncharacterized target genes from the amplified and deleted regions that were revealed by our study. In tables S5 and S6, a candidate target gene could be identified within several of these regions through the use of the mutational as well as transcriptional data. Second, we used the transcript analysis to help identify genes that were differentially expressed in GBMs compared to normal brain. A large number of genes (143) were expressed on average at 10-fold higher levels in the GBMs. Among the overexpressed genes, 16 encoded proteins that are predicted to be secreted or expressed on the cell surface, suggesting new opportunities for diagnostic and therapeutic applications. Finally, we assessed whether the gene sets implicated in the pathways enriched for genetic alterations were also altered through expression changes. Notably, the gene sets in these pathways were more highly enriched for differentially expressed genes than the remaining sets ($P < 0.001$) (12). These expression data thus independently highlight the potential importance of these pathways in the development of GBMs.

High frequency alterations of IDH1 in GBM. The *CAN*-gene list (table S7) included a number of individual genes that had not previously been linked to GBMs. The most frequently

mutated of these genes, *IDH1* on chromosome 2q33, encodes isocitrate dehydrogenase 1, which catalyzes the oxidative carboxylation of isocitrate to α -ketoglutarate, resulting in the production of NADPH. Of the five isocitrate dehydrogenase proteins encoded in the human genome, four are localized to the mitochondria while only *IDH1* is localized within the cytoplasm and peroxisomes (43). The *IDH1* protein forms an asymmetric homodimer (44), and is thought to play a substantial role in cellular control of oxidative damage through generation of NADPH (45, 46). None of the other *IDH* genes were found to be genetically altered in our analysis.

IDH1 was somatically mutated in five of the 22 GBM tumors in the Discovery Screen. Surprisingly, all five had the same heterozygous point mutation, a change of a guanine to an adenine at position 395 of the *IDH1* transcript (G395A), leading to the replacement of an arginine with a histidine at amino acid residue 132 of the protein (R132H). In our prior study of colorectal cancers, this same codon was mutated in a single case through alteration of the adjacent nucleotide, resulting in a R132C amino acid change (10). Five GBMs evaluated in our Prevalence Screen were found to have heterozygous somatic R132H mutations and an additional two tumors had a third distinct somatic mutation affecting the same amino acid residue, R132S (fig. S2 and Table 4). In addition to the Discovery and Prevalence Screen samples, 44 other GBMs were analyzed for *IDH1* mutations, revealing six tumors with somatic mutations affecting R132. In total, 18 of 149 GBMs (12%) analyzed had alterations in *IDH1*. The R132 residue is conserved in all known species and is localized to the substrate binding site, where it forms hydrophilic interactions with the alpha-carboxylate of isocitrate (Fig. 1) (44, 47).

Several important observations were made about *IDH1* mutations and their potential clinical significance. First, mutations in *IDH1* preferentially occurred in younger GBM patients, with a mean age of 33 years for *IDH1*-mutated patients, as opposed to 53 years for patients with wild-type *IDH1* ($P < 0.001$, t test, Table 4). In patients under 35 years of age, nearly 50% (9 of 19) had mutations in *IDH1*. Second, mutations in *IDH1* were found in nearly all of the patients with secondary GBMs (mutations in 5 of 6 secondary GBM patients, as compared to 7 of 99 patients with primary GBMs, $P < 0.001$, binomial test). Third, patients with *IDH1* mutations had a significantly improved prognosis, with a median overall survival of 3.8 years as compared to 1.1 years for patients with wild-type *IDH1* (Fig. 2, $P < 0.001$, log-rank test). Although both younger age and mutated *TP53* are known to be positive prognostic factors for GBM patients, this association between *IDH1* mutation and improved survival was noted even in the subgroup of young patients with *TP53* mutations ($P < 0.02$, log-rank test).

Discussion. The data resulting from this integrated analysis of mutations and copy number alterations have provided a novel view of the genetic landscape of glioblastomas. Like all large-scale genetic analyses, our study has limitations. We did not assess certain molecular alterations, including chromosomal translocations and epigenetic changes. However, our large scale expression studies should have identified any genes that were differentially expressed through these mechanisms (table S9). Additionally, we focused on copy number changes that were focal amplifications or homozygous deletions, as these have historically been most useful in identifying cancer genes. The array data we have generated can also be analyzed to determine loss of heterozygosity (LOH) or low-amplitude regions of copy number gains, but such changes cannot generally be used to pinpoint new candidate cancer genes. Finally, the samples directly extracted from patient tumors contained small amounts of contaminating normal tissue, which limited our ability to detect homozygous deletions and to a lesser extent, somatic mutations, in those specific tumors.

Despite these limitations, our study provides a number of important genetic and clinical insights into GBMs. First, it revealed that some of the pathways known to be altered in GBMs affect a larger fraction of genes and patients than previously anticipated. A majority of the tumors analyzed had alterations in genes encoding components of each of the *TP53*, *RBI*, and *PI3K* pathways. The fact that all but one of the cancers with mutations in members of a pathway did not have alterations in other members of the same pathway suggests that such alterations are functionally equivalent in tumorigenesis. Second, these results have identified a variety of new genes and signaling pathways not previously implicated in GBMs (table S7 and S8). Some of these pathways were found to be altered in previous genome-wide analyses of pancreatic, breast and colorectal cancers and may represent core processes that underlie human tumorigenesis (17, 22, 36). A number of the signaling pathways mutated or altered through expression differences in GBMs appear to be involved in nervous system signaling processes and represent novel and potentially useful aspects of GBM biology.

The comprehensive nature of our study allowed us to identify *IDH1* as an unexpected target of genetic alteration in patients with GBM. All mutations in this gene resulted in amino acid substitutions at position 132, an evolutionarily conserved residue located within the isocitrate binding site (44). The recurrent nature of the mutations is reminiscent of activating alterations in oncogenes such as *BRAF*, *KRAS*, and *PIK3CA*. Our speculation that this sequence change is an activating mutation is strengthened by the absence of inactivating changes (e.g., frameshift or stop mutations), the absence of other alterations in key residues of the active site,

and the fact that all mutations observed to date were heterozygous (without any evidence of loss of the second allele through LOH). Interestingly, enzymatic studies have shown that *in vitro* engineered substitution of arginine at residue 132 with a different amino acid than observed in patients (glutamate) results in a catalytically inactive enzyme, suggesting a critical role for this residue (48). Further biochemical and molecular analyses will be needed to determine the effect of alterations of *IDH1* on enzymatic activity and cellular phenotype.

Regardless of the specific molecular consequences of *IDH1* alterations, detection of mutations in *IDH1* is likely to be clinically useful. Although significant effort has focused on the identification of characteristic genetic lesions in primary and secondary GBMs, the altered genes identified to date are not optimal for this purpose (5). Our study revealed *IDH1* mutation to be a novel and potentially more specific marker for secondary GBM. One hypothesis is that *IDH1* alterations identify a biologically specific subgroup of GBM patients, including both patients who would be classified as having secondary GBMs as well as a subpopulation of primary GBM patients with a similar tumor biology and more protracted clinical course (Table 4). Interestingly, patients with *IDH1* mutations had a very high frequency of *TP53* mutation and a very low frequency of mutations in other commonly altered GBM genes (Table 4). Patients with mutated *IDH1* also had distinct clinical characteristics, including younger age and a significantly improved clinical prognosis (Table 4). It is conceivable that new treatments could be designed to take advantage of *IDH1* alterations in these patients, as inhibition of a different *IDH* enzyme (mitochondrial *IDH2*) has recently been shown to result in increased sensitivity of tumor cells to a variety of chemotherapeutic agents (49). In summary, the discovery of *IDH1* and other genes previously not known to play a role in human tumors (table S7) validates the utility of genome-wide genetic analysis of tumors in general and opens new avenues of basic and clinical brain tumor research.

References and Notes

1. D. N. Louis *et al.*, *Acta Neuropathol* **114**, 97 (2007).
2. R. Stupp *et al.*, *N Engl J Med* **352**, 987 (2005).
3. H. Scherer, *American Journal of Cancer* **40**, 159 (1940).
4. P. Kleihues, H. Ohgaki, *Neuro Oncol* **1**, 44 (1999).
5. H. Ohgaki, P. Kleihues, *Am J Pathol* **170**, 1445 (2007).
6. H. Ohgaki *et al.*, *Cancer Res* **64**, 6892 (2004).
7. I. K. Mellinghoff *et al.*, *N Engl J Med* **353**, 2012 (2005).
8. E. A. Maher *et al.*, *Cancer Res* **66**, 11502 (2006).
9. C. L. Tso *et al.*, *Cancer Res* **66**, 159 (2006).
10. T. Sjöblom *et al.*, *Science* **314**, 268 (2006).
11. L. D. Wood *et al.*, *Science* **318**, 1108 (2007).
12. See Supporting Online Material.
13. D. P. Cahill *et al.*, *Clin Cancer Res* **13**, 2038 (2007).

14. C. Hunter *et al.*, *Cancer Res* **66**, 3987 (2006).
15. J. M. Winter, J. R. Brody, S. E. Kern, *Cancer Biol Ther* **5**, 360 (2006).
16. S. Jones *et al.*, *Proc Natl Acad Sci U S A* **105**, 4283 (2008).
17. S. Jones *et al.*, *Science*, this issue (2008).
18. R. Kraus-Ruppert, J. Laissue, H. Burki, N. Odartchenko, *J Comp Neurol* **148**, 211 (1973).
19. P. C. Ng, S. Henikoff, *Nucleic Acids Res* **31**, 3812 (2003).
20. R. Karchin (2008). Structural models of mutants identified in glioblastomas.
<http://karchinlab.org/Mutants/CAN-genes/brain/GBM.html>
21. F. J. Steemers *et al.*, *Nat Methods* **3**, 31 (2006).
22. R. J. Leary *et al.*, *Proc Natl Acad Sci U S A*, in press.
23. P. Cairns *et al.*, *Nat Genet* **11**, 210 (1995).
24. J. M. Nigro *et al.*, *Nature* **342**, 705 (1989).
25. J. Li *et al.*, *Science* **275**, 1943 (1997).
26. K. Ueki *et al.*, *Cancer Res* **56**, 150 (1996).
27. A. J. Wong *et al.*, *Proc Natl Acad Sci U S A* **84**, 6899 (1987).
28. A. J. Wong *et al.*, *Proc Natl Acad Sci U S A* **89**, 2965 (1992).
29. L. Frederick, X. Y. Wang, G. Eley, C. D. James, *Cancer Res* **60**, 1383 (2000).
30. Y. Li *et al.*, *Cell* **69**, 275 (1992).
31. G. Thiel *et al.*, *Anticancer Res* **15**, 2495 (1995).
32. Y. Samuels *et al.*, *Science* **304**, 554 (2004).
33. D. K. Broderick *et al.*, *Cancer Res* **64**, 5048 (2004).
34. G. L. Gallia *et al.*, *Mol Cancer Res* **4**, 709 (2006).
35. S. Ekins, Y. Nikolsky, A. Bugrim, E. Kirillov, T. Nikolskaya, *Methods Mol Biol* **356**, 319 (2007).
36. J. Lin *et al.*, *Genome Res* **17**, 1304 (2007).
37. V. E. Velculescu, L. Zhang, B. Vogelstein, K. W. Kinzler, *Science* **270**, 484 (1995).
38. S. Saha *et al.*, *Nat Biotechnol* **20**, 508 (2002).
39. M. Sultan *et al.*, *Science* (2008).
40. R. Lister *et al.*, *Cell* **133**, 523 (2008).
41. A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, B. Wold, *Nat Methods* **5**, 621 (2008).
42. R. Morin *et al.*, *Biotechniques* **45**, 81 (2008).
43. B. V. Geisbrecht, S. J. Gould, *J Biol Chem* **274**, 30527 (1999).
44. X. Xu *et al.*, *J Biol Chem* **279**, 33946 (2004).
45. S. M. Lee *et al.*, *Free Radic Biol Med* **32**, 1185 (2002).
46. S. Y. Kim *et al.*, *Mol Cell Biochem* **302**, 27 (2007).
47. A. Nekrutenko, D. M. Hillis, J. C. Patton, R. D. Bradley, R. J. Baker, *Mol Biol Evol* **15**, 1674 (1998).
48. G. T. Jennings, K. I. Minard, L. McAlister-Henn, *Biochemistry* **36**, 13743 (1997).
49. I. S. Kil, S. Y. Kim, S. J. Lee, J. W. Park, *Free Radic Biol Med* **43**, 1197 (2007).
50. E. F. Pettersen *et al.*, *J Comput Chem* **25**, 1605 (2004).
51. We thank N. Silliman, J. Ptak, L. Dobbyn, and M. Whalen for assistance with PCR amplification; D. Lister, L. J. Ehinger, D. L. Satterfield, J. D. Funkhouser, and P. Killela for assistance with DNA purification; T. Sjöblom with for assistance with database management; the Agencourt sequencing team for assistance with automated sequencing; and C.-S. Liu and the SoftGenetics team for their assistance with mutation detection analyses. This project was carried out under the auspices of the Ludwig Brain Tumor Initiative and was supported by The Virginia and D. K. Ludwig Fund for Cancer Research, NIH grants CA121113, NS052507, CA43460, CA 57345, CA62924, CA09547, 5P50-NS-20023, CA108786, CA11898, The Pew Charitable Trusts, The Pediatric Brain Tumor Foundation Institute, The Hirschhorn Foundation, Alex's Lemonade Stand Foundation, The American Brain Tumor Association, The American Society of Clinical Oncology, The Brain Tumor Research Fund and Beckman Coulter Corporation. Under separate licensing agreements between the Johns Hopkins University and Genzyme, Beckman Coulter, and Exact Sciences Corporations, B.V., V.E.V., and K.W.K. are entitled to a share of royalties received by the University on sales of products related to research described in this paper. These authors and the University own Genzyme and Exact Sciences stock, which is subject to certain restrictions under University policy. The terms of these arrangements are managed by the Johns Hopkins University in accordance with its conflict-of-interest policies.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1164382/DC1

Materials and Methods

Figs. S1 and S2

Tables S1 to S9

References

7 August 2008; accepted 27 August 2008

Published online 4 September 2008;

10.1126/science.1164382

Include this information when citing this paper.

Fig. 1. Structure of the active site of IDH1. The crystal structure of the human cytosolic NADP(+)-dependent IDH is shown in ribbon format (PDBID: 1T0L) (44). The active cleft of IDH1 consists of a NADP-binding site and the isocitrate-metal ion-binding site. The alpha-carboxylate oxygen and the hydroxyl group of isocitrate chelate the Ca²⁺ ion. NADP is colored in orange, isocitrate in purple and Ca²⁺ in blue. The Arg132 residue, displayed in yellow, forms hydrophilic interactions, shown in red, with the alpha-carboxylate of

isocitrate. Displayed image was created with UCSF Chimera software version 1.2422 (50).

Fig. 2. Overall survival according to *IDH1* mutation status. The hazard ratio for death among patients with wild-type *IDH1* ($n = 79$), as compared to those with mutant *IDH1* ($n = 11$), was 3.7 (95 percent confidence interval, 2.1 to 6.5; $P < 0.001$). The median survival was 3.8 years for patients with mutated *IDH1*, as compared to 1.1 years for patients with wild-type *IDH1*.

Table 1. Summary of genomic analyses

Sequencing analysis	
Number of genes successfully analyzed	20,661
Number of transcripts successfully analyzed	23,219
Number of exons successfully analyzed	175,471
Primer pairs designed for amplification	219,229
Fraction of passing amplicons*	95.0%
Total number of nucleotides successfully sequenced	689,071,123
Fraction of passing amplicon sequences successfully analyzed†	98.3%
Fraction of targeted bases successfully analyzed†	93.0%
Number of somatic mutations identified (n=22 samples)	2,325
Number of somatic mutations (excluding Br27P)	993
Missense	622
Nonsense	43
Insertion	3
Deletion	46
Duplication	7
Splice site or UTR	27
Synonymous	245
Average number of sequence alterations per sample	47.3
Copy number analysis	
Total number of SNP loci assessed for copy number changes	1,069,688
Number of copy number alterations identified (n=22 samples)	281
Amplifications	147
Homozygous deletions	134
Average number of amplifications per sample	6.7
Average number of homozygous deletions per sample	6.1

*Passing amplicons were defined as having PHRED20 scores or better over 90% of the target sequence in 75% of samples analyzed.

†Fraction of nucleotides having PHRED20 scores or better (see Supporting Online Materials for additional information).

Table 2. Most frequently altered GBM CAN-genes

Gene	Point mutations [^]		Amplifications ^{&}		Homozygous deletions ^{&}		Fraction of tumors with any alteration	Passenger Probability [*]
	Number of tumors	Fraction of tumors	Number of tumors	Fraction of tumors	Number of tumors	Fraction of tumors		
CDKN2A	0/22	0%	0/22	0%	11/22	50%	50%	<0.01
TP53	37/105	35%	0/22	0%	1/22	5%	40%	<0.01
EGFR	15/105	14%	5/22	23%	0/22	0%	37%	<0.01
PTEN	27/105	26%	0/22	0%	1/22	5%	30%	<0.01
NF1	16/105	15%	0/22	0%	0/22	0%	15%	0.04
CDK4	0/22	0%	3/22	14%	0/22	0%	14%	<0.01
RB1	8/105	8%	0/22	0%	1/22	5%	12%	0.02
IDH1	12/105	11%	0/22	0%	0/22	0%	11%	<0.01
PIK3CA	10/105	10%	0/22	0%	0/22	0%	10%	0.10
PIK3R1	8/105	8%	0/22	0%	0/22	0%	8%	0.10

The most frequently-altered CAN-genes are listed; all CAN-genes are listed in Table S7. [^]Fraction of tumors with point mutations indicates the fraction of mutated GBMs out of the 105 samples in the Discovery and Prevalence Screens. CDKN2A and CDK4 were not analyzed for point mutations in the Prevalence Screen because no sequence alterations were detected in these genes in the Discovery Screen. [&]Fraction of tumors with amplifications and deletions indicates the number of tumors with these types of alterations in the 22 Discovery Screen samples. ^{*}Passenger probability indicates the Passenger probability - Mid (12).

Table 3. Mutations of the TP53, PI3K, and RB1 pathways in GBM samples

Tumor sample	TP53 pathway				PI3K Pathway					RB1 pathway			
	TP53	MDM2	MDM4	All genes	PTEN	PIK3CA	PIK3R1	IRS1	All genes	RB1	CDK4	CDKN2A	All genes
Br02X	Del			Alt				Mut	Alt			Del	Alt
Br03X	Mut			Alt	Mut				Alt				
Br04X	Mut			Alt	Mut				Alt	Mut			Alt
Br05X			Amp	Alt		Mut			Alt			Del	Alt
Br06X												Del	Alt
Br07X	Mut			Alt	Mut				Alt	Del			Alt
Br08X												Del	Alt
Br09P	Mut			Alt							Amp		Alt
Br10P	Mut			Alt									
Br11P	Mut			Alt									
Br12P	Mut			Alt			Mut		Alt				
Br13X	Mut			Alt								Del	Alt
Br14X							Mut		Alt			Del	Alt
Br15X										Mut		Del	Alt
Br16X		Amp		Alt							Amp		Alt
Br17X					Mut				Alt			Del	Alt
Br20P													
Br23X	Mut			Alt	Del				Alt				
Br25X					Mut				Alt			Del	Alt
Br26X						Mut			Alt			Del	Alt
Br27P	Mut			Alt							Amp		Alt
Br29P	Mut			Alt									
Fraction of tumors with altered gene/pathway [#]	0.55	0.05	0.05	0.64	0.27	0.09	0.09	0.05	0.50	0.14	0.14	0.45	0.68

* Mut, mutated; Amp, amplified; Del, deleted; Alt, altered [#]Fraction of affected tumors in 22 Discovery Screen samples

Table 4. Characteristics of GBM patients with IDH1 mutations

Patient ID	Patient age (years)*	Sex	Recurrent GBM [#]	Secondary GBM [^]	Overall survival (years) [§]	IDH1 Mutation		Mutation of TP53	Mutation of PTEN, RB1, EGFR, or NF1
						Nucleotide	Amino acid		
Br10P	30	F	No	No	2.2	G395A	R132H	Yes	No
Br11P	32	M	No	No	4.1	G395A	R132H	Yes	No
Br12P	31	M	No	No	1.6	G395A	R132H	Yes	No
Br104X	29	F	No	No	4.0	C394A	R132S	Yes	No
Br106X	36	M	No	No	3.8	G395A	R132H	Yes	No
Br122X	53	M	No	No	7.8	G395A	R132H	No	No
Br123X	34	M	No	Yes	4.9	G395A	R132H	Yes	No
Br237T	26	M	No	Yes	2.6	G395A	R132H	Yes	No
Br211T	28	F	No	Yes	0.3	G395A	R132H	Yes	No
Br27P	32	M	Yes	Yes	1.2	G395A	R132H	Yes	No
Br129X	25	M	Yes	Yes	3.2	C394A	R132S	No	No
Br29P	42	F	Yes	Unknown	Unknown	G395A	R132H	Yes	No
IDH1 mutant patients (n=12)	33.2	67% M	25%	42%	3.8	100%	100%	83%	0%
IDH1 wildtype patients (n=93)	53.3	65% M	16%	1%	1.1	0%	0%	27%	60%

*Patient age refers to age at which patient GBM sample was obtained. [#]Recurrent GBM designates a GBM which was resected >3 months after a prior diagnosis of GBM. [^]Secondary GBM designates a GBM which was resected > 1 year after a prior diagnosis of a lower grade glioma (WHO I-III). [§]Overall survival was calculated using date of GBM diagnosis and date of death or last patient contact; patients Br10P and Br11P were alive at last contact. Median survival for IDH1 mutant patients and IDH1 wildtype patients was calculated using logrank test. Previous pathologic diagnoses in secondary GBM patients were oligodendroglioma (WHO grade II) in Br123X, low grade glioma (WHO grade I-II) in Br237T and Br211T, anaplastic astrocytoma (WHO grade III) in Br27P, and anaplastic oligodendroglioma (WHO grade III) in Br129X. Abbreviations: GBM (glioblastoma multiforme, WHO grade IV), WHO (World Health Organization), M (male), F (female), mut (mutant). Mean age and median survival are listed for the groups of IDH1-mutated and IDH1-wildtype patients.



