Single-cell chromatin accessibility in glioblastoma delineates cancer stem cell heterogeneity predictive of survival

Authors: P. Guilhamon[1,2], M.M. Kushida[2], A. Nikolic[3,4,5], D. Singhal[3,4,5], G. MacLeod[6], S.A. Madani Tonekaboni[1,7], F.M.G. Cavalli[2], C. Arlidge[1], N. Rajakulendran[6], N. Rastegar[2], X. Hao[3,8,13], R. Hassam[3,8,13], L.J. Smith[18], H. Whetstone[2], F.J. Coutinho[2], B. Nadorp[1], K.I. Ellestad[3,4], H.A. Luchman[3,8,13], J.A. Chan[3,4,16], M.S. Shoichet[18], M.D. Taylor[2,9], B. Haibe-Kains[1,7,10,11,12], S. Weiss[3,8,13,14], S. Angers[6,17], M. Gallo[3,4,5,14], P.B. Dirks[2,9,15,*], M. Lupien[1,7,11,*]

Affiliations:
[1]Princess Margaret Cancer Centre, University Health Network, Toronto, ON M5G 1L7, Canada.
[2]Developmental and Stem Cell Biology Program and Arthur and Sonia Labatt Brain tumor Research Centre, The Hospital for Sick Children, Toronto, ON M5G 0A4, Canada.
[3]Clark Smith Brain Tumour Centre, Arnie Charbonneau Cancer Institute, University of Calgary, Calgary, Alberta T2N 4N1, Canada.
[4]Alberta Children's Hospital Research Institute, University of Calgary, Calgary, Alberta T2N 4N1, Canada.
[5]Department of Biochemistry and Molecular Biology, Cumming School of Medicine, University of Calgary, Calgary, Alberta T2N 4N1, Canada.
[6]Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, ON M5S 3M2, Canada.
[7]Department of Medical Biophysics, University of Toronto, Toronto, ON M5S 1A8, Canada.
[8]Hotchkiss Brain Institute, University of Calgary, Calgary, AB T2N 4N1, Canada.
[9]Division of Neurosurgery, University of Toronto, Toronto, ON M5S 1A8, Canada.
[10]Department of Computer Science, University of Toronto, Toronto, ON M5S 1A8, Canada.
[11]Ontario Institute for Cancer Research, Toronto, ON M5G 0A3, Canada.
[12]Vector Institute, Toronto, ON M5G 0A3, Canada.
[13]Department of Cell Biology & Anatomy, University of Calgary, Calgary, AB T2N 4N1, Canada.
[14]Department of Physiology & Pharmacology, University of Calgary, Calgary, AB T2N 4N1, Canada.
[15]Departments of Molecular Genetics and Surgery, University of Toronto, Toronto, ON M5S 1A8, Canada.
[16]Department of Pathology and Laboratory Medicine, University of Calgary, Calgary, AB, Canada.
[17]Department of Biochemistry, Faculty of Medicine, University of Toronto, Toronto, ON, Canada.
[18]Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON, Canada.

*Correspondence to: Mathieu Lupien (mathieu.lupien@uhnresearch.ca) and Peter B Dirks (peter.dirks@sickkids.ca)

ABSTRACT

Chromatin accessibility discriminates stem from mature cell populations, enabling the identification of primitive stem-like cells in primary tumors, such as Glioblastoma (GBM) where self-renewing cells driving cancer progression and recurrence are prime targets for therapeutic intervention. We show, using single-cell chromatin accessibility, that primary GBMs harbor a heterogeneous self-renewing population whose diversity is captured in patient-derived glioblastoma stem cells (GSCs). In depth characterization of chromatin accessibility in GSCs identifies three GSC states: Reactive, Constructive, and Invasive, each governed by uniquely essential transcription factors and present within GBMs in varying proportions. Orthotopic xenografts reveal that GSC states associate with survival, and identify an invasive GSC signature predictive of low patient survival. Our chromatin-driven characterization of GSC states improves prognostic precision and identifies dependencies to guide combination therapies.

INTRODUCTION

Glioblastoma (GBM) is a lethal form of brain cancer with standard surgery and radiation giving a median survival of only 12.6 months[1]. The addition of temozolomide chemotherapy provides only an additional 2.5 months in the small subset of responsive patients[2]. Despite extensive characterization and stratification of the bulk primary tumors, no targeted therapies have been successfully developed[1,3]. GBM tumors are rooted in self-renewing tumor-initiating cells commonly referred to as glioblastoma stem cells (GSCs)[4] that drive disease progression *in vivo[5,6]* and display resistance to chemo- and radiotherapy leading to disease recurrence[7]. The promise of therapeutically targeting self-renewing tumor-initiating cancer cells depends on

63 our capacity to capture the full range of heterogeneity within this population from individual
64 tumors. Intratumoral heterogeneity within primary GBM has recently been documented
65 through single cell RNA-seq experiments and revealed a continuum between four cellular
66 states[8]: neural-progenitor-like (NPC), oligodendrocyte-progenitor-like (OPC), astrocyte-like
67 (AC), and mesenchymal-like (MES)[8]. A subsequent study[9] using single-cell gene-centric
68 enrichment analysis placed GBM cells along a single axis of variation from proneural to
69 mesenchymal transcriptional profiles, with cells expressing stem-associated genes lying at the
70 extremes of this axis. Hence, primary GBM consists of distinct states, across which stem-like
71 cells appear to be found. Whether these stem-like cells found across GBM states represent
72 functionally distinct GSC populations with tumor-initiating properties and unique dependencies
73 remains to be established to guide therapeutic progress. To address this issue, we combined
74 single-cell technologies to define GSC composition in primary GBM with functional assays to
75 reveal the unique dependencies across GSCs, reflective of invasive, constructive and reactive
76 states that relate to patient outcome.

77

78 **RESULTS**

79 Chromatin accessibility readily discriminates stem from mature cell populations[10],
80 which can be resolved at the single cell level through single-cell ATAC-seq[11,12] taking into
81 account non-gene centric features, such as accessibility of noncoding elements and total
82 amount of accessible DNA sequences. Applying single-cell chromatin accessibility profiling
83 (scATAC-seq) across four *IDH* wild-type primary adult GBM tumors (3797 cells) revealed 7-9
84 accessibility modules in each tumor based on unsupervised clustering (Fig. 1A). We assigned

85 cells to each of the four scRNA-seq-derived cellular states[8] based on individual cells' chromatin
86 accessibility enrichment scores for the promoter regions of each state's signature genes. The
87 MES state reported from scRNA-seq[8] dominates the identity of two or more modules reported
88 from chromatin accessibility in every tumor (Fig. 1B-C). In contrast, the NPC and OPC states are
89 mixed within the same module defined based on chromatin accessibility, dominating over the
90 other states typically in at least two modules. Cells assigned to the AC state did not preferentially
91 cluster within a single module reported from chromatin accessibility (Fig. 1B-C). Collectively, our
92 results suggest that chromatin accessibility reflects a greater stratification of the MES state,
93 detects similarities between the OPC and NPC states and heterogeneity within the AC state.

94 To identify putative cancer stem cells within each primary tumor, we next focused on the
95 level of chromatin accessibility at promoters of 19 transcription factors previously associated
96 with self-renewal and tumor-propagating capacity in GBM[13] (Fig. 1D). Individual cells scoring
97 as putative cancer stem cells were not restricted to a unique module defined by chromatin
98 accessibility but were distributed across a subset of modules, suggesting heterogeneity across
99 cancer stem cells in primary GBM, in agreement with reports relying on single GBM cell
100 labelling[6,10,12–20] assessing the heterogeneity of self-renewing tumor-initiating cells.
101 Putative cancer stem cells identified in primary GBM through scATAC-seq were found in
102 modules ascribed to every one of the four cellular states defined by gene expression[8],
103 predominantly within NPC and OPC containing modules and a smaller fraction (<10%) in MES-
104 specific modules across all four tumors (Fig. 1E). This suggests that the core transcriptional unit
105 of cancer stem cells in primary GBM[13] is not restricted to a unique population defined by its

106 global transcriptional or chromatin accessibility profile with the resolution achieved with current
107 single-cell technologies.

108 To further probe the heterogeneity in chromatin accessibility within the GBM cancer
109 stem cell pool, we derived GSC populations from 27 adult *IDH* wild-type GBM tumors[21] and
110 profiled their chromatin accessibility by bulk ATAC-seq (Fig. 2A). Each patient-derived GSC
111 showed a similar enrichment for accessible chromatin regions in promoters and 5'UTRs, and
112 depletion in introns and distal intergenic regions (Fig. 2B). Collectively, we uncovered 92% of
113 the total predicted regions of accessible chromatin (255,891 regions) within GSCs based on a
114 saturation analysis using a self-starting nonlinear regression model across the 27 samples (Fig.
115 2C). We next assessed the similarity between these GSCs and the putative cancer stem cells
116 found by scATAC-seq in the four primary GBMs. GSCs were identified within each tumor by
117 calculating the enrichment of accessible chromatin regions shared by a majority of GSCs
118 (>14/27) in each tumor cell (Fig. 2D). Comparing the distribution of GSCs across the 7-9 modules
119 defined by scATAC-seq to that of the 19 transcription factor-derived cancer stem cell signature
120 demonstrates concordance between the two signatures (Fig. 2E). Moreover, the enrichment z-
121 scores for both cancer stem signatures (ie stem transcription factors signature and GSC
122 chromatin accessibility signature) are significantly correlated across cells in all four tumors ($p \leq$
123 $1.6e^{-5}$) (Fig. 2F). Additionally, an average of 91.2% (85.1-100%) of the cells identified by either
124 signature display the hallmark GBM copy number changes at chromosomes 7 and 10, confirming
125 their neoplastic status (Fig. S1A-D). Collectively, these results demonstrate that the patient-
126 derived GSC populations reflect the chromatin identity of putative cancer stem cells found in
127 primary brain tumors, highlighting the value of these GSCs as models to deepen our

128 understanding of individual cells within primary GBM with features found in self-renewing
129 tumor-initiating cells. Accordingly, spectral clustering of the 27 patient-derived GSC ATAC
130 profiles identifies three distinct states of self-renewing tumor-initiating cells (Fig. 3A).
131 Expression profiling of these 27 GSCs by RNA-seq reveals GSCs significantly enriched for the
132 signatures of each of the three TCGA GBM subtypes (proneural, classical, and
133 mesenchymal)[22] (Fig. 3B, top panel). However, the assignment of the proneural, classical and
134 mesenchymal subtypes across GSCs did not match the three clusters identified from ATAC-seq
135 (Fig. 3B, top panel). Conversely, clustering the GSCs by gene expression, independently of their
136 chromatin accessibility, did largely recapitulate the GSC states defined by chromatin
137 accessibility (Fig. 3B, bottom panel). This suggests that the mismatch between the GSC cluster
138 from chromatin accessibility profiles and TCGA expression subtypes is not mainly due to
139 differences between chromatin accessibility and gene expression. A more likely possibility is
140 that given the TCGA subtypes were determined from bulk GBM, they may not fully capture the
141 nature of rarer populations found within a tumor, such as the cancer stem cell populations.

142 Gene set enrichment analysis with GSEA[23] and g:profiler[24] using genes exclusively
143 enriched for both expression and promoter chromatin accessibility in each subtype reported
144 significantly enriched terms defining the largest GSC state as a Reactive state, with terms
145 related to immune cells and response (Fig. 3C, top panel). A second GSC state was enriched for
146 Constructive gene sets involved in brain, neuron, and glial cell development (Fig. 3C, middle
147 panel). The third and smallest GSC state presented an Invasive state characterized by terms
148 relating to the extracellular matrix and angiogenesis (Fig. 3C, bottom panel). We next mapped
149 copy number alterations (CNAs) across the 27 GSCs by applying the molecular neuropathology

150 classifier tool[25] to DNA methylation data from the GSCs (Fig. 3D). While common GBM CNAs,
151 including EGFR gains and CDKN2A/B loss, were observed, the CNA-based classification of GSCs
152 failed to match the three chromatin accessibility-derived states, suggesting the three GSC
153 states are not genetically defined (Fig. 3D). Further comparison of the accessible chromatin in
154 each GSC state reveals that only a small subset of accessible chromatin regions drives the three
155 GSC states (Fig. 3E and Fig. S1E-G). Our ability to discriminate GSC state-specific regions of
156 accessible chromatin is reflective of the comprehensiveness of our cohort to saturate the
157 detection of accessible regions to 93%, 88%, and 71% across the Reactive (n=13), Constructive
158 (n=9), and Invasive (n=5) state GSCs, respectively (Fig. 3F).

159 Considering that regions of accessible chromatin serve as binding sites for transcription
160 factors engaging in gene expression regulation, we next tested for DNA recognition motif family
161 enrichment across regions exclusively accessible in Reactive, Constructive or Invasive GSC states
162 (Fig. 4A and Fig. S2A). The most enriched DNA recognition motif families in each state were
163 either depleted or showed low-level enrichment in the other states. Specifically, the DNA
164 recognition motifs for the interferon-regulatory factor (IRF) and Cys2-His2 zinc finger (C2H2 ZF)
165 transcription factor families were enriched in the Reactive state (Fig. 4A, top panel). Regulatory
166 factor X (RFX) and basic helix-loop-helix (bHLH) DNA recognition motifs were enriched in the
167 Constructive state (Fig. 4A, middle panel), while the Forkhead motif family was enriched in the
168 Invasive state (Fig. 4A, bottom panel). Genome-wide CRISPR/Cas9 essentiality screens (Fig.
169 S2B) in three Reactive, two Constructive and one Invasive GSC[26] revealed the preferential
170 requirement for expressed transcription factors (Fig. S2C-E) recognizing the enriched DNA
171 recognition motif in a state-specific manner (Fig. 4B). Specifically, the SP1 regulatory network

172 is preferentially essential in the Reactive state GSCs (Fig 4B, top panel), ASCL1, OLIG2, AHR,
173 and NPAS3 are uniquely essential in the Constructive state GSCs (Fig. 4B, middle panel) and
174 FOXD1 in essential only in the Reactive state GSC (Fig. 4B, bottom panel). Notably, SP1 itself is
175 exclusively essential in only one Reactive GSC (G564). However, of the 36 transcription factors
176 from the Reactive-enriched families (IRF and C2H2 ZF) that were essential in at least one
177 Reactive GSC and not in any of the Constructive or Invasive GSCs, 13 are directly regulated by
178 SP1 (Fig. S2C), thus suggesting that the SP1 regulatory network as a whole, rather than SP1 on
179 its own, is key in the Reactive GSC state. Notably, all six transcription factors display significantly
180 higher expression in GBM compared to normal brain[27] (Fig. S2F), further supporting their
181 function as key regulators of tumor initiation and development. An additional gene set
182 enrichment analysis combining genes exclusively essential to each state with the putative
183 targets of the key transcription factors outlined above identifies additional enriched terms
184 supporting the identities of the three GSC states as Reactive, Constructive, and Invasive (Fig.
185 S2G).

186 Previous work suggests that GBM tumors harbor a heterogeneous population of
187 GSCs[14,15,28]. We therefore quantified the presence of Reactive, Constructive and Invasive
188 cancer stem cells in our four primary GBM based on their scATAC-seq profiles. The Constructive
189 state was dominant in every primary tumor, ranging from 9-21% of all cells captured by scATAC-
190 seq (Fig. 4C). The Reactive and Invasive states accounted for only 0-9% of all cells, in varying
191 proportions from one tumor to another (Fig. 4C). Collectively, our results further support the
192 heterogeneous nature of cancer stem cells that populate primary tumors.

While various classifications of GBMs and/or their constitutive bulk and stem tumor cells have been reported, with some associating with patient survival[8,22,28–31], no molecular signature in GBM has so far been reported that can significantly stratify the poorer prognosis *IDH* wild-type patients by survival. We performed intracranial xenografts of 37 *IDH* wild-type GSC populations and classified the transplanted cells by their GSC state to perform a differential survival analysis (Fig. 4D). The overall survival times of the transplanted mice grouped by GSC state were significantly different (LogRank test $p = 0.041$), with the Invasive state GSCs leading to the worst prognosis. Relative to the mice injected with Reactive state GSCs, mice with Constructive state GSCs and Invasive state GSCs had hazard ratios of 1.3 (95% CI: 0.57-2.97) and 3.5 (95% CI: 1.2-10.49), respectively. Next, we investigated the prognostic value of the GSC states using the TCGA GBM cohort (*IDH* wild-type, $n = 144$). When classified by dominant GSC state, TCGA tumors display the same trend as the xenografts with Invasive state-dominated tumors showing the lowest survival (Fig. S3A). However, with only two tumors classified as Invasive-dominant, the difference in survival between the three patient groups was not statistically significant ($p = 0.3$) (Fig. S3A). We proceeded to rank the TCGA tumors solely by their concordance to Invasive GSCs and classified the patient tumors into Invasive-low (z-score < 1), Invasive-mid (z-score 1-1.65), and Invasive-high (z-score ≥ 1.65) groups (Fig. 4E). With this stratification method, median patient survival per group not only decreased with increasing Invasive GSC score, but we also identified an Invasive-high subset of tumors with significantly lower survival ($p = 0.019$, HR = 2.8, 95% CI: 1.3-5.81)) (Fig. 4F). Invasion assays using representative Invasive state GSCs additionally highlight the invasive properties of these populations both *in vitro* and *in vivo* (Fig. S4). These results show that cancer stem cell states

215 defined based on the chromatin accessibility in GSCs can identify transcriptional programs
216 associated with poor prognosis and can serve as a signature to identify high-risk patients in *IDH*
217 wild-type GBM.

218

219 **DISCUSSION**

220 Defining the nature of self-renewing tumor-initiating cells in primary GBM is required to
221 identify vulnerabilities for therapeutic intervention. Quantifying their heterogeneity within
222 tumors can guide treatment strategies and assist in predicting the course of disease progression.
223 Here we show that chromatin accessibility assays capture a heterogeneity across self-renewing
224 tumor-initiating cells in primary GBM that extends beyond their genetic diversity, and underlies
225 the heterogeneity in bulk progeny[15]. This heterogeneity aligns with diversity in the three-
226 dimensional genome organization of GSCs[32] and agrees with how the three-dimensional
227 genome organization instructs cis-regulatory plexuses underlying gene regulation[33–37]. We
228 further reveal a specific cancer stem state that is significantly predictive of patient survival and
229 can be used as a signature to identify high-risk patients. Our results also highlight dependencies
230 unique to each cancer stem state. Specifically, the Reactive GSC state-associated transcription
231 factor SP1 and its regulatory partners are involved in cellular differentiation and growth,
232 apoptosis, response to DNA damage, chromatin remodelling[38], stimulation of *TERT*
233 expression in cancer stem cells[39], and increased stemness and invasion in GBM[40]. In
234 contrast, the Constructive GSC state rely on transcription factors including OLIG2, a known GSC
235 marker[41], AHR involved in tumor microenvironment responses and metabolic adaptation[42],
236 NPAS3, a regulator of Notch signaling and neurogenesis[43] and ASCL1, a critical regulator of

237 GSC differentiation and marker of sensitivity to Notch inhibition in GSCs[19,44]. Finally, the
238 Invasive GSC state relies on FOXD1, a pluripotency regulator and determinant of tumorigenicity
239 in GSCs regulating expression of the aldehyde dehydrogenase ALDH1A3, a functional marker
240 for invasive GSCs[45,46]. Collectively, our results support developing combination therapy using
241 targeting agents against each GSC state, such as Notch inhibitors[19] and small molecule
242 inhibitors of ALDH[45], to eradicate self-renewing tumor-initiating cells with the hope to cure
243 GBM patients.

244

245 **METHODS**

246 Patient samples and cell culture

247 All tissue samples were obtained following informed consent from patients, and all
248 experimental procedures were performed in accordance with the Research Ethics Board at The
249 Hospital for Sick Children (Toronto, Canada), the University of Calgary Ethics Review Board, and
250 the Health Research Ethics Board of Alberta - Cancer Committee (HREBA). Approval to
251 pathological data was obtained from the respective institutional review boards. Patient tumor
252 tissue samples were dissociated in artificial cerebrospinal fluid followed by treatment with
253 enzyme cocktail at 37°C. Patient tumor-derived GSCs were grown as adherent monolayer
254 cultures in serum-free medium as previously described[21]. Briefly, cells were grown adherently
255 on culture plates coated with poly-L-ornithine and laminin. Serum-free NS cell self-renewal
256 media (NS media) consisted of Neurocult NS-A Basal media, supplemented with 2 mmol/L L-
257 glutamine, N2 and B27 supplements, 75 µg/mL bovine serum albumin, 10 ng/mL recombinant
258 human EGF (rhEGF), 10 ng/mL basic fibroblast growth factor (bFGF), and 2 µg/mL heparin. A

259 subset (22/37) of the GSCs used for orthotopic xenografts were grown as non-adherent spheres
260 prior to single-cell dissociation and injection into the mice. Briefly, serum-free medium (SFM)
261 was used to initiate GSC cultures. Non-adherent spheres formed after 7-21 days in culture and
262 were expanded, then cryopreserved in 10% dimethyl sulfoxide (DMSO; Sigma-Aldrich) in SFM
263 until used in experiments.
264
265 ATAC-seq
266 ATAC-seq was used to profile the accessible chromatin landscape of 27 patient tumor-
267 derived GSCs. 50,000 cells were processed from each sample as previously described[47,48]. The
268 resulting libraries were sequenced with 50 bp single-end reads which were mapped to hg19.
269 Reads were filtered to remove duplicates, unmapped or poor quality (Q <30) reads,
270 mitochondrial reads, chrY reads, and those overlapping the ENCODE blacklist. Following
271 alignment, accessible chromatin regions/peaks were called using MACS2. Default parameters
272 were used except for the following: --keep-dup all -B --nomodel --SPMR -q 0.05 --slocal 6250 --
273 llocal 6250. The signal intensity was calculated as the fold enrichment of the signal per million
274 reads in a sample over a modelled local background using the bdgcmp function in MACS2.
275 Spectral clustering implemented in the SNFtool package[49] was run on the SNF fused similarity
276 matrix to obtain the groups corresponding to k=2 to 12. Enrichment for genomic features was
277 calculated using CEAS[50].
278 A given chromatin region was considered exclusive to one of the clusters if it was called
279 as a peak in any of the cluster's samples using a q-value filter of 0.05 and was not called as a peak

280 in any of the other samples using a q-value filter of 0.2, in order to ensure stringency of
281 exclusivity.

282 The ATAC-seq saturation analysis was performed by randomizing the order of samples,
283 and successively calculating the number of additional peaks discovered with the addition of each
284 new sample. This process was repeated 10,000 times and averaged. A self-starting non-linear
285 regression model was then fitted to the data to estimate the level of saturation reached.

286 For the xenograft survival analysis, 11/37 GSCs used overlap with the cohort of 27
287 described above. The other 26/37 GSCs were profiled by ATAC-seq independently following the
288 same protocol described above and assigned to a GSC state through unsupervised hierarchical
289 clustering with the original cohort of 27 GSCs.

290

291 Single cell ATAC-seq

292 The four tumors used were G4218 (primary GBM, IDH wt, Male, 64yrs), G4250 (primary
293 GBM, IDH wt, Male, 73yrs), G4275 (primary GBM, IDH wt, Female, 52yrs), G4349 (primary GBM,
294 IDH wt, Male, 62 yrs). Fragments of tumor were received fresh from the operating room, and
295 blunt dissected into individual fragments of approximately 0.3-0.7 cm3. Each fragment was
296 placed in 1 mL of freezing media (400 µL of NeuroCult NS-A Basal medium with proliferation
297 supplement (StemCell Technologies; #05751) containing 20 µg/mL rhEGF (Peprotech, AF-100-
298 15), 10 µg/mL bFGF (StemCell Technologies, #78003), and 2 µg/mL heparin (StemCell
299 Technologies, #07980); 500 µL of 25% BSA (Millipore-Sigma; A9647) in DMEM, and 100 µL
300 DMSO (Millipore-Sigma; D2650) in a 2 mL cryotube, and placed at -80 C in a CoolCell for at least
301 24 hours. Samples were then stored at -80C until use. Cryopreserved primary GBM samples were

302 washed at 1000 RPM for 5 minutes in PBS to remove DMSO, and then transferred to 1.5 mL
303 tubes. Samples were resuspended in cold ATAC resuspension buffer (10 mM Tris-HCl pH 7.4, 10
304 mM NaCl, 3 mM MgCl2, 0.1% NP-40, 0.1% Tween-20, 0.01% Digitonin, 1% BSA in PBS) on ice
305 and dissociated using a wide-bore P1000 pipette tip and vortexing, followed by 10 minutes of
306 incubation on ice. Cells were spun down at 500x g for 5 minutes at 4 C, washed in the ATAC
307 resuspension buffer, spun down again, and resuspended in ATAC-Tween wash buffer (10 mM
308 Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% Tween-20, 1% BSA in PBS), then passed
309 through a cell strainer top FACS tube (Falcon; #38030) to remove debris. Nuclei quality and
310 quantity was evaluated using trypan blue on an Invitrogen Countess II device in duplicate, and a
311 subset of nuclei was spun down in a fresh tube and resuspended in 10X sample dilution buffer.
312 Nuclei were then used for single cell ATAC-seq library construction using the Chromium Single
313 Cell ATAC Solution v1.0 kit (10X Genomics) on a Chromium controller. Completed libraries were
314 further quality checked for fragment size and distribution using an Agilent TapeStation prior to
315 sequencing. Single-cell ATAC-seq samples were sequenced on a NextSeq 500 (Illumina)
316 instrument with 50 bp paired-end reads at the Centre for Health Genomics and Informatics
317 (CHGI) at the University of Calgary.

318 The raw sequencing data was demultiplexed using cellranger-atac mkfastq (Cell Ranger
319 ATAC, version 1.0.0, 10x Genomics). Single cell ATAC-seq reads were aligned to the hg19
320 reference genome (hg19, version 1.1.0, 10x Genomics) and quantified using cellranger-atac
321 count function with default parameters (Cell Ranger ATAC, version 1.1.0, 10x Genomics). The
322 resulting data were analysed using the chromVAR[51] and Signac[52] R packages (v1.4.1). The
323 number of accessibility modules in each sample was determined using the ElbowPlot method

324 implemented in Signac. Similarity between individual cells and GSC states was assessed using
325 the deviation scores calculated by chromVAR within the single cell data for significantly
326 differentially accessible sets of peaks (Fold Change Signal difference >2 and Wilcoxon test q-
327 value <=0.05) between the states as determined by bulk ATAC-seq. Similarity between
328 individual cells and the expression-derived cellular states was assessed using the deviation
329 scores calculated by chromVAR within the single cell data for promoter regions of the signature
330 genes of each of the cellular states[8]. A 2-fold cut-off was used to determine dominance of a
331 UMAP module by an individual or group of cellular states. Similarity between individual cells and
332 the GBM cancer stem cell signatures was assessed using the deviation scores calculated by
333 chromVAR within the single cell data for promoter regions of the 19 transcription factors
334 identified as markers of cancer stem cells in GBM[13].

335 Copy number variants in single cells were determined using CONICSmat[53] with default
336 parameters using the gene activity matrix generated by Signac as input. We focused on chr7
337 gains and chr10 losses as they are hallmark chromosomal changes in GBM and found the
338 following fractions of cells carrying these CNVs, on average across the four tumors: 76% of all
339 cells, 88% of cells allocated to scRNAseq cellular states[8], 95% of cancer stem cells based on
340 the 19 gene signature, 91% of GSCs based on shared accessible regions between 14/27 GSC
341 populations, 94% of GSCs identified based on the state-specific signatures.

342

343 <u>DNA Methylation arrays</u>

344 Bisulfite conversion of DNA for methylation profiling was performed using the EZ DNA
345 Methylation kit (Zymo Research) on 500 ng genomic DNA from all 27 samples. Conversion

346 efficiency was quantitatively assessed by quantitative PCR (qPCR). The Illumina Infinium
347 MethylationEPIC BeadChips were processed as per manufacturer's recommendations. The R
348 package ChAMP v2.6.4[54] was used to process and analyse the data. For the copy number
349 analysis, the raw IDAT files were uploaded to the MNP tool[25], which directly compares the
350 copy number profile estimated from the probe intensities on the methylation array to the
351 distribution observed across thousands of brain tumors in its database.
352
353 RNA-seq
354 RNA was extracted from GSCs using the Qiagen RNeasy Plus kit. RNA sample quality
355 was measured by Qubit (Life Technologies) for concentration and by Agilent Bioanalyzer for
356 RNA integrity. All samples had RIN above 9. Libraries were prepared using the TruSeq Stranded
357 mRNA kit (Illumina). Two hundred nanograms from each sample were purified for polyA tail
358 containing mRNA molecules using poly-T oligo attached magnetic beads, then fragmented
359 post-purification. The cleaved RNA fragments were copied into first strand cDNA using reverse
360 transcriptase and random primers. This is followed by second strand cDNA synthesis using
361 RNase H and DNA Polymerase I. A single "A" base was added and adapter ligated followed by
362 purification and enrichment with PCR to create cDNA libraries. Final cDNA libraries were verified
363 by the Agilent Bioanalyzer for size and concentration quantified by qPCR. All libraries were
364 pooled to a final concentration of 1.8nM, clustered and sequenced on the Illumina NextSeq500
365 as a pair-end 75 cycle sequencing run using v2 reagents to achieve a minimum of ~40 million
366 reads per sample. Reads were aligned to hg19 using the STAR aligner v2.4.2a [55] and transcripts
367 were quantified using RSEM v1.2.21[56] or vst transformed using DESeq2[57].

368

369 Motif Enrichment

370 Regions exclusively accessible in one of the GSC states and not the others were used as
371 input sequences for the motif enrichment, while the full ATAC-seq catalogue served as the
372 background set when running HOMER v4.7 to detect enrichments of transcription factor binding
373 motifs. Enriched motifs were then grouped into families based on similarities in DNA-binding
374 domains using the CIS-BP database[58]. Each family was assigned the fold-enrichment value of
375 the most enriched motif within the family.

376 The transcription factors whose motifs were found enriched in Reactive-exclusive
377 accessible regions, were run together through GSEA[23], and the gene set corresponding to
378 genes potentially regulated by SP1 was identified as significantly enriched (GSEA gene set
379 GGGCGGR_SP1_Q6). The expression levels of key transcription factors in tumor and normal
380 samples were analysed and displayed using GEPIA[27].

381

382 Gene essentiality screen

383 Illumina sequencing reads from genome-wide TKOv1 CRISPR screens in patient-derived
384 GSCs[26] were mapped using MAGECK[59] and analysed using the BAGEL algorithm with
385 version 2 reference core essential genes/non-essential genes[60,61]. Resultant raw Bayes Factor
386 (BF) statistics were used to determine essentiality of transcription factor genes using a minimum
387 BF of 3 and a 5% FDR cut-off. For visualisation purposes only, the essentiality scores were scaled
388 and the individual GSC essentiality thresholds subtracted from each score to obtain a common
389 threshold at 0 across GSCs.

390

391 Orthotopic xenografts

392 All animal procedures were performed according to and approved by the Animal Care
393 Committee of the Hospital for Sick Children or the University of Calgary. All attempts are made
394 to minimize the handling time during surgery and treatment so as not to unduly stress the
395 animals. Animals are observed daily after surgery to ensure there are no unexpected
396 complications. For intracranial xenografts, 100,000 GSC cells were stereotactically injected into
397 the frontal cortex of 6-8 weeks old female NOD/SCID or C17/SCID mice. Mice were monitored
398 and euthanized once neurological symptoms were observed or at the experimental endpoint of
399 12 months.

400

401 Invasion assay

402 Hydrogels were synthesized as previously described[62], with the following
403 modifications: 1% w/v hyaluronan-methyl furan, 2.3mM MMP cleavable crosslinker, and 400µM
404 fibronectin-derived peptide with sequence Mal-SKAGPHSRNGRGDSPG. Cells were plated on
405 hydrogels at a density of 3500 cells/hydrogel and allowed to adhere for 24 h. 48 h after seeding,
406 fresh media was added to each well. Cells were fixed with 4% PFA 4-5 days after seeding. Cells
407 were counterstained with Hoechst (1/500) to label nuclei, Alexa FluorTM 488 Phalloidin (1/40) to
408 label F-actin, and 15 µm FluosphereTM red beads (1/15) to label the surface of the hydrogel.
409 Hydrogels were imaged using confocal microscopy at 10x magnification, taking images every 20
410 µm on the z-axis. Imaris Bitplane 8.3.1 software was used to prepare images and to analyze the
411 positions of each cell and surface bead label. A custom Matlab script was used to calculate the

412 percent invasion, defined as the percentage of total cells located below a 75 μm threshold from
413 the surface of the hydrogel. Statistics were calculated using Graphpad Prism 7.04 software. Plots
414 are shown as mean with standard deviation, and statistics shown are from one-way ANOVA with
415 multiple comparisons with Dunnet's test correction, comparing all patient cell lines (dark grey)
416 to the hf6562 healthy control. Data were displayed with p values represented as * $p \leq 0.05$, ** p
417 ≤ 0.01, *** $p \leq 0.001$, **** $p \leq 0.0001$.
418
419 Immunohistochemistry
420 Tissue samples were formalin fixed and paraffin embedded. Serial sections
421 deparaffinized, rehydrated through an alcohol gradient to water and antigen retrieval in citrate
422 buffer pH 6.0 was used for the human nucleolin antibody at 5.0 g/mL (ab13541) (Abcam,
423 Canbridge, MA). Endogenous peroxide activity and nonspecific binding was blocked with
424 3%(v/v) peroxide and 2% (v/v) normal horse serum. Primary antibody and anti-mouse
425 ImmPRESS-HRP secondary antibody were incubated for 1 hr and visualized using DAB (3,3'-
426 diaminobenzidine) (Vectorlabs, Burlingame, CA). Normal horse serum or monoclonal IgM was
427 used in control sections.
428
429 Survival analysis
430 Survival analysis on xenografts and TCGA data was performed using R packages
431 survival[63] and survminer[64]. The LogRank test was used in every analysis. See ATAC-seq
432 section for details on how each GSC used in the orthotopic xenografts was assigned to a GSC
433 state. TCGA samples were assigned to individual GSC states in the following way. 1) Using the

434 unsupervised clustering of RNA-seq data presented in Fig. 3B, the 23/27 GSCs that displayed
435 matched GSC state assignments by RNA-seq and ATAC-seq were used in this analysis. 2) Genes
436 preferentially enriched in each GSC state were determined using DEseq2[57] ($q \leq 0.05$ and fold-
437 change ≥ 2). 3) The mean log2(FPKM+1) value for each of these genes over all GSCs in each state
438 was calculated to obtain a single representative value for each gene in each of the three GSC
439 states. 4) the concordance index was then calculated between each TCGA sample and each GSC
440 state and individual TCGA samples were assigned to the GSC state with the highest score.
441 Similarly, to assign TCGA samples to the three Invasive groups (Invasive-low, -mid, and -high),
442 the concordance to Invasive GSCs as calculated above was used. The z-score for each sample
443 was then used to classify each TCGA sample into the three subgroups of Invasive-low (Invasive
444 z-score < 1), Invasive-mid (Invasive z-score 1-1.65), and Invasive-high (Invasive z-score ≥ 1.65).
445 When changing the Invasive z-score thresholds for grouping the TCGA samples, the most
446 Invasive-high subgroup remains associated with the lowest survival (Fig. S3B-C).

447

448 Data and materials availability

449 The GSCs are available upon reasonable request from PBD and SW. The GSC ATAC-seq
450 and DNA methylation data have been deposited at GEO. The scATAC-seq data has been
451 deposited at GEO. RNA-seq data are available at EGA.

452

453 Code availability

All data analysis was performed using established methods implemented in published software or R packages. Software and package versions and parameters are detailed in the Methods section. All scripts used for the analysis are available upon request.

REFERENCES

1. Carlsson SK, Brothers SP, Wahlestedt C (2014) Emerging treatment strategies for glioblastoma multiforme. EMBO Mol Med 6: 1359–1370.
2. Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJB, Belanger K, Brandes AA, Marosi C, Bogdahn U, Curschmann J, Janzer RC, Ludwin SK, Gorlia T, Allgeier A, Lacombe D, Cairncross JG, Eisenhauer E, Mirimanoff RO, et al. (2005) Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. N Engl J Med 352: 987–996.
3. von Neubeck C, Seidlitz A, Kitzler HH, Beuthien-Baumann B, Krause M (2015) Glioblastoma multiforme: emerging treatments and stratification markers beyond new drugs. Br J Radiol 88: 20150354.
4. Venere M, Fine HA, Dirks PB, Rich JN (2011) Cancer stem cells in gliomas: Identifying and understanding the apex cell in cancer's hierarchy. Glia 59: 1148–1154.
5. Chen J, Li Y, Yu T-S, McKay RM, Burns DK, Kernie SG, Parada LF (2012) A restricted cell population propagates glioblastoma growth after chemotherapy. Nature 488: 522–526.
6. Gallo M, Coutinho FJ, Vanner RJ, Gayden T, Mack SC, Murison A, Remke M, Li R, Takayama N, Desai K, Lee L, Lan X, Park NI, Barsyte-Lovejoy D, Smil D, Sturm D, Kushida MM, Head R, Cusimano MD, et al. (2015) MLL5 Orchestrates a Cancer Self-Renewal State by Repressing the Histone Variant H3.3 and Globally Reorganizing Chromatin. Cancer Cell 28: 715–729.
7. Bao S, Wu Q, McLendon RE, Hao Y, Shi Q, Hjelmeland AB, Dewhirst MW, Bigner DD, Rich JN (2006) Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. Nature 444: 756–760.
8. Neftel C, Laffy J, Filbin MG, Hara T, Shore ME, Rahme GJ, Richman AR, Silverbush D, Shaw ML, Hebert CM, Dewitt J, Gritsch S, Perez EM, Nicolas Gonzalez Castro L, Lan X, Druck N, Rodman C, Dionne D, Kaplan A, et al. (2019) An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. Cell 0. Available: http://www.cell.com/article/S0092867419306877/abstract. Accessed 19 July 2019.
9. Wang L, Babikir H, Muller S, Yagnik G, Shamardani K, Catalan F, Kohanbash G, Alvarado B, Di Lullo E, Kriegstein A, Shah S, Wadhwa H, Chang SM, Philips JJ, Aghi MK, Diaz AA (2019) The phenotypes of proliferating glioblastoma cells reside on a single axis of variation. Cancer Discov. Available: http://dx.doi.org/10.1158/2159-8290.CD-19-0329.
10. Stergachis AB, Neph S, Reynolds A, Humbert R, Miller B, Paige SL, Vernot B, Cheng JB, Thurman RE, Sandstrom R, Haugen E, Heimfeld S, Murry CE, Akey JM, Stamatoyannopoulos JA (2013) Developmental fate and cellular maturity encoded in

494 human regulatory DNA landscapes. Cell 154: 888–903.
495 11. Buenrostro JD, Corces MR, Lareau CA, Wu B, Schep AN, Aryee MJ, Majeti R, Chang HY,
496 Greenleaf WJ (2018) Integrated Single-Cell Analysis Maps the Continuous Regulatory
497 Landscape of Human Hematopoietic Differentiation. Cell 173: 1535–1548.e16.
498 12. Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, Snyder MP,
499 Pritchard JK, Kundaje A, Greenleaf WJ, Majeti R, Chang HY (2016) Lineage-specific and
500 single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution.
501 Nat Genet 48: 1193–1203.
502 13. Suvà ML, Rheinbay E, Gillespie SM, Patel AP, Wakimoto H, Rabkin SD, Riggi N, Chi AS,
503 Cahill DP, Nahed BV, Curry WT, Martuza RL, Rivera MN, Rossetti N, Kasif S, Beik S, Kadri
504 S, Tirosh I, Wortman I, et al. (2014) Reconstructing and reprogramming the tumor-
505 propagating potential of glioblastoma stem-like cells. Cell 157: 580–594.
506 14. Lan X, Jörg DJ, Cavalli FMG, Richards LM, Nguyen LV, Vanner RJ, Guilhamon P, Lee L,
507 Kushida MM, Pellacani D, Park NI, Coutinho FJ, Whetstone H, Selvadurai HJ, Che C, Luu B,
508 Carles A, Moksa M, Rastegar N, et al. (2017) Fate mapping of human glioblastoma reveals
509 an invariant stem cell hierarchy. Nature 549: 227–232.
510 15. Meyer M, Reimand J, Lan X, Head R, Zhu X, Kushida M, Bayani J, Pressey JC, Lionel AC,
511 Clarke ID, Cusimano M, Squire JA, Scherer SW, Bernstein M, Woodin MA, Bader GD, Dirks
512 PB (2015) Single cell-derived clonal analysis of human glioblastoma links functional and
513 genomic heterogeneity. Proc Natl Acad Sci U S A. Available:
514 http://dx.doi.org/10.1073/pnas.1320611111.
515 16. Orzan F, De Bacco F, Crisafulli G, Pellegatta S, Mussolin B, Siravegna G, D'Ambrosio A,
516 Comoglio PM, Finocchiaro G, Boccaccio C (2017) Genetic Evolution of Glioblastoma Stem-
517 Like Cells from Primary to Recurrent Tumor. Stem Cells. Available:
518 http://dx.doi.org/10.1002/stem.2703.
519 17. Rheinbay E, Suvà ML, Gillespie SM, Wakimoto H, Patel AP, Shahid M, Oksuz O, Rabkin
520 SD, Martuza RL, Rivera MN, Louis DN, Kasif S, Chi AS, Bernstein BE (2013) An aberrant
521 transcription factor network essential for Wnt signaling and stem cell maintenance in
522 glioblastoma. Cell Rep 3: 1567–1579.
523 18. Miller TE, Liau BB, Wallace LC, Morton AR, Xie Q, Dixit D, Factor DC, Kim LJY, Morrow JJ,
524 Wu Q, Mack SC, Hubert CG, Gillespie SM, Flavahan WA, Hoffmann T, Thummalapalli R,
525 Hemann MT, Paddison PJ, Horbinski CM, et al. (2017) Transcription elongation factors
526 represent in vivo cancer dependencies in glioblastoma. Nature 547: 355–359.
527 19. Park NI, Guilhamon P, Desai K, McAdam RF, Langille E, O'Connor M, Lan X, Whetstone H,
528 Coutinho FJ, Vanner RJ, Ling E, Prinos P, Lee L, Selvadurai H, Atwal G, Kushida M, Clarke
529 ID, Voisin V, Cusimano MD, et al. (2017) ASCL1 Reorganizes Chromatin to Direct Neuronal
530 Fate and Suppress Tumorigenicity of Glioblastoma Stem Cells. Cell Stem Cell. Available:
531 http://dx.doi.org/10.1016/j.stem.2017.06.004.
532 20. Liau BB, Sievers C, Donohue LK, Gillespie SM, Flavahan WA, Miller TE, Venteicher AS,
533 Hebert CH, Carey CD, Rodig SJ, Shareef SJ, Najm FJ, van Galen P, Wakimoto H, Cahill DP,
534 Rich JN, Aster JC, Suvà ML, Patel AP, et al. (2017) Adaptive Chromatin Remodeling Drives
535 Glioblastoma Stem Cell Plasticity and Drug Tolerance. Cell Stem Cell 20: 233–246.e7.
536 21. Pollard SM, Yoshikawa K, Clarke ID, Danovi D, Stricker S, Russell R, Bayani J, Head R, Lee
537 M, Bernstein M, Squire JA, Smith A, Dirks P (2009) Glioma stem cell lines expanded in

538 adherent culture have tumor-specific phenotypes and are suitable for chemical and
539 genetic screens. Cell Stem Cell 4: 568–580.
540 22. Wang Q, Hu B, Hu X, Kim H, Squatrito M, Scarpace L, deCarvalho AC, Lyu S, Li P, Li Y,
541 Barthel F, Cho HJ, Lin Y-H, Satani N, Martinez-Ledesma E, Zheng S, Chang E, Sauvé C-
542 EG, Olar A, et al. (2017) Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes
543 Associates with Immunological Changes in the Microenvironment. Cancer Cell 32: 42–
544 56.e6.
545 23. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A,
546 Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a
547 knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl
548 Acad Sci U S A 102: 15545–15550.
549 24. Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, Vilo J (2016) g:Profiler-a
550 web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res 44:
551 W83–W89.
552 25. Capper D, Jones DTW, Sill M, Hovestadt V, Schrimpf D, Sturm D, Koelsche C, Sahm F,
553 Chavez L, Reuss DE, Kratz A, Wefers AK, Huang K, Pajtler KW, Schweizer L, Stichel D,
554 Olar A, Engel NW, Lindenberg K, et al. (2018) DNA methylation-based classification of
555 central nervous system tumours. Nature 555: 469–474.
556 26. MacLeod G, Bozek DA, Rajakulendran N, Monteiro V, Ahmadi M, Steinhart Z, Kushida
557 MM, Yu H, Coutinho FJ, Cavalli FMG, Restall I, Hao X, Hart T, Luchman HA, Weiss S, Dirks
558 PB, Angers S (2019) Genome-Wide CRISPR-Cas9 Screens Expose Genetic Vulnerabilities
559 and Mechanisms of Temozolomide Sensitivity in Glioblastoma Stem Cells. Cell Rep 27:
560 971–986.e9.
561 27. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z (2017) GEPIA: a web server for cancer and
562 normal gene expression profiling and interactive analyses. Nucleic Acids Res. Available:
563 http://dx.doi.org/10.1093/nar/gkx247.
564 28. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, Cahill DP, Nahed
565 BV, Curry WT, Martuza RL, Louis DN, Rozenblatt-Rosen O, Suvà ML, Regev A, Bernstein
566 BE (2014) Single-cell RNA-seq highlights intratumoral heterogeneity in primary
567 glioblastoma. Science 344: 1396–1401.
568 29. Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L,
569 Golub T, Mesirov JP, Alexe G, Lawrence M, O'Kelly M, Tamayo P, Weir BA, Gabriel S,
570 Winckler W, Gupta S, Jakkula L, et al. (2010) Integrated genomic analysis identifies
571 clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA,
572 IDH1, EGFR, and NF1. Cancer Cell 17: 98–110.
573 30. Yin W, Tang G, Zhou Q, Cao Y, Li H, Fu X, Wu Z, Jiang X (2019) Expression Profile Analysis
574 Identifies a Novel Five-Gene Signature to Improve Prognosis Prediction of Glioblastoma.
575 Front Genet 10: 419.
576 31. Zuo S, Zhang X, Wang L (2019) A RNA sequencing-based six-gene signature for survival
577 prediction in patients with glioblastoma. Sci Rep 9: 2615.
578 32. Johnston M, Nikolic A, Ninkovic N, Guilhamon P, Cavalli F, Seaman S, Zemp F, Lee J,
579 Abdelkareem A, Ellestad K, Murison A, Kushida M, Coutinho F, Ma Y, Mungall A, Moore R,
580 Marra M, Taylor M, Dirks P, et al. (2019) High-resolution structural genomics reveals new
581 therapeutic vulnerabilities in glioblastoma. Genome Res. Available:

582 http://dx.doi.org/10.1101/gr.246520.118.
583 33. Bailey SD, Desai K, Kron KJ, Mazrooei P, Sinnott-Armstrong NA, Treloar AE, Dowar M,
584 Thu KL, Cescon DW, Silvester J, Yang SYC, Wu X, Pezo RC, Haibe-Kains B, Mak TW,
585 Bedard PL, Pugh TJ, Sallari RC, Lupien M (2016) Noncoding somatic and inherited single-
586 nucleotide variants converge to promote ESR1 expression in breast cancer. Nat Genet 48:
587 1260–1266.
588 34. Sallari RC, Sinnott-Armstrong NA, French JD, Kron KJ, Ho J, Moore JH, Stambolic V,
589 Edwards SL, Lupien M, Kellis M (2017) Convergence of dispersed regulatory mutations
590 predicts driver genes in prostate cancer. bioRxiv: 097451. Available:
591 https://www.biorxiv.org/content/10.1101/097451v2. Accessed 20 November 2019.
592 35. Kim K, Jang K, Yang W, Choi E-Y, Park S-M, Bae M, Kim Y-J, Choi JK (2016) Chromatin
593 structure-based prediction of recurrent noncoding mutations in cancer. Nat Genet 48:
594 1321–1326.
595 36. Schmitt AD, Hu M, Ren B (2016) Genome-wide mapping and analysis of chromosome
596 architecture. Nat Rev Mol Cell Biol 17: 743–755.
597 37. Zheng H, Xie W (2019) The role of 3D genome organization in development and cell
598 differentiation. Nat Rev Mol Cell Biol 20: 535–550.
599 38. O'Connor L, Gilmour J, Bonifer C (2016) The Role of the Ubiquitously Expressed
600 Transcription Factor Sp1 in Tissue-specific Transcriptional Regulation and in Disease. Yale
601 J Biol Med 89: 513–525.
602 39. Liu T, Yuan X, Xu D (2016) Cancer-Specific Telomerase Reverse Transcriptase (TERT)
603 Promoter Mutations: Biological and Clinical Implications. Genes 7. Available:
604 http://dx.doi.org/10.3390/genes7070038.
605 40. Lee WS, Kwon J, Yun DH, Lee YN, Woo EY, Park M-J, Lee J-S, Han Y-H, Bae IH (2014)
606 Specificity protein 1 expression contributes to Bcl-w-induced aggressiveness in
607 glioblastoma multiforme. Mol Cells 37: 17–23.
608 41. Trépant A-L, Bouchart C, Rorive S, Sauvage S, Decaestecker C, Demetter P, Salmon I
609 (2015) Identification of OLIG2 as the most specific glioblastoma stem cell marker starting
610 from comparative analysis of data from similar DNA chip microarray platforms. Tumour
611 Biol 36: 1943–1953.
612 42. Gabriely G, Wheeler MA, Takenaka MC, Quintana FJ (2017) Role of AHR and HIF-1α in
613 Glioblastoma Metabolism. Trends Endocrinol Metab 28: 428–436.
614 43. Michaelson JJ, Shin M-K, Koh J-Y, Brueggeman L, Zhang A, Katzman A, McDaniel L, Fang
615 M, Pufall M, Pieper AA (2017) Neuronal PAS Domain Proteins 1 and 3 Are Master
616 Regulators of Neuropsychiatric Risk Genes. Biol Psychiatry 82: 213–223.
617 44. Rajakulendran N, Rowland KJ, Selvadurai HJ, Ahmadi M, Park NI, Naumenko S, Dolma S,
618 Ward RJ, So M, Lee L, MacLeod G, Pasiliao C, Brandon C, Clarke ID, Cusimano MD,
619 Bernstein M, Batada N, Angers S, Dirks PB (2019) Wnt and Notch signaling govern self-
620 renewal and differentiation in a subset of human glioblastoma stem cells. Genes Dev 33:
621 498–510.
622 45. Cheng P, Wang J, Waghmare I, Sartini S, Coviello V, Zhang Z, Kim S-H, Mohyeldin A,
623 Pavlyukov MS, Minata M, Valentim CLL, Chhipa RR, Bhat KPL, Dasgupta B, La Motta C,
624 Kango-Singh M, Nakano I (2016) FOXD1-ALDH1A3 Signaling Is a Determinant for the
625 Self-Renewal and Tumorigenicity of Mesenchymal Glioma Stem Cells. Cancer Res 76:

626 7219–7230.
627 46. Koga M, Matsuda M, Kawamura T, Sogo T, Shigeno A, Nishida E, Ebisuya M (2014) Foxd1
628 is a mediator and indicator of the cell reprogramming process. Nat Commun 5: 3197.
629 47. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ (2013) Transposition of native
630 chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding
631 proteins and nucleosome position. Nat Methods 10: 1213–1218.
632 48. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S,
633 Satpathy AT, Rubin AJ, Montine KS, Wu B, Kathiria A, Cho SW, Mumbach MR, Carter AC,
634 Kasowski M, Orloff LA, Risca VI, Kundaje A, Khavari PA, et al. (2017) An improved ATAC-
635 seq protocol reduces background and enables interrogation of frozen tissues. Nat
636 Methods 14: 959–962.
637 49. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, Haibe-Kains B, Goldenberg A
638 (2014) Similarity network fusion for aggregating data types on a genomic scale. Nat
639 Methods 11: 333–337.
640 50. Shin H, Liu T, Manrai AK, Liu XS (2009) CEAS: cis-regulatory element annotation system.
641 Bioinformatics 25: 2605–2606.
642 51. Schep AN, Wu B, Buenrostro JD, Greenleaf WJ (2017) chromVAR: inferring transcription-
643 factor-associated accessibility from single-cell epigenomic data. Nat Methods 14: 975–
644 978.
645 52. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y,
646 Stoeckius M, Smibert P, Satija R (2019) Comprehensive Integration of Single-Cell Data.
647 Cell 177: 1888–1902.e21.
648 53. Müller S, Cho A, Liu SJ, Lim DA, Diaz A (2018) CONICS integrates scRNA-seq with DNA
649 sequencing to map gene expression to tumor sub-clones. Bioinformatics 34: 3217–3219.
650 54. Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, Beck S
651 (2014) ChAMP: 450k Chip Analysis Methylation Pipeline. Bioinformatics 30: 428–430.
652 55. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,
653 Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29: 15–21.
654 56. Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with
655 or without a reference genome. BMC Bioinformatics 12: 323.
656 57. Anders S, Huber W (2010) Differential expression analysis for sequence count data.
657 Genome Biol 11: R106.
658 58. Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi
659 HS, Lambert SA, Mann I, Cook K, Zheng H, Goity A, van Bakel H, Lozano J-C, Galli M,
660 Lewsey MG, Huang E, Mukherjee T, Chen X, et al. (2014) Determination and inference of
661 eukaryotic transcription factor sequence specificity. Cell 158: 1431–1443.
662 59. Li W, Xu H, Xiao T, Cong L, Love MI, Zhang F, Irizarry RA, Liu JS, Brown M, Liu XS (2014)
663 MAGeCK enables robust identification of essential genes from genome-scale
664 CRISPR/Cas9 knockout screens. Genome Biol 15: 554.
665 60. Hart T, Tong AHY, Chan K, Van Leeuwen J, Seetharaman A, Aregger M, Chandrashekhar
666 M, Hustedt N, Seth S, Noonan A, Habsid A, Sizova O, Nedyalkova L, Climie R,
667 Tworzyanski L, Lawson K, Sartori MA, Alibeh S, Tieu D, et al. (2017) Evaluation and Design
668 of Genome-Wide CRISPR/SpCas9 Knockout Screens. G3 7: 2719–2727.
669 61. Hart T, Chandrashekhar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, Mis M,

670 Zimmermann M, Fradet-Turcotte A, Sun S, Mero P, Dirks P, Sidhu S, Roth FP, Rissland
671 OS, Durocher D, Angers S, Moffat J (2015) High-Resolution CRISPR Screens Reveal
672 Fitness Genes and Genotype-Specific Cancer Liabilities. Cell 163: 1515–1526.
673 62. Tam RY, Yockell-Lelièvre J, Smith LJ, Julian LM, Baker AEG, Choey C, Hasim MS,
674 Dimitroulakos J, Stanford WL, Shoichet MS (2019) Rationally Designed 3D Hydrogels
675 Model Invasive Lung Diseases Enabling High-Content Drug Screening. Adv Mater 31:
676 e1806214.
677 63. Therneau T (2015) A Package for Survival Analysis in S. version 2.38. Available:
678 https://CRAN.R-project.org/package=survival.
679 64. Drawing Survival Curves using "ggplot2" [R package survminer version 0.4.6] (n.d.).
680 Available: https://cran.r-project.org/web/packages/survminer/index.html. Accessed 16
681 October 2019.
682

683 **ACKNOWLEDGMENTS**

684 We would like to thank the Princess Margaret Genomics Centre for their assistance in this study
685 and 10x Genomics, especially Adam Jerauld, Ariel Royall and John Chevillet for training and
686 support. We also thank Aude Gerbaud and Stacey Krunholtz for their help with figure formatting
687 and David Vetrie for his thoughtful input on the manuscript. This work was supported by the
688 Princess Margaret Cancer Foundation, the Canadian Epigenetics, Environment, and Health
689 Research Consortium (CEEHRC) initiative from the Canadian Institute for Health Research
690 (CIHR) (TGH-158221 to ML, SA and PBD) and SU2C Canada Cancer Stem Cell Dream Team
691 Research Funding (SU2C-AACR-DT-19-15) provided by the Government of Canada through
692 Genome Canada and the Canadian Institute of Health Research, with supplemental support
693 from the Ontario Institute for Cancer Research, through funding provided by the Government
694 of Ontario. Stand Up To Cancer Canada is a Canadian Registered Charity (Reg. # 80550 6730
695 RR0001). Research Funding is administered by the American Association for Cancer Research
696 International - Canada, the Scientific Partner of SU2C Canada. This study was also conducted
697 with the support of the Ontario Institute for Cancer Research through funding provided by the

698 Government of Ontario for the Brain Cancer Translational Research Initiative. PG was supported
699 by a CIHR Fellowship (MFE 338954). Funding to HAL and SW was from CIHR. MG holds a Canada
700 Research Chair in Brain Cancer Epigenomics and is supported by CIHR, NSERC and the Alliance
701 for Cancer Gene Therapy. PBD is also supported by CIHR, OICR, the Terry Fox Research Institute,
702 and the Hospital for Sick Children Foundation, Jessica's Footprint, Hopeful Minds, and the
703 Bresler Family. PBD holds a Garron Chair in Childhood Cancer Research at the Hospital for Sick
704 Children. SA is supported by the CIHR, the Terry Fox Institute and the Canadian Cancer Society.
705 ML holds an Investigator Award from the Ontario Institute for Cancer Research and a Canadian
706 Institutes of Health Research (CIHR) New Investigator Award.

707

708 **AUTHOR CONTRIBUTIONS**

709 ML, PG, and PBD conceptualized and designed the study assisted by FJC. PG conducted the
710 genomics experiments, designed and/or implemented most of the computational and statistical
711 approaches, and made the figures. MMK and RH performed all the tissue culture, under the
712 supervision of PBD, HAL and SW. AN, KE, and DS generated the scATAC-seq data from samples
713 provided by JAC, under the supervision of MG. BN performed the alignment of the scATAC-seq
714 data. NR, XH and RH performed the xenografts, under the supervision of PBD, HAL, and SW. CA
715 generated some of the ATAC-seq data used in the xenograft classification. FMGC ran the
716 spectral clustering, under the supervision of MDT. GM, NR and SA contributed the essentiality
717 screen data. HW conducted the endpoint xenograft staining. LJS performed the invasion assay
718 under the supervision of MSS. SAMT contributed to the computational analysis design, under

719 supervision of BHK and ML. The manuscript was written by PG, PBD and ML with input from all
720 other authors.

721

722 **COMPETING INTERESTS**

723 The authors declare no competing interests.

724

725 **FIGURES**

726 **Fig. 1. The diverse GBM cancer stem cell pool. (A)** UMAP representation of chromatin
727 accessibility across four primary GBM. **(B)** UMAPs with tumor cells assigned to cellular states.
728 **(C)** UMAP modules are grouped by dominant cellular state. **(D)** UMAPs with cancer stem cells
729 highlighted based on the enrichment of GBM cancer stem transcription factor promoters. **(E)**
730 Distribution of cancer stem cells across the modules dominated by each cellular state.

731 **Fig. 2. GSCs recapitulate the GBM cancer stem cell population. (A)** Schematic representation
732 of the GSC derivation process, from patient tumor to GSC-enriched population. **(B)** Genomic
733 feature enrichment of accessible chromatin peaks. **(C)** Saturation curve for the 27 GSCs. **(D)**
734 UMAPs with GSCs highlighted based on the enrichment of shared accessible regions across
735 GSCs. **(E)** Proportion of UMAP modules assigned to cancer stem cells and GSCs. **(F)** Correlation
736 of z-scores for each signature for each cell in each primary GBM.

737 **Fig. 3. Three GSC states driven by chromatin accessibility. (A)** Spectral clustering of ATAC-
738 seq signal across peaks in 27 GSCs. **(B)** Top: enrichment of TCGA subtypes across GSCs and
739 comparison to GSC states; Bottom: spectral clustering of gene expression across GSCs and

740 comparison to chromatin-derived GSC states. **(C)** Gene set enrichment analysis in each GSC
741 state. **(D)** CNAs across GSCs identified from DNA methylation array data cluster GSCs into four
742 subgroups. **(E)** Number and percentage of peaks unique and shared in each GSC state. **(F)**
743 Saturation analysis of each individual state.

744 **Fig. 4. Functional diversity between GSC states drives survival in GBM. (A)** Motif family
745 enrichment in each cluster; log2(Fold Enrichment) > 0.5 threshold selected based on the
746 distribution of values in each cluster (Fig. S1A). **(B)** Z-score distribution of key essential genes in
747 each cluster. Red line corresponds to the empirically determined threshold for essentiality in
748 each tested line, scaled and adjusted to zero. Boxplot whiskers in this case extend to data
749 extremes. Side barplots show the total count of the key cluster-specific regulators found
750 essential in each subtype. **(C)** UMAPs with GSCs from each state highlighted based on the
751 enrichment of the top differentially accessible regions in each GSC state. **(D)** Kaplan-Meier plot
752 for orthotopic xenografts grouped by GSC state. The dotted lines indicate median survival. The
753 pairwise p-values are also significant for Invasive vs Reactive (p=0.02) and Invasive vs
754 Constructive (p=0.045) but not for Reactive vs Constructive (p=0.45). **(E)** TCGA samples ordered
755 by increasing concordance with Invasive GSCs and grouped into three subgroups: <1, 1-
756 1.65,>1.65. **(F)** Kaplan-Meier plot for TCGA samples grouped by concordance with Invasive
757 GSCs. The dotted line indicates median survival. When considering pairwise comparisons, only
758 the Invasive-high and Invasive-low subgroups were significantly different (p=0.0043). Further
759 subgrouping of the TCGA samples into smaller intervals of concordance z-score yielded no
760 benefit, preserving the Invasive-high subgroup as the only one with significantly poorer
761 prognosis (Fig. S3D-E).

762 **SUPPLEMENTARY DATA**

763 **Fig. S1. (A-D)** UMAPs showing cells confirmed to be GBM tumor cells, with gains of chr7 or
764 losses of chr10. **(E-G)** Overlaid ATAC signal tracks for the 27 GSCs at the 92 most differentially
765 accessible regions between GSC states as determined by pairwise wilcoxon test and a median
766 signal fold change equal or greater than 2. (E) Most accessible in Reactive state GSCs. (F) Most
767 accessible in Constructive state GSCs. (G) Most accessible in Invasive state GSCs.

768 **Fig. S2. (A)** The log2(fold enrichment) for all motif families were ordered and plotted. A cut-off
769 threshold to select families for follow-up was selected based on the steep inflection of the curve
770 above 0.5. **(B)** Schematic of drop-out essentiality screen using GSCs stably expressing Cas9 and
771 gRNA libraries. **(C-E)** The expression values determined by RNA-seq for all transcription factors
772 whose motif is enriched in each GSC state and exclusively essential in at least one GSC of each
773 state is plotted. Bold colors highlight factors essential in all tested GSCs in that state. In the case
774 of the Reactive state, blue boxplots correspond to SP1 and members of its regulatory network.
775 **(F)** Expression levels of key transcription factors in tumor and normal samples, analysed and
776 displayed using GEPIA[27]. **(G)** Gene enrichment analysis of subtype-specific essential genes
777 (Essentiality), upregulated and differentially accessible genes (ATAC and RNA combined), and
778 putative targets of key transcription factors.

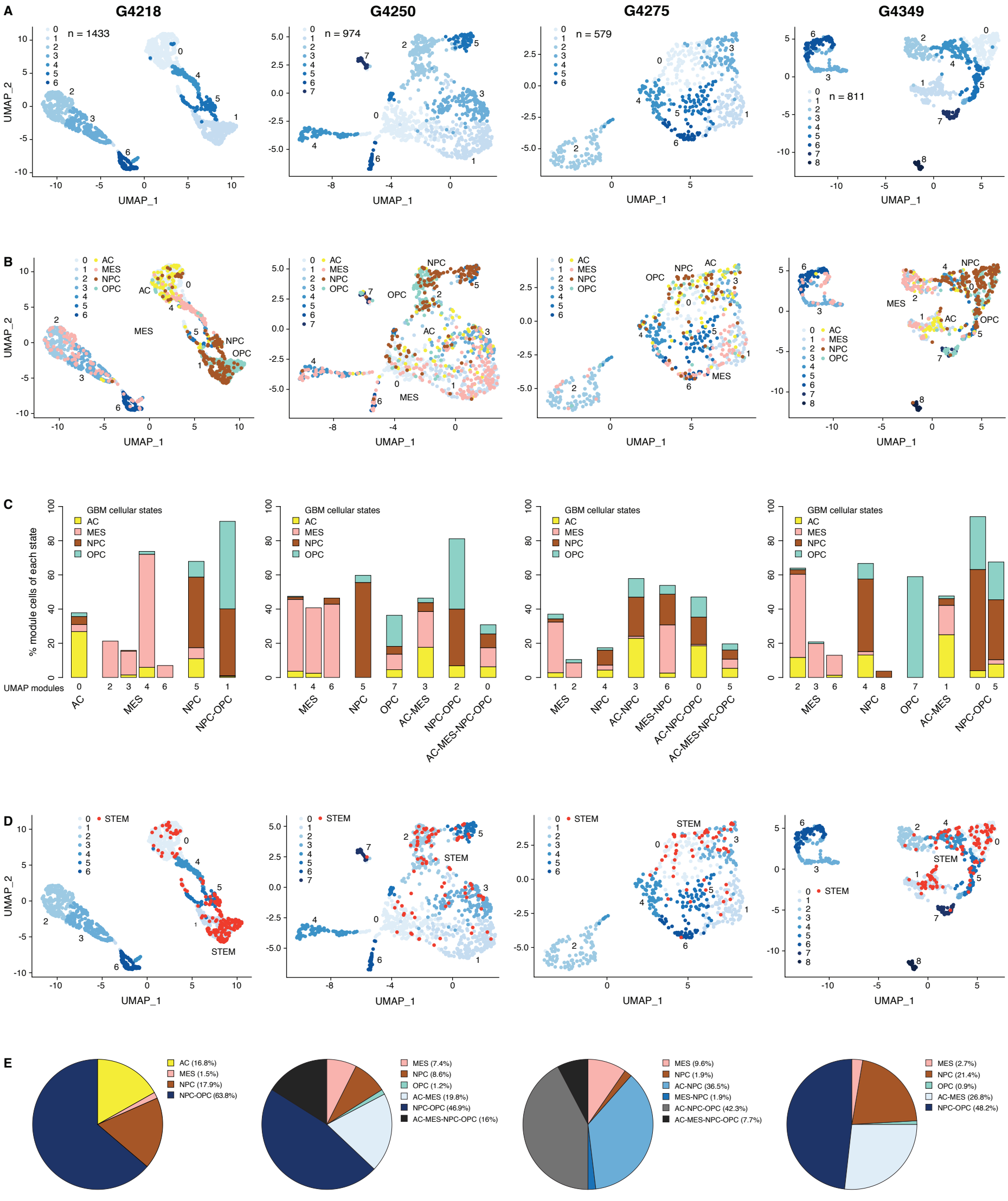
779 **Fig. S3. (A)** Kaplan-Meier plot for TCGA samples grouped by GSC state. The dotted lines indicate
780 median survival. **(B)** TCGA samples ordered by increasing concordance with Invasive GSCs and

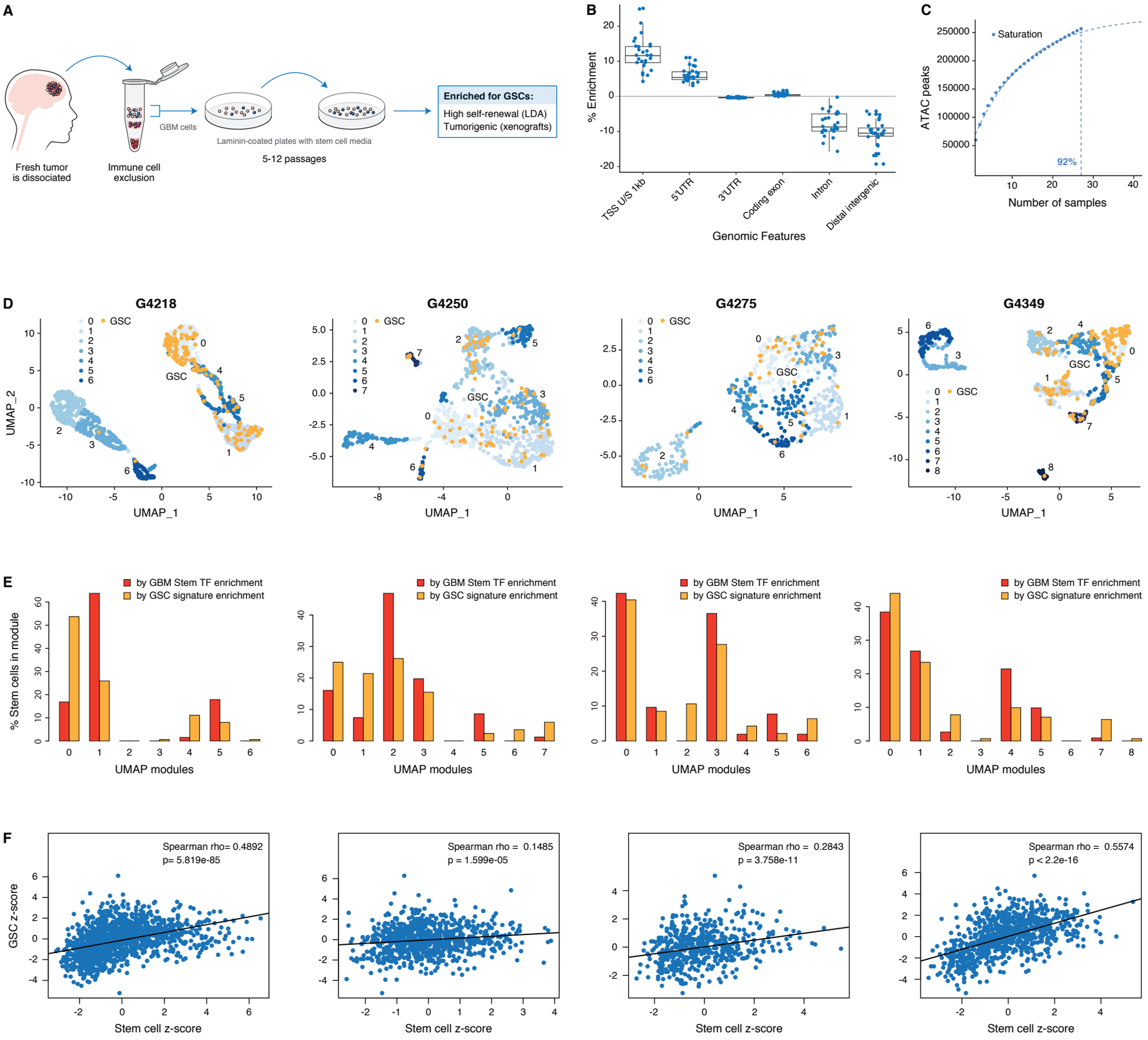
781 grouped into 8 subgroups. **(C)** Kaplan-Meier plot for TCGA samples grouped by concordance
782 with Invasive GSCs. The dotted line indicates median survival.

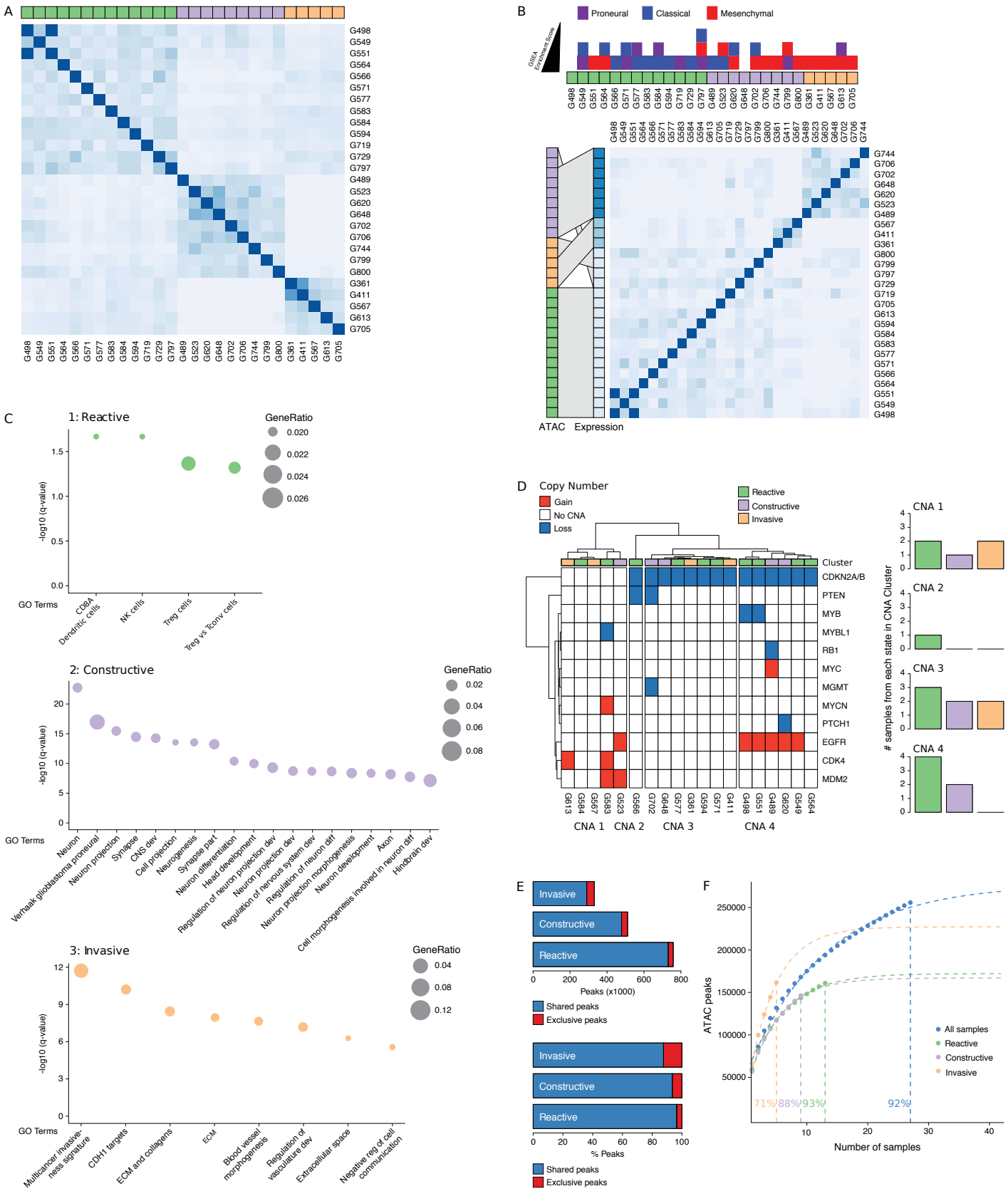
783 **Fig. S4. (A)** Left: Top and side views of confocal z-stack images of Invasive GSCs G705, G837 and
784 healthy control hf6562 cultured on hyaluronan-based hydrogels. Right: percent invasion,
785 defined as the percentage of total cells located below a 75 µm threshold from the surface of the
786 hydrogel. Plots are shown as mean with standard deviation, and statistics shown are from one-
787 way ANOVA with multiple comparisons with Dunnet's test correction, comparing all patient cell
788 lines (dark grey) to the hf6562 healthy control. Data were displayed with p values represented
789 as * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$. **(B)** Human-specific staining of mouse
790 brains injected with GSCs from G411 (upper panel) and G837 (lower panel) with higher
791 magnification (black inset) at the mass boundaries.

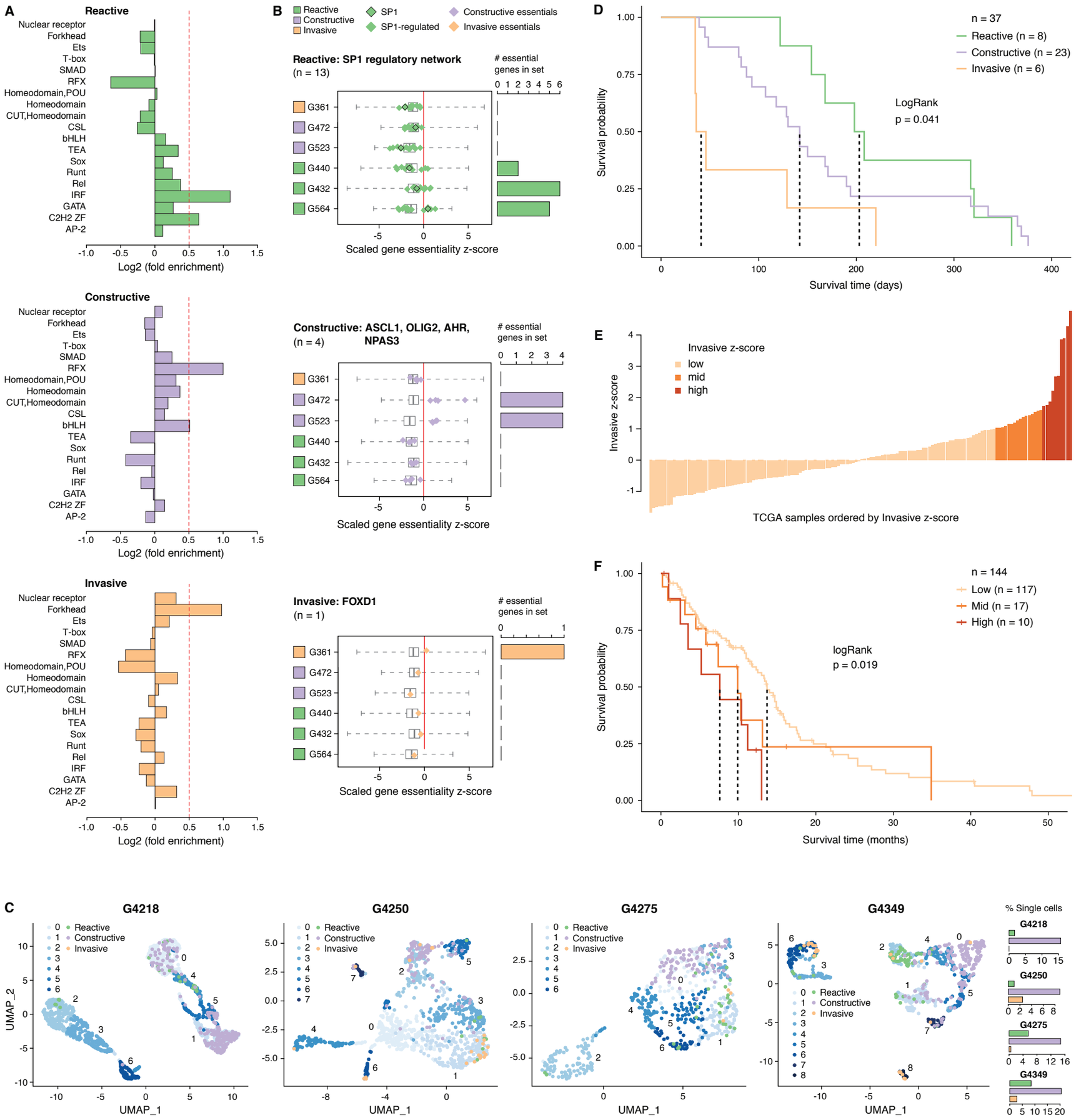
792

793


A
G4218
G4250
G4275
G4349
n = 1433
n = 974
n = 579
n = 811
UMAP_1
UMAP_2
B
AC
MES
NPC
OPC
C
GBM cellular states
% module cells of each state
UMAP modules
AC
MES
NPC
NPC-OPC
OPC
AC-MES
AC-MES-NPC-OPC
AC-NPC
MES-NPC
AC-NPC-OPC
D
STEM
E
AC (16.8%)
MES (1.5%)
NPC (17.9%)
NPC-OPC (63.8%)
MES (7.4%)
NPC (8.6%)
OPC (1.2%)
AC-MES (19.8%)
NPC-OPC (46.9%)
AC-MES-NPC-OPC (16%)
MES (9.6%)
NPC (1.9%)
AC-NPC (36.5%)
MES-NPC (1.9%)
AC-NPC-OPC (42.3%)
AC-MES-NPC-OPC (7.7%)
MES (2.7%)
NPC (21.4%)
OPC (0.9%)
AC-MES (26.8%)
NPC-OPC (48.2%)


A
Fresh tumor is dissociated
Immune cell exclusion
GBM cells
Laminin-coated plates with stem cell media
5-12 passages
Enriched for GSCs:
High self-renewal (LDA)
Tumorigenic (xenografts)
B
% Enrichment
TSS U/S 1kb
5'UTR
3'UTR
Coding exon
Intron
Distal intergenic
Genomic Features
C
Saturation
ATAC peaks
92%
Number of samples
D
G4218
G4250
G4275
G4349
GSC
UMAP_1
UMAP_2
E
by GBM Stem TF enrichment
by GSC signature enrichment
% Stem cells in module
UMAP modules
F
Spearman rho= 0.4892
p= 5.819e-85
Spearman rho = 0.1485
p = 1.599e-05
Spearman rho = 0.2843
p = 3.758e-11
Spearman rho = 0.5574
p < 2.2e-16
GSC z-score
Stem cell z-score


A
B
Proneural
Classical
Mesenchymal
GSEA Enrichment Score
ATAC
Expression
C
1: Reactive
GeneRatio
0.020
0.022
0.024
0.026
-log10 (q-value)
GO Terms
CD8A
Dendritic cells
NK cells
Treg cells
Treg vs Tconv cells
2: Constructive
GeneRatio
0.02
0.04
0.06
0.08
Neuron
Verhaak glioblastoma proneural
Neuron projection
Synapse
CNS dev
Cell projection
Neurogenesis
Synapse part
Neuron differentiation
Head development
Regulation of neuron projection dev
Neuron projection dev
Regulation of nervous system dev
Regulation of neuron diff
Neuron projection morphogenesis
Neuron development
Axon
Cell morphogenesis involved in neuron diff
Hindbrain dev
3: Invasive
GeneRatio
0.04
0.08
0.12
Multicancer invasiveness signature
CDH1 targets
ECM and collagens
ECM
Blood vessel morphogenesis
Regulation of vasculature dev
Extracellular space
Negative reg of cell communication
D
Copy Number
Gain
No CNA
Loss
Reactive
Constructive
Invasive
Cluster
CDKN2A/B
PTEN
MYB
MYBL1
RB1
MYC
MGMT
MYCN
PTCH1
EGFR
CDK4
MDM2
CNA 1
CNA 2
CNA 3
CNA 4
samples from each state in CNA Cluster
E
Invasive
Constructive
Reactive
Peaks (x1000)
Shared peaks
Exclusive peaks
% Peaks
F
ATAC peaks
All samples
Reactive
Constructive
Invasive
71%
88%
93%
92%
Number of samples


A
Reactive
Nuclear receptor
Forkhead
Ets
T-box
SMAD
RFX
Homeodomain,POU
Homeodomain
CUT,Homeodomain
CSL
bHLH
TEA
Sox
Runt
Rel
IRF
GATA
C2H2 ZF
AP-2
Log2 (fold enrichment)
Constructive
Invasive
B
Reactive
Constructive
Invasive
SP1
SP1-regulated
Constructive essentials
Invasive essentials
Reactive: SP1 regulatory network
(n = 13)
essential genes in set
G361
G472
G523
G440
G432
G564
Scaled gene essentiality z-score
Constructive: ASCL1, OLIG2, AHR, NPAS3
(n = 4)
Invasive: FOXD1
(n = 1)
C
G4218
G4250
G4275
G4349
UMAP_1
UMAP_2
% Single cells
D
n = 37
Reactive (n = 8)
Constructive (n = 23)
Invasive (n = 6)
LogRank
p = 0.041
Survival probability
Survival time (days)
E
Invasive z-score
low
mid
high
TCGA samples ordered by Invasive z-score
F
n = 144
Low (n = 117)
Mid (n = 17)
High (n = 10)
logRank
p = 0.019
Survival time (months)