

Integrated genomic characterization of *IDH1*-mutant glioma malignant progression

Hanwen Bai^{1,2}, Akdes Serin Harmancı^{2,3}, E Zeynep Erson-Omay^{2,3}, Jie Li⁴, Süleyman Coşkun^{2,3}, Matthias Simon⁵, Boris Krischek⁶, Koray Özdoğan⁷, S Bülent Omay^{2,3}, Eric A Sorensen⁸, Şevin Turcan⁹, Mehmet Bakırcıoğlu⁷, Geneive Carrión-Grant^{2,3}, Phillip B Murray¹⁰, Victoria E Clark¹⁻³, A Gulhan Ercan-Sencicek^{2,3}, James Knight^{1,11}, Leman Sencar^{2,3}, Selin Altınok^{2,3}, Leon D Kaulen^{2,3}, Burcu Gülez^{2,3}, Marco Timmer⁶, Johannes Schramm⁵, Ketu Mishra-Gorur^{1-3,12,13}, Octavian Henegariu^{1-3,12,13}, Jennifer Moliterno^{2,3}, Angeliki Louvi^{3,12,13}, Timothy A Chan^{9,14}, Stacey L Tannheimer⁸, M Necmettin Pamir⁷, Alexander O Vortmeyer⁴, Kaya Bilguvar^{1,2,11,13}, Katsuhito Yasuno^{2,3} & Murat Günel^{1-3,12,13,15}

Gliomas represent approximately 30% of all central nervous system tumors and 80% of malignant brain tumors¹. To understand the molecular mechanisms underlying the malignant progression of low-grade gliomas with mutations in *IDH1* (encoding isocitrate dehydrogenase 1), we studied paired tumor samples from 41 patients, comparing higher-grade, progressed samples to their lower-grade counterparts. Integrated genomic analyses, including whole-exome sequencing and copy number, gene expression and DNA methylation profiling, demonstrated nonlinear clonal expansion of the original tumors and identified oncogenic pathways driving progression. These include activation of the MYC and RTK-RAS-PI3K pathways and upregulation of the FOXM1- and E2F2-mediated cell cycle transitions, as well as epigenetic silencing of developmental transcription factor genes bound by Polycomb repressive complex 2 in human embryonic stem cells. Our results not only provide mechanistic insight into the genetic and epigenetic mechanisms driving glioma progression but also identify inhibition of the bromodomain and extraterminal (BET) family as a potential therapeutic approach.

Gliomas have traditionally been classified into low-grade versus malignant, high-grade tumors on the basis of cellular density, nuclear atypia, mitotic index and presence of necrosis, defining pathological grades I through IV (ref. 2). Among these, grade IV glioma, or glioblastoma multiforme (GBM), remains one of the deadliest cancers, with median survival times of only 12.2 to 18.2 months³⁻⁵. Although

most GBMs are primary and occur *de novo*, about 20% are secondary, arising through the progression of preexisting lower-grade gliomas of grade II or III (ref. 2). The genetic and epigenetic landscapes of grade II and III gliomas have been extensively studied⁶⁻⁸. Genetically, 80% harbor neomorphic *IDH1* mutations affecting Arg132, which coexist either with *TP53* and *ATRX* mutations along with chromosome 17p loss of heterozygosity (LOH) in astrocytic tumors or with *CIC* and *FUBP1* mutations as well as chromosome 1p and 19q deletions in oligodendroglial tumors⁶⁻⁸. Epigenetically, *IDH1*-mutant gliomas exhibit a CpG island methylator phenotype (CIMP), characterized by increased global DNA methylation⁹. In the absence of effective treatments, most grade II and III gliomas eventually progress to more aggressive GBMs¹⁰.

To uncover the molecular mechanisms underlying the malignant transformation of gliomas, we first sequenced the exomes of 41 *IDH1*-mutant grade II and III gliomas (20 with matching blood DNA samples) obtained at initial diagnosis and their progressed counterparts collected up to 10 years later, reaching 154-fold coverage on average (Supplementary Fig. 1 and Supplementary Tables 1 and 2). We identified a median of 31 and 80 somatic protein-altering mutations in tumors with and without matching normal DNA samples, respectively. Overall, considering all 82 sequenced gliomas, we confirmed frequent mutations in several driver genes previously reported to be mutated in gliomas⁶⁻⁸ (Fig. 1a). We also identified *NOTCH1* and *NOTCH2* mutations, occurring within sequences encoding the EGF-like domains, consistent with inactivating mutations identified in squamous cell carcinomas^{6,11-14} (Fig. 1b). Interestingly, a highly conserved Phe357 residue in EGF-like domain 9 of *NOTCH1*, near

¹Department of Genetics, Yale School of Medicine, New Haven, Connecticut, USA. ²Program in Brain Tumor Research, Yale School of Medicine, New Haven, Connecticut, USA. ³Department of Neurosurgery, Yale School of Medicine, New Haven, Connecticut, USA. ⁴Department of Pathology, Yale School of Medicine, New Haven, Connecticut, USA. ⁵Department of Neurosurgery, University of Bonn Medical School, Bonn, Germany. ⁶Department of General Neurosurgery, University Hospital of Cologne, Cologne, Germany. ⁷Department of Neurosurgery, Acibadem University School of Medicine, Istanbul, Turkey. ⁸Translational Medicine, Biomarkers, Gilead Sciences, Inc., Foster City, California, USA. ⁹Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, New York, USA. ¹⁰Department of Pharmacology, Yale School of Medicine, New Haven, Connecticut, USA. ¹¹Yale Center for Genome Analysis, Yale School of Medicine, Orange, Connecticut, USA. ¹²Department of Neurobiology, Yale School of Medicine, New Haven, Connecticut, USA. ¹³Yale Program on Neurogenetics, Yale School of Medicine, New Haven, Connecticut, USA. ¹⁴Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York, USA. ¹⁵Yale Comprehensive Cancer Center, Yale School of Medicine, New Haven, Connecticut, USA. Correspondence should be addressed to M.G. (murat.gunel@yale.edu)

Received 17 September; accepted 6 November; published online 30 November 2015; doi:10.1038/ng.3457

the calcium-binding domain, was recurrently altered in four patients¹⁵ (Fig. 1b). Additionally, several target genes of NOTCH signaling were downregulated in *NOTCH1*- and *NOTCH2*-mutant tumors (Supplementary Fig. 2). Lastly, we identified new mutations in FAT receptor genes, including *FAT1*, *FAT2* and *FAT3* (Fig. 1a)¹⁶. Besides coding mutations, we also detected copy number alterations (CNAs), including 1p and 19q deletions in oligodendroglial tumors ($q = 5.7 \times$

10^{-16} and 2.0×10^{-12} , respectively, two-sided Fisher's exact test) and copy-neutral LOH at the *TP53* locus on chromosome 17p ($q = 1.4 \times 10^{-7}$) in astrocytic tumors⁶⁻⁸ (Fig. 1c,d and Supplementary Fig. 3). Of note, deletion of the *CDKN2A-CDKN2B* tumor-suppressor locus (9p), despite being the most frequent CNA event, occurring in 44% of all tumors analyzed, did not show any enrichment within a particular glioma subtype ($q = 0.6$) (Fig. 1c,d and Supplementary Fig. 4).

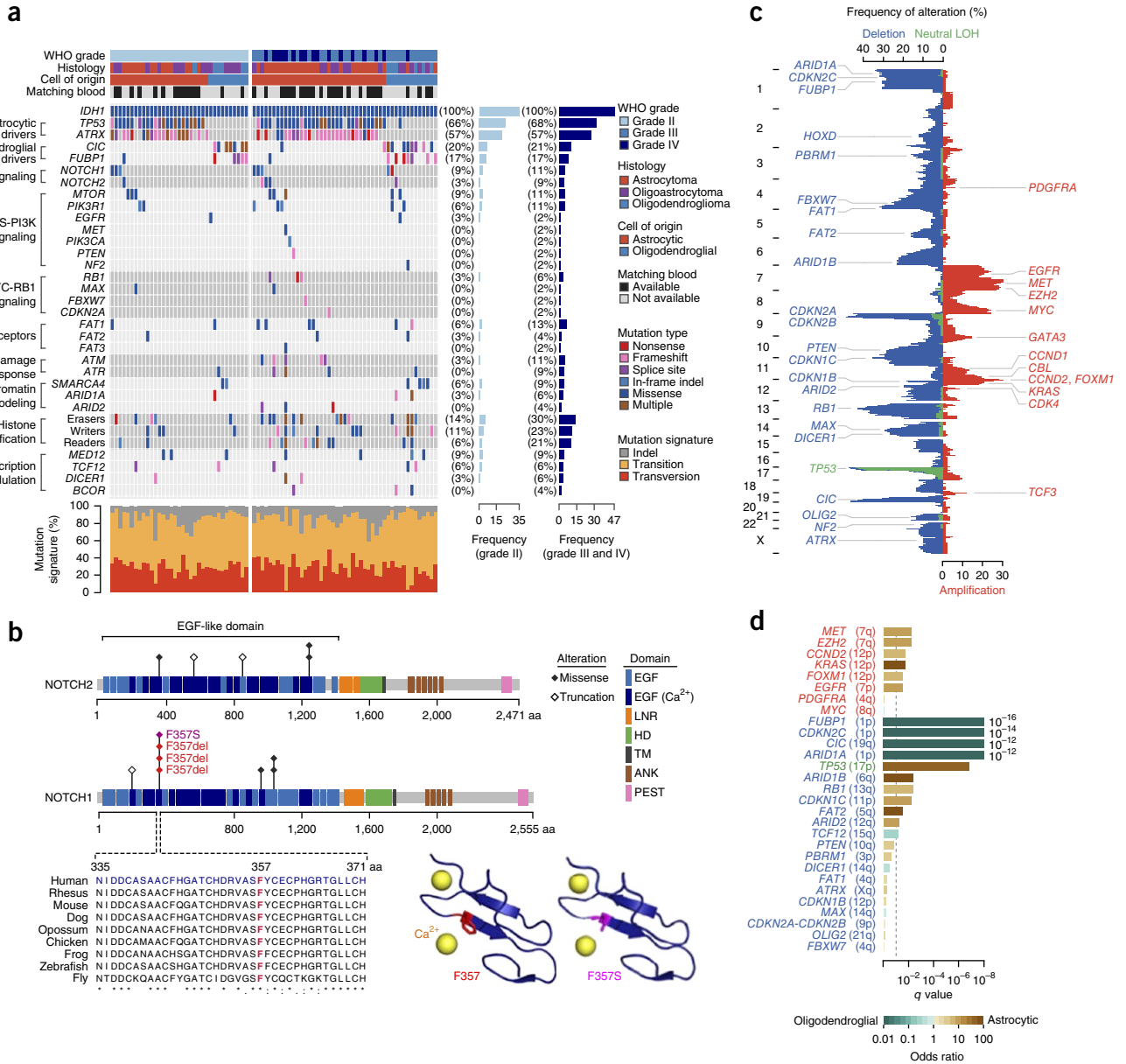


Figure 1 Genomic landscape of *IDH1*-mutant gliomas ($n = 82$ tumors). **(a)** Recurrently mutated genes. For *PIK3R1*, *PIK3CA*, *EGFR*, *MET*, *NF2* and *MAX*, only data for known hotspot mutations are plotted. Recurrently mutated genes, which are grouped on the basis of molecular pathways, are shown on the left. Their mutation frequencies in grade II tumors and in grade III and IV tumors are shown on the right. Each tumor's clinicopathological characteristics are shown at the top, and mutation signatures are indicated at the bottom. **(b)** Alterations encoded by recurrent mutations in the *NOTCH1* and *NOTCH2* genes. Protein domains and mutated residues are color-coded and depicted at the top. At the bottom left, multispecies conservation of the Phe357 mutation hotspot in EGF-like domain 9 is shown. At the bottom right, the Phe357 residue is mapped onto a crystal structure of *NOTCH1* EGF-like domain 13 (Protein Data Bank (PDB), 2VJ3). The wild-type residue is shown in red, and the p.Phe357Ser missense alteration is shown in magenta. EGF, EGF-like domain; EGF (Ca²⁺), Ca²⁺-binding EGF-like domain; LNR, Lin12/Notch repeat; HD, heterodimerization domain; TM, transmembrane domain; ANK, ankyrin repeat; PEST, proline, glutamic acid, serine, threonine-rich degradation motif. **(c)** Overall pattern of CNAs. The vertical axis represents the genome. The horizontal axis indicates the frequency of chromosomal deletions (blue) or copy-neutral LOH events (green), as well as amplifications (red). Consensus cancer-related genes located within corresponding genomic regions are indicated. **(d)** Association of CNAs with cell of origin. The horizontal axis represents association significance, calculated using two-sided Fisher's exact test. The dashed line corresponds to a q value of 0.1. The colored bars represent association odds ratios (green, enriched in oligodendroglial tumors; brown, enriched in astrocytic tumors).

Similarly, amplification of the *MYC* oncogene (8q) (26%; $q = 1.0$) was observed in both subtypes (Fig. 1c,d and Supplementary Fig. 4).

Next, to identify genomic alterations that drive *IDH1*-mutant glioma progression, we compared progressed tumors with their corresponding lower-grade counterparts with respect to mutations and CNAs (Figs. 2 and 3). Excluding three hypermutated tumors, gliomas acquired a median of 21 protein-altering mutations as they progressed (Fig. 2a). Notably, although on average 27% of the mutations and 30% of the CNAs present in the initial tumors were no longer detected during progression, the recurrent *IDH1* mutation affecting Arg132 was never lost during progression and remained clonal in all progressed tumors^{6,17} (Fig. 2 and Supplementary Fig. 5). Mutations acquired during progression also showed significantly different trinucleotide signatures from those that were lost ($P = 0.05$ and 1.0×10^{-5} for non-hypermutated and hypermutated samples, respectively, two-sided χ^2 test) (Supplementary Fig. 6). Furthermore, during progression, 27% of the higher-grade gliomas

acquired different coding mutations in the same glioma driver genes, such as *TP53*, *ATRX*, *CIC* and *FUBP1*, in which mutations were observed in their lower-grade counterparts^{6,17} (Fig. 2c and Supplementary Fig. 5). These findings suggest that genomic heterogeneity, which is known to occur in gliomas, has a substantial role during progression^{6,17}. Although patients who underwent adjunct therapy after initial diagnosis, such as chemotherapy and/or radiation therapy, tended to acquire more mutations upon progression than patients who did not ($P = 2.1 \times 10^{-3}$, two-sided Wilcoxon rank-sum test), the patterns of CNAs remained the same (Supplementary Fig. 7).

During glioma progression, the receptor tyrosine kinase (RTK)-RAS-phosphoinositide 3-kinase (PI3K) pathway, the NOTCH pathway and FAT receptors were most frequently altered (in 22%, 12% and 10% of the patients, respectively) (Fig. 3a). Genes involved in epigenetic patterning, including histone modifiers and components of the SWI/SNF chromatin-remodeling complex (44% and 5%,

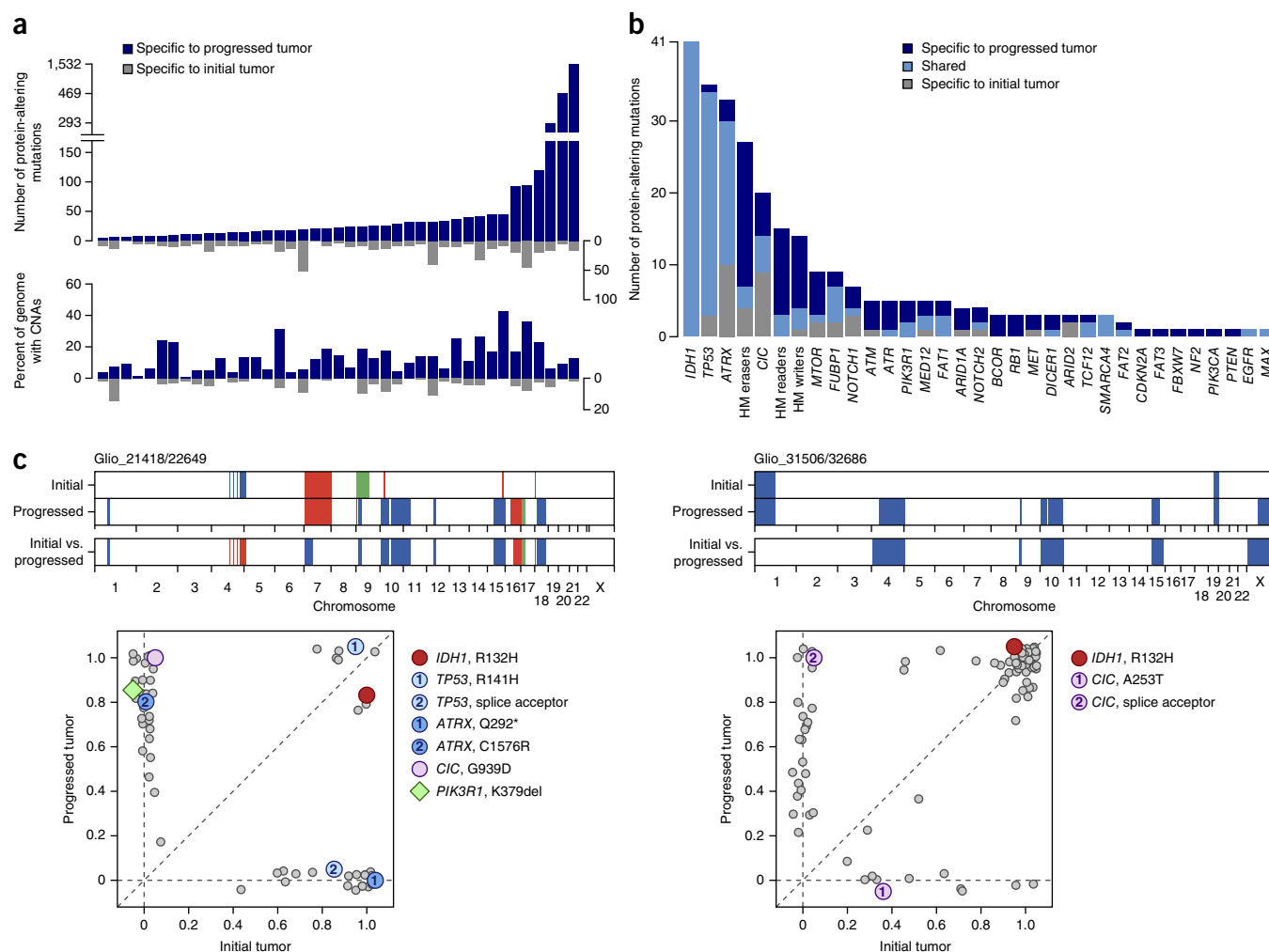


Figure 2 Patterns of nonlinear clonal evolution during glioma progression. **(a)** The number of protein-altering mutations and percentage of the genome with CNAs specific to the initial or progressed tumor are shown for all 41 patients. Alterations specific to progressed or initial tumors are shown in dark blue (above the horizontal axis) or in gray (below the horizontal axis), respectively. **(b)** Branched evolution of glioma driver genes during progression. Recurrently mutated driver genes are labeled along the horizontal axis. The vertical axis represents mutation recurrence. Mutations are colored according to whether they are specific to the initial, lower-grade tumor (gray) or the progressed, higher-grade tumor (dark blue) or are shared by the two (light blue). HM, histone modification. **(c)** Examples of branched clonal evolution during glioma progression. For each patient shown, CNAs in the initial and progressed tumors, as well as changes during progression, are indicated in the top panel (blue, deletion; red, amplification; green, copy-neutral LOH). Each point in the bottom panel represents a coding mutation. The horizontal and vertical axes represent estimated clonal frequency for each mutation (the fraction of tumor cells carrying the mutation) in the initial and progressed tumors, respectively. Oncogenic driver mutations are color-coded.

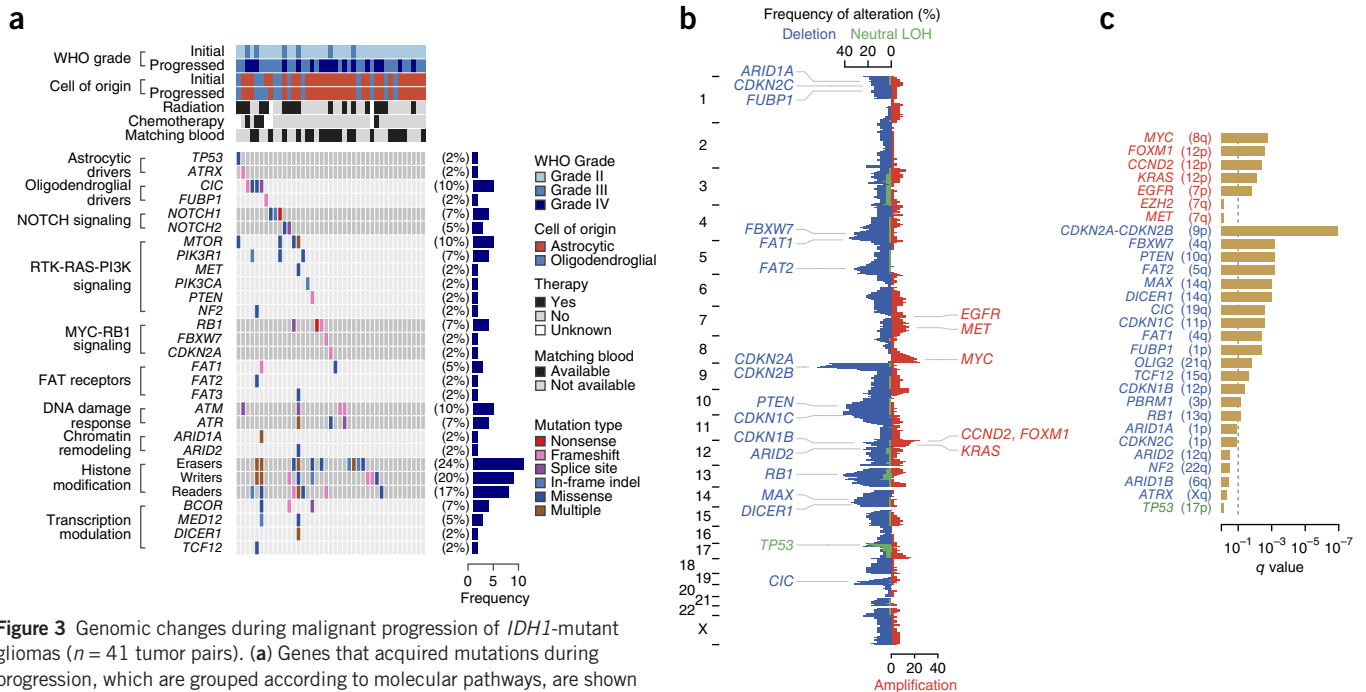


Figure 3 Genomic changes during malignant progression of *IDH1*-mutant gliomas ($n = 41$ tumor pairs). **(a)** Genes that acquired mutations during progression, which are grouped according to molecular pathways, are shown on the left. Mutation frequencies are shown on the right. Each tumor's clinicopathological characteristics are shown at the top. **(b)** Overall pattern of CNAs during malignant progression. The vertical axis represents the genome. The horizontal axis indicates the frequency of chromosomal deletions (blue), copy-neutral LOH events (green) and amplifications (red) during progression. Consensus cancer-related genes located within corresponding genomic regions are indicated. **(c)** Association of CNAs with glioma progression. The horizontal axis represents the significance of association, calculated using conditional logistic regression and likelihood-ratio test. The dashed line corresponds to a q value of 0.1.

respectively), as well as genes involved in transcriptional modulation (12%), were also mutated^{18–23} (Fig. 3a).

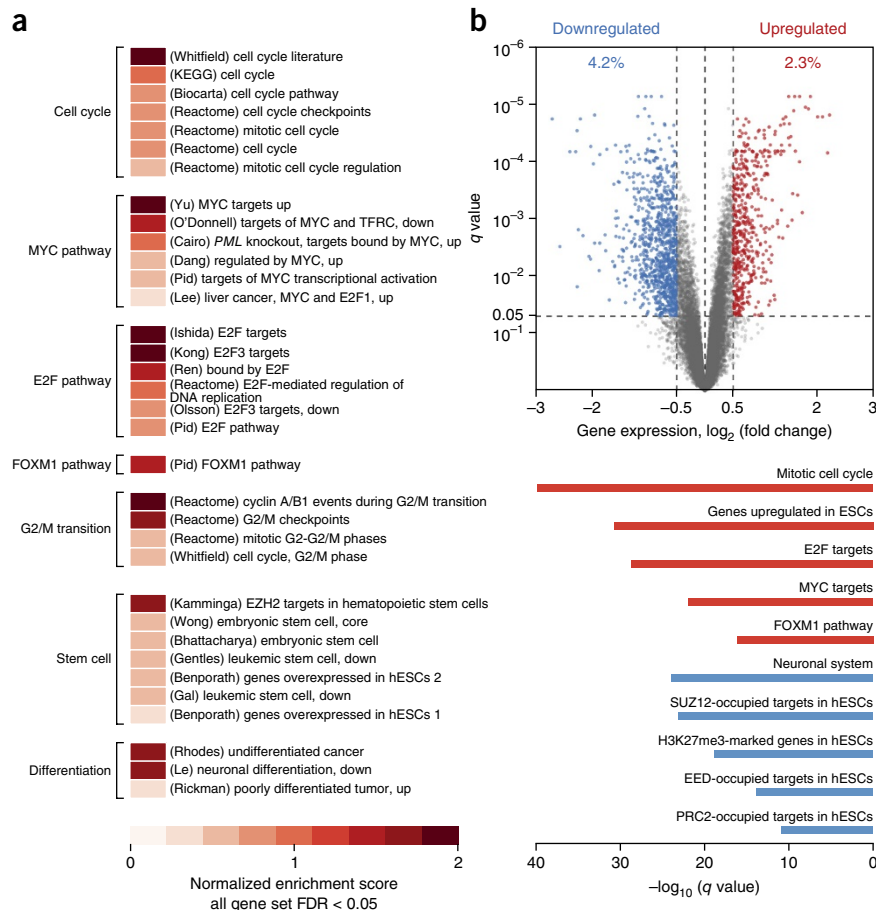
The most frequently amplified genomic region during progression was the *MYC* locus on chromosome 8q (22% of patients), and amplification was significantly associated with progression ($q = 1.7 \times 10^{-3}$, likelihood-ratio test) (Fig. 3b,c). We also observed additional genomic alterations that could potentially activate the *MYC* pathway, including deletions affecting the *FBXW7* gene, a *MYC* suppressor located on chromosome 4q (29% of patients; $q = 6.2 \times 10^{-4}$), as well as the *MAX* gene, encoding a *MYC*-associated factor and tumor suppressor, located on 14q (24%; $q = 9.5 \times 10^{-4}$)²⁴ (Fig. 3b,c). In addition to *MYC*-related loci, the *FOXM1* gene locus on 12p was amplified in 20% of patients, and amplification was significantly associated with progression ($q = 2.7 \times 10^{-3}$) (Fig. 3b,c). The most commonly deleted region during progression was chromosome 9p, encompassing the *CDKN2A-CDKN2B* locus (61%; $q = 1.1 \times 10^{-7}$) (Fig. 3b,c). *PTEN* (10q) was also frequently deleted (39%), and deletion was associated strongly with tumor progression ($q = 6.2 \times 10^{-4}$) (Fig. 3b,c). Chromosome 10q deletion has previously been identified as a marker for poor prognosis in low-grade gliomas²⁵. Finally, we observed deletions affecting the *FAT* receptor genes *FAT1* (4q; 34%; $q = 3.9 \times 10^{-3}$) and *FAT2* (5q; 29%; $q = 6.8 \times 10^{-4}$) (Fig. 3b,c).

To characterize transcriptional changes during progression, we compared the gene expression profiles of the progressed gliomas with the profiles of their corresponding initial tumors ($n = 28$ pairs) (Fig. 4). Gene set enrichment analysis (GSEA) as well as analysis of differential gene expression indicated that cell cycle processes, including the *FOXM1* and *E2F* transcription factor networks and the *MYC* signaling pathway, were enriched (Fig. 4 and Supplementary Figs. 8 and 9). This finding is consistent with the observation that genomic loci harboring several pathway components were themselves altered through CNAs during progression (Fig. 3b,c). Among these pathways,

both the *FOXM1* and *E2F* networks have important roles in promoting G1/S cell cycle transition, with the former also being a key regulator of the G2/M transition, acting synergistically with *MYC* signaling^{26–32}. Besides being a major determinant of mitotic cell fate and a transcriptional amplifier, *MYC* is also required for the maintenance of glioma stem cells^{32–36}. Indeed, genes important for stem cell identity were upregulated during progression (Fig. 4). Differentially downregulated genes that function in the neuronal system ($q = 1.1 \times 10^{-24}$, hypergeometric test) and genes that are targets of Polycomb repressive complex 2 (PRC2) in human embryonic stem cells (hESCs) ($q = 1.4 \times 10^{-11}$) (Fig. 4b).

We next compared the DNA methylation profiles of progressed gliomas with the profiles of their corresponding initial tumors ($n = 24$ pairs), correlating methylation and gene expression changes (Fig. 5). Interestingly, whereas the overall methylation level decreased during progression ($P = 2.6 \times 10^{-3}$, paired two-sided Student's t test), a small fraction (0.6%) of the 307,250 CpG sites measured became significantly hypermethylated (Fig. 5a and Supplementary Figs. 10 and 11). To understand the biological functions associated with these hypermethylated CpGs, we performed genomic region enrichment analysis, identifying enrichment for key developmental transcription factor families ($q = 4.4 \times 10^{-225}$, hypergeometric test), including homeotic genes in all four of the *HOX* clusters, as well as large subsets of the *FOX*, *PAX*, *NKX*, *SOX*, *TALE*, *TALX* and *TBX* family genes³⁷ (Supplementary Fig. 12). These transcription factors regulate a wide variety of developmental processes, including pattern formation, cell fate specification and neurogenesis. In addition, hypermethylated CpGs were also enriched in promoters marked by histone H3 lysine 27 trimethylation (H3K27me3) in hESCs ($q = 4.1 \times 10^{-234}$, hypergeometric test) as well as PRC2 targets in hESCs ($q = 7.4 \times 10^{-204}$) (Fig. 5a). Indeed, by aggregating Encyclopedia of DNA Elements (ENCODE) hESC H3K27me3 chromatin immunoprecipitation and

Figure 4 Gene expression during glioma progression ($n = 28$ tumor pairs). (a) GSEA of transcriptional changes during progression. Significantly enriched gene sets (empirical $P < 0.05$ and false discovery rate (FDR) < 0.05) are labeled and colored according to their normalized enrichment scores. (b) Differential gene expression during progression. A volcano plot comparing gene expression in progressed versus paired initial tumors is shown on the top. Significantly upregulated genes are colored in red, whereas significantly downregulated genes are colored in blue. q values were calculated using paired two-sided moderated Student's t test. Functional enrichments of differentially expressed genes are shown at the bottom. Enrichment q values were calculated by hypergeometric test.



sequencing (ChIP-seq) data, we observed enhanced H3K27me3 signal around CpGs that became hypermethylated during glioma progression³⁸ (Fig. 5b). We obtained similar results using the ChIP-seq hESC data sets for EZH2 and SUZ12 (subunits of PRC2)³⁸ (Fig. 5b). Notably, acquisition of DNA hypermethylation at PRC2 target genes in hESCs has been previously observed in colorectal, ovarian and breast cancers^{39–41}.

The observed changes in DNA methylation correlated with alterations in gene expression (Supplementary Fig. 13). Namely, PRC2 target genes in hESCs that became hypermethylated were also transcriptionally downregulated during progression (Figs. 4b and 5a, and Supplementary Fig. 13). When we restricted our analysis to CpGs located at promoter regions instead of considering all CpGs, we observed similar correlations (Supplementary Fig. 14).

In hESCs, PRC2 binds to the promoters of key developmental transcription factor genes and deposits repressive H3K27me3 marks, thereby maintaining stem cell identity⁴¹ (Fig. 5c). Our observations suggest that, during glioma progression, genes that are bound by

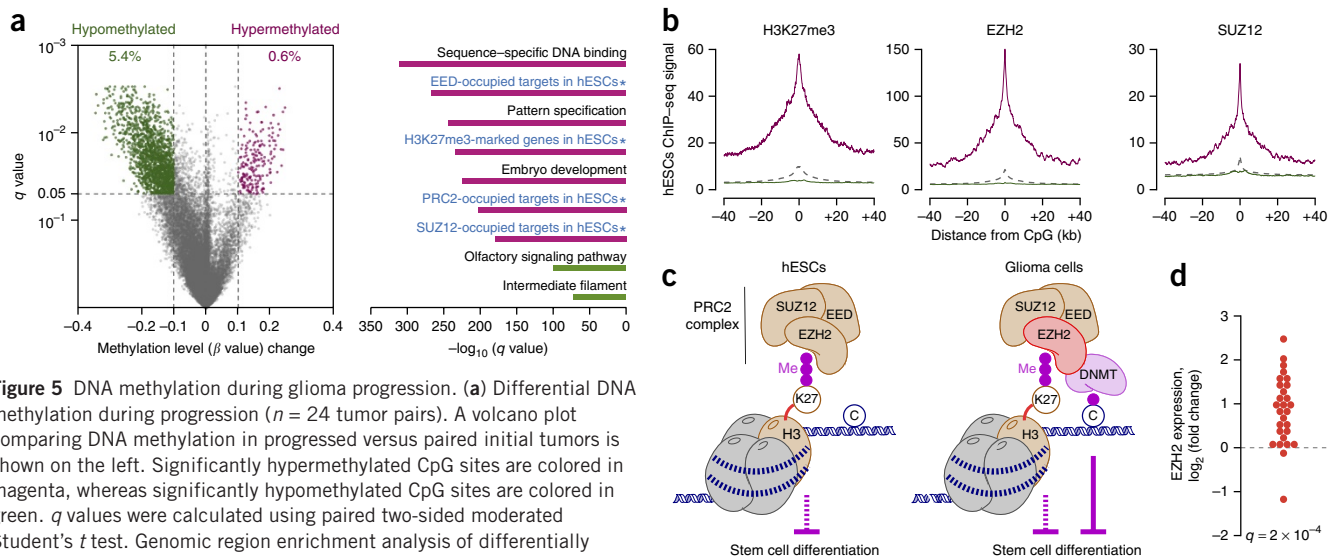


Figure 5 DNA methylation during glioma progression. (a) Differential DNA methylation during progression ($n = 24$ tumor pairs). A volcano plot comparing DNA methylation in progressed versus paired initial tumors is shown on the left. Significantly hypermethylated CpG sites are colored in magenta, whereas significantly hypomethylated CpG sites are colored in green. q values were calculated using paired two-sided moderated Student's t test. Genomic region enrichment analysis of differentially methylated CpGs is shown on the right. Enrichment q values were calculated by hypergeometric test. Gene sets marked with an asterisk and colored in blue are enriched not only in hypermethylated CpGs but also in transcriptionally downregulated genes during progression. (b) Aggregation of hESC ChIP-seq signals around DNA methylation sites. Shown are average H3K27me3, EZH2 and SUZ12 ChIP-seq signals in hESCs near hypermethylated (magenta), hypomethylated (green) and remaining (gray) CpG sites during glioma progression. (c) Schematics depicting aberrant PRC2–DNA methyltransferases (DNMT) crosstalk during glioma progression. Me, methyl group. (d) Fold change in *EZH2* expression during progression ($n = 28$ tumor pairs). The q value is from the differential gene expression analysis, calculated by paired two-sided moderated Student's t test.

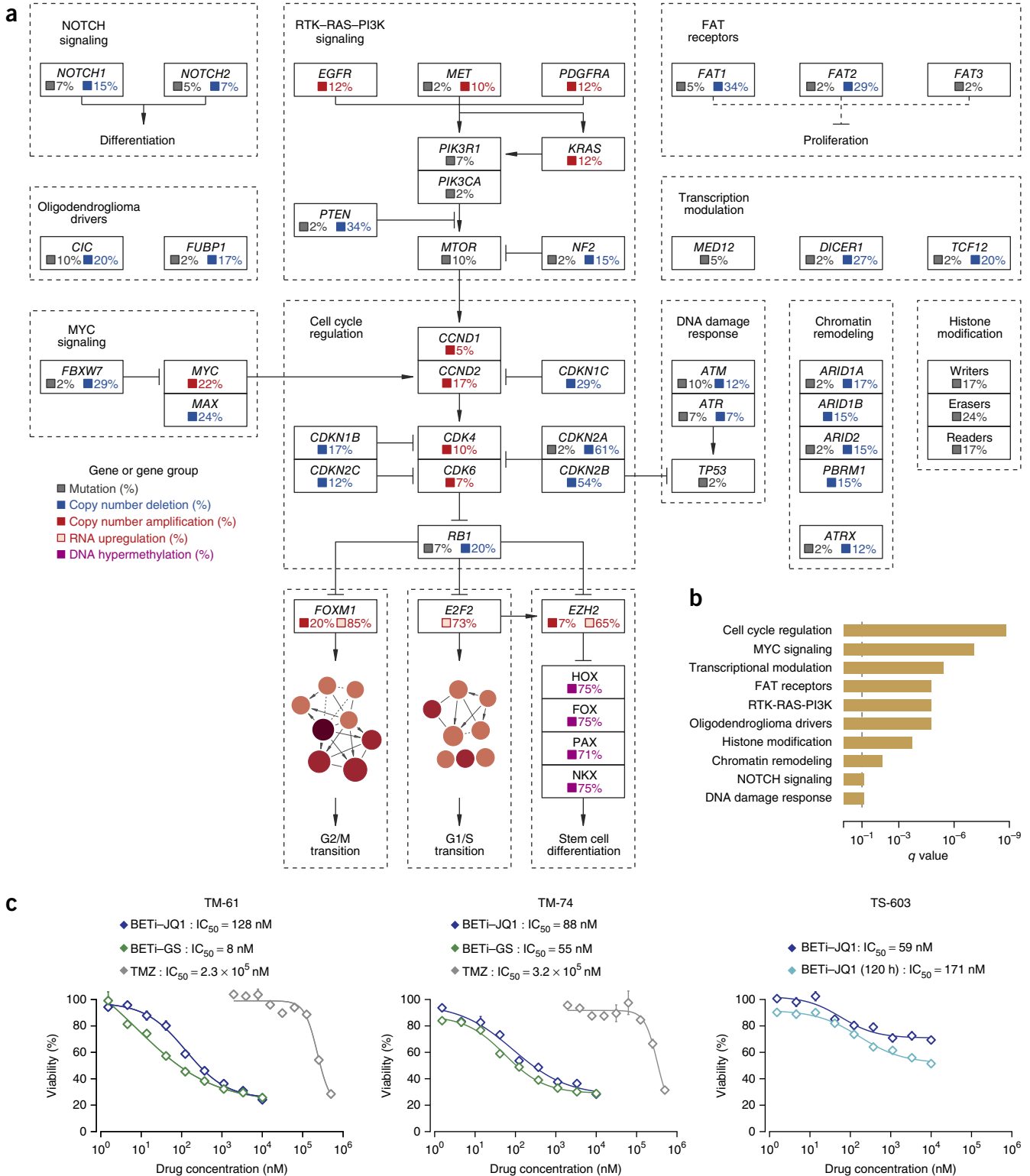


Figure 6 Oncogenic networks during glioma progression and response to BET inhibition. **(a)** The oncogenic pathways driving glioma progression are summarized. Major functional pathways and their components affected by gene mutation, CNA, or alterations in RNA expression or DNA methylation are depicted. **(b)** Association of oncogenic pathways with glioma progression. The horizontal axis represents association significance, calculated using conditional logistic regression and likelihood-ratio test. The dashed line corresponds to a *q* value of 0.1. **(c)** Therapeutic inhibition of *IDH1*-mutant glioma cell growth. Shown are 72- or 120-h survival curves of patient-derived *IDH1*-mutant glioma primary cell cultures (TM-61 and TM-74) as well as an established cell line (TS-603) after treating with BET inhibitors or TMZ. All cells were treated with compounds for 72 h unless otherwise specified. Error bars represent s.e.m., calculated over four replicates per compound per dosage. BETi-JQ1, BET inhibitor JQ1; BETi-GS, BET inhibitor GS-626510; TMZ, temozolomide.

PRC2 in hESCs become further hypermethylated, replacing reversible gene repression with permanent silencing, which in turn could 'lock' glioma cells into a perpetual state of self-renewal^{39–41} (Fig. 5c). Furthermore, in 65% of the patients, we also identified increased expression (\log_2 (fold change) >0.5) of the catalytic subunit of PRC2, *EZH2* ($q = 2 \times 10^{-4}$, paired two-sided moderated Student's *t* test) during progression (Fig. 5d). *EZH2* is known to interact with DNA methyltransferases, associating with their aberrant activities in prostate cancer^{42,43}.

By combining mutation, CNA, and gene expression and DNA methylation alteration data, we were able to uncover oncogenic pathways that drive glioma progression (Fig. 6a). The MYC signaling pathway became activated in 56% of the patients and showed one of the strongest associations with progression ($q = 8.2 \times 10^{-8}$, likelihood-ratio test) (Fig. 6a,b). As mentioned above, alterations in the individual MYC signaling components *MYC*, *FBXW7* and *MAX* were each independently associated with progression ($q < 1.0 \times 10^{-3}$ for each) (Fig. 3b,c). Additionally, the RTK-RAS-PI3K pathway and FAT receptors were altered in 68% and 54% of the patients, respectively ($q = 1.8 \times 10^{-5}$ for both), regardless of whether patients had received adjunct treatment, except for a weak association of RTK-RAS-PI3K pathway activation with treatment ($q = 0.05$, Fisher's exact test) (Fig. 6a,b and Supplementary Fig. 7). This observation suggests that alterations in these pathways are not driven by adjunct therapy but instead reflect genomic changes that underlie glioma progression. Downstream of the MYC and RTK-RAS-PI3K signaling pathways, genes involved in cell cycle regulation were altered in nearly all of the patients (88%; $q = 1.6 \times 10^{-9}$, likelihood-ratio test) (Fig. 6a,b). Further downstream, the FOXM1 and E2F2 networks were upregulated in 85% and 73% of the patients, respectively, promoting G2/M and G1/S transitions (Fig. 6a). Upregulation of *EZH2*, observed in 65% of patients, could also lead to aberrant crosstalk with DNA methyltransferases, which in turn is linked to hypermethylation of developmental transcription factors required for differentiation, as observed in 75% of patients, reprogramming glioma cells into an hESC-like state (Fig. 6a).

Having identified oncogenic pathways driving *IDH1*-mutant glioma progression and taking into account previous studies indicating the therapeutic potential of BET inhibition in primary GBMs, we tested the sensitivity of *IDH1*-mutant glioma cells to two BET inhibitors, the well-established JQ1 compound and GS-626510, a novel inhibitor^{44–47}. Using six patient-derived *IDH1*-mutant primary glioma cultures, as well as a well-characterized *IDH1*-mutant glioma cell line, we observed high sensitivity of *IDH1*-mutant cells to both BET inhibitors *in vitro*, with half-maximal inhibitory concentration (IC₅₀) values less than 250 nM and achieving maximum cell death of 31 to 77% (ref. 48) (Fig. 6c and Supplementary Table 3). Although previous studies have reported BET inhibition to target MYC signaling in hematopoietic malignancies, further studies will be needed to identify the exact therapeutic target of BET inhibition as well as the effectiveness of this inhibition *in vivo* in gliomas^{46,47,49}.

In summary, through direct comparison of the genomic landscape of progressed gliomas with that of their corresponding lower-grade counterparts, we identified a plethora of convergent alterations, including activation of the MYC and RTK-RAS-PI3K signaling pathways, alterations in cell cycle regulators such as *CDKN2A-CDKN2B*, upregulation of the FOXM1- and E2F2-mediated cell cycle transitions, and epigenetic silencing of key developmental transcription factors. Overall, our observations suggest that these pathways converge upon opposing but synergistic cellular effects that stimulate proliferation while inhibiting differentiation during glioma progression. In addition to dissecting these fundamental molecular mechanisms underlying

IDH1-mutant glioma malignant progression, our results further suggest BET inhibition as a potential therapeutic approach.

URLs. FASTX-Toolkit, http://hannonlab.cshl.edu/fastx_toolkit/; cutadapt, <https://cutadapt.readthedocs.org/en/stable/>; GRCh37 human reference genome, <http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/>; Stampy, <http://www.well.ox.ac.uk/project-stampy/>; Burrows-Wheeler Aligner (BWA), <http://bio-bwa.sourceforge.net/>; Picard tools, <http://broadinstitute.github.io/picard/>; Genome Analysis Toolkit (GATK), <https://www.broadinstitute.org/gatk/>; GATK best-practices pipeline, <https://www.broadinstitute.org/gatk/guide/best-practices>; Ensembl, <http://useast.ensembl.org/index.html>; Variant Effect Predictor (VEP), <http://useast.ensembl.org/info/docs/tools/vep/index.html>; ANNOVAR, <http://annovar.openbioinformatics.org/en/latest/>; 1000 Genomes Project, <http://www.1000genomes.org/>; Exome Variant Server, <http://evs.gs.washington.edu/EVS/>; Exome Aggregation Consortium database (ExAC), <http://exac.broadinstitute.org/>; Catalogue of Somatic Mutations in Cancer (COSMIC), <http://cancer.sanger.ac.uk/cosmic/>; Molecular Signature Database (MSigDB), <http://www.broadinstitute.org/gsea/msigdb/index.jsp>; Cytoscape, <http://www.cytoscape.org/>; ReactomeFI plugin, <http://apps.cytoscape.org/apps/reactomefiplugin>; Genomic Regions Enrichment of Annotations Tool (GREAT), <http://great.stanford.edu/great/public/html/>; ENCODE data portal, <http://genome.ucsc.edu/ENCODE/index.html>; Cis-Regulatory Element Annotation System (CEAS), <http://liulab.dfci.harvard.edu/CEAS/>; cBioPortal for Cancer Genomics, <http://www.cbioportal.org/>; R software, <http://www.r-project.org/>; DNACopy R package, <http://www.bioconductor.org/packages/release/bioc/html/DNACopy.html>; qpure R package, <http://sourceforge.net/projects/qpure/>; survival R package, <https://cran.r-project.org/web/packages/survival/index.html>; limma R package, <http://master.bioconductor.org/packages/release/bioc/html/limma.html>; sva R package, <http://www.bioconductor.org/packages/release/bioc/html/sva.html>; GSA R package, <https://cran.r-project.org/web/packages/GSA/index.html>; WGCNA R package, <http://cran.r-project.org/web/packages/WGCNA/index.html>; methylumi R package, <http://www.bioconductor.org/packages/release/bioc/html/methylumi.html>; beta-mixture quantile model (BMIQ), <https://code.google.com/p/bmiq/downloads/detail?name=BMIQ.R>; IMA R package, <https://rforge.net/IMA/>; drc R package, <http://cran.r-project.org/web/packages/drc/index.html>.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Accession codes. Progression-specific mutations identified using whole-exome sequencing, as well as gene expression and DNA methylation microarray data, have been deposited in the European Genome-phenome Archive (EGA) under accession [EGAS00001001588](#).

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

ACKNOWLEDGMENTS

We are grateful to the patients and their families for donating tissues for this research. This study was supported by the Gregory M. Kiez and Mehmet Kutman Foundation and the Yale University Department of Neurosurgery. Partial funding was provided through a research agreement between Gilead Sciences, Inc., and Yale University.

AUTHOR CONTRIBUTIONS

H.B., E.Z.E.-O., K.Y. and A.S.H. performed whole-exome sequencing analysis. H.B. and A.S.H. performed exome CNA, SNP array, DNA methylation array and ChIP-seq aggregation analyses. H.B., A.S.H. and K.Y. performed gene expression

analysis. H.B. and E.Z.E.-O. performed tumor clonal evolution analysis. H.B. performed trinucleotide mutation signature and oncogenic network analyses. H.B. and K.Y. performed statistical association analysis. A.G.E.-S. and J.K. performed mutation validation. P.B.M. performed NOTCH1 structural analyses. M.S., B.K., M.B., M.N.P., K.Ö., J.S. and M.T. provided samples and clinical data. J.L. and A.O.V. conducted neuropathological evaluations. J.L., H.B., V.E.C. and G.C.-G. prepared samples. K.M.-G., L.S. and O.H. generated patient-derived glioma cultures. S.C., S.B.O., E.A.S., S.L.T. and Ş.T. conducted *in vitro* drug testing. H.B., S.C., S.B.O., E.A.S., S.L.T., Ş.T. and T.A.C. analyzed drug testing results. K.M.-G., S.A., L.D.K. and B.G. performed RT-qPCR. A.G.E.-S. performed genomic DNA qPCR. K.B. supervised genomic experiments. H.B., M.G., J.M. and A.L. wrote the manuscript, which was reviewed and edited by the other co-authors. M.G. designed and oversaw the project.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the [online version of the paper](#).

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

- Ostrom, Q.T. *et al.* CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2007–2011. *Neuro-oncol.* **16** (suppl. 4), iv1–iv63 (2014).
- Louis, D.N. *et al.* The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.* **114**, 97–109 (2007).
- Claus, E.B. & Black, P.M. Survival rates and patterns of care for patients diagnosed with supratentorial low-grade gliomas. *Cancer* **106**, 1358–1363 (2006).
- Stupp, R. *et al.* Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N. Engl. J. Med.* **352**, 987–996 (2005).
- Stupp, R. *et al.* Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol.* **10**, 459–466 (2009).
- Suzuki, H. *et al.* Mutational landscape and clonal architecture in grade II and III gliomas. *Nat. Genet.* **47**, 458–468 (2015).
- Cancer Genome Atlas Research Network. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N. Engl. J. Med.* **372**, 2481–2498 (2015).
- Eckel-Passow, J.E. *et al.* Glioma groups based on 1p/19q, *IDH*, and *TERT* promoter mutations in tumors. *N. Engl. J. Med.* **372**, 2499–2508 (2015).
- Turcan, S. *et al.* *IDH1* mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* **483**, 479–483 (2012).
- Jaekle, K.A. *et al.* Transformation of low grade glioma and correlation with outcome: an NCCCTG database analysis. *J. Neurooncol.* **104**, 253–259 (2011).
- Artavanis-Tsakonas, S. Notch signaling: cell fate control and signal integration in development. *Science* **284**, 770–776 (1999).
- Stransky, N. *et al.* The mutational landscape of head and neck squamous cell carcinoma. *Science* **333**, 1157–1160 (2011).
- Agrawal, N. *et al.* Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in *NOTCH1*. *Science* **333**, 1154–1157 (2011).
- Pickering, C.R. *et al.* Mutational landscape of aggressive cutaneous squamous cell carcinoma. *Clin. Cancer Res.* **20**, 6582–6592 (2014).
- Cordle, J. *et al.* A conserved face of the Jagged/Serrate DSL domain is involved in Notch trans-activation and cis-inhibition. *Nat. Struct. Mol. Biol.* **15**, 849–857 (2008).
- Harvey, K.F., Zhang, X. & Thomas, D.M. The Hippo pathway and human cancer. *Nat. Rev. Cancer* **13**, 246–257 (2013).
- Johnson, B.E. *et al.* Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science* **343**, 189–193 (2014).
- Mäkinen, N. *et al.* *MED12*, the mediator complex subunit 12 gene, is mutated at high frequency in uterine leiomyomas. *Science* **334**, 252–255 (2011).
- Barbieri, C.E. *et al.* Exome sequencing identifies recurrent *SPOP*, *FOXA1* and *MED12* mutations in prostate cancer. *Nat. Genet.* **44**, 685–689 (2012).
- Lim, W.K. *et al.* Exome sequencing identifies highly recurrent *MED12* somatic mutations in breast fibroadenoma. *Nat. Genet.* **46**, 877–880 (2014).
- Hill, D.A. *et al.* *DICER1* mutations in familial pleuropulmonary blastoma. *Science* **325**, 965 (2009).
- Kumar, M.S. *et al.* Dicer1 functions as a haploinsufficient tumor suppressor. *Genes Dev.* **23**, 2700–2704 (2009).
- Labreche, K. *et al.* *TCF12* is mutated in anaplastic oligodendroglioma. *Nat. Commun.* **6**, 7207 (2015).
- Comino-Méndez, I. *et al.* Exome sequencing identifies *MAX* mutations as a cause of hereditary pheochromocytoma. *Nat. Genet.* **43**, 663–667 (2011).
- van Thuijl, H.F. *et al.* Spatial and temporal evolution of distal 10q deletion, a prognostically unfavorable event in diffuse low-grade gliomas. *Genome Biol.* **15**, 471 (2014).
- Laoukili, J., Stahl, M. & Medema, R.H. FoxM1: at the crossroads of ageing and cancer. *Biochim. Biophys. Acta* **1775**, 92–102 (2007).
- van den Boom, J. *et al.* Characterization of gene expression profiles associated with glioma progression using oligonucleotide-based microarray analysis and real-time reverse transcription–polymerase chain reaction. *Am. J. Pathol.* **163**, 1033–1043 (2003).
- Liu, M. *et al.* FoxM1B is overexpressed in human glioblastomas and critically regulates the tumorigenicity of glioma cells. *Cancer Res.* **66**, 3593–3602 (2006).
- Koo, C.-Y., Muir, K.W. & Lam, E.W.F. FOXM1: from cancer initiation to progression and treatment. *Biochim. Biophys. Acta* **1819**, 28–37 (2012).
- Wu, L. *et al.* The E2F1–3 transcription factors are essential for cellular proliferation. *Nature* **414**, 457–462 (2001).
- Hollern, D.P., Honeysett, J., Cardiff, R.D. & Andrechek, E.R. The E2F transcription factors regulate tumor development and metastasis in a mouse model of metastatic breast cancer. *Mol. Cell. Biol.* **34**, 3229–3243 (2014).
- Topham, C. *et al.* MYC is a major determinant of mitotic cell fate. *Cancer Cell* **28**, 129–140 (2015).
- Dang, C.V. *et al.* The c-Myc target gene network. *Semin. Cancer Biol.* **16**, 253–264 (2006).
- Wang, J. *et al.* c-Myc is required for maintenance of glioma cancer stem cells. *PLoS ONE* **3**, e3769 (2008).
- Zheng, H. *et al.* p53 and Pten control neural and glioma stem/progenitor cell renewal and differentiation. *Nature* **455**, 1129–1133 (2008).
- Kim, H.S. *et al.* Gliomagenesis arising from Pten- and Ink4a/Arf-deficient neural progenitor cells is mediated by the p53-Fbxw7/Cdc4 pathway, which controls c-Myc. *Cancer Res.* **72**, 6065–6075 (2012).
- McLean, C.Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- Schlesinger, Y. *et al.* Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for *de novo* methylation in cancer. *Nat. Genet.* **39**, 232–236 (2007).
- Keshet, I. *et al.* Evidence for an instructive mechanism of *de novo* methylation in cancer cells. *Nat. Genet.* **38**, 149–153 (2006).
- Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat. Rev. Genet.* **10**, 295–304 (2009).
- Viré, E. *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**, 871–874 (2006).
- Varambally, S. *et al.* The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**, 624–629 (2002).
- Cheng, Z. *et al.* Inhibition of BET bromodomain targets genetically diverse glioblastoma. *Clin. Cancer Res.* **19**, 1748–1759 (2013).
- Pastori, C. *et al.* BET bromodomain proteins are required for glioblastoma cell proliferation. *Epigenetics* **9**, 611–620 (2014).
- Shi, J. & Vakoc, C.R. The mechanisms behind the therapeutic activity of BET bromodomain inhibition. *Mol. Cell* **54**, 728–736 (2014).
- Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468**, 1067–1073 (2010).
- Rohle, D. *et al.* An inhibitor of mutant IDH1 delays growth and promotes differentiation of glioma cells. *Science* **340**, 626–630 (2013).
- Mertz, J.A. *et al.* Targeting MYC dependence in cancer by inhibiting BET bromodomains. *Proc. Natl. Acad. Sci. USA* **108**, 16669–16674 (2011).

ONLINE METHODS

Tumor sample acquisition and quality control. Tumor samples and clinical information used in this study were obtained under informed consent and approval of the institutional review board of the University of Bonn Medical School, the University Hospital of Cologne and the Acibadem University School of Medicine. Clinical information is shown in **Supplementary Table 1**. All tumor samples were collected during surgical resection and were either snap frozen or formalin fixed and paraffin embedded. Sections stained with hematoxylin and eosin or formalin-fixed, paraffin-embedded tissues were reviewed again by neuropathologists at Yale University to confirm the clinical diagnosis and to assess percentage of tumor nuclei and percentage of necrosis. For samples with a low percentage of tumor nuclei or a high percentage of necrosis, tumor regions were isolated using laser-capture microdissection.

DNA, RNA and protein extraction. Frozen tissues were first homogenized using the QIASHredder (Qiagen, 79656). Tumor DNA, RNA and protein samples were subsequently extracted using the AllPrep DNA/RNA/Protein Mini kit (Qiagen, 80004). For formalin-fixed, paraffin-embedded tissues, tumor DNA was extracted using the BiOstic FFPE Tissue DNA Isolation kit (MO BIO Laboratories, 12250-50). Extracted DNA was repaired using BioLabs PreCR Repair Mix (New England BioLabs, M0309S) with human alkyladenine DNA glycosylase (hAAG) (New England BioLabs, M0313S). Repaired DNA was subsequently purified using the QIAquick Nucleotide Removal kit (Qiagen, 28304). Normal blood DNA samples were extracted from patient-matched peripheral blood mononuclear cells using the QIAamp DNA Blood Mini kit (Qiagen, 51104).

Whole-exome sequencing and mutation calling. Exome capture was performed using NimbleGen SeqCap EZ Human Exome Library v2.0 (Roche NimbleGen, 05860504001), following the manufacturer's protocol with modifications, as described previously⁵⁰. The library was sequenced on Illumina HiSeq 2000 or 2500 instruments using 74-bp paired-end reads. Sequencing reads were evaluated using the FASTX-Toolkit and trimmed using cutadapt (version 1.2.1) to remove low-quality 3'-end bases and PCR primer-contaminated sequences⁵¹. Preprocessed paired-end reads were aligned to the human reference genome (GRCh37) using Stampy (version 1.0.21) in a hybrid mode with Burrows-Wheeler Aligner (BWA; version 0.5.9-r16)^{52,53}. All aligned reads were subsequently processed using both the Picard suite (version 1.86) and the Genome Analysis Toolkit (GATK; version 2.5-2), including deduplication, alignment quality assessment, multisequence local realignment and base quality recalibration⁵⁴. Quality control metrics are shown in **Supplementary Table 2**.

Mutation calling was performed using the GATK HaplotypeCaller tool for tumor samples with matching normal DNA samples and the UnifiedGenotyper tool for tumor samples without matching normal samples⁵⁴. For blood-matched tumor samples, genotype likelihood-based somatic scores were also calculated⁵⁵. Detected variants were annotated on the basis of Ensembl annotations (version 69) using Variant Effect Predictor (VEP; version 2.7)⁵⁶. Variants were filtered on the basis of allelic and overall depths of coverage, strand bias, somatic score, mapping quality and coverage-adjusted variant quality, as well as location within repeat or segmental duplicated regions, as described previously^{57,58}. Variants with overall alternate allele frequency >0.01 in the 1000 Genome Project database or in the National Heart, Lung, and Blood Institute (NHLBI) Exome Variant Server (ESP6500SI-V2) database were also excluded⁵⁹. Variants detected in >4 individuals in the Exome Aggregation Consortium database (ExAC; release 0.3) or mapping to genes with >100 conserved alleles in an internal database of over 2,000 exomes were excluded as well. Additionally, consensus oncogenic genes and variants were identified on the basis of the Catalogue of Somatic Mutations in Cancer (COSMIC) database⁶⁰. Lastly, for tumor samples without matched normal DNA samples, variants detected in >2 individuals in the ExAC database, unless also reported in the COSMIC database, were excluded⁶⁰. All detected mutations are shown in **Supplementary Table 4**.

Validation of mutations identified by whole-exome sequencing. To confirm mutations identified by exome sequencing, we made use of either Sanger sequencing or high-throughput sequencing of the NimbleGen SeqCap EZ Comprehensive Cancer Design (Roche NimbleGen, 4000007080), which

captures 578 cancer genes. High-throughput sequencing was performed on an Illumina HiSeq 2500 instrument, reaching 617-fold coverage on average. Read alignment and mutation calling were performed using the GATK 'best-practices' pipeline^{54,61}. Detected mutations were annotated using ANNOVAR (version 2015Mar22)⁶². Sequencing quality control metrics as well as results are shown in **Supplementary Tables 5 and 6**. For Sanger sequencing, target genomic DNA encompassing mutations was PCR amplified. We also tested recurrent *TERT* promoter mutations in 16 tumor pairs⁶⁻⁸. Amplicons were cycle sequenced on Applied Biosystems 9800 Fast Thermal Cyclers and analyzed on an Applied Biosystems 3730xl DNA Analyzer. Primer sequences as well as the results of the Sanger sequencing are shown in **Supplementary Table 7**. The results of mutation validation are summarized in **Supplementary Table 8**.

Trinucleotide mutation signatures. Each single-nucleotide substitution was categorized into one of 96 possible trinucleotide mutations, C>A/G>T, C>G/G>C, C>T/G>A, T>A/A>T, T>C/A>G or T>G/A>C, taking into account the bases immediately 5' and 3' to each substitution. Patterns of trinucleotide signatures for mutations specific to progressed tumors were compared to those of mutations specific to initial tumors by two-sided χ^2 test with 100,000 simulations.

Copy number analysis and tumor purity estimation. For all samples, the whole-exome sequencing read depth was calculated for each captured target region. For 43 glioma DNA samples and 15 blood DNA samples, genome-wide SNP array analysis was also performed using the HumanOmniExpress-24 v1.1 BeadChip (Illumina, WG-315-1101), following the manufacturer's instructions. CNAs were detected on the basis of either sequencing depth estimates or SNP array signal intensities using circular binary segmentation (CBS) as implemented in the DNACopy package in R. For each CNA segment, the minor allele frequency (MAF) was calculated to exclude false positive CNA events. On the basis of MAF within deleted chromosomal regions, tumor purity was also estimated using the qpure package in R (ref. 63). By chromosomal amplifications and deletions, we refer to copy number gains and losses, respectively, of any length and amplitude⁶⁴. Lastly, consensus cancer-related genes within each CNA were identified on the basis of the COSMIC database⁶⁰.

Validation of MYC copy number by genomic DNA quantitative PCR. MYC gene copy number was assessed in 24 randomly selected glioma samples by quantitative PCR (qPCR), using the Applied Biosystems 7900HT Fast Real-Time PCR System with iQ SYBR Green Supermix (Bio-Rad, 1708880). Copy number of the *ZNF423* locus, a region with normal copy number, was assessed as well. Additionally, to determine the qPCR threshold cycle, copy number in female reference DNA (Promega, G1521) in four serial dilutions was also measured. All reactions were run in triplicate. For each sample, the copy number of MYC was calculated and normalized to that of *ZNF423*. Primer sequences and results are shown in **Supplementary Figure 4c**.

Mutation clonal frequency estimation. Each mutation's clonal frequency—the fraction of tumor cells carrying the mutation—was estimated on the basis of its mutant allele frequency, taking into account the chromosomal copy number at the mutation site, as well as tumor purity, as described previously⁶⁵. We considered a mutation to be specific to the progressed tumor if there were no mutant reads in the paired initial tumor and vice versa.

Association of gene alterations with cell of origin. Association of 28 recurrent CNAs with cell of origin was investigated by two-sided Fisher's exact test. Multiple testing was corrected for using the Benjamini-Hochberg method; $q < 0.1$ was considered significant. In addition, we considered an event to be enriched in astrocytic tumors, as opposed to oligodendroglial ones, if the odds ratio was larger than 1 and vice versa.

Association of gene alterations with glioma progression. For 29 recurrent CNAs, we tested the statistical association of each with progression using conditional logistic regression, matching initial and progressed tumors from the same patient. Likelihood-ratio test *P* values were calculated using the clogit function in the survival package in R and were corrected for multiple testing using the Benjamini-Hochberg method; CNAs with $q < 0.1$ were considered significant⁶⁶. Using a similar approach, we tested the association of pathways

with progression, by first grouping 77 genes into ten categories on the basis of the pathways to which they belong.

Statistical relationships between gene alterations during glioma progression. We tested the statistical relationships among 42 genes altered through mutation and/or CNA during progression. The results are shown in **Supplementary Figure 15**.

Pairwise co-occurrence and mutual exclusivity. For every gene pair, we performed a Fisher's exact test to discover co-occurring or mutually exclusive relationships. We tested for these relationships using genomic alterations acquired during progression. Gene pairs with nominal $P < 0.05$ were considered co-occurring or mutually exclusive and were excluded from the following bivariate conditional logistic regression analysis.

Two-gene models. For every gene pair, we tested whether one gene contributed independently to progression by conditioning on the other gene. Specifically, we performed bivariate conditional logistic regression without an interaction term. Likelihood-ratio test P values were calculated using the clogit function in R and were corrected for multiple testing using the Benjamini-Hochberg method⁶⁶. For each gene pair, if both genes showed $q < 0.1$, we considered them to each independently contribute to progression, whereas, if only one gene showed $q < 0.1$, we considered this gene to have 'masked' the effect of the other gene.

Gene expression analysis. High-quality RNA (RNA integrity number (RIN) ≥ 7.5) from 62 exome-sequenced glioma samples (28 initial-progressed tumor pairs plus three initial and three progressed tumors) was used for gene expression array analysis, on the HumanHT-12 v4 Expression BeadChip (Illumina, BD-103-0204). Gene expression data were processed using the limma and sva packages in R, including background correction, quantile normalization and batch effect removal^{67,68}. After excluding undetected microarray probes (detection $P > 0.05$ in over 90% of the samples), a total of 18,416 probes, representing 16,202 genes, remained.

Consensus clustering and multidimensional scaling. The top 1,000 probes showing large variation, as measured by standard deviation, across the data set were selected for consensus clustering and multidimensional scaling (MDS). Consensus clustering was performed using a hierarchical clustering method based on the Pearson correlation coefficients of the selected probes, using the ConsensusClusterPlus package in R (ref. 69). MDS was performed using the R command cmdscale on the Euclidean distances of the selected probes. The results are shown in **Supplementary Figure 16**.

Gene set enrichment analysis. GSEA was conducted using the GSA package in R with the Molecular Signatures Database (MSigDB; version 5.0) c2 gene set collection^{70,71}. Empirical P values of enrichment were calculated by 100,000 gene set permutations^{70,71}. FDRs were also estimated^{70,71}.

Differential expression analysis. Differentially expressed genes were identified by comparing the expression profiles of 28 progressed tumors with the profiles of their corresponding initial tumors, using linear models and empirical Bayes methods, as implemented in the limma package in R (ref. 72). Gene significance was determined on the basis of logarithm of odds (LOD) scores, as well as P values calculated by paired two-sided moderated Student's t test, corrected for multiple testing using the Benjamini-Hochberg method (**Supplementary Table 9**)⁷². Functional enrichment of differentially expressed genes was determined using the hypergeometric test in R with the MSigDB c2 gene sets⁷⁰. Multiple testing was adjusted for using the Benjamini-Hochberg method. Gene interaction networks were constructed using the ReactomeFI plugin (version 4.1.1.beta) in Cytoscape (version 3.0.1)^{73,74}. FDRs for identified interaction networks were also calculated using ReactomeFI^{73,74}.

Co-expression network analysis. A co-expression network was constructed on the basis of the pairwise Pearson correlation coefficients of all genes, using the WGCNA package in R (ref. 75). Modules of highly correlated genes in gliomas were identified⁷⁵. The first principal component of each module, also known as the module eigengene, was calculated⁷⁵. Association of module eigengenes with glioma progression or with cell of origin was assessed using paired and unpaired two-sided Student's t tests, respectively, to identify relevant gene clusters⁷⁵. The results are shown in **Supplementary Figure 17**.

RT-qPCR validation of gene expression microarrays. RNA from eight randomly selected initial-progressed glioma pairs was reverse transcribed using SuperScript III Reverse Transcriptase (Thermo Fisher Scientific, 18080044). qPCR was performed using the Applied Biosystems 7900HT Fast Real-Time PCR System with iQ SYBR Green Supermix. Primer sequences are shown in **Supplementary Table 10**. Each primer pair had three technical replicates per tumor sample. Fold change in mRNA levels during progression was calculated and normalized to the level of *ACTB*. P values were calculated by two-sided Wilcoxon rank-sum test, with a null distribution of zero. The results are shown in **Supplementary Figure 18**.

DNA methylation analysis. Genome-wide DNA methylation profiling was performed on 48 glioma samples (24 initial-progressed tumor pairs) using the Infinium HumanMethylation450 BeadChip (Illumina, WG-314-1003). Generated raw intensity (idat) files were converted into R software-compatible format using the methylumi package⁷⁶. DNA methylation sites with missing values, located within repeat regions or on sex chromosomes were excluded. Probes containing common SNPs were also filtered out, with a total of 307,250 probes retained. Methylation data were subsequently converted into β values, ranging from 0 (unmethylated) to 1 (fully methylated), and these values were normalized using a beta-mixture quantile normalization method (BMIQ)⁷⁷.

Consensus clustering and multidimensional scaling. The top 12,000 most variable probes were selected and analyzed using the same methods as with gene expression data. The results are shown in **Supplementary Figure 19**.

Differential CpG methylation analysis. Differentially methylated CpGs were identified by comparing the methylation profiles of 24 progressed tumors with the profiles of their corresponding initial tumors, using the same methods as with differential gene expression analysis^{72,78}. The results are shown in **Supplementary Table 11**. Analysis of functional enrichment of differentially methylated CpGs was performed using the Genomic Regions Enrichment of Annotations Tool (GREAT), taking all 307,250 methylation sites as the background³⁷.

Analysis of differential promoter methylation. Promoter CpGs were defined as CpGs no more than 200 bp upstream of transcription start sites. For genes with more than one promoter CpG, mean β values were calculated⁷⁸. The results are shown in **Supplementary Table 12**. Functional enrichment was determined using the MSigDB c2 gene sets, as with differential gene expression analysis⁷⁰.

Gene-level correlation of promoter methylation with gene expression. For each gene, promoter methylation alterations during progression were correlated with downstream gene expression changes ($n = 22$ tumor pairs). Spearman's correlation coefficients (ρ), as well as P values, were calculated. Multiple testing was adjusted for using the Benjamini-Hochberg method. The results are shown in **Supplementary Figure 20**.

Human embryonic stem cell ChIP-seq signal aggregation. hESC (H1-hESC) H3K27me3, EZH2 and SUZ12 ChIP-seq data (bigwig files) were downloaded from the ENCODE data portal³⁸. Average ChIP-seq signals surrounding hypermethylated, hypomethylated and all remaining DNA methylation sites were calculated using the *Cis*-Regulatory Element Annotation System (CEAS)⁷⁹.

Survival analysis of the TCGA cohort. Genetic and clinical data for grade II and III gliomas in The Cancer Genome Atlas (TCGA) cohort were accessed through the cBioPortal for Cancer Genomics^{7,80,81}. Grade II and III gliomas carrying *IDH1* mutations affecting Arg132 ($n = 219$) were selected. Using cBioPortal, we generated Kaplan-Meier curves, comparing overall survival rates in patients with and without genetic alterations of interest^{80,81}. Log-rank test P values were also calculated^{80,81}. The results are shown in **Supplementary Figure 21**.

Pharmacological inhibition of glioma cell growth. Glioma cultures were derived from tumor samples different from the whole exome-sequenced samples. This was because the paired tumors were obtained from international collaborators in Germany and Turkey, who collected the samples from 6 to 27 years ago. Because viable cultures could not be established from these patients, we instead derived glioma cell cultures from freshly resected glioma tissues at the Yale-New Haven Hospital. In addition, TS-603 and TS-667, previously

established *IDH1*-mutant and wild-type glioma cell lines, respectively, were kindly provided by the laboratory of T.A.C. (ref. 48). Specifically, the TS-603 line was derived from a grade III oligodendroglioma, which harbored the *IDH1* mutation encoding p.Arg132His, whereas the TS-667 line was derived from a primary GBM⁴⁸. Both cell lines have been karyotyped, and their *IDH1* mutation status was confirmed using Sanger sequencing. The cell lines have also tested negative for mycoplasma contamination. In summary, a total of seven *IDH1*-mutant glioma cell cultures and 13 cultures with wild-type *IDH1* were tested for sensitivity to BET inhibition.

Cell culture. Glioma cells were treated with Accutase (EMD Millipore, SCR005) and triturated to generate single-cell suspensions. Cells were assessed for viability by Trypan blue exclusion and counted using the Countess automated cell counter (Thermo Fisher Scientific, C10227). Cells were grown either in medium containing 10% or 15% serum (DMEM/F-12 with 10% or 15% FBS and 1× penicillin-streptomycin) or in serum-free medium, which was composed of complete NS-A medium containing a mixture of NeuroCult NS-A basal medium (Stemcell Technologies, 05750) and proliferation supplement (Stemcell Technologies, 05753) at a 9:1 ratio, supplemented with 20 ng/ml recombinant human epidermal growth factor (EGF) (Stemcell Technologies, 02633), 10 ng/ml recombinant human basal fibroblast growth factor (bFGF) (Stemcell Technologies, 02634), 2 µg/ml heparin (Stemcell Technologies, 07980), 1× Antibiotic-Antimycotic (Thermo Fisher Scientific, 15240-062) and 2 µg/ml Plasmocin (Invivogen, ant-mpp)⁸².

Neurosphere formation assays. Glioma cells were seeded in 96-Well Clear Round-Bottom Ultra-Low-Attachment Microplates (Corning, 7007) at 500 to 2,000 cells per well. BET inhibitor or an equivalent volume of DMSO (0.025%) was added to each well. Final BET inhibitor concentrations ranged from 14 nM to 1.0×10^4 nM. Microtiter plates were imaged, and the surface areas of neurosphere formation were assessed over 14 d using a CloneSelect Imager (Molecular Devices). Every 4–6 d, 50% of the medium was replaced. The results are shown in **Supplementary Figure 22**.

Cell viability and caspase activity assays. Glioma cells were seeded in 96-well microtiter plates at 2,000 cells per well in 135 µl of serum-free medium or medium containing 10% or 15% serum. After overnight incubation at 37 °C, serially diluted compounds or an equivalent volume of DMSO (0.025%) were added to the cells, resulting in a total volume of 150 µl per well. Final TMZ concentrations ranged from 2.0×10^3 nM to 5.0×10^5 nM, whereas JQ1 (BETi-JQ1) and GS-626510 (BETi-GS) concentrations ranged from 1.5 nM to 1.0×10^4 nM. Each compound had four replicates per dose, whereas DMSO-only control had eight replicates. After incubation for 72 or 120 h at 37 °C, cell viability and caspase-3 and caspase-7 activities were assessed using the CellTiter-Glo Luminescent Cell Viability Assay (Promega, G7570) and the Caspase-Glo 3/7 Assay (Promega, G8090), respectively. Luminescence was measured using the GloMax-Multi Detection System (Promega, E7041). Measurements were normalized with respect to the average signal for DMSO-only control wells, which represented 100% viability or 100% caspase-3 and caspase-7 activity. Dose response curves were fitted using four-parameter log-logistic models with the drc package in R or with GraphPad Prism software (version 6.07)⁸³. IC₅₀ values were also calculated⁸³. The results are shown in **Supplementary Figure 23** and **Supplementary Table 3**.

50. Bilgüvar, K. *et al.* Whole-exome sequencing identifies recessive *WDR62* mutations in severe brain malformations. *Nature* **467**, 207–210 (2010).
 51. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
 52. Lunter, G. & Goodson, M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* **21**, 936–939 (2011).
 53. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

54. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
 55. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
 56. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–2070 (2010).
 57. Clark, V.E. *et al.* Genomic analysis of non-*NF2* meningiomas reveals mutations in *TRAF7*, *KLFA4*, *AKT1*, and *SMO*. *Science* **339**, 1077–1080 (2013).
 58. Erson-Omay, E.Z. *et al.* Somatic *POLE* mutations cause an ultramutated giant cell high-grade glioma subtype with better prognosis. *Neuro-oncol.* **17**, 1356–1364 (2015).
 59. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
 60. Forbes, S.A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39**, D945–D950 (2011).
 61. Van der Auwera, G.A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **11**, 11.10.1–11.10.33 (2013).
 62. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
 63. Song, S. *et al.* qpure: a tool to estimate tumor cellularity from genome-wide single-nucleotide polymorphism profiles. *PLoS ONE* **7**, e45835 (2012).
 64. Zack, T.I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).
 65. Bolli, N. *et al.* Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat. Commun.* **5**, 2997 (2014).
 66. Therneau, T.M. & Grambsch, P.M. *Modeling Survival Data: Extending the Cox Model* (Springer Science & Business Media, 2000).
 67. Shi, W., Oshlack, A. & Smyth, G.K. Optimizing the noise versus bias trade-off for Illumina whole genome expression BeadChips. *Nucleic Acids Res.* **38**, e204 (2010).
 68. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E. & Storey, J.D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
 69. Wilkerson, M.D. & Hayes, D.N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573 (2010).
 70. 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).
 71. Efron, B. & Tibshirani, R. On testing the significance of sets of genes. *Ann. Appl. Stat.* **1**, 107–129 (2007).
 72. Smyth, G.K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, Article3 (2004).
 73. Croft, D. *et al.* Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* **39**, D691–D697 (2011).
 74. Saito, R. *et al.* A travel guide to Cytoscape plugins. *Nat. Methods* **9**, 1069–1076 (2012).
 75. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
 76. Triche, T.J., Weisenberger, D.J., Van Den Berg, D., Laird, P.W. & Siegmund, K.D. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res.* **41**, e90 (2013).
 77. Teschendorff, A.E. *et al.* A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* **29**, 189–196 (2013).
 78. Wang, D. *et al.* IMA: an R package for high-throughput analysis of Illumina's 450K Infinium methylation data. *Bioinformatics* **28**, 729–730 (2012).
 79. Shin, H., Liu, T., Manrai, A.K. & Liu, X.S. CEAS: cis-regulatory element annotation system. *Bioinformatics* **25**, 2605–2606 (2009).
 80. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, p11 (2013).
 81. Cerami, E. *et al.* The cBio Cancer Genomics Portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
 82. Azari, H. *et al.* Isolation and expansion of human glioblastoma multiforme tumor cells using the neurosphere assay. *J. Vis. Exp.* **56**, e3633 (2011).
 83. Ritz, C. & Streibig, J.C. Bioassay analysis using R. *J. Stat. Softw.* **12**, 1–22 (2005).