nature genetics

The integrated landscape of driver genomic alterations in glioblastoma

Veronique Frattini[1,20], Vladimir Trifonov[2,3,20], Joseph Minhow Chan[2,3,20], Angelica Castano[1,20], Marie Lia[1,20], Francesco Abate[2–4], Stephen T Keir[5–7], Alan X Ji[8], Pietro Zoppoli[1], Francesco Niola[1,19], Carla Danussi[1], Igor Dolgalev[9], Paola Porrati[10], Serena Pellegatta[10], Adriana Heguy[9], Gaurav Gupta[11], David J Pisapia[12], Peter Canoll[12], Jeffrey N Bruce[11], Roger E McLendon[5–7], Hai Yan[5–7], Ken Aldape[13], Gaetano Finocchiaro[10], Tom Mikkelsen[14,15], Gilbert G Privé[8,16], Darell D Bigner[5–7], Anna Lasorella[1,12,17], Raul Rabadan[2,3] & Antonio Iavarone[1,12,18]

Glioblastoma is one of the most challenging forms of cancer to treat. Here we describe a computational platform that integrates the analysis of copy number variations and somatic mutations and unravels the landscape of in-frame gene fusions in glioblastoma. We found mutations with loss of heterozygosity in *LZTR1*, encoding an adaptor of CUL3-containing E3 ligase complexes. Mutations and deletions disrupt *LZTR1* function, which restrains the self renewal and growth of glioma spheres that retain stem cell features. Loss-of-function mutations in *CTNND2* target a neural-specific gene and are associated with the transformation of glioma cells along the very aggressive mesenchymal phenotype. We also report recurrent translocations that fuse the coding sequence of *EGFR* to several partners, with *EGFR-SEPT14* being the most frequent functional gene fusion in human glioblastoma. *EGFR-SEPT14* fusions activate STAT3 signaling and confer mitogen independence and sensitivity to EGFR inhibition. These results provide insights into the pathogenesis of glioblastoma and highlight new targets for therapeutic intervention.

© 2013 Nature America, Inc. All rights reserved.

Glioblastoma (GBM) is the most common primary intrinsic malignant brain tumor, affecting ~10,000 new patients each year with a median survival rate of 12–15 months[1,2]. Identifying and understanding the functional importance of genetic alterations that drive the initiation and progression of GBM is crucial to develop effective therapies. Previous efforts in GBM genome characterization have identified somatic changes in well-known GBM genes (*EGFR*, *PTEN*, *IDH1*, *TP53* and *NF1*, among others) and nominated putative cancer genes with somatic mutations, but the functional consequences of most of these alterations are unknown[3–6]. Furthermore, the abundance of passenger mutations and large regions of copy-number variations (CNVs) complicates the definition of the landscape of driver mutations in GBM. To address this challenge, we used a new statistical approach to nominate driver genes in GBM by integrating somatic mutations identified by whole-exome sequencing with a CNV analysis that prioritizes the focality and magnitude of the genetic alterations.

Recurrent and oncogenic gene fusions are hallmarks of hematological malignancies and have also been uncovered in solid tumors[7,8]. We recently reported that a small subset of GBMs harbor *FGFR-TACC* gene fusions and provided data to suggest that individuals with *FGFR-TACC*–positive tumors would benefit from targeted FGFR kinase inhibition[9]. It remains unknown whether gene fusions involving other receptor tyrosine kinase (RTK)-coding genes exist and produce oncogene addiction in GBM. Here we analyze a large RNA-sequencing (RNA-seq) data set of primary GBMs and glioma sphere cultures (GSCs) and report the global landscape of in-frame gene fusions in human GBM.

RESULTS

Nomination of candidate GBM genes

We reasoned that integration of somatic point mutations and focal CNVs would uncover candidate driver GBM genes. MutComFocal is

[1]Institute for Cancer Genetics, Columbia University Medical Center, New York, New York, USA. [2]Department of Biomedical Informatics, Columbia University Medical Center, New York, New York, USA. [3]Department of Systems Biology, Columbia University Medical Center, New York, New York, USA. [4]Department of Control and Computer Engineering, Politecnico di Torino, Torino, Italy. [5]The Preston Robert Tisch Brain Tumor Center, Duke University Medical Center, Durham, North Carolina, USA. [6]The Pediatric Brain Tumor Foundation Institute, Duke University Medical Center, Durham, North Carolina, USA. [7]Department of Pathology, Duke University Medical Center, Durham, North Carolina, USA. [8]Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada. [9]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, New York, USA. [10]Fondazione Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Istituto Neurologico C. Besta, Milan, Italy. [11]Department of Neurosurgery, Columbia University Medical Center, New York, New York, USA. [12]Department of Pathology, Columbia University Medical Center, New York, New York, USA. [13]Department of Pathology, MD Anderson Cancer Center, Houston, Texas, USA. [14]Department of Neurology, Henry Ford Health System, Detroit, Michigan, USA. [15]Department of Neurosurgery, Henry Ford Health System, Detroit, Michigan, USA. [16]Princess Margaret Cancer Center, University Health Network, Toronto, Ontario, Canada. [17]Department of Pediatrics, Columbia University Medical Center, New York, New York, USA. [18]Department of Neurology, Columbia University Medical Center, New York, New York, USA. [19]Present address: Neuroscience and Brain Technologies, Italian Institute of Technology, Genoa, Italy. [20]These authors contributed equally to this work. Correspondence should be addressed to A.L. (al2179@columbia.edu), R.R. (rabadan@dbmi.columbia.edu) or A.I. (ai2102@columbia.edu).

Received 18 June; accepted 29 July; published online 5 August 2013; doi:10.1038/ng.2734

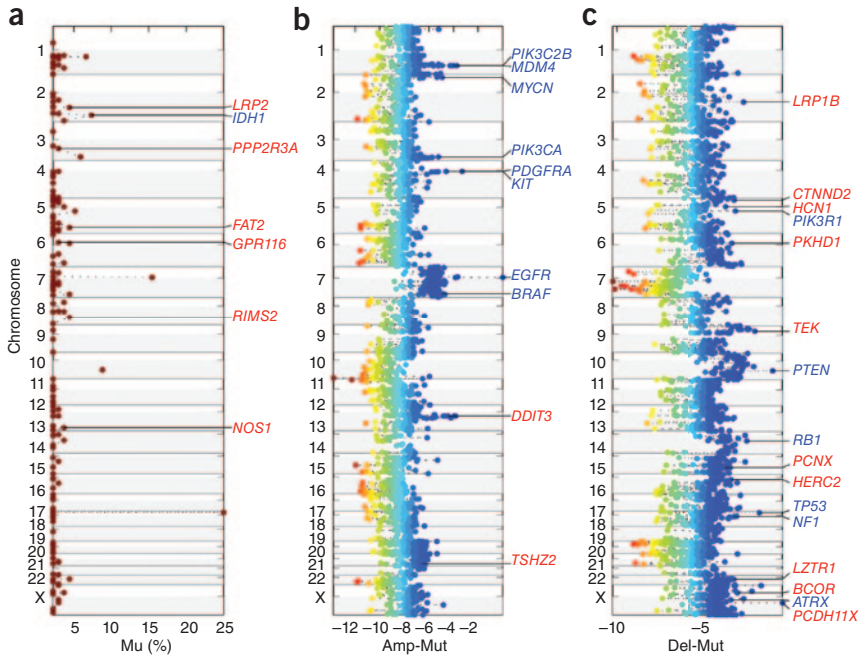


Figure 1 Chromosome view of validated GBM genes scoring at the top of each of the three categories by MutComFocal. (**a**) Mutated genes without significant copy number alterations (shown are the frequency of mutations as percentages). (**b**) Mutated genes in regions of focal and recurrent amplifications (shown are amplification/mutation scores (Online Methods)). (**c**) Mutated genes in regions of focal and recurrent deletions (shown are deletion/mutation scores (Online Methods)). Previously known GBM genes are shown in blue. New and independently validated GBM genes are shown in red. In **b** and **c**, the genes scores are colored according to their corresponding tier, with blue corresponding to high tiers and red corresponding to low tiers.

an algorithm that is designed to rank genes by an integrated recurrence, focality and mutation score (Online Methods). We applied this strategy to samples from The Cancer Genome Atlas (TCGA), 139 GBMs and matched normal DNA analyzed by whole-exome sequencing to identify somatic mutations and 469 GBMs analyzed by the Affymetrix SNP 6.0 platform to identify CNVs.

The whole-exome analysis identified a mean of 43 nonsynonymous somatic mutations per tumor sample (**Supplementary Tables 1** and **2**). The distribution of substitutions shows a higher rate of transitions as compared to transversions (67%), with a strong preference for C>T and G>A transitions (55%) (**Supplementary Fig. 1**). As has been seen in other tumor types[10], 19.2% of the mutations occurred in a CpG dinucleotide context (**Supplementary Fig. 2**). Among somatic small nucleotide variants, the most frequently mutated genes have well-established roles in cancer, including GBM (*TP53*, *EGFR*, *PTEN* and *IDH1*; **Supplementary Table 3**). In addition to known cancer genes, potentially new candidate driver genes were mutated in ~5% of the tumor samples. By integrating mutational and common focal genomic lesions, MutComFocal (**Supplementary Table 4**) stratified somatically mutated genes into three groups: recurrently mutated genes without copy-number alterations (Mut), genes in regions of focal and recurrent amplifications (Amp-Mut) and genes in regions of focal and recurrent deletions (Del-Mut). We generated a list of 67 genes that scored at the top of each of the three categories and included nearly all the genes that have been previously implicated in GBM (**Supplementary Table 4**). Among these genes (labeled in blue in **Fig. 1**) are *IDH1* (Mut; **Fig. 1a**), *PIK3C2B*, *MDM4*, *MYCN*, *PIK3CA*, *PDGFRA*, *KIT*, *EGFR*, *BRAF* (Amp-Mut; **Fig. 1b**) *PIK3R1*, *PTEN*, *RB1*, *TP53*, *NF1* and *ATRX* (Del-Mut; **Fig. 1c**). The analysis also selected 52 new candidate driver genes that were previously unreported in GBM. On the basis of their role in central nervous system development and homeostasis as well as their potential function in gliomagenesis, we selected 24 genes for resequencing in an independent data set of 83 GBMs and matched normal controls. Using Sanger sequencing, we found 18 genes that were somatically mutated in the independent panel (labeled in red in **Fig. 1** and **Supplementary Table 5**). Each validated new GBM gene is targeted by somatic mutations and CNVs in a cumulative fraction comprising between 2.9% and 45.7% of the GBMs (**Supplementary Table 4**). Furthermore, mutations of the 18 new GBM genes occur mostly in tumors with global mutation rates that are similar to the mean rate of 43 mutations per tumor and are well within the 95% confidence interval, indicating that mutations of the 18 new genes do not cluster in hypermutated tumors (**Supplementary Fig. 3**).

Among the commonly mutated and focally deleted genes with the top MutComFocal scores that we validated in the independent GBM data set, we found *BCOR*, *LRP* family members, *HERC2*, *LZTR1* and *CTNND2*. *BCOR*, an X-linked gene, encodes a component of the nuclear co-repressor complex that is essential for normal development of the neuroectoderm and normal stem cell functions[11–13]. *BCOR* mutations have recently been described in retinoblastoma and medulloblastoma[14,15]. *LRP1B*, encoding a member of the LDL receptor family, is among the most frequently mutated genes in human cancer (**Fig. 1c**)[16]. Notably, two other LDL receptor family members (*LRP2* and *LRP1*) are mutated in 4.4% and 2.9% of tumors, respectively (**Fig. 1a** and **Supplementary Table 1**). The LRP proteins are highly expressed in the neuroepithelium and are essential for forebrain morphogenesis in mice and humans[17,18]. The tumor-suppressor function of LRP proteins in GBM may relate to their ability to promote chemosensitivity and control of the Sonic hedgehog signaling pathway, which has been implicated in cancer-initiating cells in GBM[19–21]. Localized on chromosome 15q13, the HECT ubiquitin ligase gene *HERC2* is deleted and mutated in 15.1% and 2.2% of GBMs, respectively (**Supplementary Table 4**). HERC2 has been implicated in severe neurodevelopmental syndromes, and HERC2 substrates regulate genome stability and DNA damage repair[22,23].

LZTR1 mutations drive self renewal of glioma spheres

LZTR1 received one of the highest Del-Mut scores by MutComFocal (**Fig. 1c** and **Supplementary Table 4**). The *LZTR1* coding region had nonsynonymous mutations in 4.4% of GBMs, and the *LZTR1* locus (chromosome 22q11) was deleted in 22.4% of GBMs. Among the 18 new GBM genes, *LZTR1* had the highest co-occurrence score of mutations and deletions ($P = 0.0007$, Fisher's exact test; **Supplementary Table 6**). It also scored at the top of the list of genes whose CNVs were statistically correlated with expression (the Pearson correlation between *LZTR1* CNVs and expression was 0.36, $P < 10^{-6}$ by Student's t distribution; **Supplementary Table 7**). In addition, *LZTR1* emerged as the gene with the highest correlation for monoallelic expression of mutant alleles in tumors harboring *LZTR1* deletions ($P = 0.0007$; **Supplementary Table 8**). Taken together, these findings indicate that *LZTR1* is concurrently

© 2013 Nature America, Inc. All rights reserved.

targeted in GBM by mutations and copy number loss, fulfilling the two-hit model for tumor-suppressor inactivation in cancer.

LZTR1 codes for a protein with a characteristic kelch-BTB-BACK-BTB-BACK domain architecture (**Supplementary Figs. 4** and **5**) and is expressed in normal brain (**Supplementary Table 9**). *LZTR1* is highly conserved in metazoans. Although it was initially proposed that LZTR1 functions as a transcriptional regulator, this role was not confirmed in follow-up studies[24]. Most proteins with BTB-BACK domains are substrate adaptors in cullin 3 (CUL3) ubiquitin ligase complexes, in which the BTB-BACK region binds to the N-terminal domain of CUL3 and a ligand-binding domain, often a kelch six-bladed β-propeller motif, binds to substrates that have been targeted for ubiquitylation[25]. To determine whether LZTR1 directly binds CUL3, we performed coimmunoprecipitation experiments in human glioma cells. We found that CUL3 immunoprecipitates contain LZTR1 (**Fig. 2a**), suggesting that LZTR1 is an adaptor in CUL3 ubiquitin ligase complexes.

To address the function of LZTR1 mutants, we built a homology model of LZTR1 based partly on the crystal structures of the MATH-BTB-BACK protein SPOP[26], the BTB-BACK-kelch proteins KLHL3 (ref. 27) and KLHL11 (ref. 28) and the kelch domain of KEAP1 (ref. 29) (**Fig. 2b**). We predict that the second BTB-BACK region of LZTR1 binds CUL3 because of a ϕ-X-E motif in this BTB domain, followed by a 3-Box/BACK region (**Supplementary Fig. 5a**)[26]. However, we cannot exclude the possibility that the preceding BTB-BACK region also participates in CUL3 binding. Five of seven *LZTR1* mutations that we identified in GBM are located within the kelch domain and target highly conserved amino acids (**Fig. 2b** and **Supplementary Figs. 4** and **5b**). Notably, the concentration of *LZTR1* mutations in the kelch domain reflects a similar pattern of mutations in the kelch-coding region of *KLHL3*, which was recently identified in families with hypertension and electrolytic abnormalities[30,31]. The p.Arg198Gly and p.Gly248Arg alterations localize to the b-c loop of the kelch domain in a region that is predicted to provide the substrate-binding surface[29]. The p.Trp105Arg alteration targets a highly conserved anchor residue in the kelch repeats, and the p.Thr288Ile alteration disrupts a buried residue that is conserved in LZTR1 (**Fig. 2b** and **Supplementary Figs. 4** and **5b**). Both alterations are expected to perturb the folding of the kelch domain. The p.Glu353* alteration is expected to produce a misfolded kelch domain, in addition to removing the C-terminal BTB-BACK regions. Located in the BTB-BACK domains, the remaining two alterations either truncate the entire BTB-BACK-BTB-BACK region (p.Trp437*) or are predicted to disrupt the folding of the last helical hairpin in the BTB-BACK domain (p.Arg810Trp; **Fig. 2b**).

To determine whether the mutations that are predicted to affect the BTB-BACK domains perturb the interaction with CUL3, we prepared *in vitro*–translated wild-type, p.Glu353*, p.Trp437* and p.Arg810Trp MYC-tagged LZTR1 proteins and tested their ability to bind to Flag-CUL3 purified from mammalian cells. Wild-type LZTR1 bound Flag-CUL3, but the p.Glu353* and p.Trp437* mutants lost this property. However, the p.Arg810Trp mutant retained CUL3 binding in this assay (**Fig. 2c**). Besides promoting the ubiquitin-mediated degradation of substrates, cullin adaptors are short-lived proteins that undergo autoubiquitylation and destruction by the same cullin complexes that direct substrate ubiquitylation[32–34]. Thus, impaired ubiquitin ligase activity of the LZTR1-CUL3 complex should result in the accumulation of mutant LZTR1 proteins. Each of the three LZTR1 mutants

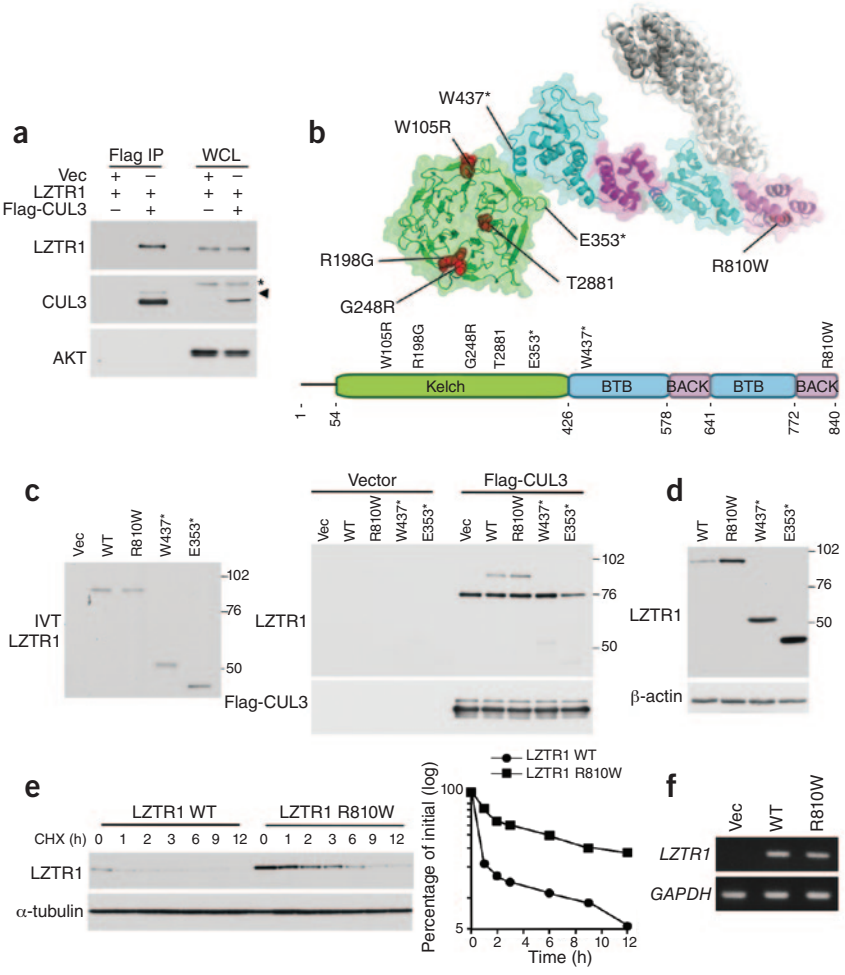


Figure 2 Interaction with CUL3 and protein stability of wild-type and mutant LZTR1. (**a**) Lysates from SF188 glioma cells transfected with vectors expressing MYC-LZTR1 and Flag-CUL3 or the empty vector were immunoprecipitated with Flag antibody and assayed by protein blot using the indicated antibodies. The asterisk indicates a nonspecific band, and the arrowhead indicates neddylated CUL3. Vec, empty vector; WCL, whole-cell lysates; IP, immunoprecipitation. (**b**) Localization of altered residues in LZTR1. Homology model of the kelch (green), BTB (cyan) and BACK (magenta) domains of LZTR1, with the CUL3 N-terminal domain (gray) docked onto the putative binding site. GBM alterations are indicated in red. The amino acids are referred to by their one-letter abbreviations (for example, W437* represents p.Trp437*). (**c**) *In vitro* analysis of the interaction between CUL3 and wild-type and GBM-related mutants of LZTR1. Left, *in vitro*–translated (IVT) MYC-LZTR1 input. Right, *in vitro*–translated MYC-LZTR1 was mixed with Flag-CUL3 that was immunoprecipitated from transfected HEK-293T cells. Bound proteins were analyzed by protein blot using the indicated antibodies. (**d**) Steady-state protein amounts of wild-type LZTR1 and GBM-related mutants. (**e**) Left, cells transfected with wild-type LZTR1 or the p.Arg810Trp mutant were treated with cycloheximide (CHX) for the indicated time. Right, quantification of wild-type LZTR1 and p.Arg810Trp mutant protein from the experiment shown to the left. (**f**) Semiquantitative RT-PCR evaluation of the expression of wild-type and mutant (p.Arg810Trp) *LZTR1* RNA in the cells transfected as in **e**.

© 2013 Nature America, Inc. All rights reserved.

that are predicted to compromise the integrity of the BTB-BACK domains had greater accumulation than wild-type LZTR1 in transient transfection assays (**Fig. 2d**). The steady state and half-life of the LZTR1 p.Arg810Trp mutant protein were markedly increased, although there were no changes in the mutant mRNA (**Fig. 2e,f**). Thus, as the two truncated mutants, the p.Arg810Trp alteration compromised protein degradation.

Next we sought to establish the biological consequences of LZTR1 inactivation in human GBM. The differential gene expression pattern of GBMs harboring mutations and deletions of *LZTR1* or normal *LZTR1* showed that tumors with genetic inactivation of *LZTR1* were enriched for genes associated with glioma sphere growth and proliferation[35] (**Fig. 3a**). Introduction of LZTR1 into three independent GBM-derived sphere cultures resulted in strong inhibition of glioma sphere formation and the expression of glioma stem cell markers (**Fig. 3b–e**). LZTR1 also decreased the size of the tumor spheres, induced a flat and adherent phenotype and reduced the abundance of proteins associated with cell-cycle progression (cyclin A, PLK1 and p107; **Fig. 3d,e**). Notably, both the p.Arg810Trp and p.Trp437* alterations abolished the ability of LZTR1 to impair glioma sphere formation (**Fig. 3f**). These experiments indicate that *LZTR1* inactivation in human GBM drives the self renewal and growth of glioma spheres.

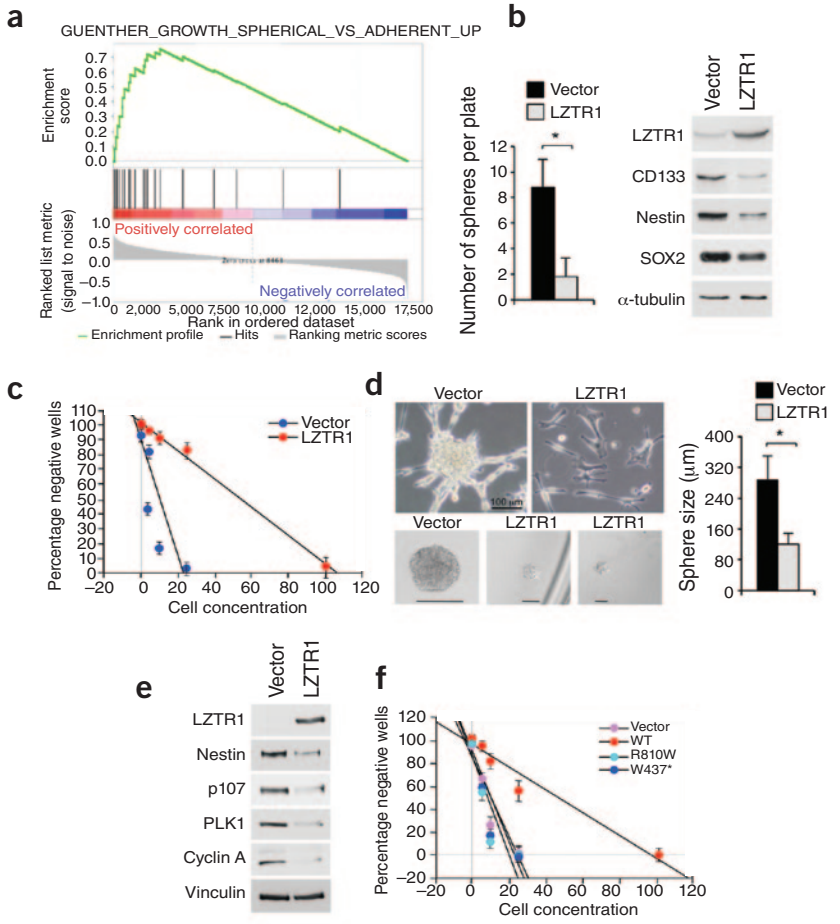


Figure 3 Functional analysis of wild-type LZTR1 and GBM-associated mutants in GBM-derived cells. (**a**) Gene Set Enrichment Analysis (GSEA) showing upregulation of genes associated with the phenotype 'spherical cultures' of glioma cells in primary human GBMs carrying mutations in *LZTR1* (enrichment score = 0.754; *P* (family-wise error rate) = 0.000; *q* (false discovery rate, FDR) = 0.000). (**b**) Sphere-forming assay (left) and protein blot analysis (right) of GBM-derived glioma spheres (GSC #48) expressing empty vector or LZTR1. The data are shown as the mean ± s.d. of triplicate samples (*$P = 0.0036$, *t* test). (**c**) Linear regression plot of an *in vitro* limiting dilution assay using GBM-derived glioma spheres from GSC #46 expressing vector or LZTR1. The frequencies of sphere-forming cells were 8.49 ± 1.04% and 1.44 ± 0.05% in vector- and LZTR1-expressing cells, respectively ($P = 0.00795$, *t* test). Each data point represents the average of triplicates. Error bars, s.d. (**d**) Top left, bright-field microphotographs of GBM-derived line #46 cells 6 d after transduction with vector- or LZTR1-expressing lentivirus. Bottom left, bright-field microphotographs of spheres from GBM-derived glioma cells from GSC #46 transduced with lentivirus expressing vector or LZTR1, as in **c**. Right, the sizes of tumor spheres from the cultures in **c** were determined by microscopy review after 14 d of culture. $n = 60$ spheres from triplicates for each condition. The data are shown as the mean ± s.d. (*$P < 0.0001$, *t* test). (**e**) Protein blot analysis of GBM-derived cells from GSC #84 expressing vector or LZTR1. (**f**) Linear regression plot of an *in vitro* limiting dilution assay using GBM-derived line #84 expressing vector, LZTR1, LZTR1 p.Arg810Trp or LZTR1 p.Trp437*. The frequencies of sphere-forming cells were 7.2 ± 0.92% for vector, 1.48 ± 0.09% for wild-type LZTR1 ($P = 0.0096$), 7.82 ± 0.99% for LZTR1 p.Arg810Trp ($P = 0.2489$) and 6.74 ± 1.07% for LZTR1 p.Trp437* ($P = 0.2269$). Error bars, s.d. Statistical analyses in **e** were performed by *t* test.

CTNND2 inactivation induces mesenchymal transformation

Among the top-ranking genes in the MutComFocal analysis, *CTNND2* is expressed at the highest levels in normal brain (**Supplementary Table 9**). *CTNND2* encodes δ-catenin, a member of the p120 subfamily of catenins that are expressed almost exclusively in the nervous system, where δ-catenin is crucial for neurite elongation, dendritic morphogenesis and synaptic plasticity[36–38]. Germline hemizygous loss of *CTNND2* impairs cognitive functions and underlies some forms of intellectual disability[39,40]. *CTNND2* shows a pronounced clustering of mutations in GBM. The observed spectrum of mutations includes four in the armadillo-coding domain and one in the region coding for the N-terminal coiled-coil domain (**Supplementary Fig. 6a**), which are the two most relevant functional domains of δ-catenin. Each mutation targets highly conserved residues with probably (p.Lys629Gln, p.Ala776Thr, p.Ser881Leu and p.Asp999Glu) and possibly (p.Ala71Thr) damaging consequences[41]. GBMs harbor focal genomic losses of *CTNND2*, and deletions correlate with loss of *CTNND2* expression (**Supplementary Fig. 6b** and **Supplementary Table 7**).

Immunostaining experiments showed that δ-catenin is strongly expressed in the normal brain, particularly in neurons, as demonstrated by co-staining with the neuronal markers β3-tubulin and MAP2 but not the astrocytic marker GFAP (**Fig. 4a,b**). Conversely, immunostaining of 69 GBMs and protein blotting of nine GSCs showed negligible or absent expression of δ-catenin in 21 tumors and in most of the GSCs (**Supplementary Fig. 6c,d**). Oncogenic transformation in the central nervous system frequently disrupts the default proneural cell fate and induces an aberrant mesenchymal phenotype that is associated with an aggressive clinical outcome[42]. Gene expression analysis of 498 GBMs from TCGA showed that low *CTNND2* expression

© 2013 Nature America, Inc. All rights reserved.

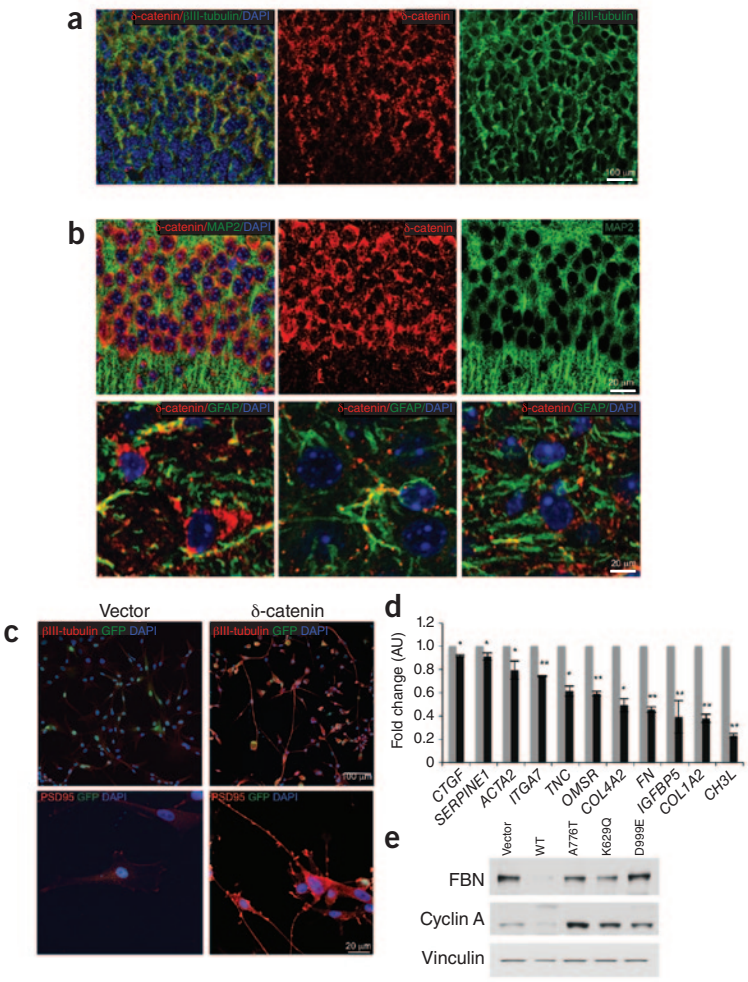


Figure 4 Expression of δ-catenin in neurons and δ-catenin–driven loss of mesenchymal markers in GBM. (**a**) Pattern of expression of δ-catenin in the developing brain, as determined by immunostaining. Double immunofluorescence staining of the brain cortex was performed using δ-catenin antibody (red) and βIII-tubulin (green); nuclei are counterstained with 4′,6-diamidino-2-phenylindole (DAPI) (blue). (**b**) Pattern of expression of δ-catenin in the adult brain, as determined by immunostaining. Top, double immunofluorescence staining of brain cortex was performed using δ-catenin antibody (red) and MAP2 (green); nuclei are counterstained with DAPI (blue). Bottom, double immunofluorescence staining of brain cortex was performed using δ-catenin antibody (red) and GFAP (green); nuclei are counterstained with DAPI (blue). (**c**) Immunofluorescence staining for βIII-tubulin (top) and PSD95 (bottom) in U87 cells expressing δ-catenin or empty vector. (**d**) Expression of mesenchymal genes in glioma cells expressing δ-catenin or empty vector (averages of triplicate quantitative RT-PCRs). Error bars, s.d. $^{*}P < 0.005$, $^{**}P < 0.001$, *t* test. AU, arbitrary units. (**e**) Protein blot using the indicated antibodies in U87 cells expressing wild-type δ-catenin, glioma-associated δ-catenin mutants or empty vector. FBN, fibronectin. Vinculin is shown as a loading control.

is strongly enriched in tumors showing the mesenchymal gene expression signature ($P = 2.4 \times 10^{-12}$, *t* test; **Supplementary Fig. 7a**). Tumors with reduced *CTNND2* expression were characterized by poor clinical outcome, among which tumors with *CTNND2* copy number loss showed the worst prognosis (**Supplementary Fig. 7b,c**). Tumors with low *CTNND2* expression showed the worst clinical outcome in mesenchymal GBM, although nonmesenchymal tumors also showed poor prognosis, albeit with reduced strength (**Supplementary Fig. 7d**). Mesenchymal transformation of GBM is associated with irreversible loss of proneural cell fate and neuronal markers[42] and is detected in most established glioma cell lines. Expression of δ-catenin in the U87 human glioma cell line reduced cell proliferation (**Supplementary Fig. 8a**), elevated the expression of the neuronal proteins βIII-tubulin, PSD95 (a postsynaptic marker) and N-cadherin (**Fig. 4c** and **Supplementary Fig. 8b**) and decreased mRNA and protein levels of mesenchymal markers (**Fig. 4d,e** and **Supplementary Fig. 8b**). These effects were associated with morphologic changes that were characterized by neurite extension and the development of branched dendritic processes (**Fig. 4c** and **Supplementary Fig. 8c,d**). Conversely, expression of the p.Ala776Thr, p.Lys629Gln and p.Asp999Glu mutants of CTNND2 did not induce neuronal features or downregulate the mesenchymal marker fibronectin (**Fig. 4e** and **Supplementary Fig. 8c,d**). Consistent with δ-catenin inhibition of cell proliferation in glioma cells, only wild-type δ-catenin decreased the expression of cyclin A, an S-phase cyclin (**Fig. 4e**). Next we analyzed the effect of expressing δ-catenin in GBM-derived sphere culture #48, which lacks endogenous δ-catenin protein (**Supplementary Fig. 6d**) and expresses high levels of mesenchymal markers[43]. Introduction of δ-catenin into sphere culture #48 strongly reduced the abundance of the mesenchymal proteins smooth muscle actin (SMA), collagen-5A1 (COL5A1) and fibronectin, as measured by quantitative immunofluorescence (**Fig. 5a,b**). It also induced βIII-tubulin more than eightfold (**Fig. 5c,d**). A time-course analysis showed that the highest degree of βIII-tubulin–positive neurite extension was present at 4–6 d after transduction, which was then followed by progressive depletion of neuronal-like cells from culture (**Fig. 5d**). We then asked whether δ-catenin affects the self renewal and growth of glioma spheres *in vitro* or their ability to grow as tumor masses *in vivo*. In a limiting dilution assay, δ-catenin inhibited glioma sphere formation more than eightfold (**Fig. 5e**). To determine the effect of δ-catenin on brain tumorigenesis *in vivo*, we generated #48 GSCs expressing luciferase and conducted bioluminescence imaging at different times after stereotactic transduction of control and δ-catenin–expressing cells in mouse brains. When compared to controls, we observed a fivefold inhibition of tumor growth by δ-catenin at each time point analyzed (**Fig. 5f** and **Supplementary Fig. 8e**). These results identify *CTNND2* inactivation as a key genetic alteration driving the aggressive mesenchymal phenotype of GBM.

Recurrent EGFR fusions in GBM

To identify gene fusions in GBM, we analyzed RNA-seq data from a total of 185 GBM samples (161 primary GBMs from TCGA plus 24 short-term GSCs freshly isolated in our laboratory from individuals carrying primary GBMs). The analysis of the RNA-seq data led to the discovery of 92 candidate rearrangements that give rise to in-frame fusion transcripts (**Supplementary Table 10**). Besides the previously reported *FGFR3-TACC3* fusions events, the most frequent recurrent in-frame fusions involved *EGFR*, which were present in 7.6% of the samples (14/185, 95% confidence interval (CI) 3.8–11.3%). Nine of 14 *EGFR*

© 2013 Nature America, Inc. All rights reserved.

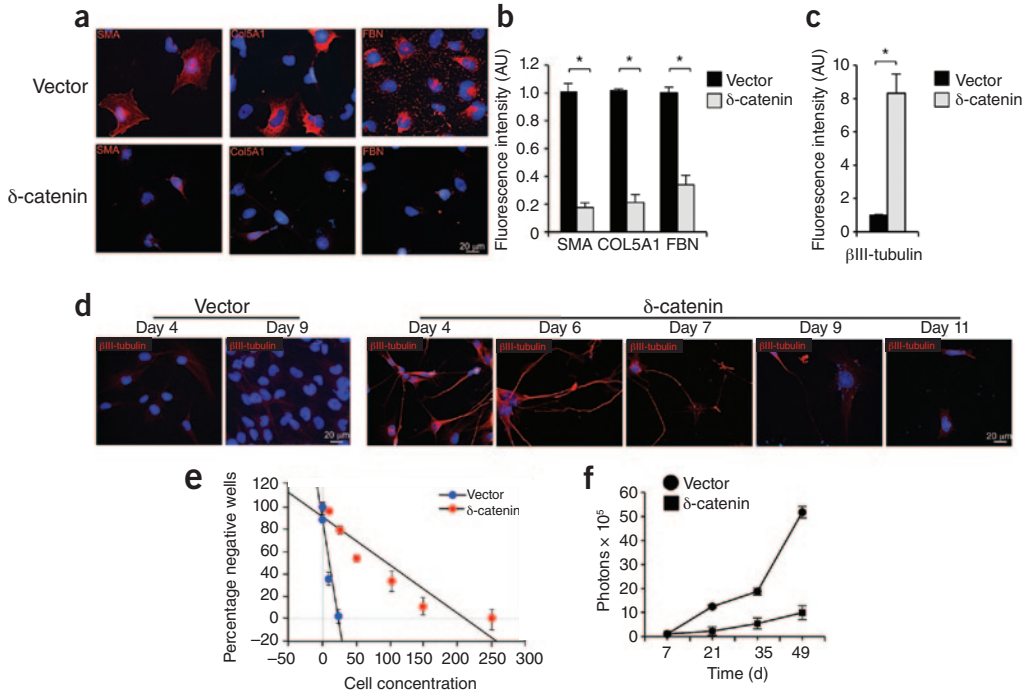


Figure 5 Functional analysis of δ-catenin in mesenchymal GBM. (**a**) Immunofluorescence for FBN, collagen-5α1 (COL5A1) and SMA in glioma spheres from GSC #48 4 d after infection with lentiviruses expressing δ-catenin or empty vector. Nuclei are counterstained with DAPI. (**b**) Quantification of the fluorescence intensity of SMA, COL5A1 and FBN in cultures treated as in **a**. $n = 3$ independent experiments; data are shown as the mean ± s.d. $^*P < 0.05$, *t* test. (**c**) Quantification of fluorescence intensity for βIII-tubulin in cells from GSC #48 infected with lentiviruses expressing *CTNND2* or empty vector. Error bars, s.d. $^*P < 0.05$, *t* test. (**d**) Time-course analysis of βIII-tubulin expression in glioma spheres from GSC #48 transduced with lentiviruses expressing *CTNND2* or empty vector. There was a loss from the advanced culture of βIII-tubulin–expressing cells. (**e**) Linear regression plot of an *in vitro* limiting dilution assay using GBM-derived cells from GSC #48 expressing empty vector or δ-catenin. The frequencies of sphere-forming cells were 7.42 ± 1.16% and 0.88 ± 0.02% (percentage of sphere-forming cells) for vector and δ-catenin, respectively ($P = 0.0098$, *t* test). Error bars, s.d. (**f**) Longitudinal analysis of bioluminescence imaging in mice injected intracranially with GBM-derived line #48 expressing vector or δ-catenin. $n = 3$ mice for vector; $n = 5$ mice for δ-catenin. Symbols are the mean, and bars are the s.e.m. of the photon counts.

fusions included the recurrent partner *SEPT14* (6/185, 3.2%) or *PSPH* (3/185, 1.6%) as the 3′ gene segment in the fusion. All *EGFR-SEPT14* and two of three *EGFR-PSPH* gene fusions occurred within amplified regions of the fusion genes (**Supplementary Fig. 9**). The quantitative analysis of expressed reads spanning the fusion breakpoint compared to reads spanning *EGFR* exons not implicated in the fusion transcripts showed that the *EGFR* fusion genes were expressed at higher levels in five of nine tumors (**Supplementary Table 11**). We also found two highly expressed in-frame fusions involving *NTRK1*, the neurotrophic tyrosine kinase receptor 1 gene, as the 3′ gene with two different 5′ partners (*NFASC-NTRK1* and *BCAN-NTRK1*). Fusions involving *NTRK1* are common in papillary thyroid carcinomas[44]. Using EXomeFuse, an algorithm that reconstructs genomic fusions from whole-exome data, we confirmed that the *EGFR-SEPT14* and *NTRK1* fusions result from recurrent chromosomal translocations and reconstructed the corresponding genomic breakpoints (**Supplementary Table 12**).

The sequence of the PCR products spanning the fusion breakpoint validated all three types of recurrent in-frame fusion predictions (*EGFR-SEPT14*, *EGFR-PSPH* and *NTRK1* fusions; **Fig. 6** and **Supplementary Figs. 10** and **11**). In **Figure 6a,b**, we show the prediction and cDNA sequence validation, respectively, for one tumor harboring an *EGFR-SEPT14* fusion (TCGA-27-1837). The amplified cDNA contained an ORF for a 1,041-residue that resulted from the fusion of EGFR residues 1–982 with SEPT14 residues 373–432 (**Fig. 6c**). Thus, the structure of EGFR-SEPT14 fusions involve EGFR at the N terminus, providing a receptor tyrosine kinase domain that is fused to a coiled-coil domain from SEPT14. Exon-specific RNA-seq expression in TCGA-27-1837 demonstrated that the *EGFR* and *SEPT14* exons that are implicated in the fusion are highly expressed as compared to mRNA sequences that are not included in the fusion event (**Supplementary Fig. 12**). Using PCR, we mapped the genomic breakpoint to chromosome 7 (55,268,937 for *EGFR* and 55,870,909 for *SEPT14*, genome build GRCh37/hg19) within *EGFR* exon 25 and *SEPT14* intron 9, creating a transcript in which the 5′ *EGFR* exon 24 is spliced to the 3′ *SEPT14* exon 10 (**Fig. 6d**). Notably, the fused *EGFR-PSPH* cDNA and predicted fusion protein in sample TCGA-06-5408 involved the same *EGFR* N-terminal region that was implicated in *EGFR-SEPT14*, with *PSPH* providing a 35-residue C-terminal portion (**Supplementary Fig. 10**). An example of a fusion in which the EGFR-TK region is the 3′ partner is the *CAND1-EGFR* fusion in GSC #16 (**Supplementary Fig. 13**). Each fusion transcript includes the region of the *EGFR* mRNA coding for the TK domain (**Supplementary Table 10**). RT-PCR and genomic PCR followed by Sanger sequencing of GBM TCGA-06-5411 validated the *NFASC-NTRK1* fusion in which the predicted fusion protein includes the TK domain of the high-affinity NGF receptor (TRKA, also called NTRK1) fused downstream to the immunoglobulin-like region of the cell-adhesion and ankyrin-binding region of neurofascin (**Supplementary Fig. 11**).

To confirm that GBM harbors recurrent *EGFR* fusions and to determine their frequency in an independent data set, we screened cDNA from a panel of 248 GBMs and discovered ten additional cases with *EGFR-SEPT14* fusions (4%). Conversely, we did not find *NFASC-NTRK1* fusions in this data set. We also determined the frequency of *EGFR-PSPH* fusions to be 2.2% (3/135).

The discovery of recurrent *EGFR* fusions in GBM is of particular interest. EGFR is activated in a fraction of primary GBMs (~25%) by an in-frame deletion of exons 2–7 (EGFRvIII)[45]. However, seven of nine tumors harboring *EGFR-SEPT14* and *EGFR-PSPH* gene fusions lacked the EGFRvIII rearrangement (**Supplementary Table 13**). We sought to determine whether the most frequent *EGFR* fusion in GBM (*EGFR-SEPT14*) provides an alternative mechanism of EGFR activation and confers sensitivity to EGFR inhibition. First we asked whether *EGFR* gene fusions cluster into any gene expression subtype of GBM (proneural,

© 2013 Nature America, Inc. All rights reserved.

neural, classical or mesenchymal). Although no individual subtype showed a statistically significant enrichment of *EGFR* fusions, eight of nine GBMs harboring *EGFR-SEPT14* or *EGFR-PSPH* belonged to the classical and mesenchymal subtype (Fisher's $P = 0.05$ for classical or mesenchymal enrichment; **Supplementary Table 14**). Next we compared the effects of ectopic *EGFR-SEPT14*, EGFRvIII or wild-type *EGFR* on glioma cells. Lentiviral transduction of human GSC #48 (which lacks genomic alteration of *EGFR*) showed that cells expressing *EGFR-SEPT14* or EGFRvIII but not those expressing wild-type *EGFR* or an empty vector retained growth and self renewal in the absence of EGF and basic FGF (bFGF) (**Fig. 7a**). Accordingly, established glioma cell lines expressing *EGFR-SEPT14* or EGFRvIII proliferated at a higher rate than control cells or cells expressing wild-type *EGFR* (**Supplementary Fig. 14a,b**). Furthermore, *EGFR-SEPT14* and EGFRvIII markedly enhanced the migration of glioma cells in a wound assay (**Supplementary Fig. 14c**). These findings suggest that *EGFR-SEPT14* might constitutively activate signaling events downstream of EGFR. When analyzed in the presence or absence of mitogens, the expression of *EGFR-SEPT14* (or EGFRvIII) in GSC #48 triggered constitutive activation of phosphorylated STAT3 (pSTAT3) but had no effects on pERK or pAKT (**Fig. 7b,c**). Differential gene expression analysis identified a set of nine genes upregulated in *EGFR-SEPT14*–positive tumors compared with EGFRvIII-positive GBMs (**Supplementary Fig. 15**). These genes broadly relate to the inflammatory and immune response, and some code for chemokines that have been associated with aggressive glioma phenotypes (CXCL9, CXCL10 and CXCL11)[46].

We then sought to investigate whether *EGFR-SEPT14* fusions confer sensitivity to inhibition of EGFR-TK. Treatment of GSC #48 expressing *EGFR-SEPT14*, EGFRvIII, wild-type *EGFR* or vector control with lapatinib, an irreversible EGFR inhibitor that was recently proposed to target EGFR alterations in GBM[47], showed that *EGFR-SEPT14* and EGFRvIII but not wild-type *EGFR* sensitized glioma cells to pharmaceutical EGFR inhibition (**Fig. 7d**). We obtained similar effects after treatment of derivatives of GSC #48 with erlotinib, another inhibitor of EGFR-TK (**Fig. 7e**). To determine whether sensitivity to EGFR-TK inhibition is retained in human glioma cells naturally harboring *EGFR-SEPT14 in vivo*, we used an *EGFR-SEPT14*–positive GBM xenograft (D08-0537 MG) that was established from a heavily pretreated patient. Treatment of D08-0537 MG tumors with lapatinib or erlotinib showed that both drugs significantly ($P = 0.003$ for lapatinib, $P = 0.037$ for erlotinib) delayed tumor growth, with lapatinib showing the strongest antitumor effects. Conversely, EGFR inhibitors were ineffective against GBM xenograft D08-0714 MG, which lacks *EGFR* genomic alterations (**Fig. 7f** and **Supplementary Fig. 14d**). Taken together, these data show that *EGFR-SEPT14* fusions confer mitogen-independent growth, constitutively activate STAT3 signaling and impart sensitivity to EGFR kinase inhibition to glioma cells harboring the fusion gene.

DISCUSSION

We describe a computational pipeline that computes the frequency, magnitude and focality of CNVs at any locus in the human genome with the somatic mutation rate for genes residing at that genomic location, thus integrating two genetic hallmarks of driver cancer genes (focality of CNVs and point mutations) into a single score. Besides recognizing nearly all genes that are known to have functional relevance in GBM, we discovered and validated somatic mutations in 18 new genes that also harbor focal and recurrent CNVs in a fraction of GBMs. The importance of some of these genes extends beyond GBM, as is underscored by cross-tumor relevance (for example, *BCOR*) and protein family recurrence (for example, *LRP* family members). Also, the *LZTR1* mutations that target highly conserved residues in the kelch domain (Trp105, Gly248 and Thr288) and in the second BTB-BACK domain (Arg810) are recurrent events in other tumor types[48]. Thus, understanding the nature of the substrates of LZTR1-CUL3 ubiquitin ligase activity will provide important insights into the pathogenesis of multiple cancer types. The importance of *LZTR1* genetic alterations in GBM is underscored by concurrent targeting of *LZTR1* by mutations and deletions, supporting a two-hit mechanism of tumor-suppressor gene inactivation, as well as the impact of mutations targeting the BTB-BACK domains on CUL3 binding and/or protein stability and their ability to release glioma cells

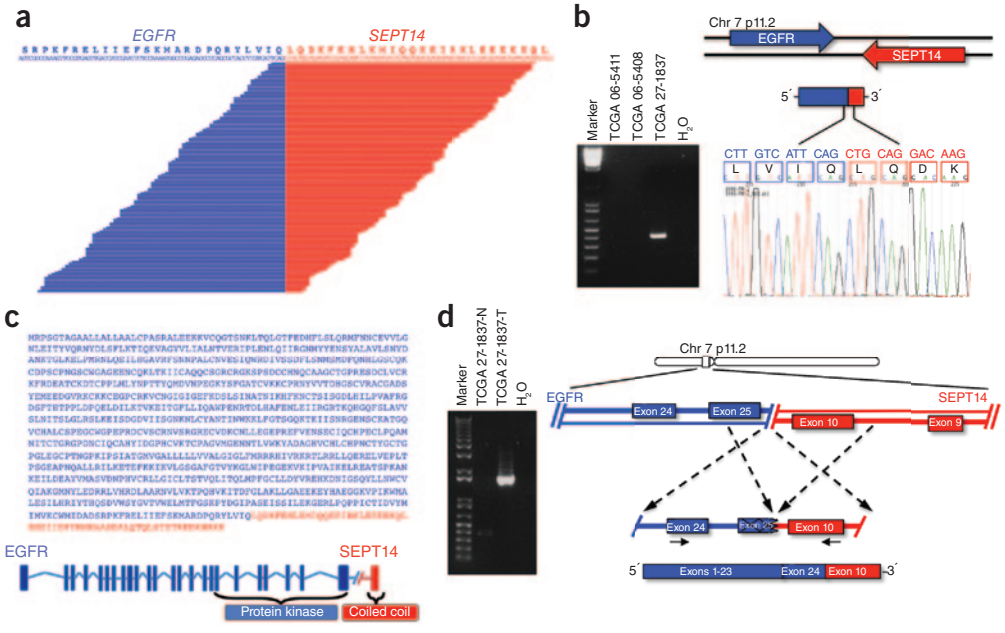
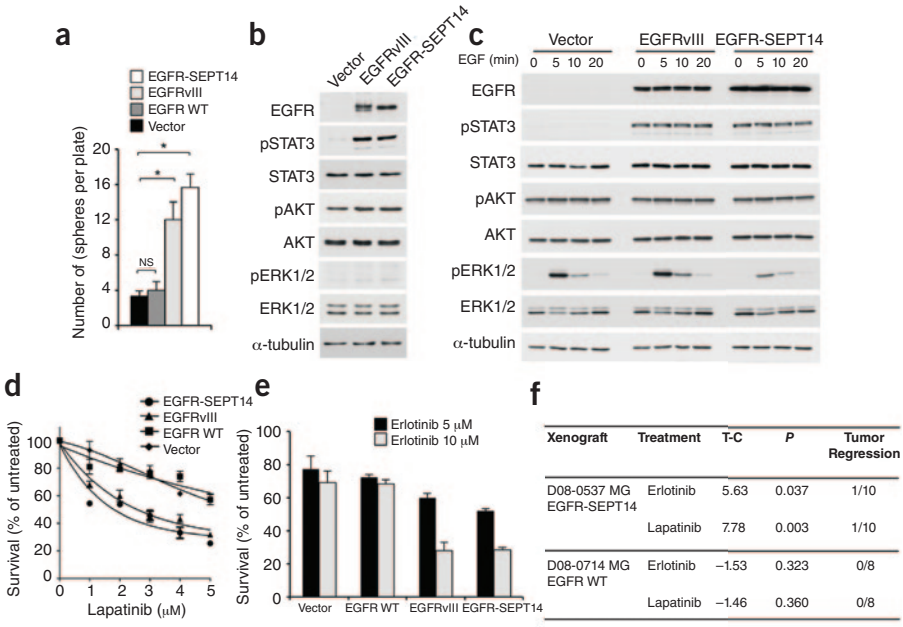


Figure 6 The *EGFR-SEPT14* gene fusion identified by whole-transcriptome sequencing. (**a**) Split reads are shown aligning on the breakpoint. The predicted reading frame at the breakpoint is shown at the top, with *EGFR* sequences in blue and *SEPT14* in red. (**b**) Left, *EGFR-SEPT14*–specific PCR from cDNA derived from GBMs. Marker, 1-kb ladder. Right, Sanger sequencing chromatogram showing the reading frame at the breakpoint and putative translation of the fusion protein in the positive sample. (**c**) EGFR-SEPT14 fusion protein sequence and schematics. Regions corresponding to EGFR and SEPT14 are shown in blue and red, respectively. The fusion joins the tyrosine kinase domain of EGFR and the coiled-coil domain of SEPT14. (**d**) Genomic fusion of *EGFR* exon 25 with intron 9 of *SEPT14*. In the fused mRNA, exon 24 of *EGFR* is spliced 5′ to exon 10 of *SEPT14*. Solid arrows indicate the position of the fusion genome primers that generate a fusion-specific PCR product in the GBM sample TCGA-27-1837.

© 2013 Nature America, Inc. All rights reserved.



f

Xenograft	Treatment	T-C	*P*	Tumor Regression
D08-0537 MG EGFR-SEPT14	Erlotinib	5.63	0.037	1/10
	Lapatinib	7.78	0.003	1/10
D08-0714 MG EGFR WT	Erlotinib	–1.53	0.323	0/8
	Lapatinib	–1.46	0.360	0/8

Figure 7 Functional analysis of the EGFR-SEPT14 fusion and the effect of inhibition of EGFR kinase on glioma growth. (**a**) Sphere-forming assay in the absence of EGF in GBM-derived primary cells (GSC #48) expressing vector, wild-type EGFR, EGFRvIII or the EGFR-SEPT14 fusion. The data are shown as the mean ± s.d. of triplicate samples (**P* = 0.0051 and **P* = 0.027 for cells expressing the EGFR-SEPT14 fusion or EGFRvIII compared with those expressing vector, respectively, *t* test). NS, not significant. (**b**) Protein blot analysis of GBM-derived primary cells (GSC #48) expressing vector, EGFRvIII or the EGFR-SEPT14 fusion cultured in the presence of EGF. ERK1/2, ERK1 and ERK2. (**c**) GBM-derived cells (GSC #48) expressing vector, EGFRvIII or the EGFR-SEPT14 fusion were cultured in the absence of EGF for 48 h and then stimulated with 20 ng/ml EGF for the indicated times. Cells were assayed by protein blot analysis using the indicated antibodies. (**d**) Survival of GBM-derived cells (GSC #48) expressing vector, wild-type EGFR, EGFRvIII or the EGFR-SEPT14 fusion after treatment with lapatinib for 48 h at the indicated concentrations. The data are shown as the mean ± s.d. of triplicate samples. (**e**) Survival of GBM-derived cells (GSC #48) expressing vector, wild-type EGFR, EGFRvIII or the EGFR-SEPT14 fusion after treatment with erlotinib for 48 h at the indicated concentrations. The data are shown as the mean ± s.d. of triplicate samples. The experiments were repeated three times. (**f**) *In vivo* inhibition of tumor growth by EGFR kinase inhibitors in xenografts derived from patients with gliomas carrying the EGFR-SEPT14 fusion ($n = 10$) but not wild-type EGFR ($n = 8$). T-C indicates the median difference in survival between drug-treated and vehicle-treated (control) mice.

from the restraining activity of the wild-type protein on growth and self renewal.

The finding that loss-of-function mutations in *CTNND2* cluster in mesenchymal GBM provides a clue to the genetic events driving this aggressive GBM subtype. The function of δ-catenin for crucial neuronal morphogenesis indicates that full-blown mesenchymal transformation in the brain requires loss of master regulators that constrain cell determination along the neuronal lineage. Introduction of δ-catenin into human glioma spheres collapsed the mesenchymal phenotype and inhibited sphere formation and tumor growth. Thus, the ability of δ-catenin to reprogram glioma cells expressing mesenchymal genes toward a neuronal fate unravels an unexpected plasticity of mesenchymal GBM that might be exploited therapeutically.

In this study we also report the landscape of gene fusions from a large data set of GBMs analyzed by RNA-seq. In-frame gene fusions retaining the RTK-coding domain of *EGFR* emerged as the most frequent gene fusion in GBM. In this tumor, *EGFR* is frequently targeted by focal amplifications, and our finding underscores the strong recombinogenic probability of focally amplified genes, as was recently reported for the *MYC* locus in medulloblastoma[49]. Resembling intragenic rearrangements that generate the EGFRvIII allele, we found that *EGFR-SEPT14* fusions impart to glioma cells the ability to self renew and grow in the absence of mitogens, constitutively activate STAT3 signaling and confer sensitivity to EGFR inhibition. These findings highlight the relevance of fusions implicating RTK-coding genes in the pathogenesis of GBM[9]. They also provide a strong rationale for the inclusion of patients with GBM who harbor *EGFR* fusions in clinical trials based on EGFR inhibitors.

URLs. DNA and RNA sequencing and copy number variant data in TCGA, http://cancergenome.nih.gov; glioma patient survival data from the Repository for Molecular Brain Neoplasia Data (REMBRANDT), https://caintegrator.nci.nih.gov/rembrandt/; sequence data deposition in database of Genotypes and Phenotypes (dbGaP), http://www.ncbi.nlm.nih.gov/gap; gene fusion annotation software package Pegasus, http://sourceforge.net/projects/pegasus-fus/; NIH ImageJ software, http://rsb.info.nih.gov/ij/.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. RNA sequencing of 24 human GBM sphere cultures in this study have been deposited under the dbGaP study accession phs000505.v2.p1. We also analyzed RNA and DNA sequencing of TCGA GBM samples from the dbGaP study accession phs000178.v1.p1.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This work was supported by National Cancer Institute grants R01CA101644 and R01CA131126 (A.L.) and R01CA085628 and R01CA127643 (A.I.), the Stewart Foundation (R.R.), the Partnership for Cure (R.R.), US National Institutes of Health (NIH) grant NIH 1 P50 MH094267-01 (R.R.), the Lymphoma Research Foundation (R.R.), NIH 1 U54 CA121852-05 (R.R.), NIH 1R01CA164152-01 (R.R.), the Leukemia and Lymphoma Society (R.R.), the Canadian Cancer Society (G.G.P.), the Cancer Research Society (G.G.P.), the National Institute of Neurological Disorders and Stroke R01NS061776 (A.I.) and a grant from The Chemotherapy Foundation (A.I.). G.F. was supported by grants from the Associazione Italiana per la Ricerca sul Cancro and from the Italian Ministry of Health. V.F., P.Z., C.D. and F.N. are supported by fellowships from the Italian Ministry of Welfare/Provincia di Benevento and the Federazione Italiana Associazioni Genitori Oncoematologia Pediatrica (FIAGOP) (C.D.). We thank J. Parkinson for helpful discussions on the phylogeny of *LZTR1* genes, L. Bertin for help with protein blots, J. Kroll (Tumor Biology Center, Freiburg) for the *LZTR1* plasmids and M. Pagano (New York University) for CUL3 expression plasmids.

AUTHOR CONTRIBUTIONS

A.L., R.R. and A.I. conceived the ideas for this study. R.R. designed and supervised the computational approach, and A.L. and A.I. designed and supervised the

© 2013 Nature America, Inc. All rights reserved.

experimental platform. A.L. performed or assisted in each step of the experimental platform. V.F., A.C., M.L., F.N. and C.D. conducted biological experiments. V.T. performed the MutComFocal analysis. J.M.C. and F.A. performed the gene fusion analysis, allele-specific expression and most of the bioinformatics analyses. P.Z. performed bioinformatics and statistical analyses. S.T.K., H.Y., R.E.M. and D.D.B. performed the human glioma xenograft analyses to evaluate the effects of EGFR inhibitors and provided human GBM specimens. A.X.J. and G.G.P. performed the modeling analysis of LZTR1. I.D. and A.H. conducted the targeted sequencing analysis. P.P., S.P., D.J.P., P.C., J.N.B., K.A., G.G., G.F. and T.M. provided tissue materials from study subjects. A.L., R.R. and A.I. wrote the manuscript with contributions from all other authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Porter, K.R., McCarthy, B.J., Freels, S., Kim, Y. & Davis, F.G. Prevalence estimates for primary brain tumors in the United States by age, gender, behavior, and histology. *Neuro. Oncol.* **12**, 520–527 (2010).
2. Stupp, R. *et al.* Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N. Engl. J. Med.* **352**, 987–996 (2005).
3. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
4. Noushmehr, H. *et al.* Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* **17**, 510–522 (2010).
5. Parsons, D.W. *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807–1812 (2008).
6. Verhaak, R.G. *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in *PDGFRA*, *IDH1*, *EGFR*, and *NF1*. *Cancer Cell* **17**, 98–110 (2010).
7. Bass, A.J. *et al.* Genomic sequencing of colorectal adenocarcinomas identifies a recurrent *VTI1A-TCF7L2* fusion. *Nat. Genet.* **43**, 964–968 (2011).
8. Chinnaiyan, A.M. & Palanisamy, N. Chromosomal aberrations in solid tumors. *Prog. Mol. Biol. Transl. Sci.* **95**, 55–94 (2010).
9. Singh, D. *et al.* Transforming fusions of *FGFR* and *TACC* genes in human glioblastoma. *Science* **337**, 1231–1235 (2012).
10. Rubin, A.F. & Green, P. Mutation patterns in cancer genomes. *Proc. Natl. Acad. Sci. USA* **106**, 21766–21770 (2009).
11. Fan, Z. *et al.* BCOR regulates mesenchymal stem cell function by epigenetic mechanisms. *Nat. Cell Biol.* **11**, 1002–1009 (2009).
12. Wamstad, J.A. & Bardwell, V.J. Characterization of Bcor expression in mouse development. *Gene Expr. Patterns* **7**, 550–557 (2007).
13. Wamstad, J.A., Corcoran, C.M., Keating, A.M. & Bardwell, V.J. Role of the transcriptional corepressor Bcor in embryonic stem cell differentiation and early embryonic development. *PLoS ONE* **3**, e2814 (2008).
14. Pugh, T.J. *et al.* Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. *Nature* **488**, 106–110 (2012).
15. Zhang, J. *et al.* A novel retinoblastoma therapy from genomic and epigenetic analyses. *Nature* **481**, 329–334 (2012).
16. Beroukhim, R. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905 (2010).
17. Kantarci, S. *et al.* Mutations in *LRP2*, which encodes the multiligand receptor megalin, cause Donnai-Barrow and facio-oculo-acoustico-renal syndromes. *Nat. Genet.* **39**, 957–959 (2007).
18. Willnow, T.E. *et al.* Defective forebrain development in mice lacking gp330/megalin. *Proc. Natl. Acad. Sci. USA* **93**, 8460–8464 (1996).
19. Christ, A. *et al.* LRP2 is an auxiliary SHH receptor required to condition the forebrain ventral midline for inductive signals. *Dev. Cell* **22**, 268–278 (2012).
20. Cowin, P.A. *et al.* *LRP1B* deletion in high-grade serous ovarian cancers is associated with acquired chemotherapy resistance to liposomal doxorubicin. *Cancer Res.* **72**, 4060–4073 (2012).
21. Lima, F.R. *et al.* Glioblastoma: therapeutic challenges, what lies ahead. *Biochim. Biophys. Acta* **1826**, 338–349 (2012).
22. Bekker-Jensen, S. *et al.* HERC2 coordinates ubiquitin-dependent assembly of DNA repair factors on damaged chromosomes. *Nat. Cell Biol.* **12**, 80–86 (2010).
23. Harlalka, G.V. *et al.* Mutation of *HERC2* causes developmental delay with Angelman-like features. *J. Med. Genet.* **50**, 65–73 (2013).
24. Nacak, T.G., Leptien, K., Fellner, D., Augustin, H.G. & Kroll, J. The BTB-kelch protein LZTR-1 is a novel Golgi protein that is degraded upon induction of apoptosis. *J. Biol. Chem.* **281**, 5065–5071 (2006).
25. Stogios, P.J., Downs, G.S., Jauhal, J.J., Nandra, S.K. & Prive, G.G. Sequence and structural analysis of BTB domain proteins. *Genome Biol.* **6**, R82 (2005).
26. Errington, W.J. *et al.* Adaptor protein self-assembly drives the control of a cullin-RING ubiquitin ligase. *Structure* **20**, 1141–1153 (2012).
27. Ji, A.X. & Prive, G.G. Crystal structure of KLHL3 in complex with Cullin3. *PLoS ONE* **8**, e60445 (2013).
28. Canning, P. *et al.* Structural basis for Cul3 assembly with the BTB-Kelch family of E3 ubiquitin ligases. *J. Biol. Chem.* **288**, 7803–7814 (2013).
29. Lo, S.C., Li, X., Henzl, M.T., Beamer, L.J. & Hannink, M. Structure of the Keap1:Nrf2 interface provides mechanistic insight into Nrf2 signaling. *EMBO J.* **25**, 3605–3617 (2006).
30. Boyden, L.M. *et al.* Mutations in kelch-like 3 and cullin 3 cause hypertension and electrolyte abnormalities. *Nature* **482**, 98–102 (2012).
31. Louis-Dit-Picard, H. *et al. KLHL3* mutations cause familial hyperkalemic hypertension by impairing ion transport in the distal nephron. *Nat. Genet.* **44**, 456–460 (2012).
32. Emanuele, M.J. *et al.* Global identification of modular cullin-RING ligase substrates. *Cell* **147**, 459–474 (2011).
33. Galan, J.M. & Peter, M. Ubiquitin-dependent degradation of multiple F-box proteins by an autocatalytic mechanism. *Proc. Natl. Acad. Sci. USA* **96**, 9124–9129 (1999).
34. Zhang, D.D. *et al.* Ubiquitination of Keap1, a BTB-Kelch substrate adaptor protein for Cul3, targets Keap1 for degradation by a proteasome-independent pathway. *J. Biol. Chem.* **280**, 30091–30099 (2005).
35. Günther, H.S. *et al.* Glioblastoma-derived stem cell–enriched cultures form distinct subgroups according to molecular and phenotypic criteria. *Oncogene* **27**, 2897–2909 (2008).
36. Abu-Elneel, K. *et al.* A δ-catenin signaling pathway leading to dendritic protrusions. *J. Biol. Chem.* **283**, 32781–32791 (2008).
37. Arikkath, J. *et al.* δ-catenin regulates spine and synapse morphogenesis and function in hippocampal neurons during development. *J. Neurosci.* **29**, 5435–5442 (2009).
38. Kosik, K.S., Donahue, C.P., Israely, I., Liu, X. & Ochiishi, T. δ-catenin at the synaptic-adherens junction. *Trends Cell Biol.* **15**, 172–178 (2005).
39. Israely, I. *et al.* Deletion of the neuron-specific protein δ-catenin leads to severe cognitive and synaptic dysfunction. *Curr. Biol.* **14**, 1657–1663 (2004).
40. Jun, G. *et al.* δ-catenin is genetically and biologically associated with cortical cataract and future Alzheimer-related structural and functional brain changes. *PLoS ONE* **7**, e43728 (2012).
41. Hicks, S., Wheeler, D.A., Plon, S.E. & Kimmel, M. Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. *Hum. Mutat.* **32**, 661–668 (2011).
42. Phillips, H.S. *et al.* Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* **9**, 157–173 (2006).
43. Carro, M.S. *et al.* The transcriptional network for mesenchymal transformation of brain tumours. *Nature* **463**, 318–325 (2010).
44. Pierotti, M.A. & Greco, A. Oncogenic rearrangements of the NTRK1/NGF receptor. *Cancer Lett.* **232**, 90–98 (2006).
45. Dunn, G.P. *et al.* Emerging insights into the molecular and cellular basis of glioblastoma. *Genes Dev.* **26**, 756–784 (2012).
46. Liu, C. *et al.* Chemokine receptor CXCR3 promotes growth of glioma. *Carcinogenesis* **32**, 129–137 (2011).
47. Vivanco, I. *et al.* Differential sensitivity of glioma- versus lung cancer-specific EGFR mutations to EGFR kinase inhibitors. *Cancer Discov.* **2**, 458–471 (2012).
48. Forbes, S.A. *et al.* COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Res.* **38**, D652–D657 (2010).
49. Northcott, P.A. *et al.* Subgroup-specific structural variation across 1,000 medulloblastoma genomes. *Nature* **488**, 49–56 (2012).

© 2013 Nature America, Inc. All rights reserved.

ONLINE METHODS

SAVI (statistical algorithm for variant frequency identification). We estimated the frequencies of variant alleles in 139 paired tumor and normal whole-exome samples from TCGA using the SAVI pipeline[50]. The algorithm estimates the frequency of variant alleles by constructing an empirical Bayesian prior for those frequencies, using data from the whole sample, and obtains a posterior distribution and high credibility intervals for each allele[50]. The prior and posterior are distributed over a discrete set of frequencies with a precision of 1% and are connected by a modified binomial likelihood, which allows for some error rate. More precisely, we assume a prior distribution $p(f)$ of the frequency f and a prior for the error e uniform on the interval $[0,E]$ for a fixed $0 \le E \le 1$. We think of the sequencing data at a particular allele as a random experiment producing a string of m (the total depth at the allele) bits with n '1's (the variant depth at the allele). Assuming a binomial likelihood of the data and allowing for bits being misread because of random errors, the posterior probability $P(f)$ of the frequency f is

$$P(f) = \frac{p(f)}{C} \times \frac{1}{b-a} \int_{f}^{f+E-2Ef} x^n (1-x)^{m-n}\, dx$$

where C is a normalization constant. For a particular allele, the value of E is determined by the quality of the nucleotides sequenced at that position as specified by their Phred scores. The SAVI pipeline takes as input the reads produced by the sequencing technology, filters out low-quality reads and maps the rest onto a human reference genome. After mapping, a Bayesian prior for the distribution of allele frequencies for each sample is constructed by an iterative posterior update procedure starting with a uniform prior. To genotype the sample, we use the posterior high-credibility intervals for the frequency of the alleles at each genomic location. Alternatively, combining the Bayesian priors from different samples, we obtain posterior high-credibility intervals for the difference between the samples of the frequencies of each allele. Finally, the statistically significant differences between the tumor and normal samples are reported as somatic variants. The results are shown in **Supplementary Table 1**. To estimate the positive prediction value of SAVI in the TCGA GBM samples, we selected 41 mutations for independent validation by Sanger sequencing. We confirmed 39 of the 41 mutations using Sanger sequencing, resulting in a 0.95 (95% CI 0.83–0.99) validation rate (**Supplementary Table 2**).

Supplementary Table 3 shows the candidate genes ranked by the number of somatic nonsynonymous mutations. A robust fit of the ratio of nonsynonymous to synonymous mutations was generated with a bisquare weighting function. The excess of nonsynonymous alterations was estimated using a Poisson distribution with a mean equal to the product of the ratio from the robust fit and the number of synonymous mutations. Genes in highly polymorphic genomic regions were filtered out based on an independent cohort of normal samples. The list of these regions includes families of genes known to generate false positives in somatic predictions (for example, the *HLA*, *KRT* and *OR* gene families).

MutComFocal. Key cancer genes are often amplified or deleted in chromosomal regions containing many other genes. Point mutations and gene fusions, conversely, provide more specific information about which genes may be implicated in the oncogenic process. We developed MutComFocal, a Bayesian approach that assigns a driver score to each gene by integrating point mutations and CNV data from 469 GBMs (Affymetrix SNP6.0). In general, MutComFocal uses three different strategies. First, the focality component of the score is inversely proportional to the size of the genomic lesion to which a gene belongs and thus prioritizes more focal genomic lesions. Second, the recurrence component of the MutComFocal score is inversely proportional to the total number of genes altered in a sample, which prioritizes samples with a smaller number of altered genes. Third, the mutation component of the score is inversely proportional to the total number of genes mutated in a sample, which achieves the twofold goal of prioritizing mutated genes on one hand and prioritizing samples with a smaller number of mutations on the other.

More specifically, for a particular sample, let $(c_1,N_1),\ldots,(c_k,N_k)$ describe the amplification lesions in that sample so that N_i is the number of genes in the ith lesion and c_i is its copy number change from normal. For a gene belonging to the ith lesion, the amplification recurrence sample score is defined as $(c_1,N_1),\ldots,(c_k,N_k)$, and its amplification focality sample score is defined as $(c_i/\sum_j c_j) \times (1/N_i)$. To obtain the amplification recurrence and focality scores for a particular gene, we sum the corresponding sample scores over all the samples and normalize the result so that each score sums to 1. The deletion and recurrence scores are defined in a similar manner. The mutation score is analogous to a recurrence score in which we assume that mutated genes belong to lesions with only one gene.

The amplification/mutation score is defined as the product of the two amplification scores and the mutation score, whereas the deletion/mutation score is defined as the product of the two deletion scores and the mutation score. The amplification/mutation and deletion/mutation scores are normalized to 1, and for each score, genes are divided into tiers iteratively so that the top 2^X remaining genes are included in the next tier, where H is the entropy of the scores of the remaining genes normalized to 1. On the basis of their tier across the different types of scores, genes are assigned to being either deleted/mutated or amplified/mutated, and genes in the top tiers are grouped into contiguous regions. The top genes in each region are considered manually and selected for further functional validation.

The recurrence and focality scores can be interpreted as the posterior probabilities that a gene is driving the selection of the disease under two different priors, one global and one local in nature. The recurrence score is higher if a gene participates in many samples that do not have too many altered genes, whereas the focality score is higher if the gene participates in many focal lesions. Besides lending strong support to the inference of a gene as a potential driver, the directionality of the copy number alteration (amplification or deletion) informs us of the probable behavior of the candidate gene as an oncogene or tumor suppressor, respectively.

The genes displayed in **Figure 1** were selected on the basis of the MutComFocal ranking (top 250 genes), the size of the minimal region (less than 10 genes) and the frequency of mutations (more than 2% for deletion/mutations and at least 1% for amplification/mutations).

RNA-seq bioinformatics analysis. We analyzed 161 RNA-seq GBM tumor samples from TCGA, a public repository containing large-scale genome sequencing of different cancers, plus 24 patient-derived GSCs. Nine GSC samples reported in previous studies were kept in our analysis to evaluate recurrence[9]. We analyzed the samples using the ChimeraScan[51] algorithm to detect a list of gene fusion candidates. Briefly, ChimeraScan detects those reads that discordantly align to different transcripts of the same reference (split inserts). These reads provide an initial set of putative fusion candidates. The algorithm then realigns the initially unmapped reads to the putative fusion candidates and detects those reads that align across the junction boundary (split reads). These reads provide the genomic coordinates of the breakpoint.

RNA-seq analysis detected a total of 39,329 putative gene fusion events. To focus the experimental analysis on biologically relevant fused transcripts, we applied the Pegasus annotation pipeline (http://sourceforge.net/projects/pegasus-fus/). For each putative fusion, Pegasus reconstructs the entire fusion sequence on the basis of the genomic fusion breakpoint coordinates and gene annotations. Pegasus also annotates the reading frame of the resulting fusion sequences as either in frame or a frame shift. Moreover, Pegasus detects the protein domains that are either conserved or lost in the new chimeric event by predicting the amino acid sequence and automatically querying the UniProt web service. On the basis of the Pegasus annotation report, we selected relevant gene fusions for further experimental validation according to the reading frame and the conserved and lost domains. The selected list (**Supplementary Table 7**) was based on in-frame events expressed by ten or more reads, with at least one read spanning the breaking point. To filter out candidate *trans*-splicing events, we focused on events with putative breakpoints at a distance of at least 25 kb.

Identification of genetic rearrangements using whole-exome data. Although whole-exome sequencing data contain low intronic coverage that reduces the sensitivity for fusion discovery, they are readily available through the TCGA database. To characterize the genomic breakpoint of the chromosomal rearrangement, we therefore designed EXome-Fuse, a new gene fusion discovery pipeline that is designed particularly to analyze whole-exome data. For the samples harboring *EGFR-SEPT14*, *EGFR-PSPH*, *NFASC-NTRK1* and *BCAN-NTRK1* fusions in RNA, we applied EXome-Fuse to the corresponding whole-exome sequencing data

© 2013 Nature America, Inc. All rights reserved.
npg

deposited in TCGA. This algorithm can be divided into three stages: split-insert identification, split-read identification and virtual reference alignment. Mapping against the human genome reference hg18 with BWA, we first identify all split inserts to compile a preliminary list of fusion candidates. We prune this list of any false positives produced from paralogous gene pairs using the Duplicated Genes Database and the EnsemblCompara GeneTrees[52]. Pseudogenes in the candidate list were annotated using the list from the HUGO Gene Nomenclature Committee (HGNC) database[53] and were given lower priority. We also filtered out candidates between homologous genes, as well as those with homologous or low-complexity regions around the breakpoint. For the remaining fusion candidates, we probed for any supporting split reads and their mates using BLAST with a word size of 16, identity cutoff of 90% and an expectation cutoff of 10^{-4}. We then create a virtual reference for each fusion transcript and realign all reads to calculate a final tally of split inserts and split reads such that all aligning read pairs maintain forward-reverse directionality.

Targeted exon sequencing. All protein-coding exons for the 24 genes of interest were sequenced using genomic DNA extracted from frozen tumors and matched blood. Five-hundred nanograms of DNA from each sample were sheared to an average size of 150 bp in a Covaris instrument for 360 s (duty cycle, 10%; intensity, 5; cycles per burst, 200). Bar-coded libraries were prepared using the Kapa High-Throughput Library Preparation Kit Standard (Kapa Biosystems). Libraries were amplified using the KAPA HiFi Library Amplification kit (Kapa Biosystems) (eight cycles). Libraries were quantified using Qubit Fluorimetric Quantitation (Invitrogen), and the quality and size was assessed using an Agilent Bioanalyzer. An equimolar pool of the four bar-coded libraries (300 ng each) was created, and 1,200 ng was input to exon capture using one reaction tube of the custom Nimblegen SeqCap EZ (Roche) with custom probes targeting the coding exons of the 38 genes. Capture by hybridization was performed according to the manufacturer's protocols with the following modifications: 1 nmol of a pool of blocker oligonucleotides (complementary to the bar-coded adapters) was used, and post-capture PCR amplification was done using the KAPA HiFi Library Amplification kit, instead of the Phusion High-Fidelity PCR Master Mix with HF Buffer Kit, in a 60 µl volume, as we found that the Kapa HiFi kit greatly reduced or eliminated the bias against GC-rich regions. The pooled capture library was quantified by Qubit (Invitrogen) and Bioanalyzer (Agilent) and sequenced in on an Illumina MiSeq sequencer using the 2 × 150 paired-end cycle protocol. Reads were aligned to the hg19 build of the human genome using BWA with duplicate removal using SAMtools as implemented by Illumina MiSeq Reporter. Variant detection was performed using GATK UnifiedGenotyper. Somatic mutations were identified for paired samples using SomaticSniper and filtered for frequency of less than 3% in normal samples and over 3% in tumor samples. Variants were annotated with the Charity annotator to identify protein-coding changes and cross referenced against known dbSNP, 1000 Genomes and COSMIC variants. Sanger sequencing was used to confirm each mutation from normal and tumor DNA. The complete list of Sanger-validated somatic mutations is reported in **Supplementary Table 5**.

Enrichment of amplified and deleted genes for single-nucleotide variants (SNVs). Although MutComFocal combines SNV and CNV data to identify genes driving oncogenesis, it does not explicitly determine whether amplified or deleted genes are enriched for SNVs within the same sample. Deletions and SNVs of a gene within the same sample might indicate a two-hit model of a tumor suppressor. Alternatively, amplifications and gain-of-function mutations of an oncogene within the sample might further promote oncogenesis. For each MutComFocal candidate gene, we determined the number of TCGA samples with both amplification and SNVs, amplification alone, SNVs alone or neither. We calculated the corresponding Fisher's *P* value. We performed a similar analysis for deletions.

Correlation between copy number and expression. One method of assessing the functional relevance of an amplified or deleted gene is to assess the effect of gene dosage. For each gene nominated by MutComFocal, we calculated the Pearson's correlation coefficient between copy number and expression. The corresponding *P* values were computed using paired Student's *t* test.

Allele-specific expression of SNVs. For a given gene nominated by MutComFocal, RNA sequencing can determine whether the mutant or wild-type allele is expressed. Toward this end, we applied VCFtools[54] to the TCGA BAM RNA-seq files produced by TopHat, which produces the depth of reads calling the reference (R) and variant (V) allele. A measure of relative expression of the variant allele is then V/(V + R). For each mutation, we calculated the binomial *P* value of observing more than V out of V + R reads, assuming that it is equally probable for a read to call the variant or reference. We then pooled the binomial *P* values of each mutation using the Stouffer's *Z*-score method to calculate the combined *P* value per gene.

Ruling out passenger mutations in hypermutated samples. To rule out the possibility that MutComFocal candidates tend to be passenger mutations in hypermutated samples, we compared the number of mutations in samples harboring a MutComFocal mutation to the distribution *N* of the number of mutations in each TCGA sample. Because the number of TCGA samples was well above 30, we assumed that *N* was well approximated by the normal distribution and calculated the mean, µ, and s.d., σ. For each MutComFocal mutation, we performed the *Z*-test and found that all mutations failed statistical significance after correction for multiple hypotheses by the Benjamini-Hochberg method.

Determining the presence of EGFRvIII transcripts. To determine the prevalence of EGFRvIII transcripts, we created an in-house script to calculate the number of split inserts and split reads supporting the junction between *EGFR* exons 1 and 8. We considered the EGFRvIII isoform to be expressed if there were more than five split reads or five split inserts in a sample.

Calculating the relative expression of *EGFR* fusions compared to wild-type *EGFR*. To determine the functional relevance of *EGFR-SEPT14* and *EGFR-PSPH* fusions, we determined the relative expression between the fusion and wild-type transcripts within each sample on the basis of BAM files mapped by TopHat and provided by TCGA. As a proxy for expression of the transcript, we calculated the depth of reads covering either a mutant or wild-type junction. In particular, we considered the depth of reads covering the fusion breakpoint of *EGFR-SEPT14* or *EGFR-PSPH* to estimate the expression of the fusion transcript. Because all *EGFR* fusions stereotypically involved exon 24 joined to either *SEPT14* or *PSPH*, we considered the depth of reads covering the junctions between *EGFR* exons 25–26, 26–27 and 27–28 to be a specific gauge of wild-type *EGFR* expression.

Enrichment of the classical and mesenchymal subtype among samples with *EGFR* fusions. To assess whether samples with *EGFR* fusions tended to occur in a particular GBM subtype, we first classified each TCGA GBM sample by expression according to the methods of Verhaak *et al.*[6]. We then tallied the number of classical, mesenchymal, proneural and neural samples with and without *EGFR* gene fusions. We found that the combined class of classical and mesenchymal phenotype was enriched for *EGFR* fusions according to the Fisher's exact test.

Copy number variation in *EGFR* fusions. Gene fusions often arise from genomic instability. Motivated by this observation, we downloaded segmented SNP array data from TCGA and calculated the log2 ratio between the tumor and normal copy numbers. We plotted this along the chromosomal neighborhood of *EGFR*, *SEPT14* and *PSPH* (chr7:55,000,000–56,500,000).

GSEA. To determine the biological impact of *LZTR1* mutations, we used GSEA[55], which is an analytical tool that harnesses expression data to nominate gene sets enriched for a particular phenotype. Having identified TCGA samples with *LZTR1* SNVs, we applied GSEA to the TCGA expression data. We first compared samples with *LZTR1* SNVs against those with wild-type *LZTR1* (excluding *LZTR1* deletions). To assess statistical significance, we randomized our data set by permuting gene sets 500 times and considered only gene sets with an FDR $q < 0.05$.

Differential expression between samples with *EGFR-SEPT14* and EGFRvIII. We also performed in-house differential expression analysis to determine a distinct molecular signature distinguishing the *EGFR-SEPT14* and EGFRvIII phenotypes. Toward this end, we performed a *t* test comparing the expression of the two groups of samples for each gene. Correcting for multiple hypotheses using the Benjamini-Hochberg method, we considered only genes with FDR < 0.05. In addition, we excluded genes with a variance less than the tenth percentile or absolute value lower than two across all samples. These filters left a predictive set of ten genes. We then performed hierarchical clustering on the expression of these ten genes using Euclidean distance and average linkage.

© 2013 Nature America, Inc. All rights reserved.
npg

doi:10.1038/ng.2734

Modeling of LZTR1. Structural templates for the kelch and BTB-BACK regions of human LZTR1 were identified with HHpred[56]. An initial three-dimensional model was generated with the I-TASSER server[57]. The CUL3 N-terminal domain was docked onto the model by superposing the $KLHL3^{BTB\text{-}BACK}/CUL3^{NTD}$ crystal structure[27] onto the second LZTR1 BTB-BACK domain. The model does not include higher quaternary structure, although many BTB domains, and many kelch domains, are known to self associate[25]. The short linkage between the end of the first BACK domain and the beginning of the second BTB domain would seem to preclude an intrachain BTB-BTB pseudo-homodimer, and we expect that LZTR1 self associates and forms higher-order assemblies. Both BACK domains are the shorter, atypical form of the domain and consist of two helical hairpin motifs, as in SPOP[26,58], and not the four-hairpin motif seen in most BTB-BACK-kelch proteins[28,58]. The model from the kelch domain predicts an unusual 1+3 velcro arrangement[59], with the N-terminal region contributing strand d of blade 1 and the C-terminal region contributing strands a, b and c of the same blade, although we cannot rule out an alternative 2+2 velcro model.

Cell culture. U87 cells were obtained from ATCC. SNB19, U87 and HEK-293T cells were cultured in DMEM supplemented with 10% fetal bovine serum (FBS). Growth rates were determined by plating cells in six-well plates at 3 d after infection with the lentivirus indicated in the figure legends. The number of viable cells was determined by Trypan blue exclusion in triplicate cultures obtained from triplicate independent infections. For the wound assay testing migration, confluent cells were scratched with a pipette tip and cultured in 0.25% FBS. After 16 h, images were taken using the Olympus IX70 connected to a digital camera. Images were processed using the ImageJ64 software. The area of the cell-free wound was assessed in triplicate samples. Experiments were repeated twice.

GBM-derived primary cultures were grown in DMEM:F12 medium containing N2 and B27 supplements and human recombinant FGF-2 and EGF (50 ng/ml each; Peprotech). For sphere formation, cells were infected with lentiviral particles. Four days later, single cells were plated at density of 1 cells per well in triplicate in low-attachment 96-well plates. The number and the size of spheres were scored after 10–14 d. Limiting dilution assays were performed as described previously[60]. Spheres were dissociated into single cells and plated in low-attachment 96-well plates in 0.2 ml of medium containing growth factors (EGF and FGF-2), except for the EGFR-transduced cells, which were cultured in the absence of EGF. Cultures were left undisturbed for 10 d, and then the percentage of wells not containing spheres for each cell dilution was calculated and plotted against the number of cells per well. Linear regression lines were plotted, and the number of cells required to generate at least one sphere in every well (the stem cell frequency) was calculated. The experiment was repeated twice. Treatment of GBM primary cultures with erlotinib or lapatinib was performed in cells transduced with the pLOC vector, wild-type pLOC-EGFR, EGFRvIII or EGFR-SEPT14 and selected with blasticidin for 5 d. Cells were seeded on 6-cm dishes in the absence of EGF and treated with the indicated drugs at the indicated doses for 48 h. Each treatment group was seeded in triplicate. Absolute viable cell counts were determined by Trypan blue exclusion and counted on a hemocytometer. EGF stimulation of EGFR-transduced primary glioma cells was performed in cells deprived of growth factors for 48 h. Cells were collected at the indicated times and processed for protein blot analysis.

Immunofluorescence. Immunofluorescence staining on normal mouse and human brain and brain tumor tissue microarrays were performed as previously described[43,61,62]. Immunofluorescence microscopy was performed on cells fixed with 4% paraformaldehyde in phosphate buffer. Cells were permeabilized using 0.2% Triton X-100. The antibodies and concentrations used in the immunofluorescence staining are detailed in **Supplementary Table 15**.

Secondary antibodies conjugated to Alexa Fluor 594 (1:300, A11037, Molecular Probes) or Alexa 488 (1:500, A11008, Molecular Probes) were used. DNA was stained with DAPI (Sigma). Fluorescence microscopy was performed on a Nikon A1R MP microscope. Quantification of the fluorescence intensity staining in primary or established glioma cells was performed using NIH ImageJ software (see URLs). A histogram of the intensity of fluorescence of each point of a representative field for each condition was generated. The fluorescence intensity of ten fields from three independent experiments was scored, standardized to the number of cells in the field and divided by the intensity of the vector.

Protein blotting, immunoprecipitation and *in vitro* binding. Protein blot analysis and immunoprecipitation were performed using the antibodies detailed in **Supplementary Table 16**. For the *in vitro* binding between CUL3 and LZTR1, wild-type and mutant LZTR1 were translated *in vitro* using the TNT Quick Coupled Transcription/Translation System (Promega). Flag-CUL3 was immunoprecipitated from transfected HEK-293T cells with Flag-M2 beads (Sigma) using RIPA buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate (DOC), 0.1% SDS, 1 mM phenylmethylsulfonyl fluoride (PMSF), 10 mM NaF, 0.5 M Na_3OV_4 (sodium orthovanadate) and Complete Protease Inhibitor Cocktail, Roche). Binding was performed in 200 mM NaCl plus 0.5% NP-40 for 2 h at 4 °C. Immunocomplexes were analyzed by SDS-PAGE and immunoblot.

Cloning and lentiviral production. The lentiviral expression vectors pLOC-GFP and pLOC-CTNND2 were purchased from Open Biosystems. Full-length *EGFR-SEPT14* cDNA was amplified from tumor sample TCGA-27-1837. Wild-type *EGFR*, EGFRvIII and *EGFR-SEPT14* cDNAs were cloned into the pLOC vector. pCDNA-MYC-Hist-LZTR1 was a kind gift from J. Kroll (Tumor Biology Center, Freiburg, Germany)[24]. pCDNA-Flag-CUL3 was kindly provided by M. Pagano (New York University, New York, USA). Wild-type and mutant cDNAs for *LZTR1* and *CTNND2* obtained by site-directed mutagenesis (QuikChange II, Agilent) were cloned into the pLOC vector. Lentiviral particles were produced using published protocols[43,61–64].

Genomic PCR and RT-PCR. Total RNA was extracted from cells using an RNeasy Mini Kit (QIAGEN) following the manufacturer's instructions. Five-hundred nanograms of total RNA was retrotranscribed using the Superscript III kit (Invitrogen) following the manufacturer's instructions. The cDNAs obtained after the retrotranscription were used as templates for quantitative PCR as described[43,64]. The reaction was performed with a Roche480 thermal cycler using the Absolute Blue QPCR SYBR Green Mix from Thermo Scientific. The relative amount of specific mRNA was normalized to *GAPDH*. Results are presented as the mean ± s.d. of triplicate amplifications. The validation of fusion transcripts was performed using both genomic PCR and RT-PCR with forward and reverse primer combinations designed within the margins of the paired-end read sequences detected by RNA-seq. Expressed fusion transcript variants were subjected to direct sequencing to confirm the sequence and translation frame. The primers used for the screening of gene fusions are detailed in **Supplementary Table 17**. The primers used for genomic detection of gene fusions are listed in **Supplementary Table 18**. Semiquantitative RT-PCR to detect exogenous wild-type MYC-LZTR1 and mutant p.Arg801Trp LZTR1 was performed using the primers listed in **Supplementary Table 19**.

Subcutaneous xenografts and drug treatment. Female athymic mice (*nu/nu* genotype, BALB/c background, 6–8 weeks old) were used for all antitumor studies. All animal procedures conformed to guidelines of the Institutional Animal Care and Use Committee (IACUC) of Duke University (protocol number A107-11-04 (3)) and the US National Institutes of Health. Patient-derived adult human glioblastoma xenografts were maintained at the Duke Brain Tumor Center Biorepository (Institutional Review Board number Pro00007434) Xenografts were excised from host mice under sterile conditions and homogenized with the use of a tissue press and modified tissue cytosieve (Biowhitter Inc.), and tumor homogenate was loaded into a repeating Hamilton syringe (Hamilton, Co.) dispenser. Cells were injected subcutaneously into the right flank of the athymic mouse at an inoculation volume of 50 µl with a 19-gauge needle[65]. Subcutaneous tumors were measured twice weekly with hand-held vernier calipers (Scientific Products). Tumor volumes (V) were calculated with the following formula: $((\text{width})^2 \times (\text{length}))/2 = V\ (\text{mm}^3)$. For the subcutaneous tumor studies, groups of mice randomly selected by tumor volume were treated with EGFR kinase inhibitors when the median tumor volumes were an average of 150 mm^3 and were compared with control animals receiving vehicle (saline). Erlotinib was administered at 100 mg per kg body weight orally once per day for 10 d. Lapatinib was administered at 75 mg per kg body weight orally twice per day for 20 d. Response to treatment was assessed by a delay in tumor growth and tumor regression. Growth delay, expressed as a T-C value, is defined as the difference in days between the median time required for tumors in treated and control animals to reach a volume five times greater than that measured at the start of

© 2013 Nature America, Inc. All rights reserved.
npg

the treatment. Tumor regression is defined as a decrease in tumor volume over two successive measurements. Statistical analysis was performed using a SAS statistical analysis program, the Wilcoxon rank-order test for growth delay and Fisher's exact test for tumor regression.

Intracranial injection. GBM-derived primary cells were first infected with a lentivirus expressing luciferase and subsequently transduced with the pLOC vector or pLOC-CTNND2 lentiviral particles. Intracranial injection was performed in 9-week-old male *nu/nu* mice (Charles River Laboratories) in accordance with guidelines of IACUC Committee. Briefly, $5 \leq 10^5$ cells were resuspended in 2.5 μl of PBS and injected into the caudate putamen using a stereotaxic frame (coordinates relative to the bregma: 0.6 mm anterior; 1.65 mm medium-lateral; 3 mm depth-ventral). Tumor growth was monitored using the IVIS Imaging system. Briefly, mice were anesthetized with 3% isoflurane before intraperitoneal injection of 100 mg per kg body weight n-luciferin (Xenogen). Ten minutes after injection of n-luciferin, images were acquired for 1 min with the Xenogen IVIS system (Xenogen) using Living Image acquisition and analysis software (Xenogen). The bioluminescent signal was expressed in photons per second and displayed as a pseudo-color image representing the spatial distribution of photon counts.

Statistics and survival analyses. The Kaplan-Meier survival analysis for *CTNND2* CNVs and *CTNND2* expression were obtained using the REMBRANDT glioma data set.

50. Tiacci, E. *et al. BRAF* mutations in hairy-cell leukemia. *N. Engl. J. Med.* **364**, 2305–2315 (2011).
51. Iyer, M.K., Chinnaiyan, A.M. & Maher, C.A. ChimeraScan: a tool for identifying chimeric transcription in sequencing data. *Bioinformatics* **27**, 2903–2904 (2011).
52. Vilella, A.J. *et al.* EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.* **19**, 327–335 (2009).
53. Seal, R.L., Gordon, S.M., Lush, M.J., Wright, M.W. & Bruford, E.A. genenames.org: the HGNC resources in 2011. *Nucleic Acids Res.* **39**, D514–D519 (2011).
54. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
55. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
56. Söding, J. Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**, 951–960 (2005).
57. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2010).
58. Zhuang, M. *et al.* Structures of SPOP-substrate complexes: insights into molecular architectures of BTB-Cul3 ubiquitin ligases. *Mol. Cell* **36**, 39–50 (2009).
59. Fülop, V. & Jones, D.T. Beta propellers: structural rigidity and functional diversity. *Curr. Opin. Struct. Biol.* **9**, 715–721 (1999).
60. Tropepe, V. *et al.* Distinct neural stem cells proliferate in response to EGF and FGF in the developing mouse telencephalon. *Dev. Biol.* **208**, 166–188 (1999).
61. Niola, F. *et al.* Id proteins synchronize stemness and anchorage to the niche of neural stem cells. *Nat. Cell Biol.* **14**, 477–487 (2012).
62. Niola, F. *et al.* Mesenchymal high-grade glioma is maintained by the ID-RAP1 axis. *J. Clin. Invest.* **123**, 405–417 (2013).
63. Zhao, X. *et al.* The N-Myc-DLL3 cascade is suppressed by the ubiquitin ligase Huwe1 to inhibit proliferation and promote neurogenesis in the developing brain. *Dev. Cell* **17**, 210–221 (2009).
64. Zhao, X. *et al.* The HECT-domain ubiquitin ligase Huwe1 controls neural differentiation and proliferation by destabilizing the N-Myc oncoprotein. *Nat. Cell Biol.* **10**, 643–653 (2008).
65. Friedman, H.S. *et al.* Experimental chemotherapy of human medulloblastoma cell lines and transplantable xenografts with bifunctional alkylating agents. *Cancer Res.* **48**, 4189–4195 (1988).

© 2013 Nature America, Inc. All rights reserved.

doi:10.1038/ng.2734