

1 Detection of glioma and prognostic subtypes by non-invasive circulating cell-free DNA
2 methylation markers

3

4 Noushmehr H^{1,2,^}, Sabedot TS^{1,2,*}, Malta TM^{1,2,*}, Nelson K^{1,*}, Snyder J^{1,2,*}, Wells M¹,
5 deCarvalho A¹ Mukherjee A³, Chitale D³, Mosella M^{1,2}, Asmaro K^{1,2}, Robin A¹, Rosenblum M¹,
6 Mikkelsen T¹, Rock J¹, Poisson LM⁴, Lee I¹, Walbert T¹, Kalkanis S¹, Castro AV^{1,2,*}

7

8 ¹Department of Neurosurgery, Hermelin Brain Tumor Center, Henry Ford Health System,
9 Detroit, MI, USA

10 ²Department of Neurosurgery, Omics Laboratory, Henry Ford Health System, Detroit, MI, USA

11 ³Department of Pathology, Henry Ford Health System, Detroit, MI, USA

12 ⁴Department of Biostatistics, Henry Ford Health System, Detroit, MI, USA

13

14 *Contributed equally as second author

15

16 ^Corresponding author

17 Houtan Noushmehr, PhD

18 Associate Scientist/Professor

19 Department of Neurosurgery

20 Hermelin Brain Tumor Center

21 Henry Ford Health System

22 2799 West Grand Blvd, E&R 3096

23 Detroit, MI, 48202

24 houtan.noushmehr@hfhs.org

25

26

27 **SUMMARY**

28 Genome-wide DNA methylation profiling has shown that epigenetic abnormalities are
29 biologically important in glioma and can be used to classify these tumors into distinct prognostic
30 groups. Thus far, DNA profiling has required surgically resected glioma tissue; however,
31 gliomas release tumoral material into biofluids, such as blood and cerebrospinal fluid, providing
32 an opportunity for a minimally invasive testing. While prior studies have shown that genetic and
33 epigenetic markers can be detected in blood or cerebrospinal fluid (e.g., liquid biopsy [LB]),
34 there has been low sensitivity for tumor-specific markers. We hypothesize that the low
35 sensitivity is due to the targeted assay methods. Therefore, we profiled the genome-wide CpG
36 methylation levels in DNA of tumor tissue and cell-free DNA in serum of glioma patients, to
37 identify non-invasive epigenetic LB (eLB) markers in the serum that reflect the characteristics of
38 the tumor tissue. From the epigenetic profiles of serum from patients diagnosed with glioma
39 (N=15 *IDH* mutant and N=7 *IDH* wildtype) and with epilepsy (N=3), we defined glioma-specific
40 and *IDH*-specific eLB signatures (Glioma-eLB and *IDH*-eLB, respectively). The epigenetic
41 profiles of the matched tissue demonstrate that these eLB signatures reflected the signature of the
42 tumor. Through cross-validation we show that Glioma-eLB can accurately predict a patient's
43 glioma from those with other neoplasias (N=6 Colon; N=14 Pituitary; N=3 Breast; N=4 Lung),
44 non-neoplastic immunological conditions (N=22 sepsis; N=9 pancreatic islet transplantation),
45 and from healthy individuals (sensitivity: 98%; specificity: 99%). Finally, *IDH*-eLB includes
46 promoter methylated markers associated with genes known to be involved in glioma
47 tumorigenesis (*PVT1* and *CXCR6*). The application of the non-invasive eLB signature discovered
48 in this study has the potential to complement the standard of care for patients harboring glioma.

49

50 INTRODUCTION

51 Gliomas are a heterogenous group of intracranial tumors that are constantly evolving, generally
52 recur, and frequently progress to more malignant subtypes. Recently, genomic and epigenomic
53 alterations have defined subtypes of glioma (e.g., *IDH* mutation, 1p19q chromosomal deletion,
54 and Glioma-CpG Island Methylator Phenotype [G-CIMP]) with distinct prognostic outcomes ^{1–6}.
55 Currently, this molecular diagnosis and classification, which guides clinical management, is
56 dependent on tissue profiling obtained by invasive surgical approaches (tissue biopsy or
57 excision). However, this surgery-centered approach does not allow serial tissue evaluation to
58 capture the dynamic molecular evolution of these tumors, may not be feasible in surgically
59 inaccessible tumors, requires the risk of an invasive procedure in an often comorbid population,
60 and may delay the diagnosis of this disease to later stages due to procedure risks and diagnostic
61 sensitivity. MRI is a relevant non-invasive approach to diagnose and follow patients with glioma;
62 however, limitations remain for differential diagnosis (e.g., lymphoma), detection of minimal or
63 remnant tumoral burden, and in distinguishing progression from pseudo-progression caused by
64 radiation-induced necrosis or treatments such as immunotherapy ^{7,8}. In addition, serial
65 assessments may be costly and cumbersome procedures for patients. Therefore, the discovery of
66 a minimally or non-invasive approach that allow earlier identification of sensitive and specific
67 molecular biomarkers that reflect tumor burden and its dynamic evolution in real-time is
68 desirable. An approach that meets the above criteria is liquid biopsy (LB) of biofluids (e.g.,
69 blood, and cerebrospinal fluid [CSF]) which detect materials shed by the tumors such as
70 circulating tumor cells and genomic specimens (e.g., circulating tumor DNA) ^{9,10}.
71 In the past decade, investigation of the diagnostic, prognostic and predictive applications of LB
72 throughout a patient's disease course has been feasible in many tumors ^{11–16}. For instance, in

73 central nervous system (CNS) neoplasms, including gliomas, CSF has been a relevant source of
74 molecular markers^{10,17–29} and can be used to track the glioma tumor evolution³⁰. However,
75 obtaining CSF is an invasive, complex and risky procedure which can cause, for instance,
76 brainstem herniation due to increased intracranial pressure and/or risk of bleeding due to
77 thrombocytopenia caused by chemotherapy in patients harboring CNS tumors. In addition, serial
78 assessment of CSF markers throughout a person's disease follow-up as standard clinical
79 management raises additional concerns of patient compliance, impact on quality of life,
80 feasibility to perform test (e.g., patients on anticoagulation or those unable to lie flat) and
81 additional risks secondary to repeated use of this procedure. In contrast, blood LB is minimally
82 invasive, quick and feasible to perform longitudinally; however, one limitation of blood-derived
83 LB is the dismal and often low yield of molecular material released into the blood by CNS
84 tumors (likely due to the blood brain barrier) which may hinder the detection of molecular
85 features, such as specific and rare tumor mutations or novel and clinically relevant molecular
86 markers shed by the tumor^{10,21,31,32}. To overcome these limitations, the performance of a more
87 comprehensive “omics” approach (e.g., DNA methylomic or genomic) has been proposed^{10,33}
88 and implemented for certain cancer types^{31,34}. Although genomic LB associated with gliomas
89 has been performed^{10,26}, the anticipated success of highly specific LB markers has been
90 hampered by the low mutation frequency of the associated glioma tumor (0.5-2.6%)^{3,35}. In
91 addition, genetic alterations are generally site specific (e.g., frame-shift, point mutation) which
92 may limit their detection in the fragmented DNA released into the circulation³⁶. The low
93 likelihood of detecting at least one of the 75 significantly mutated genes (somatic)³ associated
94 with glioma in any cell-free DNA assay renders the feasibility of genomic-specific LB a
95 challenge. On the other hand, DNA methylation is a stable marker that is tissue specific,

96 clinically relevant to gliomas, and altered in large regions of the genome. Thus, DNA methylome
97 profiling is an attractive approach for the identification of diagnostic, prognostic and predictive
98 markers in LB^{34,37}. In the tumoral tissue, genome-wide DNA methylation profiling has shown
99 that epigenetic abnormalities play important biological and clinical roles in CNS tumors,
100 particularly in gliomas^{3,38–41}. For instance, G-CIMP is a subset of glioma with extensive
101 epigenomic alterations that confers a stronger prognostic value than age, grade and histology
102^{1,3,41}. However, the relevance of comprehensive DNA methylation profiling in the blood-based
103 circulating free tumor DNA of patients with gliomas has not been explored. Herein, we
104 hypothesized that genome-wide methylome profiling is a useful approach to identify methylome-
105 specific markers associated with glioma in cfDNA. To address this hypothesis, we profiled the
106 genome-wide CpG methylation landscape of matching serum and tissue from 22 patients
107 diagnosed with gliomas (N=15 *IDH* mutant and N=7 *IDH* wildtype) and from 3 patients with a
108 non-neoplastic brain disease (i.e., epilepsy). We identified a set of epigenetic signatures in the
109 serum LB (herein referred to as eLB) that resembles the tissue epigenomic landscape associated
110 with glioma. We showed that the eLB could differentiate glioma from non-tumoral brain tissue
111 and stratify gliomas based on prognostic class (e.g., *IDH* mutation status). We further observed
112 that the specificity of the eLB allowed accurate discrimination of patients with glioma from
113 patients with tumors of other origins and from patients with immune-related disease states
114 (pancreatic islet transplantation and sepsis). The *IDH*-eLB signature includes promoter
115 methylated markers associated with genes known to be involved in glioma tumorigenesis (e.g.,
116 *PVT1* and *CXCR6*). Finally, we propose a novel clinical approach to apply the eLB panels to
117 complement the standard of care in the diagnosis and follow-up. The ability to monitor patients

118 by eLB has the potential to improve the pre- and post-surgical quality of care for patients
119 harboring gliomas.

120 **RESULTS**

121 **Glioma cell-free DNA methylome**

122 In this study, we selected 22 matching pairs of primary glioma tissue and serum, stored at the
123 Hermelin Brain Tumor Center (HBTC) bank from patients who underwent neurosurgery at the
124 Henry Ford Health System, Detroit, Michigan. Serum was collected immediately prior to dura
125 incision at the time tumor tissue was surgically resected. According to the current World Health
126 Organization (WHO) 2016 criteria for glioma classification, our HBTC cohort comprised 3
127 grade IV, 11 grade III, and 8 grade II gliomas among which 15 *IDH* mutants (*IDHmut*) and 7
128 *IDH* wildtypes (*IDHwt*) were included (Table 1, Extended Data Table S1). As expected, we
129 observed a significant overall survival difference between patients with *IDHmut* and *IDHwt*
130 tumors (mean (95% CI): 62.62 months (53.72-73.32) *versus* 26.86 months (8.85-33.6), Extended
131 Data Fig. S1A). Total extracted serum cfDNA quantity, normalized to the genomic size
132 (Genomic Equivalents [GE]/ml, see Methods), showed that patients with glioma had
133 significantly higher serum cfDNA in relation to patients who underwent surgery for epilepsy in
134 the absence of tumor (mean \pm s.e.: 12268.8 \pm 9269.1 *versus* 3777.4 \pm 2324.7 GE/ml, respectively,
135 student t-test p-value=0.003216; Fig. 1A). This is consistent with the hypothesis that tumors tend
136 to release more cfDNA. Aligned with the hypothesis that increased cfDNA is associated with
137 aggressive tumors, likely due to brain-blood barrier breakdown^{42,43}, we also observed a trend,
138 albeit not statistically significant, wherein *IDHwt* serum had more cfDNA than *IDHmut*
139 (mean \pm s.e.: 15799.3 \pm 8645.4 vs 10621.2 \pm 9364.9 GE/ml; respectively, student t-test p-

140 value=0.2255, Extended Data Fig. S1B) suggesting that cfDNA is altered during gliomagenesis
141 and may be released more abundantly in more aggressive subtypes.

142 We performed an epigenome-wide profile of the glioma cfDNA using Infinium Human
143 Methylation 850K (HM850K). Filtering and pre-processing steps were taken to align these data
144 with that of serum methylome data from colorectal cancer (N=2)⁴⁴ and pituitary adenoma
145 (N=14, unpublished data), as well as from glioma tissue³ (see Methods). Principal component
146 analysis (PCA) showed a distinct separation between gliomas and non-tumor specimens as well
147 as other neoplasms (Fig. 1B). We estimated the source or content of the released cfDNA, and
148 DNA of the matched tissue specimens, by deconvoluting the methylome using primary non-
149 cancer cell-type methylation-based signatures⁴⁵. We observed that in relation to non-tumor
150 specimens, the methylomes from both serum and tissue of glioma patients tended to present less
151 glial (52% lower proportion on average in tumor compared to non-tumor) and more neuronal
152 cell-related (79% greater proportion on average in tumor compared to non-tumor) epigenetic
153 signatures (Fig. 1C, Extended Data Table S2). Interestingly, for specific signatures related to the
154 immune cells, the serum from glioma patients showed a distinct makeup in relation to their
155 matching tissue and non-tumor serum. For instance, B-cell and CD8 T-cell-related signatures in
156 the glioma-patient serum were lower than in the tissue, but glial-related specimens were higher
157 than the associated non-tumor specimens (1.9 and 22 fold increase in means for serum,
158 respectively; 12 and 130 fold increase in means for tissue, respectively; Fig. 1C). On the other
159 hand, CD4 T-cell-related signatures were depleted in the tumor serum in relation to non-tumor
160 serum (0.38 fold decrease) and undetectable in the tumor and non-tumor tissue. Notably, the
161 serum from both non-tumor and glioma patients included signatures associated with higher
162 proportions of neutrophil- and monocyte-related cell types. Together, these results indicate that

163 the cfDNA methylome contains signatures that are specifically related to glioma as well as to the
164 immune system, which may be related to a response to changes in the tumor microenvironment
165 or release of immune invasion from the glioma tumor.

166 Considering the full methylome profile, we investigated whether glioma subtype-specific
167 epigenetic signatures defined at the tumor tissue level could be detected in the serum of glioma
168 patients. Interestingly, despite evidence that brain tumors release circulating tumor DNA into the
169 blood and that the release mechanism is dependent on a phenotype modification (e.g.,
170 mesenchymal glioma subtype)⁴⁶, the published epigenetic signatures^{3,6} were undetectable in the
171 serum methylome (Fig. 1D & Extended Data Fig. S1C-F). This is consistent with previous
172 findings using whole-genome sequencing that have shown a low detection of tumor genetic
173 hotspot mutations associated with gliomas in the cfDNA^{10,26}. However, given that we observed
174 epigenome-wide differences in the tissue of glioma patients in relation to other tumor types and
175 non-tumor tissue samples (Fig. 1E), it is plausible that other relevant CpG methylation sites, not
176 previously included in glioma marker panels, could be released into the serum by the brain tumor
177 and, consequently, be detectable *via* analysis of the serum methylome.

178 **Glioma eLB derived from cell-free DNA in serum can identify patients with glioma**

179 We performed an epigenome-wide differential analysis to identify specific, serum-based,
180 epigenetic markers associated with glioma compared to non-glioma serum methylome, which we
181 named Glioma-eLB (N=1075 differential CpG sites overlapping autosomal chromosomes,
182 Wilcoxon-rank sum test p-value < 0.001, Extended Data Fig. S2A-B). We confirmed the origin
183 of the Glioma-eLB by evaluating the measured DNA methylation for each CpG in the matching
184 glioma tissue. The glioma tissue-derived methylome was analyzed as part of TCGA project using
185 the Illumina HM450K array, a reduced content compared to the HM850K. We identified

186 567/1075 differential CpGs measured on the HM450K array, of which 384 (68%) of the
187 detectable Glioma-eLB were present in the matching tissue and distinct from serum methylome
188 of other neoplasias (Fig. 2A-B, Extended Data Fig. S2C, Extended Data Table S3). We explored
189 the specificity of the Glioma-eLB signature using independent primary glioma tissue and other
190 non-glioma tumor methylomes and confirmed that glioma serum methylome clustered with
191 primary glioma (Fig. 2C, Extended Data Fig. S2D), corroborated our initial observation that the
192 eLB measurements reflected the glioma-tissue methylome. To further investigate the content of
193 the Glioma-eLB signature, we analyzed the similarity matrix across available non-cancer cell-
194 type signatures based on methylation (brain-, neural- and immune-associated cell types) and
195 observed that Glioma-eLB was more similar to neutrophils, monocytes, and normal glia- and
196 neuronal-cell-related signatures than any other normal cell types originating from different cell
197 lineages (Fig. 2D, Extended Data Fig. S2C). Interestingly, the Glioma-eLB signature segregated
198 with non-tumor serum that in turn clustered with brain and glial signatures, suggesting that
199 Glioma-eLB signature also captures tissue-of-origin signatures. Overall, we found that non-
200 invasive serum-derived Glioma-eLB signature is detectable and may reflect the expected
201 heterogeneous cell population present in glioma tissue.

202 Next we annotated the genomic location of the Glioma-eLB. We observed 167 CpG probes (see
203 Extended Data Table S3) that overlapped with known promoters; 64 were hypermethylated and
204 not significantly enriched in promoters ($OR=0.92$, 95% CI: [0.69, 1.21], chi-square test p-
205 value=0.564) while 103 were hypomethylated and this observation was enriched in promoters
206 ($OR=2.36$, 95% CI: [1.84, 3.04], chi-square test p-value=4.49E-12) above expected distribution
207 for the methylation platform (Extended Data Table S4). The lack of matching gene expression

208 data to the normal brain tissue DNA methylation limited our ability to investigate the biological
209 context of this Glioma-eLB signature.

210 Given the specificity of the detectable Glioma-eLB, we developed a machine learning (ML)
211 model to predict the presence of glioma. To determine the robustness of our eLB signature, we
212 applied a cross-validation method as follows: First, we redefined a new set of Glioma-eLB
213 relative to non-tumor controls using 11 randomly selected cases from the initial cohort (N=22).
214 Next, we trained a ML model (RandomForest) using the Glioma-eLB on the training set. This
215 provided us with a Glioma-eLB Index (GI), which estimates the probability that a sample is
216 likely a glioma (high GI or close to 1) or non-glioma (low GI or close to 0). We evaluated the
217 performance of the ML model and GI by applying it to the serum methylome of gliomas of the
218 test set (serum specimens left out of the training set, N=11), as well as to other serum cfDNA and
219 plasma methylomes from patients with non-tumor conditions (sepsis, pancreatic islet transplant
220 recipient) and other neoplasias (colorectal, breast, lung, pituitary and cancers of unknown
221 primary). To assess the stability of the GI development method, we replicated the test-set
222 selection, ML generation, and application steps 1000 times. The averaged GI for each of the
223 1000 models revealed glioma tissue and serum (test set) methylome samples carried the highest
224 GI (>0.5), whereas the plasma and/or serum of other non-glioma tumors (pituitary tumor, CRC,
225 breast carcinoma, etc.) carried a lower GI (<0.5) (Fig. 2E, Extended Data Table S5). Notably,
226 inflammatory conditions such as sepsis which carry a higher immune response, were not
227 classified as glioma (Fig. 2E), suggesting that despite the close association between Glioma-eLB
228 and immune cell signatures (Fig. 2D), the immune response captured in the cfDNA is specific to
229 gliomas. We evaluated each model's GI value at increments of 0.05 and determined that a GI of
230 0.60 accurately (mean±s.e.: 99.3%±0.000278) classified a glioma with 98% sensitivity and 99%

231 specificity (\pm s.e.: 0.001507 and 0.0002612, respectively, Fig. 2F, Extended Data Table S5). In
232 summary, we conclude that a classification model based on the Glioma-eLB signature can
233 predict a patient's glioma-like status from other neoplasias, non-cancer diseases or conditions
234 using the methylome profiles identified from a serum sample.

235 **Identification of prognostic glioma classes by non-invasive eLB**

236 Somatic mutation in one of the isocitrate dehydrogenase genes (*IDH1*, *IDH2*, [IDH]) is a
237 prognostic marker for adult glioma (WHO Grade II-IV) which is traditionally identified from
238 excised glioma tissue. We sought to define *IDH* mutation status by analyzing the cfDNA
239 methylation data of serum from *IDH*wt and *IDH*mut tumors, using the same approach used to
240 define the Glioma-eLB. Since our cohort carried a low sample size for the *IDH*mut 1p19q
241 codeletion (N=5) and *IDH*mut 1p19q intact (N=10), we combined these two good prognostic
242 subtypes into one class, *IDH*mut (N=15). Applying a supervised method and restricting our
243 analysis to CpGs within autosomal chromosomes (chrom 1-22), we identified 2647 *IDH*-eLB
244 that distinguished *IDH*mut from *IDH*wt gliomas (Fig. 3A, Extended Data Table S6) further
245 refined by selecting those with a similar methylation pattern in the matching tissue (CpGs
246 overlap HM450K = 1525/2647) which generated specific *IDH*mut-eLB and *IDH*wt-eLB
247 signatures (N=114/1525, 7.5% and N=124/1525, 8%, respectively, Fig. 3A, Extended Data Fig.
248 S3A). Harnessing the matching tissue methylome as well as pan-glioma methylome data from
249 adult patients, we observed that the *IDH*-specific eLB discriminates the two *IDH* subtypes at the
250 primary tissue level and the respective *IDH* serum methylome (*IDH*wt and *IDH*mut) clusters
251 with the respective tissue subtype (Fig. 3B-C, Extended Data Fig. S3B-D) corroborating the
252 specificity of the identified *IDH*-eLB.

253 We then investigated the potential functional or biological role of the *IDH*-eLB by analyzing the
254 methylome and transcriptome from the matching glioma tissue (see Methods). *IDH*wt specific
255 eLB overlapping CpG islands, shores and open seas were significantly enriched or depleted
256 when compared to the expected distribution set by the platform (chi-square test p-value 3.7E-04
257 enriched, 7.8E-05 enriched, 8.9E-05 depleted, respectively, Extended Data Fig. S3E). We
258 identified 28 *IDH*-eLB-specific signatures linked to a gene promoter (Table 2) of which, 14
259 transcripts were differentially expressed in *IDH*mut vs *IDH*wt tissues. Ten out of 14 transcripts
260 were inversely expressed in relation with the promoter methylation state (i.e. hypermethylated
261 promoter and down-regulated expression or vice-versa) (summarized in Table 2, Extended Data
262 Fig. S4A). For instance, *CXCR6* is a chemokine related to *CSCL16*, that is overexpressed in
263 glioma and associated with poor prognosis^{47–49}. *CXCR6* expression has been reported as a
264 predictor of recurrence and survival in hepatocellular carcinoma, in addition to intratumoral
265 neutrophils⁵⁰ and interestingly, *CXCR6* knockout glioma mice survived longer⁴⁷. Congruent
266 with these reports, *IDH*-eLB signature include the promoter hypomethylated state of *CXCR6* and
267 this gene is overexpressed in both TCGA and HBTC *IDH*wt glioma tissues (Fig. 3D). *PVT1* is a
268 long noncoding RNA (lncRNA) that when highly expressed is associated with progression and
269 poor prognosis in a pan-cancer cohort from TCGA. *PVT1* overexpression has also been
270 associated with poor response to chemotherapy in gliomas and squamous cell carcinoma of the
271 head and neck^{51–53}. In line with these findings, hypomethylation of the *PVT1* promoter was
272 detected in our *IDH*-eLB in association with overexpression of the correspondent gene in the
273 worst prognostic subtype (*IDH*wt), in both the TCGA and the current cohort samples (Fig. 3E).
274 Altogether, these findings suggest that with the selected *IDH*-eLB were able to differentiate the
275 various prognostic subtypes of glioma by utilizing the serum methylome. Moreover, the

276 associated CpG methylation in promoter genes which carry a biological and prognostic value in
277 gliomas, support the idea that these non-invasive eLB signatures are feasible for detection and
278 specific to glioma tumors.

279 **DISCUSSION**

280 In a review by Wan et al ³³, studies to define non-invasive blood-based markers for cancer has
281 mainly focused on identifying circulating tumor DNA (ctDNA) somatic mutations. And as such,
282 the focus has been on obtaining ctDNA from plasma since blood cell lysis during the preparation
283 of serum samples could release DNA from non-cancerous cells and thus dilute any ctDNA
284 markers and complicate the detection of ctDNA ³³. However, in glioma in particularly, low
285 sensitivity and reliability of extracting any glioma-specific ctDNA from plasma remains a
286 challenge because of the low frequency in somatic mutation and the targeted approach ³⁶. Since
287 epigenetic reflect the cell-of-origin ⁵⁴ and in glioma, somatic DNA methylation aberrations are
288 widespread ^{1,3,39,41}, we focused on profiling the DNA methylation of the released DNA to detect
289 glioma status (Glioma-eLB) and associated prognostic subtypes (*IDH*-eLB) by blood
290 (summarized in Fig. 4A). In addition, epigenetic profiling has the advantage of providing
291 information about the tumor microenvironment, which usually lacks somatic mutations ³³.
292 Deconvolution analysis allowed us to estimate that the Glioma-eLB were associated with
293 signatures from immune, neuronal, and glial cells; however, specific contribution of the
294 individual cells that comprise glioma could not be fully assessed as the information regarding the
295 methylome of the individual tumor cells is currently lacking. The signatures that characterize
296 Glioma-eLB are evidence in favor of leakage of the tumor microenvironment, or from the tumor
297 itself, as well as of a systemic immune response due to the presence of the glioma ⁵⁵. The
298 application of a variety of integrative approaches using matching methylome and transcriptome

299 of the primary tumor as well as available serum/plasma methylome from other non-glioma
300 patients highlighted the sensitivity and specificity of the detected Glioma-eLB. For instance, the
301 ML algorithm using the Glioma-eLB as the input correctly classified glioma tissues as such, in
302 contrast to the serum or tissue of other tumors and non-neoplastic conditions that were not
303 classified as gliomas (Fig. 2E). Despite Glioma-eLB comprising immune-cell-signatures,
304 patients with pro-inflammatory processes, such as sepsis, were not classified as glioma
305 suggesting that Glioma-eLB signatures related to the immune cells were related to the presence
306 of glioma (Fig. 2E). Many previous studies had limited success in detecting well known
307 prognostic markers, currently used in clinical practice and accrued from tissue analysis (such as
308 *MGMT* status, *IDH*, *PTEN*, *EGFR*, etc.)^{10,56,57}; however, our holistic approach of unbiased
309 screening allowed us to detect *IDH*-eLB overlapping promoters of genes which transcripts have
310 been well characterized as having a functional role in glioma tumorigenesis (e.g., *PVT1*, *LOXL*,
311 *CXCR6*, etc.) even though they are not currently utilized as part of the clinical treatment strategy.
312 These results also highlight that the *IDH*-eLB signatures are capturing glioma-tissue specific
313 markers.
314 Once validated in the proper settings (e.g., independent cohort), the potential clinical
315 implications of our findings could pave the way for altering current standard diagnostic measures
316 and therapies for glioblastoma, and possibly other brain tumors. Although still in its infancy, the
317 detection of glioma in the blood using a probability score generated by a ML algorithm based on
318 the methylome profile of the patient could challenge the current paradigm of a high-risk tissue-
319 dependent diagnosis prior to definitive treatment. For instance, the application of this approach
320 to patients with tumors not amenable to a meaningful resection due to comorbidities or
321 neuroanatomical surgical limitation such as deep or eloquent regions in the brain could have a

322 major impact in their management. In such scenarios, patients and physicians could proceed to
323 chemoradiotherapy without tissue diagnosis, negating the morbidity associated with biopsy or
324 subtotal resections that offer limited therapeutic benefit while starting definitive treatment
325 sooner. In our proposed model (Fig. 4B-C) and as described by others, the liquid biopsy could
326 also help differentiate glioma from other conditions (e.g., CNS lymphoma; demyelinating
327 disease or metastatic disease) and solve the dilemma of radiographic confounders, such as
328 pseudoprogression from radiation necrosis or immunotherapy agents from true disease
329 progression. The ability to diagnose and characterize the type of glioma prior to surgical
330 procedures could also aid neurosurgeons in tailoring the surgical approach to offer optimal
331 benefit such as surgical planning for maximal resection when typically a diagnostic biopsy
332 would be performed or clinical trial screening prior to initial surgical intervention which may
333 include intraoperative surgical trials and neoadjuvant therapies. In addition, a possible advantage
334 of detecting and characterizing the *IDH* status of aggressive gliomas rekindles the concept of
335 neoadjuvant therapy prior to surgical resection, a frontier that has yet to be fully explored in this
336 type of pathology. Recently, *IDH* mutation status of a solid tumor has been shown to change
337 during tumor progression possibly due to drug response⁴⁰, which is only confirmed during
338 recurrence by biopsy or resection. However, as suggested by Mazor et al⁴⁰, longitudinal
339 monitoring of *IDH* status during treatment offers significant opportunities to understand the role
340 of *IDH1* inhibitors. Last, but not least, although the cost of profiling whole-epigenome arrays
341 may currently limit its potential application, the discovery of an alternative, sensitive and more
342 cost-effective method to profile the epigenome of cfDNA has shown promise in several tumors
343³⁴. Once validated, the ML approach using detected eLB may be carried out after the detection
344 and initial treatment stages and used as real-time surveillance where frequent blood sampling

345 could aid the treatment team in monitoring disease status, progression to a more or less
346 aggressive phenotype, and response to specific treatment modalities. This could be done from the
347 comfort of the physician's office or anywhere a blood sample can be obtained rather than relying
348 solely on diagnostic imaging or surgical excision, as these require accumulation of tissue prior to
349 recognition of tumor activity, and offer the possibility of early intervention and preventative
350 approaches. Serial real-time monitoring of gliomas with the relative safety of a blood test has the
351 potential to redesign primary brain tumor diagnosis, surveillance, and therapeutic endpoints
352 which may shed light on new opportunities that improve outcomes for patients with malignant
353 primary brain tumors.

354 In summary, we developed a glioma index based on the glioma-specific CpG methylation
355 landscape that accurately predicted the presence of these tumors. These encouraging results
356 provide the framework to develop a blood-based epigenetic panel marker that will not only
357 provide real-time information regarding tumor features and burden but can also be used to
358 monitor disease progression and treatment response using a minimally invasive approach in ever-
359 evolving tumors such as malignant gliomas.

360

361 **SUBJECTS AND METHODOLOGY**

362 **EXPERIMENTAL DESIGN**

363 **Patients**

364 We performed a retrospective study entailing archival serum and tissue from patients who
365 received surgery to resect gliomas (N=22; composed of 15 *IDH*mut and 7 *IDH*wt; 4 *MGMT*-
366 negative and 18 *MGMT*-positive, Extended Data Table S1) at the Henry Ford Health System
367 (HFHS). The samples selected for this study had both serum and primary tissue available at the
368 tumor bank of the Hermelin Brain Tumor Center (HFHS, Detroit, MI). As an early tumor tissue
369 source site, matching tissue was submitted to TCGA from the HFHS and analyzed along with
370 1,122 other primary gliomas in a comprehensive pan-glioma study reported by our team³.
371 Additional serum samples from 3 non-tumor subjects were included (N=3, epilepsy). Clinical
372 data comprised of demographic features, and pathology report (e.g., grade, histology, molecular
373 markers, imaging, time to death, time to recurrence/progression, and treatment type) were
374 extracted from patient medical records. The project was approved by the HFHS Institutional
375 Review Board (IRB# 12490) and patients consented to have their specimens used for research
376 purposes. Tissue samples were blindly reviewed by two neuropathologists (A.M. and D.C.) to
377 determine which part of the tumor was feasible for DNA extraction (based on the percentage of
378 necrosis, hemorrhage, infiltration, and adjacent brain, etc.) and to confirm the molecular status
379 by immunohistochemistry and/or PCR (e.g., *IDH* and *MGMT* status). Detailed information about
380 our cohort demographics is depicted in Table 1 and Extended Data Table S1.

381 **Serum collection and processing**

382 Peripheral blood (15 mL) was drawn for each subject into 2 BD Vacutainer SSTs (Becton
383 Dickinson) at the time of surgery, before the opening of the dura-mater. Serum sample was

384 separated within 1 hour from collection by centrifugation at 1,300 x g for 10 minutes at 20°C;
385 aliquoted into up to five 2 mL cryovials and stored at -80°C until processing.

386 **cfDNA isolation, quantification, and DNA methylation data generation**

387 cfDNA was extracted from 2.2-9.3 mL aliquots of serum (Extended Data Table S1) using the
388 Quick-cfDNA Serum & Plasma Kit according to the manufacturer protocol (Zymo Research -
389 catalog # D4076). DNA concentration was measured with Qubit (Thermo Fisher Scientific). The
390 concentration of cfDNA in the serum was calculated by dividing the total amount of cfDNA
391 extracted by the amount of serum used for extraction. We then converted the concentration of
392 cfDNA in the serum (ng/mL) into haploid genome equivalents/mL by multiplying by a factor of
393 303 (assuming the mass of a haploid genome 3.3 pg)⁴⁵.

394 The extracted DNA (30-300 ng) was bisulfite-converted (Zymo EZ DNA methylation Kit; Zymo
395 Research) and profiled using an Illumina Human EPIC array (HM850K), at the USC Epigenome
396 Center, Keck School of Medicine, University of Southern California, Los Angeles, California.

397 The amount of bisulfite-converted DNA as well as the completeness of bisulfite conversion for
398 each sample was assessed using a panel of MethylLight-based real-time PCR quality control
399 assays as described previously⁵⁸. Bisulfite-converted DNAs were then repaired using the

400 Illumina Restoration Kit as recommended by the manufacturer (Zymo EZ DNA methylation Kit;
401 Zymo Research). The repaired DNA was used as a substrate for the Illumina EPIC BeadArrays,

402 as recommended by the manufacturer and first described in Moran et al., 2016⁵⁹. The raw DNA
403 methylation data reported in this paper has been deposited to Mendeley Data (CURRENTLY
404 embargo) at <https://data.mendeley.com/datasets/cgrz6zztfg>.

405 **DNA methylation pre-processing**

406 Methylation array data were processed with the minfi package in R. The raw signal intensities
407 were extracted from the *.IDAT files and corrected for background fluorescence intensities and
408 red-green dye-bias using the ‘noob’ function (preprocessNoob) as described by Triche et al.,
409 2013⁶⁰. The beta-values were calculated as (M/(M+U)), in which M and U refer to the (pre-
410 processed) mean methylated and unmethylated probe signal intensities, respectively.
411 Measurements in which the fluorescent intensity was not statistically significant above
412 background signal (detection p value > 10⁻¹⁶) were removed from the data set. Before the
413 analyses, we filtered out probes that were designed for sequences with known polymorphisms or
414 probes with poor mapping quality (complete list of masked probes provided by Zhou et al.⁶¹)
415 and the X and Y chromosomes.

416 **Deconvolution**

417 To tease out the origin of the cfDNA in the serum, we applied a previously described
418 methodology⁴⁵ to deconvolute the relative contribution of cell types to a given sample. Briefly,
419 we selected 100 of the most specific hypermethylated and hypomethylated CpG probes for each
420 cell line of interest. Given the availability of public data and the nature of our study, we included
421 healthy cell lines and isolated cells from the blood, brain, vascular endothelial cells and
422 adipocytes. We then used our DNA methylation derived signature and the non-negative least
423 squares method to deconvolute our serum cohort using the standalone program provided by
424 Moss and colleagues⁴⁵. We then normalized the percentages generated by the standalone
425 program for each cell type from 0 to 100 by serum and tissue separately.

426 **Supervised analysis**

427 We performed a supervised analysis taking into account clinical-pathological attributes in both
428 tissue or serum and used the Wilcoxon rank-sum test to identify differentially methylated sites
429 between two groups of study (i.e., glioma vs non-glioma, *IDHmut* vs *IDHwt*)
430 Probes were considered differentially methylated when p-values were less than 0.01 in the
431 comparison between *IDHmut* vs *IDHwt* glioma samples (*IDH-eLB*), or 0.001 between glioma
432 with non-tumor samples (Glioma-eLB). To identify differentially methylated probes in the serum
433 that were tissue-specific, we calculated the differences in DNA methylation between the serum
434 and tissue, by patient. Next we calculated the mean of the difference for each probe across
435 glioma samples. For the *IDH*-specific analysis we calculated the mean of the difference across
436 *IDHmut* and *IDHwt* samples. We then selected probes with less than 5% difference between
437 tissue and serum and considered them tissue-specific.

438 **Random Forest**

439 We used a random forest ML model with the aim to classify available cfDNA methylation
440 (serum or plasma and tissue) derived from tumor patients, patients with metastasis of unknown
441 primary, non-tumor conditions (sepsis, pancreatic islet transplantation recipient) and non-
442 tumor/non-diseased cell-free DNA and non-tumor brain and glioma tissue from TCGA.
443 To validate the performance of our ML model, for this specific approach, we randomly selected
444 11 glioma serum samples and used this set to compare to the 3 non-tumor serum samples and
445 derived a new glioma-tissue specific signature (top 1000 most significant probes). We then used
446 this set of 11 glioma serum samples and 3 non-tumor serum samples to train the random forest
447 ML model. We tested the remaining 11 glioma serum samples using our model, ensuring to
448 bootstrap 1000 times to reduce training/test biases. We named the output probability score of the
449 random forest model Glioma-eLB index (GI).

450 **Gene Expression and integrative analysis**

451 First, we stratified the *IDH*-eLB CpG probes into distinct genomic regions^{41,54} defined as
452 “OpenSeas” (N=51 *IDH*mut, N=33 *IDH*wt), “Shores” (N=26 *IDH*mut, N=44 *IDH*wt), “Shelves”
453 (N=10 *IDH*mut, N=4 *IDH*wt) and “CpG Islands” (N=27 *IDH*mut, N=43 *IDH*wt) (Extended Data
454 Fig. S3E). We annotated our *IDH*-specific eLB prognostic probes (from the comparison *IDH*mut
455 vs *IDH*wt serum samples) to their genomic location and selected probes mapped to the promoter
456 of genes (defined as 2000 base pairs upstream and 200 base pairs downstream the TSS). We then
457 looked into the DNA methylation levels of promoter CpGs in the corresponding glioma tissue
458 and selected the probes in which the methylation status (i.e. hypermethylated or hypomethylated)
459 was the same in both tissue and serum samples. Finally, we investigated the expression levels of
460 the corresponding gene in glioma tissue and annotated their identified roles in prior cancer
461 studies.

462 **ACKNOWLEDGMENTS**

463 The authors are grateful to the patients who participated on this study. This work was supported
464 by the Henry Ford Health System, Department of Neurosurgery and the Hermelin Brain Tumor
465 Center Foundation. Additionally, LMP, HN, ACD, MW, and AM are supported by NCI
466 R01CA222146 and HN, TSS, TMM, LMP, ACD, and AVC are supported by Department of
467 Defense CA170278. We would like to thank Nancy Takacs for administrative support. We
468 would like to thank Susan MacPhee and Kate Lawrenson for critical review of the manuscript.

469

470 **AUTHOR CONTRIBUTIONS**

471 The contributions of the authors are as follows: Serum and tissue collection and storage: KN and
472 AD. surgical procedure and pathology review: JS, AM, DC, AR, MR, TM, JR, IL, TW, and SK.

473 cfDNA extraction and methylome profile: TMM, KN, and AD. Glioma-eLB and *IDH*-eLB
474 signatures, HN, TSS, TMM, and AVC; methodology, HN, TSS, TMM, and AVC; statistical
475 analysis, HN, TSS, TMM, and LP; pituitary tissue and serum curation, HN, TMM, MW, MM,
476 and AVC; clinical analysis and interpretation, HN, TSS, TMM, LP, JS, KPA, TM, IL, TW, SK,
477 and AVC; data interpretation, HN, TSS, TMM, and AVC; visualization: TSS and TMM with
478 AVC and HN input; original draft: HN and AVC; and overall concept and coordination, HN. All
479 authors read and approved the final manuscript.

480

481 **COMPETING INTERESTS**

482 The authors declare to have no competing interests.

483

484 **BIBLIOGRAPHY**

- 485 1. Noushmehr, H. *et al.* Identification of a CpG island methylator phenotype that defines a
486 distinct subgroup of glioma. *Cancer Cell* **17**, 510–522 (2010).
- 487 2. Cancer Genome Atlas Research Network *et al.* Comprehensive, Integrative Genomic
488 Analysis of Diffuse Lower-Grade Gliomas. *N. Engl. J. Med.* **372**, 2481–2498 (2015).
- 489 3. Ceccarelli, M. *et al.* Molecular profiling reveals biologically discrete subsets and pathways
490 of progression in diffuse glioma. *Cell* **164**, 550–563 (2016).
- 491 4. Brennan, C. W. *et al.* The somatic genomic landscape of glioblastoma. *Cell* **155**, 462–477
492 (2013).
- 493 5. Yan, H. *et al.* IDH1 and IDH2 mutations in gliomas. *N. Engl. J. Med.* **360**, 765–773 (2009).
- 494 6. Sturm, D. *et al.* Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and
495 biological subgroups of glioblastoma. *Cancer Cell* **22**, 425–437 (2012).

- 496 7. Verma, N., Cowperthwaite, M. C., Burnett, M. G. & Markey, M. K. Differentiating tumor
497 recurrence from treatment necrosis: a review of neuro-oncologic imaging strategies. *Neuro*
498 *Oncol.* **15**, 515–534 (2013).
- 499 8. Parvez, K., Parvez, A. & Zadeh, G. The diagnosis and treatment of pseudoprogression,
500 radiation necrosis and brain tumor recurrence. *Int. J. Mol. Sci.* **15**, 11832–11846 (2014).
- 501 9. Joosse, S. A. & Pantel, K. Tumor-Educated Platelets as Liquid Biopsy in Cancer Patients.
502 *Cancer Cell* **28**, 552–554 (2015).
- 503 10. Fontanilles, M., Duran-Peña, A. & Idbaih, A. Liquid biopsy in primary brain tumors:
504 looking for stardust! *Curr. Neurol. Neurosci. Rep.* **18**, 13 (2018).
- 505 11. Dor, Y. & Cedar, H. Principles of DNA methylation and their implications for biology and
506 medicine. *Lancet* **392**, 777–786 (2018).
- 507 12. Chen, X. Q. *et al.* Microsatellite alterations in plasma DNA of small cell lung cancer
508 patients. *Nat. Med.* **2**, 1033–1035 (1996).
- 509 13. Bettegowda, C. *et al.* Detection of circulating tumor DNA in early- and late-stage human
510 malignancies. *Sci. Transl. Med.* **6**, 224ra24 (2014).
- 511 14. Li, J. *et al.* Clinical applications of liquid biopsy as prognostic and predictive biomarkers in
512 hepatocellular carcinoma: circulating tumor cells and circulating tumor DNA. *J. Exp. Clin.*
513 *Cancer Res.* **37**, 213 (2018).
- 514 15. Fernandez-Cuesta, L. *et al.* Identification of Circulating Tumor DNA for the Early
515 Detection of Small-cell Lung Cancer. *EBioMedicine* **10**, 117–123 (2016).
- 516 16. Vatandoost, N. *et al.* Early detection of colorectal cancer: from conventional methods to
517 novel biomarkers. *J. Cancer Res. Clin. Oncol.* **142**, 341–351 (2016).
- 518 17. Martínez-Ricarte, F. *et al.* Molecular Diagnosis of Diffuse Gliomas through Sequencing of

- 519 Cell-Free Circulating Tumor DNA from Cerebrospinal Fluid. *Clin. Cancer Res.* **24**, 2812–
520 2819 (2018).
- 521 18. Best, M. G. *et al.* Liquid biopsies in patients with diffuse glioma. *Acta Neuropathol.* **129**,
522 849–865 (2015).
- 523 19. Shankar, G. M., Balaj, L., Stott, S. L., Nahed, B. & Carter, B. S. Liquid biopsy for brain
524 tumors. *Expert Rev. Mol. Diagn.* **17**, 943–947 (2017).
- 525 20. Adamczyk, L. A. *et al.* Current Understanding of Circulating Tumor Cells - Potential Value
526 in Malignancies of the Central Nervous System. *Front. Neurol.* **6**, 174 (2015).
- 527 21. Heitzer, E., Haque, I. S., Roberts, C. E. S. & Speicher, M. R. Current and future
528 perspectives of liquid biopsies in genomics-driven oncology. *Nat. Rev. Genet.* **20**, 71–88
529 (2019).
- 530 22. Mouliere, F. *et al.* Detection of cell-free DNA fragmentation and copy number alterations in
531 cerebrospinal fluid from glioma patients. *EMBO Mol. Med.* **10**, (2018).
- 532 23. Panditharatna, E. *et al.* Clinically Relevant and Minimally Invasive Tumor Surveillance of
533 Pediatric Diffuse Midline Gliomas Using Patient-Derived Liquid Biopsy. *Clin. Cancer Res.*
534 **24**, 5850–5859 (2018).
- 535 24. Pan, W., Gu, W., Nagpal, S., Gephart, M. H. & Quake, S. R. Brain tumor mutations
536 detected in cerebral spinal fluid. *Clin. Chem.* **61**, 514–522 (2015).
- 537 25. De Mattos-Arruda, L. *et al.* Cerebrospinal fluid-derived circulating tumour DNA better
538 represents the genomic alterations of brain tumours than plasma. *Nat. Commun.* **6**, 8839
539 (2015).
- 540 26. Wang, J. & Bettegowda, C. Applications of DNA-Based Liquid Biopsy for Central Nervous
541 System Neoplasms. *J. Mol. Diagn.* **19**, 24–34 (2017).

- 542 27. Barault, L. *et al.* Discovery of methylated circulating DNA biomarkers for comprehensive
543 non-invasive monitoring of treatment response in metastatic colorectal cancer. *Gut* **67**,
544 1995–2005 (2018).
- 545 28. Li, W. *et al.* 5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic
546 biomarkers for human cancers. *Cell Res.* **27**, 1243–1257 (2017).
- 547 29. Kang, S. *et al.* CancerLocator: non-invasive cancer diagnosis and tissue-of-origin prediction
548 using methylation profiles of cell-free DNA. *Genome Biol.* **18**, 53 (2017).
- 549 30. Miller, A. M. *et al.* Tracking tumour evolution in glioma through liquid biopsies of
550 cerebrospinal fluid. *Nature* **565**, 654–658 (2019).
- 551 31. Zill, O. A. *et al.* The Landscape of Actionable Genomic Alterations in Cell-Free Circulating
552 Tumor DNA from 21,807 Advanced Cancer Patients. *Clin. Cancer Res.* **24**, 3528–3538
553 (2018).
- 554 32. Jiang, R. *et al.* A comparison of isolated circulating tumor cells and tissue biopsies using
555 whole-genome sequencing in prostate cancer. *Oncotarget* **6**, 44781–44793 (2015).
- 556 33. Wan, J. C. M. *et al.* Liquid biopsies come of age: towards implementation of circulating
557 tumour DNA. *Nat. Rev. Cancer* **17**, 223–238 (2017).
- 558 34. Shen, S. Y. *et al.* Sensitive tumour detection and classification using plasma cell-free DNA
559 methylomes. *Nature* **563**, 579–583 (2018).
- 560 35. Sun, J. *et al.* Examination of Plasma Cell-Free DNA of Glioma Patients by Whole Exome
561 Sequencing. *World Neurosurg.* (2019). doi:10.1016/j.wneu.2019.01.092
- 562 36. Lavon, I., Refael, M., Zelikovitch, B., Shalom, E. & Siegal, T. Serum DNA can define
563 tumor-specific genetic and epigenetic markers in gliomas of various grades. *Neuro Oncol.*
564 **12**, 173–180 (2010).

- 565 37. Uehiro, N. *et al.* Circulating cell-free DNA-based epigenetic assay can detect early breast
566 cancer. *Breast Cancer Res.* **18**, 129 (2016).
- 567 38. Capper, D. *et al.* DNA methylation-based classification of central nervous system tumours.
568 *Nature* **555**, 469–474 (2018).
- 569 39. de Souza, C. F. *et al.* A Distinct DNA Methylation Shift in a Subset of Glioma CpG Island
570 Methylator Phenotypes during Tumor Recurrence. *Cell Rep.* **23**, 637–651 (2018).
- 571 40. Mazor, T. *et al.* Clonal expansion and epigenetic reprogramming following deletion or
572 amplification of mutant IDH1. *Proc Natl Acad Sci USA* **114**, 10743–10748 (2017).
- 573 41. Malta, T. M. *et al.* Glioma CpG island methylator phenotype (G-CIMP): biological and
574 clinical implications. *Neuro Oncol.* **20**, 608–620 (2018).
- 575 42. Newman, A. M. *et al.* An ultrasensitive method for quantitating circulating tumor DNA
576 with broad patient coverage. *Nat. Med.* **20**, 548–554 (2014).
- 577 43. Cheng, F., Su, L. & Qian, C. Circulating tumor DNA: a promising biomarker in the liquid
578 biopsy of cancer. *Oncotarget* **7**, 48832–48841 (2016).
- 579 44. Gallardo-Gómez, M. *et al.* A new approach to epigenome-wide discovery of non-invasive
580 methylation biomarkers for colorectal cancer screening in circulating cell-free DNA using
581 pooled samples. *Clin. Epigenetics* **10**, 53 (2018).
- 582 45. Moss, J. *et al.* Comprehensive human cell-type methylation atlas reveals origins of
583 circulating cell-free DNA in health and disease. *Nat. Commun.* **9**, 5068 (2018).
- 584 46. Sullivan, J. P. *et al.* Brain tumor cells in circulation are enriched for mesenchymal gene
585 expression. *Cancer Discov.* **4**, 1299–1309 (2014).
- 586 47. Lepore, F. *et al.* CXCL16/CXCR6 axis drives microglia/macrophages phenotype in
587 physiological conditions and plays a crucial role in glioma. *Front. Immunol.* **9**, 2750 (2018).

- 588 48. Darash-Yahana, M. *et al.* The chemokine CXCL16 and its receptor, CXCR6, as markers
589 and promoters of inflammation-associated cancers. *PLoS ONE* **4**, e6695 (2009).
- 590 49. Lee, J. T., Lee, S. D., Lee, J. Z., Chung, M. K. & Ha, H. K. Expression analysis and clinical
591 significance of CXCL16/CXCR6 in patients with bladder cancer. *Oncol. Lett.* **5**, 229–235
592 (2013).
- 593 50. Gao, Q. *et al.* CXCR6 upregulation contributes to a proinflammatory tumor
594 microenvironment that drives metastasis and poor patient outcomes in hepatocellular
595 carcinoma. *Cancer Res.* **72**, 3546–3556 (2012).
- 596 51. Zhang, Y., Yang, G. & Luo, Y. Long non-coding RNA PVT1 promotes glioma cell
597 proliferation and invasion by targeting miR-200a. *Exp. Ther. Med.* **17**, 1337–1345 (2019).
- 598 52. Yu, C. *et al.* LncRNA PVT1 promotes malignant progression in squamous cell carcinoma
599 of the head and neck. *J. Cancer* **9**, 3593–3602 (2018).
- 600 53. Zou, H. *et al.* lncRNAs PVT1 and HAR1A are prognosis biomarkers and indicate therapy
601 outcome for diffuse glioma patients. *Oncotarget* **8**, 78767–78780 (2017).
- 602 54. Shen, H. & Laird, P. W. Interplay between the cancer genome and epigenome. *Cell* **153**,
603 38–55 (2013).
- 604 55. Thorsson, V. *et al.* The immune landscape of cancer. *Immunity* **48**, 812–830.e14 (2018).
- 605 56. DeWeerdt, S. The genomics of brain cancer. *Nature* **561**, S54–S55 (2018).
- 606 57. Ramkissoon, L. A. *et al.* Precision Neuro-oncology: the Role of Genomic Testing in the
607 Management of Adult and Pediatric Gliomas. *Curr. Treat. Options Oncol.* **19**, 41 (2018).
- 608 58. Campan, M., Weisenberger, D. J., Trinh, B. & Laird, P. W. MethyLight. *Methods Mol. Biol.*
609 **507**, 325–337 (2009).
- 610 59. Moran, S., Arribas, C. & Esteller, M. Validation of a DNA methylation microarray for

- 611 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics* **8**,
612 389–399 (2016).
- 613 60. Triche, T. J., Weisenberger, D. J., Van Den Berg, D., Laird, P. W. & Siegmund, K. D. Low-
614 level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res.* **41**,
615 e90 (2013).
- 616 61. Zhou, W., Laird, P. W. & Shen, H. Comprehensive characterization, annotation and
617 innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res.* **45**, e22
618 (2017).

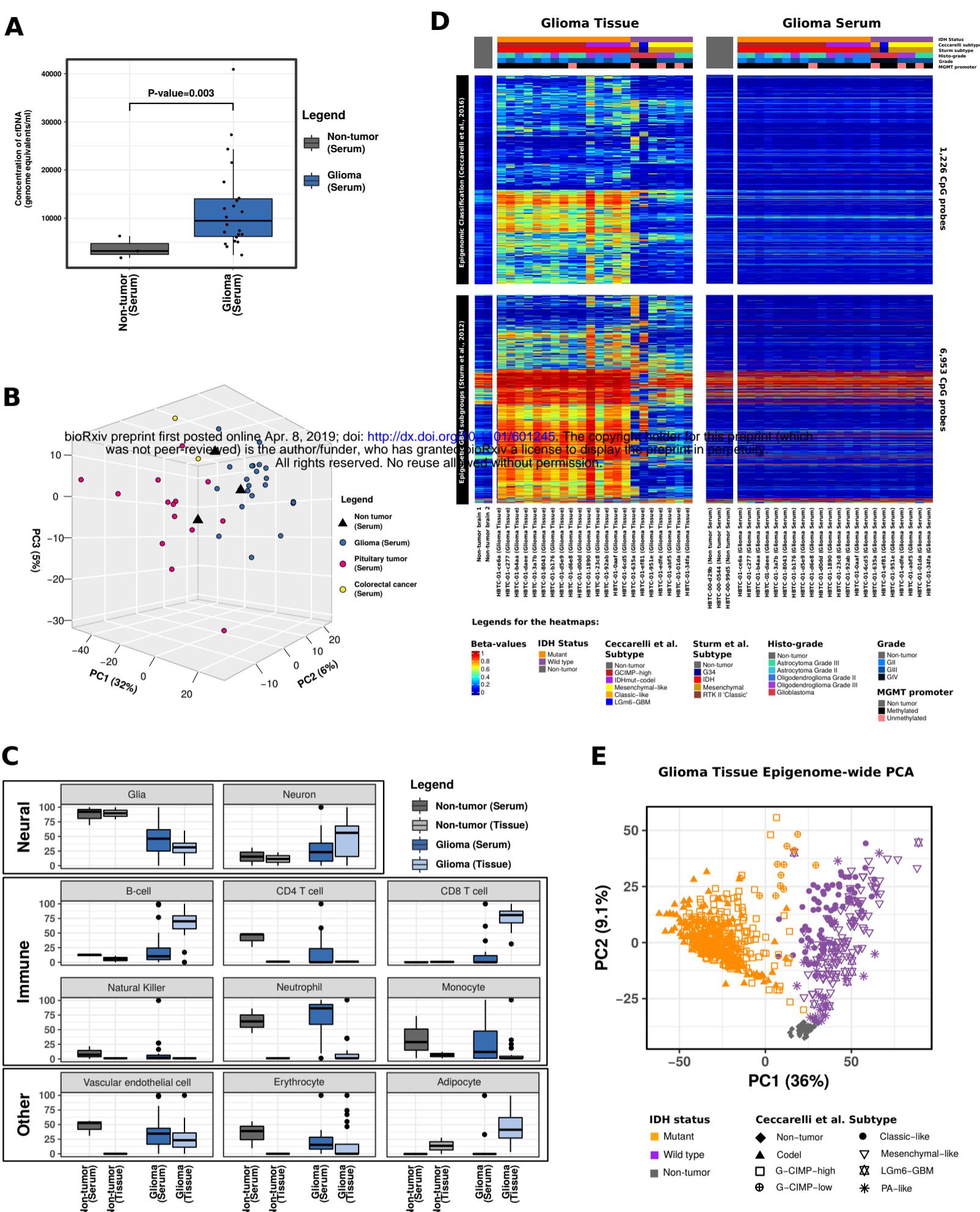


Fig. 1: Genome-wide DNA methylation profile of Glioma serum cfDNA.

A) Total serum cfDNA concentration normalized to the genomic size (Genomic Equivalents/ml) is represented in two boxplots (non-tumor vs glioma). B) Epigenome-wide cell-free DNA methylation (serum-based) derived from Glioma, Pituitary tumor, Colorectal carcinoma (CRC), and non-tumor patients presented by a similarity method: Principal Components Analysis (PCA). Each dot represents a sample cfDNA methylation (epigenome-wide) and colored based on tissue/tumor of origin. Total percent variance is indicated along all three axes. C) Cell-type CpG methylation-based deconvolution of our patients' cfDNA methylome is divided into three relevant categories: Neural, Immune and Other cell types. Y-axis represents normalized percent of each cell-type and separately by serum and tissue present in the cfDNA methylome. Eleven sets of boxplots are each divided into two categories; non-tumor (grey) vs glioma (blue). The plots are further divided by tissue (non-tumor-light grey or glioma- light blue) and cfDNA serum (non-tumor- dark grey or glioma- dark blue). D) Published tissue-derived epigenetic signatures from Ceccarelli et al. 2016 (top heatmaps) and Sturm et al. 2012 (bottom heatmaps) are undetectable in Glioma cfDNA methylome. Levels of CpG methylation in our HBTC cohort divided by glioma tissue (left two heatmaps, respectively) vs matching serum cfDNA (right two heatmaps). Columns indicate patients with clinical and molecular annotation tracks listed on top of heatmaps and rows indicates CpG probe. Beta-value (DNA methylation levels) are indicated in the legend from 0 (low CpG methylation) to 1 (high CpG methylation). E) Glioma-tissue epigenome-wide PCA highlights the difference between glioma and non-tumor and between IDHmut and IDHwt patients. TCGA Pan-Glioma DNA methylation data from Ceccarelli et al. 2016 is represented in a Principal Component Analysis to evaluate similarities genome-wide. The first two principal components are plotted using all available DNA methylation data points (~400,000 CpGs, unfiltered). The variance total percentage is labeled along both axes. Fifty-four total non-glioma tissue (grey) were also included in this analysis to highlight the epigenome-wide difference between glioma and non-tumor. The glioma cohort is further divided by available known IDH status and by our recent epigenomic subtypes.

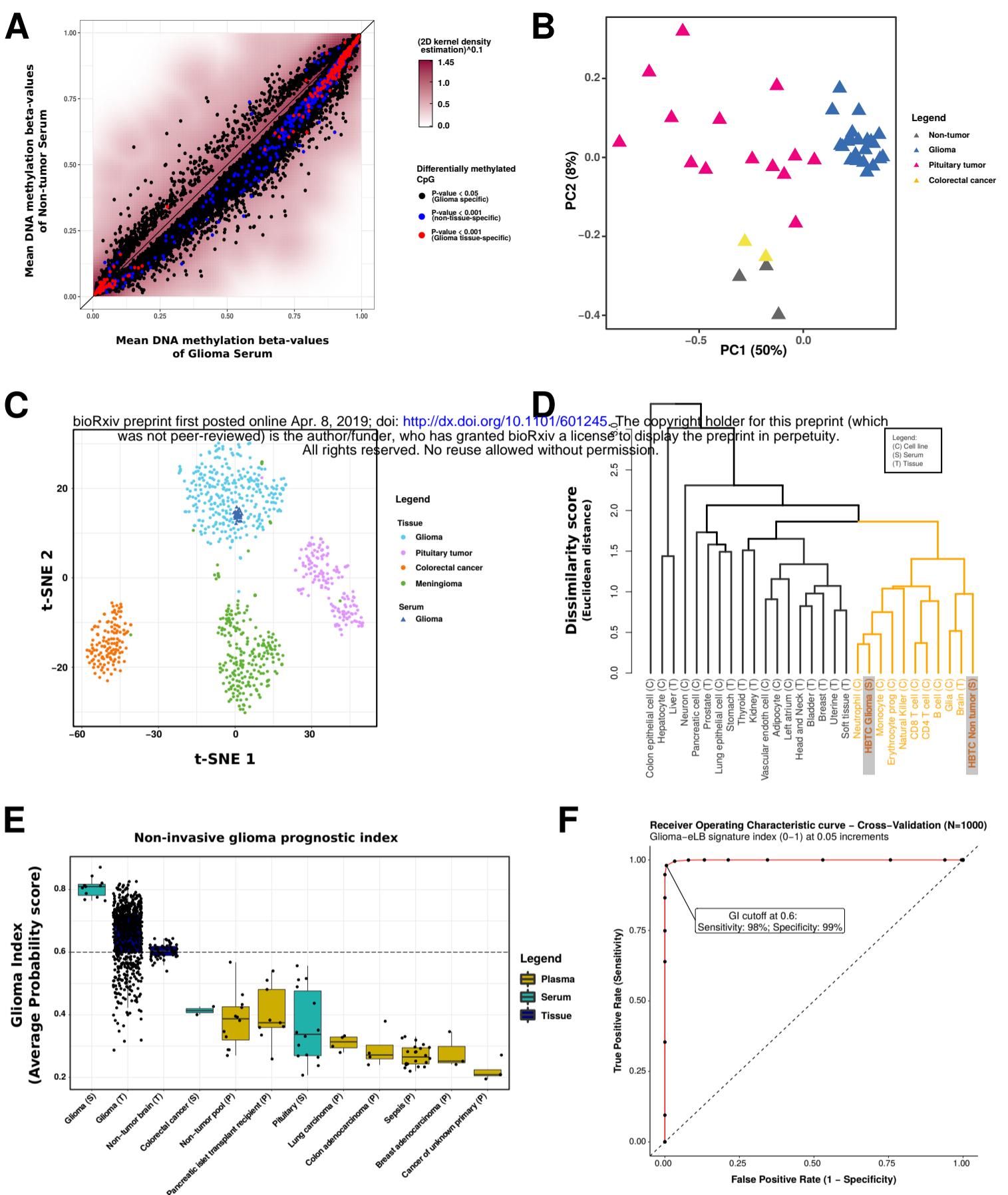
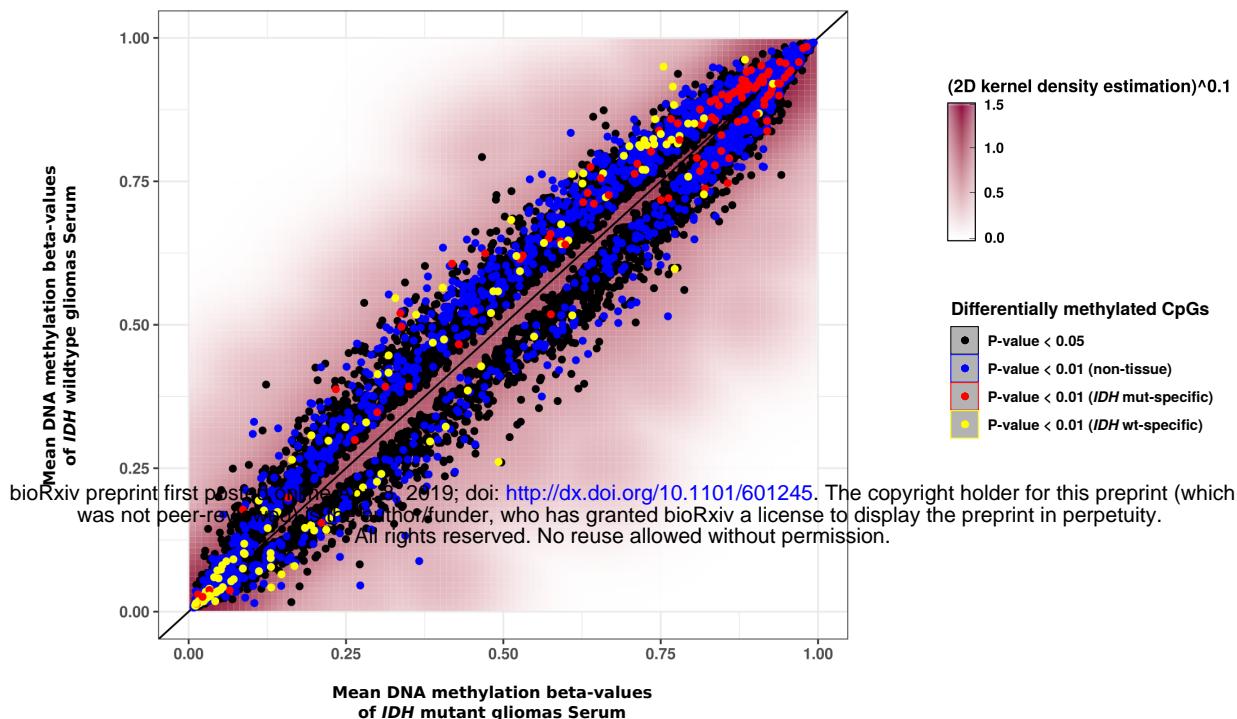
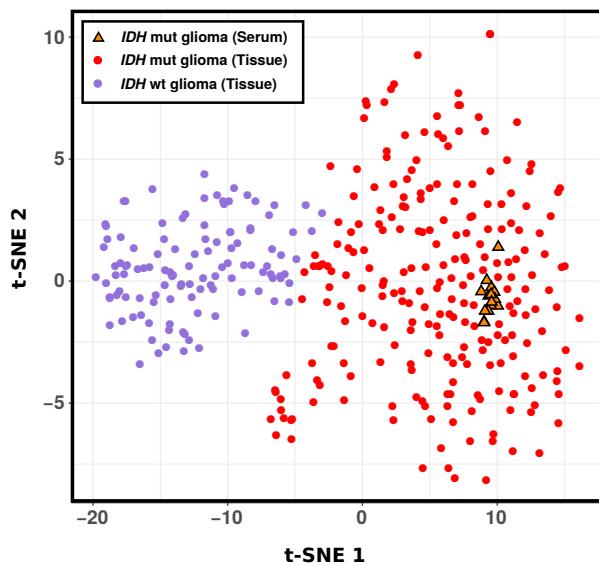
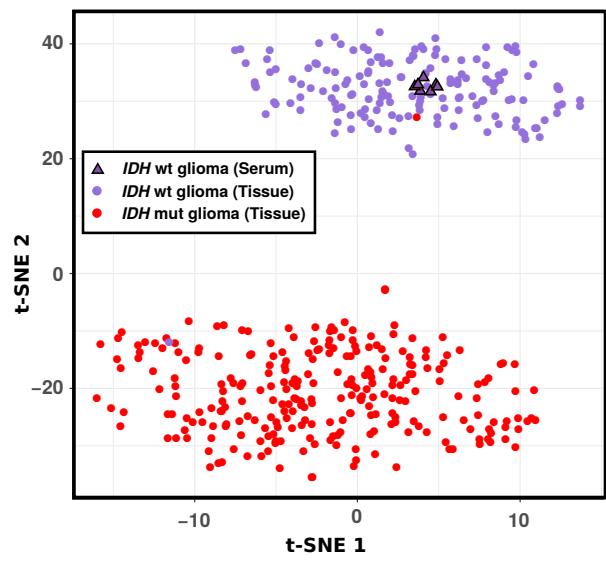
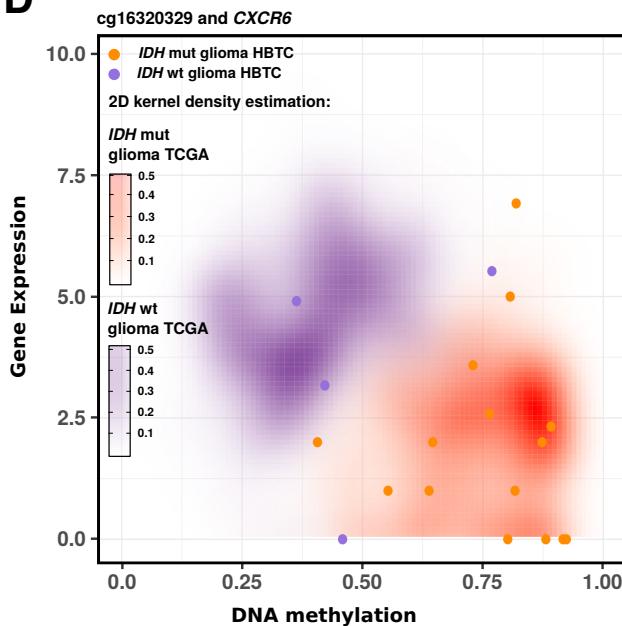
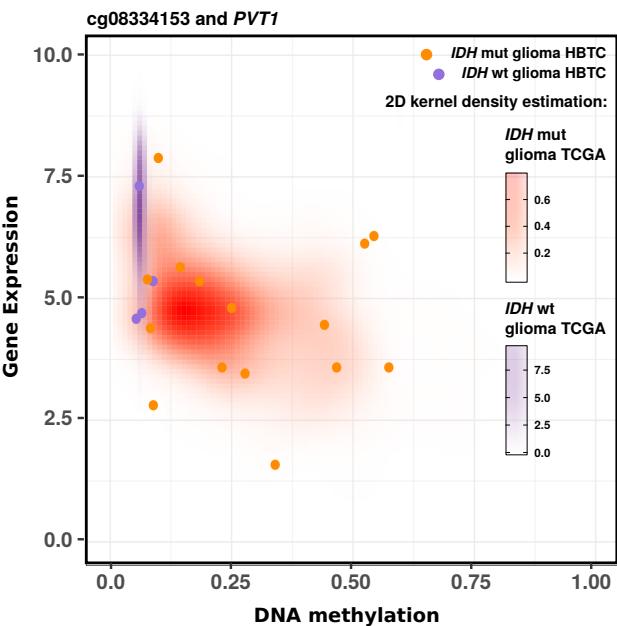


Fig. 2: Identification of Glioma-specific Epigenetic Liquid Biopsy (eLB) as a Diagnostic Marker for Gliomas.

A) Epigenome-wide mean DNA methylation across our patient cohort's serum cfDNA methylation (y-axis: three non-tumor samples and x-axis: 22 glioma-patient-derived serum samples). Non-significant CpG methylation probes (p -value $> 5\%$) are condensed into a density heatmap by calculating the 2D kernel density estimation to the power of 0.1. Identified cfDNA methylation signatures associated with glioma patients ($N=1,075$) are selected by different p -values (black, p -value < 0.05 ; blue, and red, p -value < 0.001). Glioma-eLB signatures (p -value < 0.001) are further divided into CpGs that are measurable (Glioma tissue specific, red) or not measurable in the matching glioma tissue (Non-tissue specific, blue). B) Principal Component Analysis using the Glioma-eLB signatures as input. Serum methylome from glioma, pituitary tumors, CRC and non-tumor samples are represented. C) Glioma-specific tissue-matching eLB ($N=384$) was used to subset the published primary tumor tissue DNA methylation and using t-SNE (t-distributed stochastic neighbour embedding) dimensionality reduction to visualize the similarities of each sample. As expected, each primary tumor type (circles) clusters with its known cell-of-origin. Serum cfDNA methylation of our patients cohort (triangles) clusters with the primary glioma tissue DNA methylation profiles. D) Dendrogram of non-tumor cell types in comparison to Glioma-eLB ($N=384$). Based on the mean DNA methylation across each cell type, this dendrogram shows that glioma serum cfDNA clusters with relevant immune cell-types along with glial-derived cells and bulk brain (non-tumor) samples. E) Machine learning (ML) application (Random Forest) using our defined Glioma tissue-specific eLB to classify tumors and available cfDNA methylation (serum or plasma) derived from tumor patients, patients with metastasis of unknown primary, non-tumor conditions (e.g., sepsis, pancreatic islet transplantation recipient) and non-tumor/non-diseased cell-free DNA. Y-axis represents the ML similarity index based on Glioma-eLB signatures averaged across 1000 iterations. Zero indicates low probability of a sample being a glioma, while 1 indicates high probability of a sample being a glioma. Dash line indicates cutoff to determine glioma classification. F) Receiver operating characteristic curve derived from the average across 1000 cross validation. Specificity and sensitivity calculated at 0.05 increments ($N=21$) from 0 to 1 Glioma-eLB signature index.

A**B****C****D****E****Fig. 3: IDH Specific eLB Prognostic Markers.**

A) Mean DNA methylation of 750,000 CpG across 15 *IDH* mut-patient derived serum (x-axis) vs 7 *IDH* wt-patient derived serum (y-axis). Non-significant CpG methylation probes are condensed into a density heatmap by calculating the 2D kernel density estimation to the power of 0.1. Proposed prognostic glioma-specific eLB (N=1,075) selected by p-values are represented by colored dots (black, p-value < 0.05; blue, red and yellow, p-value < 0.01). Similarities across selected tumor tissue and serum, using *IDH* mut- tissue specific eLB levels (red circles) and *IDH* wt- tissue specific eLB levels (yellow circles). B-C) Similarities across Pan-Glioma tissue (N=259 *IDH* mut and 160 *IDH* wt) and *IDH* glioma cfDNA methylation (N=15 *IDH* mut, 7 *IDH* wt), using *IDH* mut- tissue specific eLB signatures B) and *IDH* wt- tissue specific eLB signatures C) as input in a t-SNE analysis to visualize the similarities by sample. Circles represent primary tissue and triangles represents tumor serum cfDNA. Red indicates *IDH* mut and purple indicates *IDH* wt. D-E) scatter plot between DNA methylation (x-axis) and Gene expression (y-axis) for all pan-glioma primary tumor tissue. 2D kernel density indicates all glioma samples divided by *IDH* status (purple = *IDH* wt, orange = *IDH* mut). Circles indicate the HBTC primary tumor tissue DNA methylation and expression values. D) DNA methylation and expression scatter plot for promoter CpG associated with *CXCR6*. E) DNA methylation and expression scatter plot for promoter CpG associated with *PVT1*.

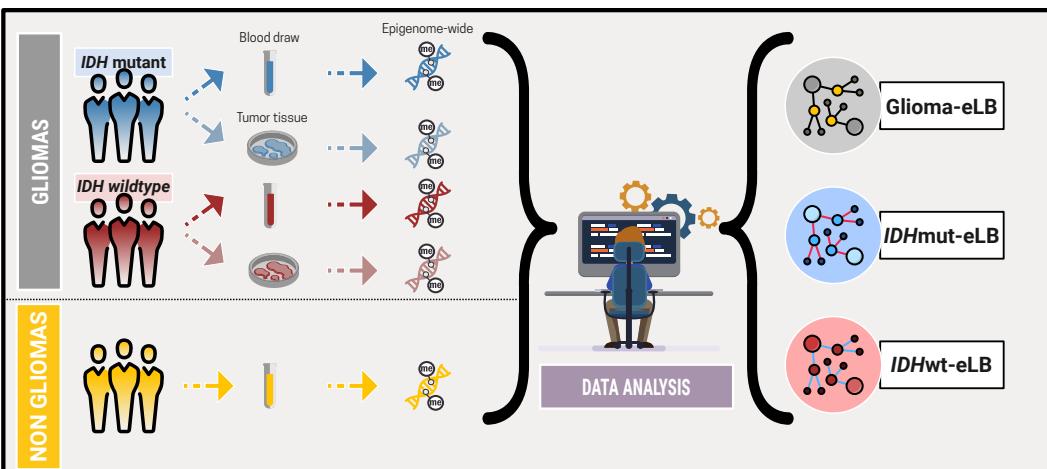
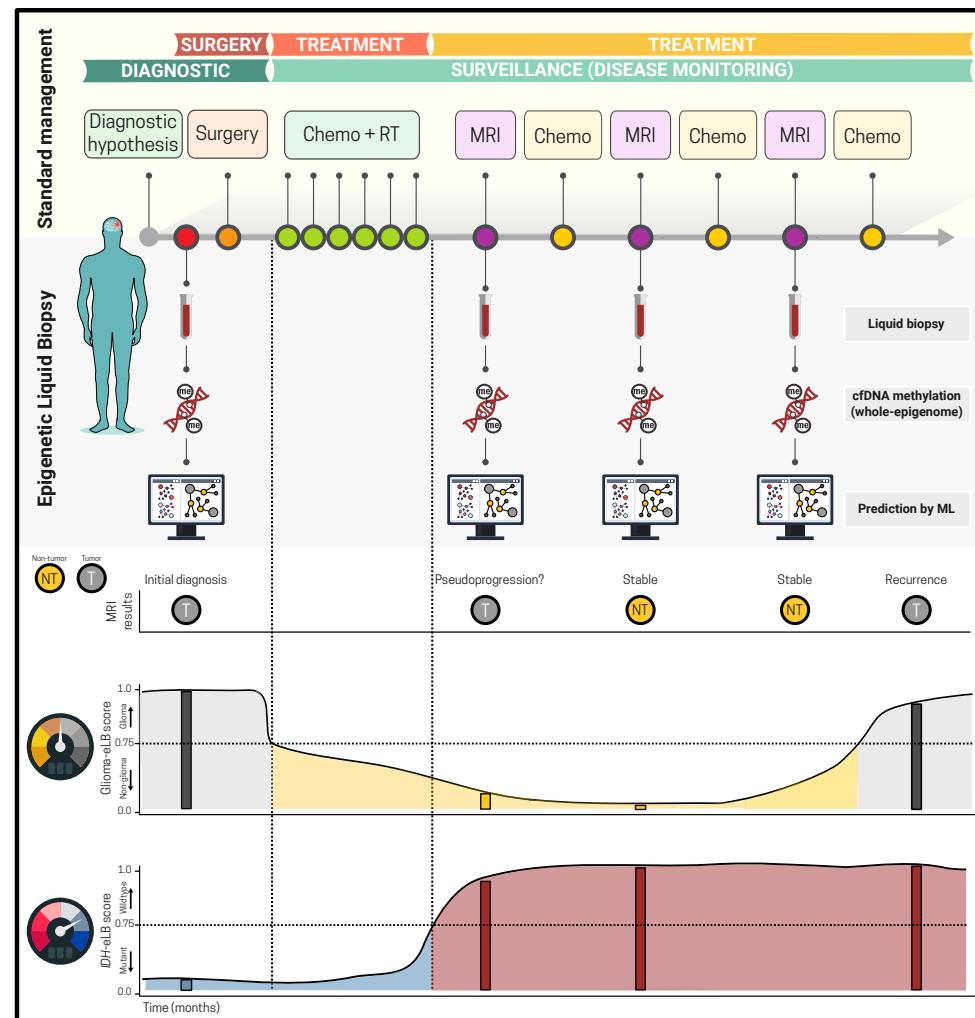
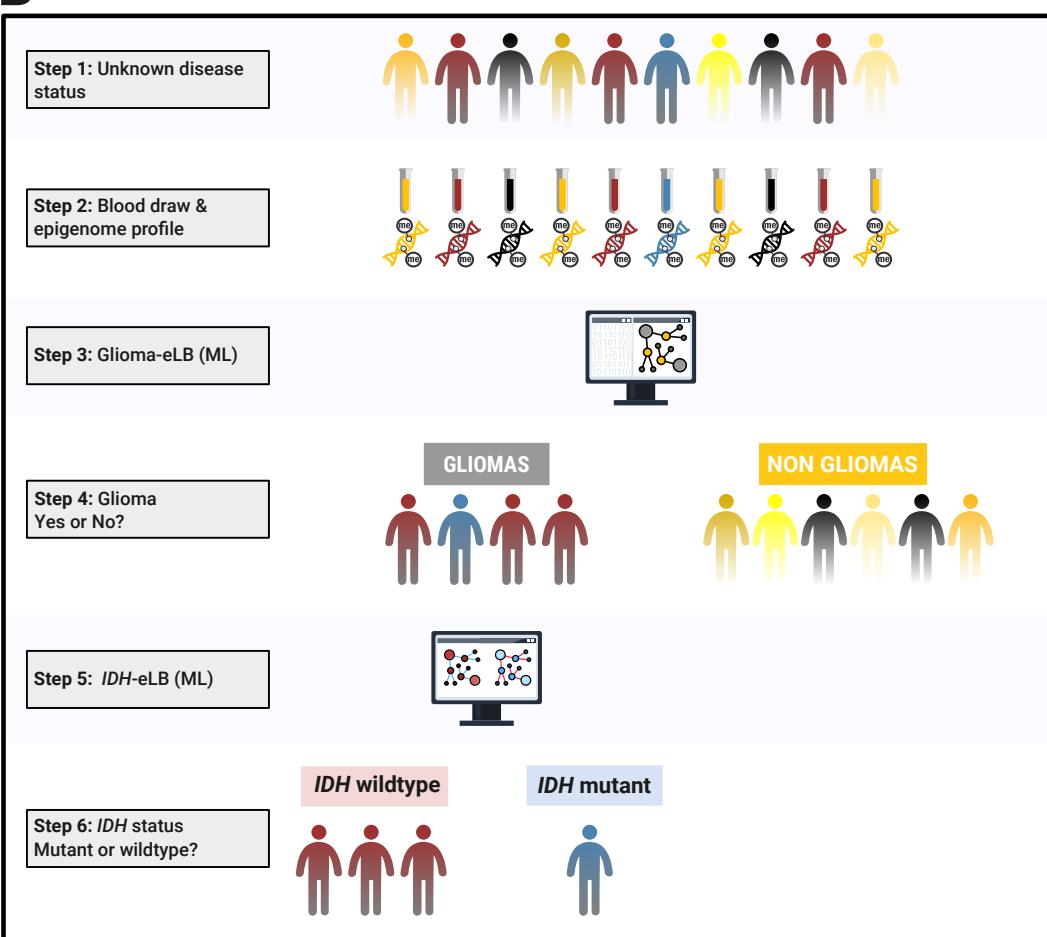
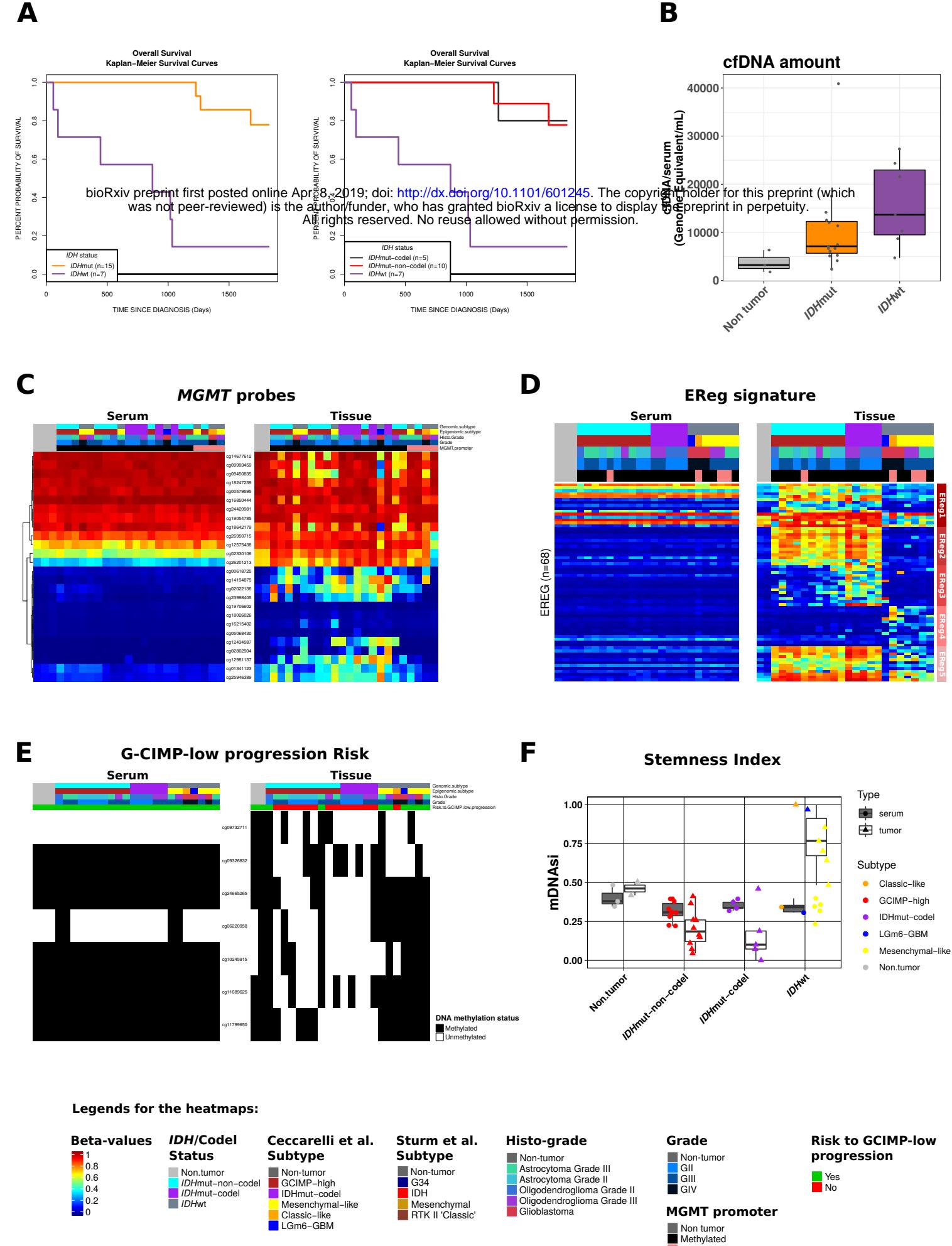
A**C****B**

Fig. 4: Proposed clinical application of Glioma- and IDH eLB - Real time diagnosis and, surveillance of early tumor progression or recurrence.

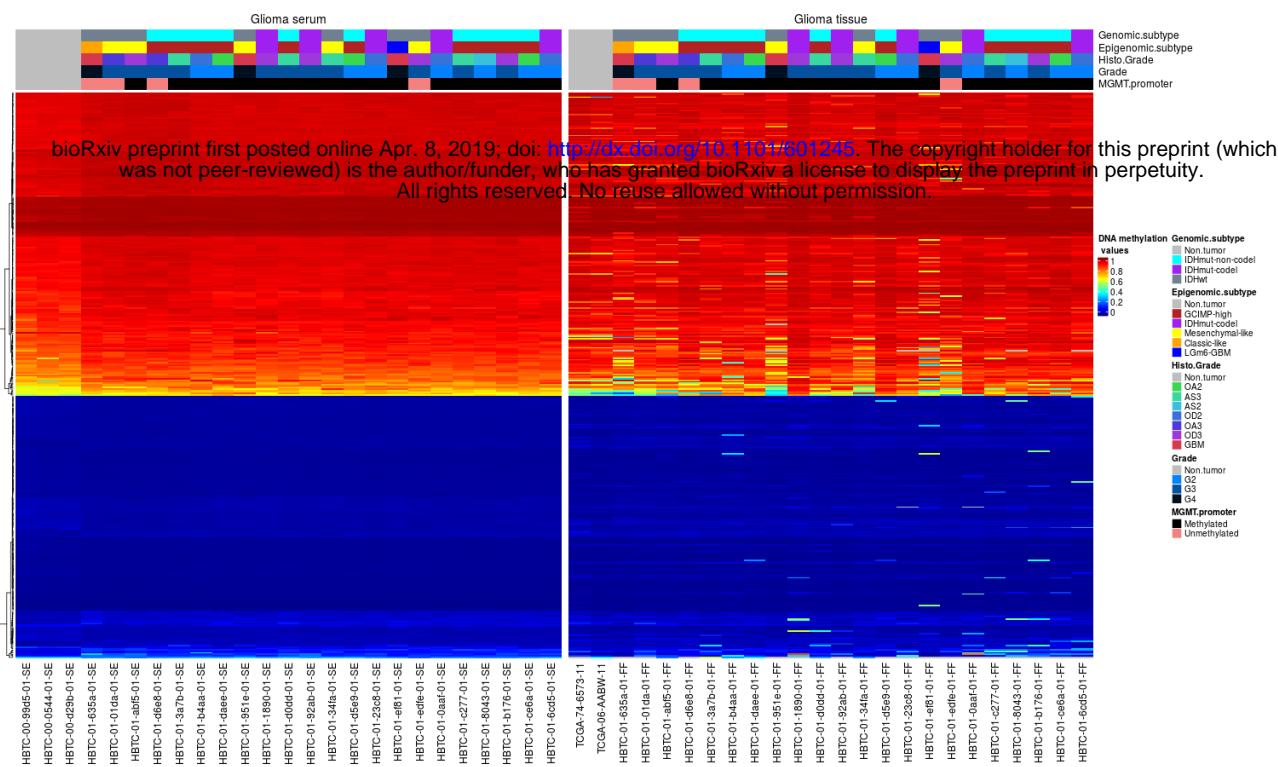
A) Steps used to generate the eLB and the application of a machine learning (ML) model to predict glioma and glioma subtypes. B) At specific intervals, currently determined by magnetic resonance imaging (MRI) visits, patients' whole-blood is collected and serum/plasma is immediately processed. cfDNA is isolated and profiled using DNA methylation microarray (profile >700,000 CpGs across the entire human genome) and entered into a ML algorithm to generate a index (Glioma-index). C) According to the predetermined index threshold (0.6, sensitivity: 98%; specificity: 99%), the new sample is classified as glioma or non-glioma and according to IDH status (mutant or wildtype). The discovered glioma specific eLB (N=1,075) can be used to complement current clinical diagnostic and monitoring events. Prognostic IDH-eLB could be used at time of diagnosis and for monitoring during active treatment through survivorship care. Complementing the MRI findings, eLB could improve detection, reduce false-positive and increase tumor identification (glioma vs necrosis vs non glioma conditions), assess treatment outcome and help tailor treatment options for patients with glioma. eLB could also foster early detection of glioma progression to improve treatment outcomes. Future clinical trials are needed to evaluate the robustness of our eLB signatures for clinical application.



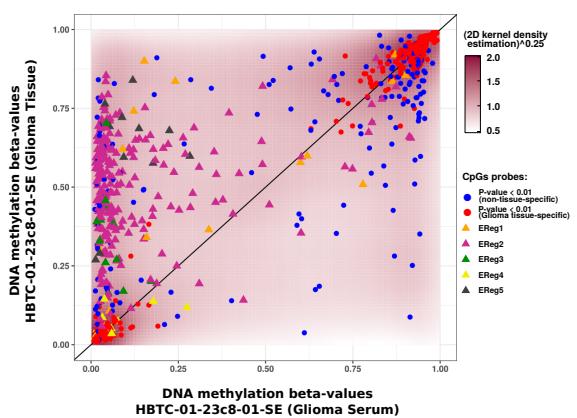
Extended Data Fig. 1.

A) Kaplan-Meier survival curves showing samples separated by *IDH* status (left) and *IDH* status combined with 1p/19q co-deletion (right). Tick represents censorship. B) Total serum cfDNA concentration normalized to the genomic size (Genomic Equivalents/ml) in non-tumor, *IDHmut* and *IDHwt* samples. C) Heatmap of DNA methylation probes mapped the promoter region of *MGMT* gene. DNA methylation beta-values are represented as a color gradient from low (blue) to high (red) in C, D, and E. D) Heatmap of DNA methylation of probes that define epigenetically regulated genes in glioma subtypes. Rows represent EReg probes described by Ceccarelli et al., 2016. E) Predictive biomarkers for glioma progression. Rows represent probes that stratify gliomas into risk for aggressive recurrence (deSouza et al., 2018). Each marker was coded as white if methylated and black if unmethylated according to the published cutoffs. F) Stemness index defined by DNA methylation as described by Malta et al. 2018. Samples are stratified by genomic group and by sample type.

A

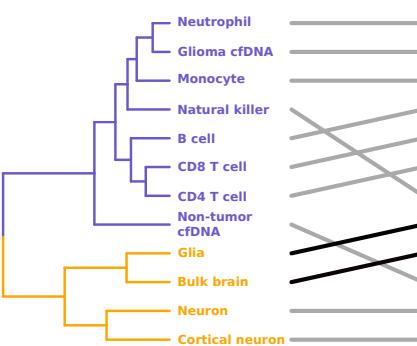


B

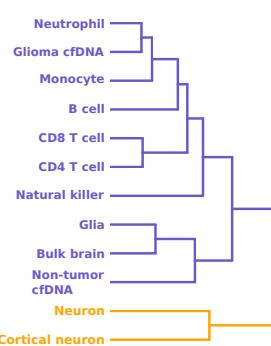


C

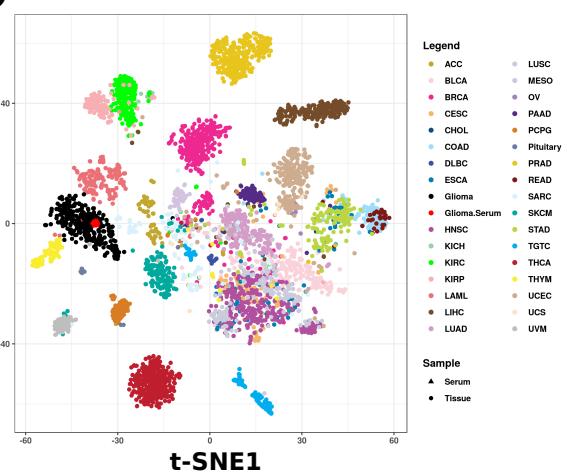
Glioma non-tissue-specific



Glioma Tissue-specific

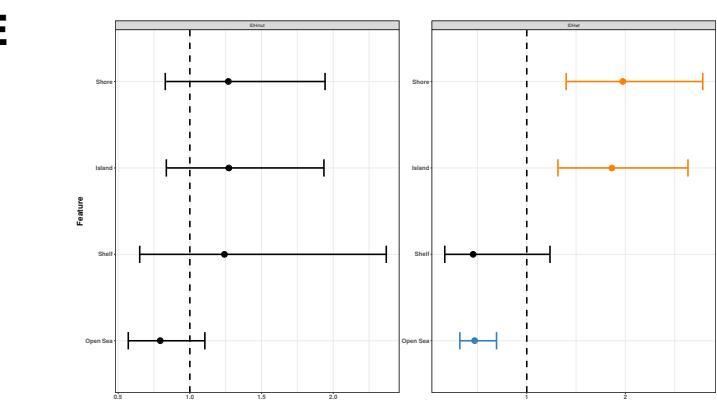
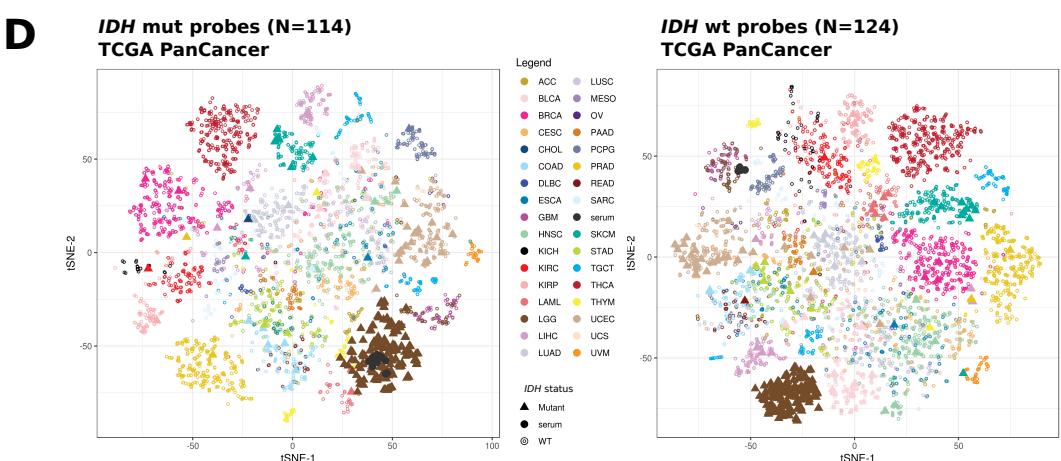
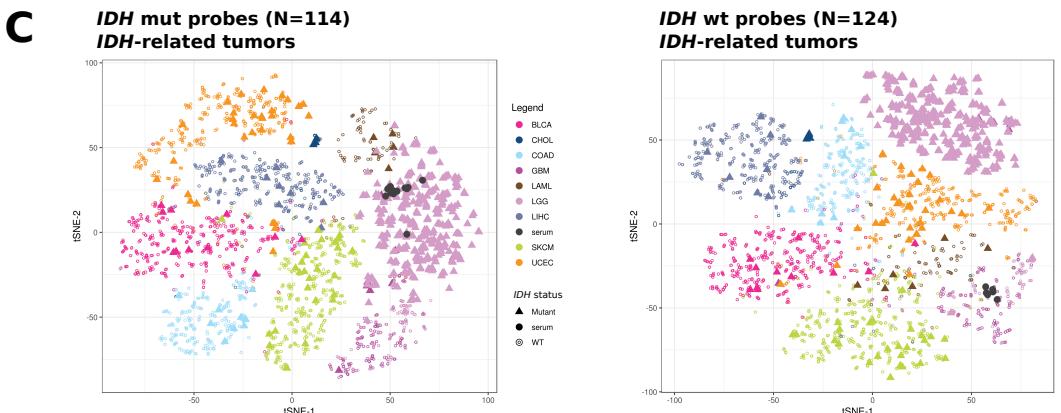
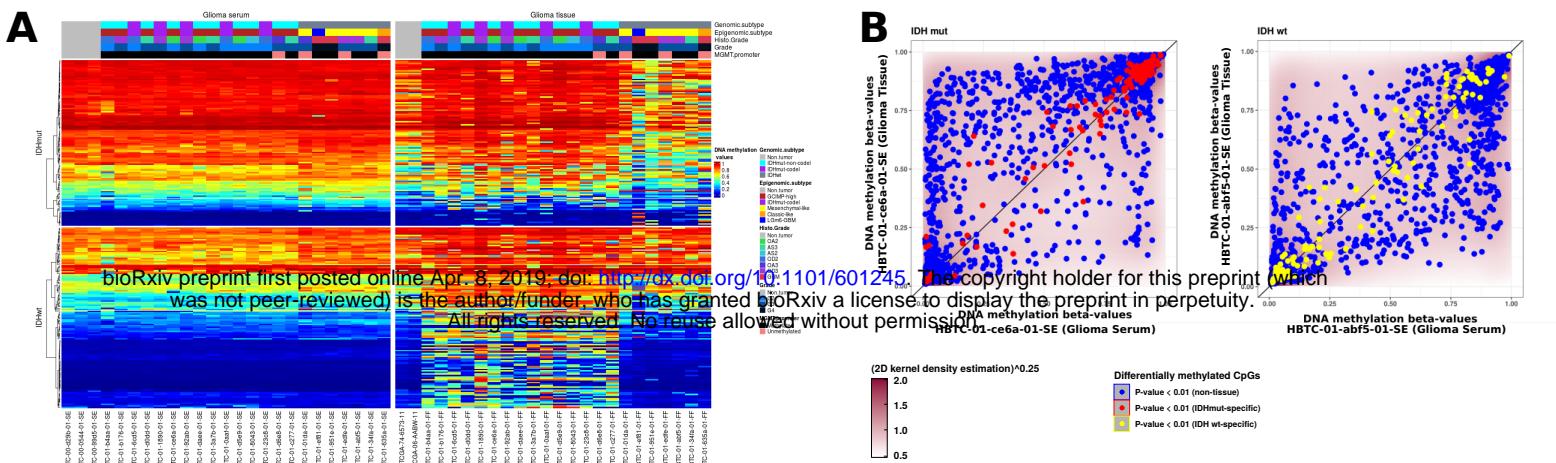


D



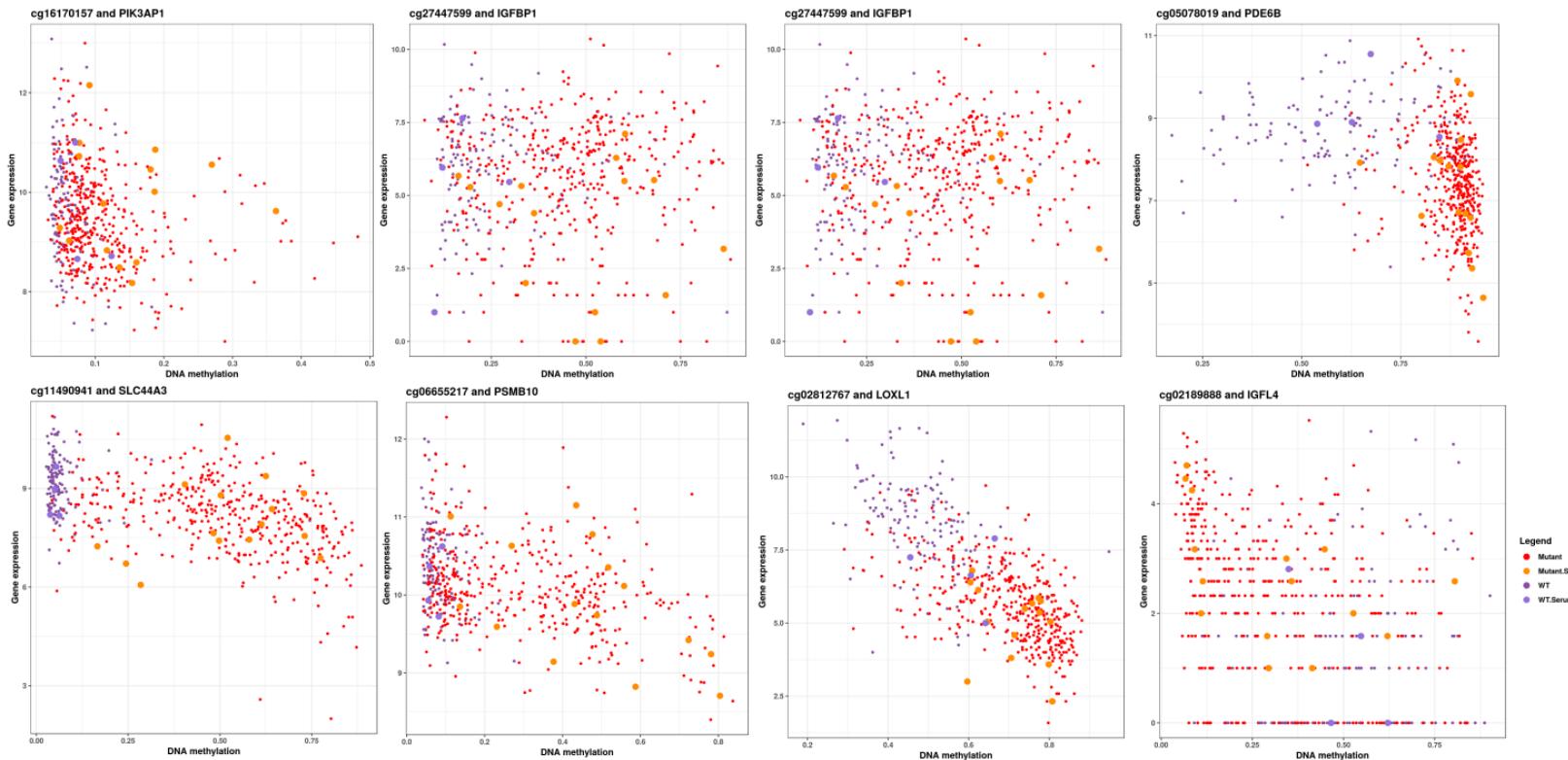
Extended Data Fig. 2.

A) Heatmap of DNA methylation of Glioma-eLB probes (N=1075 CpG sites). B) DNA methylation levels of Glioma-eLB and previously published probes (Ceccarelli et al., 2016) in the serum and in the tissue of a representative patient. C) Dendograms of non-tumor cell types and serum using Glioma non-tissue-specific CpGs (N=186, left) in comparison to tissue-specific (N=384, right) CpGs. D) Glioma-specific tissue-matching eLB (N=384) was used to subset the published primary tumor tissue DNA methylation (N=33 tumor types) and using t-SNE (t-distributed stochastic neighbour embedding) dimensionality reduction to visualize the similarities of each sample. As expected, each primary tumor type (circles) clusters with its known cell-of-origin. Serum cfDNA methylation of our patients cohort (triangles) clusters with the primary glioma tissue DNA methylation profiles.



Extended Data Fig. 3.

A) Heatmap of DNA methylation of *IDH*-eLB probes (N=114 *IDH*mut-eLB and 124 *IDH*wt-eLB CpG sites). B) Comparison between DNA methylation levels of glioma tissue (y-axis) and serum (x-axis) of one representative *IDH*mut glioma patient on the left and one representative *IDH*wt glioma patient on the right. C-D) *IDH*-eLB (N=114 *IDH*mut-eLB and 124 *IDH*wt-eLB CpG sites) was used to subset the published primary tumor tissue DNA methylation data and using t-SNE (t-distributed stochastic neighbour embedding) dimensionality reduction to visualize the similarities of each sample. C) t-SNE using *IDH*mut-eLB CpGs on the left and *IDH*wt-eLB CpGs on right with primary TCGA tumor tissue from tumor types (N=9) with known *IDH* mutation. As expected, each primary tumor type (circles) clusters with its known cell-of-origin. Serum cfDNA methylation of our *IDH*mut patients cohort (triangles) clusters with the *IDH*mut primary glioma tissue and serum cfDNA methylation of our *IDH*wt patients cohort (triangles) clusters with the *IDH*wt primary glioma tissue. D) t-SNE using *IDH*mut-eLB CpGs on the left and *IDH*wt-eLB CpGs on right with primary TCGA tumor tissue (N=33 tumor types). E) Odds-ratio for the frequencies of *IDH*mut-eLB probes (left) and *IDH*wt-eLB (right), respectively, that overlap a particular molecular feature relative to the expected genome-wide distribution of the methylation platform.

A**Extended Data Fig. 4:**

A) DNA methylation (x-axis) and expression (y-axis) scatter plot for promoter CpG associated with the corresponding gene. Each dot represents a sample. Red represents *IDHmut* glioma tissue samples, dark purple represents *IDHwt* glioma tissue samples, orange represents *IDHmut* glioma serum cfDNA and light purple *IDHwt* glioma serum cfDNA.