Published in final edited form as:

Nat Genet. 2018 January ; 50(1): 62–72. doi:10.1038/s41588-017-0001-z.

Mutant-IDH1-dependent chromatin state reprogramming, reversibility, and persistence

Sevin Turcan^{1,12,*}, Vladimir Makarov¹, Julian Taranda², Yuxiang Wang¹, Armida W. M. Fabius^{iD,1,13}, Wei Wu¹, Yupeng Zheng³, Nour El-Amine², Sara Haddock^{1,4}, Gouri Nanjangud^{iD,5}, H. Carl LeKaye⁶, Cameron Brennan⁷, Justin Cross⁸, Jason T. Huse¹, Neil L. Kelleher^{iD,9}, Pavel Osten², Craig B. Thompson¹⁰, and Timothy A. Chan^{iD,1,4,11,*}

¹Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

²Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA.

³Sanofi Genzyme, Waltham, MA, USA.

⁴Weill Cornell School of Medicine, New York, NY, USA.

⁵Molecular Cytogenetics Core Facility, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

⁶Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

⁷Department of Neurosurgery, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

⁸Donald B. and Catherine C. Marron Cancer Metabolism Center, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

⁹Department of Chemistry, Northwestern University, Evanston, IL, USA.

¹⁰Cancer Biology and Genetics Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

¹¹Immunogenomics and Precision Oncology Platform, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

Reprints and permissions information is available at www.nature.com/reprints.

*Correspondence and requests for materials should be addressed to S.T. or T.A.C. sevin.turcan@med.uni-heidelberg.de; chant@mskcc.org.

¹²Present addresses: Department of Neurology, University Hospital Heidelberg, Heidelberg, Germany.

¹³Department of Ophthalmology, VU Medical Center, Amsterdam, The Netherlands.

Armida W. M. Fabius <http://orcid.org/0000-0002-1220-1868>

Gouri Nanjangud <http://orcid.org/0000-0002-8547-1957>

Neil L. Kelleher <http://orcid.org/0000-0002-8815-3372>

Timothy A. Chan <http://orcid.org/0000-0002-9265-0283>

Author contributions

S.T. and T.A.C. conceived of the study. S.T., V.M., J.T., Y.W., A.W.M.F., W.W., Y.Z., N.E.-A., S.H., G.N., H.C.L., C.B., J.C., and J.T.H. performed the experiments. S.T., V.M., J.T., Y.W., A.W.M.F., Y.Z., N.E.-A., S.H., G.N., H.C.L., C.B., J.C., J.T.H., N.L.K., P.O., and T.A.C. analyzed the results. T.A.C. and C.B.T. supervised the project. All authors contributed to the writing or editing of the manuscript.

Competing interests

The authors declare no competing financial interests.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41588-017-0001-z>.

Abstract

Mutations in *IDH1* and *IDH2* (encoding isocitrate dehydrogenase 1 and 2) drive the development of gliomas and other human malignancies. Mutant *IDH1* induces epigenetic changes that promote tumorigenesis, but the scale and reversibility of these changes are unknown. Here, using human astrocyte and glioma tumorsphere systems, we generate a large-scale atlas of mutant-*IDH1*-induced epigenomic reprogramming. We characterize the reversibility of the alterations in DNA methylation, the histone landscape, and transcriptional reprogramming that occur following *IDH1* mutation. We discover genome-wide coordinate changes in the localization and intensity of multiple histone marks and chromatin states. Mutant *IDH1* establishes a CD24⁺ population with a proliferative advantage and stem-like transcriptional features. Strikingly, prolonged exposure to mutant *IDH1* results in irreversible genomic and epigenetic alterations. Together, these observations provide unprecedented high-resolution molecular portraits of mutant-*IDH1*-dependent epigenomic reprogramming. These findings have substantial implications for understanding of mutant *IDH* function and for optimizing therapeutic approaches to targeting *IDH*-mutant tumors.

Mutations in *IDH1* and *IDH2* are found in over 80% of lowergrade gliomas (LGGs) and secondary glioblastomas^{1,2}. *IDH1* mutations commonly occur at codon 132 and lead to the production of 2-hydroxyglutarate (2HG)³. In experimental models, introduction of mutant *IDH1* and 2HG production lead to global DNA hypermethylation, histone methylation alterations, and differentiation block^{4–6}. Pharmacological lowering of 2HG levels with inhibitors selective for mutant *IDH* does not completely reverse mutant-*IDH*-dependent epigenetic changes in vitro or in vivo, at least for the treatment durations tested^{7,8}. A definitive therapeutic effect for *IDH* inhibitors is yet to be established, as both growth suppression and acceleration have been observed upon treatment in various solid tumor models^{7,8}. These studies raise important questions that need to be resolved. First, a comprehensive understanding of how mutant *IDH* alters chromatin states is lacking and needs to be established. Second, it is not known whether epigenomic reprogramming caused by mutant *IDH* is reversible and, if so, what the kinetics of reversibility are. Here we present a comprehensive picture of transcriptional and epigenomic dynamics that dissects the complex molecular reprogramming associated with mutant *IDH1*. The data show that mutant *IDH1* reprogramming of the epigenome and transcriptome is dynamic, characterized by reversibility and partial persistence despite withdrawal of mutant *IDH1*. Furthermore, a decoupling between normal transcriptional and epigenetic states occurs upon loss of mutant *IDH1* expression.

Results

Transcriptional dynamics associated with *IDH1* R132H expression

We performed multi-‘omic’ analyses of engineered immortalized human astrocytes (IHAs) and patient-derived glioma tumorspheres to characterize the dynamics and reversibility of epigenetic and transcriptional changes induced by *IDH1* Arg132His (R132H) (Fig. 1a). We constructed a tetracycline-inducible expression system (Fig. 1b,c) and passaged IHAs in the presence of doxycycline for 30 passages (considered as baseline). We profiled these cells for

1, 5, 10, 20, and 40 passages following doxycycline withdrawal with gene expression and methylation arrays (Fig. 1d). For IHAs with inducible expression of mutant IDH1 at baseline (Dox⁺), starting at ten passages following doxycycline withdrawal, cells in the doxycycline withdrawal state (Dox^{off}) were similar to cells that had never been exposed to doxycycline (Dox⁻) (Fig. 1e). Control IHAs with doxycycline-inducible empty vector did not exhibit significant doxycycline-dependent changes in gene expression (Supplementary Fig. 1a,b). To dissect gene expression dynamics, we identified clusters of genes with transient, gradual, and persistent changes in gene expression following doxycycline withdrawal (Supplementary Table 1). Coherent clusters included 193 downregulated probes (151 genes) (73.6% transient, 16.6% gradual, and 9.8% persistent) (Fig. 1f, left) and 243 upregulated probes (205 genes) (64.2% transient, 30% gradual, and 5.8% persistent) (Fig. 1f, right). Persistently upregulated genes—whose expression was independent of IDH1 R132H after doxycycline withdrawal—included *L1CAM*, *MTUS1*, and *WNT6* (Fig. 1g, top). *L1CAM*, a neural cell adhesion molecule with roles in normal brain development, is upregulated in gliomas and preferentially expressed in glioma stem cells^{9–12}. Several transcription factors, including *MAF*, *MEOX2*, and *NKX2-1*, were persistently downregulated (Fig. 1g, bottom). *L1CAM* was upregulated ($P = 7.5 \times 10^{-8}$) and *MEOX2* was downregulated ($P = 4.6 \times 10^{-9}$) in IDH-mutated LGGs, and these genes may have a role in glioma pathogenesis (Supplementary Fig. 1c). The magnitude of the fold change in expression at baseline was a major determinant of gene expression reversibility ($P = 0.0026$ for upregulated genes and $P < 2.2 \times 10^{-16}$ for downregulated genes) (Supplementary Fig. 1d). To assess the relevance of our findings, we compared gene expression and methylation changes to those observed in IDH-mutant gliomas⁵, finding similarly deregulated pathways in IHAs inducibly expressing mutant IDH1 and CpG island methylator phenotype positive (CIMP⁺) gliomas (Supplementary Fig. 2a). Furthermore, hypermethylated loci in IHAs with inducible expression of mutant IDH1 displayed increased median methylation in CIMP⁺ LGGs as compared to CIMP⁻ LGGs (Supplementary Fig. 2b).

Interestingly, our study identified mutant-IDH1-dependent upregulation of *CD24*, which encodes a putative cell-surface marker for stem-like cell populations¹³ that is overexpressed in tumors and broadly expressed in many tissues, including the brain¹⁴ (Fig. 1h). In IHAs inducibly expressing mutant IDH1, *CD24* was upregulated as compared to non-induced cells, and its expression was decreased upon doxycycline withdrawal (Fig. 1h and Supplementary Fig. 2c). We confirmed higher *CD24* expression in a cohort of CIMP⁺ LGGs (Supplementary Fig. 2d). We then sorted CD24⁻ and CD24⁺ subpopulations from Dox⁺ IHAs expressing mutant IDH1 (~2–6% of these IHAs were CD24⁺). CD24⁺ cells formed significantly more colonies in soft agar than CD24⁻ cells (Supplementary Fig. 2e,f). Transcriptome analysis of sorted CD24 subpopulations indicated 285 upregulated and 293 downregulated genes in CD24⁺ relative to CD24⁻ cells (Supplementary Table 2). Gene set enrichment analysis (GSEA) showed positive enrichment of cholesterol biosynthesis, MYC target, WNT signaling, and embryonic stem cell core pathways, demonstrating a stem cell-like gene expression program in CD24⁺ cells (Supplementary Fig. 2g). Methylation analysis identified 16,525 hypermethylated and 20,260 hypomethylated loci in CD24⁺ IHAs relative to CD24⁻ IHAs (Supplementary Table 2). Overall, 137 downregulated and 189 upregulated genes had a CpG probe whose methylation was inversely correlated with their expression in

CD24⁺ and CD24⁻ cells (Fig. 1i and Supplementary Table 2). We used RNA-seq data from the BrainSpan Developmental Transcriptome¹⁵ to determine the expression patterns of genes upregulated in CD24⁺ IHAs during human brain development. Several genes, including *CD24P4*, *L1CAM*, and *GNAL*, were expressed at higher levels in the prenatal brain and the brain during early infancy as compared to adult brain (Fig. 1j). Notably, *L1CAM* is an interaction partner of CD24 and can promote self-renewal, tumorigenesis, and proliferation in gliomas^{9,16,17}. Taken together, these data suggest that CD24⁺ cells emerge in a progenitor-like state from astrocytes expressing IDH1 R132H and this process may contribute to epigenetic and transcriptional reprogramming.

Dynamic reshaping of chromatin landscape by IDH1 R132H

ChIP-seq analysis of major histone modifications showed progressive enrichment of H3K4me3 (associated with actively transcribed genes), H3K36me3 (preventing intragenic cryptic transcript initiation during active transcription), and H3K9me3 and H4K20me3 (repressive marks enriched in heterochromatic regions) in successive passages of IHAs with stable expression of IDH1 R132H but not in parental astrocytes (Fig. 2a,b). We inspected the chromatin reorganization of H3K9me3, H4K20me3, and H3K36me3 using Hilbert curves¹⁸, which showed mutant-IDH1-specific spread over additional genomic regions (Supplementary Fig. 3a). To determine histone methylation kinetics, we grouped the union of peaks across all cell lines with stable mutant IDH1 expression per histone mark by *k*-means clustering into six clusters. Regions within H3K4me3 cluster 1 exhibited the highest intensity across all passages and were mostly located within 5' UTRs or promoter regions (Supplementary Fig. 3b), whereas regions in cluster 6, which showed enrichment in mutant-expressing astrocytes at passage 40, were mostly located in introns or intergenic regions (Fig. 2c and Supplementary Fig. 3b). Clustering of H3K9me3 and H3K36me3 peaks identified regions with similar enrichment across all samples (cluster 5 for both H3K9me3 and H3K36me3) and regions with uniformly high enrichment for H3K9me3 (cluster 2 and cluster 6) or H3K36me3 (cluster 1 and cluster 6) in mutant-expressing astrocytes at passage 40 (Fig. 2c,d). Next, we divided the genome into 5-kb sliding windows, calculated H3K9me3 and H3K36me3 enrichment for each window, and identified six distinct groups using *k*-means clustering. This unbiased approach also identified genomic regions with increased H3K9me3 and H3K36me3 signal in mutant-expressing astrocytes at passage 40, whereas larger regions were quiescent for these histone marks at the same passage (Supplementary Fig. 3c–e). These results indicate that increased H3K9me3 and H3K36me3 methylation occurs at specific regions, rather than uniformly across the genome (Supplementary Fig. 3c–e).

While there was an overall increase in H3K36me3, H3K9me3, and H4K20me3 in IHAs expressing mutant IDH1, the proportions of overlapping genomic features for these marks remained approximately equivalent to those in parental cells (Supplementary Fig. 4a). In contrast, the distribution of H3K4me3 peak locations extended to intronic and intergenic regions in IHAs expressing IDH1 R132H (Fig. 2e). Notably, we observed focal H3K4me3 gains at 4q12 in mutant-expressing astrocytes, including increased H3K4me3 within the *DANCR* locus and the transcriptional start site (TSS) of *PDGFRA*, an oncogene intimately linked to gliomagenesis¹⁹ (Fig. 2f).

We analyzed bulk histone methylation levels in IHAs with inducible expression of mutant IDH1 by liquid chromatography coupled with mass spectrometry of purified histones, confirming significant increases in H3K4me2, H3K4me3, H3K9me2, H3K9me3, and H3K36me3 in Dox⁺ IHAs (Supplementary Fig. 4b). To determine the reversibility of the changes in the histone methylation landscape, we performed ChIP-seq on H3K4me3, H3K27me3, H3K9me3, and H3K36me3 at baseline and at 40 passages following doxycycline withdrawal. As for IHAs stably expressing mutant IDH1, we observed substantial H3K4me3 enrichment along with increases in H3K36me3 and H3K9me3 in Dox⁺ IHAs relative to Dox⁻ IHAs at baseline (although the increases were more modest than those in IHAs with stable expression) (Fig. 2g).

To determine the chromatin state organization in IHAs expressing mutant IDH1, we used ChromHMM²⁰ to learn genome-wide combinations of H3K4me3, H3K36me3, H3K27me3, and H3K9me3 in IHAs with stable and inducible expression (Fig. 2h,i). In IHAs stably expressing mutant IDH1, we identified states largely devoid of histone marks (states 1–3) with differential genome coverage and enrichment. Predominant features (>20% emission probability, defined as the probability of finding a histone mark in a given state) for other states included H3K36me3 + H3K9me3 (state 4), H3K36me3 + H3K9me3 + H3K27me3 (states 5 and 6), all features (state 7), H3K9me3 + H3K27me3 (state 8), H3K9me3 + H3K27me3 + H3K4me3 (state 9), and H3K4me3 (state 10). Similarly, in IHAs with inducible expression, states 1, 3, and 6 did not have any histone marks (Fig. 2i), states 4 and 5 had features of all histone marks, and state 2 was marked by H3K36me3 + H3K9me3 + H3K27me3. Features for other states included H3K36me3 + H3K27me3 (state 7), H3K36me3 with high enrichment in exonic regions (state 8), H3K9me3 + H3K36me3 (state 9), and H3K4me3 (state 10) (Fig. 2i). Chromatin state analysis showed a time-dependent increase in genome coverage of state 5 in IHAs with stable expression of mutant IDH1 (6.7% of the genome covered after 40 passages) (Fig. 2h). A similar chromatin state in IHAs with inducible IDH1 expression (state 5) exhibited higher genome coverage in Dox⁺ IHAs (3.2%) (Fig. 2i). In addition, there was a greater percentage of IHAs stably expressing mutant IDH1 in state 2, which is marked by low levels of H3K36 signal (Fig. 2h). Interestingly, as compared to Dox⁻ and Dox^{off} cells, Dox⁺ IHAs had a greater percentage of their genomes in state 6 (largely devoid of histone marks), whose regions frequently transition to state 5 or state 8 (marked by high H3K36me3 signal), indicating that these regions are located near active elements (Fig. 2i). These results show dramatic, specific, and dynamic histone mark reorganization in the presence of IDH1 R132H.

Mutant-IDH1-induced DNA methylation is gradually reversible, but a fraction of the genome exhibits persistence

Next, we compared the DNA methylation profiles of Dox⁺ and Dox⁻ inducible IHAs at baseline. Overall, 41,209 loci were differentially methylated, of which 67.4% were hypermethylated ($n = 27,792$). Notably, 18,023 of the hypermethylated loci displayed at least 70% methylation in a patient-derived glioma tumorsphere harboring mutation of endogenous *IDH1* encoding the R132H mutant (TS603) ($P = 1 \times 10^{-6}$), suggesting methylation gain across similar loci (Supplementary Fig. 5a,b). Unsupervised clustering of DNA methylation showed that, starting at 20 passages following doxycycline withdrawal,

Dox^{off} astrocytes were more similar to Dox⁻ than to Dox⁺ astrocytes (Fig. 3a). Control IHAs with empty vector did not undergo doxycycline-dependent changes in methylation (Supplementary Fig. 5c–e).

While methylation at most loci eventually returned to baseline levels following doxycycline withdrawal, 25% of differentially methylated loci maintained their aberrant methylation levels (Fig. 3b). Using an approach similar to that applied to delineate gene expression dynamics, we identified methylation clusters with time-dependent changes (Supplementary Table 3). Overall, 19,528 hypermethylated loci (13.8% transient, 62.5% gradual, and 23.7% persistent) and 8,342 hypomethylated loci (52.4% transient, 29.1% gradual, and 18.5% persistent) showed coherent time-dependent kinetics (Fig. 3c,d). As for the gene expression data, we observed a significant positive correlation between the magnitude of methylation changes and the degree of persistence ($P < 2.2 \times 10^{-16}$). Overall, 49% of loci with intermediate and 64.4% of loci with high gains in methylation were persistently hypermethylated (Fig. 3e). Next, we binned persistently methylated loci by their baseline β values (DNA methylation percentage) in Dox⁻ IHAs (Supplementary Fig. 6a,b). Most of these loci had intermediate methylation levels while 765 loci had less than 20% methylation in the baseline state (Supplementary Fig. 6a,b). The median β -value difference (β) for persistently methylated loci compared to baseline was a gain of 21% at the end of the doxycycline withdrawal time course (Supplementary Fig. 6c). Select loci ($n = 29$) were unmethylated at baseline (<10%) and maintained a > 30% increase at passage 40 after doxycycline withdrawal (Supplementary Fig. 7a,b).

Overlap of differentially methylated loci with Encyclopedia of DNA Elements (ENCODE) states²¹ showed hypermethylation proximal to TSSs (states 1 and 2) and enhancer regions (state 7) and hypomethylation of quiescent regions (state 15) in IHAs with inducible expression of mutant IDH1 (Fig. 3f). In the persistent state, hypermethylated loci were preferentially localized in promoter regions (states 1 and 2), whereas 52.3% of hypomethylated loci were located within the quiescent genome (state 15).

We sought to determine whether any correlation existed between altered gene expression and the epigenetic state in IHAs inducibly expressing mutant IDH1 (Fig. 3g). Downregulated genes in the gradual and persistent gene expression clusters tended to have a tighter hypomethylated β -value distribution within their intergenic regions (Fig. 3g). We plotted ChIP-seq signals ± 1 kb around CpG sites corresponding to either all probes, to generate background signal, or methylation clusters (Supplementary Fig. 8a,b). Persistently hypermethylated loci exhibited slightly increased surrounding H3K4me3 signal in Dox⁻ cells relative to Dox⁺ cells at baseline and were largely depleted of H3K27me3, H3K36me3, and H3K9me3 (Supplementary Fig. 8b). In contrast, persistently hypomethylated loci had low H3K4me3 signal (similar to other clusters) but exhibited increased signal (with low enrichment) for H3K27me3, H3K36me3, and H3K9me3 in Dox⁻ cells (Fig. 3h). It is important to note that all three gene expression clusters had an overall higher H3K4me3 signal surrounding hypomethylated CpG sites in Dox⁺ versus Dox⁻ cells, suggesting a shift in H3K4me3 toward regions that lose methylation (Fig. 3h). Overlap of chromatin states with aberrantly methylated loci indicated differences between the chromatin states associated with hypermethylated and hypomethylated loci; however, we did not observe a

chromatin state common among persistent clusters (hypo- and hypermethylated), with the exception that hypomethylated loci in the persistent cluster overlapped minimally (1% of persistently hypomethylated loci) with chromatin state 10 (marked by high H3K4me3 enrichment at TSSs) (Fig. 3i). These findings suggest that the small set of genes whose expression continues to be perturbed after withdrawal of IDH1 R132H are likely to be indirectly regulated by mutant IDH and that maintenance of an aberrant chromatin state is not necessary for persistent changes in gene expression.

Gain of H3K4me3 and loss of DNA methylation at regions with quiescent chromatin states

ChromHMM analysis showed low-to-intermediate enrichment levels of H3K4me3 co-occurring with other states such as H3K27me3 + H3K9me3 + H3K36me3, pointing to aberrant coexistence of chromatin states in cells expressing mutant IDH1 (Fig. 2h,i). One such state, state 5 with higher genome coverage in Dox⁺ IHAs, was largely distributed among intronic and intergenic regions (Fig. 4a). While several Dox⁻ chromatin states (states 1, 6, and 10) were maintained in Dox⁺ samples, others showed shifts in their composition (Fig. 4b). Overall, in Dox⁻ and Dox⁺ inducible IHAs, H3K4me3 signal displayed a positive correlation with gene expression (Supplementary Fig. 9a) and, when TSSs were ranked by H3K4me3 occupancy across both inducible lines, the expression of several genes, including *PDGFRA* and *CD9* with higher H3K4me3 ranking in Dox⁺ samples, was also upregulated by more than 1.5-fold (Supplementary Table 4).

Although most H3K27me3, H3K9me3, and H3K36me3 peaks reverted to baseline levels following doxycycline withdrawal (Supplementary Fig. 9b), a distinct proportion of the sites at which H3K4me3 peaks were gained in Dox⁺ IHAs maintained their aberrant histone methylation (Supplementary Fig. 9c). Of note, 37% of genes with decreased H3K4me3 signal were also persistently hypermethylated, including *MAP1LC3A*, *MARVELD2*, and *TET1* (Supplementary Fig. 9d). RNA-seq data indicated a strong reduction in expression of the autophagy-related gene *MAP1LC3A* that was maintained upon doxycycline withdrawal (Fig. 4c and Supplementary Table 5).

We derived the 11,443 regions with the most significant H3K4me3 gain in Dox⁺ IHAs inducibly expressing mutant IDH1 relative to Dox⁻ IHAs (Supplementary Table 6). These regions were largely devoid of H3K4me3 in Dox⁻ cells and maintained low-level H3K4me3 signal in the Dox^{off} state (Fig. 4d). A majority of regions that gained H3K4me3 with mutant IDH1 expression were enriched for intergenic and intronic regions (Fig. 4e,f).

To further explore IDH1 R132H-specific chromatin alterations associated with H3K4me3, we performed ChIP-seq for H3K4me3 on TS603 and two IDH-wild-type tumorspheres (TS543 and TS667). GSEA using the GREAT toolbox of the genes corresponding to the top 2,000 H3K4me3 peaks at TSSs for all profiled cells identified several enriched gene sets in IDH-mutant but not IDH-wild-type cells, including MYC/MAX targets and ZNF143 binding, implicating MYC signaling^{22,23} and aberrant chromatin interactions in IDH-mutant cell lines (Fig. 4g). Furthermore, differential enrichment analysis of models with mutant versus wild-type IDH1 identified 1,159 H3K4me3-binding sites that separated cell lines on the basis of IDH1 R132H status, and aberrant H3K4me3 remained at lower enrichment

levels in the Dox^{off} state for these differentially enriched regions (Fig. 4h, Supplementary Fig. 9e, and Supplementary Table 6).

Upregulated viral defense signatures, endogenous retroviruses, and genome instability in IDH1 R132H-expressing astrocytes

We hypothesized that H3K4me3 gain and DNA methylation loss at intergenic regions might poise repetitive elements located at methylation-dense, quiescent genomic regions for activation. GSEA indicated a highly significant positive enrichment of genes within the interferon (IFN)- α response pathway gene set in astrocytes with inducible expression of mutant IDH (Fig. 5a,b). Given the upregulated interferon response, we considered whether endogenous retroviruses (ERVs) might be reactivated in response to IDH1 R132H expression. As described previously²⁴, we mapped our RNA-seq data to a family of expressed ERVs and identified specific ERVs (for example, ERVH-2, ERVH-4, ERVH-6, and ERVH-5) with elevated expression in mutant-IDH-expressing samples, but not in control cells, upon long-term culturing (at the end of the time course, 70 passages) (Fig. 5c). We validated several of these ERVs by qPCR (Fig. 5d). Additionally, a previously described viral defense signature²⁵ was upregulated in IHAs inducibly expressing mutant IDH1, as well as in a subset of IDH1-mutant LGGs (Fig. 5e and Supplementary Fig. 10a). We tested expression of ERVs previously described to trigger an immune response²⁶ and observed a modest but significant increase in expression for several of these ERVs (Fig. 5f). Therefore, it appears that IDH1 R132H may have a role in upregulating ERVs, which may contribute to genome instability or immune activation²⁵.

On the basis of the observations above, we wondered whether mutant IDH1 could promote genetic instability. GSEA indicated aberrant enrichment of several positional gene sets in Dox⁺ IHAs, including negative enrichment of chromosome 19q13 genes (Fig. 5g). Analysis of copy number alterations using methylation data identified amplifications and deletions of several regions, including a higher frequency of broad deletions of chromosome 19q, in Dox⁺ IHAs inducibly expressing mutant IDH1 as compared to Dox⁻ IHAs, but not in controls with empty vector (Fig. 5h and Supplementary Fig. 10b,c). Chromosome 19q loss (with intact 1p) occurs in gliomas and may be a negative prognostic marker²⁷. Metaphase FISH confirmed our finding of an increased frequency of 19q loss in mutant-IDH1-expressing IHAs (Fig. 5i). Loss of 19q occurred primarily via unbalanced translocation or rarely by isochromosome of 19p.

In vitro and in vivo phenotypic persistence associated with IDH1 R132H expression

We asked whether mutant-IDH1-associated phenotypes required continued presence of IDH1 R132H. Consistent with previous studies²⁸, Dox⁺ IHAs exhibited defective contact inhibition and increased proliferation despite confluence (Fig. 6a). Strikingly, loss of contact inhibition was highly dependent on IDH1 R132H expression in vitro and returned to the normal state within six passages of doxycycline withdrawal, at a faster pace than for epigenetic and transcriptional reversion (Fig. 6a). This suggests that contact inhibition abrogation requires continued mutant IDH1 expression. To examine whether mutant IDH1 is required for tumor maintenance, we implanted the inducible IHAs into the forebrains of NOD-SCID mice. Animals were placed on a doxycycline-containing diet to induce IDH1

R132H expression. Two months after implantation, mice were randomized on the basis of bioluminescence signal (BLI): ten mice continued on doxycycline and ten mice returned to regular mouse feed. BLI measurements indicated a small (albeit not statistically significant) reduction in tumor growth after doxycycline withdrawal (Fig. 6b). None of the mice injected with control IHAs developed tumors (Supplementary Fig. 11a), while several mice (~40%) injected with Dox⁺ IHAs displayed intracranial lesions on magnetic resonance imaging (MRI) (Supplementary Fig. 11b,c).

To better understand the morphology of these tumors, we used serial two-photon tomography (STPT) to image ZsGreen fluorescence (marker for injected inducible IHAs) throughout the brain. Cells positive for green fluorescence were automatically detected (Fig. 6c) and reconstructed in 3D (Fig. 6d and Supplementary Videos 1–3), and their spatial information was registered to a 3D reference brain based on the Allen Brain Atlas to extract structural information (Fig. 6e)^{29,30}. While clusters of injected Dox⁺ cells were not apparent by MRI, mouse brains injected with these cells displayed several bright areas on STPT slices (Supplementary Fig. 11d). Semiquantitative analysis of a representative STPT image showed that the majority of cells were present at the injection site (somatosensory cortex) and spread into nearby regions, with growth in dorsal striatum (Fig. 6e). Notably, we identified several brains with parenchymal infiltration of mutant-IDH1-expressing astrocytes observed at cellular resolution that were not apparent on MRI images (Supplementary Fig. 11e). Overall, our results indicate that IHAs with IDH1 R132H have a greater potential to infiltrate and form tumors in an orthotopic model; however, tumors continue to grow when IDH1 R132H is removed.

Discussion

We have assembled an atlas of integrated longitudinal epigenomic and transcriptomic data from IDH1 R132H-expressing glial systems to map a detailed landscape of global remodeling and reversibility associated with *IDH1* mutations. Although *IDH1* mutation results in progressive accumulation of methylation marks on DNA and histones, subsequent suppression of mutant IDH1 expression does not result in the epigenome and transcriptome completely returning to the original state (at least over the time period tested). Our findings suggest alternate effector pathways associated with IDH mutations that may eventually be decoupled from the neomorphic enzymatic activity of mutant IDH protein. This is intriguing given previous findings that 2HG inhibition alone is not sufficient to reverse the mutant-IDH-induced epigenetic state^{7,8,31}. Our data also show the slow nature of methylation changes upon loss of mutant IDH1 and highlight the need for experiments with longer time courses using mutant IDH1 inhibitors to determine its long-term effects together with the utility of combining epigenetic drugs with small-molecule inhibitors of mutant IDH1.

We identified *LICAM* as being persistently upregulated despite long-term loss of mutant IDH1 expression, suggesting that continued *LICAM* expression might contribute to the oncogenic phenotype conferred by mutant IDH1. Previous studies have shown that *LICAM* is expressed on glioma stem cells and that its suppression leads to reduced tumor growth in intracranial brain tumor xenograft models⁹. Further research into *LICAM* inhibition in IDH-

mutant gliomas, perhaps with L1CAM-blocking antibodies, may be considered as a means to sensitize therapeutic response to mutant-IDH-specific inhibitors.

Our results show that, in addition to repressive chromatin marks such as H3K9me₃, there is enrichment of activating histone marks (such as H3K4me₃) upon expression of mutant IDH1, highlighting the importance of understanding the chromatin landscape in IDH-mutant gliomas. Our data also show that loss of DNA methylation in mutant-IDH1-expressing astrocytes occurs more frequently in regions marked by a quiescent/low chromatin state. Interestingly, DNA hypomethylation has been observed in IDH1-mutant cell line models and in recurrent IDH1-mutant gliomas, suggesting that these hypomethylated regions may have a role in initiation or progression of IDH-mutant gliomas^{6,23}. H3K4me₃ enrichment and loss of DNA methylation at quiescent genomic regions may indicate the presence of an open and dynamic chromatin state that can lead to aberrant recruitment of transcription machinery and transcriptional activation of noncoding RNAs such as ERVs. Our observation that loss of 19q is accelerated in IDH1 R132H-expressing astrocytes suggests that, collectively, deregulation of histone marks and DNA methylation, and possibly activation of ERVs, may lead to genomic instability.

We demonstrate gain in H3K4me₃ at the *PDGFRA* promoter in IDH1 R132H-expressing astrocytes as early as passage 2. *PDGFRA* is overexpressed in gliomas, and previous work has shown that IDH mutations disrupt enhancer boundary function, leading to *PDGFRA* activation³². An early H3K4me₃ gain may poise *PDGFRA* for activation before genome-wide hypermethylation is established or may act together with reduced CTCF binding to achieve *PDGFRA* activation. These data suggest that *PDGFRA* signaling may be poised for activation as an early event in IDH1-mutant gliomas.

We also identified a CD24⁺ stem cell-like subpopulation with distinct molecular alterations within IDH-mutant cells. In soft agar assays, we observed the increased potential for colony formation of the CD24⁺ subpopulation, although further experimental confirmation is needed to establish a direct link between CD24 status and increased clonogenic ability. These results highlight the importance of higher-resolution sequencing approaches to identify subpopulations within IDH-mutant tumors that may ultimately contribute to recurrence or treatment resistance. Our findings show that mutant IDH1 inhibition by doxycycline withdrawal leads to a modest decrease in tumor growth in a brain orthotopic model of IDH1 R132H-expressing IHAs. While inhibition of mutant IDH1 is likely to offer benefit, there is a need to identify and target additional pathways to control growth of IDH-mutated tumors.

In summary, the integrated data presented here provide new insights into transcriptional and epigenetic dynamics associated with *IDH1* mutations in unprecedented detail and may be useful for IDH-related studies and for the development of novel therapies for glioma.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at <https://doi.org/10.1038/s41588-017-0001-z>.

Methods

Cell culture

Immortalized human astrocytes were a gift from R. O. Pieper (UCSF) and were generated as previously described³³. IHAs were infected with pLVX-Tet-On retrovirus and selected with 800 µg/ml G418. Selected IHAs were infected with either empty pRetroX-hygro or pRetroX-hygro containing the coding sequence for IDH1 R132H cloned downstream of the Tet-responsive promoter and selected with 500 µg/ml hygromycin B. IHAs were cultured in the presence of doxycycline to induce expression of IDH1 R132H or in its absence (one passage equivalent to two doublings). Gliomaspheres (TS603, TS543, and TS667) were derived from patients undergoing tumor resection at Memorial Sloan Kettering Cancer Center (MSKCC). Tumors were obtained in accordance with institutional review board policies at MSKCC. Glioma tumorsphere lines were maintained in Neural Stem Cell (NSC) Basal Medium with NSC proliferation supplements, 10 ng/ml EGF, 20 ng/ml bFGF, and 2 µg/ml heparin (StemCell Technologies). All cell lines were routinely tested for mycoplasma and found to be negative. TS603 is a World Health Organization (WHO) grade III oligodendroglioma line with 1p/19q co-deletion and a mutation in endogenous *IDH1* encoding R132H, and TS543 and TS667 are malignant glioma tumorsphere lines and harbor *PDGFRA* amplifications^{7,34}.

FACS analysis

Inducible IHAs at passage 50 were stained for 30 min on ice in the dark with antibody to CD24 (BD Biosciences, 555427), and dead cells were excluded by DAPI staining. Cells were washed in PBS and analyzed using a FACSARIA III (Becton Dickinson).

Sample preparation

Genomic DNA was extracted with the DNeasy Blood and Tissue kit (Qiagen), and RNA was isolated with the RNeasy Plus Mini kit.

Histone extraction and mass spectrometry

Histones from 5 million cells per sample were extracted using a standard acid extraction protocol. Histone modification was quantified using a bottom-up mass spectrometry strategy as outlined previously³⁵. Briefly, the histone pellet washed in ice-cold acetone was air dried and then subjected to trypsin digestion and chemical derivatization using propionic anhydride as described previously³⁶. Histone peptides were then analyzed by nano-liquid chromatography coupled with triple-quadrupole mass spectrometry (Dionex UltiMate 3000 and Thermo Fisher Scientific TSQ Quantum) using the selected reaction monitoring (SRM/ MRM) method developed previously³⁷. Data were analyzed using Skyline software (MacCoss Lab, University of Washington)³⁸.

Proliferation assays

Cell proliferation was measured in duplicate using xCELLigence RTCA DP (ACEA Biosciences) (cell index). Cells were seeded at a density of 10,000 cells/well on 16-well plates (E-Plate, ACEA Biosciences) in a humidified chamber at 37 °C and 5% CO₂, and the

cell index was monitored for 150 h. Experimental results were exported using xCELLigence RTCA software (v 1.2.1).

Soft agar assays

Sorted cells (100,000 cells per sample) were plated in growth medium into six-well plates containing 0.65% top and bottom agar. Cells were plated in the middle layer in growth medium containing 0.40% agar. The medium covering the agar was refreshed every 3 d using medium with or without doxycycline (1 µg/ml). After 2–4 weeks, colonies were stained with 0.0005% crystal violet and quantified using a Gelcount colony counter (Oxford Optronix).

Bioluminescence imaging

Cell lines for use in orthotopic in vivo experiments were labeled with pHIV-Luc-ZsGreen (a gift from B. Welm (University of Utah), Addgene plasmid 39196). Transduced IHAs were sorted for ZsGreen expression. Bioluminescence imaging was performed weekly following intraperitoneal injection of d-luciferin, and signal was measured using the Xenogen IVIS Spectrum in vivo imaging system (PerkinElmer). Living Image software (PerkinElmer) was used to acquire and analyze the bioluminescence data.

Orthotopic transplantation experiments

All mouse experiments were approved by the Institutional Animal Care and Use Committee at MSKCC, and strict guidelines were enforced. Female *NOD-scid* IL2Rg^{null} mice (Jackson Laboratory) that were 5–6 weeks of age were intracranially injected with 5×10^5 cells overexpressing pHIV-Luc-ZsGreen using a fixed stereotactic apparatus (Stoelting). Injections were made to the right frontal cortex, 1.5 mm lateral and 1 mm caudal and at a depth of 2 mm with respect to bregma. Mice were fed either normal diet or doxycycline-containing chow (doxycycline hyclate diet formulated at 2,500 mg/kg; Envigo). Randomization of mice receiving the doxycycline diet was done as follows. Briefly, animals were ranked according to their bioluminescence signal and alternately assigned to group 1 (continued on the doxycycline diet) or group 2 (doxycycline diet withdrawn, continued on regular diet). Designation of diet conditions for either group was determined randomly in Microsoft Excel. Animals that were sick or died as a result of unrelated circumstances were excluded from further analyses. Lesions were analyzed by a board-certified neuropathologist (J.T.H.), and morphology, degree of infiltration, and cellularity were used as criteria for histopathology assessment.

FISH

Cell lines were harvested according to standard procedures. Briefly, cell lines were cultured with Colcemid (0.1 µg/ml) at 37 °C for 45 min, resuspended in 0.075 M KCl, incubated at 37 °C for 10 min, and then fixed in methanol:acetic acid (3:1) solution. FISH analysis was performed on the fixed cells using a four-color 1p36/1q23 and 19p12/19q13 homebrew probe mix. The probe mix consisted of BAC clones mapping to the following loci: 1p36.32 (RP4-785P20; labeled with Red dUTP), 1q23.3 (RP11-1038N13, RP11-1059C21; labeled with Aqua dUTP), 19p12 (RP11-359H18, CTD-2502P8; labeled with Orange dUTP), and

19q13.3 (CTD-2639E6, RP11-960B2; labeled with Green dUTP). Probe labeling, hybridization, washing, and fluorescence detection were performed according to standard procedures. Slides were scanned using a Zeiss Axioplan 2i epifluorescence microscope equipped with a megapixel CCD camera (CV-M4+CL, JAI) controlled by Isis 5.5.9 imaging software (MetaSystems Group). The entire hybridized area was scanned under 63×/100× to assess the quality of hybridization and the signal pattern. A minimum of 50 consecutive intact metaphases were captured, and signal counts (copy number) for each locus were recorded. A metaphase was considered to be positive for loss of 1p or 19q if the signal count (copy number) of the respective control locus was greater (relative loss).

Array-based genomic analysis

Expression analysis was performed using the Affymetrix U133 2.0 microarray, and methylation analysis was performed using the Illumina Infinium HumanMethylation450K bead array. Arrays were processed at the Integrated Genomics Operation (iGO) at the MSKCC according to the manufacturer's protocol. R statistical software³⁹ (v3.3.1) was used for data analysis. The ChAMP methylation pipeline (version 1.10) was used to extract and analyze data from idat files, and normalization was carried out using beta-mixture quantile normalization⁴⁰. Batch correction was performed using the ComBat algorithm⁴¹. β values were calculated by comparing Dox⁺ to Dox⁻ samples for IHAs with inducible expression of mutant IDH1 at the baseline passage. Loci with absolute $\beta > 0.1$ were considered to be aberrantly methylated and were binned as follows: low, $0.1 < \text{absolute } \beta < 0.2$; intermediate, $0.2 \leq \text{absolute } \beta < 0.4$; high, $\text{absolute } \beta \geq 0.4$. For overlap with ENCODE chromatin state models, we obtained chromatin core 15-state data for the NH-A ENCODE cell line (Epigenome ID E125).

Affymetrix CEL files were imported into R statistical software. Normalization was performed with the AffyPLM package in Bioconductor (v3.3), using RMA background correction, quantile normalization, and the Tukey biweight summary method. Batch correction was performed using the ComBat algorithm by incorporating batch information. To identify genes or loci specific to mutant IDH1, we excluded probes with equal or greater fold changes or β in Dox⁺ versus Dox⁻ inducible IHAs with empty vector.

IDH1-mutant inducible models and LGGs were correlated as follows. First, pathway enrichment of differentially expressed genes in IHAs with inducible expression of mutant IDH1 (Dox⁺ versus Dox⁻ across all passages) and CIMP⁺ versus CIMP⁻ LGGs⁵ was compared using compareCluster (with the enrichPathway function) in the clusterProfiler R package⁴². Second, hypermethylated probes ($\beta > 0.1$) at passages 1, 5, 10, 20, and 40 after baseline were used to plot the distribution of β values in CIMP⁺ and CIMP⁻ LGGs⁵. For gene expression in CD24-defined subpopulations ($n = 2$ replicates), normalization was performed as described above and differentially expressed genes were identified using the limma package. Genes with absolute fold change > 1 and FDR-adjusted P value < 0.05 were considered to be significantly differentially expressed.

To identify the CpG loci with the most impact on gene expression, the methylation probes with the most negative difference ($\beta < -0.1$) corresponding to upregulated genes or the most positive difference ($\beta > 0.1$) corresponding to downregulated genes were identified.

Genomic features associated with these probes were extracted from the Infinium HumanMethylation450K annotation file.

To derive gene expression clusters, genes with an absolute fold change > 1 ($\log_2$ scale) in Dox⁺ samples as compared to Dox⁻ samples at baseline were identified and their fold change was determined at consecutive doxycycline withdrawal time points by comparing to baseline Dox⁻ samples. We define genes that returned to baseline values (absolute fold change < 1 when compared to Dox^{off} samples) within 5 passages as ‘transient’, within 20 passages as ‘gradual’, and that remained aberrantly expressed at 40 passages as ‘persistent’. Aberrantly expressed genes at baseline (absolute fold change > 1) were binned as follows: low, absolute fold change < 1.5 ; intermediate, $1.5 < \text{absolute fold change} < 2$; high, absolute fold change > 2 . The significance of fold changes across clusters was determined using Fisher’s exact test for count data. As for gene expression clusters, we derived methylation clusters with similar kinetics following doxycycline withdrawal. To this end, we considered loci with absolute $\beta > 0.1$ as aberrantly methylated at baseline and set an absolute β cutoff of 0.1 for cluster binning. In addition, subgroups of loci within the persistent cluster that maintained absolute $\beta > 0.1$ at passage 40 following doxycycline withdrawal were identified.

Copy number alterations

Copy number alterations were called from the Infinium HumanMethylation450K array using the ChAMP or conumee package in R statistical software^{40,43}. Segmented data were visually inspected and subsequently used to identify broad and focal copy number alterations in GISTIC2.0⁴⁴ on the GenePattern server with the --armpeel and --broad parameters set to “yes”.

Magnetic resonance imaging

The brains of injected mice were scanned on a 200-MHz Bruker 4.7 T Biospec MRI scanner (Bruker Biospin) equipped with a 300 mT/m 20-cm (ID) gradient (Resonance Research). Mice were anaesthetized with 2% isoflurane in oxygen. Sedated animals were physiologically monitored during a scan period (SA Instruments). For mouse brain imaging, axial T2-weighted brain images were acquired using the fast spin-echo RARE sequence (rapid acquisition with relaxation enhancement) with TR 1.5 s, TE 50 ms, RARE factor 8, slice thickness 0.7 mm, field of view 30×20 mm, in-plane resolution $117 \times 125 \mu\text{m}$, and 24 averages.

Chromatin immunoprecipitation

Cells were cross-linked in 1% formaldehyde for 15 min at room temperature with constant agitation, followed by quenching with 125 mM glycine for 5 min at room temperature. ChIP assays were performed with the EZ-Magna ChIP kit (Millipore, 17–611). Briefly, nuclei were isolated and chromatin was sheared using a Bioruptor (Diagenode) (30-s pulses, 30-s rest between pulses, power setting 10). Shearing efficiency was confirmed with a Bioanalyzer (Agilent). Chromatin was incubated with primary antibody overnight at 4 °C with constant agitation. DNA–protein immune complexes were eluted, and cross-linking was reversed. DNA extraction was performed using a spin filter column. Immunoprecipitated and

input DNA samples were subjected to sequencing on an Illumina HiSeq 2500 instrument. Prior to sequencing, realtime qPCR was performed to confirm enrichment of targets in ChIP samples as compared to input samples. ChIP was performed using the following antibodies: H3K4me3 (Millipore, 17–614, lot NG1848343), H3K27me3 (Millipore, 17–622, lot 1987188), H3K36me3 (Abcam, ab9050, lot GR52625-1), H3K9me3 (Abcam, ab8898, lot GR47224-2), and H4K20me3 (Abcam, ab9053, lot GR2463.2).

RNA-seq

RNA-seq was performed at iGO, MSKCC. Raw reads in fastq format were aligned to hg19 using RNA-seq STAR aligner version 2.4.0d^{45,46}. The STAR aligner was chosen for mapping accuracy and speed⁴⁷. Mapped reads for each sample were counted for each gene in annotation files in GTF format (gencode.v19. annotation.gtf) using the FeatureCounts read summarization program⁴⁸. Individual count files were merged by an in-house R script and normalized using the DESeq R package. To generate heat maps, count data were transformed using the rlog (regularized log transformation) function from the DESeq2 R package, and visualized and clustered (with default parameters) using the pheatmap R package. Expression of transcribed ERVs was determined as described²⁴, and rlog counts were visualized using the pheatmap R package.

Public data acquisition and processing

Level 3 (processed and normalized) RNA-seq data along with clinical annotations for brain LGGs were downloaded from Broad GDAC Firehose. Gene expression for the viral defense signature²⁵ for each sample was visualized using the pheatmap package in R. *LICAM* and *MEOX2* expression data were extracted, and the significance of differences between IDH-mutant and IDH-wild-type LGGs was assessed by Student's *t* test.

ChIP-seq

ChIP-seq was performed at iGO, MSKCC. Raw sequencing data were aligned to the hg37 genome build using BWA (v.0.7.10)⁴⁹. Further indel realignment, base quality score recalibration, and duplicate read removal were performed using the Genome Analysis Toolkit (GATK) (version 3.2.2)^{50,51}. Multiple classes of metrics (CollectAlignmentSummaryMetrics, CollectInsertSizeMetrics, QualityScoreDistribution, MeanQualityByCycle, and CollectBaseDistributionByCycle) were collected with Picard's CollectMultipleMetrics to ensure alignment quality. Aligned ChIP-seq files from replicates ($n = 2$) were merged. To call enriched regions, we first applied MACS2 (version 2.1.0)⁵² to all histone marks. For narrow histone marks (H3K4me3), resulting raw peaks were filtered by $P = 0.01$. Next, for each peak, RPKM (reads per kilobase of transcript per million mapped reads) was calculated for immunoprecipitated (IP) and input bam files. Only peaks with RPKM(IP)/ RPKM(input) ratio greater than 0.5 are reported in the final results. For broad histone marks (H3K36me3, H3K9me3, H3K79me2, H3K27me3, and H4K20me3), resulting raw peaks were filtered by $P = 0.1$. To ensure high stringency of broad peak calling, we applied a second peak caller called SICER (version 1.1)⁵³. Two sets of peaks were merged using the bedtools merge tool, and only intersecting broad peaks are reported. BigWig files were generated using bamCoverage from the deepTools suite with options --binSize 10 --normalizeTo 1×2451960000 --extendReads --centerReads⁵⁴. The log₂-

transformed ratios of ChIP signal to input were calculated using bamCompare from deepTools. Resulting normalized BigWig files were used as input to computeMatrix to calculate scores for regions of interest (using either scale-regions or reference-point mode) and visualized using either plotHeatmap or plotProfile tools from deepTools. BigWig files were visualized using the gviz package in R. The bedtools merge tool was used to create a union of all peaks for each histone mark.

Functional enrichment based on proximal gene annotations was determined using GREAT v. 3.0.0 with default enrichment settings⁵⁵. Tag counts $\pm 2,000$ bp around TSSs were determined using the annotatePeaks HOMER script. Tag counts > 32 and with absolute fold change of at least 1 between conditions were used for further analyses. Differential peaks were analyzed using the getDifferentialPeaks HOMER script with default settings, and regions with fold change > 10 , Poisson P value < 0.0001 , and fold change > 1.5 were determined as significant. HOMER's "Basic Annotation" column was parsed, and counts per annotation were plotted as bar charts. The "noncoding" annotation in HOMER refers to exons of noncoding RNAs within the RefSeq NR subset.

Visualization of ChIP-seq peaks—Peaks were visualized by Hilbert curves, using the HilbertVis R package¹⁸. Three-channel RGB images (EBImage::rgbImage R function) of Hilbert curves for mutant-IDH1-expressing and parental cell lines were overlaid on one panel, visualized, and compared across all passages.

Chromatin state discovery—ChIP-seq data and their corresponding input bam files were binarized with a 200-bp bin size using custom script to convert a set of bam files into binarized data matrices suitable for ChromHMM model learning tools²⁰. The LearnModels tool was applied to learn a chromatin state model. Emission parameters were saved in textual and graphical format.

Functional annotation of ChIP-seq peaks—ChIP-seq peaks generated by MACS2 were annotated by HOMER (v4.4)⁵⁶. First, MACS2 bed files were reformatted to HOMER peak file format, and the annotatePeaks script was then applied to determine whether a peak was in a TSS, TTS, exonic (coding), 5' UTR exon, 3' UTR exon, intronic, intergenic, or CpG island region.

2-HG analysis

2HG levels were determined by mass spectrometry as previously described⁵.

RNA isolation, reverse transcription, and qPCR for ERV quantification

Total RNA was isolated from 5×10^6 cells (inducible IHAs at passage 70, $n = 3$ cell culture replicates per group) using the RNeasy Plus Mini kit (Qiagen). 2 μ g of total RNA was used for reverse transcription with the SuperScript III One-Step RT-PCR system with random hexamers (Invitrogen). cDNA was used as template for qPCR to determine expression of ERVs on a QuantStudio 6 Flex Real-Time PCR console (Thermo Fisher). Results were normalized to 18 S rRNA. Primer sequences are listed in Supplementary Table 7.

Statistics

To identify differentially expressed genes between sorted CD24 samples, the limma R package was used and genes with absolute fold change > 1 and FDR-adjusted P value < 0.05 were considered to be significantly differentially expressed. GSEA was performed using the GSEA module within GenePattern with either the Hallmark collection or positional gene sets (c1) and the permutation method set to gene_set^{57,58}. Gene sets with FDR q value < 0.25 were considered to be significantly enriched. Student's t tests and Fisher's exact tests were performed as two-sided tests. Differential binding analysis for comparison of H3K4me3 peaks across IDH1-wild-type and IDH1-mutant models was performed using the DiffBind R package with method = DNA_DESEQ2_BLOCK and P value < 0.05 as the cutoff. To visualize RNA-seq heat maps, rlog transformed count data were visualized and clustered (with default parameters) using the pheatmap R package. Functional enrichment based on proximal gene annotations was determined using GREAT v.3.0.0 default enrichment settings. For narrow histone marks raw peaks were filtered by $P = 0.01$ and for broad histone marks raw reads were filtered by $P = 0.1$ before further processing. For ChIP-seq heat maps, data were grouped using k -means clustering into $k = 6$ groups. Empirical P values were calculated after 10,000 independent random trials and estimated as $(r + 1)/(n + 1)$, where n is the number of simulated samples and r is the number of values for simulated samples greater than or equal to the median of a given methylation probe set.

Tracing of ZsGreen-labeled immortalized human astrocytes with serial two-photon tomography

For STPT experiments, mice orthotopically injected with ZsGreen-labeled IHAs were perfused transcardially with ice-cold PBS followed by ice-cold 4% paraformaldehyde ($n = 5$ for Dox⁻ and Dox⁺ mice, $n = 8$ for Dox^{off} mice). After a 24-h post-fix in 4% paraformaldehyde, brains were kept in 0.7% glycine solution for 48 h and then stored in PBS at 4 °C. Whole brains were embedded in 4% agarose in 0.05 M PB, and cross-linking was performed in 0.2% sodium borohydrate solution for 2–3 h. Whole brains were sliced in 50- μ m sections using a high-speed two-photon microscope with integrated vibratome sectioning (x - y resolution of 1 μ m; z step of 50 μ m; TissueCyte 1000, TissueVision-STP) as described⁵⁹. Raw image files were corrected for illumination, stitched in 2D, and aligned in 3D. We employed a two-channel configuration in which channel 1 detected background signal and channel 2 detected ZsGreen signal. The background channel was employed as a subtracted channel to reduce the signal/noise ratio for the signal channel. ZsGreen-positive cells from the subtracted channel automatically detected by a convolutional network trained to recognize nuclear cell body labeling⁶⁰ were visually validated and reconstructed in 3D, and their spatial information was registered by affine followed by B-spline transformation using the software Elastix⁶¹ to a 3D reference brain based on the Allen Brain Atlas^{29,30}. The total number of cells in each brain region was normalized by the total number of positive cells within the superior hierarchical structure as defined by the Allen Brain Atlas (for example, cortex) to account for variability in the total number dependent on where each tumor grew. Brain 3D reconstruction was done using Imaris version 7.6.5 (Bitplane).

Life Sciences Reporting Summary

Further information on experimental design is available in the Life Sciences Reporting Summary.

Data availability

All data have been deposited in the Gene Expression Omnibus under accession GSE85942.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the members of the Chan and Thompson laboratories for helpful discussions. This work was supported in part by the US National Institutes of Health (NIH; R01 CA177828) (T.A.C. and C.B.T.), the MSKCC Brain Tumor Center (S.T. and T.A.C.), the Sontag Foundation (T.A.C.), the PaineWebber Chair Endowment (T.A.C.), NIH T32 grant 5T32CA160001 (S.T.), the MSKCC Society (T.A.C.), the NIH (R01 MH096946) (P.O.), and NIH Cancer Center Support Grant P30CA008748 (G.N.). This research was carried out in collaboration with the National Resource for Translational and Developmental Proteomics under grant P41 GM108569 (N.L.K.) from the National Institute of General Medical Sciences, NIH.

References

1. Parsons DW, et al. An integrated genomic analysis of human glioblastoma multiforme. *Science*. 2008; 321:1807–1812. [PubMed: 18772396]
2. Yan H, et al. *IDH1* and *IDH2* mutations in gliomas. *N. Engl. J. Med.* 2009; 360:765–773. [PubMed: 19228619]
3. Dang L, et al. Cancer-associated *IDH1* mutations produce 2-hydroxyglutarate. *Nature*. 2009; 462:739–744. [PubMed: 19935646]
4. Lu C, et al. *IDH* mutation impairs histone demethylation and results in a block to cell differentiation. *Nature*. 2012; 483:474–478. [PubMed: 22343901]
5. Turcan S, et al. *IDH1* mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature*. 2012; 483:479–483. [PubMed: 22343889]
6. Duncan CG, et al. A heterozygous *IDH1*^{R132H/WT} mutation induces genome-wide alterations in DNA methylation. *Genome Res.* 2012; 22:2339–2355. [PubMed: 22899282]
7. Rohle D, et al. An inhibitor of mutant IDH1 delays growth and promotes differentiation of glioma cells. *Science*. 2013; 340:626–630. [PubMed: 23558169]
8. Tateishi K, et al. Extreme vulnerability of *IDH1* mutant cancers to NAD⁺ depletion. *Cancer Cell*. 2015; 28:773–784. [PubMed: 26678339]
9. Bao S, et al. Targeting cancer stem cells through L1CAM suppresses glioma growth. *Cancer Res.* 2008; 68:6043–6048. [PubMed: 18676824]
10. Ducray F, et al. Anaplastic oligodendrogliomas with 1p19q codeletion have a proneural gene expression profile. *Mol. Cancer*. 2008; 7:41. [PubMed: 18492260]
11. Izumoto S, et al. Gene expression of neural cell adhesion molecule L1 in malignant gliomas and biological significance of L1 in glioma invasion. *Cancer Res.* 1996; 56:1440–1444. [PubMed: 8640837]
12. Maness PF, Schachner M. Neural recognition molecules of the immunoglobulin superfamily: signaling transducers of axon guidance and neuronal migration. *Nat. Neurosci.* 2007; 10:19–26. [PubMed: 17189949]
13. Mimeault M, Batra SK. Molecular biomarkers of cancer stem/progenitor cells associated with progression, metastases, and treatment resistance of aggressive cancers. *Cancer Epidemiol. Biomarkers Prev.* 2014; 23:234–254. [PubMed: 24273063]

14. Fang X, Zheng P, Tang J, Liu Y. CD24: from A to Z. *Cell. Mol. Immunol.* 2010; 7:100–103. [PubMed: 20154703]
15. Miller JA, et al. Transcriptional landscape of the prenatal human brain. *Nature.* 2014; 508:199–206. [PubMed: 24695229]
16. Mohanan V, Temburni MK, Kappes JC, Galileo DS. L1CAM stimulates glioma cell motility and proliferation through the fibroblast growth factor receptor. *Clin. Exp. Metastasis.* 2013; 30:507–520. [PubMed: 23212305]
17. Kleene R, Yang H, Kutsche M, Schachner M. The neural recognition molecule L1 is a sialic acid-binding lectin for CD24, which induces promotion and inhibition of neurite outgrowth. *J. Biol. Chem.* 2001; 276:21656–21663. [PubMed: 11283023]
18. Anders S. Visualization of genomic data with the Hilbert curve. *Bioinformatics.* 2009; 25:1231–1235. [PubMed: 19297348]
19. Verhaak RG, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in *PDGFRA*, *IDH1*, *EGFR*, and *NF1*. *Cancer Cell.* 2010; 17:98–110. [PubMed: 20129251]
20. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods.* 2012; 9:215–216. [PubMed: 22373907]
21. Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015; 518:317–330. [PubMed: 25693563]
22. Bailey SD, et al. ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters. *Nat. Commun.* 2015; 2:6186. [PubMed: 25645053]
23. Bai H, et al. Integrated genomic characterization of *IDH1*-mutant glioma malignant progression. *Nat. Genet.* 2016; 48:59–66. [PubMed: 26618343]
24. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell.* 2015; 160:48–61. [PubMed: 25594174]
25. Chiappinelli KB, et al. Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell.* 2016; 164:1073. [PubMed: 27064190]
26. Roulois D, et al. DNA-demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous transcripts. *Cell.* 2015; 162:961–973. [PubMed: 26317465]
27. Scheie D, et al. Prognostic variables in oligodendroglial tumors: a single-institution study of 95 cases. *Neuro Oncol.* 2011; 13:1225–1233. [PubMed: 21856683]
28. Koivunen P, et al. Transformation by the (*R*)-enantiomer of 2-hydroxyglutarate linked to EGLN activation. *Nature.* 2012; 483:484–488. [PubMed: 22343896]
29. Kim Y, et al. Mapping social behavior-induced brain activation at cellular resolution in the mouse. *Cell Rep.* 2015; 10:292–305. [PubMed: 25558063]
30. Sunkin SM, et al. Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res.* 2013; 41:D996–D1008. [PubMed: 23193282]
31. Turcan S, et al. Efficient induction of differentiation and growth inhibition in *IDH1* mutant glioma cells by the DNMT inhibitor decitabine. *Oncotarget.* 2013; 4:1729–1736. [PubMed: 24077826]
32. Flavahan WA, et al. Insulator dysfunction and oncogene activation in *IDH* mutant gliomas. *Nature.* 2016; 529:110–114. [PubMed: 26700815]
33. Sonoda Y, et al. Formation of intracranial tumors by genetically modified human astrocytes defines four pathways critical in the development of human anaplastic astrocytoma. *Cancer Res.* 2001; 61:4956–4960. [PubMed: 11431323]
34. Silber J, et al. miR-34a repression in proneural malignant gliomas upregulates expression of its target *PDGFRA* and promotes tumorigenesis. *PLoS One.* 2012; 7:e33844. [PubMed: 22479456]
35. Zheng Y, Huang X, Kelleher NL. Epiproteomics: quantitative analysis of histone marks and codes by mass spectrometry. *Curr. Opin. Chem. Biol.* 2016; 33:142–150. [PubMed: 27371874]
36. Garcia BA, et al. Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nat. Protoc.* 2007; 2:933–938. [PubMed: 17446892]
37. Zheng Y, Tipton JD, Thomas PM, Kelleher NL, Sweet SM. Site-specific human histone H3 methylation stability: fast K4me3 turnover. *Proteomics.* 2014; 14:2190–2199. [PubMed: 24826939]

38. MacLean B, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*. 2010; 26:966–968. [PubMed: 20147306]
39. R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna; Austria: 2008.
40. Morris TJ, et al. ChAMP: 450K chip analysis methylation pipeline. *Bioinformatics*. 2014; 30:428–430. [PubMed: 24336642]
41. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007; 8:118–127. [PubMed: 16632515]
42. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012; 16:284–287. [PubMed: 22455463]
43. Feber A, et al. Using high-density DNA methylation arrays to profile copy number alterations. *Genome Biol*. 2014; 15:R30. [PubMed: 24490765]
44. Mermel CH, et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol*. 2011; 12:R41. [PubMed: 21527027]
45. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013; 29:15–21. [PubMed: 23104886]
46. Dobin A, Gingeras TR. Mapping RNA-seq reads with STAR. *Curr. Protoc. Bioinformatics*. 2015; 51:11.14.1–11.14.19. [PubMed: 26334920]
47. Engström PG, et al. Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat. Methods*. 2013; 10:1185–1191. [PubMed: 24185836]
48. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014; 30:923–930. [PubMed: 24227677]
49. Li H, Durbin R. Fast and accurate short read alignment with Burrows- Wheeler transform. *Bioinformatics*. 2009; 25:1754–1760. [PubMed: 19451168]
50. McKenna A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010; 20:1297–1303. [PubMed: 20644199]
51. DePristo MA, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet*. 2011; 43:491–498. [PubMed: 21478889]
52. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. *Nat. Protoc*. 2012; 7:1728. [PubMed: 22936215]
53. Zang C, et al. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics*. 2009; 25:1952–1958. [PubMed: 19505939]
54. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*. 2014; 42:W187–W191. [PubMed: 24799436]
55. McLean CY, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol*. 2010; 28:495–501. [PubMed: 20436461]
56. Heinz S, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*. 2010; 38:576–589. [PubMed: 20513432]
57. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA*. 2005; 102:15545–15550. [PubMed: 16199517]
58. Reich M, et al. GenePattern 2.0. *Nat. Genet*. 2006; 38:500–501. [PubMed: 16642009]
59. Ragan T, et al. Serial two-photon tomography for automated ex vivo mouse brain imaging. *Nat. Methods*. 2012; 9:255–258. [PubMed: 22245809]
60. Turaga SC, et al. Convolutional networks can learn to generate affinity graphs for image segmentation. *Neural Comput*. 2010; 22:511–538. [PubMed: 19922289]
61. Klein S, Staring M, Murphy K, Viergever MA, Pluim JP. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans. Med. Imaging*. 2010; 29:196–205. [PubMed: 19923044]

Fig. 1. Characterization of IDH1-mutant-induced gene expression reversibility

a, Summary of the cellular models used to study IDH1 R132H function and the epigenomic and transcriptomic datasets produced in this study. **b**, IHAs transduced with inducible empty vector (EV) or vector encoding wild-type (WT) or R132H (MUT) IDH1 were grown for 30 passages (baseline) in the presence (+) or absence (–) of 1 μ g/ml doxycycline. Doxycycline was withdrawn from the medium for either two (top) or ten (bottom) passages after baseline (off). Western blot analysis of inducible IHAs confirms expression of IDH1 R132H only in the presence of doxycycline. **c**, Total intracellular 2HG was measured by gas chromatography coupled with mass spectrometry and normalized to an internal standard

(D5-2HG) and cell number. 2HG levels are increased upon doxycycline administration and are completely inhibited at 15 passages following doxycycline withdrawal after baseline. **d**, Schematic of the reversibility model for inducible expression of IDH1 R132H in astrocytes. P, passage. **e**, Hierarchical clustering of global gene expression profiles of IHAs as shown in **d**. **f**, Heat maps of downregulated (left) and upregulated (right) gene expression clusters following doxycycline withdrawal at baseline. Pers., persistent. The color scale shows normalized expression levels. **g**, Gene expression time course for several persistently upregulated (top) and downregulated (bottom) genes. **h**, Line plots show average expression levels of three Affymetrix probes targeting *CD24* over time. **i**, Percentages of genomic features for the methylated loci exhibiting the most inverse correlation with gene expression data in $CD24^+$ as compared to $CD24^-$ IDH1 R132H-expressing astrocytes. Hyper/down, hypermethylated and downregulated genes; hypo/up, hypomethylated and upregulated genes. IGR, intergenic region. **j**, Average expression of example genes upregulated in $CD24^+$ cells measured in BrainSpan Developmental Transcriptome data. RPKM, reads per kilobase of transcript per million mapped reads; pcw, post-conception weeks; mos, months; yrs, years.

Fig. 2. Remodeling of chromatin states and the epigenomic landscape by mutant IDH1

a, Genome coverage (left) and number of peaks (right) for histone marks in IHAs stably expressing IDH1 R132H at passages 2, 10, and 40. **b**, Genome coverage (left) and number of peaks (right) for histone marks in parental IHAs at passages 2, 10, and 40. **c**, Heat maps displaying log₂-transformed input-normalized ChIP signal for *k*-means clusters (*k* = 6) of the union of H3K4me3 (left), H3K9me3 (middle), and H3K36me3 (right) peaks for parental IHAs (Par) and all IHAs stably expressing mutant IDH1 (Mut). Scaled regions from peak start to peak end are displayed. The color scale indicates ChIP signal ranging from no enrichment (dark blue) to high enrichment (dark red). **d**, H3K9me3 (top) and H3K36me3

(bottom) $\log_2$ -transformed input-normalized ChIP signal profiles for the clusters identified in **c**. Scaled regions from peak start to peak end are displayed. **e**, HOMER-derived annotations of the genomic loci covered by H3K4me3 peaks in parental IHAs and IHAs stably expressing IDH1 R132H at passages 2, 10, and 40. TTS, transcription termination site. **f**, H3K4me3 read coverage normalized to $1\times$ sequencing depth at the *PDGFRA* TSS and *DANCR* locus for parental IHAs and IHAs with stable expression of mutant IDH1. The range of the y axis is indicated and fixed for both plots. **g**, Number of called peaks in inducible IHAs at baseline (passage 30, Dox⁻ and Dox⁺) and 40 passages following doxycycline withdrawal from baseline (Dox^{off} 40 P). **h,i**, Chromatin states learned jointly across all IHAs stably expressing IDH1 R132H (**h**) and inducible IHAs (Dox⁻, Dox⁺, Dox^{off} 40 P) (**i**). Plots show emission parameters (%) learned on the basis of combinations of chromatin marks, genomic annotation, genome coverage (%), and transition parameters. Emission parameters correspond to the probability of a given histone mark occurring in a particular chromatin state. Shading indicates relative fold enrichment. LAD, lamina-associated domains.

Fig. 3. Reversibility and kinetics of DNA methylation changes in astrocytes inducibly expressing IDH1 R132H

a, Hierarchical clustering of global methylation profiles in inducible IHAs. **b**, Distribution of β values (compared to baseline without doxycycline) for hypo- and hypermethylated loci at baseline (top panel) and subsequent passages after doxycycline withdrawal. **c**, Heat maps of hypermethylated (left) and hypomethylated (right) clusters following doxycycline withdrawal. At passage 40 following doxycycline withdrawal, 20% of hypermethylated and 30% of hypomethylated loci within the persistent clusters maintain their altered methylation state with equal or greater absolute β -value differences as compared to baseline. The color scale shows normalized methylation values ranging from low (dark blue) to high (red). **d**,

Number of hypomethylated (blue) and hypermethylated (red) loci falling within the transient, gradual, and persistent clusters. **e**, Table showing the total number of hypermethylated (top) and hypomethylated (bottom) loci binned by β -value differences (at baseline) across the transient, gradual, and persistent methylation clusters. **f**, Overlap of hypermethylated and hypomethylated loci in Dox⁺ versus Dox⁻ IHAs at baseline with the 15-state ChromHMM annotations of the Roadmap Epigenomics project. ENCODE state annotations are noted below for the 15 states. **g**, Distribution of β values (Dox⁺ versus Dox⁻ at baseline) for the loci corresponding to up- and downregulated genes (Dox⁺ versus Dox⁻ at baseline) within the transient, gradual, and persistent clusters stratified by genomic feature. None of the persistently upregulated genes fall within 3' UTRs or intergenic regions. **h**, Log₂-transformed input-normalized H3K4me3, H3K27me3, H3K36me3, and H3K9me3 ChIP signal profiles (top) and heat maps (bottom) ± 1 kb around hypomethylated CpG sites within the transient (dark blue), gradual (light blue), and persistent (yellow) clusters for Dox⁻, Dox⁺, and Dox^{off} IHAs. The color scale indicates ChIP signal ranging from no enrichment (dark blue) to high enrichment (dark red). **i**, Overlap of hypermethylated and hypomethylated clusters with chromatin states in Dox⁻ IHAs. "Background" corresponds to overlap of all probes on the Illumina HumanMethylation450K array with Dox⁻ chromatin states.

Fig. 4. Characterization of H3K4me3 dynamics in IDH1 R132H models

a, HOMER annotation of genomic regions within chromatin state 5 in inducible IHAs. *n* indicates the number of genome segments in state 5. The color scale indicates the percentage of total segments within each annotation category and ranges from white (0%) to dark blue (100%). **b**, Chromatin state transitions from Dox⁻ to Dox⁺ IHAs. Numbers inside the cells correspond to the z score of the genomic region covered by each transitioned state in Dox⁺ cells. The color scale ranges from low coverage (dark blue) to high coverage (red). **c**, Sequence-depth-normalized H3K4me3 and RNA-seq coverage tracks for *MAP1LC3A* show persistent reduction of H3K4me3 at its promoter and loss of gene expression over the exonic

regions. **d**, Profiles (top) and heat maps (bottom) of $\log_2$ -transformed input-normalized H3K4me3 ChIP-seq signal ± 1 kb around the start and end sites of the top 11,443 regions with significant gains in H3K4me3 in Dox⁺ as compared to Dox⁻ IHAs. The color scale indicates ChIP signal ranging from no enrichment (dark blue) to high enrichment (dark red). Start, peak start; end, peak end. **e**, Distribution of the locations for the top 11,443 differentially enriched H3K4me3 peaks with respect to their nearest TSSs. **f**, Pie chart showing the distribution of HOMER-annotated genomic features for the top 11,443 differentially enriched H3K4me3 peaks. **g**, GSEA (using GREAT toolbox) of the top 2,000 TSSs with the highest H3K4me3 enrichment across parental IHAs and IHAs stably expressing IDH1 R132H (Mutant) at passage 40, in inducible Dox⁻ and Dox⁺ IHAs, in Dox^{off} passage 40 IHAs, and in TS603 tumorspheres. Color corresponds to the multiple-testing-adjusted q value of enrichment. **h**, Profiles (top) and heat maps (bottom) of $\log_2$ -transformed input-normalized H3K4me3 ChIP signal ± 1 kb around the start and end sites of the top 1,150 significant regions in the IDH1-mutant models (compared to IDH-wild-type cells) along with Dox^{off} passage 40 IHAs for the same regions. The color scale indicates ChIP signal ranging from no enrichment (dark blue) to high enrichment (dark red).

Fig. 5. Upregulated viral response and genomic instability in cells expressing IDH1 R132H
a, GSEA of the IFN- α signaling profile showing that the interferon signaling pathway is significantly upregulated in Dox⁺ as compared to Dox⁻ IHAs. NES, normalized enrichment score; FDR, false discovery rate. **b**, GSEA of the IFN- α signaling profile showing that the interferon signaling pathway is significantly upregulated in Dox^{off} states as compared to Dox⁻ IHAs. **c**, Heat map of RNA-seq-derived ERV expression for astrocytes with empty vector (green) or inducible expression of IDH1 R132H (purple) at baseline (Dox⁻, Dox⁺), 40 passages following baseline (Dox⁻ 40 P, Dox⁺ 40 P), 1 passage or 40 passages following doxycycline withdrawal (Dox^{off} 1 P, Dox^{off} 40 P), and tumorspheres (TS543, TS603,

TS667) (gray). The scale corresponds to the z scores of regularized log-transformed read counts. Regularized log-transformed (rlog) is a function that transforms RNA-seq read counts. **d**, qPCR validation of select ERVs (from **c**) with increased expression in Dox⁺ IHAs at passage 70. Error bars, s.d; $n = 3$; $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$; n.s., not significant (unpaired t test). **e**, Heat map of viral defense signature genes²⁵ for astrocytes with empty vector (green) or inducible expression of IDH1 R132H (purple) at baseline (Dox⁻, Dox⁺), 40 passages following baseline (Dox⁻ 40 P, Dox⁺ 40 P), 1 passage or 40 passages following doxycycline withdrawal (Dox^{off} 1 P, Dox^{off} 40 P), and tumorspheres (TS543, TS603, TS667) (gray). Scale corresponds to the z scores of regularized log-transformed read counts. **f**, qPCR analysis of select ERVs involved in antiviral response showing increased expression in Dox⁺ IHAs at passage 70. Error bars, s.d; $n = 3$; $*P < 0.05$, $**P < 0.01$, $***P < 0.001$; n.s., not significant (unpaired t test). **g**, GSEA showing that genes on chromosome 19q13 are concordantly and significantly downregulated in Dox⁺ as compared to Dox⁻ IHAs. **h**, Copy number profile derived from Illumina HumanMethylation450K arrays for astrocytes expressing IDH1 R132H at baseline (compared to Dox⁻ astrocytes). Red denotes amplification, and blue denotes loss. **i**, Quantification of metaphases with relative loss of 19q in Dox⁺ and Dox⁻ inducible IHAs at baseline (P30) and five passages after baseline (P35).

Fig. 6. Phenotype and tumor growth dynamics associated with mutant IDH1

a, In vitro proliferation of inducible astrocytes (Dox⁻, Dox⁺, and Dox^{off}; the number of passages following doxycycline withdrawal is specified for each plot). Error bars, s.e.m.; $n = 2$. **b**, Serial bioluminescence imaging tracking the growth of inducible astrocytes (Dox⁻, $n = 10$ mice; Dox⁺, $n = 20$ mice). The arrow indicates the point at which the Dox⁺ group was randomized into Dox^{off} (doxycycline withdrawn; $n = 10$) and Dox⁺ (continued on doxycycline diet; $n = 10$) mice. Each time point shows the average radiance across groups (mean $\pm$ s.e.m.). **c**, Inducible IHAs labeled with ZsGreen and imaged with STPT. Bright areas of coronal sections correspond to ZsGreen positivity; Allen Brain Atlas coordinates are

shown at the bottom. **d**, Left, whole brain surface reconstructions from 280 brain sections imaged with STPT using 20% resolution for raw data; scale bar, 500 μm . Middle panels, 3D tumor reconstructions for ZsGreen-labeled cells showing tumor location and growth (sagittal (left) and horizontal (right) views). Right, reconstructed tumors; scale bar, 250 μm . **e**, Semiquantitative prediction of cell numbers in various brain regions, indicating spread of fluorescent cells from a representative IDH1 R132H-expressing (Dox^+) tumor in comparison to a representative tumor where IDH1 R132H expression was eliminated (Dox^{off}). MO, somatomotor areas; SS, somatosensory areas; ILA, infralimbic area; CP, caudoputamen; ACB, nucleus accumbens; FS, fundus of striatum; OT, olfactory tubercle; AAA, anterior amygdalar area; Palv, pallidum, ventral region; LHA, lateral hypothalamic area; LPO, lateral preoptic area.