Mutant IDH1-Dependent Chromatin State Reprogramming, Reversibility, and Persistence

Sevin Turcan[1 *#], Vladimir Makarov[1], Julian Taranda[2], Yuxiang Wang[1], Armida W.M. Fabius[1¶], Wei Wu[1], Yupeng Zheng[3], Nour El-Amine[2], Sara Haddock[1, 4], , Gouri Nanjangud[5], H. Carl LeKaye[6], Cameron Brennan[7], Justin Cross[8], Jason T. Huse[1], Neil L. Kelleher[9], Pavel Osten[2], Craig B. Thompson[10], and Timothy A. Chan[1, 4, 11*]

[1]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065

[2]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724

[3]Sanofi Genzyme, 153 Second Avenue, Waltham, MA 02451

[4]Weill Cornell School of Medicine, New York, NY 10065

[5]Molecular Cytogenetics Core Facility, Memorial Sloan Kettering Cancer Center, New York, New York 10065

[6]Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY 10065

[7]Department of Neurosurgery, Memorial Sloan Kettering Cancer Center, New York, New York 10065

[8]Donald B. and Catherine C. Marron Cancer Metabolism Center, Memorial Sloan Kettering Cancer Center, New York, New York 10065

[9]Departments of Chemistry and Molecular Biosciences, Proteomics Center of Excellence, Northwestern University, Evanston, Illinois 6061

[10]Cancer Biology and Genetics Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065

[11]Immunogenomics and Precision Oncology Platform, Memorial Sloan Kettering Cancer Center, New York, NY 10065

[#]Present address: Department of Neurology and National Center for Tumor Diseases, University Hospital Heidelberg, Heidelberg, Germany
[¶] Present address: Department of Ophthalmology, VU Medical Center, Amsterdam, The Netherlands
*Correspondence to T.A.C. (chant@mskcc.org) or S.T. (sevin.turcan@med.uni-heidelberg.de)

ABSTRACT

Isocitrate dehydrogenase 1 and 2 (IDH1 and IDH2) mutations drive the development of gliomas and other human malignancies. Mutant IDH1 induces epigenetic changes that promote tumorigenesis but their scale and reversibility 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 alterations in DNA methylation, 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 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 significant implications for understanding of mutant IDH function and for optimizing therapeutic approaches to targeting IDH mutant tumors.

Mutations in isocitrate dehydrogenase 1 and 2 (IDH1 and IDH2) are found in over 80% of lower-grade gliomas (LGGs) and secondary glioblastomas [1,2]. IDH1 mutations commonly occur at Arg132, and produce 2-hydroxyglutarate (2HG) [3]. In experimental models, mutant IDH1 introduction and 2HG production leads to global DNA hypermethylation, histone methylation alterations and differentiation block [4-6]. Pharmacological inhibition of 2HG levels with mutant selective IDH inhibitors has not completely reversed mutant IDH-dependent epigenetic changes *in vitro* or *in vivo* at least for the treatment durations tested [7,8]. A definitive therapeutic effect of 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 chromatin states altered by mutant IDH 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 dissect the complex molecular reprogramming associated with mutant IDH1. Data reveals that mutant IDH1 reprogrammed epigenome and transcriptome are dynamic, characterized by reversibility and partial persistence despite withdrawal of mutant IDH1. This demonstrates a decoupling between transcriptional and epigenetic states upon loss of mutant IDH1 expression.

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 (dox) for 30 passages (considered as baseline). We profiled these cells for 1, 5, 10, 20 and 40 passages following dox withdrawal with gene expression and methylation

arrays (Fig. 1d). In inducible IDH1 mutant IHAs, starting at 10 passages following dox withdrawal (off dox), cells in off dox state are more similar to cells that have never been exposed to dox (-dox) (Fig. 1e). Control IHAs expressing dox-inducible empty vector did not exhibit significant dox-dependent gene expression changes (Supplementary Fig. 1a, b). To dissect gene expression dynamics, we identified transient, gradual and persistent gene expression clusters following dox 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 panel) and 243 upregulated probes (205 genes) (64.2% transient, 30% gradual and 5.8% are persistent) (Fig. 1f, right panel). Persistently upregulated genes - now independent of IDH1 R132H expression - include *L1CAM*, *MTUS1* and *WNT6* (Fig. 1g, top panel). 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 panel). *L1CAM* was upregulated ($P = 1 \times 7.518^{-8}$) and *MEOX2* was downregulated ($P = 1 \times 4.583^{-9}$) in *IDH* mutated LGGs and may play a role glioma pathogenesis (Supplementary Fig. 1c). Magnitude of fold change at baseline was a major determinant of gene expression reversibility ($P = 0.0026$ for upregulated genes and $P < 1 \times 2.2^{-16}$ for downregulated genes) (Supplementary Fig. 1d). To assess relevance, we compared gene expression and methylation changes to IDH mutant gliomas [5], which indicated similarly deregulated pathways in inducible IDH1 mutant IHAs and CIMP+ gliomas (Supplementary Fig. 2a). Furthermore, hypermethylated loci in inducible IDH1 mutant IHAs display increased median methylation in CIMP+ LGGs compared to CIMP- LGGs (Supplementary Fig. 2b).

Interestingly, our study revealed mutant IDH1 dependent upregulation of CD24, a putative cell surface marker for stem-like cell populations [13] that is overexpressed in tumors and broadly expressed in many tissues including the brain [14] (Fig. 1h). In inducible IDH1 mutant IHAs, *CD24* is upregulated and its expression is decreased upon dox withdrawal (Fig. 1h,

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 IDH1 mutant IHAs (~2-6% of IDH1 mutant IHAs are $CD24^+$). $CD24^+$ cells formed significantly more colonies in soft agar compared to $CD24^-$ cells (Supplementary Fig. 2e, f). Transcriptome analysis of sorted CD24 subpopulations indicated 285 upregulated and 293 downregulated genes (Supplementary Table 2). Gene set enrichment analysis (GSEA) showed positive enrichment of cholesterol biosynthesis, MYC targets, WNT signaling and embryonic stem cell core pathways, demonstrating a stem cell-like gene expression program in $CD24^+$ cells (Supplementary Fig. 2g). Methylation analysis revealed 16,525 hypermethylated and 20,260 hypomethylated loci in $CD24^+$ IHAs (Supplementary Table 2). Overall, 137 downregulated and 189 upregulated genes had an inversely correlated CpG probe (Fig. 1i, Supplementary Table 2). We utilized RNA-seq data from BrainSpan Developmental Transcriptome [15] to determine expression patterns of several upregulated genes in $CD24^+$ IHAs during human brain development. Several genes including *CD24P4*, *L1CAM*, *GNAL* expressed at higher levels during pre- and early infancy brain development 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, this data suggests that $CD24^+$ cells emerge as a progenitor-like state within IDH1 R132H expressing astrocytes and contribute to epigenetic and transcriptional reprogramming.

Dynamic reshaping of chromatin landscape by IDH1 R132H

ChIP-seq analysis of major histone modifications revealed progressive enrichment of H3K4me3 (associated with actively transcribed genes) , H3K36me3 (prevents intragenic cryptic transcript initiation during active transcription), H3K9me3 and H4K20me3 (repressive marks enriched in heterochromatic regions) in successive passages of stable IDH1 R132H IHAs, but not in parental astrocytes (Fig. 2a, b). We inspected chromatin reorganization of H3K9me3, H4K20me3 and H3K36me3 using Hilbert curves [18], which revealed mutant IDH1 specific spread

over additional genomic regions (Supplementary Fig. 3a). To determine histone methylation kinetics, we grouped union of peaks across all stable cell lines per histone mark by *k*-means clustering into six clusters. Regions within H3K4me3 cluster 1 exhibit the highest intensity across all passages, and are mostly located within 5'UTR or promoter regions (Supplementary Fig. 3b), whereas cluster 6, which showed enrichment in IDH1 mutant astrocytes at passage 40 is mostly located in introns or intergenic regions (Fig. 2c, Supplementary Fig. 3b). Clustering of H3K9me3 and H3K36me3 peaks revealed regions with similar enrichment across all samples (cluster 5 for both H3K9me3 and H3K36me3), and regions with uniformly high enrichment for H3K9me3 in IDH1 mutant astrocytes at passage 40 (cluster 2 and cluster 6) or H3K36me3 (cluster 1 and cluster 6) (Fig. 2c, d). Next, we divided the genome into 5kb sliding windows, calculated H3K9me3 and H3K36me3 enrichment for each window and identified 6 distinct groups using k-means clustering. This unbiased approach also revealed genomic regions with increased H3K9me3 and H3K36me3 signal in IDH1 mutant astrocytes at passage 40, while larger regions were quiescent for either histone mark at the same passage (Supplementary Fig. 3c-e). These results indicated that increased H3K9me3 and H3K36me3 methylation occurs at specific regions, rather than uniformly genome-wide (Supplementary Fig. 3c-e).

While there was an overall increase in H3K36me3, H3K9me3 and H4K20me3 in IDH1 mutant IHAs, their proportion of overlapping genomic features remained approximately equivalent (Supplementary Fig. 4a). In contrast, distribution of H3K4me3 peak locations included intronic and intergenic regions in IDH1 R132H IHAs (Fig. 2e). Notably, we observed focal H3K4me3 gains in 4q12 in IDH1 mutant astrocytes, including increased H3K4me3 within the *DANCR* locus and *PDGFRA* TSS, which is intimately linked to gliomagenesis [19] (Fig. 2f).

We analyzed bulk histone methylation levels in inducible IDH1 mutant IHAs with liquid chromatography/mass spectrometry of purified histones, which confirmed significant increases in H3K4me2/me3, H3K9me2/me3 and H3K36me3 in +dox relative to -dox IHAs (Supplementary Fig. 4b). To determine reversibility of histone methylation landscape, we performed ChIP-seq on

H3K4me3, H3K27me3, H3K9me3 and H3K36me3 at baseline and at 40-passages following dox withdrawal. Similar to stable mutant IDH1 IHAs, we observed substantial H3K4me3 enrichment along with increases in H3K36me3 and H3K9me3 (albeit more modest compared to stable IHAs) (Fig. 2g).

To determine chromatin state organization in IDH1 mutant IHAs, we used ChromHMM [20] to learn genome-wide combinations of H3K4me3, H3K36me3, H3K27me3 and H3K9me3 in stable and inducible IHAs (Fig. 2h, i). In stable IDH1 mutant IHAs, we identified states largely devoid of histone marks (states 1, 2 and 3) with differential genome coverages and enrichments. Predominant features (>20% emission probability) for other states included H3K36me3+H3K9me3 (state 4), H3K36me3+H3K9me3+H3K27me3 (states 5 and 6), all (state 7), H3K9me3+H3K27me3 (state 8), H3K9me3+H3K27me3+H3K4me3 (state 9) and H3K4me3 (state 10). Similarly in inducible IHAs, states 1,3 and 6 were not marked by any histone marks (Fig. 2i). States 4 and 5 had features of all histone marks and state 2 was marked by H3K36me3+H3K9me3+H3K27me3. Other states included H3K36me3+H3K27me3 (state 7), H3K36me3 (state 8) with high enrichment in exonic regions, H3K9me3 + H3K36me3 (state 9) and H3K4me3 (state 10) (Fig. 2i). Chromatin state analysis revealed a time-dependent increase in genome coverage of state 5 in IDH1 mutant stable IHAs (6.7% of genome covered after 40 passages) (Fig. 2h). A similar chromatin state in IDH1 inducible IHAs (state 5) also exhibited a higher genome coverage in +dox IHAs (3.2%) (Fig. 2i). In addition, a greater percentage of IDH1 mutant stable IHAs are present in state 2 which is marked by low levels of H3K36 signal (Fig. 2h). Interestingly, +dox IHAs also have a larger percent of their genomes in state 6, largely devoid of histone marks, which frequently transitions to state 5 or state 8 (marked by high H3K36me3 signal), indicating that this state is located near active elements (Fig. 2i). These results reveal a dramatic, specific, and dynamic histone mark re-organization in presence of IDH1 R132H.

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

Next, we compared DNA methylation profiles of +dox and –dox IDH1 mutant IHAs at baseline. Overall, 41,209 loci were aberrantly methylated, of which 67.4% were hypermethylated (n=27,792). Importantly, 18,023 of the hypermethylated loci displayed at least 70% methylation in a patient-derived glioma tumorsphere harboring endogenous IDH1 R132H mutation (TS603) ($P = 1 \times 10^{-6}$), suggesting gain of methylation across similar loci (Supplementary Fig. 5a, b). Unsupervised clustering of DNA methylation showed that starting at 20 passages following dox withdrawal, off dox astrocytes are more similar to –dox astrocytes (Fig. 3a). Control IHAs expressing empty vector did not undergo dox dependent changes (Supplementary Fig. 5c-e).

While majority of loci eventually return to baseline levels, 25% of differentially methylated loci maintain their aberrant methylation levels (Fig. 3b). Using an approach similar to that used 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) show coherent time-dependent kinetics (Fig. 3c, d). Similar to gene expression data, we observed a significant positive correlation between magnitude of methylation changes and degree of reversibility ($P < 1 \times 2.2^{-16}$). Overall, 49% of loci with intermediate and 64.4% of loci with high gains in methylation are persistently hypermethylated (Fig. 3e). Next, we binned persistently methylated loci by their baseline beta values (Supplementary Fig. 6a,b). Majority of these loci have intermediate methylation levels and 765 loci are methylated by less than 20% in the baseline state (Supplementary Fig. 6a,b). Median persistent delta-beta gain compared to baseline was 21% at the end of dox withdrawal time course (Supplementary Fig. 6c). Select loci (n=29) are unmethylated at baseline (< 10%) and maintain > 30% increase in delta-beta at off dox P40 (Supplementary Fig. 7a,b).

Overlap of differentially methylated loci with ENCODE states [21] showed hypermethylation proximal to transcription start sites (TSS; states 1 and 2) and enhancer regions (state 7) and hypomethylation of quiescent regions (state 15) in inducible IDH1 mutant IHAs (Fig. 3f). In the persistent state, hypermethylated loci had preference for 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 epigenetic state of inducible mutant IHAs (Fig. 3g). Downregulated genes in the gradual and persistent clusters tended to have a tighter hypomethylated beta value distribution within their intergenic regions (Fig. 3g). We plotted ChIP-seq signals $\pm$ 1kb around CpG sites corresponding to either the methylation clusters or all probes to generate background signal (Supplementary Fig. 8a, b). Persistently hypermethylated loci exhibited a slight increase in surrounding H3K4me3 signal in -dox cells at baseline and were largely depleted for 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 clusters had an overall higher H3K4me3 signal surrounding hypomethylated CpG site in +dox versus –dox cells, suggesting a shift in H3K4me3 towards regions that lose methylation (Fig. 3h). Overlap of chromatin states with aberrantly methylated loci indicated differences between chromatin states associated with hypermethylated and hypomethylated loci; however, we have not observed a chromatin state common among persistent clusters, with the exception that hypomethylated loci in the persistent cluster overlap minimally (1% of persistently hypomethylated loci) with chromatin state 10 (marked by high H3K4me3 enrichment at TSS) (Fig. 3i). These findings suggest that the small set of genes, which continue to be perturbed after mutant IDH withdrawal is likely to be indirect and that an aberrantly maintained chromatin state is not necessary for persistent changes.

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

ChromHMM analysis revealed low-to-intermediate enrichment levels of H3K4me3 co-occurring with other states such as H3K27me3+H3K9me3+H3K36me3, pointing to aberrant co-existence of chromatin states in IDH1 mutant cells (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, others showed shifts in their composition in +dox samples (Fig. 4b). Overall, in inducible IHAs, H3K4me3 signal displayed a positive correlation with gene expression (Supplementary Fig. 9a) and when TSS were ranked by H3K4me3 occupancy across all inducible lines, several genes, including *PDGFRA* and *CD9* with higher H3K4me3 ranking in +dox samples were also upregulated by more than 1.5-fold (Supplementary Table 4).

While majority of H3K27me3, H3K9me3 and H3K36me3 peaks reverted to baseline levels following dox withdrawal (Supplementary Fig. 9b), a distinct proportion of H3K4me3 peaks 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 and included *MAP1LC3A*, *MARVELD2* and *TET1* (Supplementary Fig. 9d). RNA-seq data indicated a strong reduction in autophagy-related gene, *MAP1LC3A* that is maintained upon dox withdrawal (Fig. 4c, Supplementary Table 5).

We derived 11,443 regions with the most significant H3K4me3 gain in +dox IDH1 mutant IHAs (Supplementary Table 6). These regions were largely devoid of H3K4me3 in -dox cells, and maintained low-level H3K4me3 signal in off dox state (Fig. 4d). Majority of regions that gain H3K4me3 were enriched at 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 WT tumorspheres (TS543, TS667). GSEA using the GREAT toolbox of top 2000 H3K4me3 peaks at TSS for all profiled

cells revealed several enriched gene sets in IDH mutant, but not in IDH wild-type cells, including Myc/Max targets and ZNF143 binding, implicating MYC signaling [22,23] and aberrant chromatin interactions in IDH-mutated cell lines (Fig. 4g). Furthermore, differential enrichment analysis of mutant versus wildtype IDH1 models revealed 1,159 H3K4me3 binding sites that separate cell lines based on IDH1 R132H status (Fig. 4h, Supplementary Fig. 9e, Supplementary Table 6) and aberrant H3K4me3 methylation remained at lower enrichment levels in the off dox state for these differentially enriched regions (Fig. 4h).

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

We hypothesized that H3K4me3 gain and DNA methylation loss at intergenic regions may poise repetitive elements that are located at methylation dense, quiescent genomic regions for activation. GSEA indicated a highly significant positive enrichment of interferon alpha response pathway gene set in inducible IDH mutant astrocytes (Fig. 5a, b). Given the upregulated interferon response, we considered whether endogenous retroviruses (ERVs) might be re-activated in response to IDH1 R132H. As described previously [24], we mapped our RNA-seq data to a family of expressed ERVs and identified specific ERVs (e.g. ERVH-2, ERVH-4, ERVH-6, ERVH-5) elevated in IDH mutant samples, but not in control cells, at passage 70 (Fig. 5c). We validated several of these ERVs by qPCR (Fig. 5d). Additionally, a previously described viral defense signature [25] was upregulated in inducible IDH1 mutant IHAs, as well as in a subset of IDH1 mutant LGGs (Fig. 5e, Supplementary Fig. 10a). We tested expression of ERVs previously described to trigger an immune response[26], and observed a modest, but significant, increase in several of selected ERVs (Fig. 5f). Therefore, it appears that IDH1 R132H may play a role in upregulating ERVs, which may contribute to genome instability or immune activation [25].

Based on observations above, we wondered whether mutant IDH1 could promote genetic instability. GSEA indicated aberrant enrichment of several positional gene sets in +dox IHAs, which included negative enrichment of chromosome 19q13 genes (Fig. 5g). Analysis of copy-number alterations using methylation data revealed amplifications and deletions of several regions, including a higher frequency of broad deletions of chromosome 19q, in +dox compared to –dox inducible IDH1 mutant IHAs, but not in empty vector controls. (Supplementary Fig. 10b, c; Fig. 5h). Chromosome 19q loss (with intact 1p) occurs in gliomas and may be a negative prognostic marker [27]. Metaphase FISH confirmed our findings of increased frequency of 19q loss in IDH1 mutant 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 studies [28], +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 normal state within 6 passages following dox withdrawal, at a faster pace compared to 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 forebrains of NOD-SCID mice. Animals were placed on dox-containing diet to induce IDH1 R132H expression. Two months after implantation, mice were randomized based on bioluminescence signal (BLI): 10 mice continued on dox, and 10 mice returned to regular mouse feed. BLI measurements indicated a small (albeit not statistically significant) reduction in tumor growth after dox withdrawal (Fig. 6b). None of the mice injected with control IHAs developed tumors (Supplementary Fig. 11a), while several mice

(~40%) injected with +dox IDH1 mutant IHAs displayed intracranial lesions on MRI (Supplementary Fig. 11b, c).

To better understand morphology of these tumors, we used Serial 2-photon tomography (STPT) to image ZsGreen fluorescence throughout the brain. Green fluorescent-positive cells were automatically detected (Fig. 6c), reconstructed in 3D (Fig. 6d, Supplementary Video 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]. Mouse brains injected with –dox cells, while not apparent on MRI, displayed several bright areas on STPT slices (Supplementary Fig. 11d). Semi-quantitative analysis of a representative STPT image revealed the majority of cells present at the injection site (somatosensory cortex), which spread into nearby regions with high growth in dorsal striatum (Fig. 6e). Importantly, we identified several brains with parenchymal infiltration of IDH1 mutant astrocytes observed at a cellular resolution level that were not apparent on MRI (Supplementary Fig. 11e). Overall, our results indicate that IDH1 R132H IHAs have a greater potential to infiltrate and form tumors in an orthotopic model; however, tumors continue to grow despite absence of IDH1 R132H expression.

DISCUSSION

We 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 original levels (at least over the time period tested). Our findings suggest alternate effector pathways associated with IDH mutations that may be eventually decoupled from its neomorphic enzymatic activity. This is intriguing given previous findings that 2HG inhibition alone is not sufficient to reverse IDH mutant induced epigenetic state [7,8,31]. Our data also shows the slow

nature of methylation changes upon loss of mutant IDH1 and highlights the need for longer time course experiments with mutant IDH1 inhibitors to determine its long-term effects along with the utility of combining epigenetic drugs with small molecule inhibitors of mutant IDH1.

We identified *L1CAM* to be persistently up-regulated despite long-term loss of mutant IDH1 expression, suggesting that continued *L1CAM* expression may contribute to oncogenic phenotype conferred by mutant IDH1 expression. Previous studies have shown that L1CAM is expressed on glioma stem cells and its suppression leads to reduced tumor growth in intracranial brain tumor xenograft models [9] . Further research into L1CAM inhibition in IDH mutant gliomas, i.e. with L1CAM blocking antibodies, may be considered as a potential vulnerability that may 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), highlighting the importance of understanding the chromatin landscape in *IDH* mutant gliomas. Our data also reveals 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 play 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 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/DNA methylation and possibly activation of ERVs may lead to genomic instability.

We demonstrate gain in H3K4me3 at *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* activation [32].

An early H3K4me3 gain may poise *PDGFRA* for activation until 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 potential for increased 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 indicate that mutant IDH1 inhibition by dox withdrawal leads to a modest decrease in tumor growth in a brain orthotopic model of IDH1 R132H 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, integrated data presented here provide novel 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 gliomas.

ACCESSION NUMBERS

All data have been deposited in NCBI's Gene Expression Omnibus under accession number GSE85942.

AUTHOR CONTRIBUTIONS

S.T. and T.A.C. conceived of the study. S.T., V.M., J.T, A.W.M.F., W.W., Y.Z., N.E., S.H., G.N., C.H.L., C.B., J.C., J.T.H., N.K., P.O. performed the experiments. S.T., V.M., J.T, Y.W., A.W.M.F., Y.Z., N.E., S.H., G.N.,C.H.L., C.B., J.C., J.T.H., N.K., P.O., 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.

ACKNOWLEDGEMENTS

We thank the members of the Chan and Thompson labs for helpful discussions. This work was supported in part by the 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.), NIH (R01 MH096946) (P.O.), NIH Cancer Center Support Grant P30CA008748 (C.T.) 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, National Institutes of Health.

FIGURE LEGENDS

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

a) Summary of cellular models used to study IDH1 R132H function and epigenomic and transcriptomic data sets produced in this study.

b) IHAs transduced with inducible empty vector, IDH1 wild-type or IDH1 R132H were grown for 30 passages (baseline) in the presence (Dox+) or absence (Dox-) of 1 μg/ml doxycycline (dox). Dox was withdrawn from the media for either 2 (top panel) or 10 passages (bottom panel) after baseline. Western blot of inducible IHAs confirms expression of IDH1 R132H only in presence of dox. EV, IHAs transduced with empty vector; WT, IHAs transduced with IDH1; MUT, IHAs transduced with IDH1 R132H; off, dox withdrawn.

c) Total intracellular 2-Hydroxyglutarate (2HG) was measured by gas chromatography-mass spectrometry (GC-MS) and normalized to internal standard (D5-2HG) and cell number. 2HG levels are increased upon dox administration and remain at complete inhibition at 15 passages following dox withdrawal after baseline.

d) Schematic of the reversibility model for inducible IDH1 R132H expressing astrocytes. P, passage.

e) Hierarchical clustering of global gene expression profiles of IHAs as shown in (d).

f) Heatmaps of downregulated (left panel) and upregulated (right panel) gene expression clusters following dox withdrawal at baseline. Pers, persistent. Color scale shows normalized expression levels.

g) Gene expression time course of several persistently upregulated genes (top) and downregulated (bottom) genes. Individual genes are labeled in the legend.

h) Line plots indicate average expression levels of 3 Affymetrix probes targeting CD24 over time.

i) Percent genomic features for methylated loci that exhibit the most inverse correlation with gene expression data across in IDH1 R132H^{CD24+} compared to IDH1 R132H^{CD24-} astrocytes. hyper/down, hypermethylated and down-regulated genes; hypo/up, hypomethylated and up-regulated genes.

j) Average expression of example genes upregulated in CD24^{+} cells measured in the BrainSpan Developmental Transcriptome data. FPKM, fragments per kilobase of transcript per million mapped reads. pcw, post-conceptional weeks; yrs, years.

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

a) Genome coverage (left) and number of peaks (right) in IHAs stably expressing IDH1 R132H at passage 2, 10 and 40. P, passage.

b) Genome coverage (left) and number of peaks (right) in parental IHAs at passage 2, 10 and 40. P, passage.

c) Heatmaps displaying $\log_2$ input-normalized ChIP signal for *k*-means clusters (*k*=6) for the union of H3K4me3 (left), H3K9me3 (middle) and H3K36me3 (right) peaks for all stable IHAs. Scaled regions from peak start to peak end are displayed. Color scale indicates ChIP signal ranging from no enrichment (dark blue) to high enrichment (dark red).

d) H3K9me3 (upper panel) and H3K36me3 (lower panel) $\log_2$ input-normalized ChIP signal profiles for the clusters identified in Fig. 2c. Scaled regions from peak start to peak end are displayed. Par, parental; Mut, mutant.

e) HOMER derived annotation of genomic loci covered by H3K4me3 peaks in parental and IDH1 R132H IHAs at passages 2, 10 and 40. P, passage; TTS, transcription termination site; UTR, untranslated region.

f) H3K4me3 read coverage normalized to 1x sequencing depth is shown at the *PDGFRA* TSS and *DANCR* locus for stable IHAs. Range of *y*-axis is indicated and fixed for both plots. P, passage; Par, parental; Mut, mutant

g) Number of called peaks in inducible IHAs at baseline (passage 30, dox- and dox+) and 40 passages following dox withdrawal from baseline (off dox 40P).

h) Chromatin states learned jointly across all stable IDH1 R132H expressing IHAs. Plots show emission parameters (%) learned on the basis of combinations of chromatin marks, genomic annotation, genome coverage (%) and transition parameters. Emission parameters indicate the probability that a given histone mark occurs in a particular chromatin state. Blue shading indicates relative fold enrichment.

i) Chromatin states learned jointly for inducible IHAs (-dox, +dox, off dox 40P). Plots show the emission parameters (%) learned on the basis of combinations of chromatin marks, annotation, genome coverage (%) and transition parameters. Emission parameters

indicate the probability that a given histone mark occurs in a particular chromatin state. Red shading indicates relative fold enrichment.

Fig. 3. Reversibility and kinetics of DNA methylation changes in IDH1 R132H-inducible astrocytes.

a) Hierarchical clustering of global methylation profiles in inducible IHAs.

b) Distribution of delta beta-values (compared to baseline –dox) of hypo- and hyper-methylated loci at baseline (first panel) and subsequent passages after dox withdrawal.

c) Heatmaps of hypermethylated (left panel) and hypomethylated (right panel) clusters following dox withdrawal. At passage 40 following dox withdrawal, 20% of hypermethylated and 30% of hypomethylated loci within the persistent clusters maintain their altered methylation state with equal or greater absolute beta-value differences compared to baseline. Color scale shows normalized methylation values ranging from low (dark blue) to high (red).

d) Number of hypomethylated (blue) and hypermethylated (red) loci that fall within transient, gradual and persistent clusters.

e) Table indicating the total number of hypermethylated (top) and hypomethylated (bottom) loci binned by beta-value differences (at baseline) across transient, gradual or persistent methylation clusters.

f) Overlap of hyper- and hypomethylated loci in +dox vs –dox IHAs at baseline with the 15-state ChromHMM annotation of the Roadmap Epigenomics project. ENCODE's state annotations are noted below for the 15 states.

g) Distribution of delta-beta values (+dox vs. –dox at baseline) for the loci corresponding to up- or downregulated genes (+dox vs. –dox at baseline) within transient, gradual and persistent clusters are stratified by their genomic features. None of the persistently up-regulated genes fall within 3'UTR or IGR regions.

h) Log_2 input-normalized H3K4me3, H3K27me3, H3K36me3 and H3K9me3 ChIP signal profiles (top) and heatmaps (bottom) ± 1kb around the hypomethylated CpG sites within transient (dark blue), gradual (light blue) or persistent (yellow) clusters for –dox, +dox and off dox IHAs. 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 state in –dox IHAs. *Background* indicates overlap of all probes on Illumina HumanMethylation 450K 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. Color scale indicates 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 indicate z-score of the genomic region covered by each transitioned state in +dox cells. Color scale ranges from low coverage (dark blue) to high coverage (red).

c) Sequence-depth normalized H3K4me3 and RNA-seq coverage tracks for *MAP1LC3A* indicate a persistent reduction of H3K4me3 at its promoter and loss of gene expression over the exonic regions.

d) Profiles (top) and heatmap (bottom) of log_2 input-normalized H3K4me3 ChIP-seq signal ±1kb around start and end of the top 11,443 regions with significant gains in H3K4me3 in +dox compared to –dox IHAs. 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 top 11,443 differentially enriched H3K4me3 peak distances from TSS. kb, kilobases; TSS, transcription start site.

f) Pie chart indicating distribution of HOMER annotated genomic features of the top 11,443 differentially enriched H3K4me3 peaks. TTS, transcription termination site.

g) Gene set enrichment analysis (as determined by GREAT toolbox) of top 2,000 TSS with the highest H3K4me3 enrichment across parental and IDH1 R132H stable IHAs at passage 40, -dox and +dox inducible IHAs, off dox 40P IHAs and TS603. Color indicates multiple testing adjusted q-value of enrichment.

h) Profiles (top) and heatmap (bottom) of $\log_2$ input-normalized H3K4me3 ChIP signal ±1kb around start and end of the top 1,150 significant regions in IDH1 mutant models (compared to IDH wild-type cells) along with off dox 40P IHAs for the same regions. Color scale indicates ChIP signal ranging from no enrichment (dark blue) to high enrichment (dark red). start, peak start; end, peak end.

Fig. 5. Upregulated viral response and genomic instability in IDH1 R132H expressing cells.

a) GSEA of interferon alpha signaling profile showing that interferon signaling pathway is significantly upregulated in +dox compared to –dox IHAs. NES, normalized enrichment score; FDR, false discovery rate.

b) GSEA of interferon alpha signaling profile showing that interferon signaling pathway is significantly upregulated in off dox states compared to –dox IHAs.

c) Heatmap of RNA-seq derived endogenous retrovirus (ERV) expression for empty vector (green) or IDH1 R132H inducible astrocytes (purple) at baseline (-dox, +dox), 40 passages following baseline (-dox 40P, +dox 40P), 1 passage or 40 passages following dox withdrawal (off 1P, off 40P) and tumorspheres (TS543, TS603, TS667) (gray). Scale represents z-scores of regularized log-transformed read counts.

d) qPCR validation of select ERVs (from Fig. 5c) with increased expression in IDH1 R132H +dox IHAs at passage 70. Error bars, s.d; *n*=3 ; * *P* < 0.05, ** *P* < 0.01, *** *P* < 0.001, **** *P* < 0.0001; n.s, not significant (unpaired t-test).

e) Heatmap of viral defense signature genes[25] for empty vector (green) or IDH1 R132H inducible astrocytes (purple) at baseline (-dox, +dox), 40 passages following baseline (-dox 40P, +dox 40P), 1 passage or 40 passages following dox withdrawal (off 1P, off 40P) and tumorspheres (TS543, TS603, TS667) (gray). Scale represents z-scores of regularized log-transformed read counts.

f) qPCR analysis of select ERVs involved in antiviral response indicates increased expression in IDH1 R132H +dox IHAs at passage 70. Error bars, s.d; *n*=3 ; * *P* < 0.05, ** *P* < 0.01, *** *P* < 0.001, **** *P*< 0.0001; n.s, not significant (unpaired t-test).

g) GSEA indicating that genes on chr19q13 are concordantly and significantly downregulated in IDH1 R132H +dox compared to -dox IHAs.

h) Copy number profile derived from Illumina Human Methylation 450K array for IDH1 R132H expressing astrocytes at baseline (compared to –dox). Red signifies amplifications and blue signifies losses. Chr, chromosome.

i) Quantification of metaphases with relative loss of 19q in +dox and –dox IDH1 R132H inducible IHAs at baseline (P30) or 5 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 1, 5, 12 or 15 passages off dox), Error bars, s.e.m; *n*=2. P, passage.

b) Serial bioluminescence imaging tracking growth of inducible astrocytes (10 mice in –dox group; 20 mice in +dox group). Arrow indicates randomization of +dox group into off dox (dox withdrawn, 10 mice) and +dox (continued on dox diet, 10 mice). Each time point is average radiance across groups (mean ± s.e.m).

c) Inducible IHAs labeled with ZsGreen are imaged with STPT microscope. Bright areas of coronal sections depict ZsGreen expression and Allen Brain Atlas coordinates are shown at the bottom.

d) Whole brain surface reconstruction from 280-brain sections imaged with STPT with 20% resolution of raw data (first panel). 3D tumor reconstruction of ZsGreen-labeled cells showing tumor location and growth (sagittal and horizontal views; second and third panel). Last panel shows reconstructed tumor.

e) Prediction of cell numbers at various brain regions, indicating semi-quantitative spread of fluorescent cells from a representative IDH1 R132H (+dox) and off dox tumor. 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.

METHODS

Cell Culture

Immortalized human astrocytes were a gift of Dr. Russell O. Pieper (UCSF) and were generated as previously described [33]. IHAs were infected with pLVX-Tet-On retrovirus and selected with 800 ug/ml of G418. Selected IHAs were next infected with either empty pRetroX-hygro or pRetroX-hygro containing IDH1 R132H cloned downstream of Tet-reponsive promoter and selected with 500 ug/ml pf Hygromycin B. IHAs were cultured in the absence or presence of doxycycline to induce expression of IDH1 R132H (one passage equivalent to 2 doublings). Glioma spheres (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 basic-FGF and 2 μg/ml Heparin (StemCell Technologies). All cell lines were routinely tested for Mycoplasma and found to be negative for Mycoplasma contamination. TS603 is a WHO grade III oligodendroglioma line with 1p/19q codeletion and an endogeneous IDH1 R132H mutation and TS543/TS667 are malignant glioma tumorsphere lines and harbor *PDGFRA* amplifications [7,34].

FACS Analysis

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

Sample preparation

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

Histone Extraction and Mass Spectrometry

Histones from 5 million cells per sample were extracted using standard acid extraction protocol. Histone modification was quantified using bottom-up mass spectrometry strategy as outlined previously [35]. Briefly, ice-cold acetone washed histone pellet was air-dried and then subjected to trypsin digestion and chemical derivatization using propionic anhydride as described previously [36]. Histone peptides were then analyzed by nano-liquid chromatography coupled with triple quadrupole mass spectrometer (Dionex UltiMate 3000 and ThermoFisher Scientific TSQ Quantum) using selected reaction monitoring (SRM/MRM) method developed previously [37]. Data were analyzed using Skyline software (MacCoss Lab, University of Washington) [38].

Proliferation Assay

Cell proliferation was measured in duplicates 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% CO_2 and cell index was monitored for 150 hours. 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 media into 6-well plates containing 0.65% top and bottom agar. Cells were plated in the middle layer in growth media containing 0.40% agar. Medium covering the agar was refreshed every 3 days with or without Doxycycline-containing medium (1µg/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 (gift from Bryan Welm, Addgene plasmid # 39196). Transduced IHAs were sorted for ZsGreen expression. Bioluminescent imaging was performed weekly following intraperitoneal injection of D-luciferin and measured using Xenogen IVIS Spectrum *in vivo* imaging system (PerkinElmer). Living Image software (PerkinElmer) was used to acquire and analyze the BLI data.

Orthotopic transplantation experiments

All mouse experiments were approved by Institutional Animal Care and Use Committee at MSKCC and strict guidelines were enforced. Female NOD-*scid* IL2Rg^{null} (Jackson Laboratory) between 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 either fed doxycycline chow or normal diet (dox hyclate diet formulated at 2500 mg/kg; Envigo, Cambridgeshire, UK). Randomization of mice receiving dox diet were done as follows: Briefly, animals were ranked according to their BLI signal and alternatingly assigned to group 1 (continued on dox diet) or group 2 (dox diet withdrawn, continued on regular diet). Designation of diet conditions to either group was determined by randomly in Microsoft Excel (Microsoft). Animals that were sick or died due to unrelated circumstances were excluded from further analyses. Lesions were analysed by a board certified neuropathologist (J.H) and morphology, degree of infiltration and cellularity were used as criteria for histopathology assessment.

Fluorescence in situ hybridization

Cell lines were harvested according to standard procedures. Briefly, cell lines were cultured with Colcemid (0.1 µg/mL) at 37°C for 45 minutes, resuspended in 0.075 mol/L KCl, incubated at 37°C for 10 minutes, and then fixed in methanol-acetic acid (3:1). FISH analysis was performed on the fixed cells using a 4-color 1p36/1q23 and 19p12/19q13 home-brew probe. 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 Inc, Waltham, MA). The entire hybridized area was scanned under 63x/100x to assess the quality of hybridization and signals pattern. A minimum of 50 consecutive intact metaphases were captured and signal counts/copy

number for each locus was recorded. A metaphase was considered be positive for loss of 1p or 19q if the signal count/copy number of respective control locus was greater (relative loss).

Array based genomic analysis

Expression analysis was performed using Affymetrix U133 2.0 microarray (Affymetrix) and methylation analysis was performed using the Infinium HumanMethylation450K bead array (Illumina). Arrays were processed at the Integrated Genomics Operation (iGO), Memorial Sloan Kettering Cancer Center as per manufacturers' protocol. R statistical software[39] (v3.3.1) was used for data analysis. ChAMP methylation pipeline (version 1.10) was used to extract and analyze data from idat files and normalization was done using beta-mixture quantile normalization [40]. Batch correction was performed using the ComBat algorithm [41]. Delta beta values were calculated by comparing +dox to –dox inducible IDH1 mutant IHA samples at baseline passage. Loci with absolute deltaBeta > 0.1 were considered to be aberrantly methylated and were binned as: *low*, 0.1 < absolute delta-beta < 0.2; *intermediate*, 0.2 ≤ absolute delta-beta < 0.4 or *high*, absolute delta-beta ≥ 0.4. To overlap with ENCODE's annotation chromatin state models, we obtained chromatin core 15-state data for ENCODE's NH-A cell line (Epigenome ID E125).

Affymetrix CEL files were imported into the R statistical software. Normalization was performed with the AffyPLM package in BioConductor (v3.3), using RMA background correction, quantile normalization, and Tukey biweight summary method. Batch correction was performed using the ComBat algorithm by incorporating batch information. To identify aberrantly expressed genes or methylated loci or aberrantly specific to mutant IDH1 in inducible mutant IDH1 IHAs, we excluded the probes with equal or greater fold-changes or delta-beta changes in +dox vs –dox empty vector expressing inducible IHAs.

IDH1 mutant inducible models and LGGs were correlated as follows: 1) Pathway enrichment of differentially expressed genes in IDH1 mutant inducible IHAs (+dox vs –dox

across all passages) and CIMP+ and CIMP- LGGs [5] were compared using *compareCluster* (with *enrichPathway* function) in *clusterProfiler R*-package [42], 2) Hypermethylated probes (deltaBeta > 0.1) at passages 1, 5, 10, 20 and 40 after baseline were used to plot the distribution of beta-values in CIMP+ or CIMP- LGGs [5]. For CD24 gene expression data (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 in $CD24^+$ compared to $CD24^-$ samples.

To identify the CpG loci with the most impact on gene expression within the CD24 subpopulation, the methylation probe with the most negative change (deltaBeta < -0.1) corresponding to upregulated or the most positive change change (deltaBeta > 0.1) corresponding to downregulated genes were identified. Genomic features associated with these probes were extracted from the Infinium 450K HumanMethylation annotation file.

To derive gene expression clusters, genes with an absolute fold change > 1 (log_2 scale) in +dox samples copared to –dox samples at baseline were identified and their fold-change determined at consecutive dox withdrawal time points by comparing to –dox samples at baseline. We define genes that return to baseline values (absolute fold change <1 when compared to off dox samples) within 5 passages as *transient*, 20 passages as *gradual* and aberrantly expressed at 40 passages as *persistent.* Aberrantly expressed genes at baseline (absolute fold-change > 1) were binned as: *low*, absolute fold change < 1.5; *intermediate*, 1.5 < absolute fold change < 2 or *high*, absolute fold change >2. Significance of distribution of fold change levels across three clusters were determined using Fisher's exact test for count data. Similar to gene expression clusters, we derived methylation clusters with similar kinetics following dox withdrawal. To this extent, we considered loci with absolute delta-beta values > 0.1 as aberrantly methylated at baseline and set an absolute delta-beta cut-off of 0.1 for cluster

binning. In addition subgroup of loci within the persistent cluster that maintain an absolute delta-beta value ≥ 0.1 at passage 40 following dox withdrawal were identified.

GSEA was performed using GSEA module within GenePattern with either Hallmark collection or positional gene sets (c1) and permutation method set to *gene_set*.[43,44]

Copy number alterations

Copy number alterations were called from Infinium 450K HumanMethylation array using ChAMP or conumee packages in R statistical software [40,45]. Segmented data were visually inspected and subsequently used to identify broad and focal CNAs using GISTIC2.0 [46] on GenePattern server with *–armpeel* and *–broad* parameters set to *yes*.

Magnetic Resonance Imaging.

Brains of injected mice were scanned on a 200 MHz Bruker 4.7T Biospec MRI scanner (Bruker Biospin Corp., Ettlingen, Germany); equipped with a 300 mT/m ID 20-cm gradient (Resonance Research Inc., Billerica, MA). Mice were anaesthetized by 2% isoflurane in oxygen. Sedated animals were physiologically monitored during scan period (SA Instruments Inc., Stony Brook, NY). For mouse brain imaging, brain axial T2-weighted images using fast spin-echo RARE sequence (Rapid Acquisition with Relaxation Enhancement) were acquired with TR 1.5 s, TE 50 ms, RARE factor of 8, slice thickness of 0.7 mm, field of view 30×20mm, in-plane resolution of 117×125μM, and 24 averages.

Chromatin Immunoprecipitation

Cells were cross-linked in 1% formaldehyde for 15 minutes at room temperature, with constant agitation, followed by quenching with 125 mM Glycine for 5 minutes at room temperature. ChIP assay was performed with EZ-Magna ChIP kit (Millipore, # 17-611). Briefly, nuclei were isolated and chromatin was sheared using Bioruptor (Diagenode) (30-s pulses, 30-s rest between pulses,

power setting 10). Shearing efficiency was confirmed with Bioanalyzer (Agilent). Chromatin was incubated with primary antibody overnight at 4 degrees with constant agitation. DNA and protein immune complexes were eluted and reverse cross-linked. DNA extraction was performed using a spin filter column. Immunoprecipitated and Input DNA were subjected to sequencing on Illumina HiSeq 2500. Prior to sequencing, quantitative real-time PCR 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), H4K20me3 (Abcam, ab9053, Lot GR2463.2).

RNA-sequencing

RNA Sequencing was performed at the Integrated Genomics Operation (iGO), Memorial Sloan Kettering Cancer Center. Raw reads in fastq format were aligned to Human Genome HG19 using RNAseq STAR aligner version 2.4.0d [47,48]. STAR aligner was chosen for mapping accuracy and speed [49]. Mapped reads for each sample were counted for each gene in annotation files in GTF format (gencode.v19.annotation.gtf) using FeatureCounts read summarization program [50]. Individual count files were merged to the matrix by an in-house R script and normalized using DESeq R package. To generate heatmaps, count data were transformed using *rlog* (regularized log transformation) function from *DESeq2* R package, visualized and clustered (with default parameters) using *pheatmap* R package. Expression of transcribed ERVs were determined as described[24] and rlog counts were visualized using *pheatmap* R package.

Public data acquision and processing

Level 3 data (processed and normalized) RNA-seq data along with clinical annotation for Brain Lower Grade Gliomas were downloaded from Broad GDAC Firehose. Gene expression of the

viral defense signature[25] per sample was visualized using *pheatmap* package in R. *L1CAM* and *MEOX2* expression were extracted and significance between IDH mutant and IDH wildtype LGGs were assessed by Student's t-test.

ChIP-sequencing

ChIP-Seq sequencing was performed at the Integrated Genomics Operation (iGO), Memorial Sloan Kettering Cancer Center. Raw sequencing data were aligned to hg37 genome build using BWA v.0.7.10 [51]. Further indel realignment, base-quality score recalibration and duplicate-read removal were performed using Genome Analysis Toolkit (GATK) version 3.2.2 [52,53]. 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 [54] 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 IP (immuno-precipitated) and INPUT bam files. Only peaks with RPKM(IP)/RPKM(INPUT) ratio greater than 0.5 were reported to final result. For broad histone marks (H3K36me3, H3K9me3, H3K79me2, H3K27me3, H4K20me3) resulting raw peaks were filtered by by $P = 0.1$. To ensure high stringency of broad peak calling, we applied the second peak caller called SICER version 1.1 [55]. Two sets of peaks were merged using bedtools merge tool, only intersecting broad peaks were reported. BigWig files were generated using *bamCoverage* from deepTools suite with options --binSize 10 --normalizeTo1x 2451960000 --extendReads --centerReads[56]. Log_2 ratio of ChIP signal and 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 usig *gviz* package in R. Bedtools merge tool was used to create a union of all peaks per histone mark.

Functional enrichment based on proximal gene annotations was determined using GREAT v.3.0.0 default enrichment settings [57]. Differential binding analysis for comparison of H3K4me3 peaks across IDH1 wildtype and IDH1 mutant models was performed using DiffBind R package with following parameters*: method=DNA_DESEQ2_BLOCK* and p-value < 0.05 as cut-off) . Tag counts ± 2000bp around TSS were determined using *annotatePeaks* HOMER script. Tag counts > 32 and with at least 1-fold absolute change between conditions were used for further analyses. Differential peaks were analyzed using *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 was plotted as barcharts. The 'non-coding' annotation in HOMER refers to exons of non-coding RNAs within RefSeq NR subset.

Visualization of ChIP-seq peaks: Peaks were visualized using Hilbert curves (HilbertVis R package) which allows the compact display of the entire chromosome while maintaining its spatial structure, thereby allowing direct comparison of multiple datasets [18]. Three channel RGB images (EBImage::rgbImage R function) of Hilbert curves for IDH1 mutant and parental cell lines were overlaid on one panel, visualized and compared across all passages.

***Chromatin state discovery*.** ChIP-Seq and their corresponding input bam files were binarized with 200 bp bin size using custom script to convert a set of bam files into binarized data matrices suitable for ChromHMM model learning tools [20]. 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) [58]. First, MACS2 bed files were reformatted to HOMER peak files format and then, annotatePeaks script was applied to determine whether a peak is in the

TSS (transcription start site), TTS (transcription termination site), Exon (Coding), 5' UTR Exon, 3' UTR Exon, Intronic, Intergenic or CpG islands region.

2-HG analysis. 2HG levels were determined by mass spectrometry as previously described [5].

RNA Isolation, Reverse Transcription and Quantitative PCR for ERV quantification

Total RNA was isolated from $5x10^6$ cells (inducible IHAs at P70) by using RNeasy Plus Mini Kit (Qiagen). 2μg of total RNA was used for reverse transcription with SuperScript III One-Step RT-PCR system with random hexamer (Invitrogen). cDNA was used as template for quantitative PCR to determine expression of ERVs on a QuantStudio 6 Flex Real-Time PCR console (Thermo Fisher). Results were normalized to 18S rRNA. Primer sequences are listed in Supplementary Table 7.

Tracing of zsGreen labeled immortalized human astrocytes with Serial 2-photon Tomography

For serial 2-photon tomography (STPT) experiments, mice orthotopically injected with zsGreen labeled immortalized human astrocytes were perfused transcardially with ice-cold PBS followed by 4% ice-cold PFA. After a 24 hour post-fix in 4% PFA, brains were kept in 0.7% glycine solution for 48 hours and then store in PBS in 4°. Whole brain was embedded in 4% agarose in 0.05M PB, cross-linked in 0.2% sodium borohydrate solution for 2-3hrs. Whole brain sample was sliced in 50 μm sections using high-speed two photons microscope with integrated vibratome sectioning (x-y resolution of 1 μm; z step of 50 μm; TissueCyte 1000, TissueVision-STP) as described before [59]. Raw image files were corrected for illumination, stitched in 2D, and aligned in 3D. We employed two channels configuration using channel 1 (background signal) and channel 2 detected ZsGreen signal. Background channel was employed as subtracted channel

to reduce signal/noise from the signal channel. ZsGreen-positive cells from the subtracted channel automatically detected by a convolutional network trained to recognize nuclear cell body labeling [60] were visually validated, reconstructed in 3D, and their spatial information was registered by affine followed by B-spline transformation using the software Elastix [61] to a 3D reference brain based on the Allen Brain Atlas [29,30]. Total number of cells in each brain region was normalized by total number of positive neurons in the belonging macro structure (e.g. cortex) to account for variability in the total number depending on where each tumor has grown. Brain 3D reconstruction was done using Imaris version 7.6.5 (Bitplane, Zurich, CH, Switzerland).

REFERENCES

1 Parsons, D. W. *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807-1812, doi:10.1126/science.1164382 (2008).

2 Yan, H. *et al.* IDH1 and IDH2 mutations in gliomas. *The New England journal of medicine* **360**, 765-773, doi:10.1056/NEJMoa0808710 (2009).

3 Dang, L. *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739-744, doi:10.1038/nature08617 (2009).

4 Lu, C. *et al.* IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* **483**, 474-478, doi:10.1038/nature10860 (2012).

5 Turcan, S. *et al.* IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* **483**, 479-483, doi:10.1038/nature10866 (2012).

6 Duncan, C. G. *et al.* A heterozygous IDH1R132H/WT mutation induces genome-wide alterations in DNA methylation. *Genome research* **22**, 2339-2355, doi:10.1101/gr.132738.111 (2012).

7 Rohle, D. *et al.* An inhibitor of mutant IDH1 delays growth and promotes differentiation of glioma cells. *Science* **340**, 626-630, doi:10.1126/science.1236062 (2013).

8 Tateishi, K. *et al.* Extreme Vulnerability of IDH1 Mutant Cancers to NAD+ Depletion. *Cancer cell* **28**, 773-784, doi:10.1016/j.ccell.2015.11.006 (2015).

9 Bao, S. *et al.* Targeting cancer stem cells through L1CAM suppresses glioma growth. *Cancer research* **68**, 6043-6048, doi:10.1158/0008-5472.CAN-08-1079 (2008).

10 Ducray, F. *et al.* Anaplastic oligodendrogliomas with 1p19q codeletion have a proneural gene expression profile. *Molecular cancer* **7**, 41, doi:10.1186/1476-4598-7-41 (2008).

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 research* **56**, 1440-1444 (1996).

12 Maness, P. F. & Schachner, M. Neural recognition molecules of the immunoglobulin superfamily: signaling transducers of axon guidance and neuronal migration. *Nature neuroscience* **10**, 19-26, doi:10.1038/nn1827 (2007).

13 Mimeault, M. & Batra, S. K. Molecular biomarkers of cancer stem/progenitor cells associated with progression, metastases, and treatment resistance of aggressive cancers. *Cancer Epidemiol Biomarkers Prev* **23**, 234-254, doi:10.1158/1055-9965.EPI-13-0785 (2014).

14 Fang, X., Zheng, P., Tang, J. & Liu, Y. CD24: from A to Z. *Cellular & molecular immunology* **7**, 100-103, doi:10.1038/cmi.2009.119 (2010).
15 Miller, J. A. *et al.* Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199-206, doi:10.1038/nature13185 (2014).
16 Mohanan, V., Temburni, M. K., Kappes, J. C. & Galileo, D. S. L1CAM stimulates glioma cell motility and proliferation through the fibroblast growth factor receptor. *Clin Exp Metastasis* **30**, 507-520, doi:10.1007/s10585-012-9555-4 (2013).
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. *The Journal of biological chemistry* **276**, 21656-21663, doi:10.1074/jbc.M101790200 (2001).
18 Anders, S. Visualization of genomic data with the Hilbert curve. *Bioinformatics* **25**, 1231-1235, doi:10.1093/bioinformatics/btp152 (2009).
19 Verhaak, R. G. *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer cell* **17**, 98-110, doi:10.1016/j.ccr.2009.12.020 (2010).
20 Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nature methods* **9**, 215-216, doi:10.1038/nmeth.1906 (2012).
21 Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).
22 Bailey, S. D. *et al.* ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters. *Nature communications* **2**, 6186, doi:10.1038/ncomms7186 (2015).
23 Bai, H. *et al.* Integrated genomic characterization of IDH1-mutant glioma malignant progression. *Nature genetics* **48**, 59-66, doi:10.1038/ng.3457 (2016).
24 Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48-61, doi:10.1016/j.cell.2014.12.033 (2015).
25 Chiappinelli, K. B. *et al.* Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell* **164**, 1073, doi:10.1016/j.cell.2015.10.020 (2016).
26 Roulois, D. *et al.* DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. *Cell* **162**, 961-973, doi:10.1016/j.cell.2015.07.056 (2015).
27 Scheie, D. *et al.* Prognostic variables in oligodendroglial tumors: a single-institution study of 95 cases. *Neuro-oncology* **13**, 1225-1233, doi:10.1093/neuonc/nor114 (2011).
28 Koivunen, P. *et al.* Transformation by the (R)-enantiomer of 2-hydroxyglutarate linked to EGLN activation. *Nature* **483**, 484-488, doi:10.1038/nature10898 (2012).
29 Kim, Y. *et al.* Mapping social behavior-induced brain activation at cellular resolution in the mouse. *Cell reports* **10**, 292-305, doi:10.1016/j.celrep.2014.12.014 (2015).
30 Sunkin, S. M. *et al.* Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic acids research* **41**, D996-D1008, doi:10.1093/nar/gks1042 (2013).
31 Turcan, S. *et al.* Efficient induction of differentiation and growth inhibition in IDH1 mutant glioma cells by the DNMT Inhibitor Decitabine. *Oncotarget* **4**, 1729-1736, doi:10.18632/oncotarget.1412 (2013).
32 Flavahan, W. A. *et al.* Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110-114, doi:10.1038/nature16490 (2016).
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 research* **61**, 4956-4960 (2001).

34 Silber, J. *et al.* miR-34a repression in proneural malignant gliomas upregulates expression of its target PDGFRA and promotes tumorigenesis. *PloS one* **7**, e33844, doi:10.1371/journal.pone.0033844 (2012).
35 Zheng, Y., Huang, X. & Kelleher, N. L. Epiproteomics: quantitative analysis of histone marks and codes by mass spectrometry. *Curr Opin Chem Biol* **33**, 142-150, doi:10.1016/j.cbpa.2016.06.007 (2016).
36 Garcia, B. A. *et al.* Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nature protocols* **2**, 933-938, doi:10.1038/nprot.2007.106 (2007).
37 Zheng, Y., Tipton, J. D., Thomas, P. M., Kelleher, N. L. & Sweet, S. M. Site-specific human histone H3 methylation stability: fast K4me3 turnover. *Proteomics* **14**, 2190-2199, doi:10.1002/pmic.201400060 (2014).
38 MacLean, B. *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966-968, doi:10.1093/bioinformatics/btq054 (2010).
39 Team, R. D. C. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.r-project.org/. . (2008).
40 Morris, T. J. *et al.* ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics* **30**, 428-430, doi:10.1093/bioinformatics/btt684 (2014).
41 Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127, doi:10.1093/biostatistics/kxj037 (2007).
42 Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-287, doi:10.1089/omi.2011.0118 (2012).
43 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).
44 Reich, M. *et al.* GenePattern 2.0. *Nature genetics* **38**, 500-501, doi:10.1038/ng0506-500 (2006).
45 Feber, A. *et al.* Using high-density DNA methylation arrays to profile copy number alterations. *Genome biology* **15**, R30, doi:10.1186/gb-2014-15-2-r30 (2014).
46 Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome biology* **12**, R41, doi:10.1186/gb-2011-12-4-r41 (2011).
47 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
48 Dobin, A. & Gingeras, T. R. Mapping RNA-seq Reads with STAR. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* **51**, 11 14 11-19, doi:10.1002/0471250953.bi1114s51 (2015).
49 Engstrom, P. G. *et al.* Systematic evaluation of spliced alignment programs for RNA-seq data. *Nature methods* **10**, 1185-1191, doi:10.1038/nmeth.2722 (2013).
50 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).
51 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

52 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).
53 DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* **43**, 491-498, doi:10.1038/ng.806 (2011).
54 Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using MACS. *Nature protocols* **7**, 1728-1740, doi:10.1038/nprot.2012.101 (2012).
55 Zang, C. *et al.* A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25**, 1952-1958, doi:10.1093/bioinformatics/btp340 (2009).
56 Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic acids research* **42**, W187-191, doi:10.1093/nar/gku365 (2014).
57 McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501, doi:10.1038/nbt.1630 (2010).
58 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).
59 Ragan, T. *et al.* Serial two-photon tomography for automated ex vivo mouse brain imaging. *Nature methods* **9**, 255-258, doi:10.1038/nmeth.1854 (2012).
60 Turaga, S. C. *et al.* Convolutional networks can learn to generate affinity graphs for image segmentation. *Neural Comput* **22**, 511-538, doi:10.1162/neco.2009.10-08-881 (2010).
61 Klein, S., Staring, M., Murphy, K., Viergever, M. A. & Pluim, J. P. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging* **29**, 196-205, doi:10.1109/TMI.2009.2035616 (2010).