

1 **TITLE: Single-cell multimodal glioma analyses reveal epigenetic regulators of**  
2 **cellular plasticity and environmental stress response**

3  
4  
5 **AUTHORS:**

6 Kevin C. Johnson<sup>1,12\*</sup>, Kevin J. Anderson<sup>1,12</sup>, Elise T. Courtois<sup>1</sup>, Floris P. Barthel<sup>1,2</sup>, Frederick S.  
7 Varn<sup>1</sup>, Diane Luo<sup>1</sup>, Martine Seignon<sup>1</sup>, Eunhee Yi<sup>1</sup>, Hoon Kim<sup>1</sup>, Marcos RH Estecio<sup>3</sup>, Ming Tang<sup>4</sup>,  
8 Nicholas E. Navin<sup>5</sup>, Rahul Maurya<sup>1</sup>, Chew Yee Ngan<sup>1</sup>, Niels Verburg<sup>6</sup>, Philip C De Witt Hamer<sup>6</sup>,  
9 Ketan Bulsara<sup>7</sup>, Michael L. Samuels<sup>1</sup>, Sunit Das<sup>8,9,10</sup>, Paul Robson<sup>1,11</sup>, Roel GW Verhaak<sup>1\*</sup>

10  
11 **AUTHOR AFFILIATIONS:**

12 <sup>1</sup>The Jackson Laboratory for Genomic Medicine, Farmington, CT, 06032, USA.

13 <sup>2</sup>Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Pathology, Brain Tumor Center  
14 Amsterdam, Amsterdam UMC, Vrije Universiteit, Amsterdam, The Netherlands

15 <sup>3</sup>Department of Epigenetics and Molecular Carcinogenesis, The University of Texas MD  
16 Anderson Cancer Center, Houston, TX, US

17 <sup>4</sup>Department of Data Science, Dana-Farber Cancer Institute, Cambridge, MA, US

18 <sup>5</sup>Department of Genetics, The University of Texas MD Anderson Cancer Center, Houston, TX,  
19 US

20 <sup>6</sup>Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Neurosurgery, Brain Tumor  
21 Center Amsterdam, de Boelelaan 1117, Amsterdam, The Netherlands

22 <sup>7</sup>Division of Neurosurgery, The University of Connecticut Health Center, Farmington, CT, US

23 <sup>8</sup>Arthur and Sonia Labatt Brain Tumour Research Centre, Hospital for SickKids, University of  
24 Toronto.

25 <sup>9</sup>Institute of Medical Science, University of Toronto.

26 <sup>10</sup>Division of Neurosurgery, Li Ka Shing Knowledge Institute, St. Michael's Hospital, University of  
27 Toronto.

28 <sup>11</sup>Genetics and Genome Sciences, University of Connecticut School of Medicine

29 <sup>12</sup>Co-first author

30 \* Co-corresponding authors: [kevin.c.johnson@jax.org](mailto:kevin.c.johnson@jax.org) and [roel.verhaak@jax.org](mailto:roel.verhaak@jax.org).

31 Lead Contact: [roel.verhaak@jax.org](mailto:roel.verhaak@jax.org)

32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

50 **ABSTRACT:**

51 Glioma intratumoral heterogeneity enables adaptation to challenging microenvironments and  
52 contributes to universal therapeutic resistance. Here, we integrated 914 single-cell DNA  
53 methylomes, 55,284 single-cell transcriptomes, and bulk multi-omic profiles across 11 adult  
54 IDH-mutant or IDH-wild-type gliomas to delineate sources of intratumoral heterogeneity. We  
55 found that local DNA methylation instability, or epimutation burden, was elevated in more  
56 aggressive tumors, reflected intratumoral variability, linked with transcriptional disruption, and  
57 associated with environmental stress response. We show that the activation of cell-state specific  
58 transcription factors is impacted by epimutations and that loosened epigenetic control may  
59 facilitate cellular plasticity. Our analyses support that somatic copy number alterations (SCNAs)  
60 promote epigenetic instability and that SCNAs largely precede epigenetic and transcriptomic  
61 diversification during glioma evolution. We confirmed the link between genetic and epigenetic  
62 instability by analyzing larger cohorts of bulk longitudinally collected and spatially separated  
63 DNA methylation data. Increased DNA methylation instability was associated with accelerated  
64 disease progression, and recurrently selected DNA methylation changes were enriched for  
65 environmental stress response pathways. Our work provides an integrative framework to better  
66 understand glioma evolution and highlights the importance of epigenetic heterogeneity in  
67 shaping therapeutic response.

68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83

## 84 INTRODUCTION:

85 Diffuse gliomas are the most common malignant brain tumors in adults and remain  
86 incurable. Extensive molecular characterization of glioma has defined genomic drivers and  
87 clinically relevant subtypes based on the presence of *IDH1/2* gene mutations (i.e., IDH-mutant  
88 and IDH-wild-type) (Cancer Genome Atlas Research et al., 2015; Ceccarelli et al., 2016; Louis  
89 et al., 2016). Inter- and intra-tumoral heterogeneity are salient features across glioma subtypes  
90 that contribute to the universal therapeutic resistance. The heterogeneity observed in surgical  
91 resection specimens reflects each tumor's evolutionary path that is driven by competition  
92 between subpopulations harboring diverse genetic, epigenetic, and transcriptional aberrations  
93 (Barthel et al., 2019; Klughammer et al., 2018; Korber et al., 2019; Mazor et al., 2015; Wang et  
94 al., 2017). Thus, understanding how these different layers of heterogeneity integrate to define  
95 clonal lineages and drive glioma evolution may provide insights into treatment failure.

96 The study of tumor heterogeneity is complicated by cellular plasticity that enables cancer  
97 cells to reversibly transition between distinct cellular states in response to genetic,  
98 microenvironmental, and therapeutic stimuli (Flavahan et al., 2017). Single-cell RNA sequencing  
99 studies have previously identified such dynamic cellular states in IDH-wild-type gliomas  
100 (Bhaduri et al., 2020; Nefitel et al., 2019; Wang et al., 2019; Yuan et al., 2018). Cell states of  
101 IDH-mutant gliomas were found to display a more restricted plasticity along a hierarchical  
102 differentiation axis (Tirosh et al., 2016; Venteicher et al., 2017). Epigenetic modifications, such  
103 as DNA methylation at cytosine followed by guanine dinucleotides (i.e., CpGs), are mitotically  
104 heritable marks and regulate cellular states (Easwaran et al., 2014). For example, the transition  
105 from a differentiated-like state to an undifferentiated, or stem-like, state following chemotherapy  
106 in glioma was accompanied by epigenetic reprogramming (Liau et al., 2017). However, the  
107 epigenetic mechanisms that enable cellular plasticity and regulate glioma cell states remain  
108 poorly understood.

109 Epimutation is aberrant DNA methylation resulting from errors in the placement or  
110 removal of epigenetic marks. These stochastic errors in DNA methylation replication can  
111 accumulate in cancer cells as passenger events or be evolutionarily selected by destabilizing  
112 gene expression programs. Accordingly, epimutations provide genetically identical tumor cells  
113 with greater plasticity to respond to environmental stressors (Flavahan et al., 2017). Previous  
114 studies of glioma have demonstrated associations between bulk tumor epigenetic heterogeneity  
115 metrics and clinical outcomes (Ceccarelli et al., 2016; Klughammer et al., 2018). Together,  
116 these findings suggest that stochastic DNA methylation alterations contribute to tumor

117 heterogeneity and cellular plasticity and may drive clonal evolution of treatment-resistant  
118 phenotypes.

119         Single-cell DNA methylation technologies have recently emerged as tools to further  
120 dissect heterogeneous cell populations (Angermueller et al., 2016; Argelaguet et al., 2019;  
121 Farlik et al., 2016; Zhu et al., 2018b) and define epigenetic states that contribute to tumor  
122 evolution (Bian et al., 2018; Gaiti et al., 2019). Here, we integrated single-cell DNA methylomes,  
123 single-cell transcriptomes, and single-cell copy number profiles with bulk genomic profiles  
124 across a cohort of 11 glioma patient samples to deconstruct the sources of glioma  
125 heterogeneity. These analyses identified the gene regulatory regions most susceptible to  
126 stochastic DNA methylation alterations, the epigenetic modulation of transcriptional networks  
127 involved in glioma cellular identity, and that genetic driver events largely precede DNA  
128 methylation diversification during glioma evolution. We confirmed these single-cell findings  
129 through the association of DNA methylation instability across spatially separated and  
130 longitudinally collected bulk glioma tissue samples. Collectively, our work provides insights into  
131 the sources of intratumoral heterogeneity that fuel glioma evolution.

132

## 133 **RESULTS:**

### 134 **Single-cell DNA methylation highlights inter- and intratumoral heterogeneity at gene** 135 **regulatory regions.**

136 To investigate glioma heterogeneity we performed single-cell DNA methylation, single-cell gene  
137 expression, and accompanying bulk tumor profiling in 11 adult patients with glioma (Figure 1A).  
138 This cohort was representative of two principal molecular subtypes (IDH-mutant and IDH-wild-  
139 type) and captured distinct clinical time points (i.e., unmatched initial and recurrent tumors,  
140 Table S1 and Figure S1). We mechanically dissected tumor specimens from the same  
141 geographic region dissociating tissue for single-cell protocols and flash freezing tissue for bulk  
142 genomic assays (Figure 1A). We implemented an established single-cell DNA methylation  
143 protocol, reduced representation bisulfite sequencing (scRRBS), and 10X Genomics' single-cell  
144 gene expression protocol on cells from the same dissociation (Figure S2A) (Guo et al., 2015;  
145 Guo et al., 2013). Viable CD45<sup>-</sup> (i.e., pan-immune cell marker) cells were plated for scRBBS,  
146 while single-cell transcriptomics was performed on all viable cells deriving a set of 914 single-  
147 cell methylomes and 55,284 single-cell transcriptomes (Methods). On average, ~150,000  
148 unique CpG dinucleotides covering representative chromosomal regions were measured per  
149 cell (Figure S2B-E), and expression was measured on an average of 2,340 genes per cell.  
150 Tumor cells were defined based on the detection of inferred copy number alterations in both

151 datasets resulting in a final set of 844 tumor cells for single-cell DNA methylation and 30,831  
152 tumor cells for single-cell transcriptomics (Methods, Figure S3A-I).

153 Unsupervised clustering and multidimensional scaling of the pairwise distances between  
154 single-cell genome-wide DNA methylation patterns grouped tumor cells by *IDH1* mutation status  
155 consistent with IDH-mutant tumors displaying greater genome-wide DNA methylation levels  
156 (Figure 1B and Figure S4A, Wilcoxon  $p < 2.2e-16$ ). The co-localization of cells from different  
157 patients suggested some shared epigenetic states, while the isolated patient-specific grouping  
158 of 1 of 6 IDH-mutant and 2 out of 5 IDH-wild-type tumors suggested that genetic intertumoral  
159 heterogeneity identified by whole genome sequencing was also observed at the epigenetic level  
160 (Figure S1 and Figure 1B).

161 We next sought to determine the extent of intratumoral epigenetic heterogeneity by  
162 quantifying stochastic DNA methylation alterations in each single cell. In a non-diseased gene  
163 regulatory context, there is a general DNA methylation congruence in nearby CpGs reflecting  
164 tightly ordered gene regulation (Figure 1C top panel) (Kelsey et al., 2017). Epimutations reflect  
165 local DNA methylation disorder and may disrupt both local and distant gene regulation (Figure  
166 1C bottom panel) (Easwaran et al., 2014). We constructed an epimutation burden metric per cell  
167 measured by the proportion of sequencing reads discordant for DNA methylation status as  
168 previously described (Gaiti et al., 2019; Klughammer et al., 2018; Landau et al., 2014). Cell-to-  
169 cell variation in epimutation burden was tumor dependent (Figure 1D) and was increased in  
170 tumor cells compared with non-tumor cells across IDH-mutant and IDH-wild-type glioma  
171 subtypes (Wilcoxon  $p < 0.0001$ , Figure S4B). The mean epimutation burden across a tumor's  
172 single cells was not associated with the total somatic single nucleotide variant burden inferred  
173 through whole genome sequencing (Spearman correlation  $\rho = 0.26$ ,  $p = 0.43$ ), independent of  
174 sequence context (Figure S4C). However, there was a positive association between the fraction  
175 of the genome with somatic copy number alterations (SCNA burden) and epimutation burden  
176 (Spearman correlation  $\rho = 0.66$ ,  $p = 0.03$ ). Mutation burden reflects patient age (Alexandrov et  
177 al., 2020) and mutational processes (Figure S4D), while SCNA burden is associated with  
178 severed cell cycle checkpoints that compromises a cell's ability to correct mis-segregations (Zhu  
179 et al., 2018a). The stronger relationship with SCNA burden suggested that epimutation burden  
180 increases with advanced disease rather than being elevated in the tumor cell of origin.

181 We next examined whether stochastic DNA methylation changes might impact levels of  
182 DNA methylation and transcriptional output. First, we determined both the epimutation and DNA  
183 methylation levels per gene, observing significant positive correlations across gene regions  
184 (Spearman correlation  $p < 2.2e-16$ , Figure S5A-B). An increase in stochastic DNA methylation

185 might exert its functional impact by disrupting transcription programs (Gaiti et al., 2019). We  
186 leveraged companion single-cell RNAseq data to examine the association between epimutation  
187 burden, gene expression, and transcriptional variability. We observed both a reduction in mean  
188 expression (Kruskal Wallis  $p < 2.2e-16$ , Figure 1E) and an increase in gene dispersion  
189 (expression variability normalized for mean expression level) with increasing levels of  
190 epimutation burden in both IDH-mutant and IDH-wild-type tumors (Kruskal Wallis  $p < 2.2e-16$ ,  
191 Figure 1F, Figure S5C-D), implying that epimutation contributes to gene expression  
192 dysregulation. To understand how aberrant DNA methylation could rewire broader regulatory  
193 networks, we performed a Gene Ontology enrichment analysis on genes with high epimutation  
194 (i.e., epimutation burden  $> 0.5$ ) and genes with low epimutation (i.e., epimutation burden 0.0-  
195 0.1) (Methods), revealing that high epimutation genes were associated with processes involved  
196 in cellular differentiation and low epimutation genes were related with critical metabolic  
197 processes (Fisher's Exact test,  $p < 0.05$ , Figure S5E-F). The enrichment results were consistent  
198 when using epimutation burden groupings from the promoter or gene body (Figure S5G-H).  
199 Together, these results suggest that cells may acquire adaptive cell states through the  
200 accumulation of epimutations that impair normal transcriptional and differentiation programs.

201 Beyond proximal dysregulation of gene expression, epimutations may impact the binding  
202 of key transcription factors, as changes in DNA methylation at DNA-binding motifs can positively  
203 or negatively impact transcription factor binding (Yin et al., 2017). To identify regulatory  
204 elements more prone to stochastic DNA methylation changes we determined the epimutation  
205 burden across the transcription factor binding sites (TFBS) listed in the JASPAR database  
206 (Methods, Figure 1G). Consistent with observed subtype-specific differences in methylation  
207 disorder, the majority of transcription factor binding sites had increased epimutation burden in  
208 IDH-wild-type compared with IDH-mutant cells. In both subtypes, transcription factors shown to  
209 be essential for glioma stem cell maintenance (e.g., SOX2, SOX9, etc. (MacLeod et al., 2019))  
210 had lower than the median binding site epimutation burden suggesting that selection may act to  
211 deplete stochastic DNA methylation at these target regions (Figure 1G). In contrast,  
212 transcription factors that displayed high epimutation levels (Methods, Figure S6A) were  
213 associated with response to extracellular stimuli (Fisher's Exact test,  $p < 0.05$ , Figure S6B).  
214 These findings suggest that increased epimutation levels at these environmental stress  
215 response regulators may facilitate an adaptive response to stressors such as hypoxia, which is  
216 commonly observed in glioma (Jin et al., 2017). To substantiate this association in the bulk  
217 glioma tissues, we performed single-sample Gene Set Enrichment Analyses (ssGSEA,  
218 Methods) using bulk RNAseq data and demonstrated robust associations between tumor

219 average epimutation and positive stress response regulation (Spearman correlation  $\rho = 0.9$ ,  $p$   
220  $< 0.01$ ) or cellular response to hypoxia (Spearman correlation  $\rho = 0.98$ ,  $p < 0.001$   
221 respectively), but not randomly selected genes (Spearman correlation  $\rho = -0.05$ ,  $p > 0.05$ ,  
222 Figure 1H). Taken together, these results suggest that intratumoral variability in single-cell DNA  
223 methylation disorder may facilitate the adoption of distinct phenotypic states in response to  
224 stress stimuli.

225

## 226 **Integrative single-cell gene expression and DNA methylation analyses nominate** 227 **epigenetic regulators of glioma cell states and stress response.**

228 To further examine the association between DNA methylation, stress response, and cellular  
229 states, we defined each tumor's cellular composition from the single-cell transcriptional profiles.  
230 We performed unsupervised clustering analysis of all single cells and annotated clusters using  
231 marker genes (Figure 2A, Figure S7A-D) that revealed glial, immune, stromal, and malignant  
232 populations previously identified in glioma (Bhaduri et al., 2020; Wang et al., 2019). Malignant  
233 cells were broadly distributed over three cell states that all expressed canonical stem cell  
234 marker *SOX2* (Figure S7B). These pan-glioma states exist across both IDH-mutant and IDH-  
235 wild-type tumors, which we labelled as 1. differentiated-like, 2. stem-like, and 3. proliferating  
236 stem-like tumor cells, on the basis of marker gene expression (Figure 2A, Figure S7B, Table  
237 S2). Enumerating the proportion of pan-glioma malignant states by tumor of origin showed that  
238 IDH-mutant gliomas contained high fractions of stem-like cells (median 61%), while IDH-wild-  
239 type gliomas were dominated by differentiated-like cells (median 83%) and significantly higher  
240 fractions of proliferating stem-like cells (16% IDH-wild-type vs. 2% IDH-mutant, Wilcoxon  $p =$   
241  $0.01$ , Figure 2B). Malignant cell state diversity (Shannon diversity index) was not associated  
242 with epigenetic (Spearman correlation  $\rho = 0.12$ ,  $p > 0.05$ ) or genetic burden metrics  
243 (Spearman correlation  $\rho = -0.18$ ,  $p > 0.05$ , Figure 2B). Previously described malignant  
244 signatures of IDH-mutant glioma included Astrocyte-like and Oligodendrocyte-like cell types  
245 (Venteicher et al., 2017), which correspond to “differentiated-like” cells here. IDH-wild-type  
246 glioma cellular states (Neffel et al., 2019) included the “Astrocyte-like” and “Mesenchymal-like”,  
247 which were identified as “differentiated-like” in our clustering (Figure 2B, Figure S7D-F). The  
248 “proliferating stem-like” and “stem-like” states in our pan-glioma classification align closely with  
249 the “Undifferentiated” cells in IDH-mutant tumors and “Oligodendrocyte progenitor-like” and  
250 “Neural progenitor-like” in the IDH-wild-type tumors (Figure S7D-F), thus highlighting  
251 consistency of these pan-glioma signatures with previously reported IDH-subtype specific  
252 signatures (Neffel et al., 2019; Venteicher et al., 2017).

253 We next inferred gene regulatory networks from IDH-mutant and IDH-wild-type single-  
254 cell gene expression data to identify transcription factors (TFs) governing cell states (Methods)  
255 (Aibar et al., 2017). The inferred TF activity demonstrated that the three pan-glioma cell states  
256 are each regulated by a small set of TFs (Figure 2C-D). For example, stem-like tumor cells  
257 demonstrated the highest activity for known stem-cell regulators such as SOX2, SOX8, and  
258 OLIG2 in both the IDH-mutant and IDH-wild-type tumors (Figure 2C-D). In addition to high  
259 activity for these transcription factors, proliferating stem-like cells also had an  
260 overrepresentation of gene regulatory networks involved in chromatin remodeling and DNA  
261 repair such as those directed by EZH2 and BRCA1 (Figure 2C-D). In contrast, differentiated-like  
262 cells had higher transcription factor activities involved in astrocyte differentiation (i.e., SOX9)  
263 and response to stress stimuli (i.e., JUND, PPARA, HIF2A). We then tested whether the  
264 epimutation burden differed between cell state-specific transcription factors and did not find  
265 significant differences (Kolmogorov-Smirnov test  $p > 0.05$  Figure S7G-H). However, several  
266 transcription factors associated with the differentiated-like cell state (e.g., JUND, TFE3, and  
267 SREBF1) were characterized by high epimutation levels, nominating them as regulators of  
268 cellular fitness (Figure 2C-D).

269 To define the epigenetic states of stem-like and differentiated-like cells in glioma, we  
270 used the linked inference of genomic experimental relationships (LIGER) method to identify  
271 shared properties between single-cell gene expression and DNA methylation data (Methods,  
272 Figure 2E) (Welch et al., 2019). We found that the distribution of tumor cell states within each  
273 sample was consistent between the two methods, as expected from the same tissue  
274 dissociation (Figure S8). We next investigated whether there were different levels of DNA  
275 methylation and epimutation between the two broad cell state classifications of stem-like  
276 (combining stem-like and proliferating stem-like) and differentiated-like. In IDH-mutant tumors,  
277 stem-like cells had significantly higher levels of both epimutation burden ( $p = 2.4e-13$ ; Figure 2F  
278 left panel, Figure S9A) and DNA methylation ( $p = 6.0e-04$ , Figure 2G left panel, Figure S9B)  
279 likely reflecting elevated DNA methylation at genes responsible for cellular differentiation (Figure  
280 S5H). In IDH-wild-type tumors, which are marked by higher levels of epimutation and lower  
281 levels of DNA methylation compared with IDH-mutants, the differences between differentiated-  
282 like and stem-like cell populations demonstrated greater variability in both epimutation ( $p = 0.51$ ;  
283 Figure 2F, Figure S9C) and DNA methylation ( $p = 0.48$ ; Figure 2G and Figure S9D) suggesting  
284 loosened epigenetic control over cell states. To identify changes in DNA methylation between  
285 differentiated-like and stem-like cells in both IDH-wild-type and IDH-mutant glioma, we  
286 compared DNA methylation between cell states using a linear mixed effect model with tumor of

287 origin as the random effect (Methods). Regions with increases in DNA methylation in stem-like  
288 cells were enriched for binding sites of SP1 and TFAP2A, two transcription factors that  
289 frequently cooperate in regulation of development associated genes (Figure 2H) (Orso et al.,  
290 2010). In addition, the analysis identified enrichment of increased DNA methylation at binding  
291 sites of the HIF1A/ARNT master transcriptional regulator of hypoxic response, in stem-like cells  
292 (Figure 2H). As increased DNA methylation at binding sites may result in reduced transcription  
293 factor binding efficiency, these results suggest that elevated cell stress transcription factor  
294 activity in differentiated-like cells may occur via dynamic epigenetic remodeling (Figure 2H). We  
295 found only a few regions where there was an increase in DNA methylation in differentiated-like  
296 cells (Figure 2H). Together, these results suggest that perturbing epigenetic control via  
297 epimutation may promote the adaptive cell states necessary to tolerate diverse stressful  
298 microenvironments, including hypoxia (Li et al., 2009) and therapy (Liau et al., 2017; Shaffer et  
299 al., 2017; Sharma et al., 2010).

300

301 **Somatic copy number alterations are associated with stochastic DNA methylation**  
302 **changes during disease evolution.**

303 We next investigated whether genetic stimuli could help explain the variability of  
304 epimutation burden across glioma cells. The fraction of genome with SCNAs was associated  
305 with epimutation burden at the bulk level (Spearman correlation  $\rho = 0.66$ ,  $p = 0.03$ , Figure 1D)  
306 and we confirmed this broad observation at the single-cell level (Spearman correlation  $\rho =$   
307  $0.50$ ,  $p < 2.2e-16$  IDH-mutant and  $\rho = 0.72$ ,  $p < 2.2e-16$  IDH-wild-type, Figure 3A). To  
308 determine whether this relationship was driven by greater epimutation burden in copy number  
309 altered regions, we calculated the epimutation burden for each cell in copy number altered and  
310 non-altered regions. We did not observe a consistent relationship between epimutation burden  
311 and the copy number status in single-cell DNA methylation data (paired Wilcoxon  $p > 0.05$ ,  
312 Figure S10A). Instead, most genomic regions displayed a similar epimutation burden  
313 independent of copy number status (e.g., SM012, Figure S10A) suggesting that aneuploid  
314 regions do not directly account for increases in epigenetic diversity, but that these somatic  
315 events are likely shaped by similar biological processes (e.g., replication errors). Late replicating  
316 regions of the genome tend to accumulate more DNA mutations and structural rearrangements  
317 (Koren et al., 2012), and we discovered that single-cell epimutation burden across both  
318 promoter and gene body regions was positively associated with later replication regions in both  
319 subtypes (Kruskal Wallis  $p < 1e-04$ , Figure 3B). Late replicating genomic regions may have

320 reduced capacity to correct epimutations leading to their accumulation in a largely stochastic  
321 manner.

322 To validate the relationship between SCNA and epimutation burden in a larger cohort,  
323 we re-analyzed the bulk RRBS and copy number profiles of initial ( $n = 255$  patients) and  
324 recurrent ( $n = 152$  patients) IDH-wild-type gliomas, including a subset of longitudinally collected  
325 samples ( $n = 98$  patients) (Klughammer et al., 2018). We confirmed our findings by  
326 demonstrating that SCNA burden was positively associated with epimutation burden at both  
327 initial and recurrent timepoints (Spearman correlation  $\rho = 0.43$ ,  $p = 3.5e-13$  initial;  $\rho = 0.36$ ,  
328  $p = 6.2e-06$  recurrence, Figure 3C). A multivariable linear regression verified that this positive  
329 association between epimutation and SCNA burden was independent of subject age, tumor  
330 timepoint, and cellular proliferation (Figure S10B). To assess the relationship between  
331 longitudinal changes in SCNA burden and epimutation, we restricted our analysis to paired  
332 initial and recurrent samples and observed a positive association between increases in SCNA  
333 burden and epimutation (Spearman's correlation  $\rho = 0.37$ ,  $p = 0.0002$ , Figure 3D).  
334 Furthermore, the highest increases in epimutation burden between initial and the recurrent  
335 tumor were associated with a shorter time to second surgery in both univariate (log-rank test  $p =$   
336  $0.02$ , Figure 3F) and multivariate survival analyses (Cox proportional hazard model, HR = 1.69  
337 95% CI (1.09 – 2.62),  $p = 0.02$ , Table S3) supporting that increased epigenetic instability is  
338 associated with accelerated disease progression. SCNA burden or aneuploidy results from mis-  
339 segregation during cell cycle, which can further perpetuate epimutations through aneuploid-  
340 induced metabolic and replication stress (Zhu et al., 2018a). The association between  
341 aneuploidy and epimutation uncovered here implicates that defective cell cycle checkpoints  
342 compromises genomic but also epigenetic integrity.

343  
344 **Clonal evolution analyses highlight early somatic copy number evolution followed by**  
345 **epigenetic and transcriptomic diversification.**

346 The processes resulting in genetic, epigenetic, and transcriptomic heterogeneity may act  
347 at different times during tumor development. To evaluate somatic alteration timing and delineate  
348 intratumoral heterogeneity in the 11 glioma specimens, we reconstructed each tumor's  
349 evolutionary history from bulk tumor whole genome sequencing data. Briefly, we determined the  
350 clonality of SCNAs and somatic point mutations assigning each genomic alteration to a tumor  
351 subclone (Methods). One to four genetic subclonal populations were detected per tumor, with  
352 linear and branched evolutionary patterns consistent with those previously observed in glioma  
353 (Kim et al., 2015; Korber et al., 2019). Assessment of the timing of genetic events revealed that

354 chromosomal arm-level SCNA events were more likely to be classified as clonal (Fisher's exact  
355  $p = 0.03$ ), while mutations at genes significantly mutated in glioma were more evenly distributed  
356 across subclones (56.1% classified as clonal in non-hypermutable tumors) (Methods, Figure 4A,  
357 Figure S11A-H). To identify any copy number alterations associated with DNA methylation  
358 states we compared phylogenetic and phyloepigenetic trees derived from single-cell DNA  
359 methylation data using an entanglement coefficient where a low value (i.e., 0 entanglement)  
360 indicates complete single cell alignment across the tree structures (Methods, Figure S12). We  
361 did not observe strong alignments between phylogenetic and phyloepigenetic trees (median  
362 value = 0.46, Figure 4B); instead DNA methylation profiles grouped by similar cell states,  
363 showing that, like transcriptomic cell states, epigenetic clones are distributed across genetic  
364 clones. Moreover, these results suggest that many large-scale copy number alterations occur as  
365 early events which propagate broad epigenetic diversification rather than simply affecting  
366 methylation state in proximal regions (Figure 4B, Figure S12A-K).

367 We next asked whether genetic subclones within a tumor were associated with  
368 transcriptional diversity. We first used single-cell transcriptome inferred copy number profiles  
369 and found that three of eleven tumors (SM001, SM006, and SM012) had at least two distinct  
370 clones with chromosome arm-level alterations (Figure 4C, Figure S3). These tumors  
371 demonstrated significant shifts in cell state distributions across clones suggesting that the  
372 genetic heterogeneity also increases transcriptomic heterogeneity (per sample Fisher's Exact  
373 test,  $p < 0.05$ , Figure 4D). Changes in transcriptional states may reflect cellular behaviors  
374 required to adapt to the varied microenvironmental niches, such as hypoxia, within a tumor.

375 Previous studies have demonstrated that *EGFR*-amplifying extrachromosomal DNA  
376 (ecDNA) elements are common in IDH-wild-type gliomas and enable widespread genomic  
377 heterogeneity through both the amplification of oncogenes as well as enhancer elements  
378 (deCarvalho et al., 2018; Morton et al., 2019; Wu et al., 2019). Therefore, we hypothesized that  
379 ecDNA may represent a particularly potent contributor to genomic heterogeneity whose impact  
380 extends to epigenetic and transcriptomic diversity (Verhaak et al., 2019; Wu et al., 2019). We  
381 detected ecDNAs by analyzing whole-genome sequencing data for our cohort and validated the  
382 variable distribution of extrachromosomal *EGFR* elements within a tumor using fluorescence *in*  
383 *situ* hybridization for *EGFR* (Figure 4E, Figure S13A-D). *EGFR* ecDNAs, like chromosomal arm  
384 level events (e.g., chr7 amplification in SM001) were able distinguish subsets of tumor cells  
385 (e.g. *EGFR* ecDNA in SM006) (Figure 4C, Figure S11F-G). We classified both single-cell DNA  
386 methylation and RNA profiles as ecDNA+ or ecDNA- based on *EGFR* copy number level  
387 (Figure S13E). We observed ecDNA+ cells had increased genome-wide DNA methylation in 3

388 of 4 cases (Wilcoxon  $p < 0.05$ , Figure S13F) and greater transcriptional diversity using gene  
389 count signatures compared with ecDNA- cells (Wilcoxon  $p < 0.05$ , Figure S13G, Methods)  
390 (Gulati et al., 2020). The tumor with the highest number of genetic subclones and epimutation  
391 burden (SM012) contained an *EGFR* amplifying ecDNA assigned to subclones 3 and 4 which  
392 were marked by differential expression of a receptor tyrosine kinase gene signature. The  
393 ecDNA(-) subclone 2 most closely associated with hypoxia gene expression (Wilcoxon  $p < 2.2e$ -  
394 16, Figure 4F), providing an example of how genetic heterogeneity may influence epigenetic  
395 and transcriptional reprogramming.

396 Taken together, our evolutionary analyses show that genetic evolution largely precedes  
397 epigenomic and transcriptomic diversification, and that intratumoral genetic heterogeneity  
398 influences but does not determine cell states.

399  
400 **Integrated molecular trajectories supports adaptive DNA methylation changes under**  
401 **microenvironmental and therapeutic pressures.**

402 Our observation that genetic events likely precede epigenetic and transcriptomic diversification  
403 led us to ask whether epigenetic diversity accelerates tumor evolution by promoting cell survival  
404 in resource-deprived tumor environments (e.g., hypoxia or therapeutic exposures). To address  
405 this question and extend the generalizability of our findings, we sought to determine variable  
406 intratumoral DNA methylation levels in large-scale bulk glioma studies (Barthel et al., 2019;  
407 Ceccarelli et al., 2016; Verburg et al., 2020). Since these datasets were generated using DNA  
408 methylation microarrays, we used our single-cell DNA methylation data to define a microarray  
409 metric that quantified the DNA methylation instability of gene regions prone to epimutation  
410 (Figure 1E and Figure 5A). We reasoned that regions most susceptible to DNA methylation  
411 changes would reflect this stochasticity in bulk data by taking on intermediate DNA methylation  
412 values (Figure 5A). We confirmed that this epimutation, or DNA instability metric, approximated  
413 that of single-cell epimutation averages from the same tumor by comparing to microarray-  
414 derived profiles across the 11 tumors in our cohort (Spearman correlation  $\rho = 0.65$   $p = 0.02$ ,  
415 Figure S14A). We first applied this metric to The Cancer Genome Atlas (TCGA) data and found  
416 that the DNA methylation instability metric demonstrated differences across the TCGA-defined  
417 subtypes (Ceccarelli et al., 2016), with IDH-wild-type tumors displaying the highest levels  
418 (Kruskal Wallis  $p < 2.2e$ -16, Figure 5B). Integrating matching DNA methylation and RNAseq  
419 samples from 568 TCGA samples, we found that samples with higher levels of DNA methylation  
420 instability levels showed increased transcriptional activity of oxidative stress response genes,  
421 which corroborated our earlier finding of stronger positive associations between epigenetic

422 instability and stress response regulation than randomly selected genes (Spearman  $\rho = 0.47$ ,  
423  $p < 2.2e-16$ ,  $n = 516$  IDH-mutant initial tumors,  $\rho = 0.31$ ,  $p = 0.03$ ,  $n = 52$ , IDH-wild-type initial  
424 tumors, Figure S14B-C).

425 We next applied the DNA methylation instability metric to 119 image-guided stereotactic  
426 biopsies taken from spatially distinct regions across IDH-wild-type ( $n = 57$  biopsies, 6 patients)  
427 and IDH-mutant ( $n = 62$  biopsies,  $n = 8$  patients) tumors (Verburg et al., 2020). This enabled us  
428 to quantify the physical distance between a biopsied sample and specific radiographic features  
429 that delineate the tumor's center (e.g., magnetic resonance imaging contrast-enhanced region,  
430 Figure S14D). We found an increase in DNA methylation instability closer to the tumor's center  
431 across IDH-wild-type tumors while adjusting for patient (multivariable linear regression  $p = 0.02$ ,  
432 Figure 5C), a region frequently characterized by hypoxia. The link between radiographic  
433 features and epigenetic shifts supports the association between cellular fitness and increased  
434 epigenetic plasticity. We did not observe a consistent relationship between tumor location and  
435 DNA methylation instability in IDH-mutant tumors (multivariable linear regression  $p = 0.31$ ,  
436 Figure 5D) where hypoxia is less common.

437 The environmental pressures that tumors face may vary over time. To assess DNA  
438 methylation instability dynamics and its relationship with genetic alterations, we analyzed initial  
439 and recurrent tumor samples from the Glioma Longitudinal AnalySiS (GLASS) consortium for  
440 which DNA sequencing and DNA methylation data were available ( $n = 102$  tumors,  $n = 51$   
441 patients). For each sample, we catalogued the specific copy number and DNA methylation  
442 alterations at individual CpG sites that changed between an initial tumor and its matched  
443 recurrence. Overall, we observed that DNA methylation changes were mostly decreases in DNA  
444 methylation consistent with previous findings (de Souza et al., 2018; Mazor et al., 2015), and  
445 that DNA methylation changes mainly occurred in regions that remained copy number stable  
446 between timepoints (Figure 5E). We then tested for DNA methylation changes following  
447 treatment while accounting for differences in cellular composition of the tumor microenvironment  
448 (Methods, Figure S14E). We discovered that regions with consistently altered DNA methylation  
449 independent of changes in microenvironment cell type distribution were enriched for the binding  
450 sites of transcription factors that regulate cellular stress response, particularly hypoxia  
451 (Methods, Figure 5F-G). We also observed the enrichment for differential DNA methylation  
452 among TFs that differed between stem-like and differentiated-like states in our single-cell data  
453 (e.g., SP1 and TFAP2A, Figure 5F and Figure 2H). These observations support our single-cell  
454 findings that regions with greatest epimutation levels are involved with processes regulating  
455 cellular differentiation and stress signaling. In summary, these results suggest that stochastic

456 DNA methylation alterations can provide the variability necessary to enable transition to  
457 adaptive epigenetic phenotypes that are responsive to cellular stress (Figure 6).

458  
459

## 460 **DISCUSSION:**

461 Here, we integrated multimodal single-cell DNA methylation and transcriptomic profiles  
462 along with bulk profiles to interrogate the association between genetic tumor subclones, cellular  
463 states and epigenetic heterogeneity of glioma. We found that early genetic alterations largely  
464 precede epimutations, or stochastic changes in DNA methylation, whose accumulation  
465 throughout the genome led to dysregulated transcription and altered cellular states. Despite  
466 extensive intertumoral heterogeneity, we found recurrent epimutations localized to cellular  
467 differentiation genes and higher epimutation levels were associated with environmental  
468 pressures, such as hypoxia, highlighting a mechanism to overcome cell stress and enhance  
469 treatment resistance. Taken together, epigenetic intratumoral heterogeneity provides a plastic  
470 intermediate between genetic subclones and adaptive phenotypic cell states.

471 Epimutations increase a tumor population's epigenetic diversity through random errors in  
472 the DNA methylation replication machinery (Klughammer et al., 2018; Landan et al., 2012;  
473 Landau et al., 2014). We found that genetic and environmental stimuli further induce epigenetic  
474 variability likely through altered cellular metabolism. Deregulated metabolism is a hallmark of  
475 glioma characterized by somatic mutations in the metabolic isocitrate dehydrogenase (IDH)  
476 genes and a hypoxic microenvironment in IDH-wild-type tumors. IDH-mutant glioma cells  
477 produce the oncometabolite 2-hydroxyglutarate (2HG) that interferes with DNA demethylation  
478 (Ceccarelli et al., 2016; Dang et al., 2009; Losman and Kaelin, 2013; Noushmehr et al., 2010;  
479 Turcan et al., 2012) leading to the observed high promoter epimutation levels at cellular  
480 differentiation genes and the predominance of a stem-like cell state. Across both subtypes,  
481 epimutation level was positively associated with broad chromosomal alterations, such as arm-  
482 level gains and losses, but not mutational burden. Copy number alterations occur during  
483 replicative crises that originate early in a tumor's life history through punctuated evolution (Gao  
484 et al., 2016; Gerstung et al., 2020). We used a multimodal approach to link genetic clones  
485 across platforms and found that chromosomal alterations precede epigenetic and transcriptomic  
486 heterogeneity. The chromosomal imbalances may potentiate non-genetic diversity by  
487 accelerating cell proliferation (Taylor et al., 2018) and generating metabolic disruption via  
488 reactive oxygen species (Zhu et al., 2018a), thereby increasing the likelihood of aberrant DNA  
489 methylation. We also found that environmental stimuli, such as hypoxia, increase the rate of

490 epimutation and is supported by a previous study that demonstrated hypoxia reduced the  
491 enzymatic activity of DNA methylation regulators (Thienpont et al., 2016). Tumor hypoxia is  
492 common across many cancers and could more broadly shape the phenotype of cells resistant to  
493 therapy through epimutation (Heddleston et al., 2010). Collectively, increased chromosomal  
494 alterations and more adverse microenvironments may explain the greater cell state diversity that  
495 exists in IDH-wild-type compared with IDH-mutant gliomas.

496 In a non-tumor setting, a cell's epigenome reflects the tissue of origin and serves to  
497 stabilize cell state-specific gene expression (Roadmap Epigenomics et al., 2015). Epimutations  
498 may occur when this homeostasis is disrupted, enabling cells to acquire a de-differentiated  
499 malignant cell state or create an altered epigenetic landscape permissive to cell state transitions  
500 (Flavahan et al., 2017). Glioma cell states have been described to fall along axes of  
501 differentiation and proliferating potential (Bhaduri et al., 2020; Neftel et al., 2019; Venteicher et  
502 al., 2017; Wang et al., 2019). In accordance with prior reports, we observed pan-glioma  
503 malignant cell states that were found within each tumor. Our epigenetic single-cell profiles  
504 revealed that cell state-defining transcription factor activity can be perturbed by epimutation.  
505 Thus, diverse DNA methylation marks help to sustain multiple cell states that each confer their  
506 own fitness advantages and together accelerate disease progression.

507 Intratumoral heterogeneity in glioma reflects the Darwinian process of subclonal  
508 competition driven by limited nutrient access. While single-cell transcriptome-based phenotypes  
509 have investigated glioma transcriptomic heterogeneity (Bhaduri et al., 2020; Neftel et al., 2019;  
510 Tirosh et al., 2016; Venteicher et al., 2017; Wang et al., 2019), we have only limited knowledge  
511 on the degree of epigenetic variability. The intratumoral epigenetic variation defined here  
512 provides a link between Darwinian clone wars and phenotypic state changes by enabling  
513 diverse responses to selective pressures such as hypoxia and treatment. A better  
514 understanding of how to reprogram the glioma epigenome toward a more therapeutically  
515 vulnerable cell state will be needed to develop more effective interventions. In summary, single-  
516 cell epigenetic profiles show that each cell contains a unique set of methylation marks with  
517 distinct patterns regulating cellular states and reflecting variable levels of environmental stress.

518

#### 519 **ACKNOWLEDGEMENTS:**

520 We would like to thank the patients and their families for their generous donation to biomedical  
521 research. We would also like to thank the staff in the following groups at The Jackson  
522 Laboratory for Genomic Medicine: single cell biology laboratory, flow cytometry core, and  
523 genomic technology core for assistance in data generation. We thank Matt Wimsatt for

524 assistance in graphic design. We thank the University of Texas MD Anderson Epigenomics  
525 Profiling Core for their assistance in helping troubleshoot the scRRBS protocol. We thank  
526 Sheng Li and Samirkumar B. Amin for helpful feedback that strengthened the work. This work  
527 was supported by NIH grants R01 CA237208, R21 NS114873 and Cancer Center Support  
528 Grant P30 CA034196; Department of Defense W81XWH1910246 (R.G.W.V). F.S.V. is  
529 supported by a postdoctoral fellowship from The Jane Coffin Childs Memorial Fund for Medical  
530 Research. F.P.B. is supported by the National Cancer Institute (K99 CA226387). E.Y. is a fellow  
531 of the American Brain Tumor Association. K.C.J. is the recipient of an American Cancer Society  
532 Fellowship (130984-PF-17-141-01-DMC).

533

#### 534 **AUTHOR CONTRIBUTIONS:**

535 K.C.J. and R.G.W.V. conceived the project and designed the experiments.

536 K.B. and S.D. curated patient samples and patient annotation.

537 K.C.J., M.R.H.E., M.T., N.N., R.M., C.N., M.S, and P.R performed single-cell library optimization  
538 and sequencing.

539 K.C.J. led data production and performed experiments with D.L., E.Y., and E.T.C.

540 K.C.J and K.J.A. led data analysis in collaboration with F.P.B., F.S.V., M.S., E.Y., and H.K.

541 K.C.J., K.J.A., and R.G.W.V. wrote the manuscript with input from all authors.

542

#### 543 **DECLARATIONS OF INTEREST:**

544 R.G.W.V. is a co-founder of and has received research support from Boundless Bio, Inc.

545

#### 546 **FIGURE LEGENDS:**

547 **Figure 1. Single-cell DNA methylation sequencing highlights intratumoral heterogeneity**  
548 **and disruption of epigenetic regulatory mechanisms.**

549 (A) Schematic diagram detailing tumor sample processing and molecular profiling of single cells  
550 and bulk tumor samples ( $n = 11$  subjects). (B) Multidimensional scaling (MDS) analysis using  
551 pairwise individual CpG distance metrics calculated between individual cells. Shapes represent  
552 whether a sample was a single tumor cell ( $n = 844$  cells) or 50-tumor cells,  $n = 9/11$  subjects).  
553 Colors indicate individual subjects, shaded regions indicated *IDH1*-mutation status of tumor, and  
554 annotation is provided indicating clinical timepoint (I = initial, R = recurrence). (C) Schematic  
555 depiction of local DNA methylation disorder in different genomic contexts. Left panel  
556 demonstrates epimutation, or local DNA methylation disorder, at the promoter region, where  
557 gene expression is disrupted by epimutation. The right panel provides an example of disrupted  
558 transcription factor binding due to epimutation. (D) Boxplots of tumor cell epimutation burden  
559 grouped by sample. Each boxplot spans the interquartile range with the whiskers representing  
560 the absolute range, excluding outliers. Wilcoxon rank sum  $p$ -value represents comparison  
561 between IDHmut and IDHwt epimutation burden. Each sample is annotated with clinical and  
562 molecular metrics with  $p$ -values indicating the relationship between sample mean epimutation

563 burden and whole-genome sequencing derived somatic mutation burden or somatic alteration  
564 burden (Spearman correlation). (E) Boxplots of gene expression values, as log<sub>2</sub> (counts per  
565 million), from single-cell RNAseq data across different gene epimutation groups. Gene  
566 epimutation groups are defined by the determining the mean epimutation value across a single  
567 gene. Color indicates *IDH1* mutation status. (F) Boxplots of gene expression dispersion.  
568 Expression profiles were mean-expression scaled to account for expression level-dependent  
569 variability across the same gene epimutation groups defined in panel E. (G) Scatterplot of the  
570 mean single-cell epimutation burden metric calculated across transcription factor binding sites  
571 (TFBSs) within a subtype, ordered by IDHwt TFBS epimutation. Each column represents a  
572 single transcription factor (TF) with a colored dotted line connecting IDHmut and IDHwt values.  
573 Names of TFs previously indicated to confer fitness advantages to glioma cells (MacLeod et al.)  
574 are listed above their TFBS epimutation burden estimate. (H) Scatterplot depicting the  
575 association between average single-cell epimutation burden estimate and single-sample Gene  
576 Set Enrichment Score for stress response, hypoxia, and random genes from bulk RNAseq data.  
577 Spearman correlation coefficient and *p*-values are indicated.  
578

579 **Figure 2. Integrative single-cell gene expression and DNA methylation analyses nominate**  
580 **epigenetic regulators of glioma cell state variability.**

581 (A) Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction plot of  
582 scRNAseq data ( $n = 55,284$  tumor cells,  $n = 11$  subjects) showing the clustering of cell  
583 populations by transcriptionally defined cell state (point color) and labelled according to marker  
584 gene expression (Figure S6B). (B) Stacked bar plots representing the proportion of cellular  
585 states per tumor for pan-glioma classification. Each sample is annotated with molecular metrics  
586 with *p*-values indicating the relationship between cell type diversity, measured by Shannon's  
587 entropy, and sample mean epimutation burden, whole-genome sequencing derived somatic  
588 alteration burden, or whole-genome sequencing derived somatic mutation burden (Spearman  
589 correlation). (C-D) Enriched transcription factor activity across pan-glioma cellular states  
590 determined using SCENIC algorithm and displayed as a heatmap of z-score enrichment values.  
591 Visualization is presented for the hierarchical clustering of 5,000 randomly selected tumor cells  
592 in both (C) IDHmut and (D) IDHwt tumors. (E) Schematic diagram representing LIGER workflow  
593 to jointly cluster single-cell RNAseq and DNA methylation data generated from the same tumor  
594 dissociation. (F) Boxplots representing the average epimutation burden in differentiated-like and  
595 stem-like populations in IDHmut (left panel) and IDHwt (right) tumors. (G) Boxplots representing  
596 the 10-kb tiled DNA methylation levels in differentiated-like and stem-like populations in IDHmut  
597 (left panel) and IDHwt (right) tumors. (H) Region set enrichment analysis for 10-kb tiles with  
598 higher DNA methylation in Stem-like (left panel) or Differentiated-like cells (right panel).  
599 Enrichment was determined by Locus Overlap Analysis (LOLA). Individual points represent  
600 enrichment of specific TFs in differentially methylated regions, color indicates results for specific  
601 IDH-mutant subtype, point size indicates log-odds ratio, and dotted line represents the statistical  
602 significance threshold (adjusted *p*-value < 0.05).  
603

604 **Figure 3. Somatic copy number alterations are associated with stochastic DNA**  
605 **methylation changes during disease evolution.**

606 (A) Scatterplot depicting the association between single-cell ( $n = 844$  tumor cells) somatic copy  
607 number alteration (SCNA) and epimutation burden estimates by IDHmut (left panel) and IDHwt  
608 (right panel) subtypes. Points are colored by patient. Spearman correlation coefficients  
609 represent subtype-specific estimates. (B) Boxplots of epimutation burden calculated across the  
610 promoter (left panel) and gene body regions (right panel) based on different DNA replication  
611 times in IDHmut ( $n = 510$ ) and IDHwt ( $n = 334$ ) single cells. Kruskal-Wallis *p*-values indicate a  
612 test for differences across the replication time groupings separately for IDHmut and IDHwt cells  
613 (C) Scatterplot depicting the re-analysis of bulk promoter epimutation burden and SCNA burden

614 in IDHwt initial ( $n = 255$ ) and recurrent ( $n = 152$ ) tumors (Klughammer et al.). Spearman  
615 correlation coefficients and  $p$ -values are presented for each independent timepoint. (D)  
616 Scatterplot depicting the association between bulk delta (subject-specific recurrence – initial  
617 estimates) SCNA burden and delta promoter epimutation burden in longitudinally profiled IDHwt  
618 tumors ( $n = 98$  subjects, Klughammer et al.) Spearman correlation coefficient and  $p$ -value are  
619 presented. (E) Kaplan-Meier curve depicting time to second surgery in subjects where the  
620 change in epimutation burden between initial and recurrent disease was above (high, red) and  
621 below (low, blue) the median. Log-rank  $p$ -value for univariate analysis is presented within the  
622 figure. Hazard Ratio and  $p$ -value for change in epimutation burden are presented below for  
623 multi-variate Cox proportional hazard model including subject age and sex as predictors.

624  
625 **Figure 4. Clonal evolution analyses highlight early copy number evolution followed by**  
626 **epigenetic and transcriptomic diversification.**

627 (A) Stacked bar plots representing the proportion of whole-genome sequencing (WGS) derived  
628 somatic copy number alteration (SCNA) burden attributed to clonal vs. subclonal events. (B)  
629 Scatterplot depicting entanglement coefficients for tanglegrams comparing cluster dendrograms  
630 of scRRBS derived copy number and DNA methylation profiles. A coefficient of 0 indicates  
631 complete alignment of the tree structures, whereas a 1 indicates random association. Color  
632 indicates *IDH1* mutation status. (C) Examples of phylogenetic trees constructed from whole  
633 genome sequencing data (mutations and SCNAs) and further annotated using single-cell  
634 inferred copy number alterations (scRRBS + scRNAseq). Tree nodes represent alterations  
635 specific to the given clone, with node size corresponding to the fraction of cells with the  
636 associated alterations. Branch length scales with the number of mutations attributed to that  
637 clone. Clonal alterations are colored in blue, with subclonal alterations colored in gold. Genes  
638 considered significantly mutated in TCGA analyses (Ceccarelli et al., 2016) and chromosomal  
639 arm-level events are presented. (D) Single-cell RNAseq-derived cellular proportions separated  
640 by copy number-defined tumor subclone (Figure S3). Reported  $p$ -values represent Fisher's  
641 exact test comparing the cellular state distributions across tumor subclones. (E) Representative  
642 Fluorescence *in situ* hybridization (FISH) images for IDHwt tumors computationally predicted to  
643 harbor *EGFR* extrachromosomal DNA (ecDNA) by whole genome sequencing ( $n = 4$  patients).  
644 FISH images show *EGFR* amplifications (red) that occur distal to control chromosome 7 probes  
645 (green) indicating extrachromosomal status and high variability in copy number status across  
646 tumor cells. Scale bars = 10 microns. (F) Ridge plots of SM012 single-cell expression of  
647 receptor tyrosine kinase and hypoxia-associated genes, grouped by copy number-defined  
648 subclones. Reported  $p$ -values represent Wilcoxon Rank Sum tests comparing the gene  
649 expression of cells across tumor subclones.

650  
651 **Figure 5. Integrated molecular trajectories supports adaptive DNA methylation changes**  
652 **under microenvironmental and therapeutic pressures.**

653 (A) Schematic workflow for construction of a DNA methylation instability metric in bulk cohorts  
654 informed by regions of high epimutation in single-cell DNA methylation data. The DNA  
655 methylation instability metric was calculated across bulk DNA methylation microarray data in a  
656 primary tumor cohort (TCGA), a cohort of multiple, spatially distinct biopsies from the same  
657 tumor (Verburg et al.), and a longitudinal cohort with accompanying genomic sequencing data  
658 (Glioma Longitudinal AnalySiS (GLASS), Barthel et al.). (B) Boxplots displaying the bulk DNA  
659 methylation instability metric calculated across previously described DNA-methylation based  
660 TCGA tumor classifications (Ceccarelli et al). Colors represent *IDH1/2* mutation status, and  
661 Kruskal-Wallis  $p$ -value testing for differences in distributions across classification is reported ( $n$   
662 = 615 primary gliomas,  $p < 2.2e-16$ ). (C-D) Scatterplots depicting distance from radiographic  
663 features plotted against the DNA methylation instability metric. Colors represent spatially  
664 separated biopsies from a single patient at initial clinical timepoint for (C) IDHwt tumors ( $n = 57$

665 biopsies,  $n = 6$  subjects) and (D) IDHmut tumors ( $n = 62$  biopsies,  $n = 8$  subjects). Linear  
666 regression lines colored by patient demonstrate the relationship between DNA methylation  
667 instability and radiographic features (i.e., contrast enhancement surface). The  $p$ -value reported  
668 from a multivariable linear regression model adjusting for subject represents the subtype-  
669 specific association between DNA methylation instability and radiographic feature. Biopsies  
670 taken closer to the tumor's center (i.e., core) have the lowest value (left hand side of plot). (E)  
671 Each column represents an individual patient sampled across initial and recurrent timepoints  
672 and is separated into IDHmut ( $n = 24$  subjects) and IDHwt ( $n = 27$  subjects). Top panel, stacked  
673 bar plot represents the proportion of CpGs sites that experienced DNA methylation change  
674 associated with a subject-specific copy number change (defined by DNA sequencing data)  
675 between primary and recurrent disease (red), DNA methylation gain not associated with a CNV  
676 change (orange), DNA methylation loss not associated with a CNV change (blue), and no  
677 longitudinal DNA methylation change (gray). Middle panel, heatmap of DNA methylation  
678 instability metric in primary and recurrent disease (blue = low, red = high). Bottom panel,  
679 differences in SCNA burden between primary and recurrent tumor. All associated  $p$ -values  
680 represent Spearman correlations between absolute change in associated metric and the fraction  
681 of longitudinal DNA methylation differences. (F) Enrichment analysis for differentially methylated  
682 CpGs between primary and recurrent timepoints when adjusting for cellular composition, glioma  
683 subtype, and subject included as a random effect. Individual points represent enrichment of  
684 specific TFs in differentially methylated positions, color indicates the average TFBS epimutation  
685 burden from single-cell RRBS data (Figure 1G), point size indicates log-odds ratio, and dotted  
686 line represents the statistical significance threshold ( $Q$ -value  $< 0.05$ ). (G) Gene Ontology  
687 enrichment of transcription factors associated with longitudinal DNA methylation changes.  
688 Dotted line represents threshold for statistical significance (Fisher's exact test,  $p < 0.05$ ).  
689

690 **Figure 6. Model of epigenetic heterogeneity and tumor evolution.** Schematic depiction of  
691 tumor evolution with general DNA methylation patterns represented by methylated (5-  
692 methylcytosine, 5mC) and unmethylated (5C) regions of the epigenome. Initiating genetic  
693 events such as *IDH1* and other driver mutations as well as somatic copy number alterations  
694 represent early stresses in glioma evolution that precipitate epigenetic heterogeneity. Both  
695 mutations in epigenetic enzymes and SCNAs can increase the likelihood of heritable DNA  
696 methylation alterations (i.e., epimutations). *IDH1* mutations result in the production of the  
697 oncometabolite 2-Hydroxyglutarate (2-HG) that leads to failure to remove aberrant DNA  
698 methylation while SCNAs can generate mitotic stress leading to the erosion of ordered DNA  
699 methylation. Non-genetic determinants further shape epigenetic heterogeneity as tumors evolve  
700 by exposing cells to spatially distinct microenvironmental stresses that impact the DNA  
701 methylation replication machinery. The subsequent epigenetic diversity provides an additional  
702 layer on which clonal evolution acts to select those cells with fitness-conferring epigenetic  
703 alterations. Ultimately, the loosened epigenetic control allows tumor cells to transition to cell  
704 states responsive to different selective pressures.

705  
706  
707 **Figure S1. Integrated molecular profiles of patient samples. Related to Figure 1.**  
708 Each patient is in a single column with data presented to indicate clinical features (top), followed  
709 by genetic alterations defined from bulk whole genome sequencing data, bulk RNA sequencing  
710 based subtype classification probabilities (Wang et al.,  $n = 8$  available), single-cell RNA tumor  
711 cellular state proportions, bulk DNA methylation microarray subtype classification probabilities  
712 (Capper et al.), and boxplots of single-cell epimutation burden with samples colored by clinical  
713 timepoint.  
714

715 **Figure S2. Sample pre-processing and metrics related to single-cell DNA methylation**  
716 **data assessment. Related to Figure 1.**

717 (A) Representative fluorescence activated cell sorting (FACS) data and strategy for viable cell  
718 enrichment for both single-cell protocols, and tumor cell enrichment in scRRBS. (B) The same  
719 multidimensional scaling (MDS) analysis using pairwise distance metrics calculated between  
720 individual cells as in Figure 1B, except colored by bisulfite conversion efficiency. (C) The  
721 number of unique CpGs detected per single cell, with the red line indicating the threshold  
722 (minimum 40,000 unique CpGs) for inclusion in the dataset presented herein. (D)  
723 Representative distribution of single locus DNA methylation estimates for a single cell. DNA  
724 methylation percentage of 0 represents an unmethylated locus, while a percentage of 100  
725 represents a methylated locus. (E) Representative genomic distribution of DNA methylation  
726 values within a single cell.

727  
728 **Figure S3. Somatic copy number alteration examples estimated from whole genome**  
729 **sequencing, single-cell Reduced Representation Bisulfite Sequencing, and single-cell**  
730 **RNA-sequencing. Related to Figure 1.**

731 (A-C) Representative images of copy number alterations derived from SM012 (IDHwt initial)  
732 whole genome sequencing (WGS) data. (A) Depth ratio for each segment with copy number  
733 status determined as compared with germline (normal blood) WGS data. (B) SM012 Single-cell  
734 DNA methylation-based copy number estimates ( $n = 69$  tumor cells) with copy number integer  
735 depicted by color (blue = CN loss, white = neutral CN, and red = CN gain). Each row is a single  
736 cell with annotation for epimutation burden provided. (C) SM012 Single-cell RNAseq based  
737 copy number inference ( $n = 5,489$ ) identifying major copy number events found in WGS with  
738 labelled subclones as presented in Figure 4D. (D-F) Similar example profiles as presented in (A-  
739 C) for tumor sample SM006 (IDHwt initial,  $n = 82$  scRRBS cells,  $n = 3,310$  scRNAseq cells). (G-  
740 I) Similar example profiles as presented in (A-C) for tumor sample SM001 (IDHmut recurrence,  
741  $n = 181$  scRRBS cells,  $n = 5,713$  scRNAseq cells).

742  
743 **Figure S4. Distribution and relationship of DNA methylation and epimutation throughout**  
744 **the glioma genome. Related to Figure 1.**

745 (A) Boxplots representing average 10-kb tiled DNA methylation values per single tumor cell. (B)  
746 Boxplots highlighting the single-cell epimutation burden estimates calculated across different  
747 genomic contexts. (C) Scatterplots showing the relationship between genomic context-specific  
748 single-cell epimutation burden (sample-specific scRRBS average) and genomic context-specific  
749 mutation burden derived from whole genome sequencing ( $n = 10$  excluding hypermutant  
750 sample). Panels are separated into global (i.e., all regions), promoter, gene body, and intergenic  
751 regions (Spearman correlations  $p > 0.05$  for all comparisons). (D) The dominant Catalogue of  
752 Somatic Mutations in Cancer (COSMIC v3) mutational signatures are presented for each  
753 subject. The stacked bar plots represent the relative contribution of each mutational signature to  
754 the tumor's mutational burden. Colors indicate distinct mutational signatures, which are further  
755 annotated with their proposed etiology.

756  
757 **Figure S5. Association between epimutation and disrupted transcriptional programs.**  
758 **Related to Figure 1.**

759 (A-B) Scatterplots depicting single-cell gene-level epimutation average plotted against the gene-  
760 level methylation estimates in both (A) promoter regions and (B) gene body regions. (C)  
761 Boxplots of gene expression values, in log<sub>2</sub> (counts per million), from single-cell RNAseq data  
762 across different sets of promoter regions defined by gene-derived epimutation groups. Gene  
763 epimutation groups are defined by the determining the mean epimutation value across a single  
764 gene. Color indicates *IDH1* mutation status. (D) Boxplots of gene expression dispersion that  
765 were mean-expression scaled to account for expression level-dependent variability across the

766 same promoter-based gene epimutation groups defined in panel C. (E-F) Gene Ontology  
767 enrichment analyses for low epimutation genes (Figure 1E, mean epimutation across all tumor  
768 cells: 0.0 - 0.1) and high epimutation genes (Figure 1F, mean epimutation across all tumor cells:  
769 > 0.5) using gene body estimates. A meta biological process is placed next to significant Gene  
770 Ontology terms. (G-H) Same analyses presented in panels E-F, but for gene-level epimutation  
771 estimates determined in promoters.  
772

773 **Figure S6. Enrichment of high epimutation transcription factors and association with**  
774 **environmental stress response. Related to Figure 1.**

775 (A) Computational approach to defining TFBSs with high epimutation burden (red tangent line at  
776 0.399 TFBS epimutation burden). X-axis represents each TF ordered by mean epimutation  
777 burden in IDHwt single-cells ( $n = 334$  cells). (B) Gene Ontology enrichment analysis of TFs with  
778 high epimutation burden in their binding sites. A meta biological process is placed next to  
779 significant Gene Ontology (GO) terms.  
780

781 **Figure S7. Pan-glioma cell state assignment and characteristics. Related to Figure 2.**

782 (A) UMAP dimensionality reduction plot of all scRNAseq data, including tumor and non-tumor  
783 cells ( $n = 55,248$  cells). Each dot depicts a single cell and colors represents the tumor of origin.  
784 Shaded regions represent cell state classification. (B) Stacked violin plots of average single-cell  
785 gene expression for cells presented in Figure S7A. Selected genes presented are informative  
786 for cell state classification. (C) Stacked bar plots representing the proportion of non-tumor  
787 cellular states (D) Stacked bar plots representing the proportion of tumor cellular states per  
788 tumor for pan-glioma classification (top row) and previously published classifications (middle  
789 row; Venteicher et al. and Neftel et al.) (E) Sankey plot representing the proportion of IDHmut  
790 tumor cells with pan-glioma classification and associated classification described in Venteicher  
791 et al. (F) Sankey plot representing the proportion of IDHwt tumor cells with pan-glioma  
792 classification and associated classification described in Neftel et al. (G) Density plots  
793 representing TFBS epimutation burden (scRRBS data) in IDHmut single-cell DNA methylation  
794 data for TFs whose activity (scRNAseq based SCENIC analysis) characterizes a specific cell  
795 state ( $n = 20$  TFs per cell state). Kolmogorov-Smirnov  $p$ -value tests for differences in TFBS  
796 epimutation burden across the cellular states. (H) Density plots representing TFBS epimutation  
797 burden (scRRBS data) in IDHwt single-cell DNA methylation data for TFs whose activity  
798 (scRNAseq based SCENIC analysis) characterizes a specific cell state ( $n = 20$  TFs per cell  
799 state). Dotted lines represent the median TFBS value for cell state defining TFs. The  
800 Kolmogorov-Smirnov  $p$ -value corresponds to differences in TFBS epimutation burden across  
801 the cellular states.  
802

803 **Figure S8. LIGER integrated tumor-specific clustering of single-cell RNA and single-cell**  
804 **DNA methylation data. Related to Figure 2.**

805 Joint single-cell RNAseq (scRNA) and single-cell DNA methylation (scDNAm) clustering and  
806 UMAP projections highlighting similar cellular state distributions across platforms. Sample  
807 annotation is presented on the left of each paired UMAP plot, each dot is an individual single  
808 cell, and cell number for each technology is presented in the lower-left hand corner. UMAP  
809 coordinate space remains the same for both scRNA and scDNAm visualizations with cellular  
810 states for that platform represented by a colored dot and data for the other platform represented  
811 by a gray dot. Stacked bar plots enumerating the proportion of cellular states detected by each  
812 platform are presented to the right of each paired UMAP plot. ``\*`` indicate specimens in which  
813 the cellular proportions across the two platforms are significantly different (Fisher's Exact test,  $p$   
814 < 0.05).  
815

816 **Figure S9. Sample-specific differences in DNA methylation and epimutation burden**  
817 **across different cellular states. Related to Figure 2.**

818 (A-B) Boxplots showing sample-specific differences in (A) epimutation burden and (B) 10-kb  
819 tiled DNA methylation across LIGER-defined cellular states in IDHmut tumors. Wilcoxon Rank  
820 Sum  $p$ -values are presented comparing cells from a given tumor. (C-D) Boxplots showing  
821 sample-specific differences in (C) epimutation burden and (D) 10-kb tiled DNA methylation  
822 across LIGER-defined cellular states in IDHwt tumors. Wilcoxon Rank Sum  $p$ -values are  
823 presented comparing cells from a given tumor. Samples with only one defined cell state are not  
824 visualized.

825  
826 **Figure S10. Relationships between epimutation burden and genetic alterations. Related**  
827 **to Figure 3.**

828 (A) Single-cell epimutation burden estimates were calculated across genomic regions with (teal)  
829 and without (gray) copy number alterations. The paired-sample Wilcoxon test  $p$ -value for each  
830 subtype represents the statistical difference of epimutation burden across these two regions. (B)  
831 Visualized results from multi-variable linear regression model testing for association with  
832 epimutation burden. Dot size indicates  $-\log_{10}(p\text{-value})$  for each predictor and color represents  
833 direction of association with epimutation burden (red = negative association, blue = positive  
834 association). Explanatory variables included subject age, timepoint (pre- and post-treatment),  
835 level of cellular proliferation determined by histological marker (MIB staining), and somatic copy  
836 number alteration burden (SCNA, total number of bases altered / total number of bases  
837 measured).

838  
839 **Figure S11. Whole genome sequencing phylogenetic inference of tumor samples.**  
840 **Related to Figure 4.**

841 (A-H) Phylogenetic trees constructed from whole genome sequencing data (mutations and  
842 somatic copy number alterations) using phyloWGS and further annotated using single-cell  
843 inferred copy number alterations (scRRBS + scRNAseq). Tree nodes represent alterations  
844 specific to the given clone, with node size corresponding to the fraction of cells with the  
845 associated alterations. Branch length scales with the number of mutations attributed to that  
846 clone. Clonal alterations are colored in blue, with subclonal alterations colored in gold. Genes  
847 considered significantly mutated in TCGA analyses (Ceccarelli et al., 2016) and chromosomal  
848 arm-level events are presented. Arm-level events are defined as spanning at least 80 percent of  
849 the chromosome arm, while partial events span at least 40 percent.

850  
851 **Figure S12. Tumor-specific comparisons of phylogenetic and phyloepigenetic trees.**  
852 **Related to Figure 4.**

853 (A-K) Tanglegrams highlighting the relationship between single-cell copy number and single-cell  
854 DNA methylation tree diagrams. Phylogenetic trees (left cluster) were calculated from copy  
855 number profiles (scRRBS data) and phyloepigenetic trees were constructed from the same cells  
856 across 10-kb tiled DNA methylation values. Cluster labels are connected with solid lines and are  
857 colored by cellular states determined by LIGER. Entanglement scores are listed above the  
858 phylogenetic and phyloepigenetic trees and indicate whether labels share the same structure  
859 (score = 0) or exhibit unrelated structures (score = 1).

860  
861 **Figure S13: Focal extrachromosomal DNA amplifications generate greater levels of**  
862 **epigenetic and transcript diversity in glioma single cells. Related to Figure 4.**

863 (A-D) Extrachromosomal DNA circular amplicon reconstruction displaying genomic  
864 rearrangements predicted from whole genome sequencing. Coverage depth is represented as a  
865 histogram across a genomic interval with segment copy number (CN) estimation provided on  
866 the right y-axis. Discordant read pair clusters are indicated by arcs and colors highlight read pair

867 orientation (e.g., brown = everted read pairs, (Deshpande et al., 2019). Amplicon intervals are  
868 provided at the bottom of the plot with annotation for known oncogenes (e.g., *EGFR*). (E) *EGFR*  
869 copy number estimation from single-cell RRBS data in ecDNA+ tumors. Cells with *EGFR* copy  
870 number greater than 7 were classified as *EGFR* ecDNA+ (blue). (F) Single-cell 10-kb tiled DNA  
871 methylation separated by *EGFR* ecDNA status. Single cells with inferred copy number status  
872 greater than 7 were classified as ecDNA+ (blue). Wilcoxon rank sum test  $p$ -values comparing  
873 DNA methylation across ecDNA status are reported for each patient tumor. (D) Boxplots  
874 depicting transcriptional diversity using gene count signatures calculated in scRNAseq data for  
875 each tumor, with cells separated based on inferred *EGFR* copy number status (gray = *EGFR*  
876 ecDNA-, blue = *EGFR* ecDNA+). Transcriptional diversity was compared based on predicted  
877 ecDNA status within each tumor subclone. Stars (\*) indicate statistically significant differences  
878 based on Wilcoxon Rank Sum test ( $p < 0.05$ ).

879  
880 **Figure S14. DNA methylation instability metrics calculated in primary tumor, spatial, and**  
881 **longitudinal cohorts. Related to Figure 5.**

882 (A) Scatterplot highlights the significant positive correlation between the single-cell epimutation  
883 burden metric and the bulk microarray-based DNA methylation instability metric ( $n = 11$  tumors,  
884 Spearman correlation). (B-C) Scatterplot between DNA methylation instability and (bulk  
885 RNAseq) ssGSEA enrichment scores for (B) oxidative stress response genes and (C) a  
886 randomly selected gene set substantiates finding that epigenetic instability is associated with  
887 stress response. (D) Schematic depiction of magnetic resonance image-guided biopsies and  
888 radiographic features used in spatial cohort (Verburg et al.). (E) Workflow for linear-mixed effect  
889 model identifying differentially methylated CpG sites that are selected for during tumor evolution  
890 when adjusting for estimated cellular proportions, glioma subtype, and a random effect for  
891 patient ( $n = 102$  tumor samples,  $n = 51$  patients).

892  
893

## 894 **METHODS:**

### 895 **EXPERIMENTAL METHODS**

#### 896 **Description of human tumor specimens.**

897 Human glioma resection specimens were obtained from the University of Connecticut Health  
898 Center and from St. Michael's Hospital. All tissue donations were approved by the Institutional  
899 Review Board of the Jackson Laboratory and clinical institutions involved. This work was  
900 performed in accordance with the Declaration of Helsinki principles. Initial pathological  
901 diagnosis was confirmed with tumor DNA methylation classification according to the Molecular  
902 Neuropathology Tool (Capper et al., 2018). Clinical characteristics for this population are  
903 provided in Table S1.

904

#### 905 **Sample preparation, partitioning, and fluorescence activated cell sorting for single-cell** 906 **experiments.**

907 Tumor specimens were collected directly from the operating room and immediately placed into  
908 MACS tissue storage solution at 4C (Miltenyi, Cat. no. 130-100-008). Tumor specimens from  
909 the same spatial region were then minced and partitioned into single-cell and bulk fractions

910 (Figure 1A). Any remaining tumor tissue was deposited into freezing media consisting of 90%  
911 heat-inactivated fetal bovine serum (FBS) (Invitrogen) and 10% dimethyl sulfoxide (Sigma-  
912 Aldrich), and gradually frozen in a freezing container (Mr. Frosty, Corning) over 24 hours before  
913 being stored in liquid nitrogen for future experiments (i.e., Fluorescence *in situ* hybridization).  
914 Bulk tissue specimens were immediately flash frozen for subsequent DNA and RNA extraction.  
915 The specimen fraction for single cell analyses was further mechanically and enzymatically  
916 dissociated using the Brain Tumor Dissociation Kit (P) according to the manufacturer's protocol  
917 (Miltenyi Cat. No. 130-095-942) and as previously reported (Neftel et al., 2019; Tirosh et al.,  
918 2016; Venteicher et al., 2017).

919 Single cell suspensions were blocked with human BD Fc Block (BioLegend) for 5 min on  
920 ice, prior to antibody staining, and labelled via incubation with 1:100 dilution of Alexa Fluor 488  
921 conjugated anti-CD45 antibody (Cat. no. 304017, BioLegend) and 1:100 dilution of PECy7-  
922 conjugated anti-CD31 antibody (Cat. no. 303117, BioLegend) for 30 minutes at 4C. Cells were  
923 washed with Hank's buffered saline solution and resuspended in 2mM EDTA/ 2% BSA/ PBS  
924 buffer containing [2µg/mL] propidium iodide (PI) (BD Biosciences, Cat. No. 556364) and [1µM]  
925 Calcein violet (Invitrogen) for 20 minutes at 4C. Fluorescence activated cell sorting (FACS) was  
926 performed using a BD FACSAria Fusion instrument with an 130µm nozzle and using the lowest  
927 event rate. Single cell mode was selected to further ensure stringency of sorting. For the  
928 generation of 10X sequencing libraries, 50,000-150,000 PI<sup>-</sup>, Calcein<sup>+</sup> viable single cells were  
929 collected in 20% FBS/HBSS buffer. CD45<sup>+</sup> cells were limited to no more than 20% of the total  
930 viable sort to enrich for tumor cells (Figure S2A). For the generation of single-cell DNA  
931 methylation libraries, we sorted viable (PI<sup>-</sup> and Calcein<sup>+</sup>), non-immune (CD45<sup>-</sup>), and non-  
932 endothelial (CD31<sup>-</sup>) cells into 96-well plates that were pre-loaded with 5 µL of 1X Tris-EDTA  
933 buffer (Figure S2A). Once the cells had been sorted, 96-well plates were either immediately  
934 processed through the single-cell DNA methylation protocol or flash frozen and stored at -80C.

935

### 936 **scRRBS library preparation.**

937 Single-cell DNA methylation profiling was performed using a modified version of a previous  
938 scRRBS protocol (Guo et al., 2015; Guo et al., 2013). Single-cell DNA methylation experiments  
939 were performed with sorted viable, non-immune, non-endothelial (PI<sup>-</sup>, Calcein<sup>+</sup>, CD45<sup>-</sup>, CD31<sup>-</sup>)  
940 cells in a 96-well plate containing 5 µL pre-loaded Tris-EDTA buffer with an empty well control.  
941 For 9 out of 11 tumors, the protocol was also applied to a small population control of 50-cells  
942 (PI<sup>-</sup>, Calcein<sup>+</sup>, CD45<sup>-</sup>, CD31<sup>-</sup>). Sorted 96-well plates were frozen at -80 C until processing when  
943 cells were lysed with 0.2 µL 1 M KCl (Millipore Sigma), 0.2µL 10% Triton X-100 (Millipore

944 Sigma), 0.3  $\mu$ L 20mg/mL protease (Qiagen), and nuclease-free water in a total volume of 6  $\mu$ L  
945 for 3 hours at 50 C. The protease was then heat-inactivated at 75 C for 15 minutes. The DNA  
946 was incubated with 50 units of MspI (NEB) and TaqI (NEB) with CutSmart buffer (NEB) for 3  
947 hours at 37 C. 60fg of unmethylated Lambda bacteriophage DNA (Promega) was added to each  
948 well to serve as a control for bisulfite conversion efficiency assessment. The solution was  
949 heated to 80 C for 20 minutes to heat-inactivate the restriction enzymes and placed on ice. 5  
950 units of Klenow Fragment (3'  $\rightarrow$  5' exo-, NEB), CutSmart buffer (NEB), and end-repair dNTP mix  
951 (40uM dATP, 4uM dGTP, and 4uM dCTP; NEB) totaling 2  $\mu$ L per reaction were added to  
952 perform end-repair and dA-tailing. 1:250X diluted NEXTflex methylated adapters (BiooScientific)  
953 were added to each quadrant of the 96-well plate ( $n = 24$  unique adapters) with a ligation  
954 mixture of 40 Weiss U T4 ligase (NEB), 1mM ATP (ThermoFisher Scientific), and nuclease-free  
955 water to a final volume of 4  $\mu$ L per reaction. TruSeq methylated adapters (Illumina) were also  
956 used in a single sample (SM001) using the same protocol. The ligation reaction proceeded at 16  
957 C for 30 minutes followed by an incubation of 4 C for at least 8 hours. The ligation reaction was  
958 stopped by heat-inactivation at 65 C for 20 minutes. Post-adapter ligation, 24 individual cells  
959 with unique ligated adapters were pooled from each plate quadrant for the protocol's remainder.  
960 Excess adapter was removed using a 1:1 volumetric ratio of Ampure beads (Beckman Coulter).  
961 Bisulfite conversion was performed using the EZ-DNA methylation kit (Zymo) according to the  
962 manufacturer's instructions except with one-half volumes due to reduced DNA input. The  
963 solution was incubated at 98 C for 8 minutes, 64 C for 3.5 hours, and held at 4 C once the  
964 reaction was complete. 10ng of tRNA (Roche) was added prior to column elution to serve as a  
965 protective carrier. PCR enrichment was performed using the PfuTurbo Cx hotstart (Agilent),  
966 PfuTurbo Cx hotstart buffer (Agilent), primer mix (Bioo Scientific), dNTP mix (Promega), and  
967 nuclease-free water under the following conditions: 95 degrees Celsius for 2 minutes, 32 cycles  
968 of 95 C for 20 seconds, 60 C for 30 seconds, and 72 C for 60 seconds. The PCR reaction was  
969 terminated by incubating at 72 C for 5 minutes. The libraries were purified in a 1:1 volumetric  
970 ratio of Ampure beads (Beckman Coulter), Pippin size selection was performed between 200-  
971 1000bp (Sage Science), and quantified by qPCR (Kapa Biosystems / Roche). scRRBS libraries  
972 were paired-end sequenced alongside bulk whole genome libraries on an Illumina HiSeq4000  
973 using 1% PhiX spike-in and 75bp reads.

974

#### 975 **Single-cell RNA library preparation.**

976 Sorted cells were washed and resuspended in 0.04% BSA/PBS buffer. Cells were counted on a  
977 Countess II automated cell counter and were loaded on a 10X Chromium chip with a target cell

978 recovery of 6,000 cells per lane. Sequencing libraries were performed using the single-cell 3'  
979 mRNA kit according to the manufacturer's protocol (10X Genomics). cDNA and library quality  
980 were examined on an Agilent 4200 TapeStation and quantified by qPCR (Kapa Biosystems /  
981 Roche). Illumina sequencing was performed using a paired-end 100bp protocol. Libraries were  
982 sequenced to an average depth of 50,000 unique reads per cell.

983

#### 984 **Whole genome sequencing of tumors and matched normal blood.**

985 Genomic DNA was extracted from the same tumor region as the single-cell analyses using the  
986 Qiagen AllPrep kit and matched normal blood using DNeasy kit (Qiagen). Briefly, 400ng of DNA  
987 was sheared to 400bp using a LE220 focused-ultrasonicator (Covaris) and size selected using  
988 SPRI beads (Beckman Coulter). The fragments were treated with end-repair, A-tailing, and  
989 ligation of unique adapters (Illumina) using the KAPA HyperPrep Kit (Roche). This was followed  
990 by 5 cycles of PCR amplification when necessary. DNA sequencing was performed using  
991 paired-end 75bp protocol according to the manufacturer's protocol (Illumina). The tumor  
992 samples were sequenced to an average depth of 44X and tumor-matched normal blood to 30X.

993

#### 994 **Bulk Illumina EPIC DNA methylation microarrays.**

995 250 ng of genomic tumor DNA was subject to bisulfite conversion using the EZ DNA Methylation  
996 kit (Zymo) and genome-wide DNA methylation was assessed by the Infinium MethylationEPIC  
997 kit according to the manufacturer's protocol (Illumina).

998

#### 999 **Bulk RNA sequencing.**

1000 Bulk tumor RNA was extracted from samples with sufficient tissue using the AllPrep kit  
1001 (Qiagen). Samples with RIN values > 5 as assessed by TapeStation (Agilent Technologies)  
1002 were prepared with KAPA mRNA HyperPrep kit (Roche). Libraries were sequenced using a  
1003 paired-end 150bp protocol on a NovaSeq to 50 million reads according to the manufacturer's  
1004 protocol (Illumina).

1005

#### 1006 **Fluorescent in situ hybridization (FISH) analysis.**

1007 Tissue slides were prepared by tumor touch prep method (deCarvalho et al., 2018). Positively  
1008 charged glass slides were pressed against the surface of slightly thawed tissues. The slides  
1009 were then immediately fixed by cold Carnoy's fixative (3:1 methanol:glacial acetic acid, v/v) for  
1010 30 min and then air-dried. Slides were denatured in Hybridization buffer (Empire Genomics)  
1011 mixed with EGFR-Chr7 probe (EGFR-CHR07-20-ORGR, Empire Genomics) at 75°C for 5 min

1012 and then incubated at 37°C overnight. The posthybridization wash was with 0.4x SSC at 75°C  
1013 for 3 min followed by a second wash with 2x SSC/0.05% Tween20 for 1 min. The slides were  
1014 then briefly rinsed by water and air dried. The VECTASHIELD mounting medium with DAPI  
1015 (Vector Laboratories) was applied and the coverslip was mounted onto a glass slide. Tissue  
1016 images were scanned under Leica STED 3X/DLS Confocal with 100x magnification.

1017

## 1018 **ANALYTICAL METHODS.**

### 1019 **Single-cell DNA methylation processing.**

1020 Raw sequencing reads were trimmed to remove adapters and low-quality bases using  
1021 trim\_galore with the `-- rrbs` and `-- paired` parameters (version 0.4.0  
1022 <https://github.com/FelixKrueger/TrimGalore>). The trimmed reads were then aligned to the  
1023 GRCh37 (hg19) genome using Bismark (version 0.19.1) with parameters ` -N 1 -- bowtie2 --  
1024 score\_min L,0,-0.4` (Krueger and Andrews, 2011). PCR duplicates were removed with the  
1025 `deduplicate\_bismark` command. Bisulfite conversion efficiency was determined using the  
1026 spike-in unmethylated lambda DNA. Cells with fewer than 40,000 unique CpGs detected and  
1027 bisulfite conversion rates below 95% were removed from analysis. 914 single cells were  
1028 retained for downstream analysis ( $n = 914 / 1,076$  total cells sequenced) with a mean of  
1029 145,000 CpGs per cell and mean bisulfite conversion rate of 98.4% (Table S4).

1030

### 1031 **Unsupervised clustering of scRRBS data.**

1032 Unsupervised clustering of the DNA methylation data was performed using pairwise  
1033 comparisons of individual CpGs across all cell-to-cell comparisons as previously described  
1034 (PDclust) (Hui et al., 2018). Briefly, this method performs pairwise comparisons of single-CpG  
1035 methylation measurements to create a pairwise dissimilarity (PD) value that reflects the average  
1036 absolute difference in methylation values at CpGs covered in any two cells. The pairwise  
1037 dissimilarity values were used as input features for the Multidimensional Scaling (MDS) analysis  
1038 for which visualization of cells in close proximity reflects greater similarity than cells further apart  
1039 (Figure 1B).

1040

### 1041 **Epimutation burden as a measure of epigenome instability.**

1042 Epimutation burden was determined by identifying DNA methylation concordance of nearby  
1043 CpGs on a single sequencing read as previously described for bulk and single-cell DNA  
1044 methylation data (Gaiti et al., 2019; Landan et al., 2012; Landau et al., 2014). Briefly, in order for  
1045 a sequencing read to be considered for this analysis it required a minimum of 4 CpGs located

1046 on the same sequencing read. These sequencing reads are referred to as “epialleles” and  
1047 represent a subset of a cell’s total sequencing reads. An epiallele read is considered discordant  
1048 if any of the CpGs on that sequencing read have different methylation states (e.g., three  
1049 methylated CpGs and an unmethylated CpG). The epimutation burden metric reflects the sum  
1050 of discordant epialleles divided by the total number of epialleles considered for analysis as  
1051 previously described (Gaiti et al., 2019; Landan et al., 2012; Landau et al., 2014). The  
1052 epimutation burden metric can be calculated across the entire genome (i.e., “epimutation  
1053 burden”) or restricted to specific genomic regions where the metric considers only the epialleles  
1054 overlapping that particular genomic context. A linear regression model was used to assess the  
1055 impact of the total number of epialleles considered for analysis on the epimutation burden. The  
1056 epimutation metric was very weakly associated with epiallele count in that an additional 10,000  
1057 epialleles was associated with an 0.001 increase in the epimutation burden metric. For analyses  
1058 associating epimutation burden with metrics derived from bulk WGS data, sample-level  
1059 epimutation burden was calculated as the median of single-cell epimutation values.

1060

#### 1061 **DNA methylation and epimutation over genomic annotations.**

1062 To determine region-specific DNA methylation or epimutation burden, each cell’s measured  
1063 CpGs or epialleles were intersected with the genomic coordinates of interest before methylation  
1064 value or epimutation burden calculation, respectively. All coordinates were mapped against the  
1065 hg19 human genome assembly. Regions of interest considered for analyses included promoter,  
1066 gene body, intergenic, and DNaseI regions, TF binding sites, replication timing domains, and  
1067 5kb and 10kb tiled regions. Promoters were defined as 1kb upstream and 500 bp downstream  
1068 of FANTOM5 (Forrest et al., 2014) TSS that mapped to Ensembl genes. If multiple TSSs  
1069 mapped to a given gene, the TSS with the lowest genomic coordinate was selected. Gene body  
1070 annotations were obtained from the Ensembl Genome Browser (Hunt et al., 2018). Intergenic  
1071 regions were annotated by selecting regions not overlapping Ensembl gene body coordinates.  
1072 DNaseI hypersensitivity region annotations were obtained from the UCSC Genome Browser  
1073 (Raney et al., 2013). TF binding sites were obtained from the JASPAR 2020 Core Vertebrate  
1074 database (Fornes et al., 2020) of non-redundant TF binding motifs. Each binding site is  
1075 assigned a score of 0-1000, which corresponds to the p-value for the relative position weight  
1076 matrix score of a TF binding site prediction. For a given TF, all identified target binding site  
1077 coordinates were aggregated, and binding sites were excluded if they had a relative score less  
1078 than 400, corresponding to a p-value greater than 0.0001. Replication timing of genes was  
1079 retrieved from MutSigCV (Lawrence et al., 2013), and annotations for replication timing domains

1080 were generated by binning gene coordinates into quartiles based on the replication timing score.  
1081 Methylation values were also calculated for non-overlapping windows of 5kb or 10kb. Ranks of  
1082 high epimutation levels were determined by applying the ROSE software  
1083 ([https://bitbucket.org/young\\_computation/rose](https://bitbucket.org/young_computation/rose)) for both gene-level and transcription factor  
1084 binding sites.

1085

### 1086 **SCNA estimation from single-cell DNA methylation data.**

1087 To provide evidence for somatic copy number alterations in single-cell DNA methylation  
1088 sequencing data, the Gingko algorithm (Garvin et al., 2015) was applied to single cells that  
1089 passed the scRRBS quality control filters mentioned above. Briefly, this method bins mapped  
1090 reads by chromosomal location, performs Lowess normalization to correct for GC biases,  
1091 adjusts for potential amplification artifacts, and segments the genome to determine  
1092 chromosomal regions with consistent copy number states. Here, the genome for each sample  
1093 was divided into 2,597 variable-length bins with a median length of 1Mb. Segmentation was  
1094 performed using independent normalized read counts and the parameter `mask bad bins` (i.e.,  
1095 bins with consistent pileups) was enabled. Cells were considered “non-tumor” if less than 1% of  
1096 the genome had a copy number state that was not 2. Copy number plots were generated using  
1097 the R package “gplots”. Phylogenetic and phyloepigenetic trees were constructed for the same  
1098 cells (scRRBS data) using Euclidean distance between profiles and clustered with the R  
1099 function hclust using “ward.D2” linkage. The concordance between these two trees for each  
1100 sample was determined using the tanglegram function in the dendextend R package and 10  
1101 random tree rotations were used to minimize artificial branch crossing (Galili, 2015).

1102

### 1103 **Single-cell RNA processing and analysis.**

1104 The Cell Ranger pipeline (v3.0.2) was used to convert Illumina base call files to FASTQ files  
1105 and align FASTQs to hg19 10X reference genome (version 1.2.0). Preprocessing was  
1106 performed using the Scanpy package (1.3.7) (Wolf et al., 2018). The gene expression profiles of  
1107 each cell at the 1500 most highly variable genes (as measured by dispersion (Satija et al.,  
1108 2015)) were used for neighborhood graph generation (using 33 nearest neighbors) and  
1109 dimensionality reduction with UMAP (Becht et al., 2018). Clustering was performed on this  
1110 neighborhood graph using the Leiden community detection algorithm (Traag et al., 2019). The  
1111 neighborhood graph was batch-corrected using the batch correction software BBKNN (Polanski  
1112 et al., 2020). These defined clusters were then labelled with particular cell states based on  
1113 marker gene expression and previously described cell states (Bhaduri et al., 2020; Neftel et al.,

1114 2019; Tirosh et al., 2016). Cell state classification of malignant cells was also performed using  
1115 previously developed classifiers for both IDH-wild-type (Neftel et al., 2019) and IDH-mutant  
1116 tumors (Venteicher et al., 2017). The Seurat R package was also used for downstream  
1117 analyses and visualizations (Stuart et al., 2019). Inference of gene regulatory networks was  
1118 performed using SCENIC for a random set of 5,000 cells per subtype to permit heatmap  
1119 visualization (Aibar et al., 2017). SCNA estimation from single-cell RNAseq data was performed  
1120 as previously reported (Neftel et al., 2019; Tirosh et al., 2016; Venteicher et al., 2017). Briefly,  
1121 the InferCNV method provides evidence for large-scale somatic copy number alterations by  
1122 comparing averaged gene expression intensity values compared with normal cells (based on  
1123 marker gene expression) from the same specimen. Subclusters of cells were partitioned into  
1124 clones on the basis of shared copy number patterns  
1125 (<https://github.com/broadinstitute/inferCNV>). Single-cell RNA diversity comparisons using gene  
1126 count signatures were performed using the R package CytoTRACE across cells from the same  
1127 tumor clone (Gulati et al., 2020).

1128

#### 1129 **Joint scRNA and scDNA methylation integration.**

1130 Single-cell RNA and DNA methylation malignant cell profiles were integrated within the same  
1131 specimen based on the differentially expressed across the pan-glioma RNA cell states (Table  
1132 S2). The single-cell RNA data were jointly clustered with the gene-level methylation features as  
1133 previously reported (Welch et al., 2019) using the R package liger (linked inference of genomic  
1134 experimental relationships).

1135

#### 1136 **Analysis of publicly available brain tumor DNA methylation data.**

1137 Data re-analysis of longitudinal glioma resources was accessed for Klughammer et al.  
1138 (<http://www.medical-epigenomics.org/papers/GBMatch/>) (Klughammer et al., 2018) and the  
1139 Glioma Longitudinal AnalySiS consortium (GLASS, <http://synapse.org/glass>) (Barthel et al.,  
1140 2019). Magnetic Resonance Imaging guided biopsies taken from spatially distinct regions and  
1141 subjected to bulk DNA methylation Illumina methylation microarray collected by our group was  
1142 also accessed (Verburg et al.). DNA methylation microarrays (450K) were also retrieved The  
1143 Cancer Genome Atlas initial glioma samples (Ceccarelli et al., 2016). All Illumina methylation  
1144 microarrays were processed using the R package minfi. The recurrent DNA methylation  
1145 changes between the initial and recurrent tumors were determined by fitting a linear mixed  
1146 effect model (R nlme package) to each individual CpG modeled as a logit transformed M-value  
1147 with independent variables of timepoint, subtype, cancer cell proportion, immune proportion,

1148 and subject included as the random effect. The cancer and immune cell proportions in the  
1149 GLASS bulk Illumina methylation microarray data were determined using the glioma signature in  
1150 the R package MethyLICIBERSORT as previously described (Chakravarthy et al., 2018).

1151

### 1152 **Gene and genomic region enrichment analyses.**

1153 Enrichment of genes were performed using the R package topGO. Enrichment of genomic  
1154 regions were determined using the Locus Overlap Analysis (LOLA) R package (Sheffield and  
1155 Bock, 2016). LOLA enrichment analyses used all features considered for analysis as the  
1156 background sets.

1157

### 1158 **Variant detection and copy number calling.**

1159 Variant detection and bulk copy number determination was performed in accordance to the  
1160 GATK Best practices using GATK 4.1.0.0 (Mutect2) and as previously described (Barthel et al.,  
1161 2019).

1162

### 1163 **Mutational signature identification**

1164 Mutational signatures were identified in bulk WGS samples using the MutationalPatterns R  
1165 package (Blokzijl et al., 2018). The trinucleotide context of single base substitutions was  
1166 extracted for each sample in order to construct a mutational profile. For each mutational profile,  
1167 the contribution of mutational signatures from the Catalogue of Somatic Mutations in Cancer  
1168 (COSMIC v3) was quantified. Known signatures were ranked by order of relative contribution to  
1169 the sample mutational profile; for visualization the top 5 signatures per sample were listed, with  
1170 the remaining signatures collapsed into an “Other” category.

1171

### 1172 **Phylogenetic reconstruction copy number / mutation clonality.**

1173 To reconstruct the evolutionary history and subclonal composition of tumors, PhyloWGS  
1174 (Deshwar et al., 2015) was applied to bulk WGS samples. PhyloWGS incorporates SCNAs  
1175 with simple somatic mutations (SSMs) in inferred phylogenies by converting SCNAs into pseudo  
1176 SSMs prior to subclonal reconstruction. For input, phyloWGS requires VCF format variant calls,  
1177 SCNA segments, and estimates of tumor purity, which were generated using Mutect2 (Cibulskis  
1178 et al., 2013), TITAN (Ha et al., 2014), and Sequenza (Favero et al., 2015), respectively. If a  
1179 tumor contained more than 5000 variants, input variants were subsampled to 5000, ensuring all  
1180 variants overlapping previously identified significantly mutated genes were included (Barthel et  
1181 al., 2019; Ceccarelli et al., 2016). For each phyloWGS run, multiple Markov chain Monte Carlo

1182 chains were initiated with differing start values, and the optimum solution was selected based on  
1183 negative normalized log likelihood. Cancer cell fractions (CCF) were calculated for each tumor  
1184 subpopulation as the cellular prevalence for a given subpopulation divided by the maximum  
1185 cellular prevalence for that tumor, which corresponds to the estimated tumor purity. Events were  
1186 defined as clonal if they have a CCF of 1 or subclonal otherwise. SCNA subpopulation  
1187 assignments and cellular prevalence estimates derived from phyloWGS were further informed  
1188 by scRNAseq and scRRBS-derived copy number profiles.

1189

### 1190 **Bulk RNA sequencing processing.**

1191 FASTQ files were pre-processed with fastp v0.20.0 to assess quality control and were aligned to  
1192 the hg19 genome using kallisto v0.46.0 with default parameters (Bray et al., 2016). The bulk  
1193 RNA Verhaak classification and simplicity scores were determined as previously reported  
1194 (Wang et al., 2017). Single sample gene set enrichment analysis for particular pathways was  
1195 performed using the GVSAs R package (Hanzelmann et al., 2013).

1196

### 1197 **Detection of extrachromosomal DNA.**

1198 Amplicon architect was used to detect extrachromosomal DNA in tumor whole genome  
1199 sequencing data as previously described (Deshpande et al., 2019). Briefly, this method  
1200 characterizes the architecture of amplified regions that are larger than 10kb and have more than  
1201 four copies greater than the median sample ploidy.

1202

### 1203 **DNA methylation-based tumor classification.**

1204 Probabilistic estimates of tumor classification were defined both by the Molecular  
1205 Neuropathology classification tool (version 11b4) as previously reported (Capper et al., 2018).

1206

### 1207 **Statistical methods.**

1208 All data analyses were conducted in R 3.6.1. Statistical analyses are described in the respective  
1209 Methods subsection and briefly described in the figure legends. No statistical methods were  
1210 used to predetermine study sample size.  $p$ -values of  $< 0.05$  were considered statistically  
1211 significant.

1212

### 1213 **REFERENCES:**

1214 Aibar, S., Gonzalez-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G.,  
1215 Rambow, F., Marine, J.C., Geurts, P., Aerts, J., *et al.* (2017). SCENIC: single-cell regulatory  
1216 network inference and clustering. *Nat Methods* 14, 1083-1086.

- 1217  
1218 Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y., Boot, A.,  
1219 Covington, K.R., Gordenin, D.A., Bergstrom, E.N., *et al.* (2020). The repertoire of mutational  
1220 signatures in human cancer. *Nature* 578, 94-101.  
1221  
1222 Angermueller, C., Clark, S.J., Lee, H.J., Macaulay, I.C., Teng, M.J., Hu, T.X., Krueger, F.,  
1223 Smallwood, S., Ponting, C.P., Voet, T., *et al.* (2016). Parallel single-cell sequencing links  
1224 transcriptional and epigenetic heterogeneity. *Nat Methods* 13, 229-232.  
1225  
1226 Argelaguet, R., Clark, S.J., Mohammed, H., Stapel, L.C., Krueger, C., Kapourani, C.A., Imaz-  
1227 Rosshandler, I., Lohoff, T., Xiang, Y., Hanna, C.W., *et al.* (2019). Multi-omics profiling of mouse  
1228 gastrulation at single-cell resolution. *Nature* 576, 487-491.  
1229  
1230 Barthel, F.P., Johnson, K.C., Varn, F.S., Moskalik, A.D., Tanner, G., Kocakavuk, E., Anderson,  
1231 K.J., Abiola, O., Aldape, K., Alfaro, K.D., *et al.* (2019). Longitudinal molecular trajectories of  
1232 diffuse glioma in adults. *Nature*.  
1233  
1234 Becht, E., McInnes, L., Healy, J., Dutertre, C.A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and  
1235 Newell, E.W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. *Nat*  
1236 *Biotechnol.*  
1237  
1238 Bhaduri, A., Di Lullo, E., Jung, D., Muller, S., Crouch, E.E., Espinosa, C.S., Ozawa, T.,  
1239 Alvarado, B., Spatazza, J., Cadwell, C.R., *et al.* (2020). Outer Radial Glia-like Cancer Stem  
1240 Cells Contribute to Heterogeneity of Glioblastoma. *Cell Stem Cell* 26, 48-63 e46.  
1241  
1242 Bian, S., Hou, Y., Zhou, X., Li, X., Yong, J., Wang, Y., Wang, W., Yan, J., Hu, B., Guo, H., *et al.*  
1243 (2018). Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science*  
1244 (New York, NY) 362, 1060-1063.  
1245  
1246 Blokzijl, F., Janssen, R., van Boxtel, R., and Cuppen, E. (2018). MutationalPatterns:  
1247 comprehensive genome-wide analysis of mutational processes. *Genome Med* 10, 33-33.  
1248 Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-  
1249 seq quantification. *Nat Biotechnol* 34, 525-527.  
1250  
1251 Cancer Genome Atlas Research, N., Brat, D.J., Verhaak, R.G., Aldape, K.D., Yung, W.K.,  
1252 Salama, S.R., Cooper, L.A., Rheinbay, E., Miller, C.R., Vitucci, M., *et al.* (2015).  
1253 Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N Engl J Med*  
1254 372, 2481-2498.  
1255  
1256 Capper, D., Jones, D.T.W., Sill, M., Hovestadt, V., Schrimpf, D., Sturm, D., Koelsche, C., Sahm,  
1257 F., Chavez, L., Reuss, D.E., *et al.* (2018). DNA methylation-based classification of central  
1258 nervous system tumours. *Nature* 555, 469-474.  
1259  
1260 Ceccarelli, M., Barthel, F.P., Malta, T.M., Sabedot, T.S., Salama, S.R., Murray, B.A., Morozova,  
1261 O., Newton, Y., Radenbaugh, A., Pagnotta, S.M., *et al.* (2016). Molecular Profiling Reveals  
1262 Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell* 164, 550-  
1263 563.  
1264  
1265 Chakravarthy, A., Furness, A., Joshi, K., Ghorani, E., Ford, K., Ward, M.J., King, E.V., Lechner,  
1266 M., Marafioti, T., Quezada, S.A., *et al.* (2018). Pan-cancer deconvolution of tumour composition  
1267 using DNA methylation. *Nat Commun* 9, 3220.

1268  
1269 Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S.,  
1270 Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations  
1271 in impure and heterogeneous cancer samples. *Nature Biotechnology* 31, 213-219.  
1272  
1273 Dang, L., White, D.W., Gross, S., Bennett, B.D., Bittinger, M.A., Driggers, E.M., Fantin, V.R.,  
1274 Jang, H.G., Jin, S., Keenan, M.C., *et al.* (2009). Cancer-associated IDH1 mutations produce 2-  
1275 hydroxyglutarate. *Nature* 462, 739-744.  
1276  
1277 de Souza, C.F., Sabedot, T.S., Malta, T.M., Stetson, L., Morozova, O., Sokolov, A., Laird, P.W.,  
1278 Wiznerowicz, M., Iavarone, A., Snyder, J., *et al.* (2018). A Distinct DNA Methylation Shift in a  
1279 Subset of Glioma CpG Island Methylator Phenotypes during Tumor Recurrence. *Cell Rep* 23,  
1280 637-651.  
1281  
1282 deCarvalho, A.C., Kim, H., Poisson, L.M., Winn, M.E., Mueller, C., Cherba, D., Koeman, J.,  
1283 Seth, S., Protopopov, A., Felicella, M., *et al.* (2018). Discordant inheritance of chromosomal and  
1284 extrachromosomal DNA elements contributes to dynamic disease evolution in glioblastoma. *Nat*  
1285 *Genet* 50, 708-717.  
1286  
1287 Deshpande, V., Luebeck, J., Nguyen, N.D., Bakhtiari, M., Turner, K.M., Schwab, R., Carter, H.,  
1288 Mischel, P.S., and Bafna, V. (2019). Exploring the landscape of focal amplifications in cancer  
1289 using AmpliconArchitect. *Nat Commun* 10, 392.  
1290  
1291 Deshwar, A.G., Vembu, S., Yung, C.K., Jang, G.H., Stein, L., and Morris, Q. (2015). PhyloWGS:  
1292 Reconstructing subclonal composition and evolution from whole-genome sequencing of tumors.  
1293 *Genome Biology* 16, 35.  
1294  
1295 Easwaran, H., Tsai, H.C., and Baylin, S.B. (2014). Cancer epigenetics: tumor heterogeneity,  
1296 plasticity of stem-like states, and drug resistance. *Mol Cell* 54, 716-727.  
1297  
1298 Farlik, M., Halbritter, F., Muller, F., Choudry, F.A., Ebert, P., Klughammer, J., Farrow, S.,  
1299 Santoro, A., Ciaurro, V., Mathur, A., *et al.* (2016). DNA Methylation Dynamics of Human  
1300 Hematopoietic Stem Cell Differentiation. *Cell Stem Cell* 19, 808-822.  
1301  
1302 Favero, F., Joshi, T., Marquard, A.M., Birnbak, N.J., Krzystanek, M., Li, Q., Szallasi, Z., and  
1303 Eklund, A.C. (2015). Sequenza: allele-specific copy number and mutation profiles from tumor  
1304 sequencing data. *Ann Oncol* 26, 64-70.  
1305  
1306 Flavahan, W.A., Gaskell, E., and Bernstein, B.E. (2017). Epigenetic plasticity and the hallmarks  
1307 of cancer. *Science (New York, NY)* 357, eaal2380.  
1308  
1309 Fornes, O., Castro-Mondragon, J.A., Khan, A., van der Lee, R., Zhang, X., Richmond, P.A.,  
1310 Modi, B.P., Correard, S., Gheorghe, M., Baranasic, D., *et al.* (2020). JASPAR 2020: update of  
1311 the open-access database of transcription factor binding profiles. *Nucleic Acids Res* 48, D87-  
1312 D92.  
1313  
1314 Forrest, A.R.R., Kawaji, H., Rehli, M., Kenneth Baillie, J., de Hoon, M.J.L., Haberle, V.,  
1315 Lassmann, T., Kulakovskiy, I.V., Lizio, M., Itoh, M., *et al.* (2014). A promoter-level mammalian  
1316 expression atlas. *Nature* 507, 462-470.  
1317

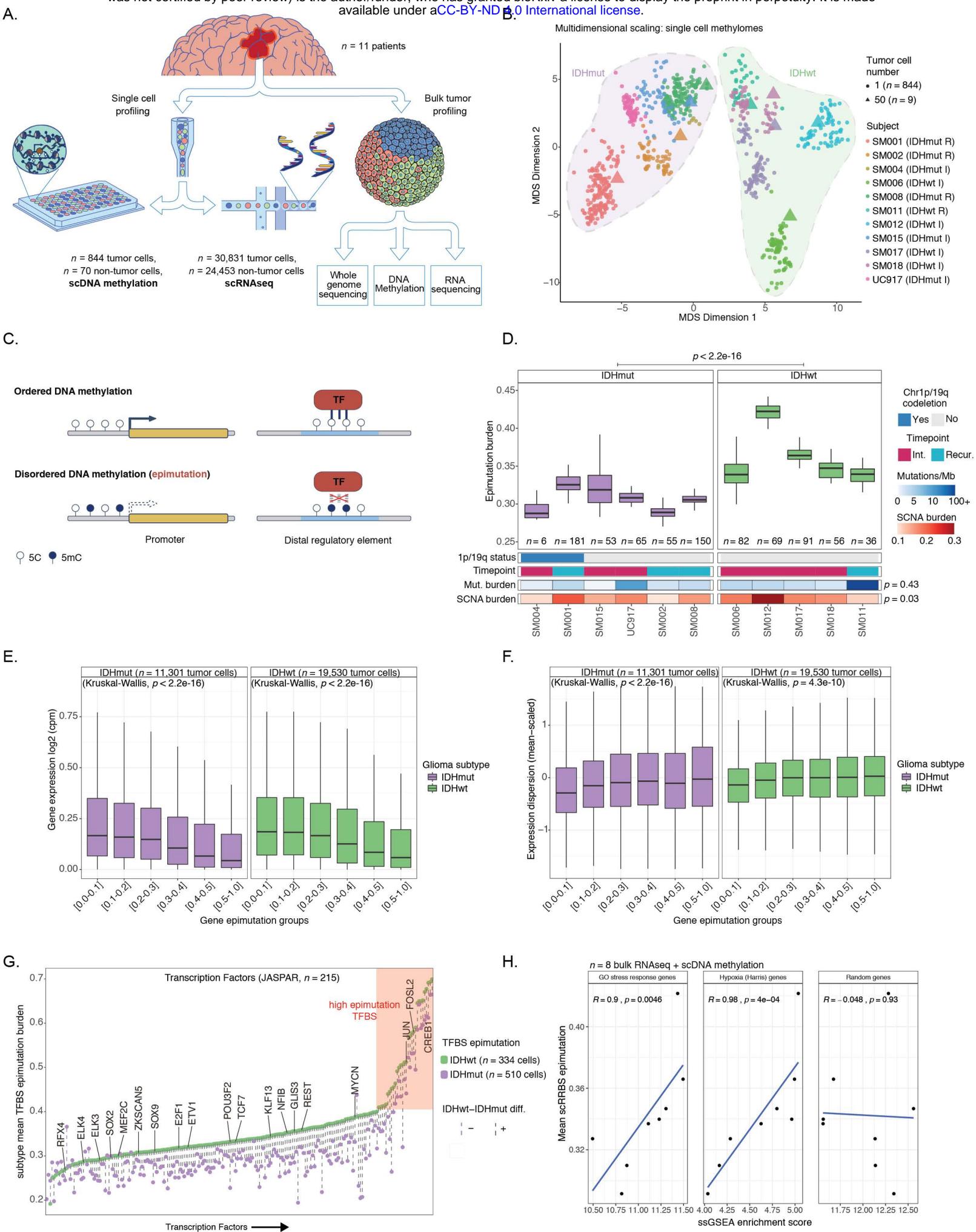
- 1318 Gaiti, F., Chaligne, R., Gu, H., Brand, R.M., Kothen-Hill, S., Schulman, R.C., Grigorev, K.,  
1319 Risso, D., Kim, K.T., Pastore, A., *et al.* (2019). Epigenetic evolution and lineage histories of  
1320 chronic lymphocytic leukaemia. *Nature* *569*, 576-580.
- 1321  
1322 Galili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of  
1323 hierarchical clustering. *Bioinformatics* *31*, 3718-3720.
- 1324  
1325 Gao, R., Davis, A., McDonald, T.O., Sei, E., Shi, X., Wang, Y., Tsai, P.C., Casasent, A., Waters,  
1326 J., Zhang, H., *et al.* (2016). Punctuated copy number evolution and clonal stasis in triple-  
1327 negative breast cancer. *Nat Genet* *48*, 1119-1130.
- 1328  
1329 Garvin, T., Aboukhalil, R., Kendall, J., Baslan, T., Atwal, G.S., Hicks, J., Wigler, M., and Schatz,  
1330 M.C. (2015). Interactive analysis and assessment of single-cell copy-number variations. *Nat*  
1331 *Methods* *12*, 1058-1060.
- 1332  
1333 Gerstung, M., Jolly, C., Leshchiner, I., Dentre, S.C., Gonzalez, S., Rosebrock, D., Mitchell, T.J.,  
1334 Rubanova, Y., Anur, P., Yu, K., *et al.* (2020). The evolutionary history of 2,658 cancers. *Nature*  
1335 *578*, 122-128.
- 1336  
1337 Gulati, G.S., Sikandar, S.S., Wesche, D.J., Manjunath, A., Bharadwaj, A., Berger, M.J., Ilagan,  
1338 F., Kuo, A.H., Hsieh, R.W., Cai, S., *et al.* (2020). Single-cell transcriptional diversity is a  
1339 hallmark of developmental potential. *Science (New York, NY)* *367*, 405-411.
- 1340  
1341 Guo, H., Zhu, P., Guo, F., Li, X., Wu, X., Fan, X., Wen, L., and Tang, F. (2015). Profiling DNA  
1342 methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite  
1343 sequencing. *Nat Protoc* *10*, 645-659.
- 1344  
1345 Guo, H., Zhu, P., Wu, X., Li, X., Wen, L., and Tang, F. (2013). Single-cell methylome  
1346 landscapes of mouse embryonic stem cells and early embryos analyzed using reduced  
1347 representation bisulfite sequencing. *Genome Res* *23*, 2126-2135.
- 1348  
1349 Ha, G., Roth, A., Khattra, J., Ho, J., Yap, D., Prentice, L.M., Melnyk, N., McPherson, A.,  
1350 Bashashati, A., Laks, E., *et al.* (2014). TITAN: inference of copy number architectures in clonal  
1351 cell populations from tumor whole-genome sequence data. *Genome Res* *24*, 1881-1893.
- 1352  
1353 Hanzelmann, S., Castelo, R., and Guinney, J. (2013). GSV: gene set variation analysis for  
1354 microarray and RNA-seq data. *BMC Bioinformatics* *14*, 7.
- 1355  
1356 Heddlestone, J.M., Li, Z., Lathia, J.D., Bao, S., Hjelmeland, A.B., and Rich, J.N. (2010). Hypoxia  
1357 inducible factors in cancer stem cells. *Br J Cancer* *102*, 789-795.
- 1358  
1359 Hui, T., Cao, Q., Wegrzyn-Woltosz, J., O'Neill, K., Hammond, C.A., Knapp, D., Laks, E., Moksa,  
1360 M., Aparicio, S., Eaves, C.J., *et al.* (2018). High-Resolution Single-Cell DNA Methylation  
1361 Measurements Reveal Epigenetically Distinct Hematopoietic Stem Cell Subpopulations. *Stem*  
1362 *Cell Reports* *11*, 578-592.
- 1363  
1364 Hunt, S.E., McLaren, W., Gil, L., Thormann, A., Schuilenburg, H., Sheppard, D., Parton, A.,  
1365 Armean, I.M., Trevanion, S.J., Flicek, P., *et al.* (2018). Ensembl variation resources. *Database*  
1366 *2018*.
- 1367

- 1368 Jin, X., Kim, L.J.Y., Wu, Q., Wallace, L.C., Prager, B.C., Sanvoranart, T., Gimple, R.C., Wang,  
1369 X., Mack, S.C., Miller, T.E., *et al.* (2017). Targeting glioma stem cells through combined BMI1  
1370 and EZH2 inhibition. *Nat Med* 23, 1352-1361.  
1371  
1372 Kelsey, G., Stegle, O., and Reik, W. (2017). Single-cell epigenomics: Recording the past and  
1373 predicting the future. *Science (New York, NY)* 358, 69-75.  
1374  
1375 Kim, H., Zheng, S., Amini, S.S., Virk, S.M., Mikkelsen, T., Brat, D.J., Grimsby, J., Sougnez, C.,  
1376 Muller, F., Hu, J., *et al.* (2015). Whole-genome and multisector exome sequencing of primary  
1377 and post-treatment glioblastoma reveals patterns of tumor evolution. *Genome Res* 25, 316-327.  
1378  
1379 Klughammer, J., Kiesel, B., Roetzer, T., Fortelny, N., Nemc, A., Nenning, K.H., Furtner, J.,  
1380 Sheffield, N.C., Datlinger, P., Peter, N., *et al.* (2018). The DNA methylation landscape of  
1381 glioblastoma disease progression shows extensive heterogeneity in time and space. *Nat Med*  
1382 24, 1611-1624.  
1383  
1384 Korber, V., Yang, J., Barah, P., Wu, Y., Stichel, D., Gu, Z., Fletcher, M.N.C., Jones, D.,  
1385 Hentschel, B., Lamszus, K., *et al.* (2019). Evolutionary Trajectories of IDH(WT) Glioblastomas  
1386 Reveal a Common Path of Early Tumorigenesis Instigated Years ahead of Initial Diagnosis.  
1387 *Cancer cell* 35, 692-704 e612.  
1388  
1389 Koren, A., Polak, P., Nemesh, J., Michaelson, J.J., Sebat, J., Sunyaev, S.R., and McCarroll,  
1390 S.A. (2012). Differential relationship of DNA replication timing to different forms of human  
1391 mutation and variation. *Am J Hum Genet* 91, 1033-1040.  
1392  
1393 Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for  
1394 Bisulfite-Seq applications. *Bioinformatics* 27, 1571-1572.  
1395  
1396 Landan, G., Cohen, N.M., Mukamel, Z., Bar, A., Molchadsky, A., Brosh, R., Horn-Saban, S.,  
1397 Zalcenstein, D.A., Goldfinger, N., Zundelovich, A., *et al.* (2012). Epigenetic polymorphism and  
1398 the stochastic formation of differentially methylated regions in normal and cancerous tissues.  
1399 *Nat Genet* 44, 1207-1214.  
1400  
1401 Landau, D.A., Clement, K., Ziller, M.J., Boyle, P., Fan, J., Gu, H., Stevenson, K., Sougnez, C.,  
1402 Wang, L., Li, S., *et al.* (2014). Locally disordered methylation forms the basis of intratumor  
1403 methylome variation in chronic lymphocytic leukemia. *Cancer cell* 26, 813-825.  
1404  
1405 Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter,  
1406 S.L., Stewart, C., Mermel, C.H., Roberts, S.A., *et al.* (2013). Mutational heterogeneity in cancer  
1407 and the search for new cancer-associated genes. *Nature* 499, 214-218.  
1408  
1409 Li, Z., Bao, S., Wu, Q., Wang, H., Eyler, C., Sathornsumetee, S., Shi, Q., Cao, Y., Lathia, J.,  
1410 McLendon, R.E., *et al.* (2009). Hypoxia-inducible factors regulate tumorigenic capacity of glioma  
1411 stem cells. *Cancer Cell* 15, 501-513.  
1412  
1413 Liao, B.B., Sievers, C., Donohue, L.K., Gillespie, S.M., Flavahan, W.A., Miller, T.E., Venteicher,  
1414 A.S., Hebert, C.H., Carey, C.D., Rodig, S.J., *et al.* (2017). Adaptive Chromatin Remodeling  
1415 Drives Glioblastoma Stem Cell Plasticity and Drug Tolerance. *Cell Stem Cell* 20, 233-246 e237.  
1416  
1417 Losman, J.A., and Kaelin, W.G., Jr. (2013). What a difference a hydroxyl makes: mutant IDH,  
1418 (R)-2-hydroxyglutarate, and cancer. *Genes Dev* 27, 836-852.

1419  
1420 Louis, D.N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee,  
1421 W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., and Ellison, D.W. (2016). The 2016 World Health  
1422 Organization Classification of Tumors of the Central Nervous System: a summary. *Acta*  
1423 *neuropathologica* 131, 803-820.  
1424  
1425 MacLeod, G., Bozek, D.A., Rajakulendran, N., Monteiro, V., Ahmadi, M., Steinhart, Z., Kushida,  
1426 M.M., Yu, H., Coutinho, F.J., Cavalli, F.M.G., *et al.* (2019). Genome-Wide CRISPR-Cas9  
1427 Screens Expose Genetic Vulnerabilities and Mechanisms of Temozolomide Sensitivity in  
1428 Glioblastoma Stem Cells. *Cell Rep* 27, 971-986 e979.  
1429  
1430 Mazor, T., Pankov, A., Johnson, B.E., Hong, C., Hamilton, E.G., Bell, R.J.A., Smirnov, I.V.,  
1431 Reis, G.F., Phillips, J.J., Barnes, M.J., *et al.* (2015). DNA Methylation and Somatic Mutations  
1432 Converge on the Cell Cycle and Define Similar Evolutionary Histories in Brain Tumors. *Cancer*  
1433 *cell* 28, 307-317.  
1434  
1435 Morton, A.R., Dogan-Artun, N., Faber, Z.J., MacLeod, G., Bartels, C.F., Piazza, M.S., Allan,  
1436 K.C., Mack, S.C., Wang, X., Gimple, R.C., *et al.* (2019). Functional Enhancers Shape  
1437 Extrachromosomal Oncogene Amplifications. *Cell* 179, 1330-1341 e1313.  
1438  
1439 Neftel, C., Laffy, J., Filbin, M.G., Hara, T., Shore, M.E., Rahme, G.J., Richman, A.R.,  
1440 Silverbush, D., Shaw, M.L., Hebert, C.M., *et al.* (2019). An Integrative Model of Cellular States,  
1441 Plasticity, and Genetics for Glioblastoma. *Cell* 178, 835-849 e821.  
1442  
1443 Noushmehr, H., Weisenberger, D.J., Diefes, K., Phillips, H.S., Pujara, K., Berman, B.P., Pan, F.,  
1444 Pelloski, C.E., Sulman, E.P., Bhat, K.P., *et al.* (2010). Identification of a CpG island methylator  
1445 phenotype that defines a distinct subgroup of glioma. *Cancer cell* 17, 510-522.  
1446  
1447 Orso, F., Cora, D., Ubezio, B., Provero, P., Caselle, M., and Taverna, D. (2010). Identification of  
1448 functional TFAP2A and SP1 binding sites in new TFAP2A-modulated genes. *BMC Genomics*  
1449 11, 355.  
1450  
1451 Polanski, K., Young, M.D., Miao, Z., Meyer, K.B., Teichmann, S.A., and Park, J.E. (2020).  
1452 BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* 36, 964-965.  
1453  
1454 Raney, B.J., Dreszer, T.R., Barber, G.P., Clawson, H., Fujita, P.A., Wang, T., Nguyen, N.,  
1455 Paten, B., Zweig, A.S., Karolchik, D., *et al.* (2013). Track data hubs enable visualization of user-  
1456 defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics* 30, 1003-  
1457 1005.  
1458  
1459 Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-  
1460 Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., *et al.* (2015). Integrative analysis of 111  
1461 reference human epigenomes. *Nature* 518, 317-330.  
1462  
1463 Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction  
1464 of single-cell gene expression data. *Nat Biotechnol* 33, 495-502.  
1465  
1466 Shaffer, S.M., Dunagin, M.C., Torborg, S.R., Torre, E.A., Emert, B., Krepler, C., Beqiri, M.,  
1467 Sproesser, K., Brafford, P.A., Xiao, M., *et al.* (2017). Rare cell variability and drug-induced  
1468 reprogramming as a mode of cancer drug resistance. *Nature* 546, 431-435.  
1469

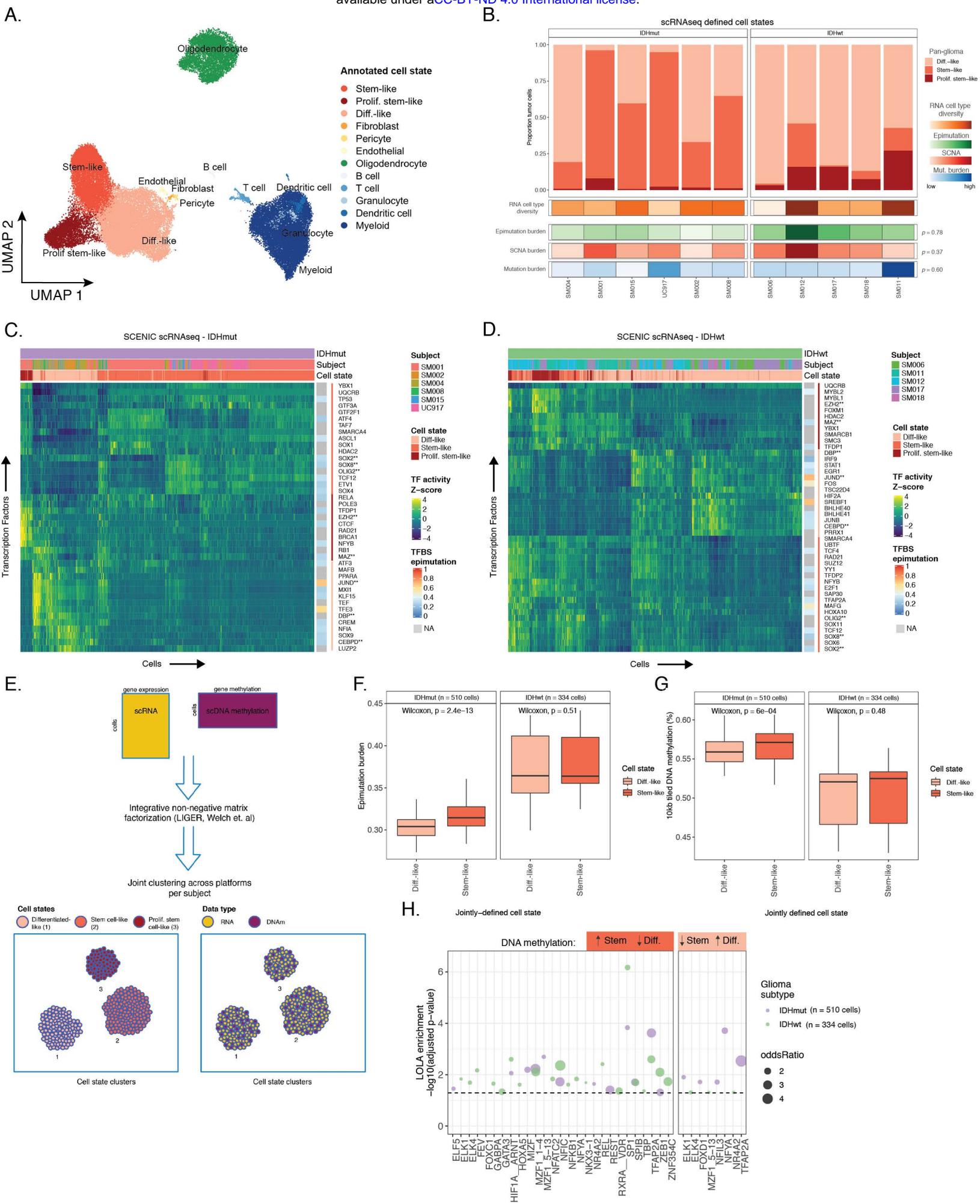
- 1470 Sharma, S.V., Lee, D.Y., Li, B., Quinlan, M.P., Takahashi, F., Maheswaran, S., McDermott, U.,  
1471 Azizian, N., Zou, L., Fischbach, M.A., *et al.* (2010). A chromatin-mediated reversible drug-  
1472 tolerant state in cancer cell subpopulations. *Cell* 141, 69-80.  
1473  
1474 Sheffield, N.C., and Bock, C. (2016). LOLA: enrichment analysis for genomic region sets and  
1475 regulatory elements in R and Bioconductor. *Bioinformatics* 32, 587-589.  
1476  
1477 Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y.,  
1478 Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell  
1479 Data. *Cell* 177, 1888-1902 e1821.  
1480  
1481 Taylor, A.M., Shih, J., Ha, G., Gao, G.F., Zhang, X., Berger, A.C., Schumacher, S.E., Wang, C.,  
1482 Hu, H., Liu, J., *et al.* (2018). Genomic and Functional Approaches to Understanding Cancer  
1483 Aneuploidy. *Cancer cell* 33, 676-689.e673.  
1484  
1485 Thienpont, B., Steinbacher, J., Zhao, H., D'Anna, F., Kuchnio, A., Ploumakis, A., Ghesquiere,  
1486 B., Van Dyck, L., Boeckx, B., Schoonjans, L., *et al.* (2016). Tumour hypoxia causes DNA  
1487 hypermethylation by reducing TET activity. *Nature* 537, 63-68.  
1488  
1489 Tirosh, I., Venteicher, A.S., Hebert, C., Escalante, L.E., Patel, A.P., Yizhak, K., Fisher, J.M.,  
1490 Rodman, C., Mount, C., Filbin, M.G., *et al.* (2016). Single-cell RNA-seq supports a  
1491 developmental hierarchy in human oligodendroglioma. *Nature* 539, 309-313.  
1492  
1493 Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-  
1494 connected communities. *Sci Rep* 9, 5233.  
1495  
1496 Turcan, S., Rohle, D., Goenka, A., Walsh, L.A., Fang, F., Yilmaz, E., Campos, C., Fabius, A.W.,  
1497 Lu, C., Ward, P.S., *et al.* (2012). IDH1 mutation is sufficient to establish the glioma  
1498 hypermethylator phenotype. *Nature* 483, 479-483.  
1499  
1500 Venteicher, A.S., Tirosh, I., Hebert, C., Yizhak, K., Neftel, C., Filbin, M.G., Hovestadt, V.,  
1501 Escalante, L.E., Shaw, M.L., Rodman, C., *et al.* (2017). Decoupling genetics, lineages, and  
1502 microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science (New York, NY)* 355.  
1503  
1504 Verburg, N., Barthel, F.B., Anderson, K.J., Johnson, K.C., Koopman, T., Yaqub, M.M., Hoekstra,  
1505 O.S., Lammertsma, A.A., Barkhof, F., Pouwels, P.J.W., *et al.* (2020). DNA methylation  
1506 classification in diffuse glioma shows little spatial heterogeneity after adjusting for tumor purity.  
1507 bioRxiv, 2020.2003.2028.012732.  
1508  
1509 Verhaak, R.G.W., Bafna, V., and Mischel, P.S. (2019). Extrachromosomal oncogene  
1510 amplification in tumour pathogenesis and evolution. *Nat Rev Cancer* 19, 283-288.  
1511  
1512 Wang, L., Babikir, H., Muller, S., Yagnik, G., Shamardani, K., Catalan, F., Kohanbash, G.,  
1513 Alvarado, B., Di Lullo, E., Kriegstein, A., *et al.* (2019). The Phenotypes of Proliferating  
1514 Glioblastoma Cells Reside on a Single Axis of Variation. *Cancer Discov* 9, 1708-1719.  
1515  
1516 Wang, Q., Hu, B., Hu, X., Kim, H., Squatrito, M., Scarpace, L., deCarvalho, A.C., Lyu, S., Li, P.,  
1517 Li, Y., *et al.* (2017). Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates  
1518 with Immunological Changes in the Microenvironment. *Cancer cell* 32, 42-56 e46.  
1519

- 1520 Welch, J.D., Kozareva, V., Ferreira, A., Vanderburg, C., Martin, C., and Macosko, E.Z. (2019).  
1521 Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. *Cell*  
1522 177, 1873-1887 e1817.  
1523
- 1524 Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene  
1525 expression data analysis. *Genome Biol* 19, 15.  
1526
- 1527 Wu, S., Turner, K.M., Nguyen, N., Raviram, R., Erb, M., Santini, J., Luebeck, J., Rajkumar, U.,  
1528 Diao, Y., Li, B., *et al.* (2019). Circular ecDNA promotes accessible chromatin and high  
1529 oncogene expression. *Nature* 575, 699-703.  
1530
- 1531 Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K.,  
1532 Kivioja, T., Dave, K., Zhong, F., *et al.* (2017). Impact of cytosine methylation on DNA binding  
1533 specificities of human transcription factors. *Science (New York, NY)* 356.  
1534
- 1535 Yuan, J., Levitin, H.M., Frattini, V., Bush, E.C., Boyett, D.M., Samanamud, J., Ceccarelli, M.,  
1536 Dovas, A., Zanazzi, G., Canoll, P., *et al.* (2018). Single-cell transcriptome analysis of lineage  
1537 diversity in high-grade glioma. *Genome Med* 10, 57.  
1538
- 1539 Zhu, J., Tsai, H.J., Gordon, M.R., and Li, R. (2018a). Cellular Stress Associated with  
1540 Aneuploidy. *Dev Cell* 44, 420-431.  
1541
- 1542 Zhu, P., Guo, H., Ren, Y., Hou, Y., Dong, J., Li, R., Lian, Y., Fan, X., Hu, B., Gao, Y., *et al.*  
1543 (2018b). Single-cell DNA methylome sequencing of human preimplantation embryos. *Nat Genet*  
1544 50, 12-19.  
1545



**Figure 1. Single-cell DNA methylation sequencing highlights intratumoral heterogeneity and disruption of epigenetic regulatory mechanisms.**

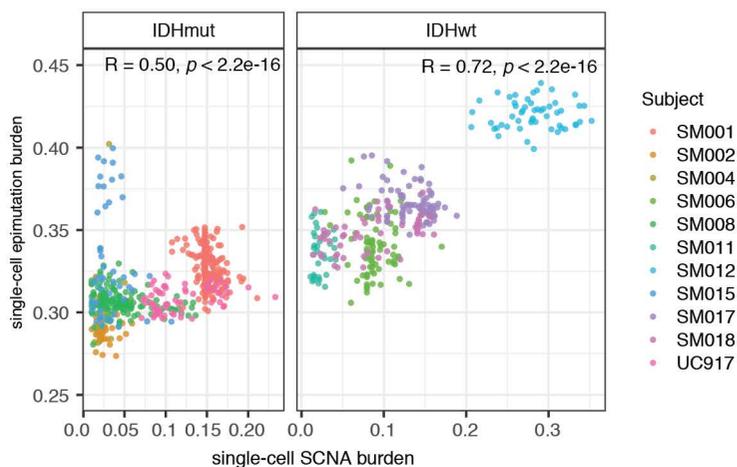
(A) Schematic diagram detailing tumor sample processing and molecular profiling of single cells and bulk tumor samples ( $n = 11$  subjects). (B) Multidimensional scaling (MDS) analysis using pairwise individual CpG distance metrics calculated between individual cells. Shapes represent whether a sample was a single tumor cell ( $n = 844$  cells) or 50-tumor cells,  $n = 9/11$  subjects). Colors indicate individual subjects, shaded regions indicated *IDH1*-mutation status of tumor, and annotation is provided indicating clinical timepoint (I = initial, R = recurrence). (C) Schematic depiction of local DNA methylation disorder in different genomic contexts. Left panel demonstrates epimutation, or local DNA methylation disorder, at the promoter region, where gene expression is disrupted by epimutation. The right panel provides an example of disrupted transcription factor binding due to epimutation. (D) Boxplots of tumor cell epimutation burden grouped by sample. Each boxplot spans the interquartile range with the whiskers representing the absolute range, excluding outliers. Wilcoxon rank sum  $p$ -value represents comparison between IDHmut and IDHwt epimutation burden. Each sample is annotated with clinical and molecular metrics with  $p$ -values indicating the relationship between sample mean epimutation burden and whole-genome sequencing derived somatic mutation burden or somatic alteration burden (Spearman correlation). (E) Boxplots of gene expression values, as  $\log_2$  (counts per million), from single-cell RNAseq data across different gene epimutation groups. Gene epimutation groups are defined by the determining the mean epimutation value across a single gene. Color indicates *IDH1* mutation status. (F) Boxplots of gene expression dispersion. Expression profiles were mean-expression scaled to account for expression level-dependent variability across the same gene epimutation groups defined in panel E. (G) Scatterplot of the mean single-cell epimutation burden metric calculated across transcription factor binding sites (TFBSs) within a subtype, ordered by IDHwt TFBS epimutation. Each column represents a single transcription factor (TF) with a colored dotted line connecting IDHmut and IDHwt values. Names of TFs previously indicated to confer fitness advantages to glioma cells (MacLeod et al.) are listed above their TFBS epimutation burden estimate. (H) Scatterplot depicting the association between average single-cell epimutation burden 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.



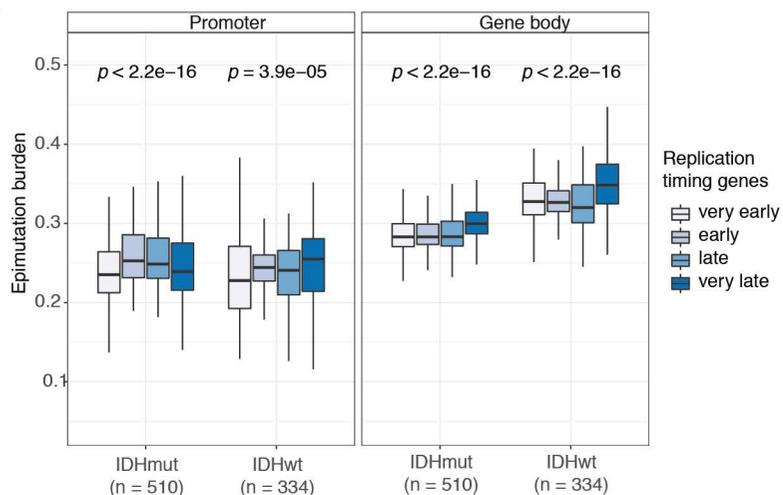
**Figure 2. Integrative single-cell gene expression and DNA methylation analyses nominate epigenetic regulators of glioma cell state variability.**

(A) Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction plot of scRNAseq data ( $n = 55,284$  tumor cells,  $n = 11$  subjects) showing the clustering of cell populations by transcriptionally defined cell state (point color) and labelled according to marker gene expression (Figure S6B). (B) Stacked bar plots representing the proportion of cellular states per tumor for pan-glioma 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 epimutation burden, whole-genome sequencing derived somatic alteration burden, or whole-genome sequencing derived somatic mutation burden (Spearman correlation). (C-D) Enriched transcription factor activity across pan-glioma cellular states determined using SCENIC algorithm and displayed as a heatmap of z-score enrichment values. Visualization is presented for the hierarchical clustering of 5,000 randomly selected tumor cells in both (C) IDHmut and (D) IDHwt tumors. (E) Schematic diagram representing LIGER workflow to jointly cluster single-cell RNAseq and DNA methylation data generated from the same tumor dissociation. (F) Boxplots representing the average epimutation burden in differentiated-like and stem-like populations in IDHmut (left panel) and IDHwt (right) tumors. (G) Boxplots representing the 10-kb tiled DNA methylation levels in differentiated-like and stem-like populations in IDHmut (left panel) and IDHwt (right) tumors. (H) Region set enrichment analysis for 10-kb tiles with higher DNA methylation in Stem-like (left panel) or Differentiated-like cells (right panel). Enrichment was determined by Locus Overlap Analysis (LOLA). Individual points represent enrichment of specific TFs in differentially methylated regions, color indicates results for specific IDH-mutant subtype, point size indicates log-odds ratio, and dotted line represents the statistical significance threshold (adjusted  $p$ -value  $< 0.05$ ).

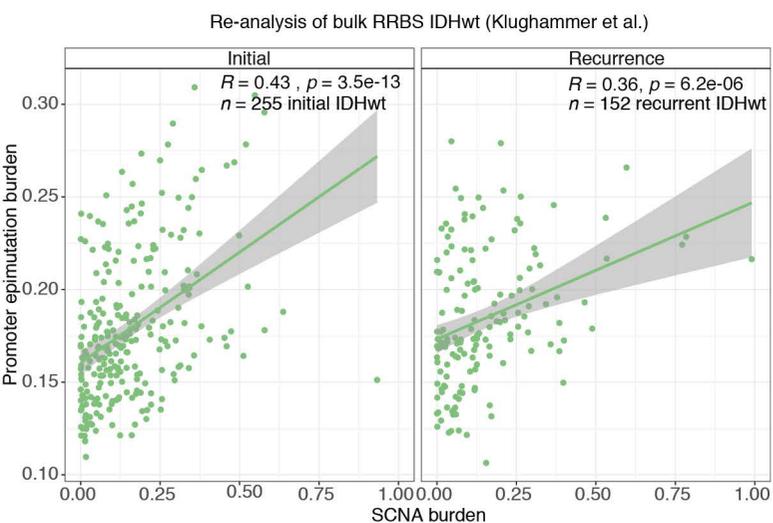
A.



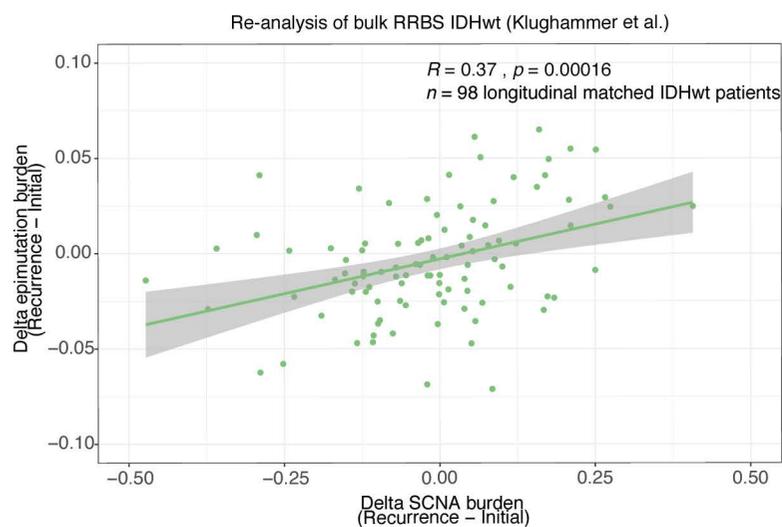
B.



C.

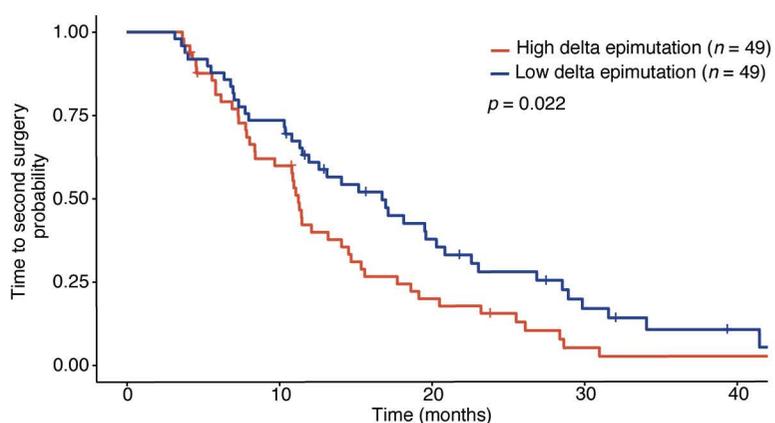


D.



E.

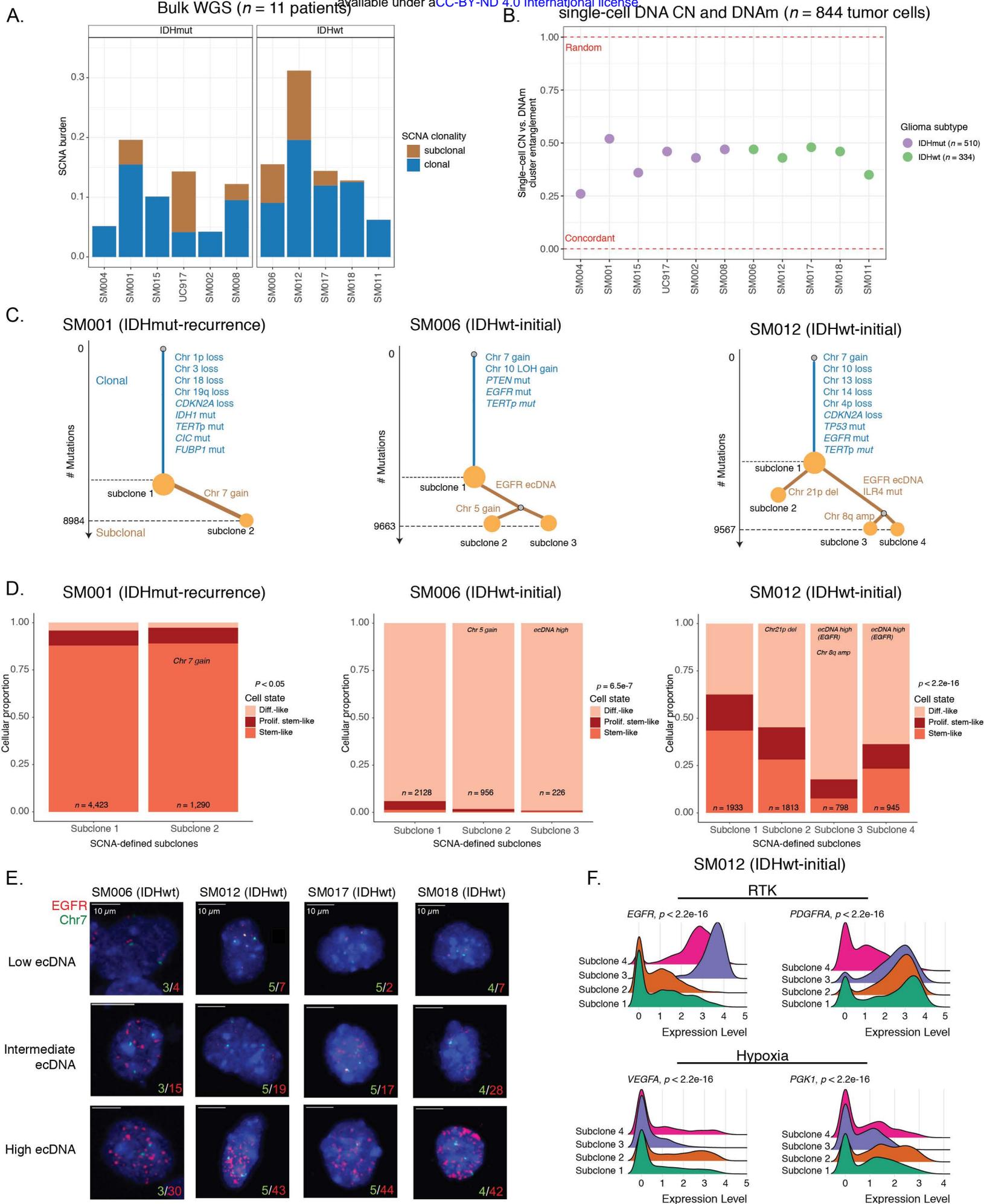
Re-analysis of bulk RRBS IDHwt (Klughammer et al.)



\*\*Epimutation mutation group was significantly associated with time to second surgery in multivariate Cox model with Age and Sex (HR = 1.69 (95% CI 1.09 - 2.62),  $p = 0.02$ )

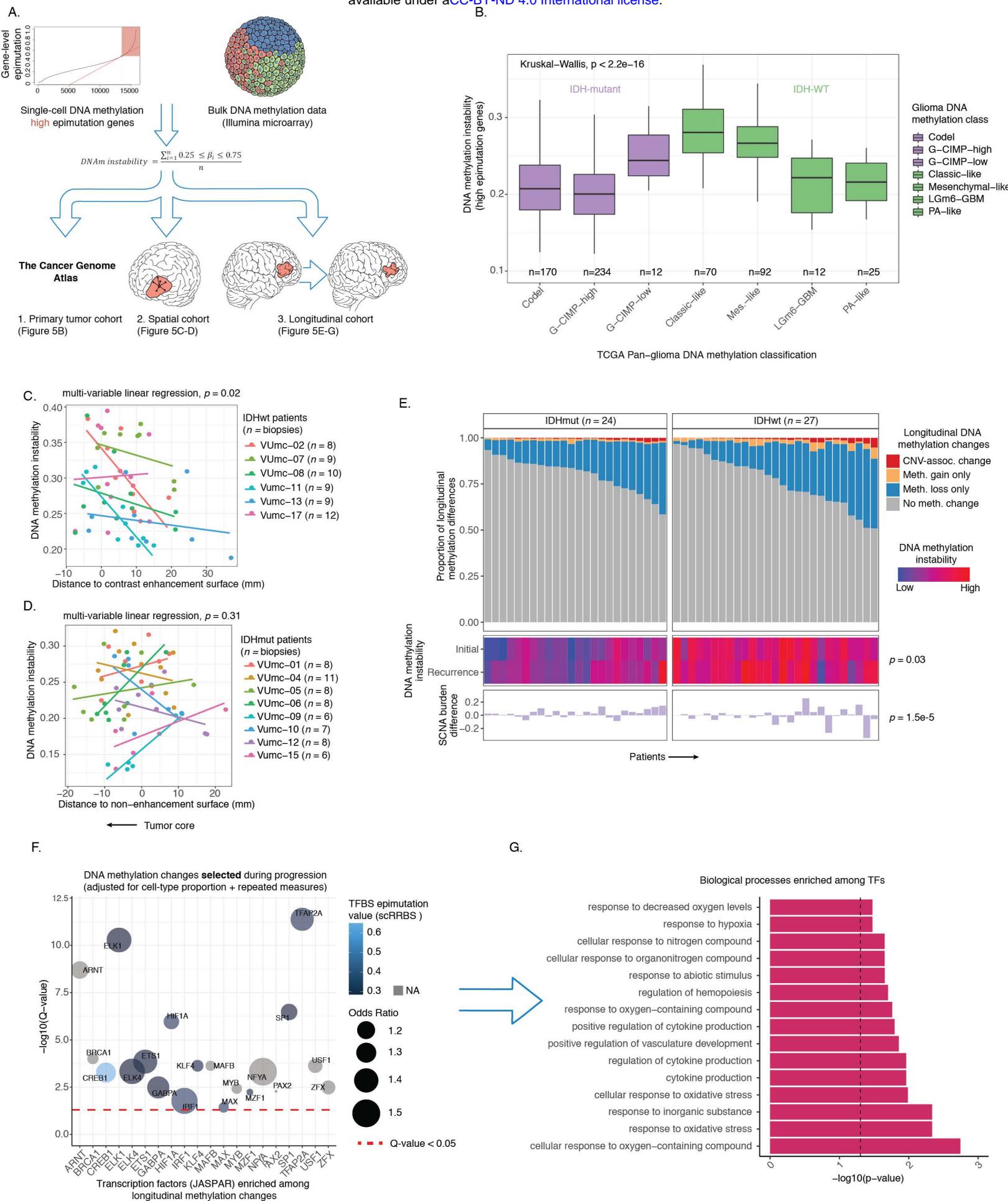
**Figure 3. Somatic copy number alterations are associated with stochastic DNA methylation changes during disease evolution.**

(A) Scatterplot depicting the association between single-cell ( $n = 844$  tumor cells) somatic copy number alteration (SCNA) and epimutation burden estimates by IDHmut (left panel) and IDHwt (right panel) subtypes. Points are colored by patient. Spearman correlation coefficients represent subtype-specific estimates. (B) Boxplots of epimutation burden calculated across the promoter (left panel) and gene body regions (right panel) based on different DNA replication times in IDHmut ( $n = 510$ ) and IDHwt ( $n = 334$ ) single cells. Kruskal-Wallis  $p$ -values indicate a test for differences across the replication time groupings separately for IDHmut and IDHwt cells. (C) Scatterplot depicting the re-analysis of bulk promoter epimutation burden and SCNA burden in IDHwt initial ( $n = 255$ ) and recurrent ( $n = 152$ ) tumors (Klughammer et al.). Spearman correlation coefficients and  $p$ -values are presented for each independent timepoint. (D) Scatterplot depicting the association between bulk delta (subject-specific recurrence – initial estimates) SCNA burden and delta promoter epimutation burden in longitudinally profiled IDHwt tumors ( $n = 98$  subjects, Klughammer et al.) Spearman correlation coefficient and  $p$ -value are presented. (E) Kaplan-Meier curve depicting time to second surgery in subjects where the change in epimutation burden between initial and recurrent disease was above (high, red) and below (low, blue) the median. Log-rank  $p$ -value for univariate analysis is presented within the figure. Hazard Ratio and  $p$ -value for change in epimutation burden are presented below for multi-variate Cox proportional hazard model including subject age and sex as predictors.



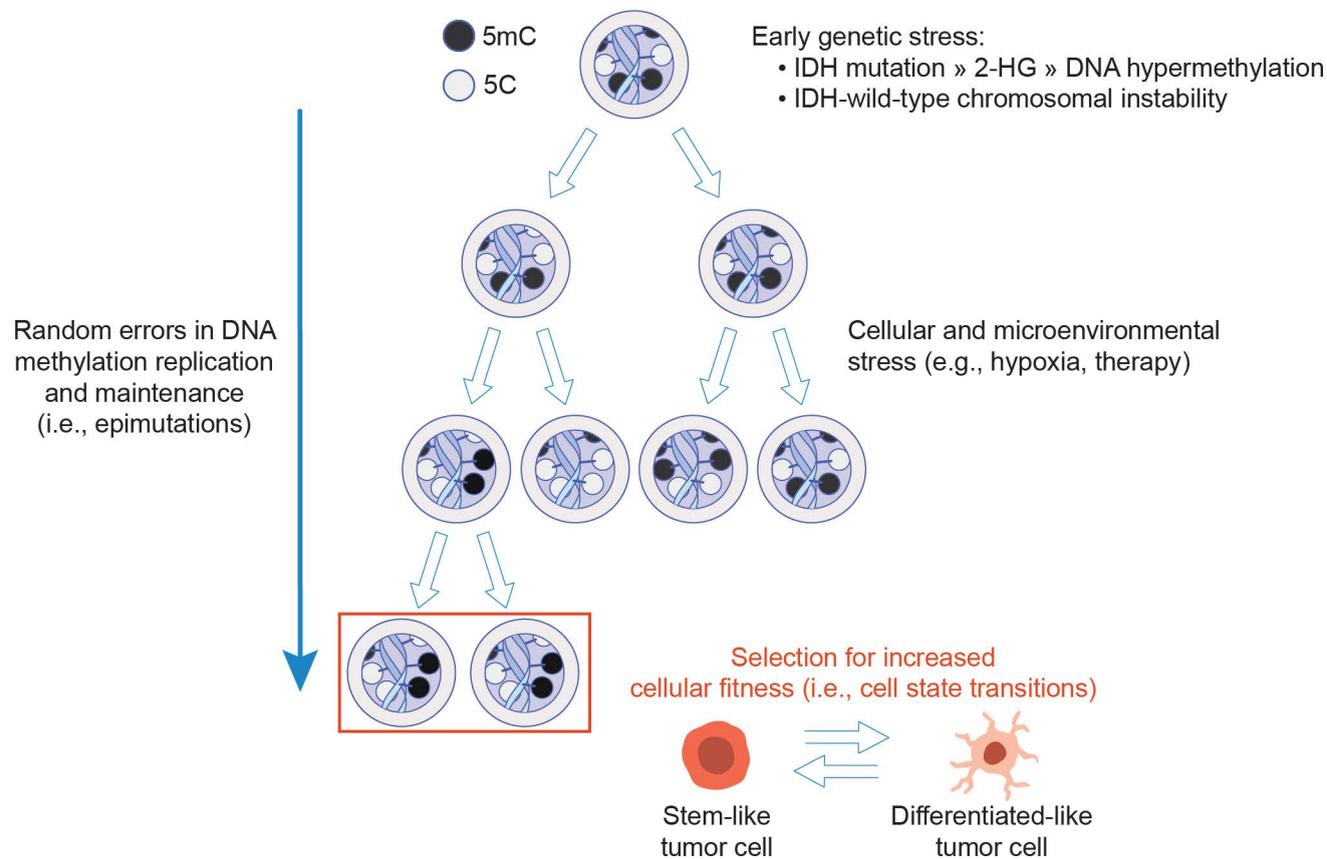
**Figure 4. Clonal evolution analyses highlight early copy number evolution followed by epigenetic and transcriptomic diversification.**

(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) Scatterplot depicting entanglement coefficients for tanglegrams comparing cluster dendrograms of scRRBS derived copy number and DNA methylation profiles. A coefficient of 0 indicates complete alignment of the tree structures, whereas a 1 indicates random association. Color indicates *IDH1* mutation status. (C) Examples of phylogenetic trees constructed from whole genome sequencing data (mutations and SCNAs) 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 analyses (Ceccarelli et al., 2016) and chromosomal arm-level events are presented. (D) Single-cell RNAseq-derived cellular proportions separated by copy number-defined tumor subclone (Figure S3). Reported *p*-values represent Fisher's exact test comparing the cellular state distributions across tumor subclones. (E) Representative Fluorescence *in situ* hybridization (FISH) images for IDHwt tumors computationally predicted to harbor *EGFR* extrachromosomal DNA (ecDNA) by whole genome sequencing (*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 microns. (F) Ridge plots of SM012 single-cell expression of receptor tyrosine kinase and hypoxia-associated genes, grouped by copy number-defined subclones. Reported *p*-values represent Wilcoxon Rank Sum tests comparing the gene expression of cells across tumor subclones.



### Figure 5. Integrated molecular trajectories supports adaptive DNA methylation changes under microenvironmental and therapeutic pressures.

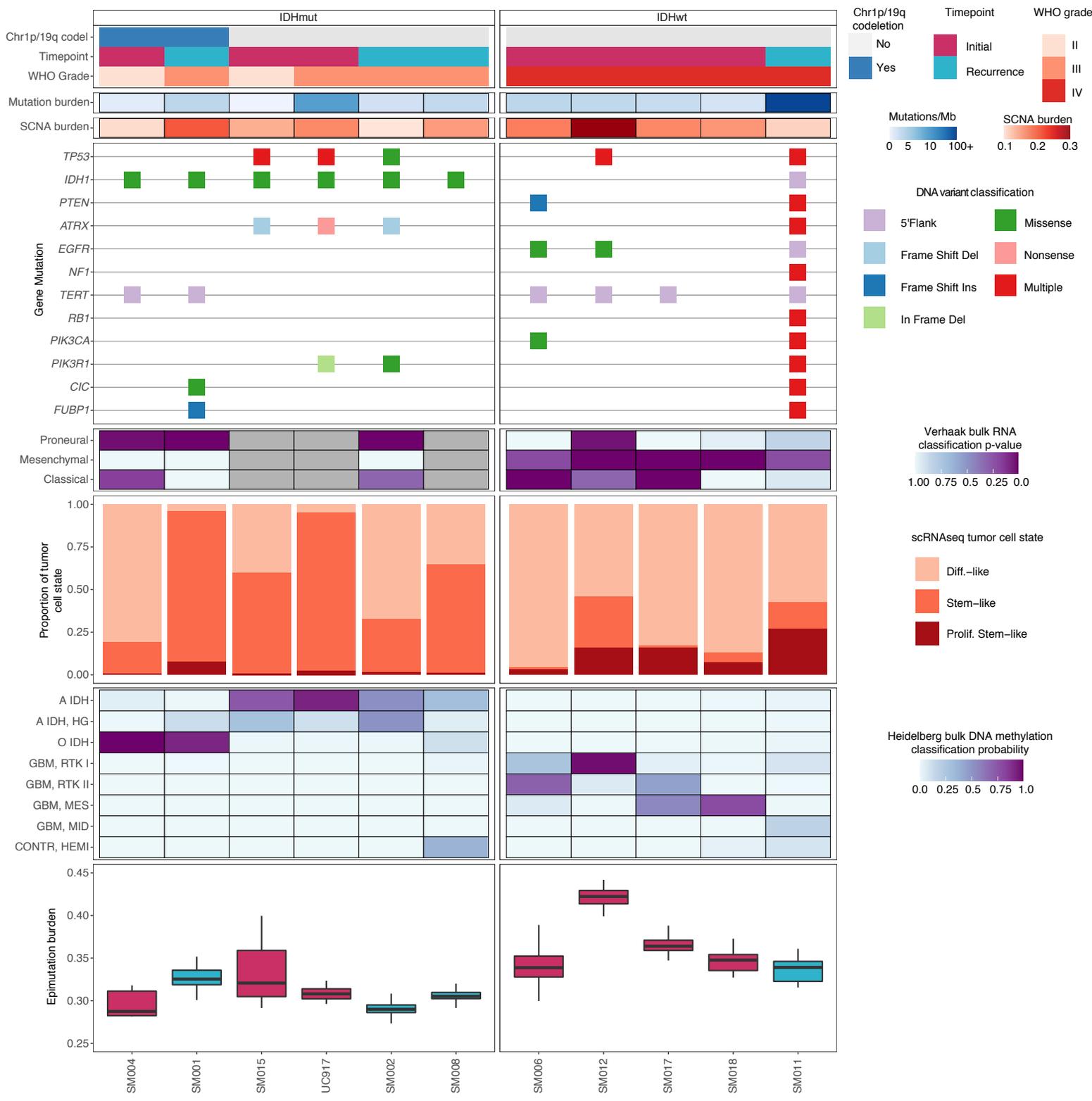
(A) Schematic workflow for construction of a DNA methylation instability metric in bulk cohorts informed by regions of high epimutation in single-cell DNA methylation data. The DNA methylation instability metric was calculated across bulk DNA methylation microarray data in a primary tumor cohort (TCGA), a cohort of multiple, spatially distinct biopsies from the same tumor (Verburg et al.), and a longitudinal cohort with accompanying genomic sequencing data (Glioma Longitudinal AnalySiS (GLASS), Barthel et al.). (B) Boxplots displaying the bulk DNA methylation instability metric calculated across previously described DNA-methylation based TCGA tumor classifications (Ceccarelli et al). Colors represent *IDH1/2* mutation status, and Kruskal-Wallis  $p$ -value testing for differences in distributions across classification is reported ( $n = 615$  primary gliomas,  $p < 2.2e-16$ ). (C-D) Scatterplots depicting distance from radiographic features plotted against the DNA methylation instability metric. Colors represent spatially separated biopsies from a single patient at initial clinical timepoint for (C) IDHwt tumors ( $n = 57$  biopsies,  $n = 6$  subjects) and (D) IDHmut tumors ( $n = 62$  biopsies,  $n = 8$  subjects). Linear regression lines colored by patient demonstrate the relationship between DNA methylation instability and radiographic features (i.e., contrast enhancement surface). The  $p$ -value reported from a multivariable linear regression model adjusting for subject represents the subtype-specific association between DNA methylation instability and radiographic feature. Biopsies taken closer to the tumor's center (i.e., core) have the lowest value (left hand side of plot). (E) Each column represents an individual patient sampled across initial and recurrent timepoints and is separated into IDHmut ( $n = 24$  subjects) and IDHwt ( $n = 27$  subjects). Top panel, stacked bar plot represents the proportion of CpGs sites that experienced DNA methylation change associated with a subject-specific copy number change (defined by DNA sequencing data) between primary and recurrent disease (red), DNA methylation gain not associated with a CNV change (orange), DNA methylation loss not associated with a CNV change (blue), and no longitudinal DNA methylation change (gray). Middle panel, heatmap of DNA methylation instability metric in primary and recurrent disease (blue = low, red = high). Bottom panel, differences in SCNA burden between primary and recurrent tumor. All associated  $p$ -values represent Spearman correlations between absolute change in associated metric and the fraction of longitudinal DNA methylation differences. (F) Enrichment analysis for differentially methylated CpGs between primary and recurrent timepoints when adjusting for cellular composition, glioma subtype, and subject included as a random effect. Individual points represent enrichment of specific TFs in differentially methylated positions, color indicates the average TFBS epimutation burden from single-cell RRBS data (Figure 1G), point size indicates log-odds ratio, and dotted line represents the statistical significance threshold ( $Q$ -value  $< 0.05$ ). (G) Gene Ontology enrichment of transcription factors associated with longitudinal DNA methylation changes. Dotted line represents threshold for statistical significance (Fisher's exact test,  $p < 0.05$ ).



**Figure 6. Model of epigenetic heterogeneity and tumor evolution.**

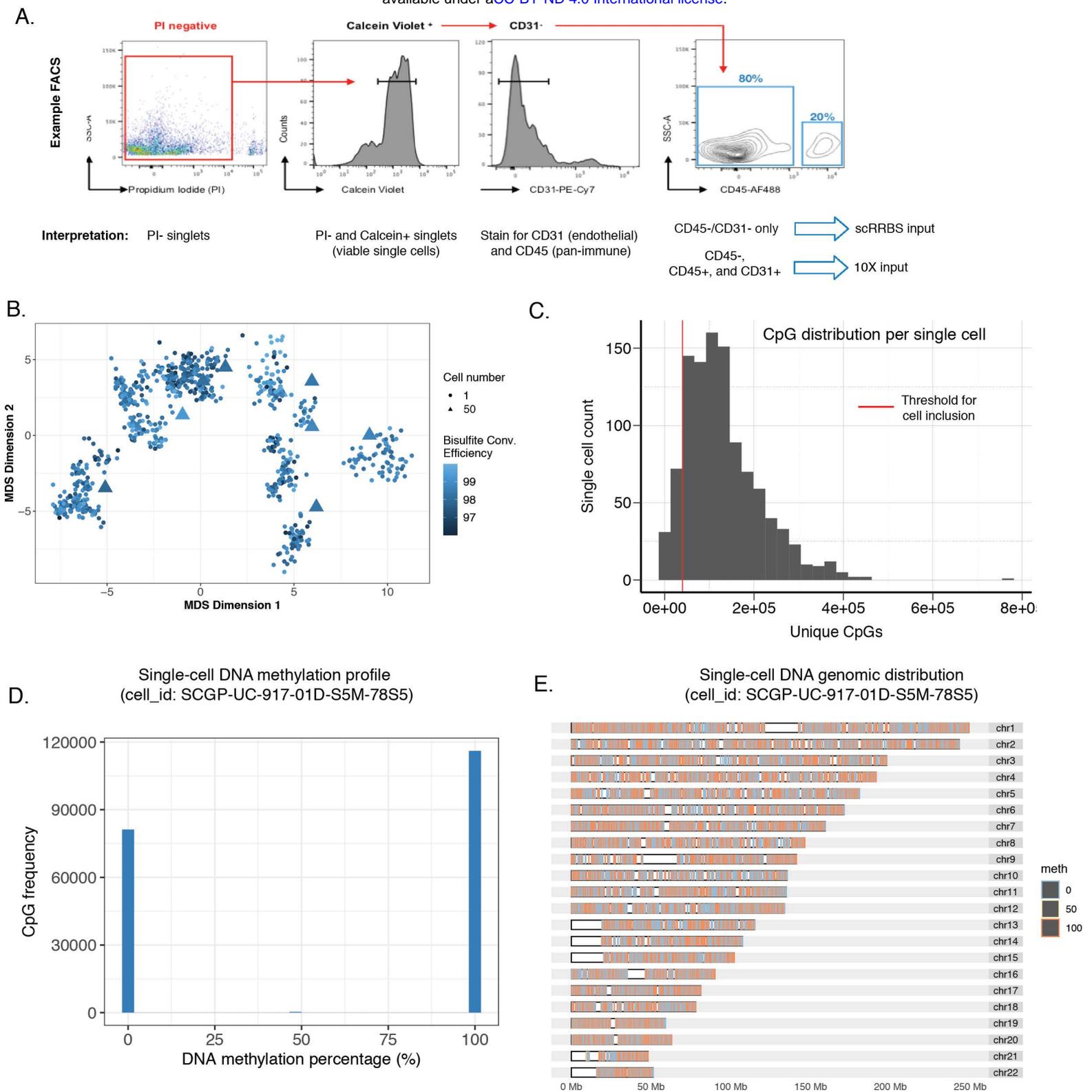
Schematic depiction of tumor evolution with general DNA methylation patterns represented by methylated (5-methylcytosine, 5mC) and unmethylated (5C) regions of the epigenome. Initiating genetic events such as *IDH1* and other driver mutations as well as somatic copy number alterations represent early stresses in glioma evolution that precipitate epigenetic heterogeneity. Both mutations in epigenetic enzymes and SCNAs can increase the likelihood of heritable DNA methylation alterations (i.e., epimutations). *IDH1* mutations result in the production of the oncometabolite 2-Hydroxyglutarate (2-HG) that leads to failure to remove aberrant DNA methylation while SCNAs can generate mitotic stress leading to the erosion of ordered DNA methylation. Non-genetic determinants further shape epigenetic heterogeneity as tumors evolve by exposing cells to spatially distinct microenvironmental stresses that impact the DNA methylation replication machinery. The subsequent epigenetic diversity provides an additional layer on which clonal evolution acts to select those cells with fitness-conferring epigenetic alterations. Ultimately, the loosened epigenetic control allows tumor cells to transition to cell states responsive to different selective pressures.

**Figure S1**



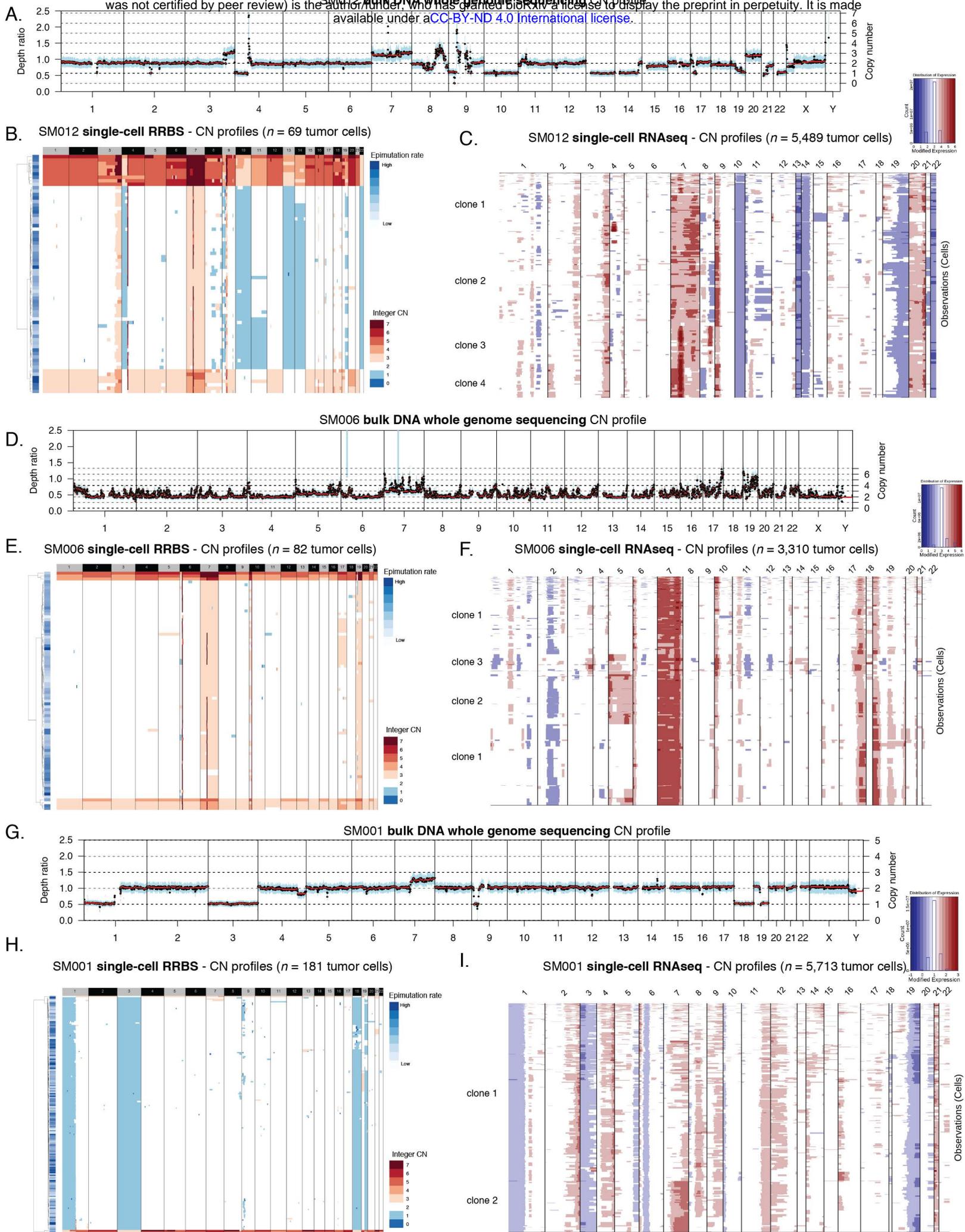
**Figure S1. Integrated molecular profiles of patient samples. Related to Figure 1.**

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 DNA methylation microarray subtype classification probabilities (Capper et al.), and boxplots of single-cell epimutation burden with samples colored by clinical timepoint.



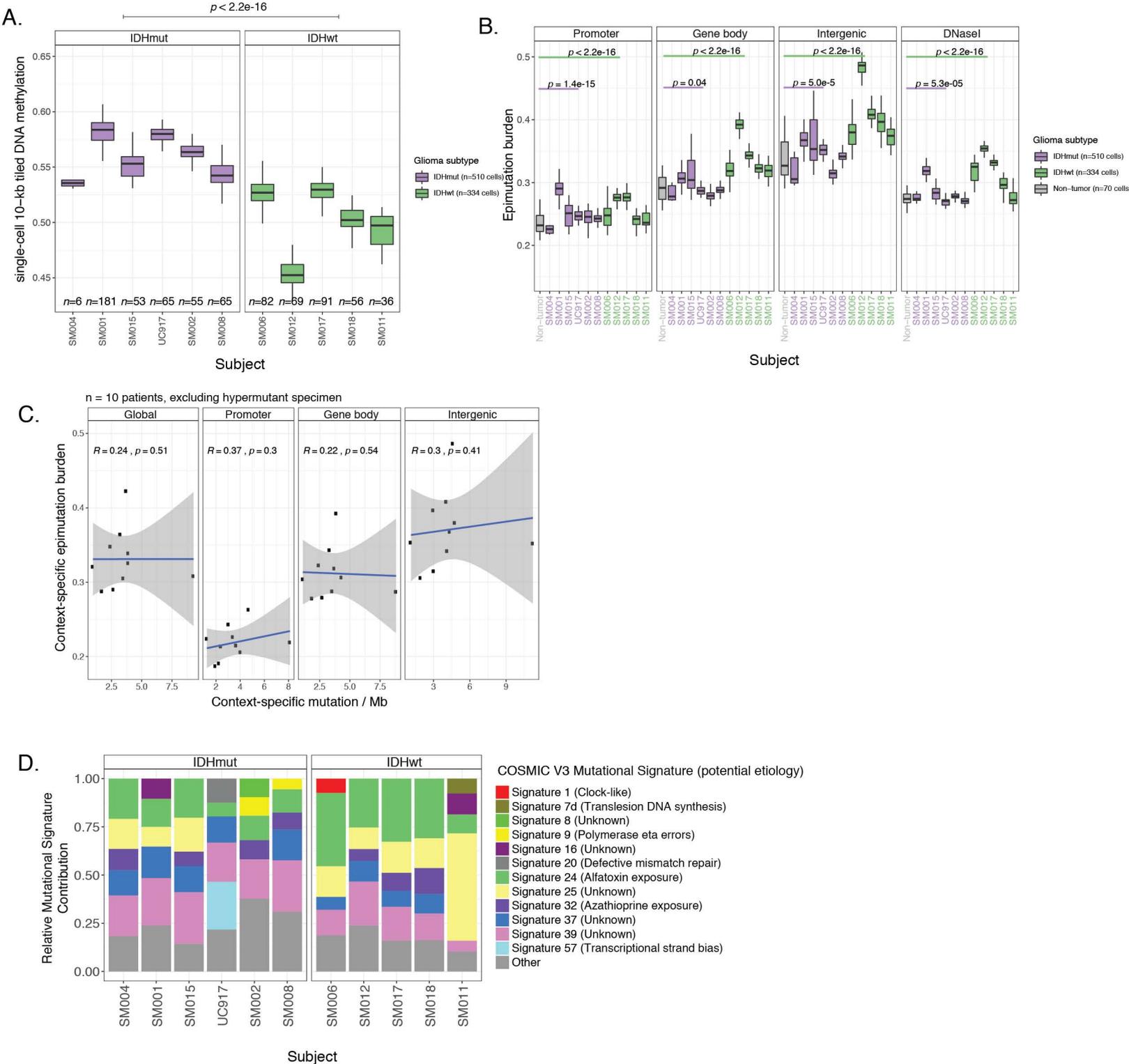
**Figure S2. Sample pre-processing and metrics related to single-cell DNA methylation data assessment. Related to Figure 1.**

(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 same multidimensional scaling (MDS) analysis using pairwise distance metrics calculated between individual cells as in Figure 1B, except colored by bisulfite conversion efficiency. (C) 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. (D) Representative distribution of single locus DNA methylation estimates for a single cell. DNA methylation percentage of 0 represents an unmethylated locus, while a percentage of 100 represents a methylated locus. (E) Representative genomic distribution of DNA methylation values within a single cell.



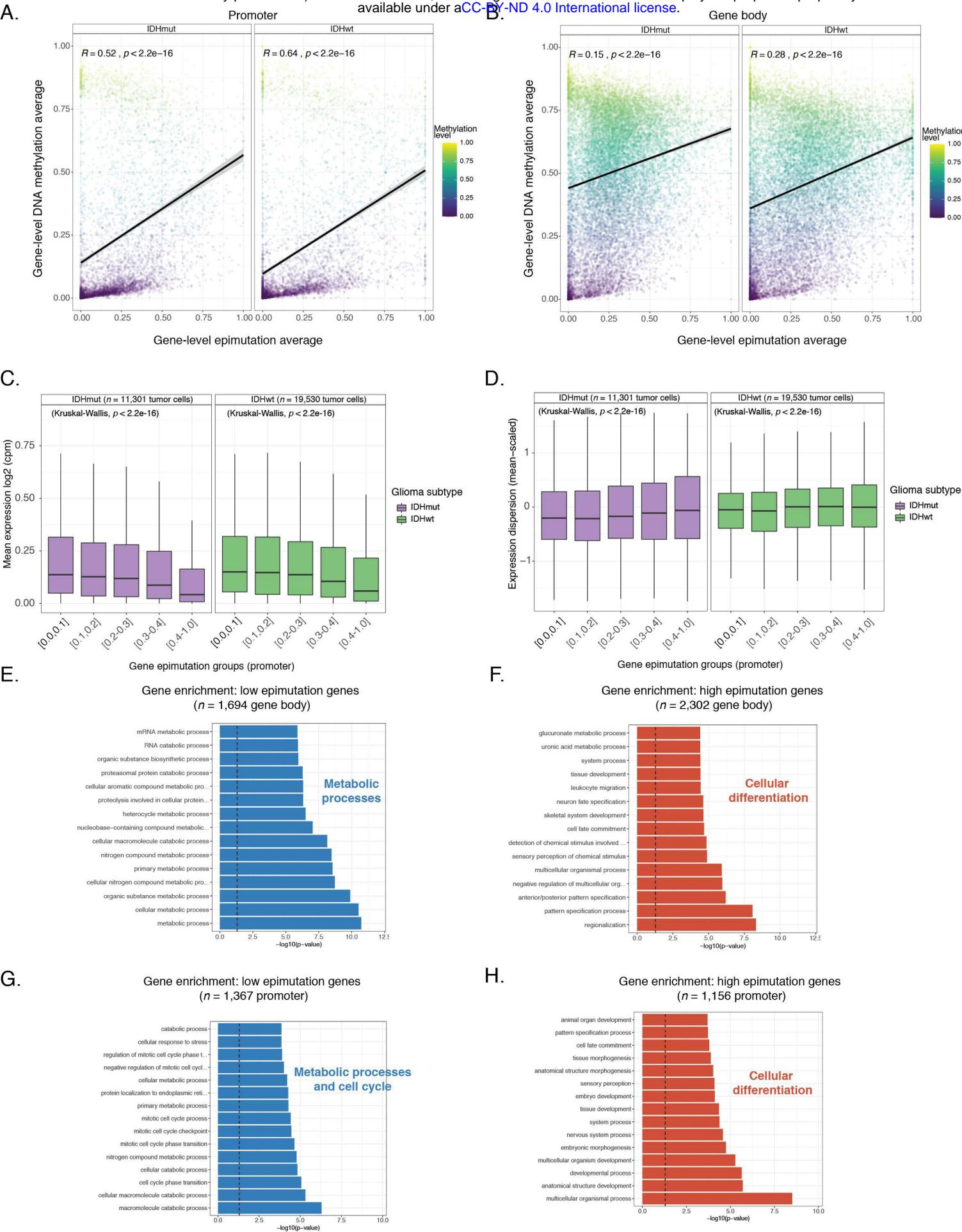
**Figure S3. Somatic copy number alteration examples estimated from whole genome sequencing, single-cell Reduced Representation Bisulfite Sequencing, and single-cell RNA-sequencing. Related to Figure 1.**

(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 DNA methylation-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 epimutation burden 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 Figure 4D. (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).



**Figure S4. Distribution and relationship of DNA methylation and epimutation throughout the glioma genome. Related to Figure 1.**

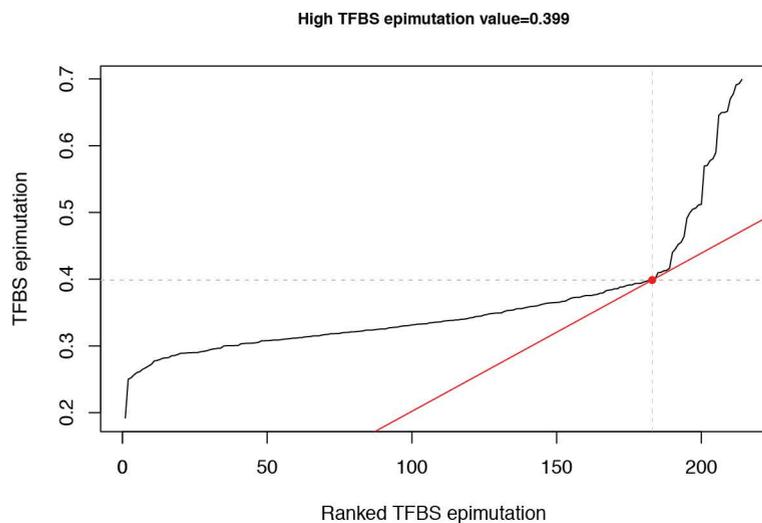
(A) Boxplots representing average 10-kb tiled DNA methylation values per single tumor cell. (B) Boxplots highlighting the single-cell epimutation burden estimates calculated across different genomic contexts. (C) Scatterplots showing the relationship between genomic context-specific single-cell epimutation burden (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 (i.e., all regions), promoter, gene body, and intergenic regions (Spearman correlations  $p > 0.05$  for all comparisons). (D) 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.



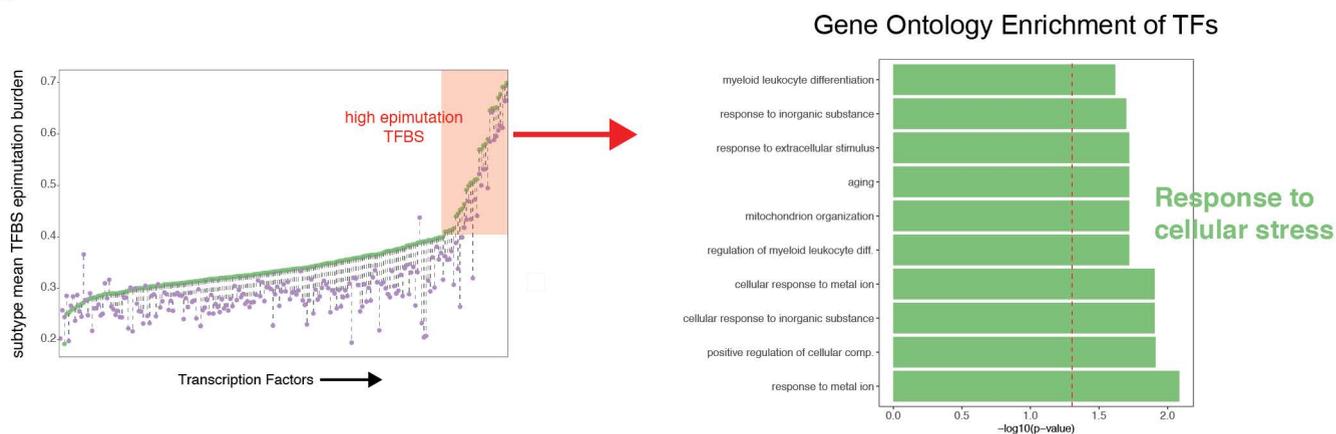
**Figure S5. Association between epimutation and disrupted transcriptional programs.  
Related to Figure 1.**

(A-B) Scatterplots depicting single-cell gene-level epimutation average plotted against the gene-level methylation estimates in both (A) promoter regions and (B) gene body regions. (C) Boxplots of gene expression values, in log<sub>2</sub> (counts per million), from single-cell RNAseq data across different sets of promoter regions defined by gene-derived epimutation groups. Gene epimutation groups are defined by the determining the mean epimutation value across a single gene. Color indicates *IDH1* mutation status. (D) Boxplots of gene expression dispersion that were mean-expression scaled to account for expression level-dependent variability across the same promoter-based gene epimutation groups defined in panel C. (E-F) Gene Ontology enrichment analyses for low epimutation genes (Figure 1E, mean epimutation across all tumor cells: 0.0 - 0.1) and high epimutation genes (Figure 1F, mean epimutation across all tumor cells: > 0.5) using gene body estimates. A meta biological process is placed next to significant Gene Ontology terms. (G-H) Same analyses presented in panels E-F, but for gene-level epimutation estimates determined in promoters.

A.



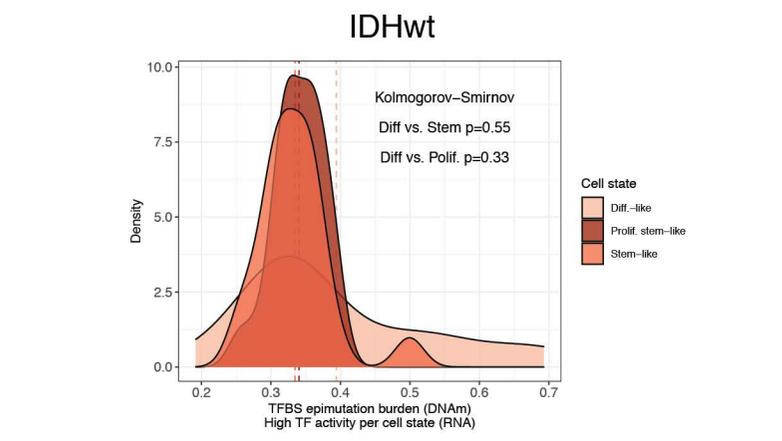
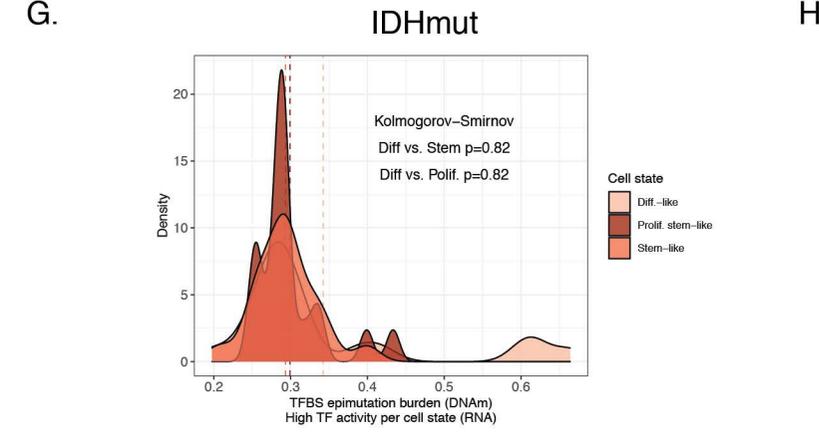
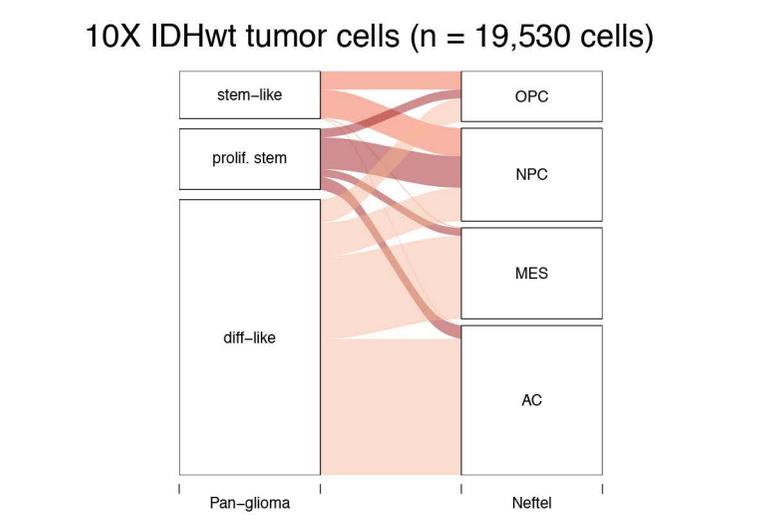
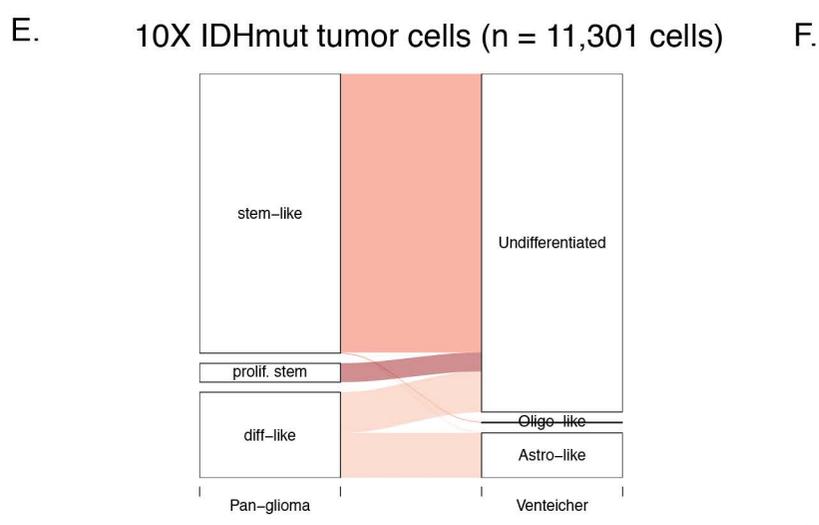
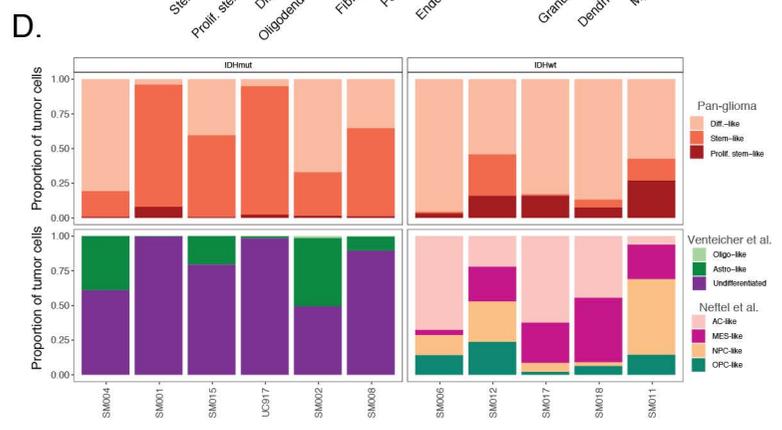
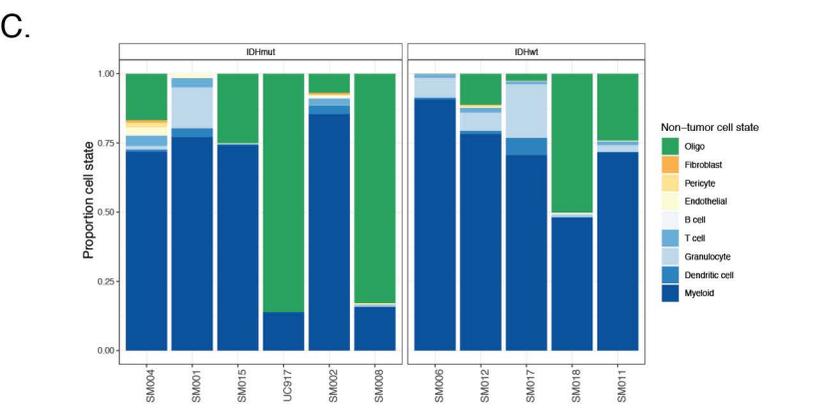
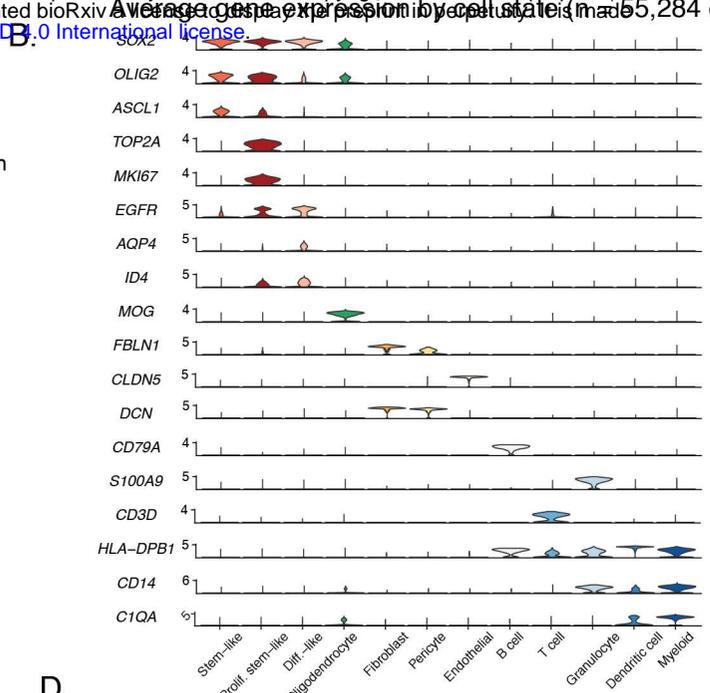
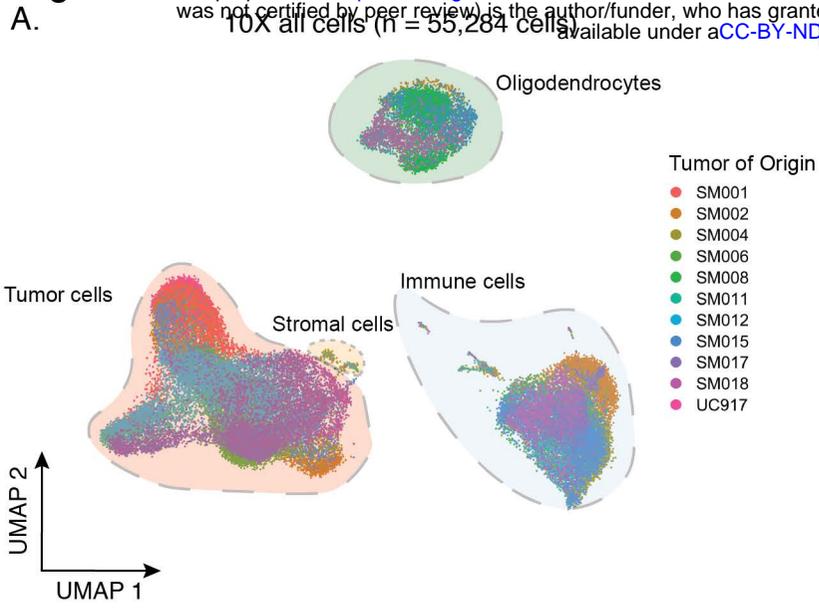
B.



**Figure S6. Enrichment of high epimutation transcription factors and association with environmental stress response. Related to Figure 1.**

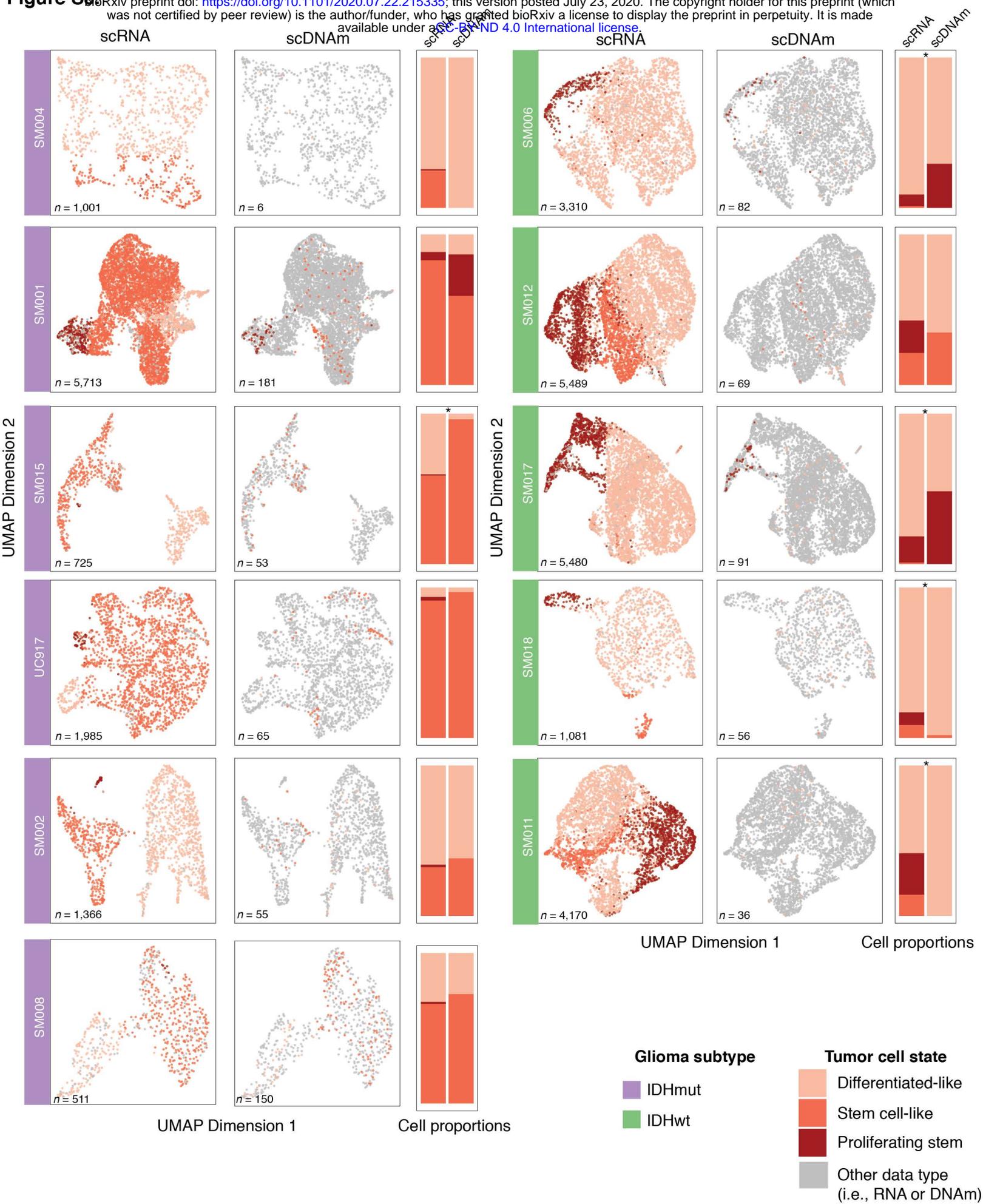
(A) Computational approach to defining TFBSs with high epimutation burden (red tangent line at 0.399 TFBS epimutation burden). X-axis represents each TF ordered by mean epimutation burden in IDHwt single-cells ( $n = 334$  cells).

(B) Gene Ontology enrichment analysis of TFs with high epimutation burden in their binding sites. A meta biological process is placed next to significant Gene Ontology (GO) terms.



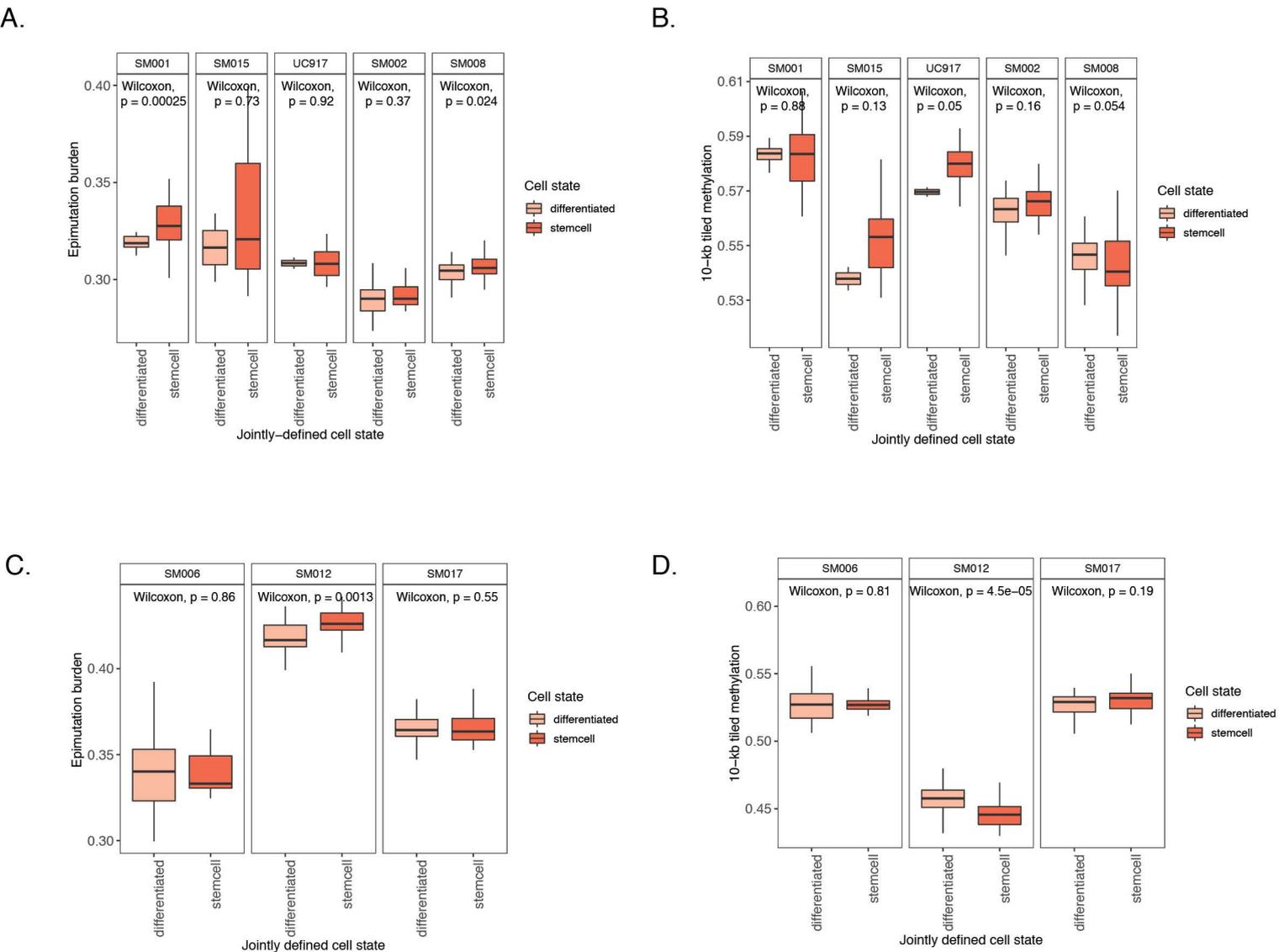
**Figure S7. Pan-glioma cell state assignment and characteristics. Related to Figure 2.**

(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 represents the tumor of origin. Shaded regions represent cell state classification. (B) Stacked violin plots of average single-cell gene expression for cells presented in Figure S7A. 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 (middle row; Venteicher et al. and 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. (F) Sankey plot representing the proportion of IDHwt tumor cells with pan-glioma classification and associated classification described in Neftel et al. (G) Density plots representing TFBS epimutation burden (scRRBS data) in IDHmut single-cell DNA methylation data for TFs whose activity (scRNAseq based SCENIC analysis) characterizes a specific cell state ( $n = 20$  TFs per cell state). Kolmogorov-Smirnov  $p$ -value tests for differences in TFBS epimutation burden across the cellular states. (H) Density plots representing TFBS epimutation burden (scRRBS data) in IDHwt single-cell DNA methylation data for TFs whose activity (scRNAseq based SCENIC analysis) characterizes a specific cell state ( $n = 20$  TFs per cell state). Dotted lines represent the median TFBS value for cell state defining TFs. The Kolmogorov-Smirnov  $p$ -value corresponds to differences in TFBS epimutation burden across the cellular states.



**Figure S8. LIGER integrated tumor-specific clustering of single-cell RNA and single-cell DNA methylation data. Related to Figure 2.**

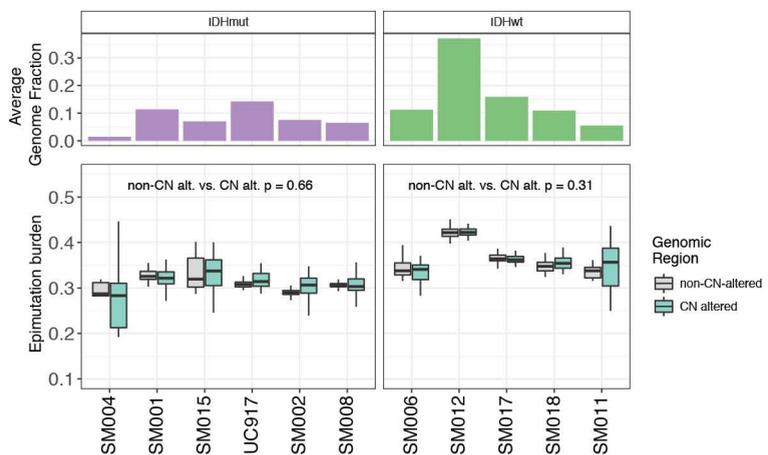
Joint single-cell RNAseq (scRNA) and single-cell DNA methylation (scDNAm) clustering and UMAP projections highlighting similar cellular state distributions across platforms. Sample annotation is presented on the left of each paired UMAP plot, each dot is an individual single cell, and cell number for 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 (Fisher's Exact test,  $p < 0.05$ ).



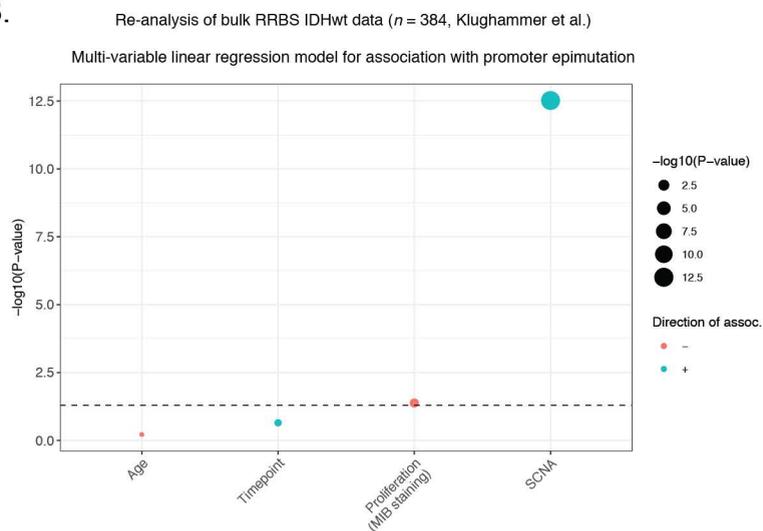
**Figure S9. Sample-specific differences in DNA methylation and epimutation burden across different cellular states. Related to Figure 2.**

(A-B) Boxplots showing sample-specific differences in (A) epimutation burden and (B) 10-kb tiled DNA methylation across LIGER-defined cellular states in IDHmut tumors. Wilcoxon Rank Sum p-values are presented comparing cells from a given tumor. (C-D) Boxplots showing sample-specific differences in (C) epimutation burden and (D) 10-kb tiled DNA methylation across LIGER-defined cellular states in IDHwt tumors. Wilcoxon Rank Sum p-values are presented comparing cells from a given tumor. Samples with only one defined cell state are not visualized.

A.

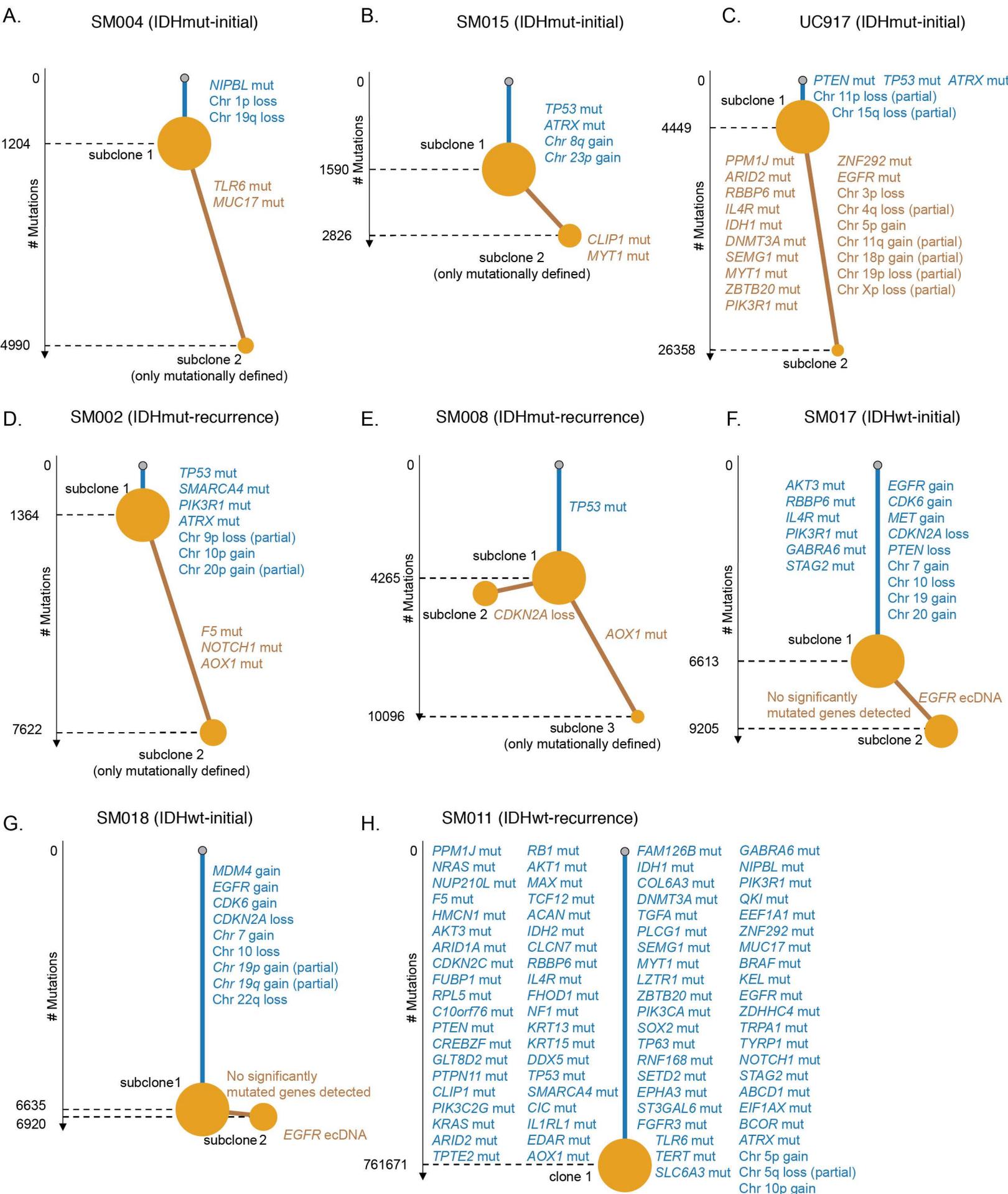


B.



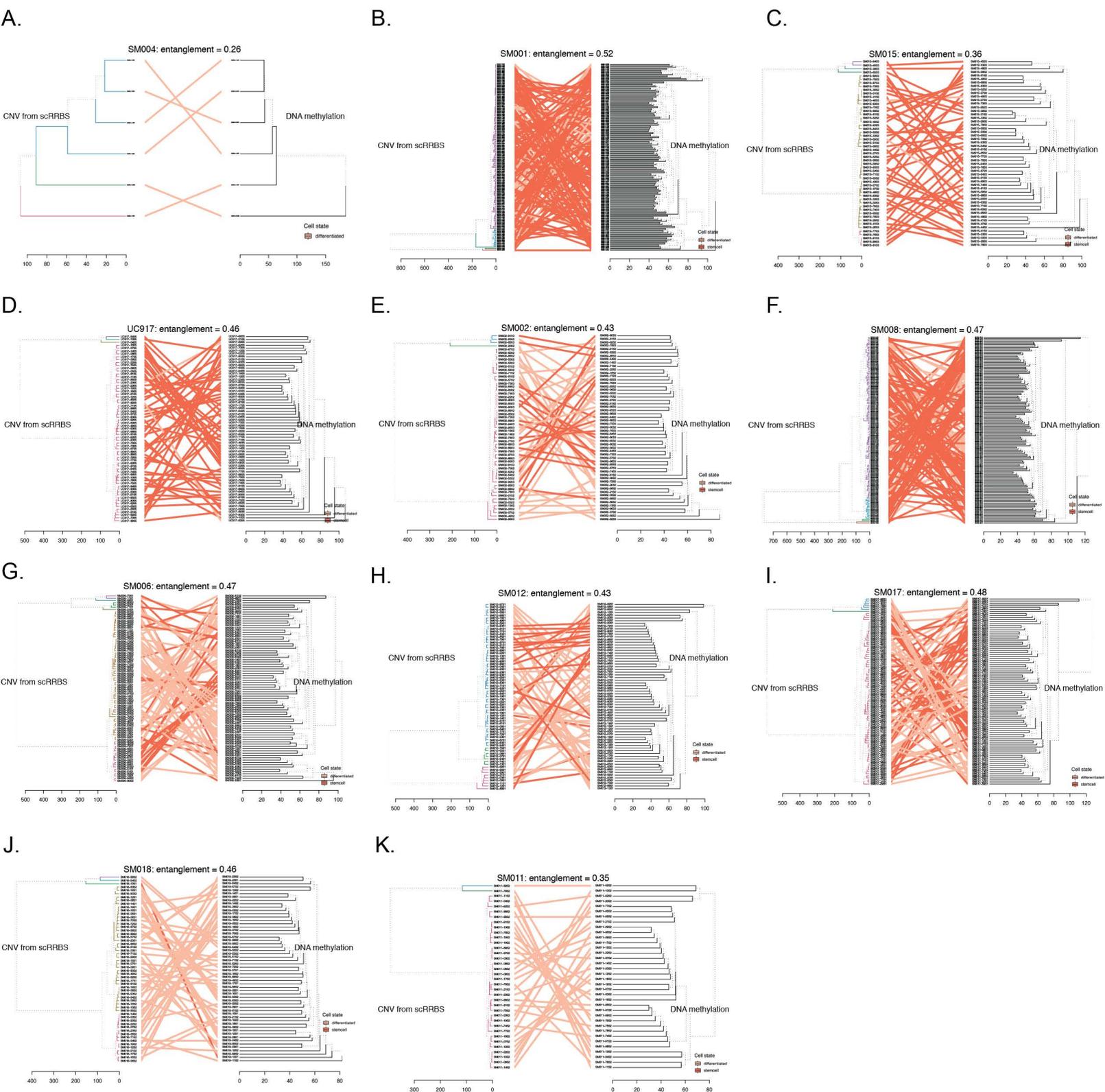
**Figure S10. Relationships between epimutation burden and genetic alterations. Related to Figure 3.**

(A) Single-cell epimutation burden estimates were calculated across genomic regions with (teal) and without (gray) copy number alterations. The paired-sample Wilcoxon test p-value for each subtype represents the statistical difference of epimutation burden across these two regions. (B) Visualized results from multi-variable linear regression model testing for association with epimutation burden. Dot size indicates  $-\log_{10}(p\text{-value})$  for each predictor and color represents direction of association with epimutation burden (red = negative association, blue = positive association). Explanatory variables included subject age, timepoint (pre- and post-treatment), level of cellular proliferation determined by histological marker (MIB staining), and somatic copy number alteration burden (SCNA, total number of bases altered / total number of bases measured).

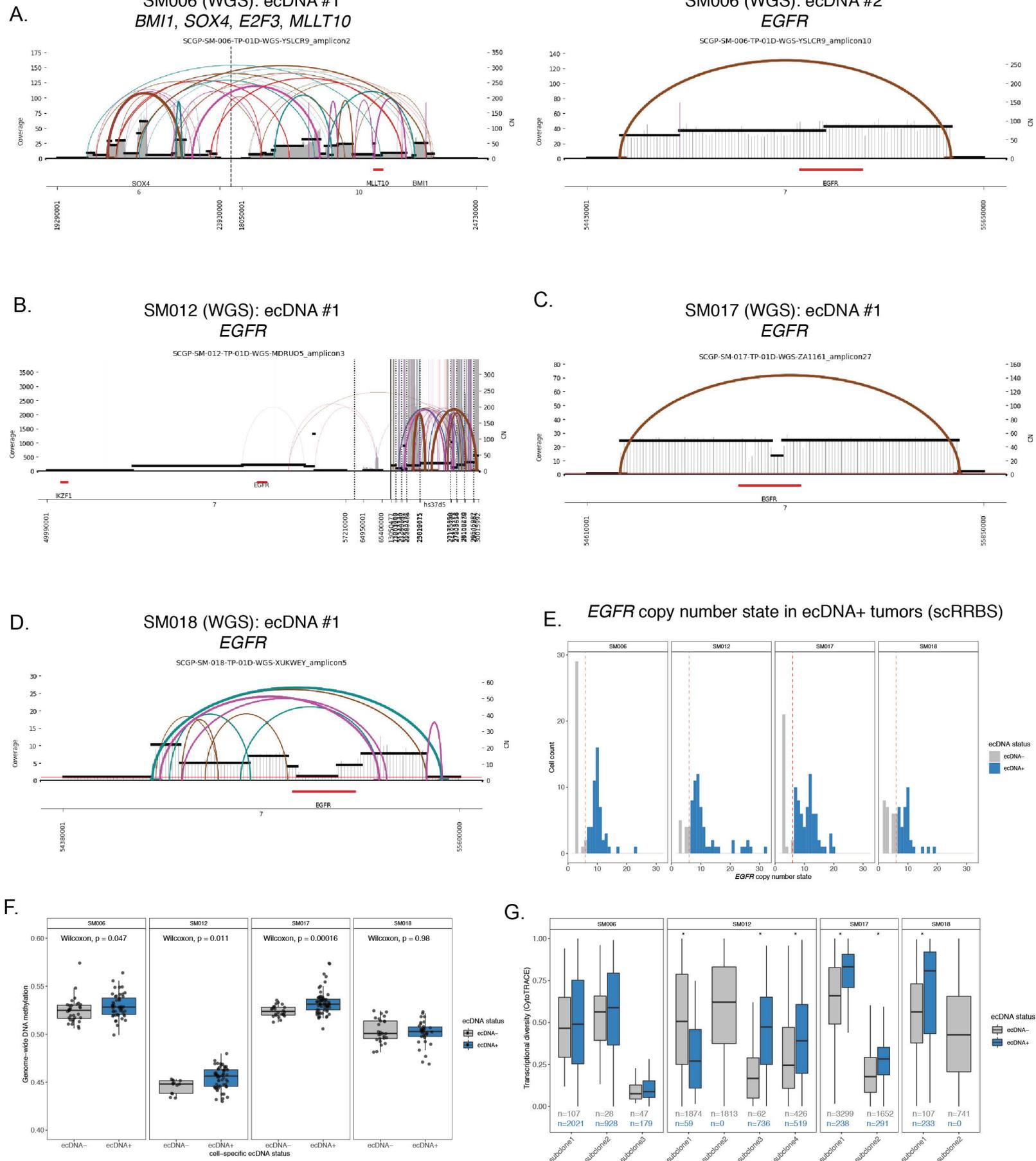


**Figure S11. Whole genome sequencing phylogenetic inference of tumor samples.  
Related to Figure 4.**

(A-H) 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 analyses (Ceccarelli et al., 2016) 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.

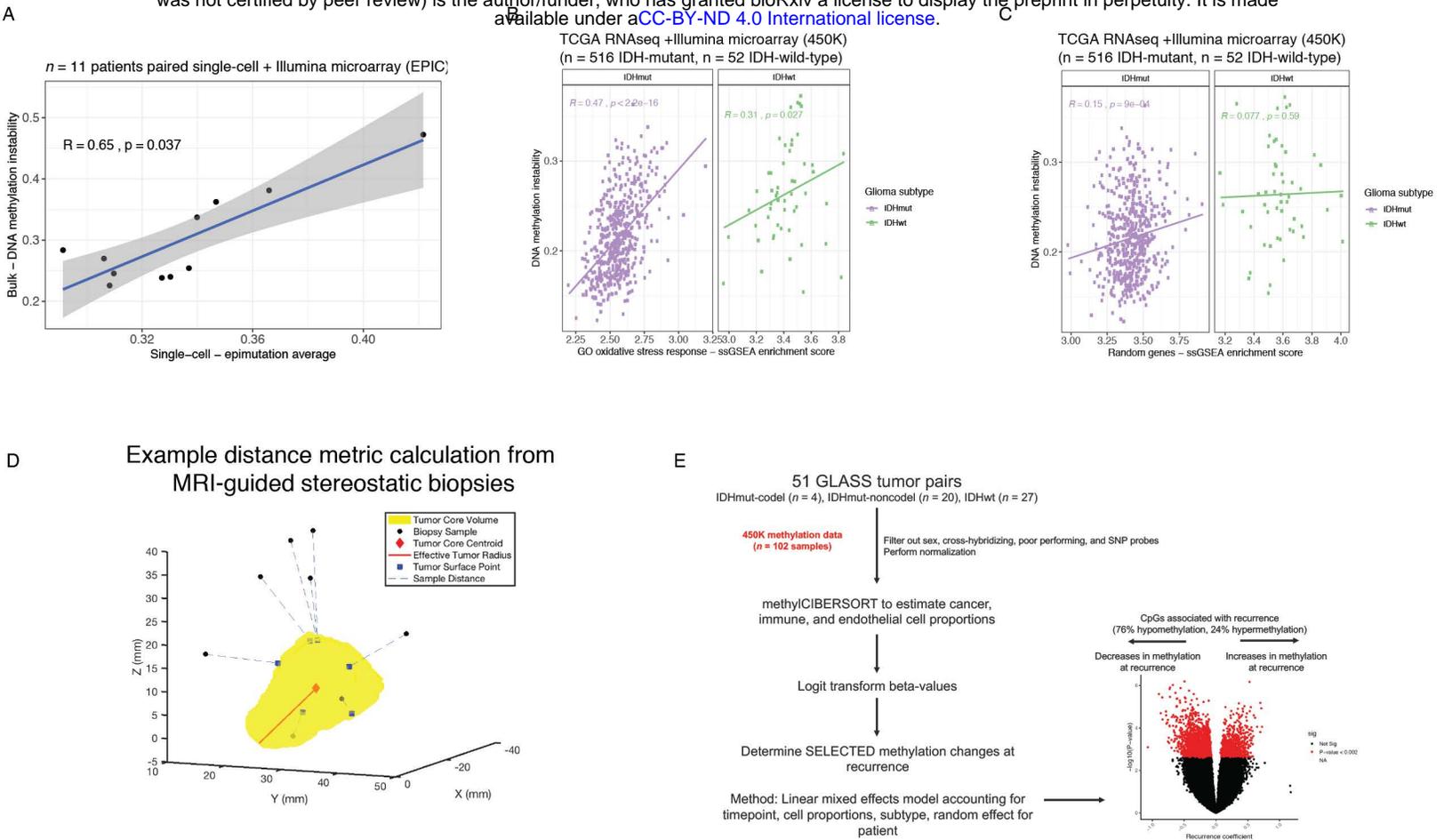


**Figure S12. Tumor-specific comparisons of phylogenetic and phyloepigenetic trees. Related to Figure 4.** (A-K) Tanglegrams highlighting the relationship between single-cell copy number and single-cell DNA methylation tree diagrams. Phylogenetic trees (left cluster) were calculated from copy number profiles (scRRBS data) and phyloepigenetic trees were constructed from the same cells across 10-kb tiled DNA methylation values. Cluster labels are connected with solid lines and are colored by cellular states determined by LIGER. Entanglement scores are listed above the phylogenetic and phyloepigenetic trees and indicate whether labels share the same structure (score = 0) or exhibit unrelated structures (score = 1).



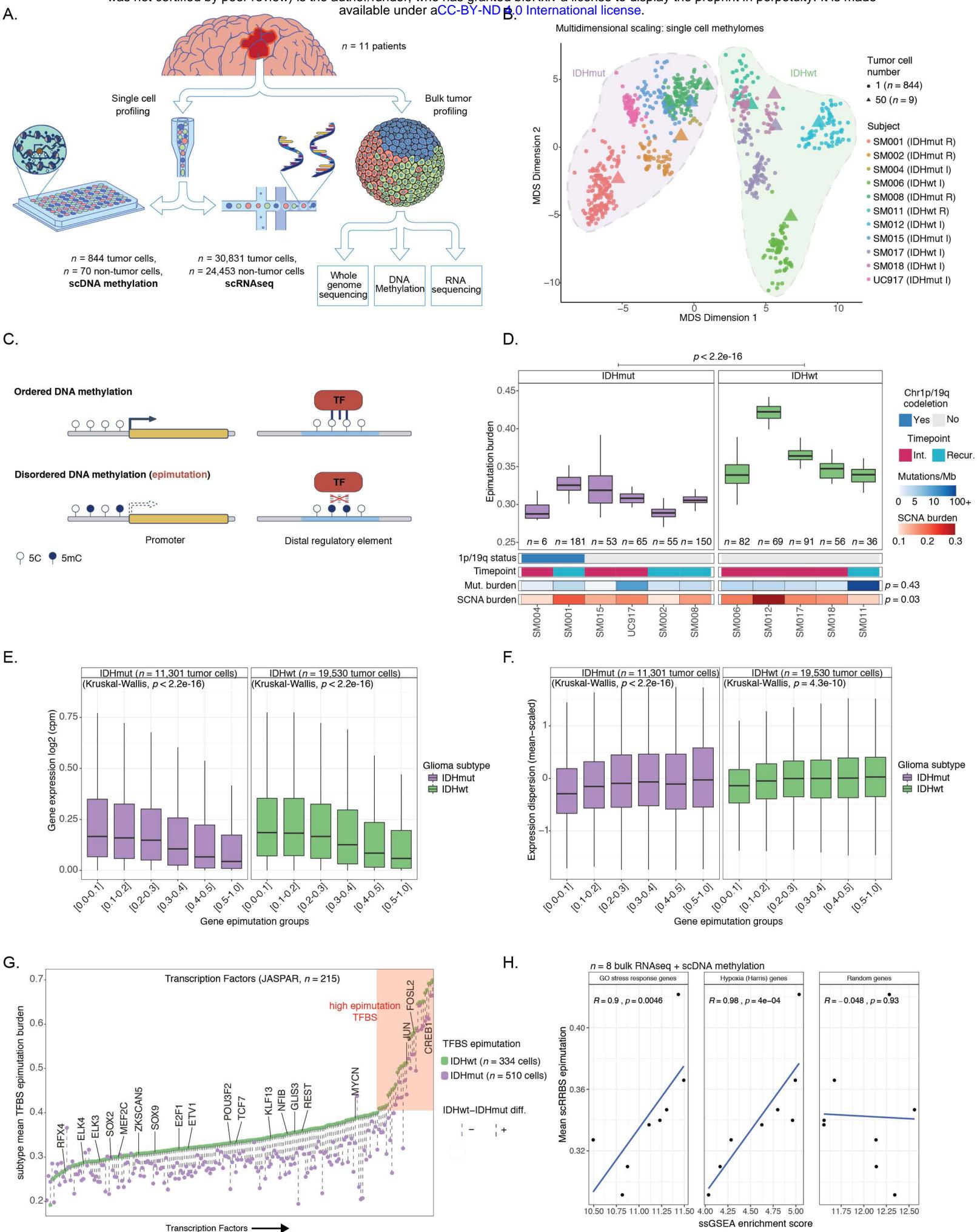
**Figure S13: Focal extrachromosomal DNA amplifications generate greater levels of epigenetic and transcript diversity in glioma single cells. Related to Figure 4.**

(A-D) 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 (e.g., brown = everted read pairs, (Deshpande et al., 2019). Amplicon intervals are provided at the bottom of the plot with annotation for known oncogenes (e.g., *EGFR*). (E) *EGFR* copy number estimation from single-cell RRBS data in ecDNA+ tumors. Cells with *EGFR* copy number greater than 7 were classified as *EGFR* ecDNA+ (blue). (F) Single-cell 10-kb tiled DNA methylation separated by *EGFR* ecDNA status. Single cells with inferred copy number status greater than 7 were classified as ecDNA+ (blue). Wilcoxon rank sum test *p*-values comparing DNA methylation across ecDNA status are reported for each patient tumor. (D) 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 Wilcoxon Rank Sum test ( $p < 0.05$ ).



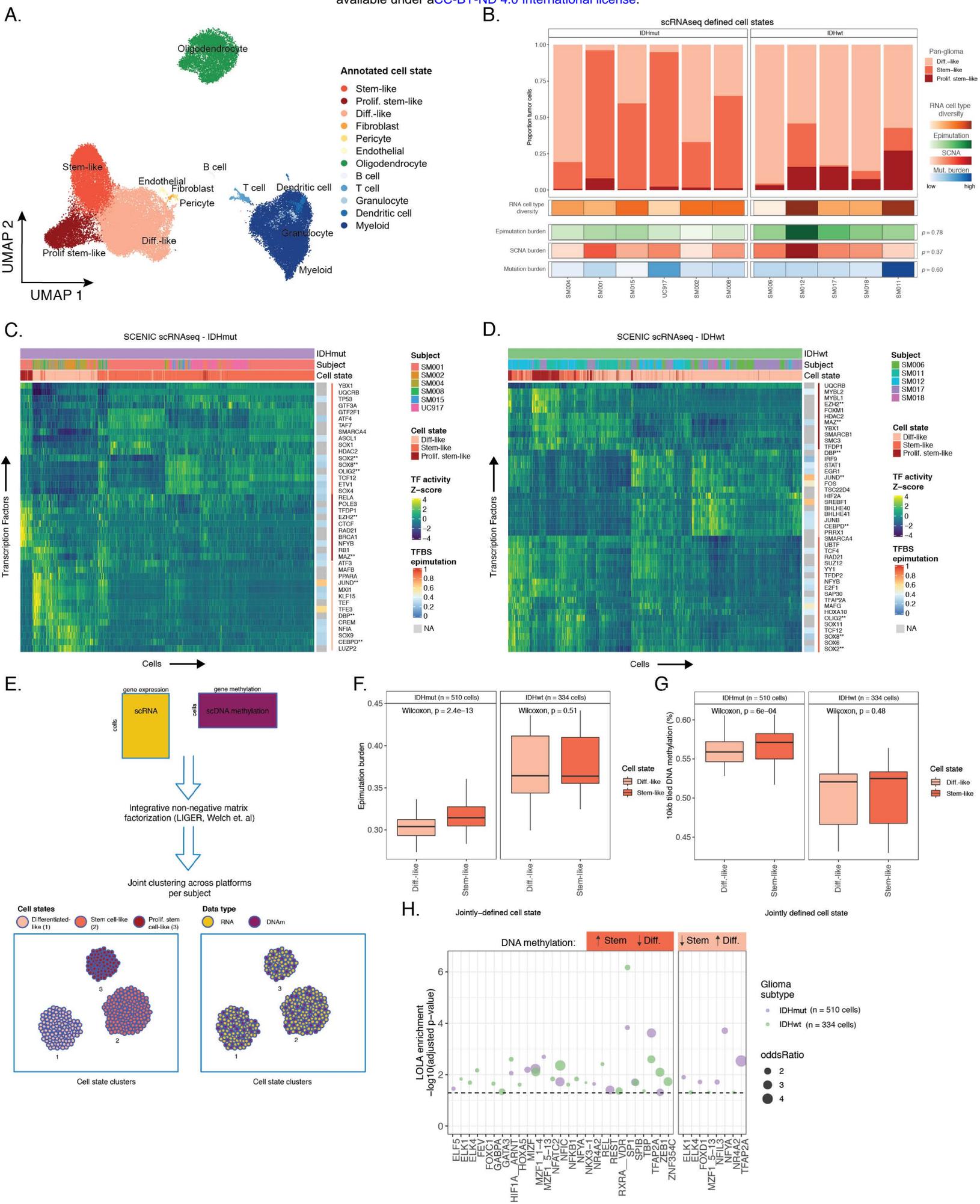
**Figure S14. DNA methylation instability metrics calculated in primary tumor, spatial, and longitudinal cohorts. Related to Figure 5.**

(A) Scatterplot highlights the significant positive correlation between the single-cell epimutation burden metric and the bulk microarray-based DNA methylation instability metric ( $n = 11$  tumors, Spearman correlation). (B-C) Scatterplot between DNA methylation instability and (bulk RNAseq) ssGSEA enrichment scores for (B) oxidative stress response genes and (C) a randomly selected gene set substantiates finding that epigenetic instability is associated with stress response. (D) Schematic depiction of magnetic resonance image-guided biopsies and radiographic features used in spatial cohort (Verburg et al.). (E) Workflow for linear-mixed effect model identifying differentially methylated CpG sites that are selected for during tumor evolution when adjusting for estimated cellular proportions, glioma subtype, and a random effect for patient ( $n = 102$  tumor samples,  $n = 51$  patients).



**Figure 1. Single-cell DNA methylation sequencing highlights intratumoral heterogeneity and disruption of epigenetic regulatory mechanisms.**

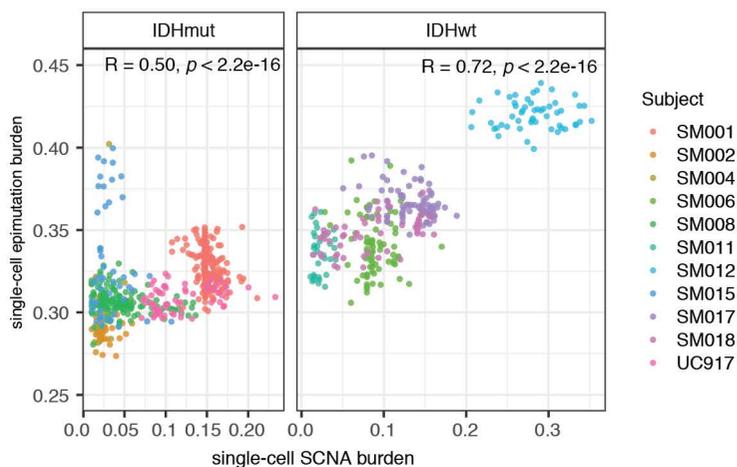
(A) Schematic diagram detailing tumor sample processing and molecular profiling of single cells and bulk tumor samples ( $n = 11$  subjects). (B) Multidimensional scaling (MDS) analysis using pairwise individual CpG distance metrics calculated between individual cells. Shapes represent whether a sample was a single tumor cell ( $n = 844$  cells) or 50-tumor cells,  $n = 9/11$  subjects). Colors indicate individual subjects, shaded regions indicated *IDH1*-mutation status of tumor, and annotation is provided indicating clinical timepoint (I = initial, R = recurrence). (C) Schematic depiction of local DNA methylation disorder in different genomic contexts. Left panel demonstrates epimutation, or local DNA methylation disorder, at the promoter region, where gene expression is disrupted by epimutation. The right panel provides an example of disrupted transcription factor binding due to epimutation. (D) Boxplots of tumor cell epimutation burden grouped by sample. Each boxplot spans the interquartile range with the whiskers representing the absolute range, excluding outliers. Wilcoxon rank sum  $p$ -value represents comparison between IDHmut and IDHwt epimutation burden. Each sample is annotated with clinical and molecular metrics with  $p$ -values indicating the relationship between sample mean epimutation burden and whole-genome sequencing derived somatic mutation burden or somatic alteration burden (Spearman correlation). (E) Boxplots of gene expression values, as  $\log_2$  (counts per million), from single-cell RNAseq data across different gene epimutation groups. Gene epimutation groups are defined by the determining the mean epimutation value across a single gene. Color indicates *IDH1* mutation status. (F) Boxplots of gene expression dispersion. Expression profiles were mean-expression scaled to account for expression level-dependent variability across the same gene epimutation groups defined in panel E. (G) Scatterplot of the mean single-cell epimutation burden metric calculated across transcription factor binding sites (TFBSs) within a subtype, ordered by IDHwt TFBS epimutation. Each column represents a single transcription factor (TF) with a colored dotted line connecting IDHmut and IDHwt values. Names of TFs previously indicated to confer fitness advantages to glioma cells (MacLeod et al.) are listed above their TFBS epimutation burden estimate. (H) Scatterplot depicting the association between average single-cell epimutation burden 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.



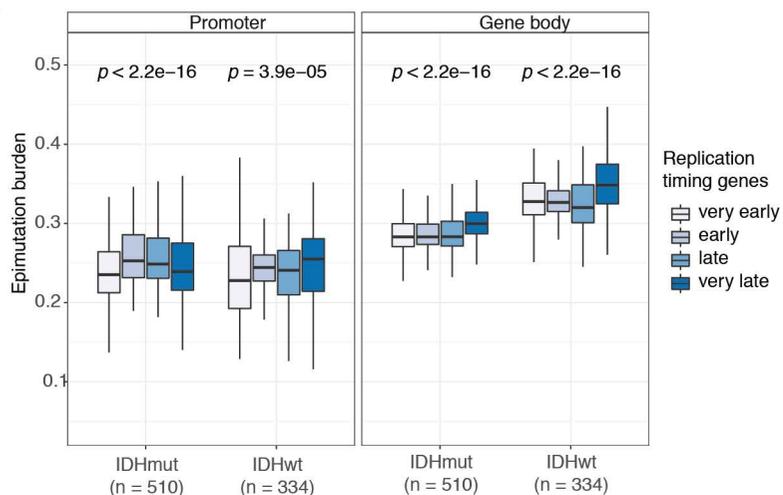
**Figure 2. Integrative single-cell gene expression and DNA methylation analyses nominate epigenetic regulators of glioma cell state variability.**

(A) Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction plot of scRNAseq data ( $n = 55,284$  tumor cells,  $n = 11$  subjects) showing the clustering of cell populations by transcriptionally defined cell state (point color) and labelled according to marker gene expression (Figure S6B). (B) Stacked bar plots representing the proportion of cellular states per tumor for pan-glioma 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 epimutation burden, whole-genome sequencing derived somatic alteration burden, or whole-genome sequencing derived somatic mutation burden (Spearman correlation). (C-D) Enriched transcription factor activity across pan-glioma cellular states determined using SCENIC algorithm and displayed as a heatmap of z-score enrichment values. Visualization is presented for the hierarchical clustering of 5,000 randomly selected tumor cells in both (C) IDHmut and (D) IDHwt tumors. (E) Schematic diagram representing LIGER workflow to jointly cluster single-cell RNAseq and DNA methylation data generated from the same tumor dissociation. (F) Boxplots representing the average epimutation burden in differentiated-like and stem-like populations in IDHmut (left panel) and IDHwt (right) tumors. (G) Boxplots representing the 10-kb tiled DNA methylation levels in differentiated-like and stem-like populations in IDHmut (left panel) and IDHwt (right) tumors. (H) Region set enrichment analysis for 10-kb tiles with higher DNA methylation in Stem-like (left panel) or Differentiated-like cells (right panel). Enrichment was determined by Locus Overlap Analysis (LOLA). Individual points represent enrichment of specific TFs in differentially methylated regions, color indicates results for specific IDH-mutant subtype, point size indicates log-odds ratio, and dotted line represents the statistical significance threshold (adjusted  $p$ -value < 0.05).

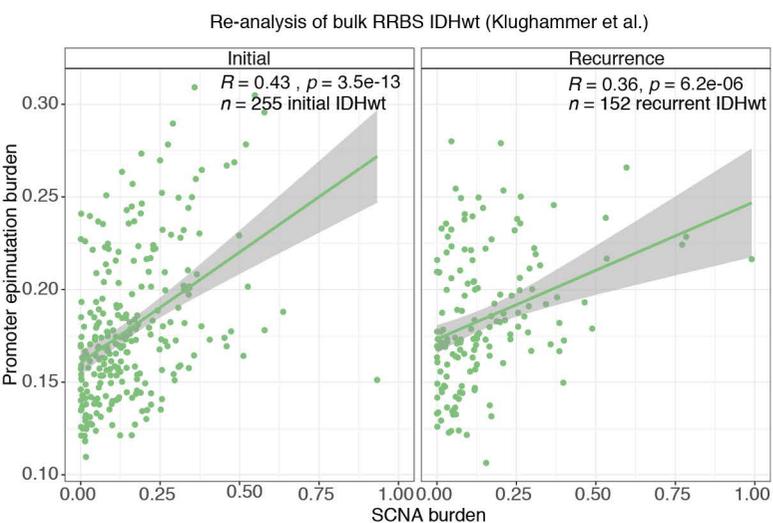
A.



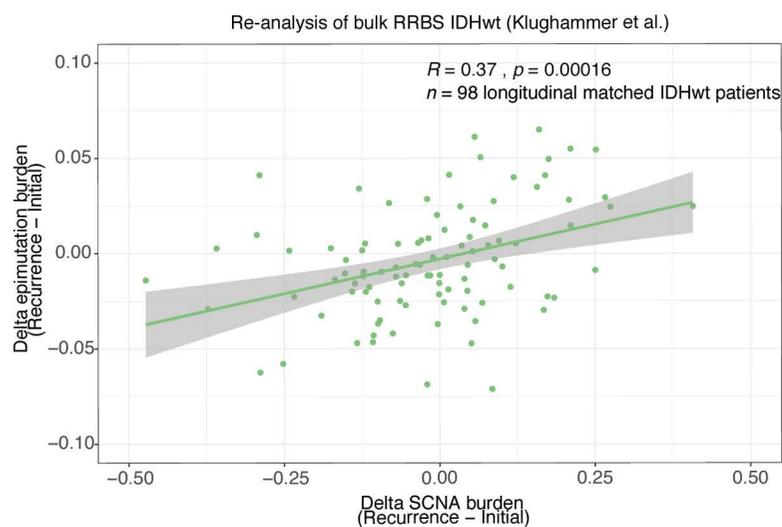
B.



C.

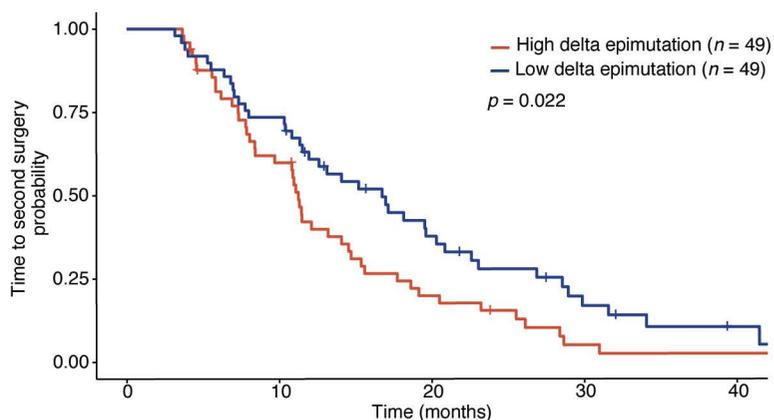


D.



E.

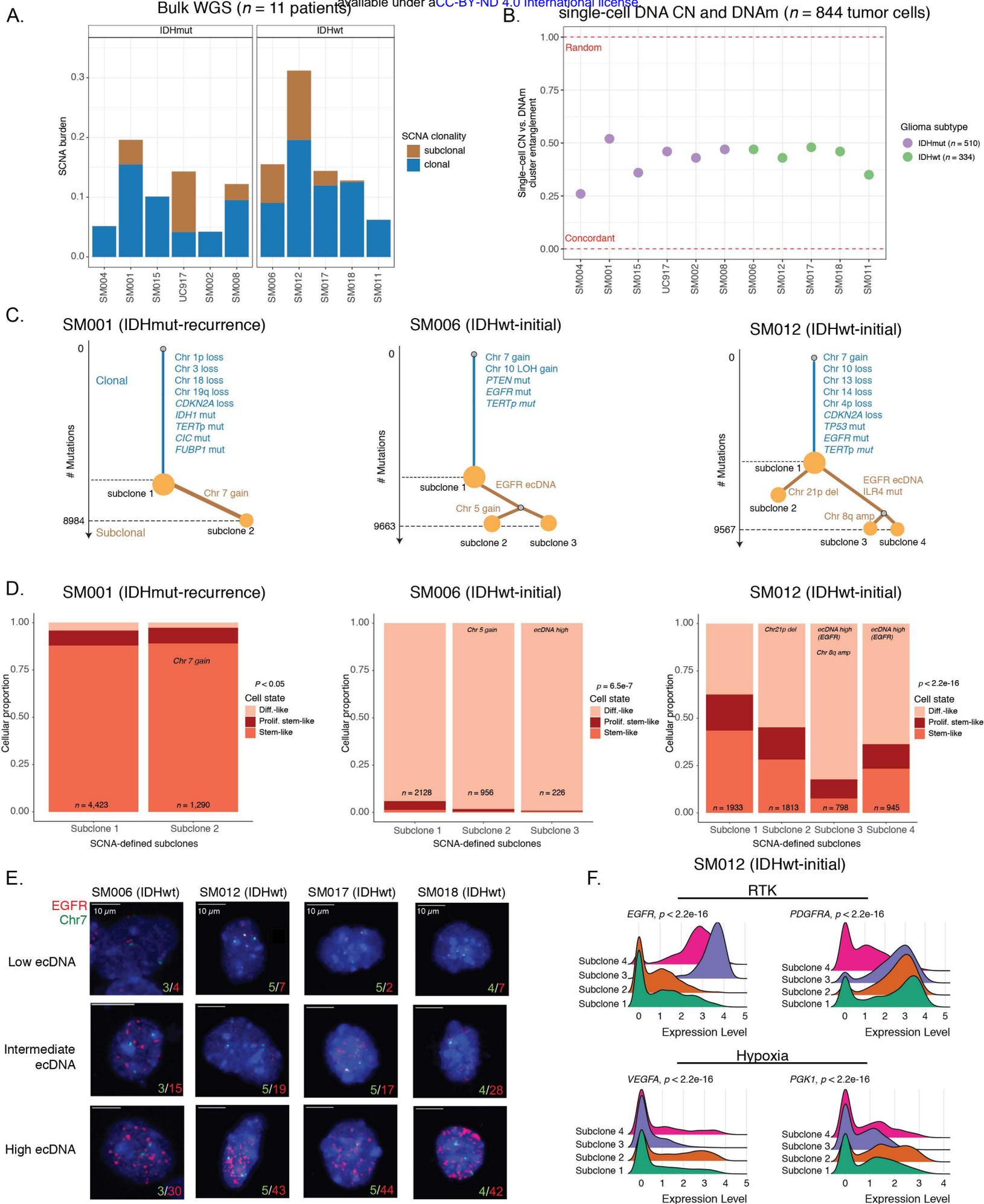
Re-analysis of bulk RRBS IDHwt (Klughammer et al.)



\*\*Epimutation mutation group was significantly associated with time to second surgery in multivariate Cox model with Age and Sex (HR = 1.69 (95% CI 1.09 - 2.62),  $p = 0.02$ )

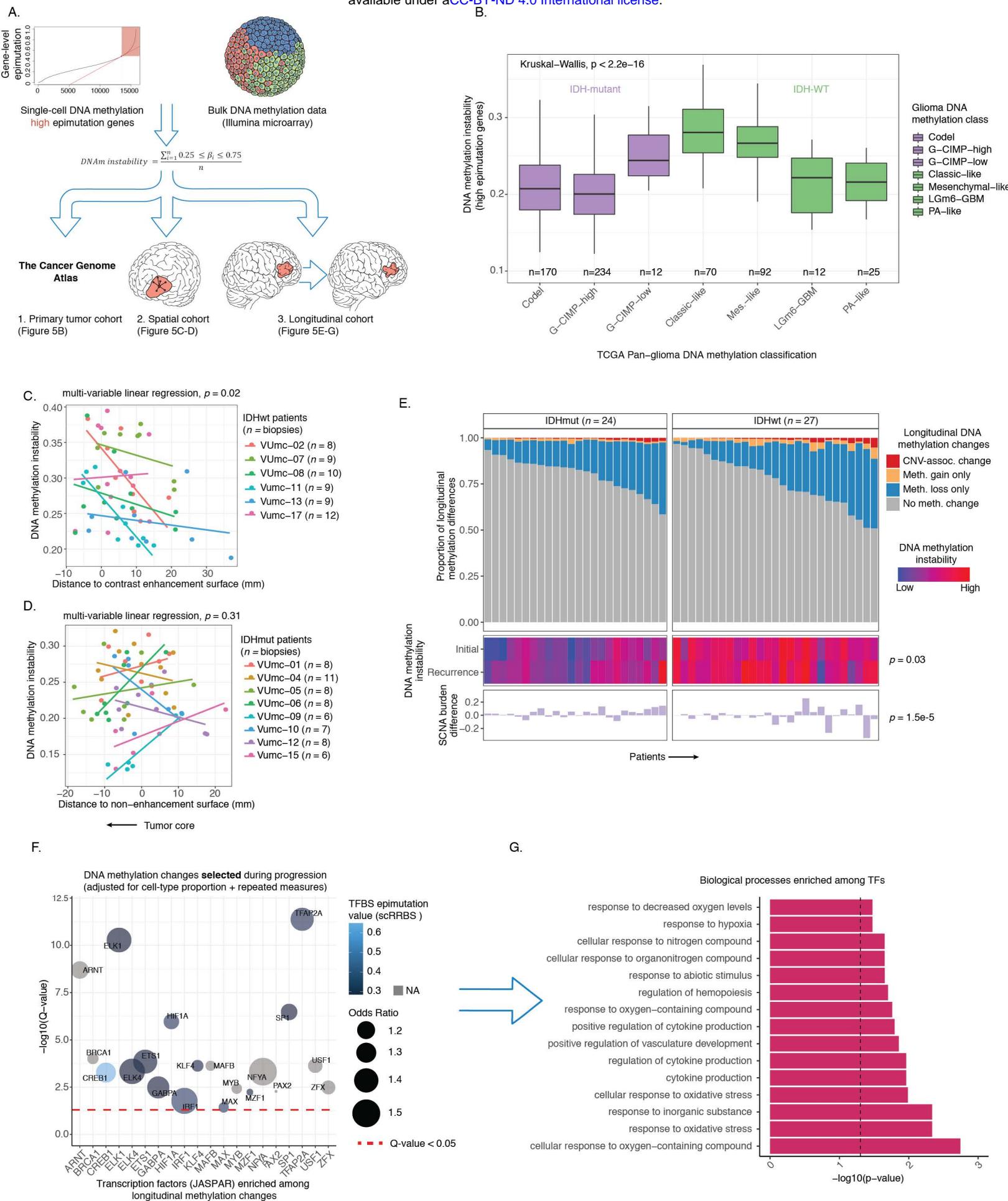
**Figure 3. Somatic copy number alterations are associated with stochastic DNA methylation changes during disease evolution.**

(A) Scatterplot depicting the association between single-cell ( $n = 844$  tumor cells) somatic copy number alteration (SCNA) and epimutation burden estimates by IDHmut (left panel) and IDHwt (right panel) subtypes. Points are colored by patient. Spearman correlation coefficients represent subtype-specific estimates. (B) Boxplots of epimutation burden calculated across the promoter (left panel) and gene body regions (right panel) based on different DNA replication times in IDHmut ( $n = 510$ ) and IDHwt ( $n = 334$ ) single cells. Kruskal-Wallis  $p$ -values indicate a test for differences across the replication time groupings separately for IDHmut and IDHwt cells. (C) Scatterplot depicting the re-analysis of bulk promoter epimutation burden and SCNA burden in IDHwt initial ( $n = 255$ ) and recurrent ( $n = 152$ ) tumors (Klughammer et al.). Spearman correlation coefficients and  $p$ -values are presented for each independent timepoint. (D) Scatterplot depicting the association between bulk delta (subject-specific recurrence – initial estimates) SCNA burden and delta promoter epimutation burden in longitudinally profiled IDHwt tumors ( $n = 98$  subjects, Klughammer et al.) Spearman correlation coefficient and  $p$ -value are presented. (E) Kaplan-Meier curve depicting time to second surgery in subjects where the change in epimutation burden between initial and recurrent disease was above (high, red) and below (low, blue) the median. Log-rank  $p$ -value for univariate analysis is presented within the figure. Hazard Ratio and  $p$ -value for change in epimutation burden are presented below for multi-variate Cox proportional hazard model including subject age and sex as predictors.



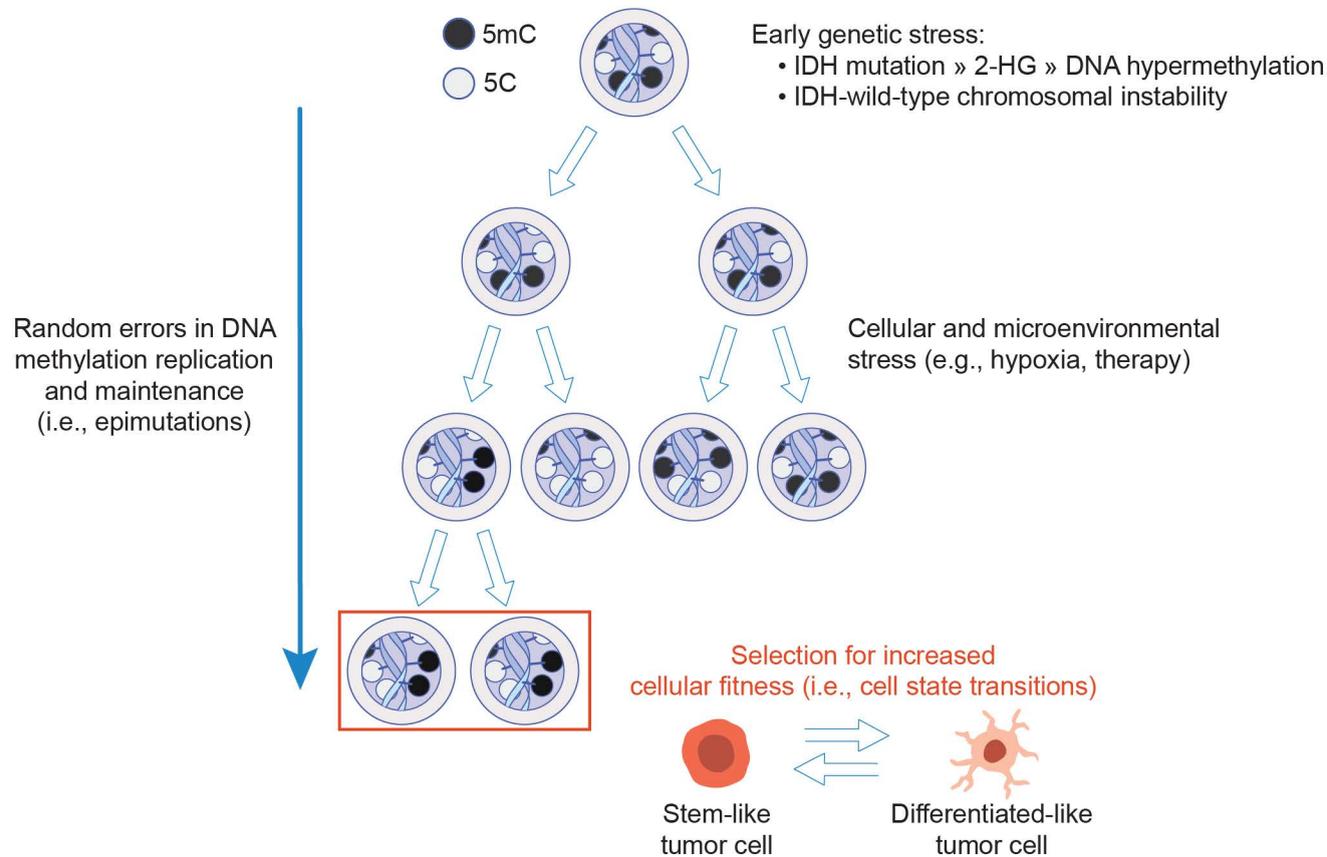
**Figure 4. Clonal evolution analyses highlight early copy number evolution followed by epigenetic and transcriptomic diversification.**

(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) Scatterplot depicting entanglement coefficients for tanglegrams comparing cluster dendrograms of scRRBS derived copy number and DNA methylation profiles. A coefficient of 0 indicates complete alignment of the tree structures, whereas a 1 indicates random association. Color indicates *IDH1* mutation status. (C) Examples of phylogenetic trees constructed from whole genome sequencing data (mutations and SCNAs) 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 analyses (Ceccarelli et al., 2016) and chromosomal arm-level events are presented. (D) Single-cell RNAseq-derived cellular proportions separated by copy number-defined tumor subclone (Figure S3). Reported *p*-values represent Fisher's exact test comparing the cellular state distributions across tumor subclones. (E) Representative Fluorescence *in situ* hybridization (FISH) images for IDHwt tumors computationally predicted to harbor *EGFR* extrachromosomal DNA (ecDNA) by whole genome sequencing (*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 microns. (F) Ridge plots of SM012 single-cell expression of receptor tyrosine kinase and hypoxia-associated genes, grouped by copy number-defined subclones. Reported *p*-values represent Wilcoxon Rank Sum tests comparing the gene expression of cells across tumor subclones.



### Figure 5. Integrated molecular trajectories supports adaptive DNA methylation changes under microenvironmental and therapeutic pressures.

(A) Schematic workflow for construction of a DNA methylation instability metric in bulk cohorts informed by regions of high epimutation in single-cell DNA methylation data. The DNA methylation instability metric was calculated across bulk DNA methylation microarray data in a primary tumor cohort (TCGA), a cohort of multiple, spatially distinct biopsies from the same tumor (Verburg et al.), and a longitudinal cohort with accompanying genomic sequencing data (Glioma Longitudinal AnalySiS (GLASS), Barthel et al.). (B) Boxplots displaying the bulk DNA methylation instability metric calculated across previously described DNA-methylation based TCGA tumor classifications (Ceccarelli et al). Colors represent *IDH1/2* mutation status, and Kruskal-Wallis  $p$ -value testing for differences in distributions across classification is reported ( $n = 615$  primary gliomas,  $p < 2.2e-16$ ). (C-D) Scatterplots depicting distance from radiographic features plotted against the DNA methylation instability metric. Colors represent spatially separated biopsies from a single patient at initial clinical timepoint for (C) IDHwt tumors ( $n = 57$  biopsies,  $n = 6$  subjects) and (D) IDHmut tumors ( $n = 62$  biopsies,  $n = 8$  subjects). Linear regression lines colored by patient demonstrate the relationship between DNA methylation instability and radiographic features (i.e., contrast enhancement surface). The  $p$ -value reported from a multivariable linear regression model adjusting for subject represents the subtype-specific association between DNA methylation instability and radiographic feature. Biopsies taken closer to the tumor's center (i.e., core) have the lowest value (left hand side of plot). (E) Each column represents an individual patient sampled across initial and recurrent timepoints and is separated into IDHmut ( $n = 24$  subjects) and IDHwt ( $n = 27$  subjects). Top panel, stacked bar plot represents the proportion of CpGs sites that experienced DNA methylation change associated with a subject-specific copy number change (defined by DNA sequencing data) between primary and recurrent disease (red), DNA methylation gain not associated with a CNV change (orange), DNA methylation loss not associated with a CNV change (blue), and no longitudinal DNA methylation change (gray). Middle panel, heatmap of DNA methylation instability metric in primary and recurrent disease (blue = low, red = high). Bottom panel, differences in SCNA burden between primary and recurrent tumor. All associated  $p$ -values represent Spearman correlations between absolute change in associated metric and the fraction of longitudinal DNA methylation differences. (F) Enrichment analysis for differentially methylated CpGs between primary and recurrent timepoints when adjusting for cellular composition, glioma subtype, and subject included as a random effect. Individual points represent enrichment of specific TFs in differentially methylated positions, color indicates the average TFBS epimutation burden from single-cell RRBS data (Figure 1G), point size indicates log-odds ratio, and dotted line represents the statistical significance threshold ( $Q$ -value  $< 0.05$ ). (G) Gene Ontology enrichment of transcription factors associated with longitudinal DNA methylation changes. Dotted line represents threshold for statistical significance (Fisher's exact test,  $p < 0.05$ ).



**Figure 6. Model of epigenetic heterogeneity and tumor evolution.**

Schematic depiction of tumor evolution with general DNA methylation patterns represented by methylated (5-methylcytosine, 5mC) and unmethylated (5C) regions of the epigenome. Initiating genetic events such as *IDH1* and other driver mutations as well as somatic copy number alterations represent early stresses in glioma evolution that precipitate epigenetic heterogeneity. Both mutations in epigenetic enzymes and SCNAs can increase the likelihood of heritable DNA methylation alterations (i.e., epimutations). *IDH1* mutations result in the production of the oncometabolite 2-Hydroxyglutarate (2-HG) that leads to failure to remove aberrant DNA methylation while SCNAs can generate mitotic stress leading to the erosion of ordered DNA methylation. Non-genetic determinants further shape epigenetic heterogeneity as tumors evolve by exposing cells to spatially distinct microenvironmental stresses that impact the DNA methylation replication machinery. The subsequent epigenetic diversity provides an additional layer on which clonal evolution acts to select those cells with fitness-conferring epigenetic alterations. Ultimately, the loosened epigenetic control allows tumor cells to transition to cell states responsive to different selective pressures.