



Single-cell multimodal glioma analyses identify epigenetic regulators of cellular plasticity and environmental stress response

Kevin C. Johnson^{1,12}✉, Kevin J. Anderson^{1,12}, Elise T. Courtois¹, Amit D. Gujar¹, Floris P. Barthel^{1,2}, Frederick S. Varn¹, Diane Luo¹, Martine Seignon¹, Eunhee Yi¹, Hoon Kim¹, Marcos R. H. Estecio³, Dacheng Zhao¹, Ming Tang⁴, Nicholas E. Navin⁵, Rahul Maurya¹, Chew Yee Ngan¹, Niels Verburg⁶, Philip C. de Witt Hamer⁶, Ketan Bulsara⁷, Michael L. Samuels¹, Sunit Das^{8,9,10}, Paul Robson^{1,11} and Roel G. W. Verhaak¹✉

Glioma intratumoral heterogeneity enables adaptation to challenging microenvironments and contributes to therapeutic resistance. We integrated 914 single-cell DNA methylomes, 55,284 single-cell transcriptomes and bulk multi-omic profiles across 11 adult IDH mutant or IDH wild-type gliomas to delineate sources of intratumoral heterogeneity. We showed that local DNA methylation disorder is associated with cell-cell DNA methylation differences, is elevated in more aggressive tumors, links with transcriptional disruption and is altered during the environmental stress response. Glioma cells under in vitro hypoxic and irradiation stress increased local DNA methylation disorder and shifted cell states. We identified a positive association between genetic and epigenetic instability that was supported in bulk longitudinally collected DNA methylation data. Increased DNA methylation disorder associated with accelerated disease progression and recurrently selected DNA methylation changes were enriched for environmental stress response pathways. Our work identified an epigenetically facilitated adaptive stress response process and highlights the importance of epigenetic heterogeneity in shaping therapeutic outcomes.

Diffuse gliomas are the most common malignant brain tumors in adults and are incurable. Extensive molecular characterization of gliomas has defined genomic drivers and clinically relevant subtypes, such as those based on the presence of *IDH1/2* gene mutations (that is, IDH mutant and IDH wild-type (WT))^{1–3}. Inter- and intratumoral heterogeneity are salient features across glioma subtypes that contribute to universal therapeutic resistance. The heterogeneity observed in surgical resection specimens reflects each tumor's evolutionary path, which is driven by competition between subpopulations harboring diverse genetic, epigenetic and transcriptional aberrations^{4–8}. Thus, understanding how these different layers of heterogeneity integrate to define clonal lineages and drive glioma evolution may provide insights into treatment failure.

The study of tumor heterogeneity is complicated by cellular plasticity that enables cancer cells to reversibly transition between distinct cellular states in response to genetic, microenvironmental and therapeutic stimuli⁹. Single-cell RNA sequencing (scRNA-seq) studies have previously identified such dynamic cellular states in IDH WT gliomas^{10–13}. Cell states of IDH mutant gliomas displayed a more restricted plasticity along a hierarchical differentiation axis^{14,15}. Epigenetic modifications, such as DNA methylation (DNAm)

at cytosine followed by guanine dinucleotides (that is, CpGs), are mitotically heritable marks and encode cellular states and dynamics¹⁶. For example, the transition from a differentiated-like state to an undifferentiated, or stem-like, state after chemotherapy in glioma was accompanied by epigenetic reprogramming¹⁷. However, the epigenetic mechanisms that enable cellular plasticity and regulate glioma cell states are poorly understood.

Aberrant DNAm resulting from errors in the placement or removal of epigenetic marks can provide genetically identical cells, the diversity needed to respond to environmental stressors. These stochastic errors in DNAm replication result in increased local DNAm disorder¹⁸. DNAm disorder is present in non-tumor cells, potentially reflecting active epigenetic remodeling, DNAm drift associated with age or environmental exposures¹⁹. DNAm disorder may accumulate in cancer cells as passenger events or be evolutionarily selected by destabilizing gene expression programs⁹. Previous studies of glioma have demonstrated associations between bulk tumor epigenetic heterogeneity metrics and clinical outcomes^{2,5,20}. Together, these findings suggest that stochastic DNAm alterations contribute to tumor heterogeneity and cellular plasticity that may drive the evolution of treatment-resistant phenotypes.

¹The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA. ²Department of Pathology, Brain Tumor Center Amsterdam, Amsterdam University Medical Center, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. ³Department of Epigenetics and Molecular Carcinogenesis, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ⁴Department of Data Science, Dana-Farber Cancer Institute, Boston, MA, USA. ⁵Department of Genetics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ⁶Department of Neurosurgery, Brain Tumor Center Amsterdam, Amsterdam University Medical Center, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. ⁷Division of Neurosurgery, The University of Connecticut Health Center, Farmington, CT, USA. ⁸Arthur and Sonia Labatt Brain Tumour Research Centre, Hospital for SickKids, University of Toronto, Toronto, Ontario, Canada. ⁹Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada. ¹⁰Division of Neurosurgery, Li Ka Shing Knowledge Institute, St. Michael's Hospital, University of Toronto, Toronto, Ontario, Canada. ¹¹Genetics and Genome Sciences, University of Connecticut School of Medicine, Farmington, CT, USA. ¹²These authors contributed equally: Kevin C. Johnson, Kevin J. Anderson. ✉e-mail: kevin.c.johnson@jax.org; roel.verhaak@jax.org

In this study, we integrated single-cell DNA methylomes, single-cell transcriptomes and single-cell copy number profiles with bulk genomic profiles across a cohort of 11 glioma patient samples to dissect heterogeneous cell populations^{21–24} and define epigenetic states that contribute to tumor evolution^{18,24}. We combined these in vivo analyses with in vitro perturbations to identify the gene regulatory regions most susceptible to stochastic DNAm alterations, the epigenetic modulation of transcriptional networks involved in glioma cellular identity and that DNAm disorder may aid the cellular stress response. Our work provides insights into the sources of intratumoral heterogeneity that fuel glioma evolution.

Results

Single-cell DNAm links DNAm disorder with epigenetic heterogeneity. To investigate glioma heterogeneity, we performed single-cell DNAm (scDNAm) and single-cell gene expression accompanied by bulk whole-genome sequencing (WGS), RNA sequencing and DNAm microarray in 11 adult patients with glioma (Fig. 1a). Both principal molecular subtypes (IDH mutant and IDH WT) and distinct clinical time points (that is, unmatched initial and recurrent tumors; Supplementary Table 1 and Extended Data Fig. 1) were represented. We mechanically dissected tumor specimens from the same geographical region dissociating tissue for single-cell protocols, flash-freezing tissue for bulk genomic assays (Fig. 1a). We applied single-cell reduced representation bisulfite sequencing (scRRBS) and 10x Genomics single-cell transcriptomics on cells from the same dissociation (Extended Data Fig. 2a)^{25,26}. Viable CD45[−] (that is, pan-immune cell marker) cells were plated for scRRBS, while single-cell transcriptomics was performed on all viable cells, arriving at a set of 914 single-cell methylomes and 55,284 single-cell transcriptomes. On average, approximately 145,000 mean unique CpG dinucleotides or 2,340 expressed genes were measured per cell. On average, approximately 8,000 mean CpGs were shared between any two cells (Extended Data Fig. 2b–h). Tumor and normal cells were grouped by inferred copy number alterations resulting in a final set of 844 DNAm and 30,831 transcriptomic tumor cell profiles (Methods and Extended Data Fig. 3a–i).

Unsupervised clustering and multidimensional scaling (MDS) of the pairwise distances between scDNAm patterns grouped tumor cells by *IDH1* mutation status (Fig. 1b) since IDH mutant tumors display greater genome-wide DNAm levels (Wilcoxon rank-sum test $P < 2.2 \times 10^{-16}$; Fig. 1c)²⁷. The colocalization of cells from different patients suggested shared epigenetic states. The isolated patient-specific grouping of one of six IDH mutant and two of five IDH WT tumors may reflect epigenetic diversity that is also influenced by genetic intertumoral heterogeneity (Fig. 1b, Extended Data Figs. 1 and 4a).

We next evaluated intratumoral epigenetic heterogeneity by quantifying stochastic DNAm alterations in each single cell. In normal cells, DNAm congruence in nearby CpGs reflects tightly ordered gene regulation (Fig. 1d, top)²⁸. Local DNAm disorder may disrupt both proximal and distal gene regulation (Fig. 1d, bottom)¹⁶. We defined DNAm disorder within a cell and across specific genomic compartments as the proportion of sequencing reads discordant for DNAm status (PDR) as described previously^{5,18,29}. Cell–cell DNAm disorder variation differed by tumor (Fig. 1e) and was increased in tumor cells compared with non-tumor cells (Wilcoxon rank-sum test $P < 0.0001$; Extended Data Fig. 4b). Total somatic single-nucleotide variant burden, reflecting patient age³⁰ and mutational processes (Extended Data Fig. 4c), was not associated with mean DNAm disorder (Spearman correlation $\rho = 0.26$, $P = 0.43$), independent of sequence context (Extended Data Fig. 4d). However, DNAm disorder was associated with the fraction of the genome with somatic copy number alteration (SCNA) burden (Spearman correlation $\rho = 0.66$, $P = 0.03$; Fig. 1e). Cell cycle checkpoint deregulation, which generates SCNAs through a cell's

compromised ability to correct mis-segregations³¹, may continue to drive stochastic DNAm replication errors during evolution rather than being elevated in the tumor cell of origin.

To examine whether local DNAm disorder is associated with broad DNAm heterogeneity, we calculated the DNAm disorder and DNAm status for each cell across specific genomic contexts including: CpG islands (CGIs) and CGI shores, Alu repeat elements and chromatin remodelers (EZH2 and CTCF; Extended Data Fig. 4e). In high DNAm regions (for example, Alu repeat elements), increased DNAm disorder was associated with decreased DNAm, while in lower DNAm regions (for example, CGIs), an increased DNAm disorder was associated with increased DNAm (Extended Data Fig. 4e; Spearman correlation $P < 0.01$). These associations persisted in individual tumors (Fig. 1f,g; Spearman correlation $P < 0.01$), highlighting how local DNAm disorder may reflect epigenetically dynamic regions that contribute to the observed intratumoral epigenetic heterogeneity^{32,33}. To compare inter- and intratumoral DNAm variation, median absolute deviations were calculated across single cells, grouping cells by subtype (Extended Data Fig. 4f) and patient (Extended Data Fig. 4g). Consistent with the results of unsupervised clustering (Fig. 1b), intertumoral heterogeneity was approximately 2–3 times greater (IDH WT) than intratumoral heterogeneity (Extended Data Fig. 4f,g), with promoters/CGIs representing variably methylated regions within a tumor. The DNAm disorder tended to increase moving away from CGI centers (Spearman correlation $R = 0.5$, $P = 3.1 \times 10^{-8}$ IDH mutant and $R = 0.6$, $P = 4.1 \times 10^{-7}$ IDH WT) suggesting that selection may reduce DNAm disorder that impairs cellular fitness at these tightly regulated regions (Fig. 1h). Together, scDNAm profiling suggests that the variability observed at critical gene regulatory regions is influenced by DNAm disorder and higher levels of disorder may reflect epigenetic remodeling.

Elevated DNAm disorder in cell identity and stress pathways.

DNAm disorder may disrupt transcriptional programs¹⁸. Using companion scRNA-seq data, we examined the association between DNAm disorder and gene expression. Mean expression was reduced (Kruskal–Wallis $P < 2.2 \times 10^{-16}$; Fig. 2a) with increased levels of DNAm disorder at both promoters and gene bodies (Kruskal–Wallis $P < 2.2 \times 10^{-16}$; Fig. 2a and Extended Data Fig. 5a). Previous CGI observations (Fig. 1g) suggest that DNAm disorder at gene regulatory regions usually results in repressive DNAm (Extended Data Fig. 5b,c), contributing to gene expression dysregulation. Gene Ontology (GO) enrichment analysis on genes with high DNAm disorder (that is, DNAm disorder > 0.4) and genes with low DNAm disorder (that is, DNAm disorder = 0–0.1) (Methods) found that high DNAm disorder genes associate with cellular differentiation processes (Fisher's exact test adjusted $P < 0.05$; Fig. 2b) and low DNAm disorder genes associate with critical cell cycle and metabolic processes (Fisher's exact test adjusted $P < 0.05$; Fig. 2c). The enrichment results were consistent when using promoter or gene body DNAm disorder groupings (Extended Data Fig. 5d,e).

Changes in DNAm patterns at DNA-binding motifs can positively or negatively impact transcription factor binding³⁴. We identified regulatory elements susceptible to DNAm changes by determining DNAm disorder of transcription factor binding site (TFBS) motifs (Fig. 2d). Most TFBS motifs showed higher DNAm disorder in IDH WT compared with IDH mutant cells, consistent with general subtype differences. Transcription factors essential for glioma stem cell maintenance (for example, SOX2, SOX9 (ref. 35)) had lower than median binding site motif DNAm disorder independent of surrounding motif CpG density, implying selection against DNAm changes at these target regions (Fig. 2d and Extended Data Fig. 5f). In contrast, transcription factors with higher binding site motif scDNAm disorder (Methods) were related to the response to extracellular stimuli (Extended Data Fig. 5g). Increased DNAm

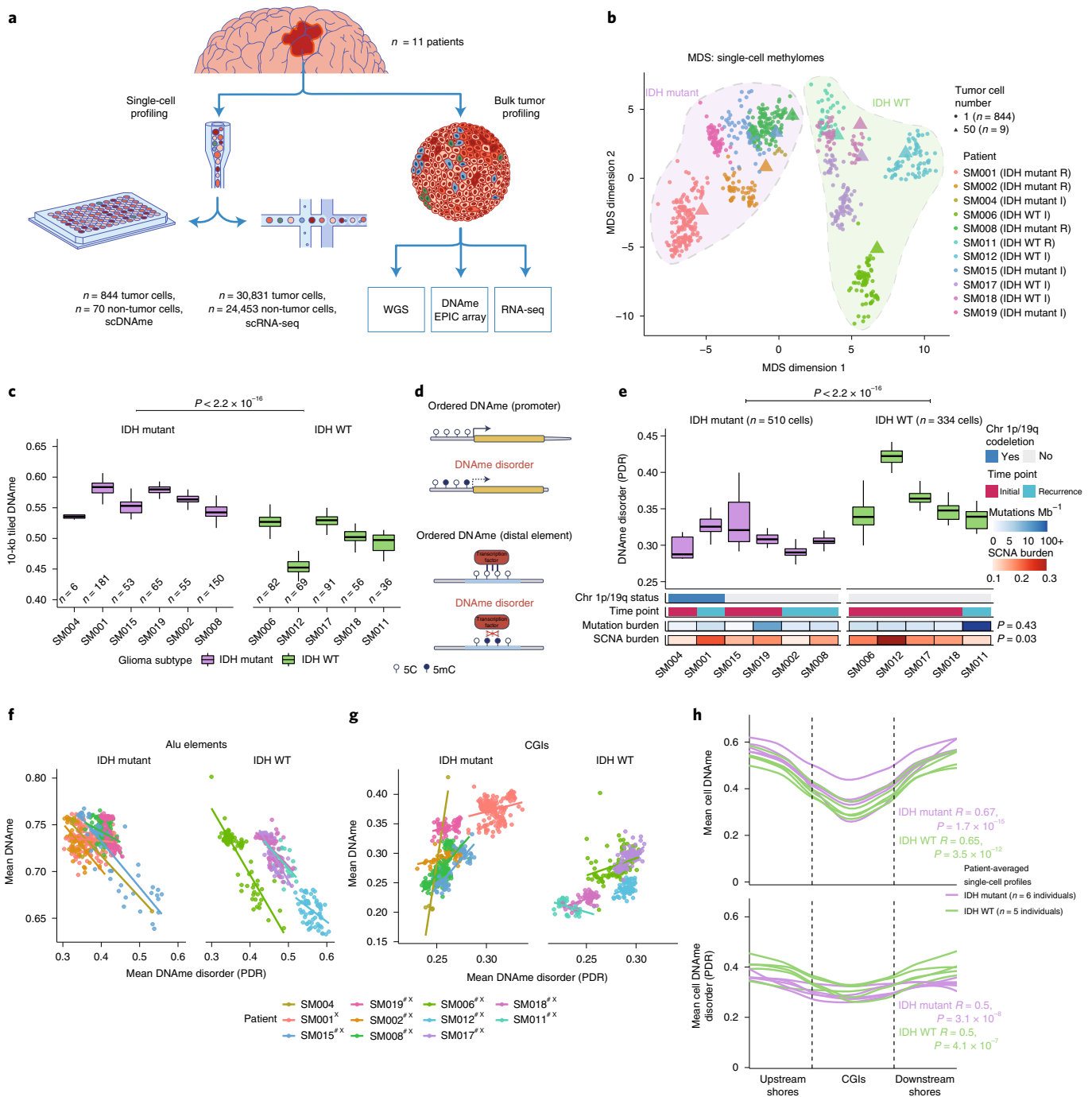


Fig. 1 | Single-cell DNA sequencing highlights the association between epigenetic intratumoral heterogeneity and local DNAm disorder. a, Schematic diagram detailing tumor sample processing and molecular profiling of single cells and bulk tumor samples ($n = 11$ individuals). **b**, MDS analysis using pairwise individual CpG distance metrics calculated between individual cells. Shapes represent whether a sample was a single tumor cell ($n = 844$ cells) or 50 tumor cells ($n = 9$ out of 11 individuals). The colors indicate individual patients, the shaded regions indicate the *IDH1* mutation status of the tumor and annotation is provided indicating clinical time point (I = initial, R = recurrence). **c**, Box plots depicting single-cell mean 10-kb tiled DNAm with individuals in columns and separated by IDH mutation status (color). Each box spans the 25th and 75th percentile, the center lines indicate the median and the whiskers represent the absolute range (minimum/maximum), excluding outliers. The two-sided Wilcoxon rank-sum test P value is presented to compare IDH mutant versus IDH WT cells. **d**, Schematic depiction of local DNAm disorder in different genomic contexts. Top: Promoter region where DNAm disorder is associated with disrupted gene expression. Bottom: Example of disrupted transcription factor binding due to DNAm disorder. **e**, Box plots of glioma cell DNAm disorder grouped by individual. The two-sided Wilcoxon rank-sum P value represents the comparison between IDH mutant and IDH WT DNAm disorder. Each individual is annotated with clinical and molecular metrics with P values indicating the relationship between sample mean DNAm disorder and WGS-derived somatic mutation burden or somatic copy number alteration burden (Spearman correlation). **f, g**, Alu repeat element (**f**) and CGI-specific (**g**) scDNAm disorder and DNAm are shown in scatter plots with linear regression lines colored per individual patient. Patient-specific Spearman correlation coefficients and P values are presented in Supplementary Table 2. Significant associations are denoted by '#' in **f** and 'X' in **g**. **h**, Mean DNAm values (top) and DNAm disorder (bottom) across CGIs with upstream and downstream CGI shores. Each individual is represented by a single curve and colored by its IDH mutation subtype. Subtype-specific Spearman correlation coefficient and P values indicate the relationship between distance from the CGI center and increase in mean cell DNAm and DNAm disorder.

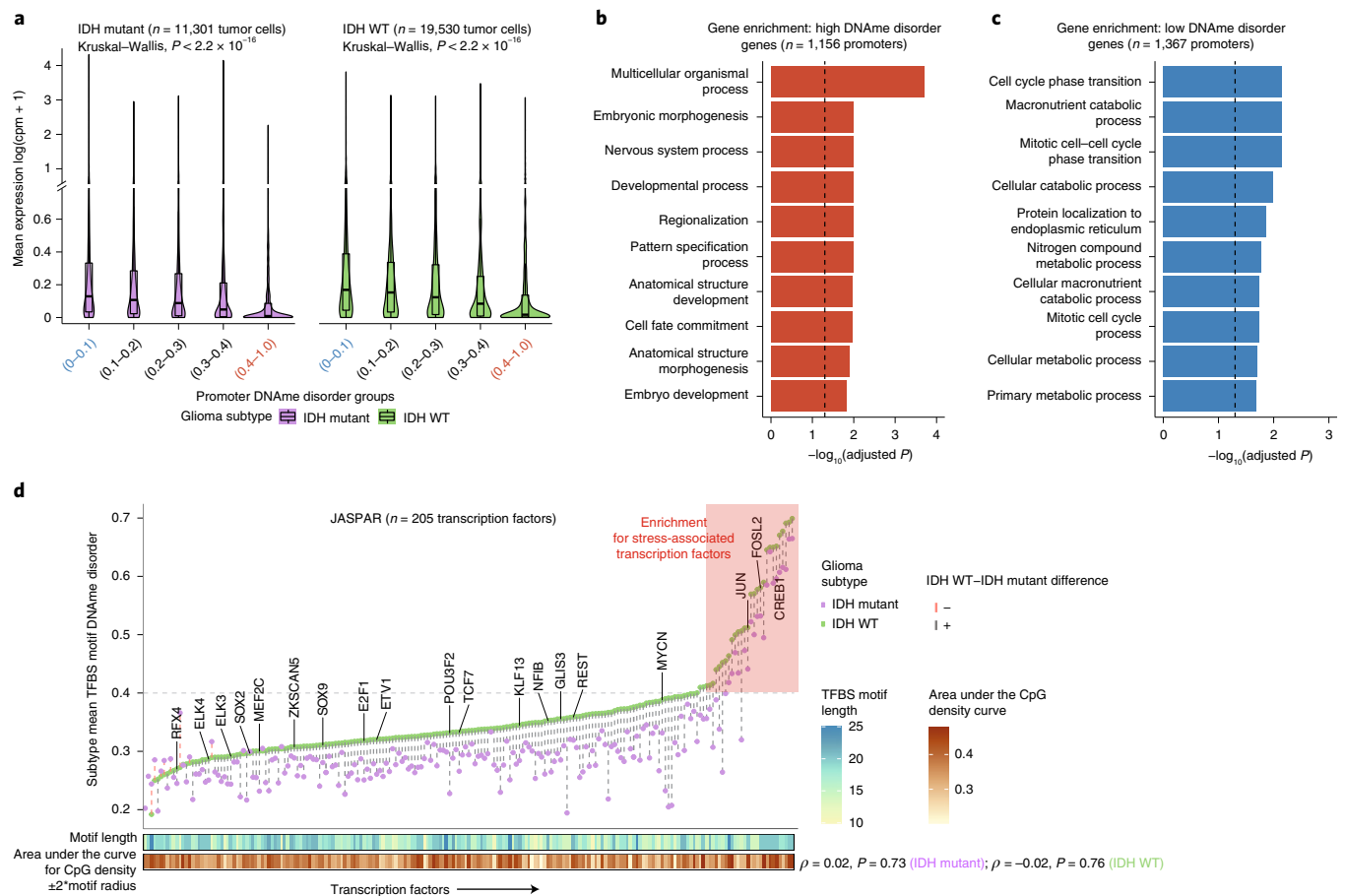


Fig. 2 | DNAmE disorder at gene regulatory elements is associated with cell identity and stress response pathways. **a**, Box plots of gene expression values from scRNA-seq data across different promoter DNAmE disorder groups. Each box spans the 25th and 75th percentile, the center lines indicate the median and the whiskers represent the absolute range (minimum/maximum), excluding outliers. The surrounding violins represent the distribution for each condition. Kruskal–Wallis test P values are presented for each subtype. The y axis includes a line break to improve the visibility of violin tails. **b,c**, GO enrichment analyses for high (**b**) and low DNAmE disorder genes (**c**) using promoter DNAmE disorder values with correction for false discovery rate. **d**, Scatter plot of mean scDNAmE disorder calculated across TFBS motifs within IDH subtype, ordered by IDH WT TFBS motif DNAmE disorder. Each column represents a single transcription factor with a colored dotted line connecting IDH mutant and IDH WT values. The names of transcription factors previously indicated to confer fitness advantages to glioma cells (MacLeod et al.)³⁵ are listed above their TFBS motif DNAmE disorder estimate. Transcription factors with high DNAmE disorder are shaded in red; annotation tracks are provided for motif length and motif CpG density (CpG density area under the curve within ± 2 times the motif radius). P values represent the Spearman correlation for IDH mutant ($P = 0.73$) and IDH WT ($P = 0.76$).

disorder levels at environmental stress response regulators may facilitate an adaptive response to stressors, such as hypoxia, which is common in glioma³⁶. To substantiate this association, we performed single-sample gene set enrichment analyses using bulk RNA-seq data and demonstrated positive associations between tumor average DNAmE disorder and upregulated stress response (Spearman correlation $R = 0.9, P < 0.01$) or cellular response to hypoxia (Spearman correlation $R = 0.98, P < 0.001$), but not randomly selected genes (Spearman correlation $R = -0.05, P > 0.05$; Extended Data Fig. 5h). Taken together, these results suggest that intratumoral variability in DNAmE disorder may facilitate the adoption of distinct epigenetic states in response to stress stimuli.

Single-cell multi-omics identifies epigenetic cell state regulators. To evaluate how DNAmE, stress response and cellular states are associated, we defined each tumor’s cellular composition using single-cell transcriptional profiles. We performed single-cell unsupervised clustering analysis and annotated clusters using marker genes (Fig. 3a and Extended Data Fig. 6a–d) to define glial, immune, stromal and malignant populations^{10,12}. Malignant cells were distributed over three canonical stem cell marker SOX2-expressing cell

states (Extended Data Fig. 6b) and existed across both IDH mutant and IDH WT tumors. We labeled these pan-glioma cell states (1) differentiated-like, (2) stem-like and (3) proliferating stem-like tumor cells (Fig. 3a, Extended Data Fig. 6b and Supplementary Table 3). Enumerating the proportion of pan-glioma malignant states showed that IDH mutant gliomas are enriched for stem-like cells (median 61%), while IDH WT gliomas contained predominantly differentiated-like cells (median 83%) and significantly more proliferating stem-like cells (median 16% IDH WT versus 2% IDH mutant, Wilcoxon rank-sum test $P = 0.02$; Fig. 3b). The previously described malignant astrocyte-like and oligodendrocyte-like IDH mutant glioma cell types¹⁵ corresponded to differentiated-like cells, as well as astrocyte-like and mesenchymal-like IDH WT glioma cellular states¹¹ (Fig. 3b and Extended Data Fig. 6e). The proliferating stem-like and stem-like states aligned closely with undifferentiated IDH mutant cells and oligodendrocyte progenitor-like, neural progenitor-like IDH WT cells, respectively (Extended Data Fig. 6e), highlighting the consistency of these pan-glioma signatures with existing glioma signatures^{11,15}.

We next inferred gene regulatory networks from single-cell expression profiles to identify transcription factors governing cell

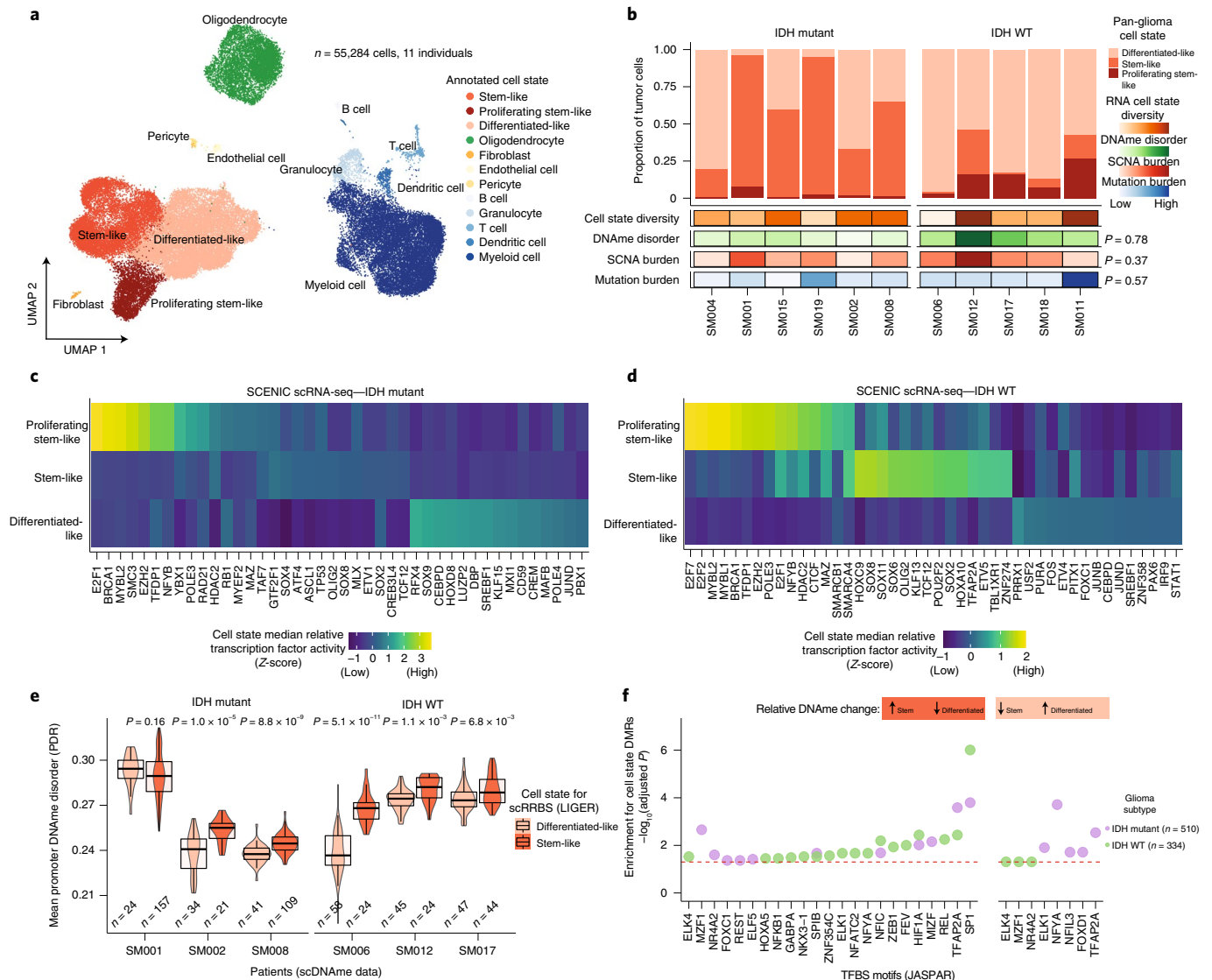


Fig. 3 | Integrative single-cell gene expression and DNAm analyses nominate epigenetic regulators of glioma cell state variability. a, UMAP

dimensionality reduction plot of scRNA-seq data ($n = 55,284$ cells, $n = 11$ individuals) showing the clustering of cell populations by transcriptionally defined cell state (point color) and labeled according to marker gene expression (Extended Data Fig. 6b). **b**, Stacked bar plots representing the proportion of cellular states per tumor for pan-glioma malignant cell classification. Each sample is annotated with molecular metrics with P values indicating the relationship between cell type diversity, measured by Shannon's entropy, and sample mean DNAm disorder, WGS-derived somatic alteration burden or WGS-derived somatic mutation burden (Spearman correlation). **c,d**, Enriched transcription factor activity across pan-glioma cellular states determined using the SCENIC algorithm and displayed as a heatmap of cell state median relative z -scores. Visualization is presented for the top 15 most active transcription factors of 5,000 randomly downsampled tumor cells in both IDH mutant (**c**) and IDH WT (**d**). **e**, Promoter DNAm disorder for tumors with at least ten cells per inferred cell state. Each box spans the 25th and 75th percentile, the center lines indicate the median and the whiskers represent the absolute range (minimum/maximum), excluding outliers. The surrounding violins represent the distribution of each condition. Two-sided Wilcoxon rank-sum test P values are presented for each tumor. **f**, Region set enrichment analysis for differentially methylated regions (DMRs, 10-kb tiles) with higher DNAm in stem-like (left) or differentiated-like cells (right). Enrichment was determined by LOLA. Individual points represent the enrichment of specific transcription factors in DMRs; the color indicates results for specific IDH subtypes; the dotted line represents the statistical significance threshold (adjusted $P < 0.05$).

states³⁷, which predicted a key set of transcription factors for each of the three pan-glioma cell states (Fig. 3c,d). Stem-like tumor cells demonstrated the highest activity for known stem cell regulators such as SOX2, SOX8 and OLIG2 (Fig. 3c,d). In addition to high SOX2/SOX8/OLIG2 activity, proliferating stem-like cells showed overrepresentation of chromatin remodeling and DNA repair gene networks as directed by EZH2 and BRCA1 (Fig. 3c,d). In contrast, differentiated-like cells demonstrated high transcription factor activity in astrocyte differentiation (that is, SOX9) and stress response (that is, JUND, FOS) processes. We confirmed that

differentiated-like cells had significantly greater stress and hypoxic transcriptional response compared with stem-like cells (Wilcoxon rank-sum test, $P < 2.8 \times 10^{-9}$; Extended Data Fig. 6f). DNAm disorder did not significantly differ between cell state-specific transcription factors (Kolmogorov–Smirnov test $P > 0.05$; Extended Data Fig. 6g). However, high binding site motif DNAm disorder levels were observed for several differentiated-like cell state transcription factors (for example, JUND and SREBF1), nominating them as cellular fitness regulators whose activity may be influenced by DNAm patterns (Fig. 3c,d).

To define the epigenetic states of stem-like and differentiated-like cells, we used the linked inference of genomic experimental relationships (LIGER) method³⁸ to identify shared properties between single-cell gene expression and DNAm data (Extended Data Fig. 7a). scDNAm and scRNA integration displayed a similar malignant cell state distribution within each sample, as expected when derived from the same tissue dissociation (Extended Data Fig. 7b). We next investigated the DNAm disorder and DNAm properties of stem-like (combining stem-like and proliferating stem-like) and differentiated-like cell state classifications. In tumors with both populations, stem-like cells displayed significantly increased promoter DNAm disorder (5 out of 6 tumors, Wilcoxon rank-sum test $P < 0.05$; Fig. 3e, left) and decreased promoter DNAm (4 out of 6 tumors, Wilcoxon rank-sum test $P < 0.05$; Extended Data Fig. 7c), potentially reflecting the greater transcriptional diversity of stem-like cells. To identify DNAm changes between stem-like and differentiated-like cells, we used a linear mixed effect model with tumor of origin as the random effect (Methods). Regions with differential DNAm across cell states were enriched for SP1 and TFAP2A binding sites, two transcription factors that frequently co-regulate developmentally associated genes (Fig. 3f)³⁹. We also identified increased DNAm at the binding sites of HIF1A/ARNT, the master transcriptional regulator of hypoxic response, in stem-like cells (Fig. 3f). Since increased DNAm at binding sites may reduce transcription factor binding efficiency, these results suggest that elevated cell stress transcription factor activity in differentiated-like cells may occur via epigenetic remodeling (Fig. 3f). Together, these results suggest that perturbing epigenetic control via DNAm disorder may promote the cell state plasticity necessary to tolerate diverse stressful microenvironments, including hypoxia⁴⁰ and therapy^{17,41,42}.

In vitro stress perturbations increase local DNAm disorder.

To directly determine whether environmental stressors impact DNAm disorder and cellular states, we subjected patient-derived glioma sphere-forming cells independently to a common tumor stress exposure (that is, hypoxia) and therapeutic exposure (that is, irradiation) (Fig. 4a). For both experiments, we used bulk RRBS with biological replicates ($n = 6$ average per condition, 60 total replicates), and gene expression with scRNA-seq. Importantly, each bulk RRBS sequencing read came from a single cell at single-allele resolution enabling DNAm disorder comparisons with our scRRBS data. We exposed two glioma cell lines to normoxic and hypoxic conditions and collected cells at 3 and 9 d. Candidate gene expression analyses via real-time PCR demonstrated that a robust cellular stress response was already present at the 3-d time point with an observed hypoxia dosage effect (Extended Data Fig. 8a,b). No hypoxia-associated DNAm disorder changes were detectable at the 3-d time point (Wilcoxon rank-sum test $P > 0.05$; Fig. 4b, left). However, there were significant hypoxia-associated DNAm disorder increases in both cell lines at the 9-day time point, suggesting that DNAm disorder accumulates with successive cell divisions (Wilcoxon rank-sum test $P < 0.05$; Fig. 4b, right). In parallel, we also irradiated the two glioma models with 2.5 Gy per day for 4 consecutive days (10 Gy total) and then collected these cells at the 9-d time point. Unlike the hypoxia exposure, the irradiation stressor was not continuous and measurements were taken after 5 d of recovery. The cells exposed to irradiation also demonstrated significant increases in DNAm disorder at CGI and promoter regions (Wilcoxon rank-sum test $P < 0.01$; Fig. 4c) compared with the 9-d normoxia (0 Gy) samples. We confirmed through WGS that irradiated and control cells shared highly similar mutational profiles, suggesting that the DNAm disorder increases were not due to underlying genetic changes (Extended Data Fig. 8c). In both hypoxia and irradiation experiments, there was reduced stress-associated DNAm disorder in regions flanking CGIs (shores) in one cell line, but no significant changes at intergenic regions indicating that DNAm disorder

may confer different selective advantages dependent on genomic context (Fig. 4b,c). DNAm disorder increases under direct stress (hypoxia) and after recent stress exposure (irradiation) suggests a common stress response mechanism that is retained even after stress removal. This is further supported by increased DNAm disorder at the binding site motifs of transcription factors whose activity is associated with cell fitness (Extended Data Fig. 8d), including upregulated ELK4, which contributes to the malignant phenotype through c-Fos regulation⁴³, and downregulated TFDP1, which promotes transcription from E2F target genes⁴⁴, whose altered activity levels may enable survival under stress (Extended Data Fig. 8e).

We next assessed whether stress-associated DNAm disorder increases are linked to cellular state shifts using scRNA-seq (10 total replicates, $n = 5$ conditions for 2 cell lines each, $n = 24,460$ cells). Unsupervised clustering by cell line demonstrated that stressed cells did not adopt new cell states but manifested as population cell state distribution shifts (Fig. 4d–g). This was supported by relatively few stress-specific differentially expressed genes (hypoxia = 166 (H2354), 68 (HF3016); irradiation = 27 (H2354), 26 (HF3016); Wilcoxon rank-sum test adjusted $P < 0.05$) that tended to be highly expressed across all states within a condition (for example, *TXNIP* in hypoxia; Extended Data Fig. 8g,h). We observed that there were hypoxia-associated increases in differentiated-like cell and reductions in proliferating stem-like cell proportions across both cell lines (chi-squared $P < 2.2 \times 10^{-16}$; Fig. 4e,g). Response to irradiation resulted in an increased stem-like compartment for HF2354 and a greater differentiated-like cell compartment for HF3016 (chi-squared test $P < 0.01$). After 9 d, cell state distributions of both irradiated cell models and the hypoxia condition for HF3016 were more comparable to controls, suggesting that stress-induced transcriptional shifts can be transient. We confirmed these stress-associated cell state shifts using a proliferation-independent IDH WT-specific cell classifier (Extended Data Fig. 8i,j). Taken together, stress-associated increases in DNAm disorder suggest that distinct microenvironmental pressures contribute to intratumoral epigenetic heterogeneity that may facilitate or stabilize adaptive cell state shifts.

SCNAs are positively correlated with DNAm disorder. We next investigated whether cellular stress resulting from genetic stimuli, in addition to environmental stimuli, could further explain DNAm disorder variability across a tumor. The fraction of the genome with SCNAs correlated with DNAm disorder at the bulk level (Spearman correlation $\rho = 0.66$, $P = 0.03$; Fig. 1e) and at the single-cell level for promoter-specific DNAm disorder and single-cell-inferred SCNAs (Spearman correlation $R = 0.70$, $P < 2.2 \times 10^{-16}$ IDH mutant and $R = 0.6$, $P < 2.2 \times 10^{-16}$ IDH WT; Fig. 5a). There were 3 significant intratumoral positive associations (Spearman correlation $P < 0.05$; Fig. 5a) indicating a weaker genetic effect or greater influence of microenvironmental stressors within a single tumor (Fig. 5a). To determine whether this relationship was driven by greater DNAm disorder in copy number altered regions, we calculated the DNAm disorder by cell in copy number-altered and non-altered regions. We did not observe a consistent relationship between DNAm disorder and the copy number status in scDNAm data (paired Wilcoxon rank-sum test $P > 0.05$). This suggests that aneuploidy does not directly account for epigenetic diversity increases but that genetic and epigenetic events are shaped by similar biological processes (for example, DNA replication stress). Late replicating regions of the genome accumulate more DNA mutations and structural rearrangements⁴⁵ and we observed a positive association between single-cell promoter and gene body DNAm disorder with later replicating regions (Kruskal–Wallis $P < 1 \times 10^{-4}$; Fig. 5b). Late replicating genomic regions may have reduced capacity to correct aberrant methylation leading to their preferential accumulation in a largely stochastic manner.

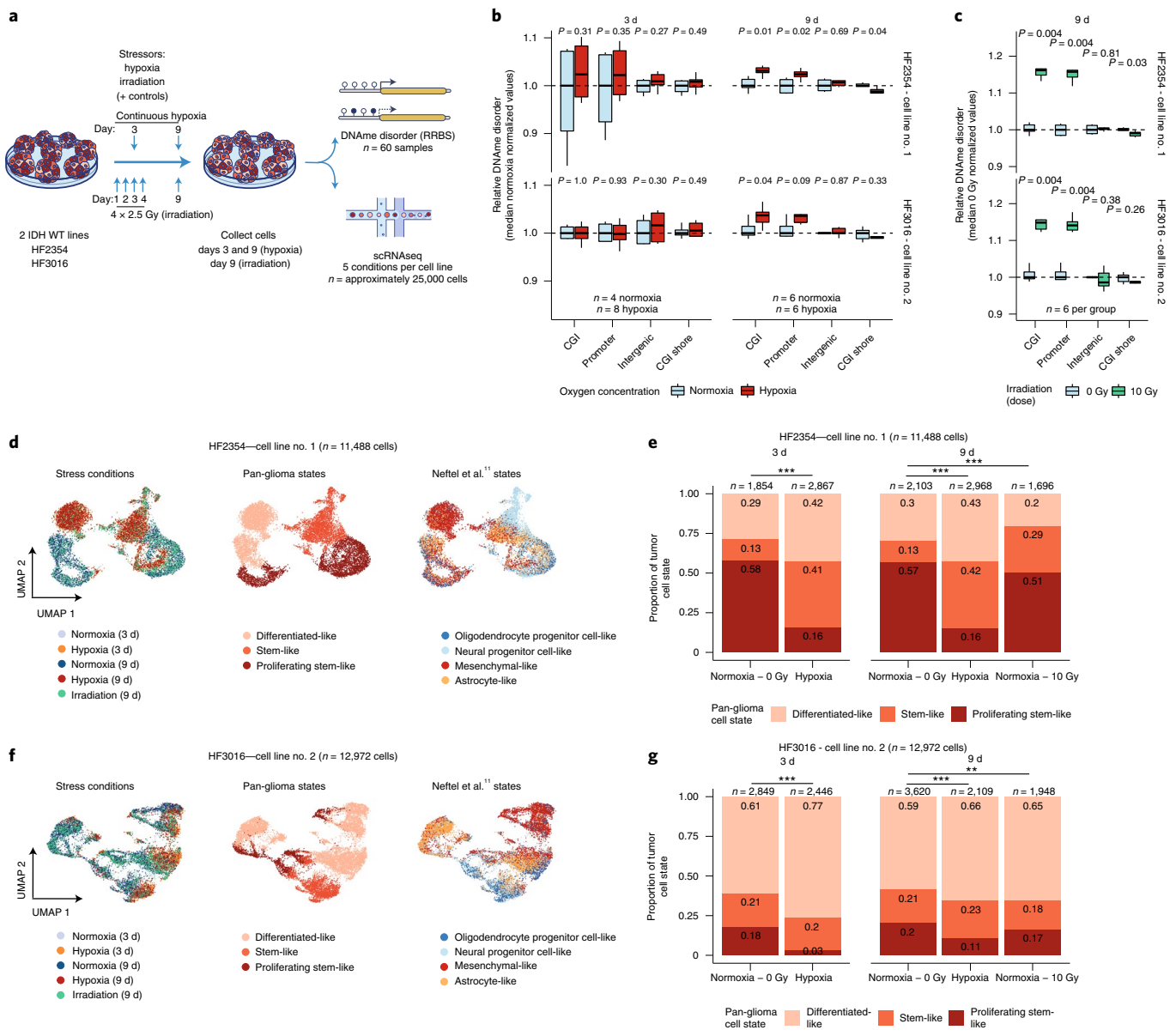


Fig. 4 | Environmental stressors increase local DNAmE disorder in vitro and are associated with cellular state shifts. **a**, In vitro experimental workflow. Patient-derived glioma sphere-forming cells were exposed to continuous stress (hypoxia, 3 and 9-d), stress followed by recovery (irradiation, 4-d stress exposure and 5-d recovery) and no stress/normoxia controls. Biological replicates were then profiled with bulk RRBS ($n = 6$ average per condition) and scRNA-seq. **b**, Box plot of relative DNAmE disorder (normalized to controls) for hypoxia at the 3-d time point (left) and 9-d time point (right). Each box spans the 25th and 75th percentile, the center lines indicate the median and the whiskers represent the absolute range (minimum/maximum), excluding outliers. Two-sided Wilcoxon rank-sum test P values are presented for different genomic contexts. Each row represents a distinct IDH WT patient-derived cell line (HF2354 and HF3016). **c**, Relative DNAmE disorder for irradiated (10 Gy) compared with no exposure controls (normoxia, 0 Gy). **d**, UMAP dimensionality reduction plot of scRNA-seq for HF2354 exposed to 3-d hypoxia, 9-d hypoxia and 9-d irradiation with no treatment controls. The inset UMAP projections are identical with different annotations to demonstrate the different stress conditions, the pan-glioma cell states described here and the proliferation-independent cell states (Neftel et al.¹¹). **e**, Stacked bar plot for the pan-glioma cell states summarized by stress exposure and time point for HF2354 ($n = 11,488$ cells). Chi-squared tests for significant changes in cellular proportions between exposures are presented ($***P < 2.2 \times 10^{-16}$, $**P < 0.01$). **f**, UMAP projection across stress conditions, pan-glioma cell states and proliferation-independent cell states for the second independent cell line HF3016 ($n = 12,972$ cells). **g**, Stacked bar plot for pan-glioma cell states for HF3016 summarized by stress exposure and time point.

To validate the relationship between SCNA and DNAmE disorder, we reanalyzed the bulk RRBS and copy number profiles of initial ($n = 255$ patients) and recurrent ($n = 129$ patients) IDH WT gliomas, including matched pairs ($n = 98$ patients)⁵. SCNA burden was positively associated with DNAmE disorder at both initial and recurrent time points, confirming our findings (Spearman correlation $R = 0.43$, $P = 3.5 \times 10^{-13}$ initial; $R = 0.33$, $P = 1.7 \times 10^{-4}$ recurrence; Fig. 5c). We repeated our analysis using only paired initial

and recurrent samples and observed a positive association between increases in SCNA burden and DNAmE disorder (Spearman correlation $R = 0.37$, $P = 0.0002$; Fig. 5d). Furthermore, the greatest changes in DNAmE disorder between initial and recurrent tumor were associated with a shorter time to second surgery in both univariate (log-rank test $P = 0.04$; Fig. 5e) and multivariate survival analyses (Cox proportional hazards model, hazard ratio (HR) = 1.55 95% confidence interval (CI) = 1.39–2.34, $P = 0.03$; Supplementary Table 3)

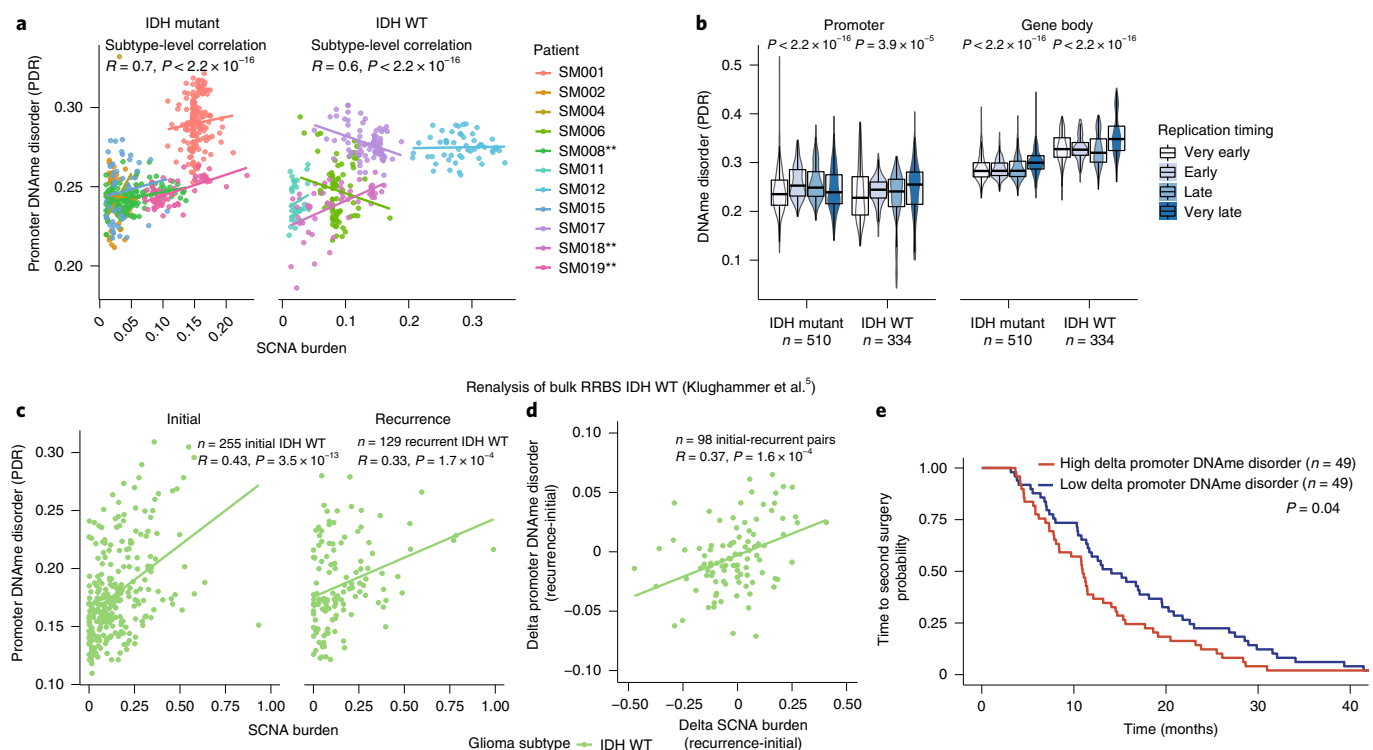


Fig. 5 | SCNAs are associated with DNAm disorder during disease evolution. **a**, Scatter plot depicting the association between single-cell-derived ($n = 790$ non-polyploid tumor cells) SCNA and promoter DNAm disorder by IDH mutant (left) and IDH WT (right) subtypes. Points and linear regression lines are colored by individual. Spearman correlation coefficients represent subtype-specific estimates. Patient-specific significant positive correlations are indicated by $**P < 0.05$ (Spearman correlation). **b**, Box plots with surrounding violins of DNAm disorder calculated across the promoter (left) and gene body regions (right) based on different DNA replication times in IDH mutant ($n = 510$) and IDH WT ($n = 334$) single cells. Each box spans the 25th and 75th percentile, the center lines indicate the median and the whiskers represent the absolute range (minimum/maximum), excluding outliers. The surrounding violins represent the distribution for each condition. The Kruskal–Wallis P values indicate a test for differences across the replication time groupings separately for IDH mutant and IDH WT cells. **c**, Scatter plot with subtype-level linear regression lines depicting the reanalysis of bulk promoter DNAm disorder and SCNA burden in IDH WT initial ($n = 255$) and recurrent ($n = 129$) tumors (Klughammer et al.⁵). Spearman correlation coefficients and P values are presented for each independent time point. **d**, Scatter plot and linear regression line depicting the association between bulk delta (patient-specific recurrence versus initial estimates) SCNA burden and delta promoter DNAm disorder in longitudinally profiled IDH WT tumors ($n = 98$ individuals, Klughammer et al.⁵). Spearman correlation coefficient and P value are presented. **e**, Kaplan–Meier curve depicting time to second surgery in individuals where the change in promoter DNAm disorder between initial and recurrent disease was above (high, red) and below (low, blue) the median. The log-rank P value for univariate analysis is presented within the figure. The HR and P value for change in DNAm disorder are presented for multivariate Cox proportional hazards model including age and sex as predictors in Supplementary Table 4.

supporting that increased epigenetic instability is associated with accelerated disease progression. We did not observe a significant positive association with overall survival (Cox proportional hazards model, HR = 1.43 95% CI = 0.93–2.20, $P = 0.10$; Supplementary Table 4). SCNA burden or aneuploidy results from errors in mitotic checkpoints, which may further perpetuate DNAm disorder and epigenetic heterogeneity through aneuploidy-induced metabolic and replication stress³¹.

Genomic alterations influence but do not define cell states. The processes driving genetic, epigenetic and transcriptomic heterogeneity may act at different times with dynamic effects on cellular state distributions. To evaluate the timing and relative impact of genetic alterations on epigenetic and transcriptomic intratumoral heterogeneity, we inferred clonal phylogenies from bulk WGS data. One to four subclonal populations were detected per tumor (Fig. 6a), with linear and branched evolutionary patterns consistent with previous reports^{1,6}. Chromosomal arm-level SCNA events were more likely to be classified as clonal/early (Fisher’s exact test $P = 0.03$; Extended Data Fig. 9a), while mutations at genes significantly mutated in glioma were more evenly distributed across subclones (56.1% classified

as clonal in non-hypermutant tumors) (Methods and Extended Data Fig. 9b–i). To determine how strongly intratumoral genetic heterogeneity is linked with epigenetic heterogeneity we compared the distribution of cell states, DNAm and DNAm disorder across single-cell copy number-based hierarchical clustering (scRRBS; Extended Data Fig. 10a–c). DNAm and DNAm disorder levels differed across copy number clusters, suggesting genetic and epigenetic coevolution (Wilcoxon rank-sum test $P < 0.05$; Extended Data Fig. 10a–c). However, LIGER-defined cell state DNAm patterns were distributed across distinct copy number profiles suggesting a convergence on shared epigenetic states. We next asked whether genetic tumor subclones were associated with transcriptional diversity. We inferred single-cell transcriptome copy number profiles and found that 3 of 11 tumors (SM001, SM006 and SM012) had at least 2 distinct clones with chromosome arm-level alterations (Fig. 6b and Extended Data Fig. 3). These tumors demonstrated significant cell state distribution shifts across clones suggesting that genetic heterogeneity also increases transcriptomic heterogeneity (per sample Fisher’s exact test $P < 0.05$; Fig. 6b). Collectively, these results suggest that large-scale copy number alterations occurring early in tumor development affect the observed epigenetic and transcriptomic diversity.

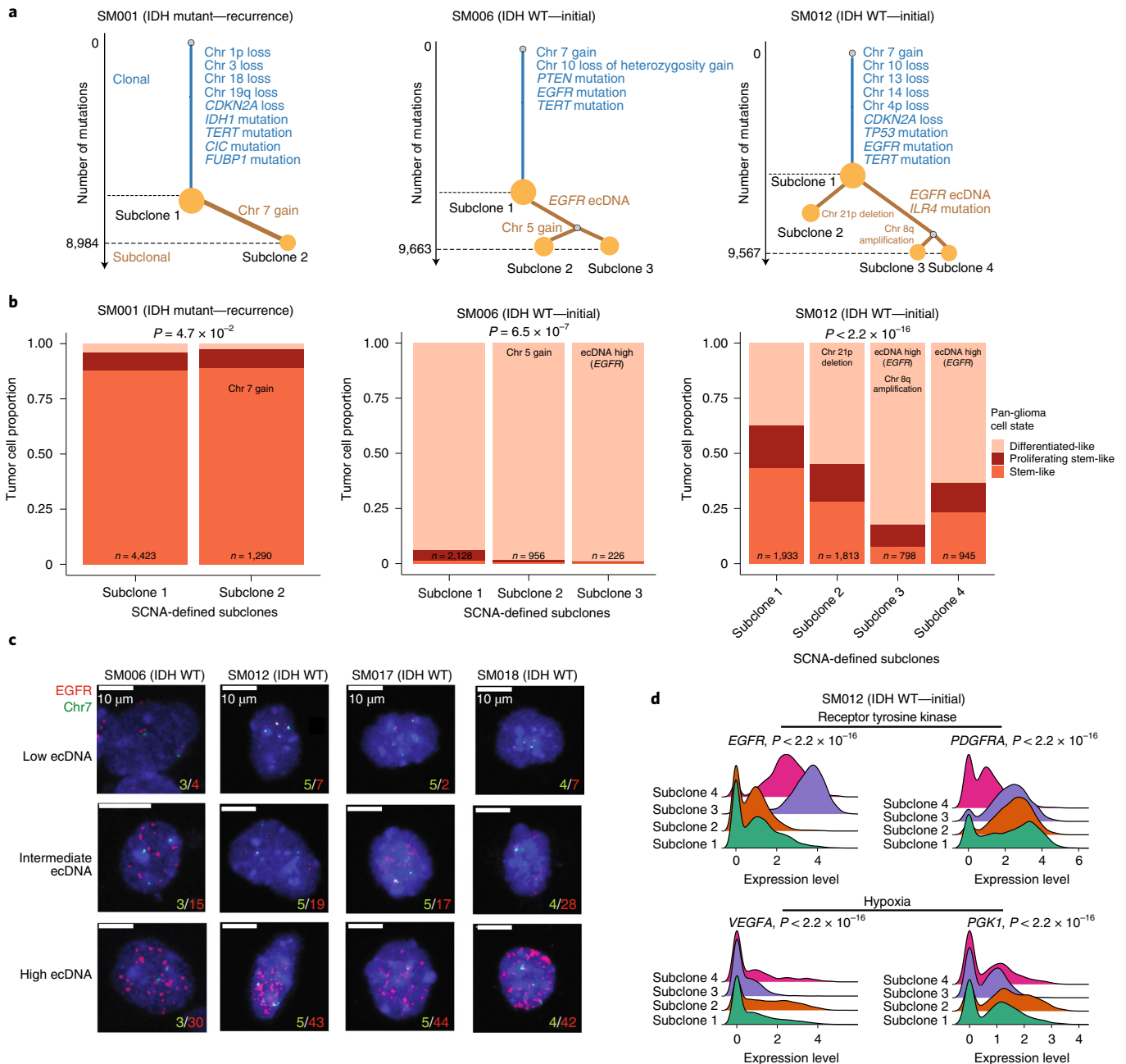


Fig. 6 | Clonal evolution analyses highlight that genetic alterations influence, but do not determine cellular states. **a**, Examples of phylogenetic trees constructed from bulk WGS data (mutations and SCNAs) and further annotated using single-cell inferred copy number alterations (scRRBS + scRNA-seq). The tree nodes represent alterations specific to the given clone, with node size corresponding to the estimated fraction of tumor cells with the associated alterations. Branch length scales with the number of mutations attributed to that clone. Clonal alterations are colored in blue, with subclonal alterations colored in gold. Genes significantly mutated in the TCGA analyses² and chromosomal arm-level gain and loss events are presented. **b**, scRNA-seq-derived cellular proportions separated by copy number-defined tumor subclone (Extended Data Fig. 3). Reported *P* values represent a two-sided Fisher's exact test comparing the cellular state distributions across tumor subclones. **c**, Representative FISH images for IDH WT tumors computationally predicted to harbor *EGFR* ecDNA by WGS ($n = 4$ patients). FISH images show *EGFR* amplifications (red) that occur distal to control chromosome 7 probes (green), indicating extrachromosomal status and high variability in copy number status across tumor cells. Scale bars, 10 μ m. **d**, Ridge plots of patient SM012 single-cell expression of receptor tyrosine kinase and hypoxia-associated genes, grouped by copy number-defined subclones. The reported *P* values represent two-sided Wilcoxon rank-sum tests comparing the gene expression of cells across tumor subclones.

Extrachromosomal DNA (ecDNA) elements in IDH WT gliomas amplify oncogenes and enhancer elements to drive genetic heterogeneity^{46–49}. We hypothesized that the impact of ecDNA on genomic heterogeneity extends to fueling epigenetic and transcriptional diversity^{48,50}. We detected ecDNAs using WGS and validated

their presence by FISH (Fig. 6c and Extended Data Fig. 10d,e). *EGFR* ecDNAs, like chromosomal arm-level events (for example, chromosome 7 amplification in SM001) distinguished subsets of tumor cells (for example, *EGFR* ecDNA in SM012) (Fig. 6b and Extended Data Fig. 10d,e). We classified both scDNA and RNA profiles as

ecDNA⁺ or ecDNA⁻ based on *EGFR* copy number level (Extended Data Fig. 10f). EcDNA⁺ cells had increased genome-wide DNAm in 3 of 4 cases (Wilcoxon rank-sum test $P < 0.05$; Extended Data Fig. 10g) and greater transcriptional diversity using gene count signatures compared with ecDNA⁻ cells (Wilcoxon rank-sum test $P < 0.05$; Extended Data Fig. 10h and Methods)⁵¹. The tumor with the highest number of genetic subclones and DNAm disorder (SM012) contained an *EGFR*-amplifying ecDNA assigned to subclones 3 and 4, which were marked by differential expression of a receptor tyrosine kinase gene signature. ecDNA⁻ subclone 2 was most closely associated with hypoxia gene expression (Wilcoxon rank-sum test $P < 2.2 \times 10^{-16}$; Fig. 6d), providing an example of how genetic heterogeneity may shape epigenetic and transcriptional reprogramming. In summary, our evolutionary analyses show that intratumoral genetic heterogeneity influences but does not determine epigenetic or transcriptomic cell states.

External pressures shape adaptive DNAm changes. We next asked whether epigenetic diversity accelerates tumor evolution by promoting cell survival in resource-deprived tumor environments (for example, hypoxia or therapeutic exposures). To address this question and extend the generalizability of our findings, we analyzed DNAm profiles from large-scale microarray-based bulk glioma studies^{3,4,52}. We inferred a microarray metric from the scDNAm data that quantified the DNAm disorder-susceptible gene regions (Fig. 7a). We reasoned that regions prone to DNAm changes would reflect this stochasticity in bulk data by assuming intermediate DNAm values (Fig. 7a). This bulk DNAm disorder metric approximated scDNAm disorder averages across our cohort (Spearman correlation $R = 0.65$ $P = 0.02$). Applying this DNAm disorder metric to The Cancer Genome Atlas (TCGA) data identified differences across TCGA-defined subtypes², with IDH WT tumors displaying the highest levels (Kruskal–Wallis $P < 2.2 \times 10^{-16}$; Fig. 7b). Integrating matching DNAm and RNA-seq samples from 568 TCGA samples showed that high bulk DNAm disorder samples showed increased transcriptional activity of oxidative stress response genes, corroborating our earlier positive associations between epigenetic instability and stress response regulation (Spearman correlation $R = 0.47$, $P < 2.2 \times 10^{-16}$, $n = 516$ IDH mutant initial tumors, $R = 0.31$, $P = 0.03$, $n = 52$, IDH WT initial tumors).

We next applied the bulk DNAm disorder metric to 119 image-guided stereotactic biopsies taken from spatially distinct regions across IDH WT ($n = 57$ biopsies, $n = 6$ patients) and IDH mutant ($n = 62$ biopsies, $n = 8$ patients) tumors⁵². This quantified the physical distance between each sample and the tumor's center, based on specific radiographic features (for example, magnetic resonance imaging contrast-enhanced region). DNAm disorder was increased closer to the tumor's center across IDH WT tumors while adjusting for patient (multivariable linear regression $P = 0.02$; Fig. 7c), a region frequently characterized by hypoxia. The link between radiographic features and epigenetic shifts supports the association between cellular fitness and increased epigenetic plasticity. We did not observe a consistent relationship between tumor location and bulk DNAm disorder in IDH mutant tumors (multivariable linear regression $P = 0.31$; Fig. 7d) where hypoxia is less prevalent.

The environmental pressures that tumors face may vary over time. We analyzed initial and recurrent tumor samples from the Glioma Longitudinal AnalySiS (GLASS) consortium for which DNA sequencing and DNAm data were available ($n = 102$ tumors, $n = 51$ patients) to relate DNAm instability to genetic alterations. We cataloged individual CpG sites where copy number or DNAm changed between the initial tumor and its matched recurrence. Overall, we observed that DNAm changes were mostly decreases in DNAm consistent with previous findings^{7,53} and that DNAm changes mainly occurred in regions that were copy number-stable

(Fig. 7e). We then tested for DNAm changes after treatment while accounting for differences in cellular composition of the tumor microenvironment (Methods). We discovered that regions with consistently altered DNAm independent of changes in microenvironment cell type distribution were enriched for the binding site motifs of transcription factors that regulate the cellular stress response, particularly hypoxia (for example, HIF1A; Fig. 7f). We also observed the enrichment for differential binding site DNAm among transcription factors that differed between stem-like and differentiated-like states in our single-cell data (for example, SP1 and TFAP2A; Figs. 3f and 7f). These observations support our single-cell findings that regions with the greatest DNAm disorder are involved with processes regulating cellular differentiation and stress signaling. In summary, we find that stochastic DNAm alterations can provide the variability necessary to enable or stabilize transition to adaptive epigenetic phenotypes that are responsive to cellular stress (Fig. 7g).

Discussion

In this study, we integrated multimodal scDNAm and transcriptomic profiles along with bulk profiles to interrogate the association between epigenetic heterogeneity, genetic alterations, cellular states and the glioma stress response. We found that early genetic alterations were associated with DNAm disorder, whose accumulation throughout the genome was linked with altered cellular states and response to environmental pressures. Elevated DNAm disorder highlights a mechanism to overcome cell stress, increase cellular plasticity and ultimately enhance treatment resistance. Taken together, epigenetic intratumoral heterogeneity provides a plastic intermediate between genetic subclones and adaptive phenotypic cell states.

Random errors in the DNAm replication machinery lead to DNAm disorder and increased intratumoral epigenetic diversity^{5,29,54}. We found that genetic and environmental stimuli further exacerbate epigenetic variability and hypothesize a convergence for both stimuli on altered cellular metabolism. Deregulated metabolism is a hallmark of both IDH mutant, which produce the oncometabolite 2-hydroxyglutarate that interferes with DNA demethylation^{2,27,55–57}, and IDH WT glioma where hypoxia is common. Additional genetic stimuli include broad chromosomal alterations that were positively associated with DNAm disorder. Through cross-platform evolutionary comparisons, we found that chromosomal alterations are early events possibly leading to the observed nongenetic diversity by generating metabolic disruption via active oxygen species³¹, thereby increasing the likelihood of aberrant DNAm. Our study shows that environmental stimuli, such as hypoxia and irradiation, increase DNAm disorder extending previous studies reporting the repressed enzymatic activity of DNAm regulators after hypoxia⁵⁸. Tumor hypoxia is common across many cancers and could more broadly shape the phenotype of cells resistant to therapy through DNAm disorder⁵⁹. Collectively, increased genomic instability and resource-poor microenvironments represent stressors that may explain the greater cell state plasticity in IDH WT relative to IDH mutant gliomas.

In a non-tumor setting, a cell's epigenome reflects the tissue of origin and serves to stabilize cell state-specific gene expression⁶⁰. A disrupted epigenetic landscape eroded by DNAm disorder may facilitate adaptive cell state transitions or increase cellular plasticity⁹. Glioma cell states fall along axes of differentiation and proliferating potential^{10–12,15}. In accordance with previous reports, we observed pan-glioma malignant cell states that were found within each tumor and in vitro models. Our single-cell epigenetic profiles revealed that cell state-defining transcription factor activity may be perturbed by DNAm disorder. Thus, diverse DNAm marks help to sustain multiple cell states that each confer their own context-dependent fitness advantages and together accelerate disease progression.

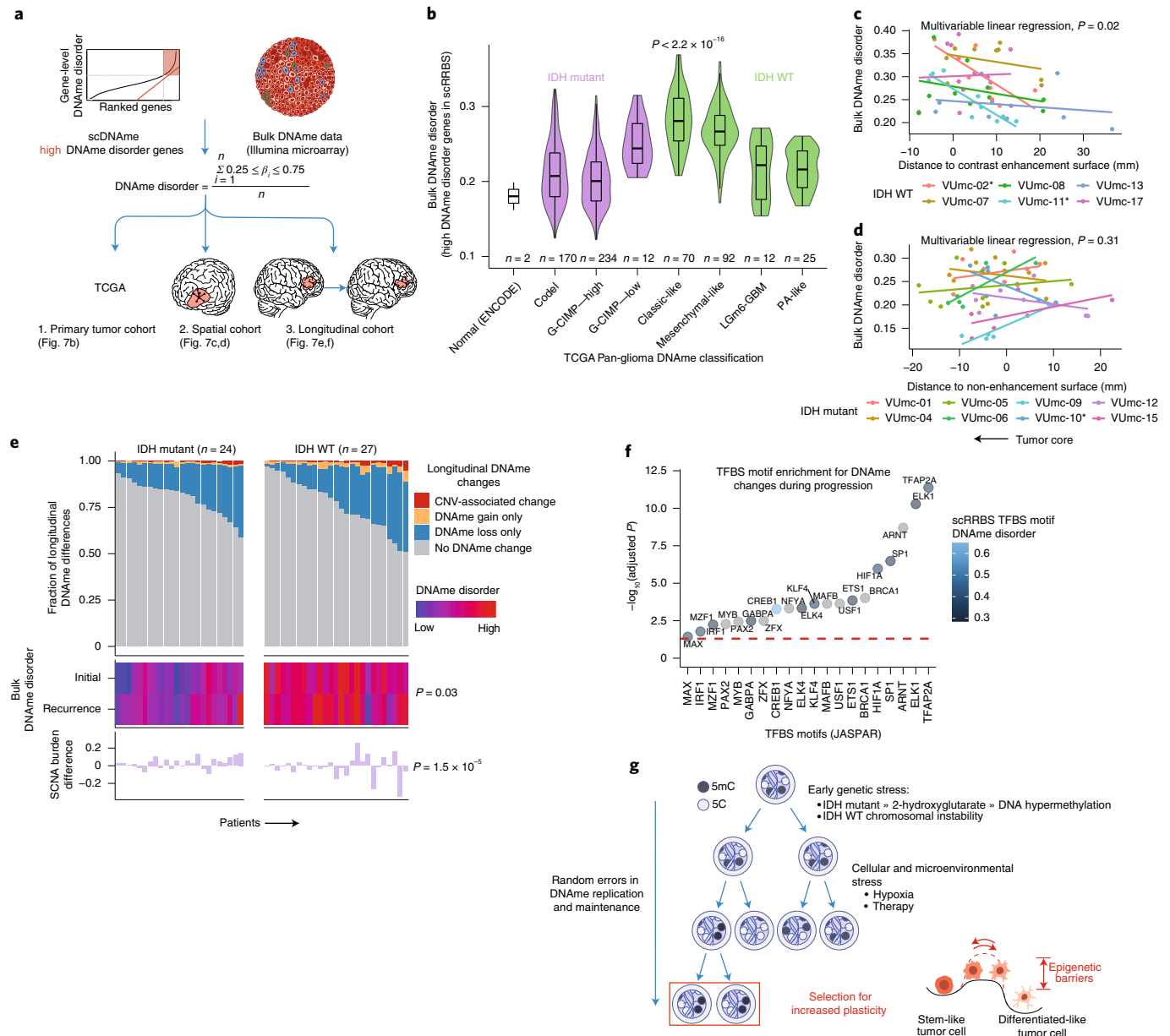


Fig. 7 | Integrated molecular trajectories support adaptive DNAm changes under microenvironmental and therapeutic stress. a, Workflow for the construction of a DNAm disorder metric in bulk cohorts informed by regions of high DNAm disorder in scDNAm data. **b**, Box plots with surrounding violins displaying the bulk DNAm disorder metric calculated across previously described DNA methylation-based TCGA classifications² and ENCODE normal cell types (astrocytes and embryonic stem cells). Each box spans the 25th and 75th percentile, the center lines indicate the median and the whiskers represent the absolute range (minimum/maximum), excluding outliers. The violins represent the distribution for each condition. A Kruskal-Wallis test for differences in distributions across the classification is reported ($n = 615$ primary gliomas, $P < 2.2 \times 10^{-16}$). **c,d**, Scatter plots depicting the distance from radiographic features plotted against the DNAm disorder. The colors represent spatially separated biopsies from a single patient at the initial clinical time point: IDH WT tumors ($n = 57$ biopsies, $n = 6$ patients) (**c**) and IDH mutant tumors ($n = 62$ biopsies, $n = 8$ patients) (**d**). The linear regression lines colored by patient demonstrate the relationship between DNAm disorder and radiographic features (that is, contrast enhancement surface). The P value reported from a multivariable linear regression model adjusting for each individual patient represents the subtype-specific association between DNAm disorder and radiographic feature. *Patient-specific correlation $P < 0.05$ (Spearman correlation). Biopsies taken closer to the tumor's center (that is, core) have the lowest x -axis value. **e**, Each column represents an individual patient sampled across initial and recurrent time points and is separated into IDH mutant ($n = 24$ patients) and IDH WT ($n = 27$ patients). Top: Stacked bar plot representing the proportion of CpGs sites that experienced DNAm change. All associated P values represent Spearman correlations between absolute change in associated metric and the fraction of longitudinal DNAm differences. **f**, Enrichment analysis for differentially methylated CpGs between initial and recurrent time points when adjusting for cellular composition, glioma subtype and patient. The individual points represent enrichment of specific transcription factors in differentially methylated positions; the color indicates the average TFBS motif DNAm disorder from scRRBS data (Fig. 2d); the dotted line represents the statistical significance threshold (adjusted $P < 0.05$). **g**, Model of DNAm disorder, stress adaptation, eroded epigenetic barriers, plasticity and glioma evolution.

Intratumoral heterogeneity in glioma reflects subclonal competition driven by limited nutrient access. While single-cell transcriptome-based phenotype studies have investigated glioma transcriptomic heterogeneity^{10–12,14,15}, we have only limited knowledge on the degree of epigenetic variability. The intratumoral epigenetic variation defined in this study provides a link between subclonal competition and phenotypic state changes by enabling diverse responses to selective pressures, such as hypoxia and treatment. Future studies are needed to fully elucidate the mechanisms by which increased DNAm disorder provides competitive advantages under stress. While we identified shared cell states that were present across different modalities, future studies employing simultaneous epigenome/transcriptome characterization will refine these cellular state classifications and identify additional determinants that shape glioma cell identity. A better understanding of therapeutically vulnerable cell states in glioma will foster development of more effective therapeutic interventions. In summary, single-cell epigenetic profiles show that diverse DNAm marks encode cellular states in glioma, permit cell state plasticity and reflect environmental stress exposures.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41588-021-00926-8>.

Received: 26 July 2020; Accepted: 27 July 2021;

Published online: 30 September 2021

References

1. Brat, D. J. et al. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N. Engl. J. Med.* **372**, 2481–2498 (2015).
2. Ceccarelli, M. et al. Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* **164**, 550–563 (2016).
3. Louis, D. N. et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol.* **131**, 803–820 (2016).
4. Barthel, F. P. et al. Longitudinal molecular trajectories of diffuse glioma in adults. *Nature* **576**, 112–120 (2019).
5. Klughammer, J. et al. The DNA methylation landscape of glioblastoma disease progression shows extensive heterogeneity in time and space. *Nat. Med.* **24**, 1611–1624 (2018).
6. Körber, V. et al. Evolutionary trajectories of IDH^{WT} glioblastomas reveal a common path of early tumorigenesis instigated years ahead of initial diagnosis. *Cancer Cell* **35**, 692–704.e12 (2019).
7. Mazar, T. et al. DNA methylation and somatic mutations converge on the cell cycle and define similar evolutionary histories in brain tumors. *Cancer Cell* **28**, 307–317 (2015).
8. Wang, Q. et al. Tumor evolution of glioma-intrinsic gene expression subtypes associates with immunological changes in the microenvironment. *Cancer Cell* **32**, 42–56.e6 (2017).
9. Flavahan, W. A., Gaskell, E. & Bernstein, B. E. Epigenetic plasticity and the hallmarks of cancer. *Science* **357**, eaal2380 (2017).
10. Bhaduri, A. et al. Outer radial glia-like cancer stem cells contribute to heterogeneity of glioblastoma. *Cell Stem Cell* **26**, 48–63.e6 (2020).
11. Neftel, C. et al. An integrative model of cellular states, plasticity, and genetics for glioblastoma. *Cell* **178**, 835–849.e21 (2019).
12. Wang, L. et al. The phenotypes of proliferating glioblastoma cells reside on a single axis of variation. *Cancer Discov.* **9**, 1708–1719 (2019).
13. Yuan, J. et al. Single-cell transcriptome analysis of lineage diversity in high-grade glioma. *Genome Med.* **10**, 57 (2018).
14. Tirosh, I. et al. Single-cell RNA-seq supports a developmental hierarchy in human oligodendrogloma. *Nature* **539**, 309–313 (2016).
15. Venteicher, A. S. et al. Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* **355**, eaai8478 (2017).
16. Easwaran, H., Tsai, H.-C. & Baylin, S. B. Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance. *Mol. Cell* **54**, 716–727 (2014).
17. Liao, B. B. et al. Adaptive chromatin remodeling drives glioblastoma stem cell plasticity and drug tolerance. *Cell Stem Cell* **20**, 233–246.e7 (2017).
18. Gaiti, F. et al. Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. *Nature* **569**, 576–580 (2019).
19. Hernandez-Herrera, I. et al. Ageing affects DNA methylation drift and transcriptional cell-to-cell variability in mouse muscle stem cells. *Nat. Commun.* **10**, 4361 (2019).
20. Johnson, K. C. et al. 5-Hydroxymethylcytosine localizes to enhancer elements and is associated with survival in glioblastoma patients. *Nat. Commun.* **7**, 13177 (2016).
21. Angermueller, C. et al. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat. Methods* **13**, 229–232 (2016).
22. Argelaguet, R. et al. Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature* **576**, 487–491 (2019).
23. Farlik, M. et al. DNA methylation dynamics of human hematopoietic stem cell differentiation. *Cell Stem Cell* **19**, 808–822 (2016).
24. Bian, S. et al. Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science* **362**, 1060–1063 (2018).
25. Guo, H. et al. Profiling DNA methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing. *Nat. Protoc.* **10**, 645–659 (2015).
26. Guo, H. et al. Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res.* **23**, 2126–2135 (2013).
27. Turcan, S. et al. IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* **483**, 479–483 (2012).
28. Kelsey, G., Stegle, O. & Reik, W. Single-cell epigenomics: recording the past and predicting the future. *Science* **358**, 69–75 (2017).
29. Landau, D. A. et al. Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. *Cancer Cell* **26**, 813–825 (2014).
30. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
31. Zhu, J., Tsai, H.-J., Gordon, M. R. & Li, R. Cellular stress associated with aneuploidy. *Dev. Cell* **44**, 420–431 (2018).
32. Hughes, L. A. E. et al. The CpG island methylator phenotype: what's in a name? *Cancer Res.* **73**, 5858–5868 (2013).
33. Luo, Y., Lu, X. & Xie, H. Dynamic Alu methylation during normal development, aging, and tumorigenesis. *Biomed. Res. Int.* **2014**, 784706 (2014).
34. Yin, Y. et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, eaaj2239 (2017).
35. MacLeod, G. et al. Genome-wide CRISPR-Cas9 screens expose genetic vulnerabilities and mechanisms of temozolomide sensitivity in glioblastoma stem cells. *Cell Rep.* **27**, 971–986.e9 (2019).
36. Jin, X. et al. Targeting glioma stem cells through combined BMI1 and EZH2 inhibition. *Nat. Med.* **23**, 1352–1361 (2017).
37. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
38. Welch, J. D. et al. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* **177**, 1873–1887.e17 (2019).
39. Orso, F. et al. Identification of functional TFAP2A and SP1 binding sites in new TFAP2A-modulated genes. *BMC Genomics* **11**, 355 (2010).
40. Li, Z. et al. Hypoxia-inducible factors regulate tumorigenic capacity of glioma stem cells. *Cancer Cell* **15**, 501–513 (2009).
41. Shaffer, S. M. et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* **546**, 431–435 (2017).
42. Sharma, S. V. et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* **141**, 69–80 (2010).
43. Peng, C. et al. Cyclin-dependent kinase 2 (CDK2) is a key mediator for EGF-induced cell transformation mediated through the ELK4/c-Fos signaling pathway. *Oncogene* **35**, 1170–1179 (2016).
44. Kent, L. N. & Leone, G. The broken cycle: E2F dysfunction in cancer. *Nat. Rev. Cancer* **19**, 326–338 (2019).
45. Koren, A. et al. Differential relationship of DNA replication timing to different forms of human mutation and variation. *Am. J. Hum. Genet.* **91**, 1033–1040 (2012).
46. deCarvalho, A. C. et al. Discordant inheritance of chromosomal and extrachromosomal DNA elements contributes to dynamic disease evolution in glioblastoma. *Nat. Genet.* **50**, 708–717 (2018).
47. Morton, A. R. et al. Functional enhancers shape extrachromosomal oncogene amplifications. *Cell* **179**, 1330–1341.e13 (2019).
48. Wu, S. et al. Circular eDNA promotes accessible chromatin and high oncogene expression. *Nature* **575**, 699–703 (2019).
49. Kim, H. et al. Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. *Nat. Genet.* **52**, 891–897 (2020).
50. Verhaak, R. G. W., Bafna, V. & Mischel, P. S. Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. *Nat. Rev. Cancer* **19**, 283–288 (2019).
51. Gulati, G. S. et al. Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* **367**, 405–411 (2020).

52. Verburg, N. et al. Spatial concordance of DNA methylation classification in diffuse glioma. *Neuro. Oncol.*, <https://doi.org/10.1093/neuonc/noab134> (2021).
 53. de Souza, C. F. et al. A distinct DNA methylation shift in a subset of glioma CpG island methylator phenotypes during tumor recurrence. *Cell Rep.* **23**, 637–651 (2018).
 54. Landan, G. et al. Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nat. Genet.* **44**, 1207–1214 (2012).
 55. Losman, J.-A. & Kaelin, W. G. Jr. What a difference a hydroxyl makes: mutant IDH, (R)-2-hydroxyglutarate, and cancer. *Genes Dev.* **27**, 836–852 (2013).
 56. Dang, L. et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739–744 (2009).
 57. Noushmehr, H. et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* **17**, 510–522 (2010).
 58. Thienpont, B. et al. Tumour hypoxia causes DNA hypermethylation by reducing TET activity. *Nature* **537**, 63–68 (2016).
 59. Heddeleston, J. M. et al. Hypoxia inducible factors in cancer stem cells. *Br. J. Cancer* **102**, 789–795 (2010).
 60. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
- Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
- © The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

Methods

Experimental methods. *Description of human tumor specimens.* Human glioma resection specimens were obtained with informed consent from the University of Connecticut Health Center and St. Michael's Hospital. All tissue donations were approved by the institutional review board of the Jackson Laboratory and clinical institutions involved. This work was performed with ethics board approval (no. 2018-NHSR-018) and in accordance with the Declaration of Helsinki principles. Patients were not compensated for their participation in this research study. Initial pathological diagnosis was confirmed with tumor DNAm classification according to the MolecularNeuropathology tool⁶¹. The clinical characteristics of this population are provided in Supplementary Table 1.

Sample preparation and sorting for the single-cell experiments. Tumor specimens were collected directly from the operating room and immediately placed into MACS tissue storage solution at 4 °C (catalog no. 130-100-008; Miltenyi Biotec). Tumor specimens from the same spatial region were then minced and partitioned into single-cell and bulk fractions (Fig. 1a). Any remaining tumor tissue was deposited into freezing media consisting of 90% heat-inactivated fetal bovine serum FBS (Invitrogen) and 10% dimethyl sulfoxide (Sigma-Aldrich), and gradually frozen in a freezing container (Mr. Frosty; Corning) over 24 h before being stored in liquid nitrogen for future experiments (that is, FISH). Bulk tissue specimens were immediately flash-frozen for subsequent DNA and RNA extraction. The specimen fraction for single-cell analyses was further mechanically and enzymatically dissociated using the Brain Tumor Dissociation Kit (P) according to the manufacturer's protocol (catalog no. 130-095-942; Miltenyi Biotec)^{11,14,15}.

Single-cell suspensions were blocked with human BD Fc Block (BD Biosciences) for 5 min on ice before antibody staining and labeled via incubation with 1:100 dilution of Alexa Fluor 488-conjugated anti-CD45 antibody (catalog no. 304017; BioLegend) and 1:100 dilution of PE-Cy7-conjugated anti-CD31 antibody (catalog no. 303117; BioLegend) for 30 min at 4 °C. Cells were washed with Hank's buffered saline solution (HBSS) and resuspended in 2 mM of EDTA/2% BSA/PBS buffer containing 2 µg ml⁻¹ of propidium iodide (PI) (catalog no. 556364; BD Biosciences) and 1 µM of calcein violet (Invitrogen) for 20 min at 4 °C. FACS was performed using a BD FACSAria Fusion instrument with a 130-µm nozzle and using the lowest event rate. Single-cell mode was selected to further ensure the stringency of sorting. Fluorescence compensation and FACS data visualization was performed using FlowJo v.10.3 (<https://www.flowjo.com/>). For the generation of 10x sequencing libraries, 50,000–150,000 PI⁻, calcein⁺ viable single cells were collected in 20% FBS/HBSS buffer. CD45⁺ cells were limited to no more than 20% of the total viable sort to enrich for tumor cells (Extended Data Fig. 2a). For the generation of scDNAm libraries, we sorted viable (PI⁻ and calcein⁺), non-immune (CD45⁻), and non-endothelial (CD31⁻) cells into 96-well plates that were preloaded with 5 µl of 1× Tris-EDTA buffer (Extended Data Fig. 2a). Once cells had been sorted, the 96-well plates were either immediately processed through the scDNAm protocol or frozen and stored at -80 °C.

scRRBS library preparation. scDNAm profiling was performed using a modified version of a previous scRRBS protocol^{25,26}. The scDNAm experiments were performed with sorted viable, non-immune, non-endothelial (PI⁻, calcein⁺, CD45⁻, CD31⁻) cells in a 96-well plate containing 5 µl of preloaded Tris-EDTA buffer with an empty well control. For 9 out of 11 tumors, the protocol was also applied to a small control population of 50 cells (PI⁻, calcein⁺, CD45⁻, CD31⁻). Sorted 96-well plates were frozen at -80 °C until processing when cells were lysed with 0.2 µl of 1 M of KCl (Sigma-Aldrich), 0.2 µl of 10% Triton X-100 (Sigma-Aldrich), 0.3 µl of 20 mg ml⁻¹ of protease (QIAGEN) and nuclease-free water in a total volume of 6 µl for 3 h at 50 °C. The protease was then heat-inactivated at 75 °C for 15 min. The DNA was incubated with 50 units of MspI (New England Biolabs) and TaqI (New England Biolabs) with CutSmart Buffer (New England Biolabs) for 3 h at 37 °C; 60 fg of unmethylated bacteriophage lambda DNA (Promega Corporation) was added to each well to serve as a control for bisulfite conversion efficiency assessment. The solution was heated to 80 °C for 20 min to heat-inactivate the restriction enzymes and placed on ice. Five units of Klenow fragment (3' → 5' exo-; New England Biolabs), CutSmart Buffer and end-repair deoxynucleoside triphosphate (dNTP) mix (40 µM of dATP, 4 µM of dGTP and 4 µM of dCTP; New England Biolabs) totaling 2 µl per reaction were added to perform end-repair and dA-tailing. Then, 1:250× diluted NEXTflex methylated adapters (Bioo Scientific) were added to each quadrant of the 96-well plate ($n = 24$ unique adapters) with a ligation mixture of 40 Weiss U T4 ligase (New England Biolabs), 1 mM of ATP (Thermo Fisher Scientific) and nuclease-free water to a final volume of 4 µl per reaction. TruSeq methylated adapters (Illumina) were also used in one sample (SM001) using the same protocol. The ligation reaction proceeded at 16 °C for 30 min followed by incubation at 4 °C for at least 8 h. The ligation reaction was stopped by heat-inactivation at 65 °C for 20 min. After adapter ligation, 24 individual cells with unique ligated adapters were pooled from each plate quadrant for the protocol's remainder. Excess adapter was removed using a 1:1 volumetric ratio of AMPure beads (Beckman Coulter). Bisulfite conversion was performed using the EZ DNA methylation kit (Zymo Research) according to the manufacturer's instructions except with one-half volumes due to reduced DNA

input. The solution was incubated at 98 °C for 8 min, 64 °C for 3.5 h and held at 4 °C once the reaction was complete. Then, 10 ng of transfer RNA (Roche) was added before column elution to serve as a protective carrier. PCR enrichment was performed using the PfuTurbo Cx HotStart (Agilent Technologies), PfuTurbo Cx HotStart Buffer (Agilent Technologies), primer mix (Bioo Scientific), dNTP mix (Promega Corporation) and nuclease-free water under the following conditions: 95 °C for 2 min, 32 cycles at 95 °C for 20 s, 60 °C for 30 s and 72 °C for 60 s. The PCR reaction was terminated by incubating at 72 °C for 5 min. The libraries were purified in a 1:1 volumetric ratio of AMPure beads. Pippin size selection was performed between 200 and 1,000 base pairs (bp) (Sage Science) and quantified by quantitative PCR (qPCR) (Kapa Biosystems). scRRBS libraries were paired-end-sequenced alongside bulk whole-genome libraries on an Illumina HiSeq 4000 using 1% PhiX spike-in and 75-bp reads.

scRNA library preparation. Sorted cells were washed and resuspended in 0.04% BSA/PBS buffer. Cells were counted on a Countess 2 automated cell counter and were loaded on a Chromium chip with a target cell recovery of 6,000 cells per lane. Sequencing libraries were performed using the single-cell 3' messenger RNA kit according to the manufacturer's protocol (10x Genomics). Complementary DNA and library quality were examined on a 4200 TapeStation (Agilent Technologies) and quantified by qPCR (Kapa Biosystems). Illumina sequencing (NovaSeq) was performed using a paired-end 100-bp protocol. Libraries were sequenced to a median depth of 50,000 unique reads per cell.

WGS of tumors and matched normal blood. Genomic DNA was extracted from the same tumor region as the single-cell analyses using the QIAGEN AllPrep kit and matched to normal blood using the DNeasy kit (QIAGEN). Briefly, 400 ng of DNA was sheared to 400 bp using an LE220 focused-ultrasonicator (Covaris) and size-selected using Solid Phase Reversible Immobilization beads (Beckman Coulter). The fragments were treated with end-repair A-tailing and ligation of unique adapters (Illumina) using the KAPA mRNA HyperPrep Kit (Roche). This was followed by five cycles of PCR amplification. DNA sequencing was performed using a paired-end 75-bp protocol according to the manufacturer's instructions (Illumina HiSeq 4000). Tumor samples were sequenced to an average depth of 44×; tumor-matched normal blood was sequenced to an average depth of 30×.

Bulk Illumina EPIC DNAm microarrays. A total of 250 ng of genomic tumor DNA was subjected to bisulfite conversion using the EZ DNA methylation kit (Zymo Research) and genome-wide DNAm was assessed by the Infinium MethylationEPIC Kit according to the manufacturer's protocol (Illumina).

Bulk RNA-seq. Bulk tumor RNA was extracted from samples with sufficient tissue using the AllPrep kit. Samples with RNA integrity number values > 5 as assessed by TapeStation were prepared with the KAPA mRNA HyperPrep Kit. Libraries were sequenced using a paired-end 150-bp protocol on a NovaSeq system to 50 million reads according to the manufacturer's protocol (Illumina).

FISH analysis. Tissue slides were prepared by tumor touch prep method⁴⁶. Positively charged glass slides were pressed against the surface of thawed frozen tissues. The slides were then immediately fixed by cold Carnoy's fixative (3:1 methanol:glacial acetic acid v/v) for 30 min and then air-dried. Slides were denatured in hybridization buffer (Empire Genomics) mixed with EGFR-Chr7 probe (EGFR-CHR07-20-ORGR; Empire Genomics) at 75 °C for 5 min and then incubated at 37 °C overnight. The posthybridization wash was with 0.4× saline-sodium citrate at 75 °C for 3 min followed by a second wash with 2× saline-sodium citrate/0.05% Tween 20 for 1 min. The slides were then briefly rinsed with water and air-dried. The VECTASHIELD mounting medium with 4',6-diamidino-2-phenylindole (Vector Laboratories) was applied and the coverslip was mounted onto a glass slide. Tissue images were scanned under a Leica STED 3X/DLS confocal microscope with 100× magnification.

Glioma sphere-forming cell lines and in vitro perturbations. Patient-derived IDH WT spheroids (HF2354 and HF3016) were cultured in neurosphere medium (NMGF): 500 ml of DMEM/F-12 medium (catalog no. 11330032; Gibco) supplemented with N-2 (catalog no. 17502-048; Gibco), 250 mg of BSA (catalog no. A4919; Sigma-Aldrich), 12.5 mg of gentamicin reagent (catalog no. 15710-064; Gibco), 2.5 ml of antibiotic/antimycotic (Invitrogen), 20 ng ml⁻¹ of EGF (catalog no. AF-100-15; PeproTech) and 20 ng ml⁻¹ of bFGF (catalog no. 100-18B; PeproTech). Previous comprehensive characterization of patient tumor and matched patient-derived spheroids demonstrated faithful propagation of genomic and transcriptomic profiles to the cell lines⁴⁶.

To induce hypoxia, glioma cells were cultured in hypoxia chambers (Thermo Fisher Scientific) under atmospheric normoxic (21% O₂) and hypoxic (2 and 1% O₂) conditions. Cells assayed at the 3-d time point were cultured under the three different oxygen conditions ($n = 4$ per group). The 9-d time point data were restricted to 21 and 1% ($n = 6$ per group) after observing oxygen concentration dosage effects at 3-d (Extended Data Fig. 9a,b). The 3 and 9-d time points were selected due to the 2–3-d doubling time of these cell lines. Irradiation was delivered using the Gammacell 1000A Blood Irradiator (Atomic Energy of Canada Limited)

at a daily dose of 2.5 Gy for 4 consecutive days for a total of 10 Gy followed by a recovery period of 5 d⁶². Irradiated cells were collected at the 9-d time point ($n=6$ per group). DNA and RNA for all conditions were isolated using the AllPrep DNA/RNA Mini Kit (QIAGEN).

Patient-derived spheroid candidate gene expression. Real-time PCR was performed with primers specific for candidate cell state (*SOX2*, *POU5F1*), stress response (*JUN*) and hypoxia marker genes (*EPAS1* (HIF2A) and *VEGFA*) at the 3-d time point. Relative gene expression was normalized to the housekeeping genes *ACTB* and *B2M*. All primers were purchased from Integrated DNA Technologies. Primer sequences are provided in Supplementary Table 5.

Cell line RRBS. Sequencing library preparation was performed using the Premium RRBS Kit (catalog no. C02030033; Diagenode) according to the manufacturer's protocol. Libraries were sequenced on a NovaSeq 6000 using a paired-end 2 × 100 strategy for all replicates ($n=60$). The mean sequencing depth was 38 million reads per sample.

Patient-derived cell line scRNA-seq. For each cell line and time point, both a perturbed (that is, hypoxia and irradiated) and a control replicate were dissociated into single cells for single-cell gene expression profiling. Briefly, cells were collected at the 3 and 9-d time points for the hypoxia experiments and the 9-d time point for the irradiation experiments. To minimize batch effects when comparing perturbed and control conditions, cells were labeled with oligonucleotide-tagged antibodies, flow sorted to enrich for viable cells and multiplexed on the same 10x Chromium lane (10x Genomics)⁶³. Libraries were sequenced using the single-cell 3' mRNA kit in the same manner described for the patient tumor specimens.

Analytical methods. Single-cell and cell line DNAm processing. Raw sequencing reads were trimmed to remove adapters and low-quality bases using TrimGalore v.0.4.0 with the `--rrbs` and `--paired` parameters (<https://github.com/FelixKrueger/TrimGalore>). The trimmed reads were then aligned to the GRCh37 (hg19) genome using Bismark v.0.19.1 with the parameters `-N 1 --bowtie2 --score_min L,0,-0.4` (ref. ⁶⁴). For single-cell data, PCR duplicates were removed with the `deduplicate_bismark` command. Bisulfite conversion efficiency was determined using the spike-in unmethylated lambda DNA. For single-cell data, cells with fewer than 40,000 unique CpGs detected and bisulfite conversion rates below 95% were removed from the analysis. A total of 914 single cells were retained for downstream analysis ($n=914$ out of 1,076 total cells sequenced) with a mean of 145,000 CpGs per cell and mean bisulfite conversion rate of 98.4% (Supplementary Table 6). All 60 cell line RRBS samples were retained for downstream analysis, with a mean of 10,228,198 unique CpGs per sample and a mean bisulfite conversion rate of 98.9% (Supplementary Table 7).

Unsupervised clustering of scRRBS data. Unsupervised clustering of the DNAm data was performed using pairwise comparisons of individual CpGs across all cell–cell comparisons (PDclust, v.0.1.0)⁶⁵. Briefly, this method performs pairwise comparisons of single-CpG methylation measurements to create a pairwise dissimilarity value that reflects the average absolute difference in methylation values at CpGs covered in any two cells. The pairwise dissimilarity values were used as input features for the MDS analysis for which visualization of cells in close proximity reflected greater similarity than cells further apart (Fig. 1b).

DNAm disorder as a measure of epigenome instability. DNAm disorder was determined by identifying DNAm concordance of nearby CpGs on a single sequencing read for bulk and scDNAm data^{18,29,54}. Briefly, for a sequencing read to be considered for this analysis, it required a minimum of four CpGs located on the same sequencing read. Sequencing reads containing at least four CpGs, referred to as 'epialleles', were extracted from aligned BAM files using SAMtools v.1.9 (ref. ⁶⁶) for downstream analysis. An epiallele was considered discordant if any of the CpGs on that sequencing read had different methylation states (for example, three methylated CpGs and an unmethylated CpG). The DNAm disorder metric reflects the sum of discordant epialleles divided by the total number of epialleles considered for analysis (that is, the proportion of discordant reads)^{18,29,54}. The DNAm disorder metric can be calculated across the entire genome (that is, DNAm disorder) or restricted to specific genomic regions where the metric considers only the epialleles overlapping that particular genomic context. A linear regression model was used to assess the impact of the total number of epialleles considered for analysis on the DNAm disorder. The DNAm disorder metric was very weakly associated with epiallele count in that an additional 10,000 epialleles were associated with an 0.001 increase in the DNAm disorder metric. For the analyses associating DNAm disorder with metrics derived from bulk WGS data, the sample-level DNAm disorder was calculated as the median of the scDNAm disorder values. For the analyses of the patient-derived cell line RRBS data, DNAm disorder was calculated separately for each CpG by determining the proportion of discordant reads overlapping the given CpG. CpGs with a sequencing depth less than 20× or at least 100 times the 95th percentile of sequencing depth were excluded from analysis.

DNAm and DNAm disorder over genomic annotations. To determine region-specific DNAm or DNAm disorder, measured CpGs or epialleles were intersected with the genomic coordinates of interest before methylation value or DNAm disorder calculation, respectively. For the analyses of patient-derived cell line RRBS data, region-specific DNAm disorder was calculated as the weighted average of per-CpG DNAm disorder values, with weights proportional to sequencing depth. All coordinates were mapped against the hg19 human genome assembly. Regions of interest considered for the analyses included CGIs, adjacent CGI shores, promoter, gene body, intergenic, Alu repeat, normal cell-specific CTCF and EZH2 binding sites (ENCODE: normal human astrocyte and H1 embryonic stem cells), DNase I hypersensitivity regions, TFBS motifs, replication timing domains and 5 and 10 kilobase (kb) tiled regions. CGI shores were defined as ±2 kb from the CGI. Promoters were defined as 1 kb upstream and 500 bp downstream of FANTOM5 (ref. ⁶⁷). Transcription start sites (TSS) were mapped to Ensembl release 96 genes. If multiple TSS mapped to a given gene, the TSS with the lowest genomic coordinate was selected. Gene body annotations were obtained from the Ensembl Genome Browser, release 96⁶⁸. Intergenic regions were annotated by selecting regions not overlapping Ensembl gene body coordinates. DNase I hypersensitivity region annotations were obtained from the UCSC Genome Browser database, 2019 update⁶⁹. TFBS motifs were obtained from the JASPAR 2020 Core Vertebrate database⁷⁰ of nonredundant transcription factor binding motifs. Each binding site was assigned a score of 0–1,000, which corresponded to the P value for the relative position weight matrix score of a TFBS motif prediction. For a given transcription factor, all identified target binding site coordinates were aggregated; binding sites were excluded if they had a relative score of less than 400, corresponding to $P > 0.0001$, or if any binding site lacked a CpG dinucleotide. TFBS motif DNAm disorder analyses required that a given epiallele included at least one CpG overlapping the TFBS motif; subsequently, epialleles considered for analysis included both CpGs within and adjacent to the motif. Analysis of DNAm disorder grouping CpGs by whether they lay at or adjacent to motifs revealed consistent DNAm disorder across epialleles overlapping TFBS motifs. Replication timing of genes was retrieved from MutSigCV v.1.0⁷¹ and gene-specific annotations for the replication timing domains were generated by binning gene coordinates into quartiles based on the replication timing score. Methylation values were also calculated for nonoverlapping windows of 5 or 10 kb. The ranks of high DNAm disorder levels were determined with the ROSE software (https://bitbucket.org/young_computation/rose) for both gene-level and TFBS.

SCNA estimation from scDNAm data. To provide evidence for SCNAs in scDNAm sequencing data, the Ginkgo algorithm⁷² was applied to single cells that passed the scRRBS quality control filters mentioned above. Briefly, this method bins mapped reads by chromosomal location, performs Lowess normalization to correct for GC biases, adjusts for potential amplification artifacts and segments the genome to determine chromosomal regions with consistent copy number states. In this study, the genome for each sample was divided into 2,597 variable-length bins with a median length of 1 megabase. Segmentation was performed using independent normalized read counts and the parameter mask bad bins (that is, bins with consistent pileups) was enabled. Cells were considered non-tumor if <1% of the genome had a copy number state that was not 2. Copy number plots were generated using the R package gplots v.3.0.1.1. Hierarchical clustering and annotation of single-cell SCNA profiles was performed using the dendextend v.1.13.4. R package⁷³.

scRNA processing and analysis. The Cell Ranger pipeline v.3.0.2 was used to convert Illumina base call files to FASTQ files and align FASTQs to the hg19 10x reference genome v.1.2.0. Preprocessing was performed using the Scanpy package v.1.3.7 (ref. ⁷⁴). The gene expression profiles of each cell at the 1,500 most highly variable genes (as measured by dispersion⁷⁵) were used for neighborhood graph generation (using 33 nearest neighbors) and dimensionality reduction with uniform manifold approximation and projection (UMAP)⁷⁶. Clustering was performed on this neighborhood graph using the Leiden community detection algorithm⁷⁷. The neighborhood graph was batch-corrected using the batch correction software BBKNN v.1.2.0⁷⁸. These defined clusters were then labeled with particular cell states based on marker gene expression and previously described cell states^{10,11,14}. A similar analytical framework was also applied to each of the two patient-derived spheroid scRNA-seq datasets, each using a different number of most variable genes and nearest neighbors. Cell state classification of malignant cells was also performed using previously developed classifiers for both IDH WT¹¹ and IDH mutant tumors¹⁵. The Seurat R package, v.3.1.1, was used for downstream analyses and visualizations⁷⁹. Inference of gene regulatory networks was performed using SCENIC, v.1.1.2-2, for a random set of 5,000 cells per analysis, with only 9-d stress cells presented in Extended Data Fig. 8d–f⁷⁷. SCNA estimation from scRNA-seq data was performed using InferCNV v.1.6.0 (refs. ^{11,14,15}). Briefly, the InferCNV method provides evidence for large-scale SCNAs by comparing averaged gene expression intensity values with non-tumor cells (based on marker gene expression) from the same specimen. Subclusters of cells were partitioned into clones on the basis of shared copy number patterns (<https://github.com/broadinstitute/inferCNV>). Single-cell gene set activity was determined using AUCell v.1.12.0 (ref. ³⁷). scRNA diversity comparisons using gene count signatures

were performed using the R package CytoTRACE v.0.1.0 across cells from the same tumor clone⁵¹.

Joint scRNA and scDNAme integration. scRNA and DNAm malignant cell profiles were integrated within the same specimen by jointly clustering gene expression with gene-level methylation features using the R package liger v.0.4.2 (ref. ³⁸).

Analysis of publicly available brain tumor DNAm data. Data reanalysis of longitudinal glioma resources was accessed for Klughammer et al.⁵ (<http://www.medical-epigenomics.org/papers/GBMatch/>) and GLASS (<http://synapse.org/glass>)⁴. Magnetic resonance imaging-guided biopsies taken from spatially distinct regions and subjected to bulk DNAm Illumina methylation microarray collected by our group can be accessed at EGAS00001005434 (ref. ⁵²). DNAm microarrays (HumanMethylation450 BeadChip) were retrieved from the TCGA initial glioma samples². All Illumina methylation microarrays were processed using the R package minfi v.1.30.0. The recurrent DNAm changes between the initial and recurrent tumors were determined by fitting a linear mixed effect model (R nlme package v.3.1-140) to each individual CpG modeled as a logit-transformed M-value with independent variables of time point, subtype, cancer cell proportion, immune proportion and patient included as the random effect. The cancer and immune cell proportions in the GLASS bulk Illumina methylation microarray data were determined using the glioma signature in the R package MethylCIBERSORT v.0.2.0 (ref. ⁸⁰).

Gene and genomic region enrichment analyses. Enrichment of genes were performed using the R package topGO v.2.42.0. Enrichment of genomic regions was determined using the Locus Overlap Analysis (LOLA) v.1.14.0 R package⁸¹. LOLA enrichment analyses used all features considered for analysis as the background sets.

Variant detection and copy number calling. Variant detection and bulk copy number determination was performed in accordance with the Genome Analysis Toolkit (GATK) best practices using GATK v.4.1.0.0 (Mutect2). Bulk tissue sequencing computational pipelines were developed using snakemake v.5.2.2 (ref. ⁸²).

Mutational signature identification. Mutational signatures were identified in bulk WGS samples using the MutationalPatterns R package v.1.10.0 (ref. ⁸³). The trinucleotide context of single base substitutions was extracted for each sample to construct a mutational profile. For each mutational profile, the contribution of mutational signatures from the Catalogue of Somatic Mutations in Cancer v3 was quantified. Known signatures were ranked by order of relative contribution to the sample mutational profile; for visualization, the top five signatures per sample were listed, with the remaining signatures collapsed into an 'Other' category.

Phylogenetic reconstruction bulk WGS clonality. To reconstruct the evolutionary history and subclonal composition of tumors, PhyloWGS v.1.0-rc2 (ref. ⁸⁴) was applied to the bulk WGS samples. PhyloWGS incorporates SCNAs with simple somatic mutations in inferred phylogenies by converting SCNAs into pseudo-simple somatic mutations before subclonal reconstruction. For input, PhyloWGS requires VCF format variant calls, SCNA segments and estimates of tumor purity, which were generated using Mutect2 v.4.1.0.0 (ref. ⁷¹), TITAN v.1.19.1 (ref. ⁸⁵) and Sequenza v.2.1.2 (ref. ⁸⁶), respectively. If a tumor contained more than 5,000 variants, input variants were subsampled to 5,000, ensuring that all variants overlapping previously identified, significantly mutated genes were included⁸⁴. For each PhyloWGS run, multiple Markov chain Monte Carlo chains were initiated with differing start values; the optimum solution was selected based on negative normalized log-likelihood. Cancer cell fractions were calculated for each tumor subpopulation as the cellular prevalence for a given subpopulation divided by the maximum cellular prevalence for that tumor, which corresponds to the estimated tumor purity. Events were defined as clonal if they had a cancer cell fraction of 1 or subclonal otherwise. SCNA subpopulation assignments and cellular prevalence estimates derived from PhyloWGS were further informed by scRNA-seq and scRRBS-derived copy number profiles.

Bulk RNA-seq processing. FASTQ files were preprocessed with fastp v.0.20.0 to assess quality control and were aligned to the hg19 genome using kallisto v.0.46.0 with default parameters⁸⁷. The bulk RNA Verhaak classification and simplicity scores were determined using the ssGSEA.GBM.classification v.1.0 R package⁸. Single-sample gene set enrichment analysis for particular pathways was performed using the GVSAs v.1.32.0 R package⁸⁸.

Detection of extrachromosomal DNA. Amplicon architect (the version used in original paper⁴⁹) was used to detect extrachromosomal DNA in tumor WGS data. Briefly, this method characterizes the architecture of amplified regions that are larger than 10 kb and have more than 4 copies greater than the median sample ploidy.

DNAm-based tumor classification. Probabilistic estimates of tumor classification were defined by the MolecularNeuropathology classification tool v.11b4 (ref. ⁶¹).

Statistics and reproducibility. All data analyses were conducted in R v.3.6.1. Statistical analyses are described in the respective Methods subsections and are briefly described in the figure legends. *P* values were false discovery rate-corrected for multiple hypotheses testing where indicated. For box plot representations, data points located outside the whisker plots are not shown to aid readability but are included in the statistical analyses. No statistical methods were used to predetermine study sample size. Data subsets are explicitly mentioned when used. The experiments were not randomized. The investigators were not blinded to allocation during the experiments and outcome assessment. *P* < 0.05 was considered statistically significant.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All de-identified, nonprotected access somatic variant calls, single-cell gene expression profiles, regional scDNAm data and scDNAm disorder data are accessible via Synapse (<https://synapse.org/singlecellglioma>). Raw bulk and single-cell sequencing data and methylation microarray data are available through the European Genome-phenome Archive under accession no. EGAS00001005300. The GRCh37 (hg19) reference genome was obtained from GATK (<https://gatk.broadinstitute.org/>).

Code availability

Major analysis scripts are available on GitHub (<https://github.com/TheJacksonLaboratory/singlecellglioma-verhaaklab>) and Zenodo (<https://doi.org/10.5281/zenodo.4967364>).

References

- Capper, D. et al. DNA methylation-based classification of central nervous system tumours. *Nature* **555**, 469–474 (2018).
- Bhat, K. P. L. et al. Mesenchymal differentiation mediated by NF- κ B promotes radiation resistance in glioblastoma. *Cancer Cell* **24**, 331–346 (2013).
- Stoeckius, M. et al. Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 224 (2018).
- Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
- Hui, T. et al. High-resolution single-cell DNA methylation measurements reveal epigenetically distinct hematopoietic stem cell subpopulations. *Stem Cell Rep.* **11**, 578–592 (2018).
- Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
- Forrest, A. R. R. et al. A promoter-level mammalian expression atlas. *Nature* **507**, 462–470 (2014).
- Hunt, S. E. et al. Ensembl variation resources. *Database (Oxford)* **2018**, bay119 (2018).
- Raney, B. J. et al. Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics* **30**, 1003–1005 (2014).
- Fornes, O. et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **48**, D87–D92 (2020).
- Lawrence, M. S. et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
- Garvin, T. et al. Interactive analysis and assessment of single-cell copy-number variations. *Nat. Methods* **12**, 1058–1060 (2015).
- Galili, T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **31**, 3718–3720 (2015).
- Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
- Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
- Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44 (2019).
- Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
- Polański, K. et al. BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* **36**, 964–965 (2020).
- Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
- Chakravarthy, A. et al. Pan-cancer deconvolution of tumour composition using DNA methylation. *Nat. Commun.* **9**, 3220 (2018).
- Sheffield, N. C. & Bock, C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics* **32**, 587–589 (2016).
- Köster, J. & Rahmann, S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **34**, 3600 (2018).

83. Blokzijl, F., Janssen, R., van Boxtel, R. & Cuppen, E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med.* **10**, 33 (2018).
84. Deshwar, A. G. et al. PhyloWGS: reconstructing subclonal composition and evolution from whole-genome sequencing of tumors. *Genome Biol.* **16**, 35 (2015).
85. Ha, G. et al. TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data. *Genome Res.* **24**, 1881–1893 (2014).
86. Favero, F. et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann. Oncol.* **26**, 64–70 (2015).
87. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
88. Hänzelmann, S., Castelo, R. & Guinney, J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
89. Deshpande, V. et al. Exploring the landscape of focal amplifications in cancer using AmpliconArchitect. *Nat. Commun.* **10**, 392 (2019).

Acknowledgements

We thank the patients and their families for their generous donation to biomedical research. We also thank the staff in the following groups at The Jackson Laboratory for Genomic Medicine: single-cell biology laboratory; flow cytometry core; and genomic technology core for assistance in data generation. We thank M. Wimsatt and Z. Reifsnnyder for assistance in graphic design. We thank the University of Texas MD Anderson Epigenomics Profiling Core for their assistance in helping troubleshoot the scRRBS protocol. We thank the Henry Ford Hospital for sharing the patient-derived glioma spheroids. This work was supported by National Institutes of Health grants R01 CA237208 and R21 NS114873, Cancer Center Support Grant P30 CA034196, Department of Defense grant no. W81XWH1910246 (to R.G.W.V.) and Jackson Laboratory Cancer Center Fast Forward funds. E.S.V. is supported by a postdoctoral fellowship from The Jane Coffin Childs Memorial Fund for Medical Research. F.P.B. is

supported by the National Cancer Institute (grant no. K99 CA226387). E.Y. is a fellow of the American Brain Tumor Association. K.C.J. is the recipient of an American Cancer Society Fellowship (no. 130984-PF-17-141-01-DMC). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

K.C.J. and R.G.W.V. conceived the project and designed the experiments. K.B. and S.D. curated the patient samples and patient annotation. N.V. and P.C.d.W.H. provided input on the multi-sector analyses. K.C.J., M.R.H.E., M.T., N.E.N., R.M., C.Y.N., M.L.S. and P.R. performed the single-cell library optimization and sequencing. K.C.J. led the data production and performed the experiments with A.D.G., D.Z., D.L., E.Y. and E.T.C. K.C.J. and K.J.A. led the data analysis in collaboration with F.P.B., E.S.V., M.S., E.Y. and H.K. K.C.J., K.J.A. and R.G.W.V. wrote the manuscript with input from all authors.

Competing interests

R.G.W.V. is a cofounder of and has received research support from Boundless Bio. The other authors declare no conflicts of interest.

Additional information

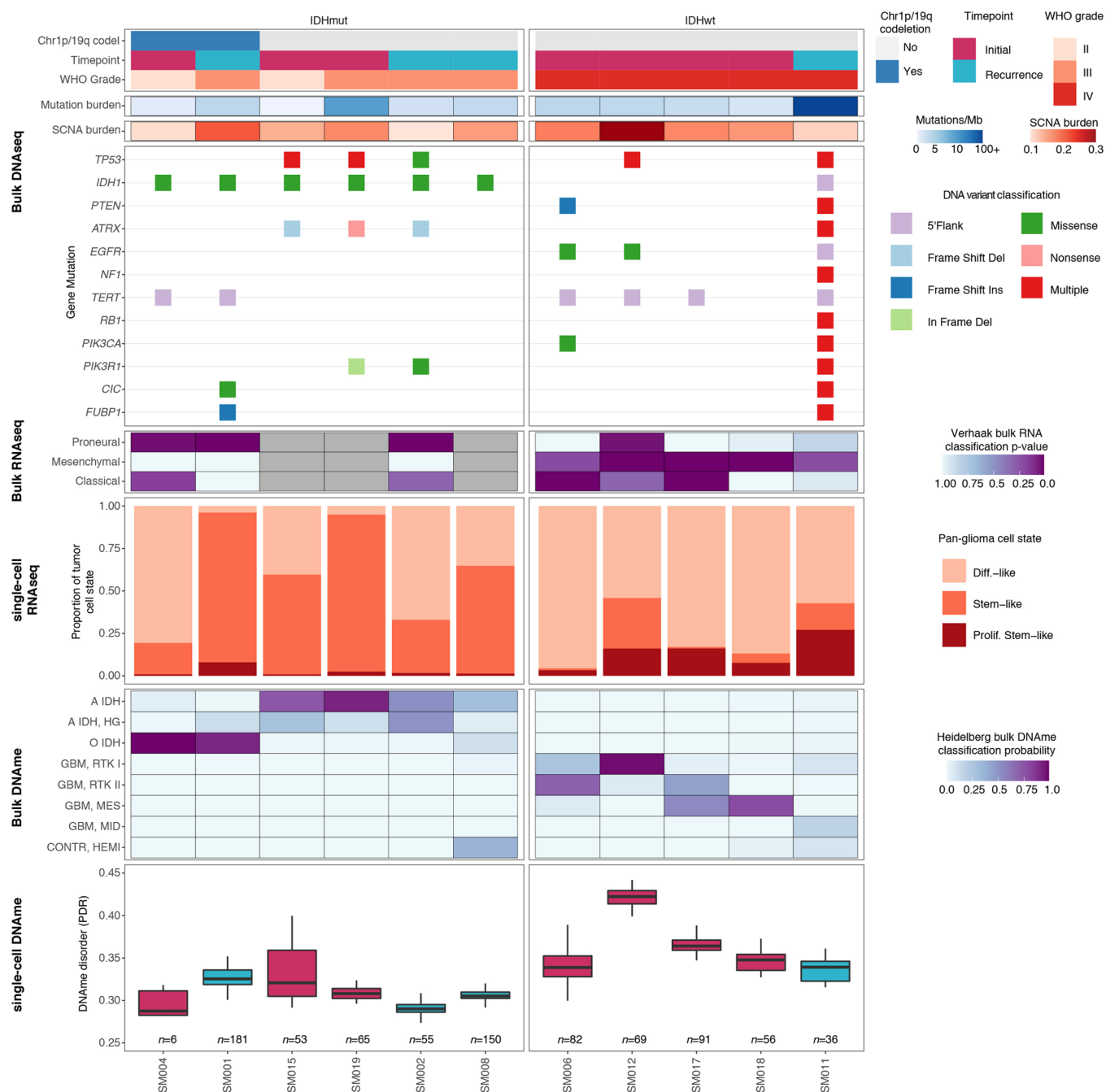
Extended data is available for this paper at <https://doi.org/10.1038/s41588-021-00926-8>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41588-021-00926-8>.

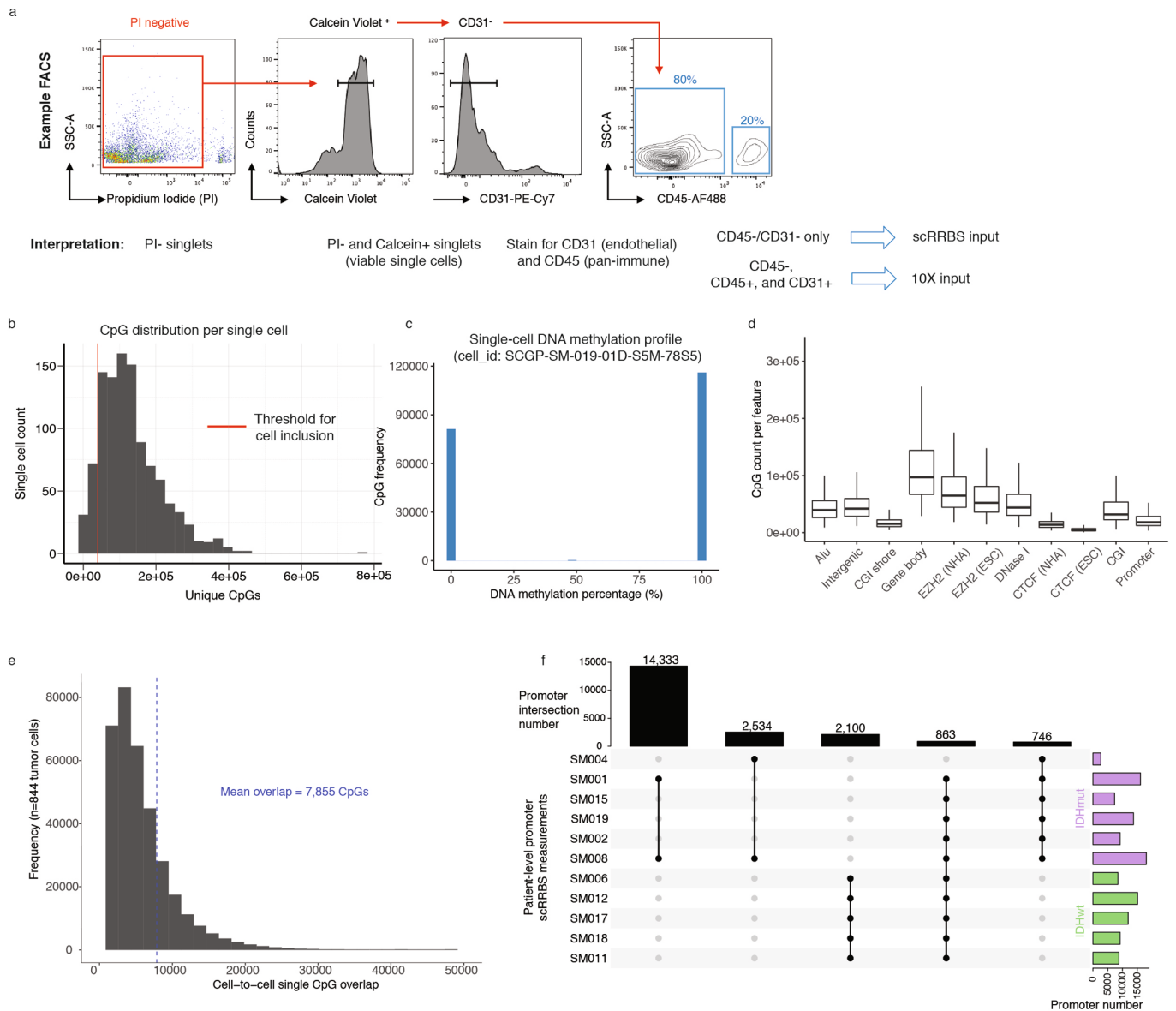
Correspondence and requests for materials should be addressed to Kevin C. Johnson or Roel G. W. Verhaak.

Peer review information *Nature Genetics* thanks David Brocks and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

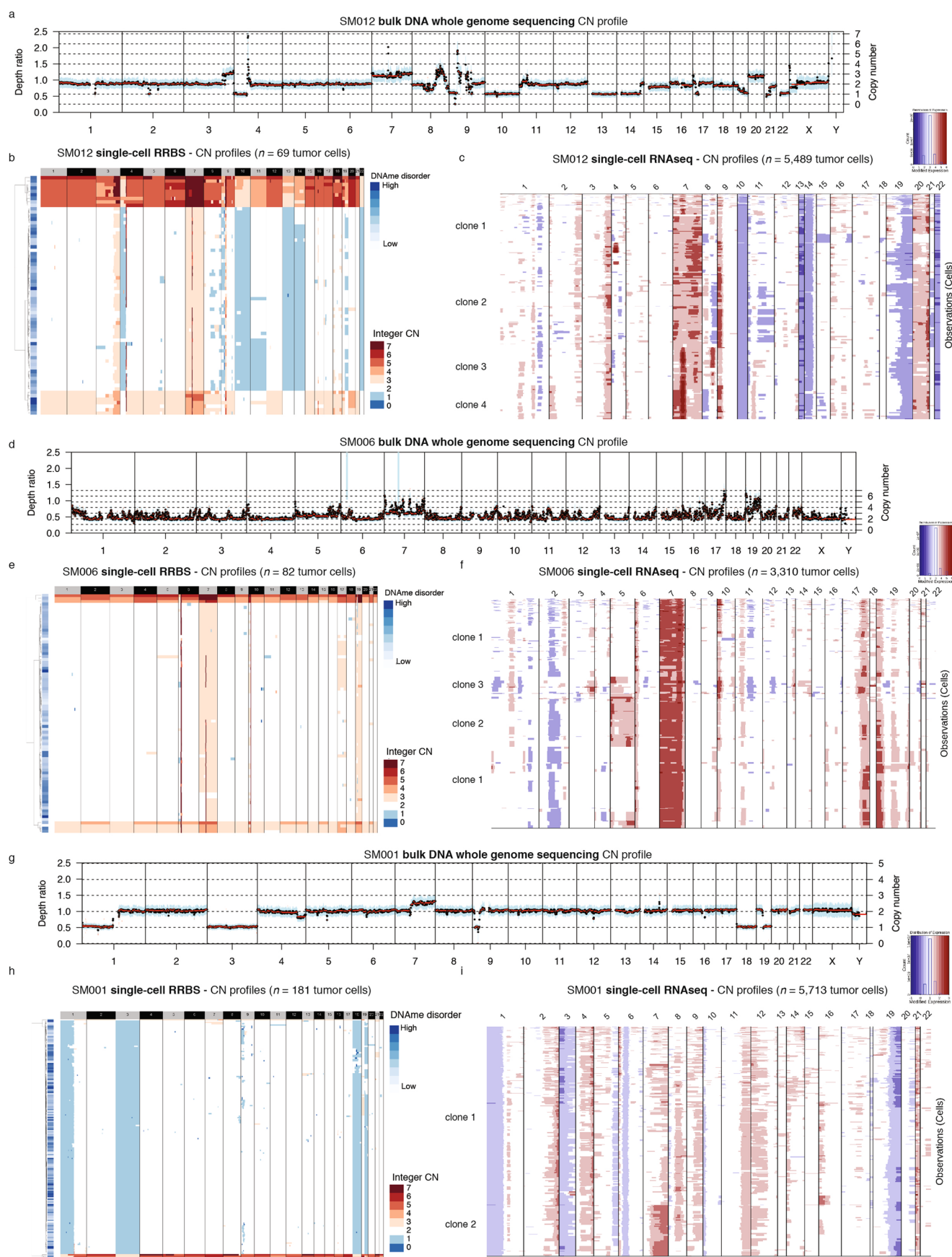
Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | Integrated molecular profiles of patient samples. Each patient is in a single column with data presented to indicate clinical features (top), followed by genetic alterations defined from bulk whole genome sequencing data, bulk RNA sequencing based subtype classification probabilities (Wang et al., $n = 8$ available), single-cell RNA tumor cellular state proportions, bulk DNAm microarray subtype classification probabilities (Capper et al.), and boxplots of single-cell DNAm disorder with samples colored by clinical timepoint. Each box spans the 25th and 75th percentile, center lines indicate the median, and the whiskers represent the absolute range (minima/maxima), excluding outliers.

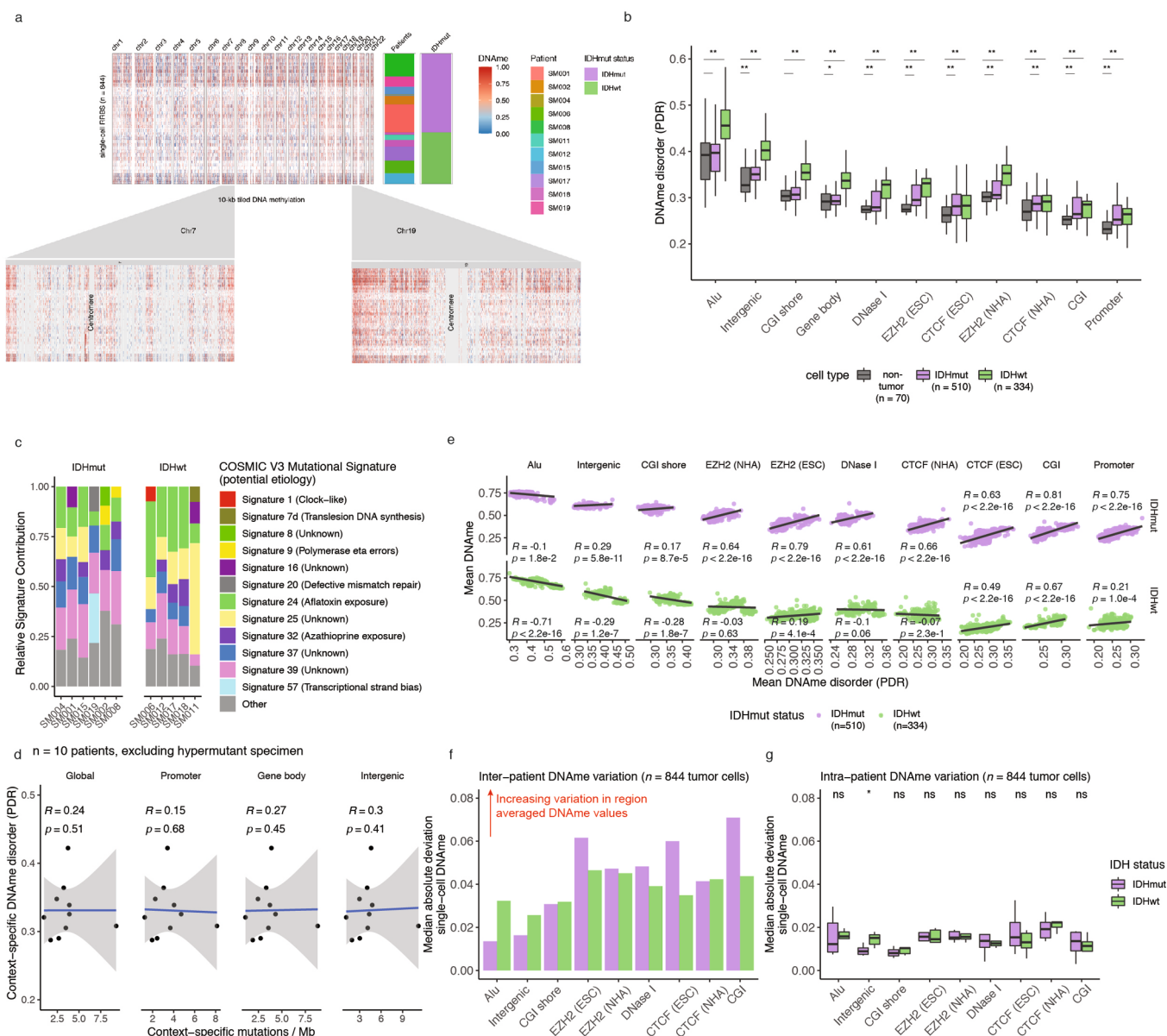


Extended Data Fig. 2 | Sample pre-processing and metrics related to single-cell DNAm data assessment. **a**, Representative fluorescence activated cell sorting (FACS) data and strategy for viable cell enrichment for both single-cell protocols, and tumor cell enrichment in scRRBS. **b**, The number of unique CpGs detected per single cell, with the red line indicating the threshold (minimum 40,000 unique CpGs) for inclusion in the dataset presented herein. **c**, Representative distribution of single locus DNAm estimates for a single cell. DNAm percentage of 0 represents an unmethylated locus, while a percentage of 100 represents a methylated locus. **d**, The CpG count per genomic features across tumor single cells. **e**, Histogram representing the cell-to-cell CpG overlap of all single cells in this dataset. **f**, Upset plot of patient-level unique gene promoter overlap. The top bar plot represents the promoter intersection number measured across all patients (center portion) indicated by a filled bullet point. The right histogram represents the total number of unique promoters measured across all cells from a given tumor with IDH mutation status indicated by color.

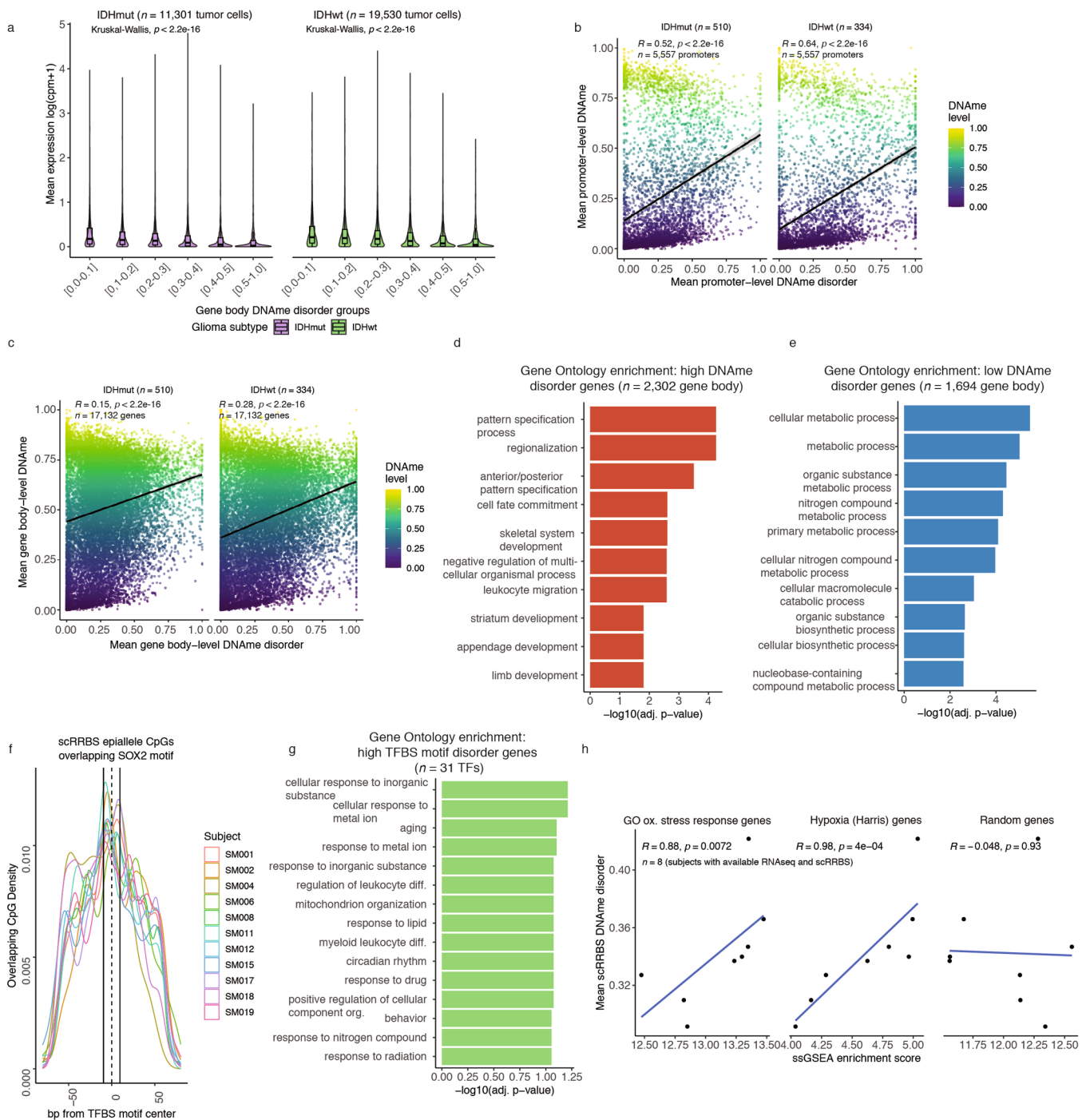


Extended Data Fig. 3 | See next page for caption.

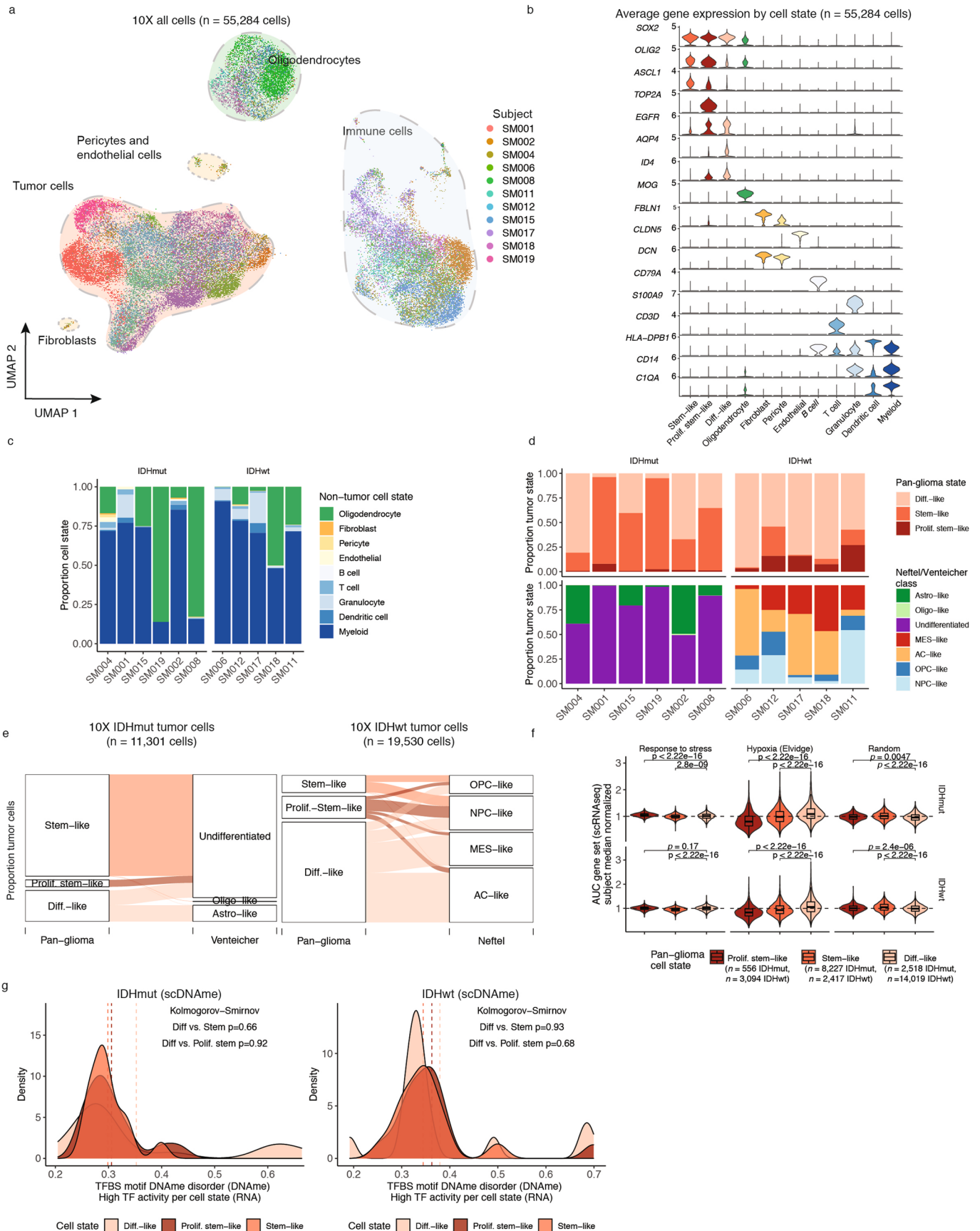
Extended Data Fig. 3 | Somatic copy number alteration examples estimated from whole genome sequencing, single-cell Reduced Representation Bisulfite Sequencing, and single-cell RNA-sequencing. a-c. Representative images of copy number alterations derived from SM012 (IDHwt initial) whole genome sequencing (WGS) data. **a**, Depth ratio for each segment with copy number status determined as compared with germline (normal blood) WGS data. **b**, SM012 Single-cell DNAm-based copy number estimates ($n = 69$ tumor cells) with copy number integer depicted by color (blue = CN loss, white = neutral CN, and red = CN gain). Each row is a single cell with annotation for DNAm disorder provided. **c**, SM012 Single-cell RNAseq based copy number inference ($n = 5,489$) identifying major copy number events found in WGS with labelled subclones as presented in Fig. 6a. **d-f**, Similar example profiles as presented in **a-c**, for tumor sample SM006 (IDHwt initial, $n = 82$ scRRBS cells, $n = 3,310$ scRNAseq cells). **g-i**, Similar example profiles as presented in **a-c** for tumor sample SM001 (IDHmut recurrence, $n = 181$ scRRBS cells, $n = 5,713$ scRNAseq cells).



Extended Data Fig. 4 | Distribution and relationship of DNAm and DNAm disorder throughout the glioma genome. a, Visualization of inter-tumoral and intra-tumoral variation in DNAm (10 kb tiled DNAm). Genome-level and chromosome-level DNAm across 844 single tumor cells. Each row represents a single cell clustered based on pairwise dissimilarity between methylomes as presented in main Fig. 1b and each column represents a single 10 kb tile over which DNAm has been averaged as indicated by heatmap color (methylated = red, unmethylated = blue). The tile color for a cell that does not have a measurement for a given tile is represented by white and a tile without a measurement across any cells is represented by grey. Row annotation both patient identifier and IDH mutation status are presented for each cell. **b**, Boxplots highlighting the single-cell DNAm disorder estimates calculated across different genomic contexts with Kruskal-Wallis p-values indicating the differences in distributions across the groups. Each box spans the 25th and 75th percentile, center lines indicate the median, and the whiskers represent the absolute range (minima/maxima), excluding outliers. **c**, The dominant Catalogue of Somatic Mutations in Cancer (COSMIC v3) mutational signatures are presented for each subject. The stacked bar plots represent the relative contribution of each mutational signature to the tumor’s mutational burden. Colors indicate distinct mutational signatures, which are further annotated with their proposed etiology. **d**, Scatterplots and linear regression lines with standard error showing the relationship between genomic context-specific single-cell DNAm disorder (sample-specific scRRBS average) and genomic context-specific mutation burden derived from whole genome sequencing (n=10 excluding hypermutant sample). Panels are separated into global (that is, all regions), promoter, gene body, and intergenic regions (Spearman correlations $\rho > 0.05$ for all comparisons). **e**, Scatterplot of the context-specific DNAm disorder (x-axis) vs. the average DNAm value (beta-value) for each genomic compartment. Subtype level Spearman correlation coefficients and p-values are presented. **f**, The median absolute deviation of DNAm across all cells from the same subtype (inter-patient heterogeneity) and **g**, all cells from the same patient (intra-patient heterogeneity). Two-sided Wilcoxon rank sum tests comparing median absolute deviation levels between IDHmut and IDHwt are presented for intra-patient DNAm heterogeneity.

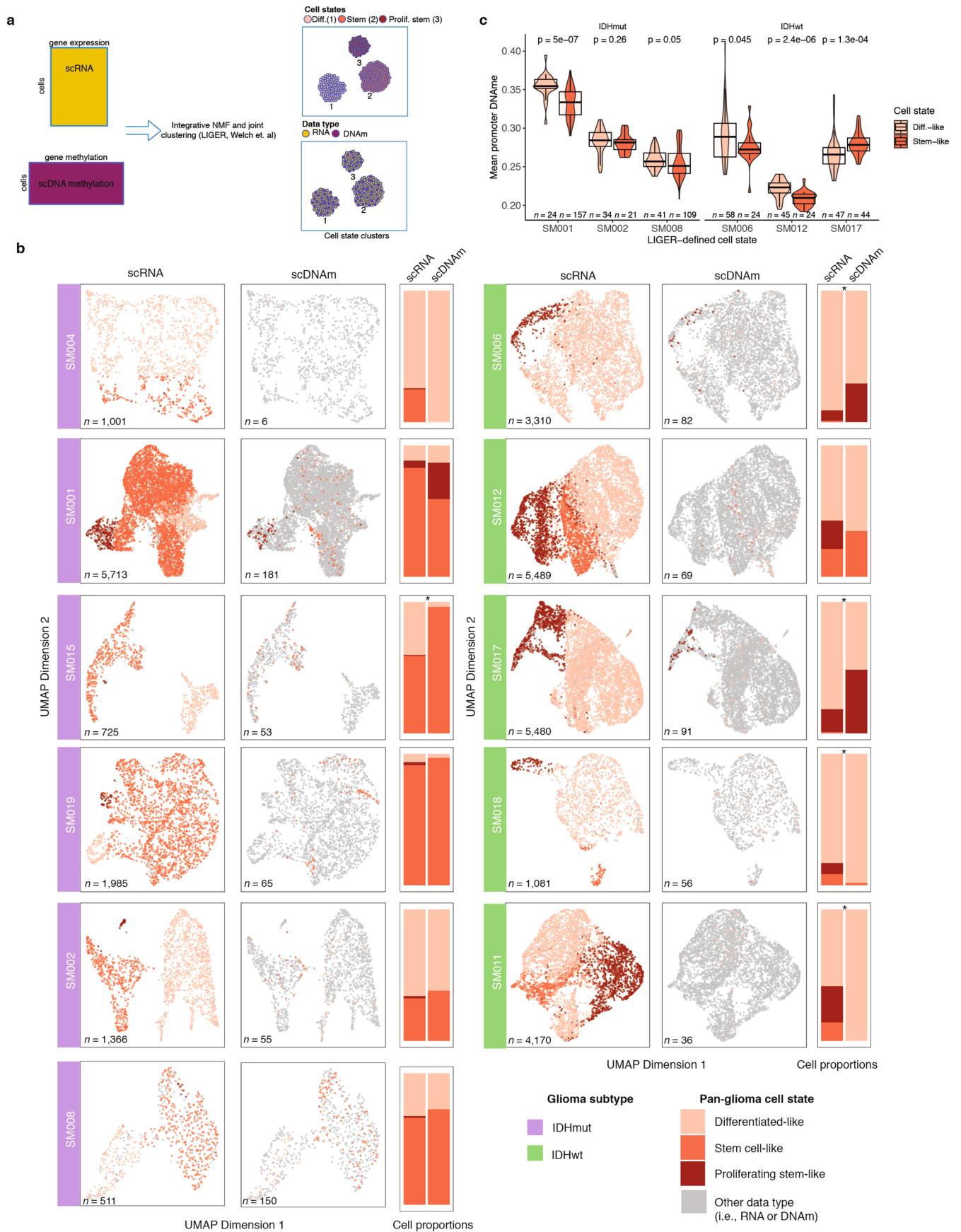


Extended Data Fig. 5 | Association between DNAmE disorder and disrupted transcriptional programs. a, Boxplots of gene expression values, in $\log(\text{counts per million}+1)$, from single-cell RNAseq data across different sets of gene body regions defined by gene-derived DNAmE disorder groups. Each box spans the 25th and 75th percentile, center lines indicate the median, and the whiskers represent the absolute range (minima/maxima), excluding outliers. Surrounding violins represent the distribution for each group. Gene DNAmE disorder groups are defined by the determining the mean DNAmE disorder value across a single gene. Color indicates *IDH1* mutation status. **b-c**, Scatterplots depicting single-cell gene-level DNAmE disorder average plotted against the gene-level methylation estimates in both **b**, promoter regions and **c**, gene body regions. **d-e**, Gene Ontology enrichment analyses with false discovery rate correction for high DNAmE genes and low DNAmE disorder genes using gene body estimates. **f**, Representative density curves of distribution of epiallele CpGs in each patient overlapping a specific TFBS motif (for example, SOX2), with curves annotated by patient identifier. **g**, Gene Ontology enrichment analysis of TFs with high DNAmE disorder in their binding sites with false discovery rate correction. **h**, Scatterplot depicting the association between average single-cell DNAmE disorder estimate and single-sample Gene Set Enrichment Score for stress response, hypoxia, and random genes from bulk RNAseq data. Spearman correlation coefficient and p -values are indicated.



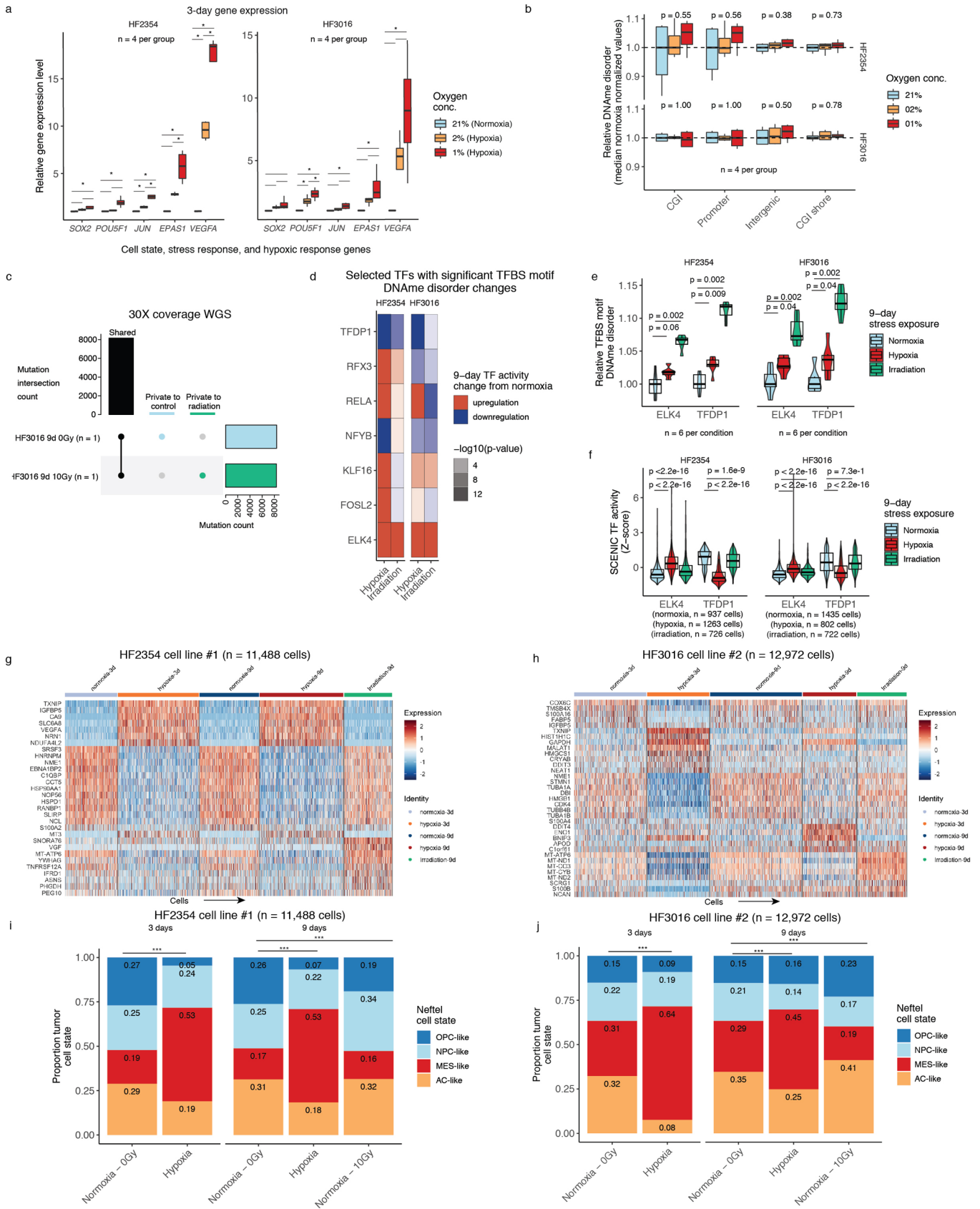
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Pan-glioma cell state assignment and characteristics. **a**, UMAP dimensionality reduction plot of all scRNAseq data, including tumor and non-tumor cells ($n = 55,248$ cells). Each dot depicts a single cell and colors represent the tumor of origin. Shaded regions represent cell state classification. **b**, Stacked violin plots of average single-cell gene expression for cells presented in Supplementary Fig. 6a. Selected genes presented are informative for cell state classification. **c**, Stacked bar plots representing the proportion of non-tumor cellular states **d**, Stacked bar plots representing the proportion of tumor cellular states per tumor for pan-glioma classification (top row) and previously published classifications (lower left row; Venteicher et al. and lower right Neftel et al.) **e**, Sankey plot representing the proportion of IDHmut tumor cells with pan-glioma classification and associated classification described in Venteicher et al. (left). Sankey plot representing the proportion of IDHwt tumor cells with pan-glioma classification and associated classification described in Neftel et al.(right). **f**, scRNAseq area under the curve estimates for selected gene sets (that is, proportion of expressed genes in signature per cell). The AUC estimates are presented for response to stress, hypoxia, and random gene set signatures summarized by pan-glioma cell state and separated by IDH mutation subtype. All cells from a single patient are normalized to its median AUC value for a given signature. Higher relative values indicate greater enrichment score for each signature. P -values represent two-sided Wilcoxon rank sum tests comparing differentiated-like tumor cells with stem-like and proliferating stem-like. **g**, Density plots representing TFBS motif DNAm disorder (scRRBS data) in IDHmut (left) and IDHwt (right) single-cell DNAm data for TFs whose activity (scRNAseq based SCENIC analysis in Fig. 6c,d) characterizes a specific cell state ($n = 15$ TFs per cell state). Kolmogorov-Smirnov p -value tests for differences in TFBS DNAm disorder across the cellular states. Dotted lines represent the median TFBS motif DNAm disorder value for cell state defining TFs.



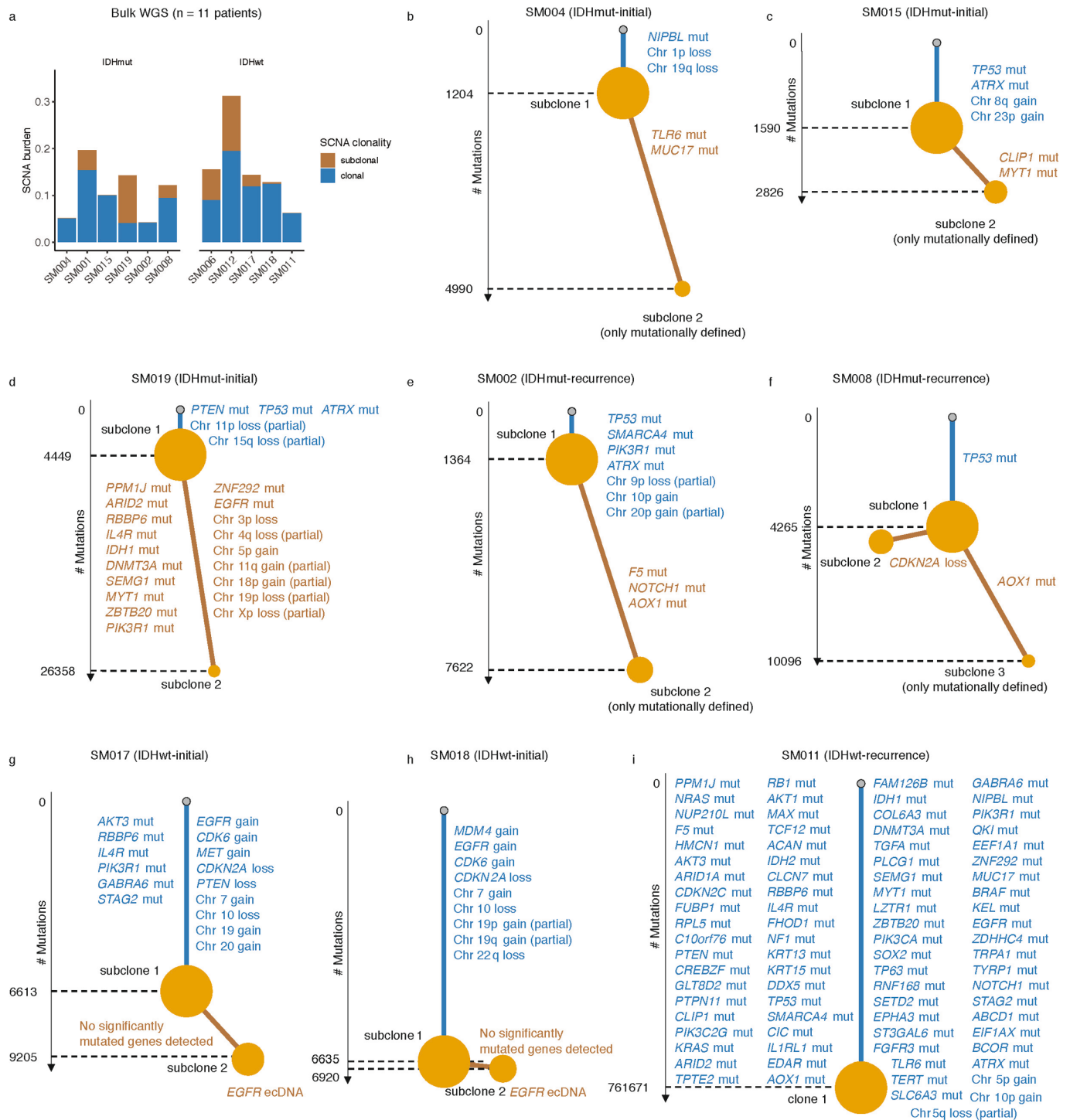
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | LIGER integrated tumor-specific clustering of single-cell RNA and single-cell DNAm data. **a**, Schematic diagram representing LIGER workflow to jointly cluster single-cell RNAseq and DNAm data generated from the same tumor dissociation. **b**, Joint single-cell RNAseq (scRNA) and single-cell DNAm (scDNAm) clustering and UMAP projections highlighting similar cellular state distributions across platforms. Sample annotation is presented on the left of each paired UMAP plot, each dot is an individual single cell, and cell number for each technology is presented in the lower-left hand corner. UMAP coordinate space remains the same for both scRNA and scDNAm visualizations with cellular states for that platform represented by a colored dot and data for the other platform represented by a gray dot. Stacked bar plots enumerating the proportion of cellular states detected by each platform are presented to the right of each paired UMAP plot. “*” indicate specimens in which the cellular proportions across the two platforms are significantly different (two-sided Fisher’s Exact test, $p < 0.05$). **c**, Promoter DNAm for samples with sufficient number of cells in each state. Each box spans the 25th and 75th percentile, center lines indicate the median, and the whiskers represent the absolute range (minima/maxima), excluding outliers. Surrounding violins represent the distribution for each condition. Two-sided Wilcoxon rank sum test p-values are presented for each tumor.

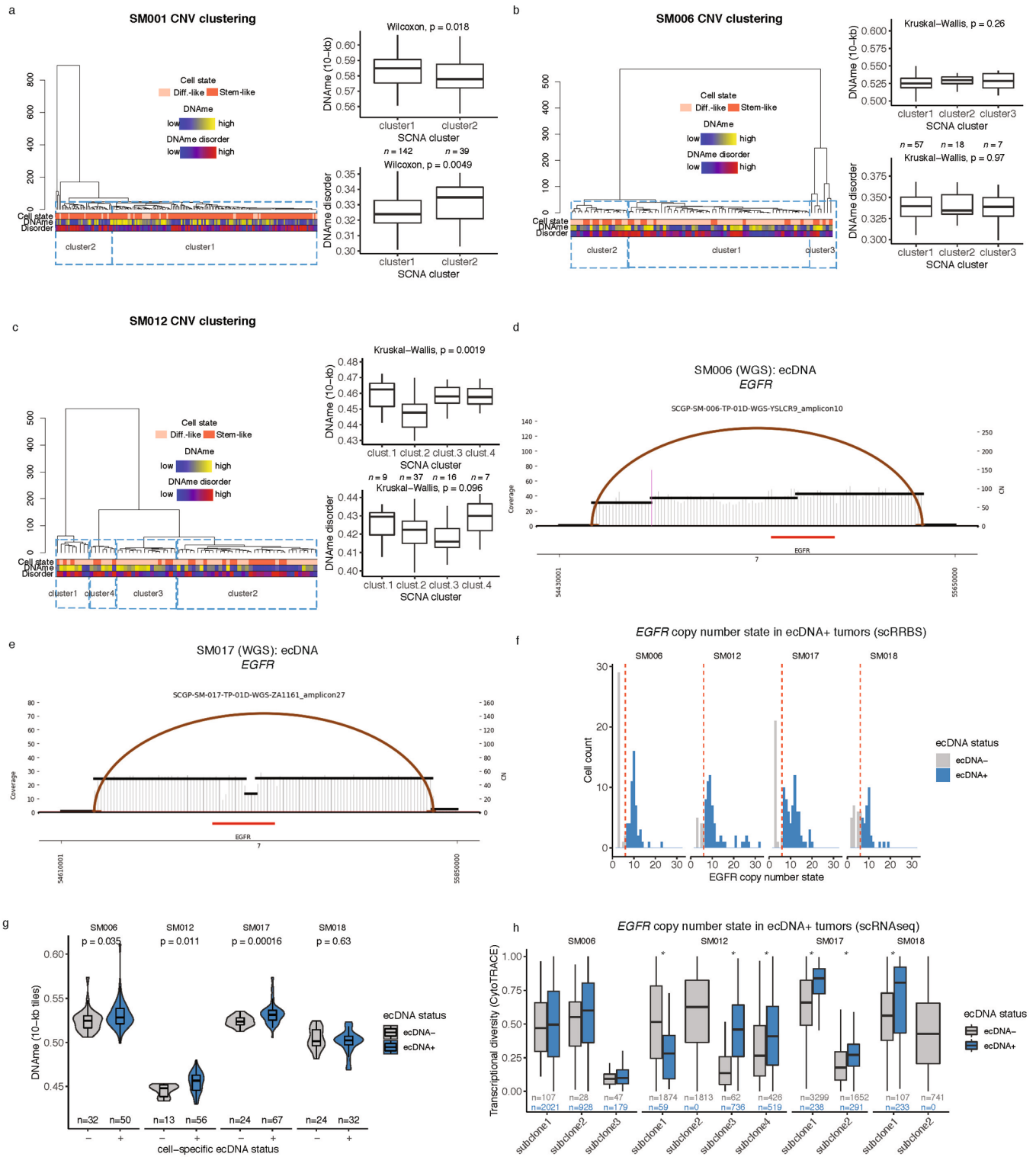


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Stress-associated changes in DNAm disorder are associated with altered population-level transcriptional dynamics and not related with genetic changes. **a**, Relative gene expression levels for two patient-derived glioma sphere-forming cells for candidate gene cell state (*SOX2*, *POU5F1*) and cell stress (*JUN*, *EPAS1* (*HIF2A*), *VEGFA*) via RT-PCR. Normoxia and varying levels of hypoxia (2% and 1% oxygen, $n = 4$ per group) were assessed. Statistical significance ($p < 0.05$, Tukey HSD) is indicated by an asterisk. **b**, Relative DNAm disorder in hypoxia conditions (2% and 1%) compared with normoxia. P -values for Kruskal-Wallis tests are presented across specific genomic contexts ($n = 4$ per group). **c**, Upset plot of shared mutations for a randomly selected replicate from cell line HF3016 cultured under normoxia and irradiation (10 Gy). Mutations were determined in reference to patient normal blood. The mutational overlap is presented by the black bar with the mutations called private to irradiation and control also presented. **d**, Heatmap representing transcription factors that were determined to have consistently different TFBS motif DNAm disorder levels in stress conditions (hypoxia and irradiation) compared with controls across both cell lines ($p < 0.1$ two-sided Wilcoxon rank sum test across all cell lines and two stressors) are presented with their change in inferred TF activity (SCENIC, methods). **e-f**, ELK4 and TFDP1 are presented for TFBS motif DNAm disorder (RRBS) and TF activity (scRNAseq), which demonstrated consistent changes in TFBS motif DNAm disorder and stress altered TF activity. Two-sided Wilcoxon rank sum test p -values are presented. **g-h**, scRNAseq scaled gene expression heatmaps for the top 5 differentially expressed genes per stress exposure and time point. **i-j**, Stacked bar plots comparing the cell state proportions for the Neftel et al. proliferation-independent IDHwt classifier across different stress conditions, time points, and cell lines. Statistical differences are presented for Chi-Square test (***) = $p < 0.001$). Oligodendrocyte progenitor cell-like (OPC-like), Neural progenitor cell-like (NPC-like), Mesenchymal-like (MES-like), and Astrocyte-like (AC-like) cell states are presented.



Extended Data Fig. 9 | Whole genome sequencing phylogenetic inference of tumor samples. **a**, Stacked bar plots representing the proportion of whole-genome sequencing (WGS) derived somatic copy number alteration (SCNA) burden attributed to clonal vs. subclonal events. **b-i**, Phylogenetic trees constructed from whole genome sequencing data (mutations and somatic copy number alterations) using phyloWGS and further annotated using single-cell inferred copy number alterations (scRRBS + scRNAseq). Tree nodes represent alterations specific to the given clone, with node size corresponding to the fraction of cells with the associated alterations. Branch length scales with the number of mutations attributed to that clone. Clonal alterations are colored in blue, with subclonal alterations colored in gold. Genes considered significantly mutated in TCGA analyses² and chromosomal arm-level events are presented. Arm-level events are defined as spanning at least 80 percent of the chromosome arm, while partial events span at least 40 percent.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Genetic influences on epigenetic and transcriptional diversity in glioma cells. **a-c**, SCNA phylogenetic trees annotated with scRRBS-derived cell state. Adjacent boxplots are presented for DNAm and DNAm disorder across cuts in the dendrograms. **d-e**, Extrachromosomal DNA circular amplicon reconstruction displaying genomic rearrangements predicted from whole genome sequencing. Coverage depth is represented as a histogram across a genomic interval with segment copy number (CN) estimation provided on the right y-axis. Discordant read pair clusters are indicated by arcs and colors highlight read pair orientation (for example, brown = everted read pairs⁶¹). Amplicon intervals are provided at the bottom of the plot with annotation for known oncogenes (for example, *EGFR*). **f**, *EGFR* copy number estimation from single-cell RRBS data in ecDNA+ tumors. Cells with *EGFR* copy number greater than 6 were classified as *EGFR* ecDNA+ (blue). **g**, Single-cell 10-kb tiled DNAm separated by *EGFR* ecDNA status. Single cells with inferred copy number status greater than 6 were classified as ecDNA+ (blue). Two-sided Wilcoxon rank sum test *p*-values comparing DNAm across ecDNA status are reported for each patient tumor. **h**, Boxplots depicting transcriptional diversity using gene count signatures calculated in scRNAseq data for each tumor, with cells separated based on inferred *EGFR* copy number status (gray = *EGFR* ecDNA-, blue = *EGFR* ecDNA+). Transcriptional diversity was compared based on predicted ecDNA status within each tumor subclone. Stars (*) indicate statistically significant differences based on two-sided Wilcoxon rank sum test ($p < 0.05$). Each box plot in this figure spans the 25th and 75th percentile, center lines indicate the median, and the whiskers represent the absolute range (minima/maxima), excluding outliers. Surrounding violins represent the distribution for each condition.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis https://github.com/TheJacksonLaboratory/singlecellglioma-verhaaklab."/>

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All deidentified, non-protected access somatic variant calls, single-cell gene expression profiles, regional single-cell DNA methylation data, and single-cell DNAm disorder data are accessible via Synapse (<https://synapse.org/singlecellglioma>) and Zenodo (<https://doi.org/10.5281/zenodo.4967364>). Bulk and single-cell sequencing data and methylation microarray data are available through the European Genome-phenome Archive (EGA) under the accession number

EGAS00001005300. Custom codes developed in the study are available at <https://github.com/TheJacksonLaboratory/singlecellglioma-verhaaklab>. The GRCh37 (hg19) reference genome was obtained from GATK (<https://gatk.broadinstitute.org/>). The JASPAR 2020 database was obtained from <http://jaspar.genereg.net/download/CORE/>. NHA (human astrocyte) and H1-ESC (human embryonic stem cell) cell line EZH2 and CTCF regions were retrieved from ENCODE (<https://www.encodeproject.org/files/ENCF001SUU/> (H1-ESC, EZH2), <https://www.encodeproject.org/files/ENCF001UBA/> (H1-ESC, CTCF), <https://www.encodeproject.org/files/ENCF001TAU/> (NHA, EZH2), <https://www.encodeproject.org/files/ENCF001TAT/> (NHA, CTCF)).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size. Sample size was based on the tissue sample/cell numbers typically presented in contemporary single cell genomic publications needed to detect statistically significant differences between conditions, experimental feasibility/cost, and tissue availability.
Data exclusions	No samples were excluded from analysis.
Replication	Across two separate cell lines and 60 total biological replicates we observed consistent experimental results for all replicates exposed to different environmental stressors. We confirmed that each genomic profiling modality was assigned to the correct tumor by analyzing common copy number alterations, which were inferred across multiple sequencing platforms and confirmed all samples were appropriately matched across datasets.
Randomization	There was no randomization in this study. Subject randomization was not performed because there were no therapeutic interventions. The stress experiments were performed across the same number of biological replicates in two separate cell lines.
Blinding	All patient samples were deidentified and were assigned a study-specific barcode. Blinding was not relevant to our study since there was no randomization of groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	Antibodies were used to select for non-immune and non-endothelial cells during the FACS sort for scRRBS experiments. Alexa Fluor 488 conjugated anti-CD45 antibody (Cat. no. 304017, BioLegend) and PECy7-conjugated anti-CD31 antibody (Cat. no. 303117, BioLegend).
Validation	Antibodies for their use in FACS with human tissue samples was confirmed according to the manufacturer's website. Alexa Fluor 488 conjugated anti-CD45 antibody (Cat. no. 304017, BioLegend) - https://www.biolegend.com/en-us/products/alexa-fluor-488-anti-human-cd45-antibody-2738?GroupID=BLG5926 PECy7-conjugated anti-CD31 antibody (Cat. no. 303117, BioLegend) - https://www.biolegend.com/en-us/products/pe-cyanine7-anti-human-cd31-antibody-6124?GroupID=BLG5721

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	The dataset includes 11 patients with data presented regarding sex, tumor grade, and treatment. This information is presented in Supplementary Table 1.
Recruitment	Adult patients diagnosed with an initial or recurrent diffuse glioma were eligible for recruitment. The material was donated for general research purposes. Informed consent was obtained from all study subjects as part of the Institutional Review Board (IRB) protocol.
Ethics oversight	The study was approved by the local ethics committee and by the IRB (The Jackson Laboratory, The University of Connecticut Health Center, and St. Michael's Hospital) and conducted in accordance to the Declaration of Helsinki protocol.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Single cell suspensions were blocked with human BD Fc Block (BioLegend) for 5 min on ice, prior to antibody staining, and labelled via incubation with 1:100 dilution of Alexa Fluor 488 conjugated anti-CD45 antibody (Cat. no. 304017, BioLegend) and 1:100 dilution of PE-Cy7-conjugated anti-CD31 antibody (Cat. no. 303117, BioLegend) for 30 minutes at 4C. Cells were washed with Hank's buffered saline solution and resuspended in 2mM EDTA/ 2% BSA/ PBS buffer containing [2µg/mL] propidium iodide (PI) (BD Biosciences, Cat. No. 556364) and [1µM] Calcein violet (Invitrogen) for 20 minutes at 4 degrees C.
Instrument	Fluorescence activated cell sorting (FACS) was performed using a BD FACSAria Fusion instrument with an 130µm nozzle and using the lowest event rate.
Software	FlowJo (10.3).
Cell population abundance	FACS was used to sort viable cells for 10X single-cell RNA sequencing and to sort non-immune and non-endothelial cells in the scRRBS experiments.
Gating strategy	Single cell mode was selected to further ensure stringency of sorting.

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.