The integrated landscape of driver genomic alterations in glioblastoma

Veronique Frattini1,20, Vladimir Trifonov2,3,20, Joseph Minhow Chan2,3,20, Angelica Castano1,20, Marie Lia1,20, Francesco Abate2–4, Stephen T Keir5–7, Alan X Ji8, Pietro Zoppoli9, Francesco Niola1,9, Carla Danussi1, Igor Dolgalev9, Paola Porrati10, Serena Pellegratta10, Adriana Heguy9, Gaurav Gupta11, David J Pisapia12, Peter Canoll12, Jeffrey N Bruce11, Roger E McLendon5–7, Hai Yan5–7, Ken Aldape13, Gaetano Finocchiaro10, Tom Mikkelsen14,15, Gilbert G Prive8,16, Darel D Bigner5–7, Anna Lasorella1,12,17, Raul Rabadaña1,12,18 & Antonio Iavarone1,12,18

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 CTNNNB2 target a neural adhesion 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.

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...
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. BCOR mutations have recently been described in retinoblastoma and medulloblastoma. LZTR1, LRP1B, and CTNND2 have 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 encoding a member of the LDL receptor family, is among the most frequently mutated genes in human cancer (Fig. 1c). 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. 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. 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.

**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
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\(^2^4\). 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 \(\beta\)-propeller motif, binds to substrates that have been targeted for ubiquitylation\(^2^5\). To determine whether LZTR1 directly binds CUL3, we performed commu- noprecipitation 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\(^2^6\), 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 \(\alpha\)-helical motif in this BTB domain, followed by a 3-Box/BACK region (Supplementary Fig. 5a)\(^2^6\). 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 hypertensive and electrolytic abnormalities\(^3^0\). 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\(^2^9\). 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\(^3^2\). 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](image-url)
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 inactivation (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.

**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 β-catenin, a member of the p120 subfamily of catenins that are expressed almost exclusively in the nervous system, where β-catenin is crucial for neurite elongation, dendritic morphogenesis and synaptic plasticity. CTNND2 encodes hemizygous loss of CTNND2 impairs cognitive functions and underlies some forms of intellectual disability. 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 β-catenin. Each mutation targets highly conserved residues with probably (p.Lys629Gln, p.Ala776Thr, p.Ser881Leu and p.Asp999Glu) and possibly (p.Ala71Thr) damaging consequences. 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 β-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 β-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 neuronal cell fate and induces an aberrant mesenchymal phenotype that is associated with an aggressive clinical outcome. Gene expression analysis of 498 GBMs from TCGA showed that low CTNND2 expression...
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 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) (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
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 carcinomas44. 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 (55,268,937–55,870,909 for SEPT14 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,
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-PSRH 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 locality 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 (locality 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 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 CTNNBD2 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 recombinogetic probability of focally amplified genes, as was recently reported for the MYC locus in medulloblastoma. 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. They also provide a strong rationale for the inclusion of patients with GBM who harbor EGFR fusions in clinical trials based on EGFR inhibitors.

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 R01 MH094267-01 (R.R.), the Lymphoma Research Foundation (R.F.), 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.L.) 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., B.Z.; C.D. and F.N.) and from the Italian Ministry of Welfare/Provincia di Benezente and the Federazione Italiana Associazioni Genitori Oncopediatria Pediatria (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
experimental platform. A.L. performed or assisted in each step of the experimental platform. V.F., A.C., M.L., E.N. and C.D. conducted biological experiments. V.F. 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. 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., Jänne, P.A., Mermel, C., educational 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. 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., Jänne, P.A., Mermel, C., educational 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. 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., Jänne, P.A., Mermel, C., educational 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. 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.
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 pipeline50. The algorithm estimates the frequency of variant alleles by constructing an empirical Bayesian prior for those regions, using data from the whole sample, and obtains a posterior distribution and high credibility intervals for each allele50. 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 \( \epsilon \) uniform on the interval \([0, E]\) for a fixed \( 0 \leq E \leq 1 \). We think of the sequencing data at a particular allele as a random experiment producing a string of \( n \) (the total depth at the allele) bits with \( n \times 1 \)’s (the variant allele at the allele). Assuming a binomial likelihood of the data and allowing for error by misread because of random errors, the posterior probability \( P(f) \) of the frequency is

\[
P(f) = \frac{P(f)C}{P(f)C} \times \frac{1}{b - n} \int_{s - n/2}^{s + n/2} f^n x^n(1 - x)^{m-n} dx
\]

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 using 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 \((c_1N_1, ..., c_iN_i)\) describe the amplification lesions in that sample so that \(N_i\) is the number of genes in the \(i\)th lesion and \(c_i\) is its copy number change from normal. For a gene belonging to the \(i\)th lesion, the amplification recurrence sample score is defined as \((\delta_i, ..., 0)\), and its amplification focality sample score is defined as \((0, ..., 0, \delta_i)\), where \(\delta_i = c_i \times \frac{N_i}{\sum \delta_j} \times (1/N)\). 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 \(2^k\) 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 recurrence51. We analyzed the samples using the Chimerascan 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 EGF-SEPT14, EGF-PSPH, NFASC-NTRK1 and BCAN-NTRK1 fusions in RNA, we applied Exome-Fuse to the corresponding whole-exome sequencing data.
Correlation between copy number and expression. The corresponding Fisher’s
P
fication and SNVs, amplification alone, SNVs alone or neither. We calculated the
within the sample might further promote oncogenesis. For each MutComFocal
gene within the same sample might indicate a two-hit model of a tumor suppres-
driving oncogenesis, it does not explicitly determine whether amplified or deleted
(SNVs).
Enrichment of amplified and deleted genes for single-nucleotide variants
list of Sanger-validated somatic mutations is reported in
was used to confirm each mutation from normal and tumor DNA. The complete
against known dbSNP, 1000 Genomes and COSMIC variants. Sanger sequencing
the bias against GC-rich regions. The pooled capture library was quantified by
kit, instead of the Phusion High-Fidelity PCR Master Mix with HF Buffer Kit, in
capture PCR amplification was done using the KAPA HiFi Library Amplification
of the 38 genes. Capture by hybridization was performed according to the manu-
factor’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
High-Throughput Library Preparation Kit Standard (Kapa Biosystems). Libraries
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%; inten-
sity, 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 manu-
factor’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 suppress-
or. 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 ampli-
fication and SNVs, amplification alone, SNVs alone or neither. We calculated the
responding 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 possi-
bility that MutComFocal candidates tend to be passenger mutations in hyper-
mutated samples, we compared the number of mutations in samples harboring a
MutComFocal mutation to the distribution of
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, \( \mu \), and s.d., \( \sigma \). 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 EGFvIII transcripts. To determine the prevalence
of EGFvIII transcripts, we created an in-house script to calculate the number of
split inserts and split reads supporting the junction between EGF exons 1 and
8. We considered the EGFvIII isoform to be expressed if there were more than
five split reads or five split inserts in a sample.
Calculating the relative expression of EGF fusion compared to wild-type
EGF. To determine the functional relevance of EGF-SEPT14 and EGF-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 EGF-SEPT14
or EGF-PSPH to estimate the expression of the fusion transcript. Because all
EGF fusion stereotypically involved exon 24 joined to either SEPT14 or PSPH, we
considered the depth of reads covering the junctions between EGF exons
25–26, 26–27 and 27–28 to be a specific gauge of wild-type EGF expression.
Enrichment of the classical and mesenchymal subtype among samples with
EGF fusion. To assess whether samples with EGF fusion tended to occur
in a particular GBM subtype, we first classified each TCGA GBM sample by expres-
sion according to the methods of Verhaak et al.
5. We then tallied the number of
classical, mesenchymal, proneural and neural samples with and without EGF
gene fusions. We found that the combined class of classical and mesenchymal
phenotype was enriched for EGF fusions according to the Fisher’s exact test.
Copy number variation in EGF fusion. 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 per-
muting gene sets 500 times and considered only gene sets with an FDR
< 0.05.
Differential expression between samples with EGF-SEPT14 and EGFvIII.
We also performed in-house differential expression analysis to determine a distinct
molecular signature distinguishing the EGF-SEPT14 and EGFvIII 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.

NATURE GENETICS
doi:10.1038/ng.2734

© 2013 Nature America, Inc. All rights reserved.
Modeling of LZTR1. Structural templates for the kelch and BTB-BACK regions of human LZTR1 were identified with HHpred55. An initial three-dimensional model was generated with the I-TASSER server37. The CUL3 N-terminal domain was docked onto the model by superposing the KLHL3/BTB-BACK/CUL3NTD crystal structure32 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 associate25,26. 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 SPOP36,58, and not the four-hairpin motif seen in most BTB-BACK-kelch proteins29,58. The model from the kelch domain predicts an unusual 1+3 velcro arrangement29, with the N-terminal region contributing strand d of b 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 culture, cells were infected with lentiviral particles. Four days later, single cells were plated at density of 1 cell 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 previously60. 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 EGF-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 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 Na3OV4 (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-CTNNBD2 were purchased from Open Biosystems. Full-length EGF-SEPT14 cDNA was amplified from tumor sample TCGA-27-1837. Wild-type EGF, EGFvIII and EGF-SEPT14 cDNAs were cloned into the pLOC vector. pCDNA-MYC-Hist-LZTR1 was a kind gift from J. Kroll (Tumor Biology Center, Freiburg, Germany)34. pCDNA-Flag-CUL3 was kindly provided by M. Pagano (New York University, New York, USA). Wild-type and mutant cDNAs for LZTR1 and CTNNBD2 obtained by site-directed mutagenesis (QuickChange II, Agilent) were cloned into the pLOC vector. Lentiviral particles were produced using published protocols43,61–64.

Genomic PCR and RT-PCR. Total RNA was extracted from cells using an RNasey 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 described43,64. 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. Semi-quantitative 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 (BioWhittaker Inc.), and tumor homogenate was loaded into a repeating Hamilton syringe (Hamilton, Co.) and dispensed. Cells were injected subcutaneously into the right flank of the athymic mouse at an inoculation volume of 50 μl with a 19-gauge needle65. Subcutaneous tumors were measured twice weekly with hand-held vernier calipers (Scientific Products). Tumor volumes (V) were calculated with the following formula: \((V = \frac{1}{2} \times (\text{width})^2 \times (\text{length}))\) 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 mm3 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

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 Na3OV4 (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.

Genomic PCR and RT-PCR. Total RNA was extracted from cells using an RNasey 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 described43,64. 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. Semi-quantitative 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 (BioWhittaker Inc.), and tumor homogenate was loaded into a repeating Hamilton syringe (Hamilton, Co.) and dispensed. Cells were injected subcutaneously into the right flank of the athymic mouse at an inoculation volume of 50 μl with a 19-gauge needle65. Subcutaneous tumors were measured twice weekly with hand-held vernier calipers (Scientific Products). Tumor volumes (V) were calculated with the following formula: \(V = \frac{1}{2} \times (\text{width})^2 \times (\text{length})\) 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 mm3 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
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-CTNN2 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 \( \mu \)l 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 CTNN2 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. Filop, 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).