Tumor evolution of glioma intrinsic gene expression subtype associates with immunological changes in the microenvironment

Qianghu Wang[1, 2, 3*], Xin Hu[1, 8*], Florian Muller[1, 4], Hoon Kim[1], Massimo Squatrito[9], Tom Mikkelsen[10], Lisa Scarpace[10], Floris Barthel[11], Yu-Hsi Lin[4], Nikunj Satani[4], Emmanuel Martinez-Ledesma[1], Edward Chang[1], Adriana Olar[5], Baoli Hu[6], Ana C. deCarvalho[10], Eskil Eskilsson[1], Siyuan Zheng[1], Amy B. Heimberger[7], Erik P. Sulman[2†], Do-Hyun Nam[12, 13, 14†], Roel G.W. Verhaak[1, 3†]

[1]Department of Genomic Medicine, [2]Department of Radiation Oncology, [3]Department of Bioinformatics and Computational Biology, [4]Department of Cancer Systems Imaging, [5]Department of Pathology, [6]Department of Cancer Biology, [7]Department of Neurosurgery, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA; [8]University of Texas-Houston Graduate School in Biomedical Sciences, Houston, TX 77030, USA; [9]Cancer Cell Biology Programme, Seve Ballesteros Foundation Brain Tumor Group, Centro Nacional de Investigaciones Oncológicas, CNIO, 28029 Madrid, Spain; [10]Departments of Neurology and Neurosurgery, Henry Ford Hospital, Detroit, MI 48202, USA; [11]Oncology Graduate School Amsterdam, VU University Medical Center, 1081 HV Amsterdam, The Netherlands; [12]Institute for Refractory Cancer Research, Samsung Medical Center, Seoul 06351, Korea; [13]Department of Health Sciences and Technology, Samsung Advanced Institute for Health Sciences and Technology, Sungkyunkwan University, Seoul 06351, Korea; [14]Department of Neurosurgery Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, 135-710, Korea.

* These authors contributed equally
[†]Co-senior authors

Correspondence: rverhaak@mdanderson.org

Keywords: glioblastoma, disease recurrence, mesenchymal subtype, proneural to mesenchymal transition, gene expression profiling, tumor microenvironment, macrophages/microglia, immune cells.

ABSTRACT

We leveraged IDH wild type glioblastomas (GBM) and derivative neurosphere models to define the tumor-intrinsic mRNA transcription phenotype. We found that intratumoral heterogeneity is reflected in the transcriptome and associated with increased tumor microenvironment presence. We performed in silico cell sorting to demonstrate that M2 macrophages/microglia are the most frequent type of immune cells in the glioma microenvironment, followed by CD4+ T lymphocytes. Longitudinal transcriptome analysis of 124 pairs of primary and recurrent glioma pairs showed expression subtype is retained in 53% of cases with no proneural to mesenchymal transition being apparent. Inference of the tumor microenvironment through gene signatures revealed a decrease in invading monocytes but a subtype dependent increase in M2 macrophages/microglia cells after disease recurrence. Our longitudinal expression dataset can be accessed at http://ackbar.cnio.es:3838/RecuR/. Our study provides a comprehensive transcriptional and cellular landscape of IDH wild type GBM during treatment modulated tumor evolution.

INTRODUCTION

The intrinsic capacity of glioblastoma (GBM) tumor cells to infiltrate normal brain impedes surgical eradication and predictably results in high rates of early recurrence. To better understand determinants of GBM tumor evolution and treatment resistance, The Cancer Genome Atlas Consortium (TCGA) performed high dimensional profiling and molecular classification of nearly 600 GBM tumors [1-5]. In addition to revealing common mutations in genes such as *TP53*, *EGFR*, *IDH1*, and *PTEN*, as well as the frequent and concurrent presence of abnormalities in the p53, RB and receptor tyrosine kinase pathways. Unsupervised transcriptome analysis identified four clusters, referred to as classical, mesenchymal, neural and proneural, that were tightly associated with genomic abnormalities[6]. The proneural and the mesenchymal expression subtypes have been most consistently described in literature with proneural relating to a more favorable outcome and mesenchymal to unfavorable survival [7-9], but these findings were affected by the relatively favorable outcome of IDH-mutant glioblastoma which are consistently classified as proneural [4,6]. Proneural to mesenchymal switching upon disease recurrence has been described as a source for treatment resistance in GBM relapse [7,10-12], but the relevance of this phenomenon in glioma progress remains ambiguous.

GBM tumor cells along with the tumor microenvironment create a complex milieu that ultimately promotes tumor cell plasticity and disease progression [13]. The presence of tumor-associated stroma results in a mesenchymal tumor gene signature and poor prognosis in colon cancers [14]. Furthermore, the association between a mesenchymal gene expression signature and reduced tumor purity has been identified as a common

69 theme across cancer [15,16]. Tumor-associated macrophages/microglia in GBM have
70 been proposed as regulators of proneural-to-mesenchymal transition through NF-kB
71 activation [10] and may provide growth factor mediated proliferative signals, which could
72 be therapeutically targeted [17-19].

73 Here, we explored the properties of the microenvironment in different GBM gene
74 expression subtypes and characterized the transition between molecular subtypes
75 before and after therapeutic intervention. In doing so, we improved the robustness of
76 gene expression subtype classification through revised gene signatures and proposed
77 analytical methodology. Our results suggested that the tumor microenvironment
78 interferes with expression based classification of GBM, both at the primary disease
79 stage as well as at disease recurrence, and suggest a role for the
80 macrophage/microglia in treatment response.

RESULTS

Harnessing glioma sphere-forming cells identifies GBM specific intertumoral transcriptional heterogeneity

We set out to elucidate the tumor-intrinsic and tumor microenvironment independent transcriptional heterogeneity of GBMs. We performed a pairwise gene expression comparison of independent set of GBMs and the derivative glioma sphere-forming cells (GSCs) (n = 37) [20]. In total, 5,334 genes were found to be significantly higher expressed in parental GBMs relative to derived GSCs that could be attributed by the tumor associated GBM microenvironment (**Figure 1A**). To focus the analysis on the tumor-intrinsic transcriptome, these genes were filtered from further analysis. GBMs with IDH mutations have distinct biological properties and favorable clinical outcomes compared to IDH wild-type GBMs [1,4,5,21]. Using the filtered gene set, we performed consensus non-negative matrix factorization clustering to identify three distinct subgroups amongst 369 IDH wild type GBMs (**Figure 1B; Figure 1C**). When comparing the clustering result with the previously defined proneural (PN), neural (NE), classical (CL) and mesenchymal (MES) classification [1,3], three subgroups were strongly enriched for CL, MES and PN GBMs, respectively (**Figure S1**). Consequently, we labeled the groups as CL, MES and PN. None of the three subgroups was enriched for the NE class, suggesting its neural phenotype is non-tumor specific. The NE group has previously been related to the tumor margin where normal neural tissue is more likely to be present [22] and such contamination might explain why the neural subtype was the only subtype to lack characteristic gene abnormalities [1,23]. To be able to classify external GBM samples, we implemented a single sample gene set enrichment analysis

104 (ssGSEA) based equivalent distribution resampling classification strategy using 70-gene
105 signatures for each subgroup (**Table S1**)(**Figure 1D**), to assign each sample three
106 empirical classification p-values by which we determined the significantly activated
107 subtype(s) in the samples. We prepared an R-library to facilitate others to evaluate our
108 approach (**Supplementary File**). Using this method we found that the stability of cluster
109 assignments of 144 TCGA GBM samples profiled using both RNA sequencing and
110 Affymetrix U133A microarrays was 95% concordance (**Figure S2, Table S2**). This was
111 an improvement over the 77% subtype concordance determined using previously
112 reported methods [3]. We evaluated the distribution of somatic variants across all the
113 three molecular subtypes (**Figure 1E)(Figure S3)** and confirmed the strong
114 associations between subtypes and genomic abnormalities in previously reported driver
115 genes [1,3].
116
117 ***Multi-activation of subtype signatures associated with intratumoral heterogeneity***
118 We observed that 34/369 (9.2%) samples showed significant enrichment of multiple
119 ssGSEA scores (empirical classification p-value<0.05), suggesting these cases activate
120 more than one transcriptional subtype (**Figure 2A**). To quantify this phenomenon, a
121 score ranging from 0 to 1 was defined to quantitatively evaluate the simplicity of subtype
122 activation based on order statistics of ssGSEA score. Samples with high simplicity
123 scores activated a single subtype and those with lowest simplicity scores activated
124 multiple subtypes. All multi-subtype TCGA samples showed simplicity scores of less
125 than 0.1 (**Figure 2A**). To determine whether transcriptional heterogeneity associated
126 with genomic intratumoral heterogeneity, we correlated simplicity scores, total mutation

127 rates and subclonal mutation rates. Included in the analysis were 224 TCGA GBMs with
128 available whole exome sequencing data [24] and ABSOLUTE [25] determined high tumor
129 purity (> 0.8) to equalize the mutation detection sensitivity [26]. Although not significant
130 (Wilcoxon rank test p-value=0.143), the total mutation rate was less in the bottom 30%
131 with lowest simplicity scores versus the top 30% samples with highest simplicity scores.
132 The subclonal mutation rate was significantly higher (p-value=0.024) in samples with
133 lowest simplicity scores (**Figure 2B; Table S3**), suggesting that increased intratumoral
134 heterogeneity associates with increased transcriptional heterogeneity.

135 We compared outcomes amongst the three transcriptional groups and observed
136 no significant differences (**Figure S4**). However, when restricting the analysis to
137 samples with high simplicity scores, a clear trend of MES showing worst survival and
138 PN the most favorable outcome became visible. For example, Kaplan-Meier analysis of
139 88 samples with simplicity scores >0.99 showed a median survival of 11.4, 14.7 and
140 16.7 months were detected in MES, CL and PN, respectively, which was significantly
141 different (log rank test, p=0.048)(**Figure 2C)(Figure S4)(Table S4**). Higher simplicity
142 scores correlated with relative favorable outcome within the PN set, non-significant in
143 the CL subtype, and correlated with relatively unfavorable survival in the MES class
144 (**Figure S5**).

145 Single GBM cell RNA sequencing recently suggested that GBMs are comprised
146 of a mixture of tumor cells with variable GBM subtype footprints [18]. We used this data to
147 classify 502 single GBM cells in addition to the bulk tumor derived from five primary
148 glioblastomas (**Table S5**). All bulk tumor samples showed simplicity scores less than
149 0.05 suggesting high transcriptional heterogeneity compared to 45 of 369 TCGA GBM

150 samples with simplicity scores below 0.05 (**Figure 2D**). In four of five cases (MGH26,
151 MGH28, MGH29 and MGH30), the bulk tumor samples were classified in the same
152 primary subtype as the majority of their single cells (**Figure 2D**). Our analysis suggests
153 that the heterogeneity observed at the single cell level is captured in the expression
154 profile of the bulk tumor, and that the five GBM samples studied at the single cell level
155 represented samples with relatively high transcriptional heterogeneity.
156
157 ***Transcriptional subtypes differentially activate the immune microenvironment***
158 Despite restricting the cluster analysis to genes exclusively expressed by GBM cells, we
159 found that tumor purity predictions based on ABSOLUTE were significantly reduced in
160 GBM classified as MES (Student T-test p-value < 10e-14; **Figure 3A**). This was
161 corroborated by gene expression based predictions of tumor purity using the
162 ESTIMATE method (Student T-test p-value < 10e-32; **Figure 3B**)[15] . The ESTIMATE
163 method has been optimized to quantify tumor-associated fibroblasts and immune cells
164 [15] and the convergence of a decreased ABSOLUTE and decreased ESTIMATE tumor
165 purity confirmed previous suggestions on the increased presence of microglial and
166 neuroglial cells in mesenchymal GBM [12,27,28]. The mean simplicity score of samples
167 classified as MES was 0.53 which was significantly lower than in PN (Wilcoxon rank test
168 p-value<0.019) and CL subtypes (Wilcoxon rank test p-value<0.0001), confirming
169 increased transcriptional heterogeneity. In order to identify genomic determinants of
170 macrophage/microglia chemoattraction, we compared genomic alterations between
171 mesenchymal class samples with high (n=51) and low (n=51) ABSOLUTE based tumor
172 purity. GBM carrying hemyzygous loss of *NF1* or somatic mutations in *NF1* showed

173 reduced tumor purity compared to GBM with wild type *NF1* (Wilcoxon rank test p-
174 value=0.0007) and this association was similarly detected when limiting the analysis to
175 MES samples (Wilcoxon rank test p-value=0.017)(**Figure S6**). Formation of dermal
176 neurofibromas in the context of *Nf1* loss of heterozygosity has been reported to be
177 context and microenvironment dependent [29]. Functional studies may clarify whether
178 NF1 deficient GBM are able to recruit cells that provide them with a proliferative
179 advantage, or whether *NF1* loss provides that proliferative advantage in a specific
180 tumor-associated microenvironment context.

181 To determine the cellular components of the tumor microenvironment across
182 different transcriptional subtypes, we used the CIBERSORT in silico cytometry [30]
183 method to evaluate absolute immune cell fractions. We evaluated 22 different immune
184 cell types in 69 PN, 137 CL and 96 MES samples, after filtering samples with
185 classification simplicity scores less than 0.1 (**Table S6**). Microglia are the resident
186 macrophages in the central nervous system. Peripheral blood monocytes also give rise
187 to tumor associated macrophages. These innate immune cells can be broadly classified
188 as the proinflammatory M1 type and the alternative tumor promoting M2 type[31]. The M2
189 macrophage gene signature showed a greater association with the MES subtype
190 (13.4%) relative to the PN (4.6%) and CL (6.0%)(**Figure 3C**), consistent with previous
191 analysis of the TCGA database[28,32]. In addition to the M2 macrophage gene signature,
192 there was also a significantly higher fraction of MES samples that expressed M1
193 macrophage (Student T-test p-value 3.20E-5) and neutrophil (Student T-test p-value
194 1.30E-9) gene signatures. In contrast, the activated natural killer T-cell gene signature
195 (Student T-test p-value 4.91E-2) was significantly reduced in the MES subtype and

196 resting memory CD4+ T cells (Student T-test p-value 5.40E-7) were less frequently
197 expressed in the PN subtype.

198 To confirm the association of macrophages/microglia with the MES GBM subtype,
199 we assessed protein expression levels of the ITGAM (alternatively known as CD11B)
200 and IBA1 (also known as AIF1) macrophage/microglial markers in a set of 18 GBM for
201 which we also characterized the expression subtype (**Figure 3D**) as well as through
202 immunohistochemistry (**Figure S7**). We confirmed the microenvironment as the main
203 source for *ITGAM/IBA1* transcription by comparing transcriptional levels in 37 GBM-
204 neurosphere pairs used for gene filtering, which showed that neurospheres do not
205 express *ITGAM/IBA1* (**Figure 3E**). The association of the MES GBM subtype with
206 increased level of M2 microglia/macrophages may suggest that in particular MES GBM
207 are candidates for therapies directed against tumor-associated macrophages[17].
208 Activated dendritic cell signatures (Student T-test p-value 7.36E-3)(**Figure 3C**) were
209 significantly higher in the CL subtype, suggesting this subtype may benefit from
210 dendritic cell vaccines [33]. Dendritic cells may require an activated phenotype in order to
211 direct the immune system. A previous study suggested that MES GBM patients treated
212 with dendritic cells were more likely to benefit [34].

213

214 ***Phenotypic plasticity upon GBM recurrence***

215 Glioblastoma has long been hypothesized to progress along a proneural to
216 mesenchymal axis[7]. To determine the relevance of this transition process in IDH
217 wildtype glioma evolution, we performed a longitudinal analysis of the subtype
218 classification and tumor-associated microenvironment in sample pairs obtained at

219 diagnosis and first disease recurrence from 124 glioma patients. The cohort included 96
220 initial GBM and first recurrence, eight pairs of primary low grade glioma and matching
221 secondary GBM, and 20 pairs of primary and recurrent low grade glioma. Gene
222 expression profiles of 78 tumor pairs were analyzed through transcriptome sequencing,
223 and remaining pairs were generated using Affymetrix (n = 31) and Illumina (n = 15)
224 microarray, respectively. To facilitate exploration of this dataset we have made it
225 available through the GlioVis portal http://ackbar.cnio.es:3838/RecuR/. We used a gene
226 expression signature[35] to determine that 33 of 124 cases were IDH-mutant/GCIMP at
227 presentation and recurrence (**Table S7**). We used the renewed gene signatures and
228 classification method to determine molecular subtype of the 91 pairs of IDH wild type
229 cases and found that expression class remained consistent after disease recurrence for
230 48 of 91 IDH-wildtype cases (52%)(**Figure 4A**). The MES subtype was most stable
231 (64%) while the CL (47%) and PN (43%) phenotypes were less frequently retained.
232 Nine, sixteen and eighteen post-treatment tumors switched subtypes to become CL,
233 MES and PN at disease recurrence, respectively, indicating that PN and MES increased
234 in frequency after recurrence while the CL subtype was least frequently found (**Figure**
235 **4A**). The CL expression class was previously found to be most sensitive to intensive
236 therapy and it is possible that therapy provides a competitive advantage for non-CL
237 cells, which could explain the reduced post-treatment incidence[3]. Our results did not
238 identify enrichment for proneural to mesenchymal transitions.

239 We observed a significant difference in transcriptional simplicity between primary
240 GBM retaining their expression class, versus those that switched to a different
241 phenotype (**Figure 4B**). GBMs with a primary tumor simplicity score greater than 0.5,

242 indicated lower transcriptional heterogeneity, were classified as the same subtype in 31
243 of 48 (64.5%) cases, compared to 15 of 41 (36.6%) cases with primary tumor simplicity
244 scores less than 0.5 (Fisher exact test p-value=0.01).
245
246 ***Microenvironment transitions upon GBM recurrence***
247 Debulking surgery, radiotherapy and chemotherapy provide therapeutic barriers but
248 nonetheless induce tumor evolution, including influences on the tumor
249 microenvironment. We explored this possibility by comparing the tumor associated
250 microenvironment in primary and recurrent GBMs using CIBERSORT (**Table S8**)[30]. A
251 comparison between 91 primary and recurrent IDH-wild type tumors revealed a
252 decrease in monocyte gene signature expression at recurrence, suggesting relative
253 depletion of circulation derived monocytes (**Figure 5A**). Next, we dissected
254 microenvironment fluctuations between diagnosis and recurrent tumors across different
255 subtype combinations. Primary non-MES (CL or PN) tumors showed relatively high
256 tumor purity and consequently, recurrent tumors classified as non-MES demonstrated a
257 relatively global decrease of immune cells while cases transitioning to MES at
258 recurrence represented a trend towards increased immune cell fractions (**Figure 5B**).
259 Gene signatures of immunosuppressive regulatory T cells showed an increase in gene
260 expression at recurrence across several primary-recurrence subtype combinations
261 although the inferred cellular fractions are relatively small (**Figure 5B**). In contrast to the
262 trend of monocyte depletion, the imputed M2 macrophage frequency was significantly
263 higher at recurrence in cases transitioning to MES (**Figure 5C**). This observation
264 converges with the higher fraction of M2 macrophages in primary MES GBM relative to

265 primary non-MES GBM. M1 macrophages and neutrophils also correlated with primary
266 MES GBM, but these associations were not confirmed for recurrent GBM. We validated
267 the increase in macrophages using immunostaining of IBA1 expression in two primary-
268 recurrent GBM pairs which were classified as CL to MES (**Figure 5D**). IBA1
269 immunoexpression was restricted to macrophages/microglia, cells exhibiting either
270 globular or filamentous/spidery morphology, with no expression in glioma tumor cells
271 (**Figure 5D**). These findings further solidify the association between MES GBM and
272 macrophage/microglia and extend this mutual relationship to disease recurrence. MES
273 tumors at recurrence compared to primary MES tumors showed an increase in
274 transcriptional activity associated with non-polarized M0 macrophages, which has been
275 previously described [28], but also dendritic cells which is potentially motivated by the
276 increased levels of neoantigens at disease recurrence [24]. In contrast, primary PN GBM
277 were found to contain significantly higher fractions of five immune cell categories
278 compared to recurrent PN GBM, indicating a relative absence of immune infiltration in
279 PN GBM upon recurrence.

280 We evaluated the effect of transcriptional class on patient survival. The analysis
281 was restricted to 50 cases for whom annotation on overall survival (OS) time and time to
282 disease progression (PFS) were available and with high simplicity scores, indicating low
283 transcriptional heterogeneity. We confirmed the worse prognosis for patients whose
284 primary tumor was classified as MES on overall survival (logrank test p=0.029 with
285 HR=1.97)(**Figure 6A; Figure 6B**). This pattern was retained in patients whose
286 secondary glioma was classified as MES (logrank test p=0.09 with HR=1.71)(**Figure 6C;**
287 **Figure 6D**). Consequently, cases for whom both primary and recurrent tumor were

288 classified as MES subtype showed the least favorable outcome, suggesting an additive

289 effect of transcriptional class at different time points (**Figure 6E, Figure 6F**)(**Figure S8**).

290 **DISCUSSION**

291 Transcriptome profiling of tumor samples is a commonly used modality for interrogating
292 pathway functionality and phenotype based patient classification. The transcriptional
293 footprint left by the tumor microenvironment, which may constitute 10-80% of cells in a
294 tumor biopsy[15], can obscure the true activity of the signaling network[14,36]. Here, we
295 employed in silico methods to integrate mRNA expression profiles from glioma samples
296 and glioma cell culture models to provide insights into glioma-intrinsic pathway activities
297 and classification, and to deconvolute the glioma associated stroma into its
298 immunological cellular components.

299 GBM expression subtype classification has emerged as an important concept to
300 better understand the biology of this devastating disease [8,37,38]. Robust classification of
301 new GBM tumors is therefore critical to ensure consistency in reporting between
302 different studies. The transcriptional glioma subtypes we discovered using tumor-
303 intrinsic gene expression values strongly overlapped with the proneural, classical and
304 mesenchymal subtypes but identified the neural class as normal neural lineage
305 contamination. Our updated methods, released through a R-library, were found to be
306 highly robust and provide the community with a standardized strategy for classification
307 of gliomas.

308 Through re-classification of primary GBM samples from TCGA and despite using
309 tumor-only transcripts, we observed that the mesenchymal GBM subtype associated
310 with the presence of tumor-associated glial and microglial cells. Mesenchymal glioma
311 cell differentiation status has been found to correlate with enrichment of
312 macrophages/microglia [10,39]. Through in silico cell type identification we additionally

313 detected enrichment of various adaptive immunity cell types, including CD4+ T
314 lymphocytes.

315 Longitudinal analysis of tumor samples is complicated by the lack of tissue
316 collections including such pairs. Through aggregation of existing and novel datasets we
317 compiled a cohort of 124 glioma pairs, including 91 pairs of IDH wild type tumors.
318 Comparison of pairs of initial gliomas and first disease recurrence did not identify the
319 trend of proneural GBM transitioning to a mesenchymal phenotype that has often been
320 suspected [7]. Mesenchymal subtype at diagnosis and at disease recurrent correlated
321 with relatively poor outcome. The recurrent IDH wild type GBM immuno environment
322 showed fewer blood derived monocytes which may reflect lower penetration through the
323 blood brain barrier. While the frequency of M2 macrophage/microglia was increased in
324 recurrent mesenchymal GBM compared to primary non-mesenchymal GBM, the overall
325 fraction of M2 macropage/microglia remained stable. This possibly suggests that the
326 majority of these cells are derived from resident CNS macrophages than through active
327 recruitment from the circulation.

328 In summary, our study defines a new strategy to determine transcriptional
329 subtype, and associated expression classes to the tumor-associated immuno-
330 environment. Our findings may aid in the implementation of immunotherapy approaches
331 [40] in a disease type with very limited treatment options. Collectively, our results have
332 improved our understanding of determinants of GBM subtype classification, the critical
333 impact of the tumor microenvironment, and provide new handles on the interpretation of
334 transcriptional profiling of glioma.

335 **Acknowledgments**

336 The authors thank Katherine Stemke-Hale for assistance in manuscript editing. The
337 results published here are in part based upon data generated by The Cancer Genome
338 Atlas project established by the National Cancer Institute (NCI) and the National Human
339 Genome Research Institute (NHGRI) of the National Institutes of Health. Information
340 about TCGA and the investigators and institutions that constitute the TCGA research
341 network can be found at http://cancergenome.nih.gov. This work is supported by grants
342 from the National Institutes of Health grants P50 CA127001 (EPS, RGWV), R01
343 CA190121 (EPS, RGWV), P01 CA085878 (RGWV); RO1 CA120813 (ABH), and P30
344 CA016672 (MD Anderson Cancer Center Support Grant for the Sequencing and
345 Microarray Facility); Cancer Prevention & Research Institute of Texas (CPRIT) grant
346 number R140606, the University Cancer Foundation via the Institutional Research
347 Grant program at the University of Texas MD Anderson Cancer Center (RGWV); the
348 National Brain Tumor Association Defeat GBM project (EPS, RGWV), the National
349 Brain Tumor Association Oligo Research Fund (RGWV).

References

1 Brennan, C. W. *et al.* The somatic genomic landscape of glioblastoma. *Cell* **155**, 462-477, doi:10.1016/j.cell.2013.09.034 (2013).

2 Cancer Genome Atlas Research, N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061-1068, doi:10.1038/nature07385 (2008).

3 Verhaak, R. G. *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer cell* **17**, 98-110, doi:10.1016/j.ccr.2009.12.020 (2010).

4 Noushmehr, H. *et al.* Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* **17**, 510-522, doi:10.1016/j.ccr.2010.03.017 (2010).

5 Ceccarelli, M. *et al.* Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell* **164**, 550-563, doi:10.1016/j.cell.2015.12.028 (2016).

6 Verhaak, C. M., Lintsen, A. M., Evers, A. W. & Braat, D. D. Who is at risk of emotional problems and how do you know? Screening of women going for IVF treatment. *Hum Reprod* **25**, 1234-1240, doi:10.1093/humrep/deq054 (2010).

7 Phillips, H. S. *et al.* Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer cell* **9**, 157-173, doi:10.1016/j.ccr.2006.02.019 (2006).

8 Huse, J. T., Phillips, H. S. & Brennan, C. W. Molecular subclassification of diffuse gliomas: seeing order in the chaos. *Glia* **59**, 1190-1199, doi:10.1002/glia.21165 (2011).

9 Zheng, S., Chheda, M. G. & Verhaak, R. G. Studying a complex tumor: potential and pitfalls. *Cancer J* **18**, 107-114, doi:10.1097/PPO.0b013e3182431c57 (2012).

10 Bhat, K. P. *et al.* Mesenchymal differentiation mediated by NF-kappaB promotes radiation resistance in glioblastoma. *Cancer cell* **24**, 331-346, doi:10.1016/j.ccr.2013.08.001 (2013).

11 Ozawa, T. *et al.* Most human non-GCIMP glioblastoma subtypes evolve from a common proneural-like precursor glioma. *Cancer Cell* **26**, 288-300, doi:10.1016/j.ccr.2014.06.005 (2014).

12 Bao, S. *et al.* Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature* **444**, 756-760, doi:10.1038/nature05236 (2006).

13 Olar, A. & Aldape, K. D. Using the molecular classification of glioblastoma to inform personalized treatment. *The Journal of pathology* **232**, 165-177, doi:10.1002/path.4282 (2014).

14 Isella, C. *et al.* Stromal contribution to the colorectal cancer transcriptome. *Nat Genet* **47**, 312-319, doi:10.1038/ng.3224 (2015).

15 Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* **4**, 2612, doi:10.1038/ncomms3612 (2013).

16 Martinez, E. *et al.* Comparison of gene expression patterns across 12 tumor types identifies a cancer supercluster characterized by TP53 mutations and cell cycle defects. *Oncogene* **34**, 2732-2740, doi:10.1038/onc.2014.216 (2015).

17 Pyonteck, S. M. *et al.* CSF-1R inhibition alters macrophage polarization and blocks glioma progression. *Nat Med* **19**, 1264-1272, doi:10.1038/nm.3337 (2013).

18 Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396-1401, doi:10.1126/science.1254257 (2014).

19 Yan, J. *et al.* FGL2 as a Multimodality Regulator of Tumor-Mediated Immune Suppression and Therapeutic Target in Gliomas. *J Natl Cancer Inst* **107**, doi:10.1093/jnci/djv137 (2015).

20 Galli, R. *et al.* Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma. *Cancer Res* **64**, 7011-7021, doi:10.1158/0008-5472.CAN-04-1364 (2004).

21 Cancer Genome Atlas Research, N. *et al.* Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N Engl J Med* **372**, 2481-2498, doi:10.1056/NEJMoa1402121 (2015).

22 Gill, B. J. *et al.* MRI-localized biopsies reveal subtype-specific differences in molecular and cellular composition at the margins of glioblastoma. *Proc Natl Acad Sci U S A* **111**, 12550-12555, doi:10.1073/pnas.1405839111 (2014).

402 23 Verhaak, R. G. *et al.* Prognostically relevant gene signatures of high-grade serous ovarian
403 carcinoma. *J Clin Invest* **123**, 517-525, doi:65833/10.1172/JCI65833 (2013).
404 24 Kim, H. *et al.* Whole-genome and multisector exome sequencing of primary and post-treatment
405 glioblastoma reveals patterns of tumor evolution. *Genome Res* **25**, 316-327,
406 doi:10.1101/gr.180612.114 (2015).
407 25 Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat*
408 *Biotechnol* **30**, 413-421, doi:10.1038/nbt.2203 (2012).
409 26 Aran, D., Sirota, M. & Butte, A. J. Systematic pan-cancer analysis of tumour purity. *Nat Commun*
410 **6**, 8971, doi:10.1038/ncomms9971 (2015).
411 27 Ye, X. Z. *et al.* Tumor-associated microglia/macrophages enhance the invasion of glioma stem-
412 like cells via TGF-beta1 signaling pathway. *J Immunol* **189**, 444-453,
413 doi:10.4049/jimmunol.1103248 (2012).
414 28 Gabrusiewicz, K. *et al.* Glioblastoma-infiltrated innate immune cells resemble M0 macrophage
415 phenotype. *JCI insight* **1**, doi:10.1172/jci.insight.85841 (2016).
416 29 Le, L. Q., Shipman, T., Burns, D. K. & Parada, L. F. Cell of origin and microenvironment
417 contribution for NF1-associated dermal neurofibromas. *Cell Stem Cell* **4**, 453-463,
418 doi:10.1016/j.stem.2009.03.017 (2009).
419 30 Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat*
420 *Methods* **12**, 453-457, doi:10.1038/nmeth.3337 (2015).
421 31 Hambardzumyan, D., Gutmann, D. H. & Kettenmann, H. The role of microglia and macrophages
422 in glioma maintenance and progression. *Nature neuroscience* **19**, 20-27, doi:10.1038/nn.4185
423 (2015).
424 32 Doucette, T. *et al.* Immune heterogeneity of glioblastoma subtypes: extrapolation from the
425 cancer genome atlas. *Cancer Immunol Res* **1**, 112-122, doi:10.1158/2326-6066.CIR-13-0028
426 (2013).
427 33 Palucka, K. & Banchereau, J. Cancer immunotherapy via dendritic cells. *Nature reviews. Cancer*
428 **12**, 265-277, doi:10.1038/nrc3258 (2012).
429 34 Prins, R. M. *et al.* Gene expression profile correlates with T-cell infiltration and relative survival
430 in glioblastoma patients vaccinated with dendritic cell immunotherapy. *Clin Cancer Res* **17**,
431 1603-1615, doi:10.1158/1078-0432.CCR-10-2563 (2011).
432 35 Baysan, M. *et al.* G-cimp status prediction of glioblastoma samples using mRNA expression data.
433 *PloS one* **7**, e47839, doi:10.1371/journal.pone.0047839 (2012).
434 36 Kim, H. & Verhaak, R. G. Transcriptional mimicry by tumor-associated stroma. *Nat Genet* **47**,
435 307-309, doi:10.1038/ng.3255 (2015).
436 37 Dunn, G. P. *et al.* Emerging insights into the molecular and cellular basis of glioblastoma. *Genes*
437 *Dev* **26**, 756-784, doi:10.1101/gad.187922.112 (2012).
438 38 Sturm, D. *et al.* Paediatric and adult glioblastoma: multiform (epi)genomic culprits emerge. *Nat*
439 *Rev Cancer* **14**, 92-107, doi:10.1038/nrc3655 (2014).
440 39 Kreutzberg, G. W. Microglia: a sensor for pathological events in the CNS. *Trends Neurosci* **19**,
441 312-318 (1996).
442 40 Blank, C. U., Haanen, J. B., Ribas, A. & Schumacher, T. N. CANCER IMMUNOLOGY. The "cancer
443 immunogram". *Science* **352**, 658-660, doi:10.1126/science.aaf2834 (2016).

Figures and Legends

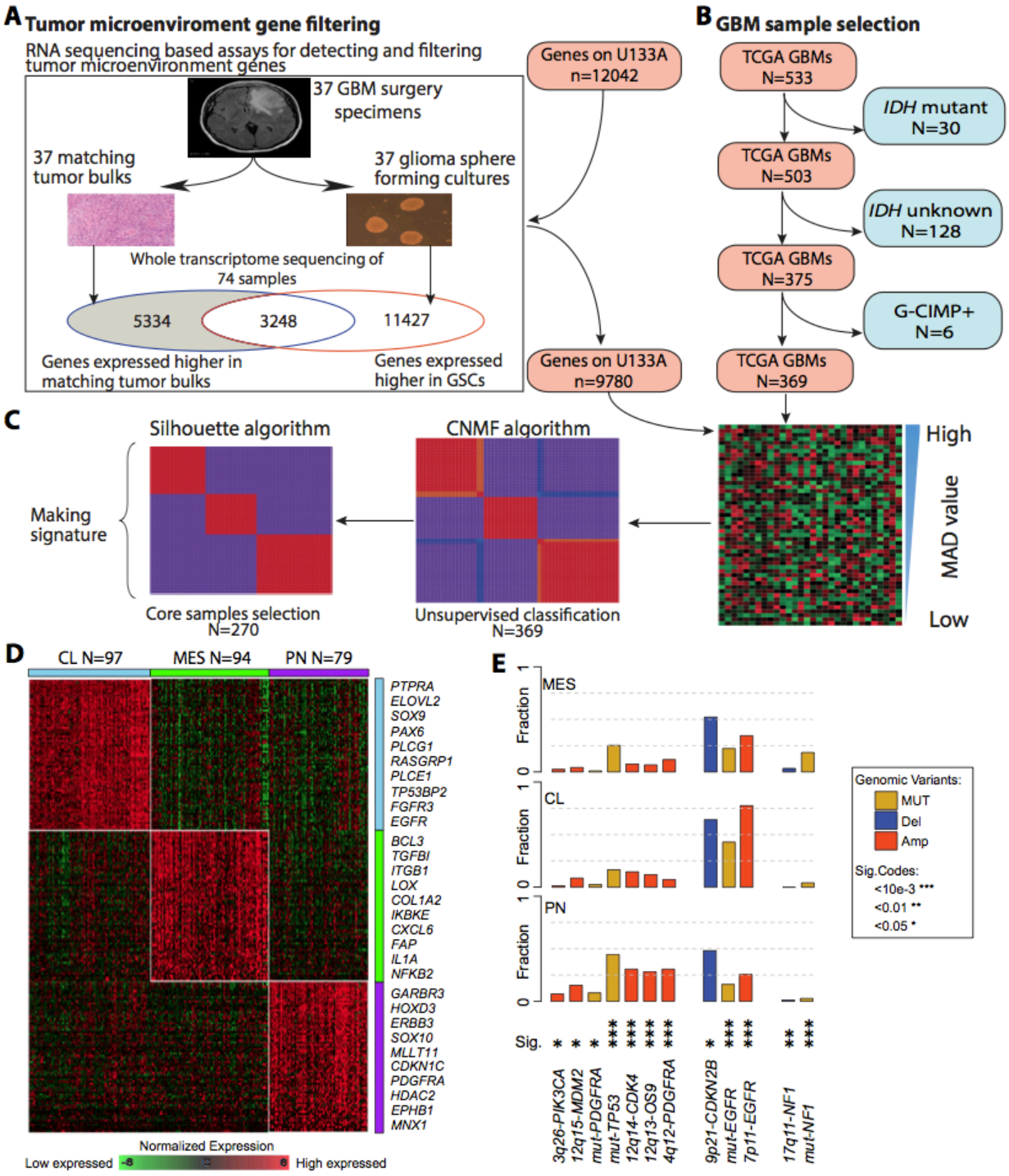


Figure 1. Molecular classification of IDHwt GBMs. (A) Filtering tumor associated microenvironment genes. (**B**) Discarding IDH mutation related GBMs. (**C**) Overview of NMF clustering. (**D**) Heatmap of 70-gene signatures by gene expression subtype were developed based on 270 GBMs. Top ten genes are shown for each subtype. (**E**) Frequency of subtype related somatic genomic alterations.

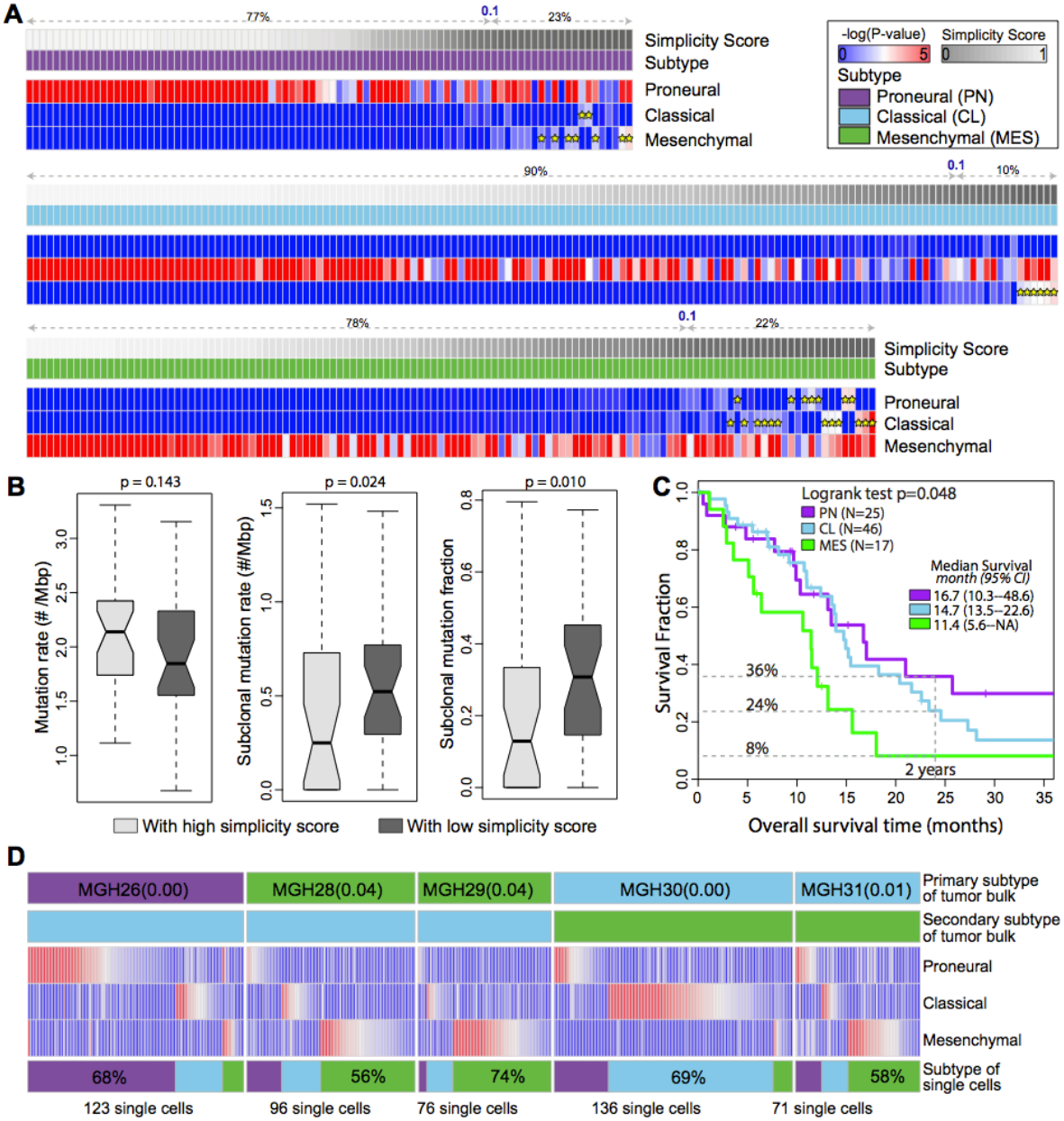


Figure 2. Multi activation of transcriptional subtypes associated with intratumoral heterogeneity. (**A**) The expression profiles of 369 IDHwt GBMs were analyzed using Affymetrix U133A. The empirical -log(P-value) of raw ssGSEA enrichment scores at each signature are shown as heatmaps, with dark blue representing no activation and bright red as highly activated. Yellow star indicates the secondary activated subtype (empirical p-value<0.05). For each panel, the first row shows simplicity score, and the second row indicates transcriptional subtype. (**B**) Comparison of mutation rate, subclonal mutation rate and subclonal mutation fraction between IDHwt GBMs with high and low simplicity scores. P-values were calculated using Wilcoxon rank test and shown at the top of each panel. (**C**) Kaplan-Meier survival curver by subtype. (**D**) Transcriptome classification of five bulk tumor samples and 502 single GBM cells derived from them. The top two row of each panel show the dominant and secondary subtype of the GBM tumor bulk. The heatmap of each panel shows the empirical -log(P-value) of the ssGSEA scores of the derived single GBM cells on each of the three subtype signatures. The bottom row shows the subtype distribution of derived single GBM cells within the same GBM tumor of origin.

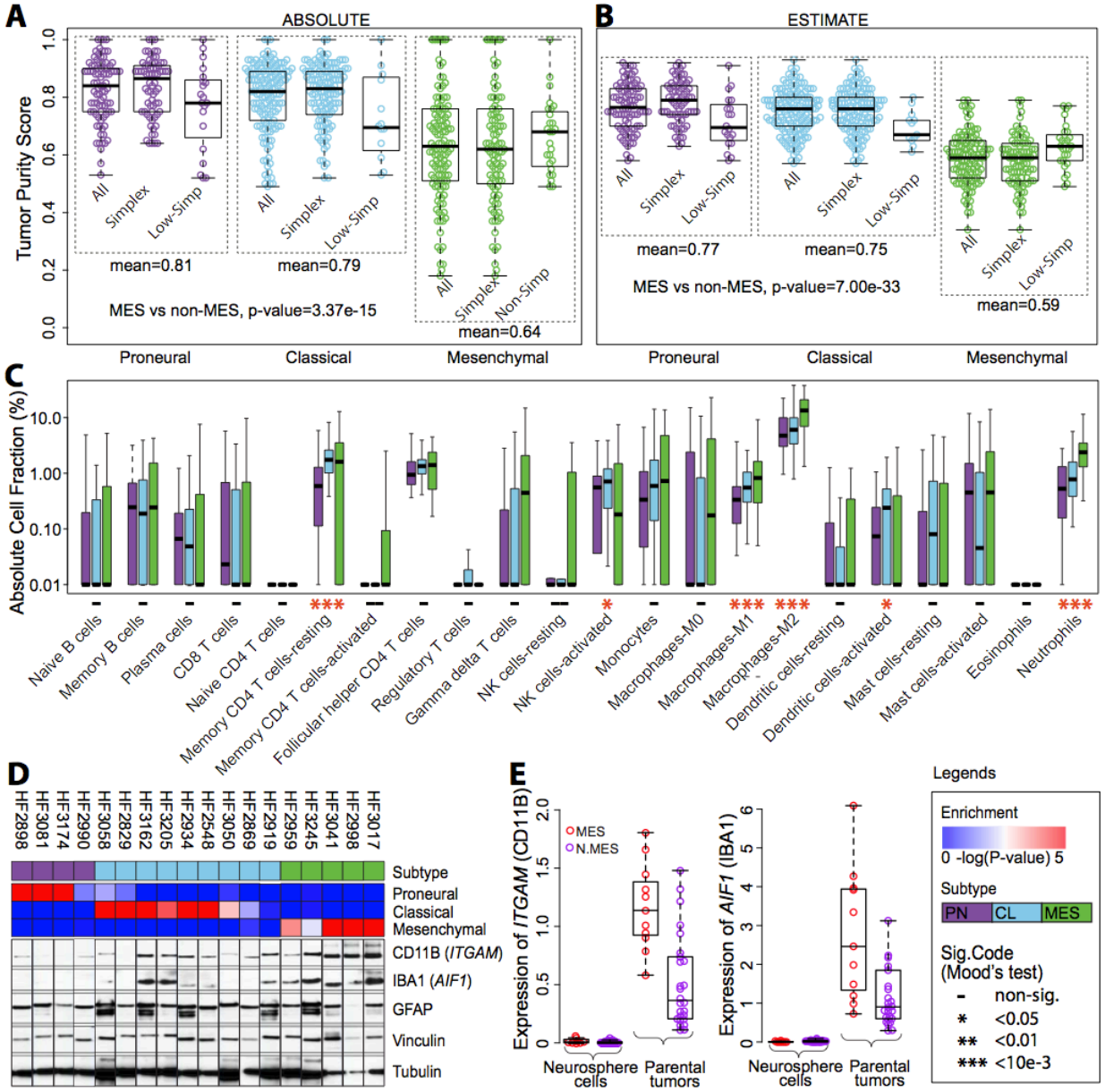


Figure 3. Transcriptional subtypes differentially activate the immune microenvironment. (**A, B**) Tumor purity of 364 respectively 369 TCGA IDHwt GBM samples was determined by ABSOLUTE and ESTIMATE. The difference in tumor purity between subtypes was evaluated using a two-sample heteroscedastic t-test. (**C**) Comparison of immune cell fraction among subtypes. Purple, skyblue and green boxplots indicate PN, CL and MES subtype, respectively. Immune cell fraction was estimated using CIBERSORT and corrected using ABSOLUTE purity scores. Difference of cell fraction between subtypes was evaluated using Mood's test. (**D**) The upper panel shows ssGSEA enrichment scores and associated expression subtype classifications. Bottom panels display protein expression of the microglial markers integrin alpha M (ITGAM) and allograft inflammatory factor 1 (IBA1), astrocyte marker glial fibrillary acidic protein (GFAP) and the loading control tubulin. (**E**) Comparison of *ITGAM* and *IBA1* gene expression levels between GBM and derived neurosphere models.

A GCIMP- Pri./Rec. Pairs

	PN	CL	MES	**Pri.**
PN	**9**	5	7	**21**
CL	9	**16**	9	**34**
MES	9	4	**23**	**36**
Rec.	**27**	**25**	**39**	91

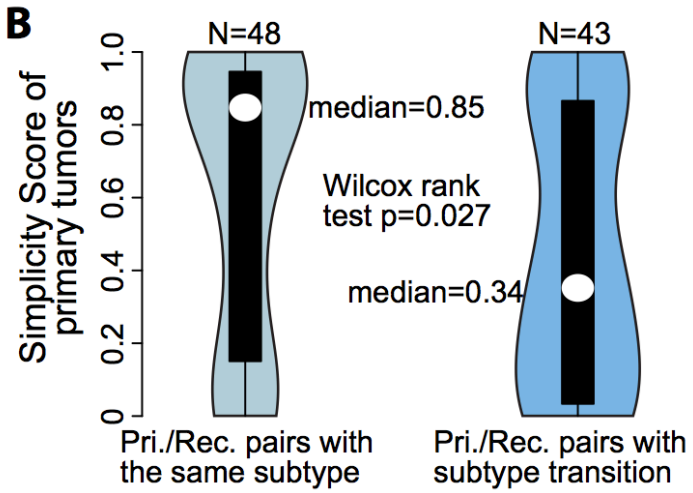


Figure 4. Comparison between transcriptional subtype of primary and paired recurrent tumors. (**A**) Rows and columns of the cross table represents subtype distribution frequency of primary and paired recurrent tumors, respectively. (**B**) Violin plots show the distribution of simplicity scores of pairs with (left) and without (right) subtype transition.

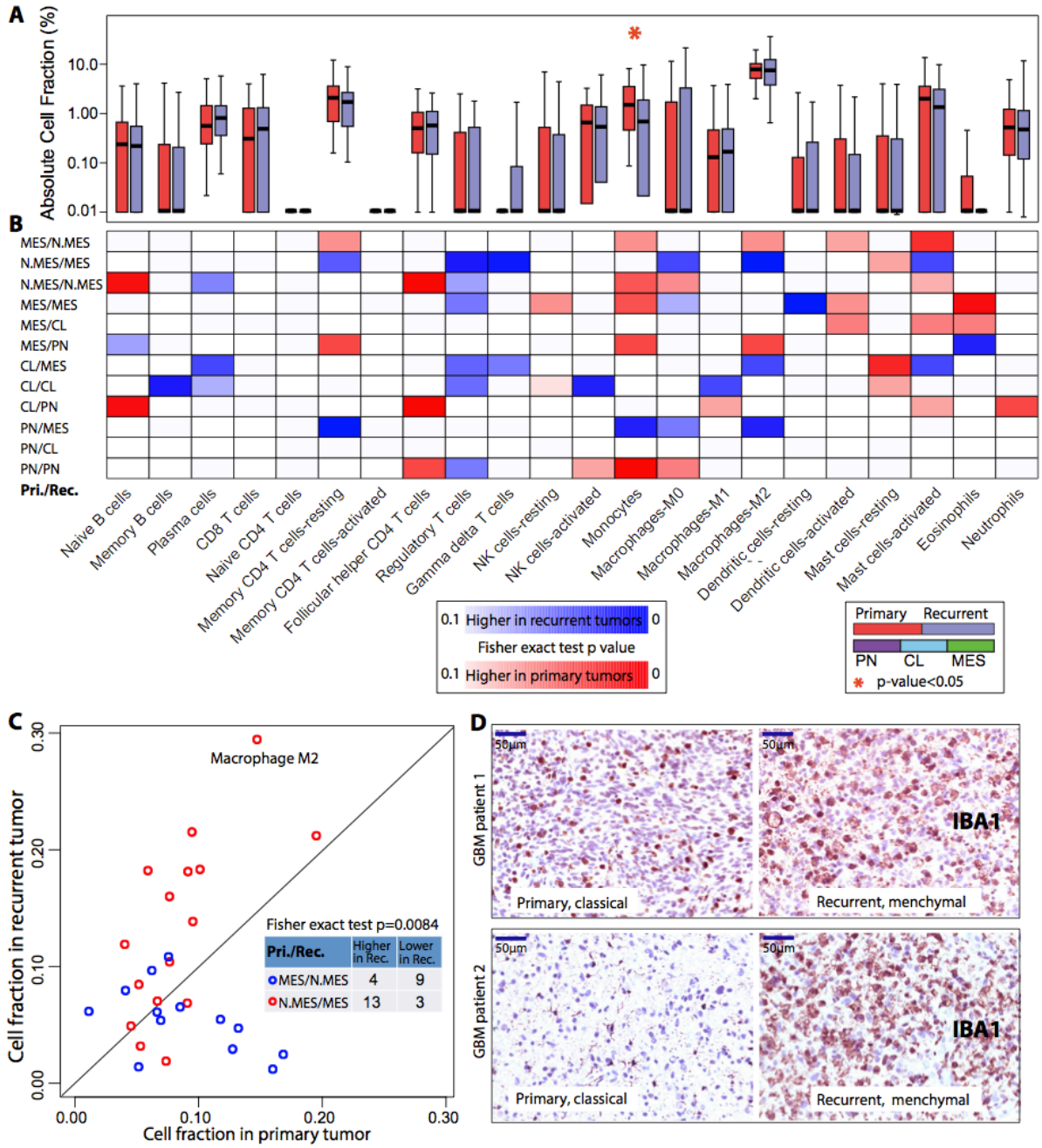


Figure 5. Microenvironment transition between primary and paired recurrent tumors. (**A**) Red and blue boxplots represent the immune cell fraction distribution of each immune cell type. Immune cell fraction was calculated using CIBERSORT and adjusted using ESTIMATE purity scores. Difference between cell fraction of primary and paired recurrent tumors was calculated using Wilcoxon rank test. (**B**) The blue-to-red heatmap represents immune cell fraction changes upon tumor recurrence per subtype transitions which were list on the left of the heatmap. (**C**) Each dot represents a pair of primary and recurrent GBM with axes indicating M2 macrophage cell fraction. (**D**) Immunohistochemical staining of the IBA1 microglia marker in two matched pairs of primary and recurrent GBM. Cellular morphology in combination with the distribution of the staining suggested exclusive microglial expression of monocyte lineage markers IBA1 in GBM tumors.

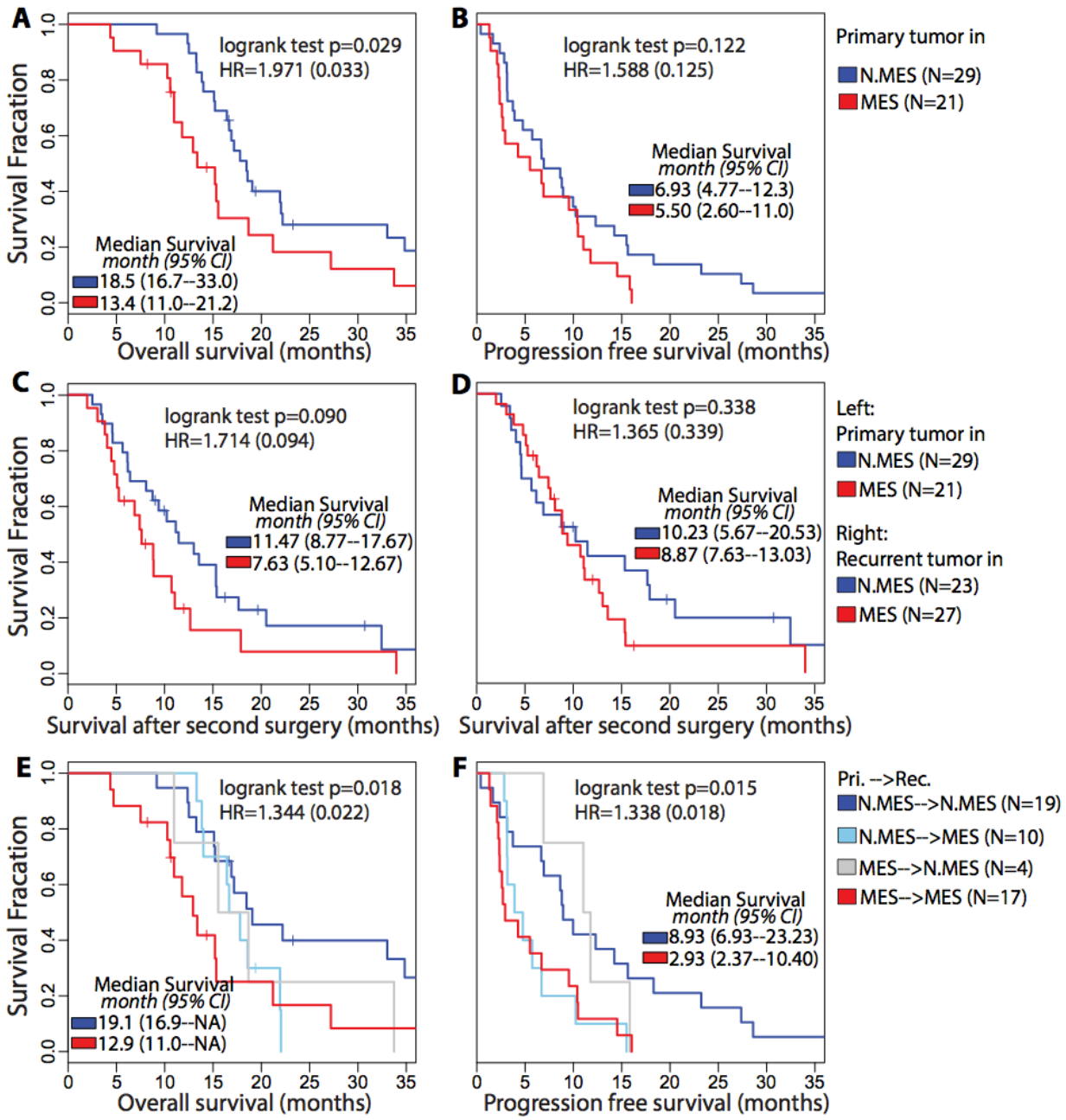
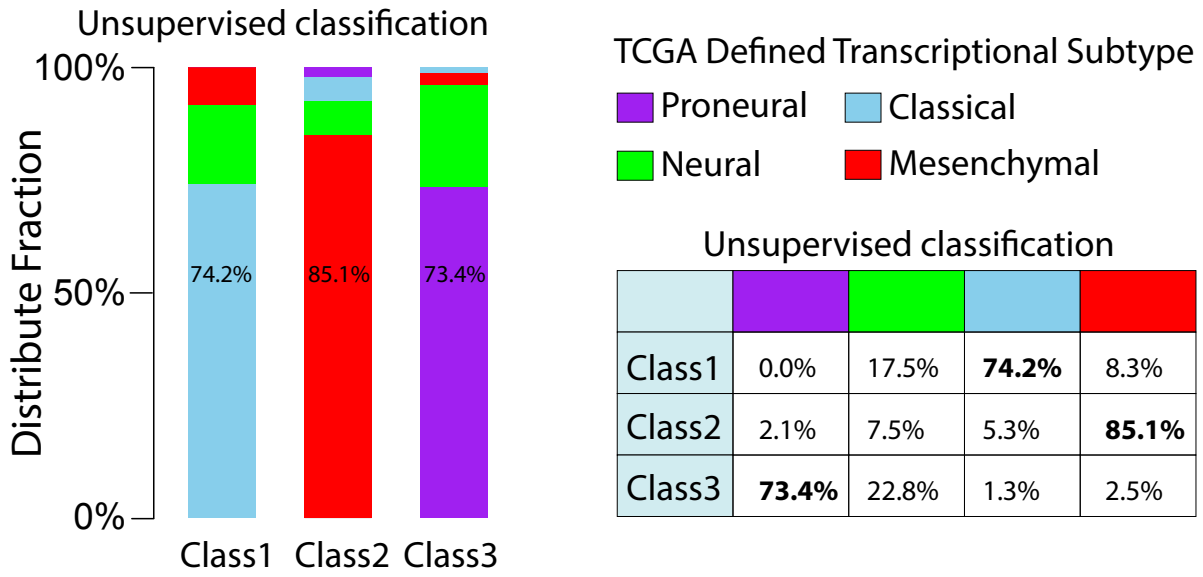


Figure 6. Survival analysis of paired IDH wild type GBM. (A, B) OS and PFS analyses between samples with different primary subtype. (C) Difference of survival time after secondary surgery between patients with non-MES and MES in primary tumors. (D) Survival analysis of time after secondary surgery between patients with non-MES and MES in recurrent tumors. (E, F) OS and PFS analyses between samples with difference recurrent subtype.

Supplementary Information

Supplementary Figures and Legends



Unsupervised classification

Class1	0.0%	17.5%	**74.2%**	8.3%
Class2	2.1%	7.5%	5.3%	**85.1%**
Class3	**73.4%**	22.8%	1.3%	2.5%

Figure S1. Comparison between GCIMP- GBM specific classification and TCGA defined GBM subtypes.

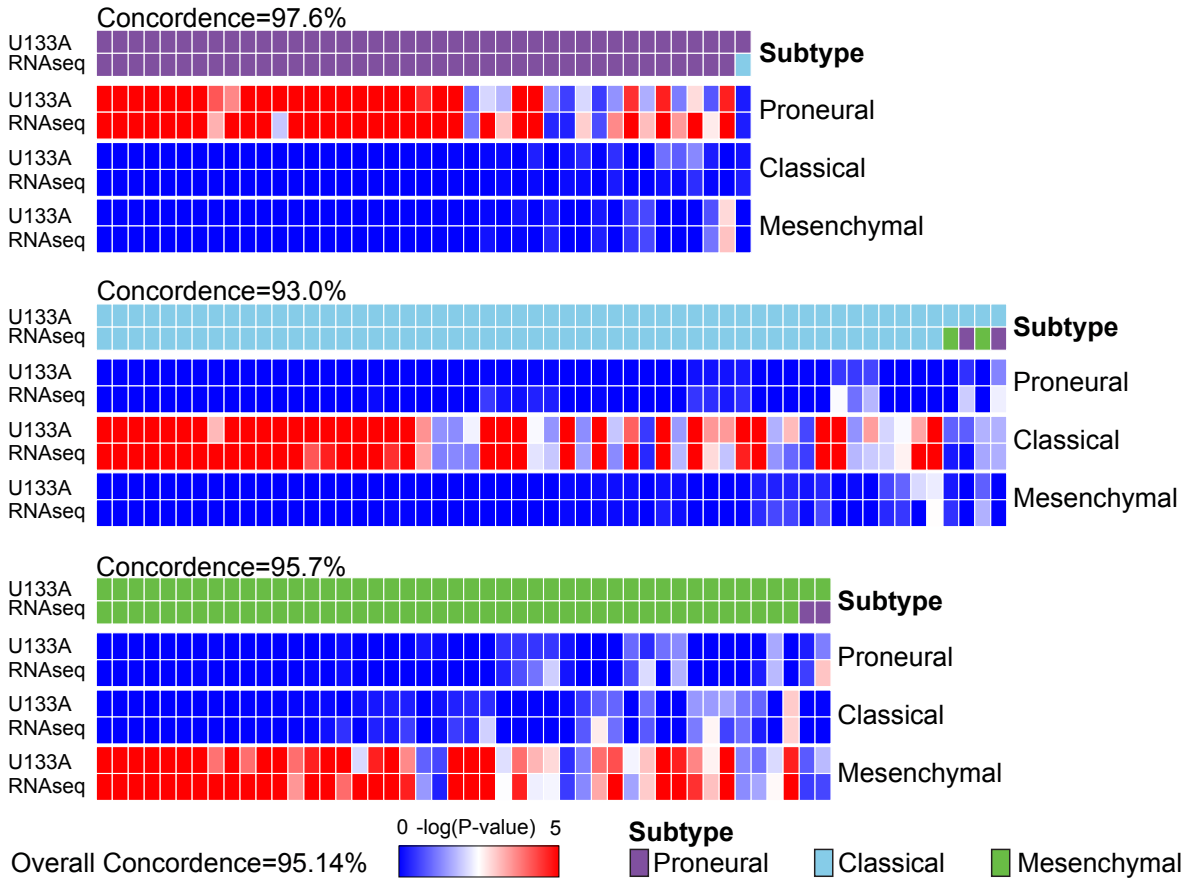


Figure S2. Concordance of transcriptional classification of GBMs cross multiple platforms. Through TCGA, the expression profiles of 144 GBM were analyzed using both Affymetrix U133A gene expression arrays and RNA sequencing. The empirical -log(P-value) of raw ssGSEA enrich- ment scores at each signature are shown as heatmaps, with dark blue representing no activation and bright red as highly activated. For each panel, the first row shows U133A based classification, and the second row indicates RNA-seq subtype classification.

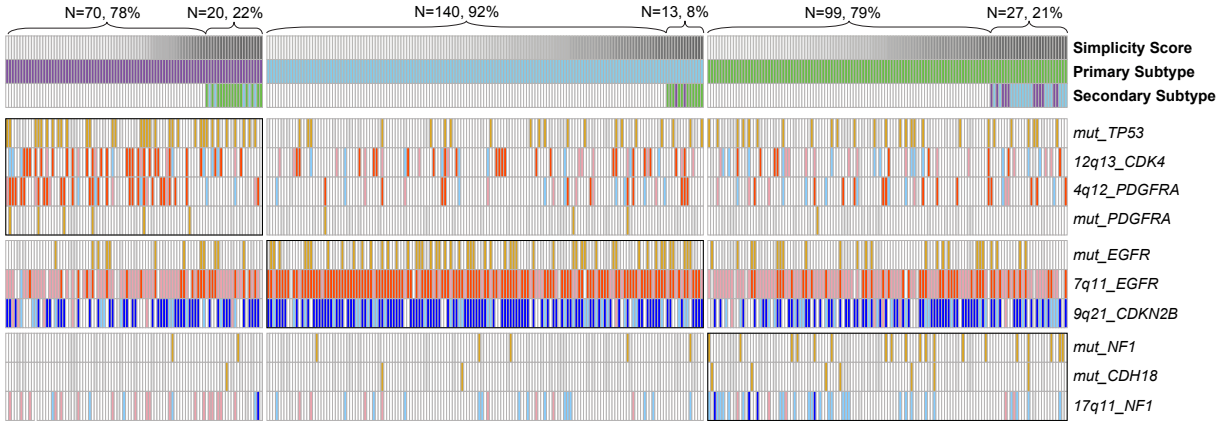


Figure S3. Genomic alteration patterns for each subtype. Ten prominent GBM somatic events were manually selected. Simplicity score, primary and secondary subtype were shown by the first three rows.

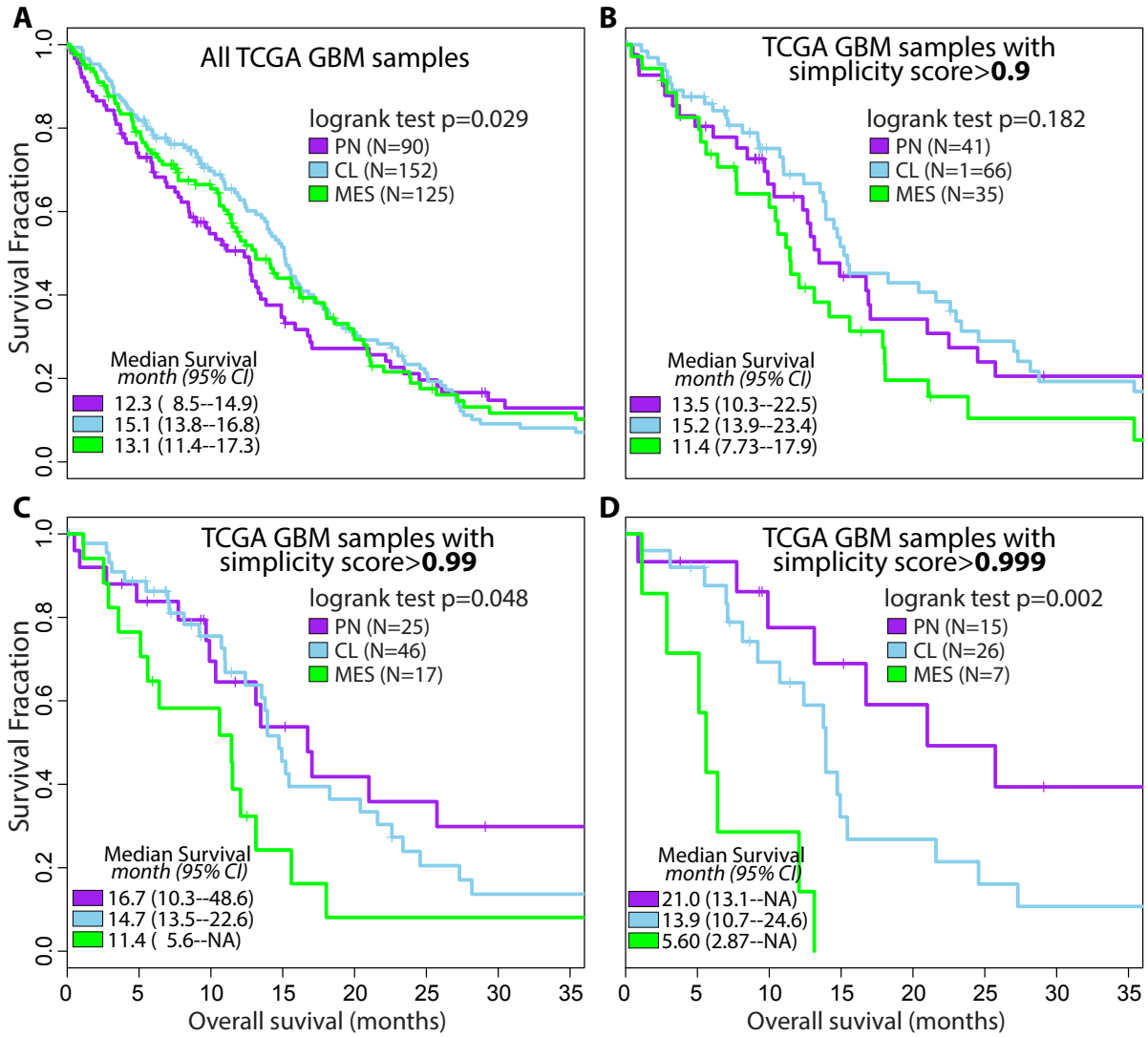
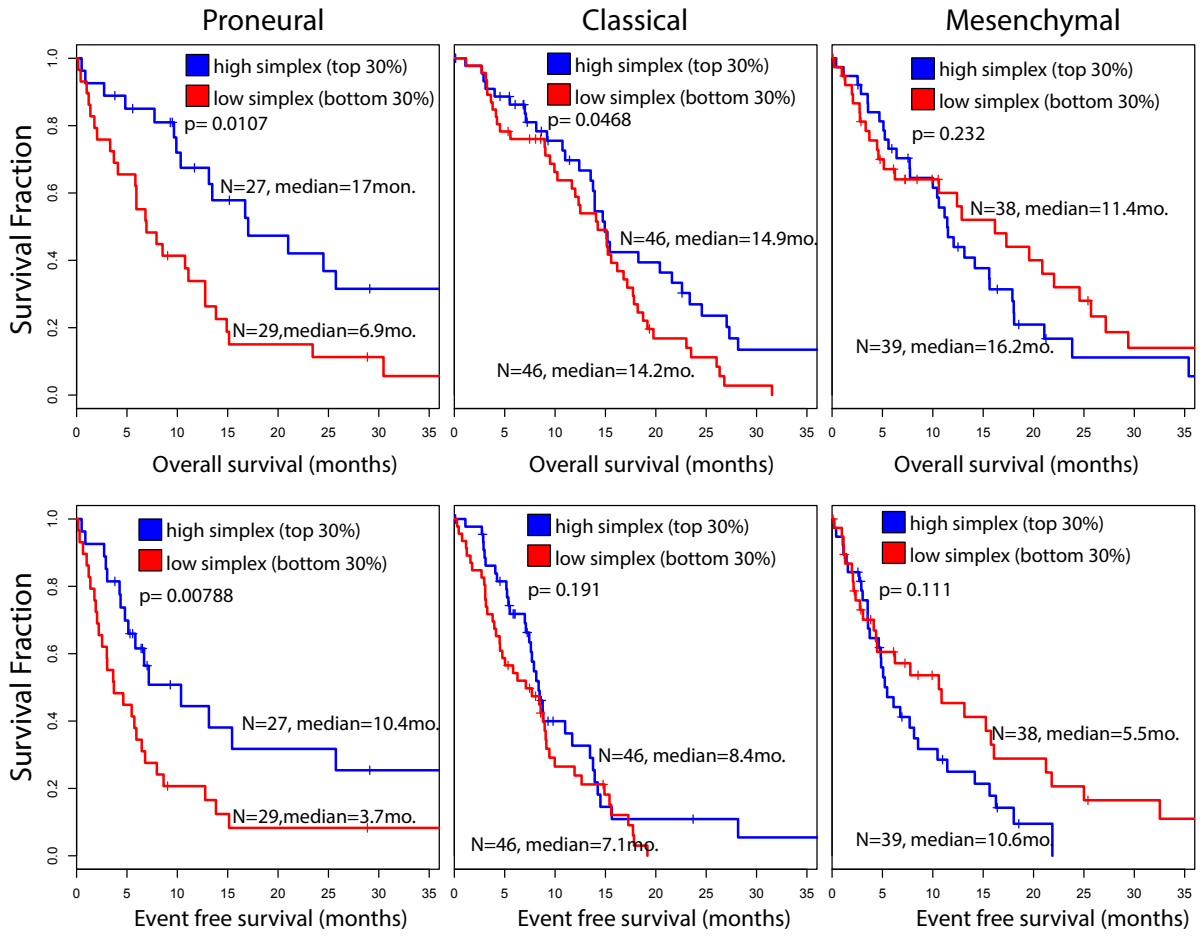


Figure S4. Patient survival differences between transcriptional subtypes. Samples were filtered using increased simplicity score as threshold from panel A to D.



523

524 **Figure S5.** Overall and event free survival analysis comparison between samples with high and
525 low simplicity scores in each subtype.

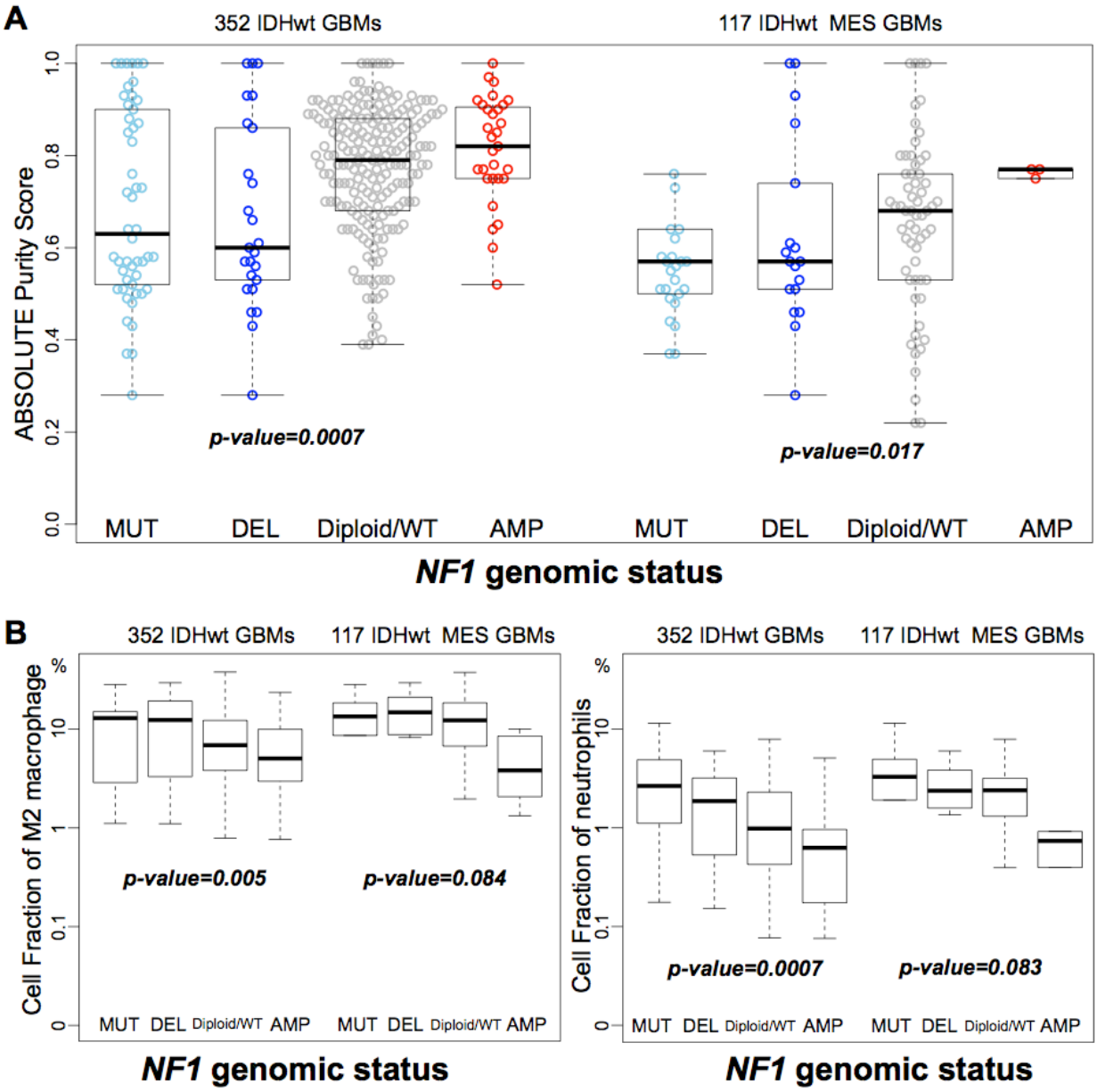


Figure S6. Comparison of tumor purity and immune cell fraction between GBMs with different *NF1* genomic status. P-values were calculated using Wilcoxon rank test between samples carrying *NF1* deletion/mutation and others.

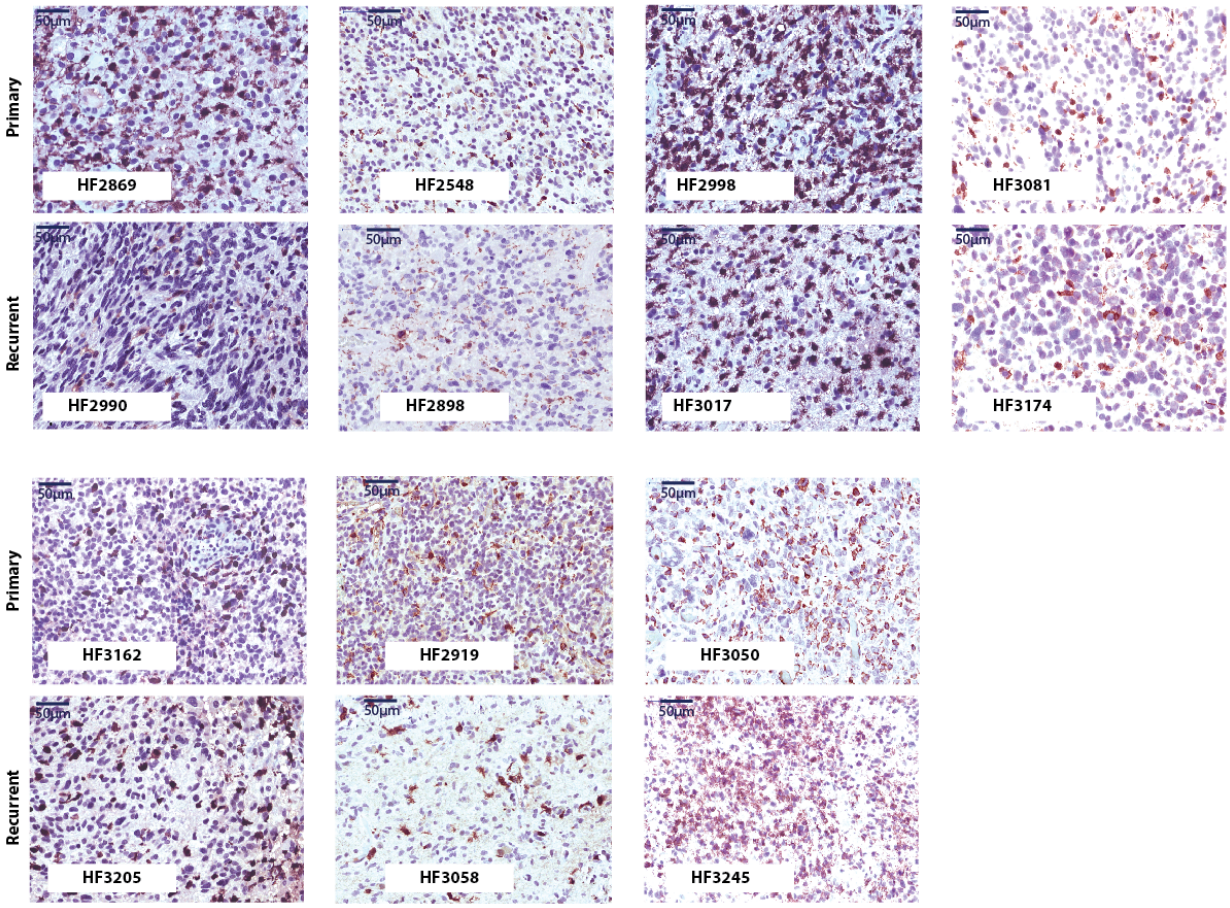


Figure S7. Immunohistochemistry staining of the IBA1 imcroglia marker.

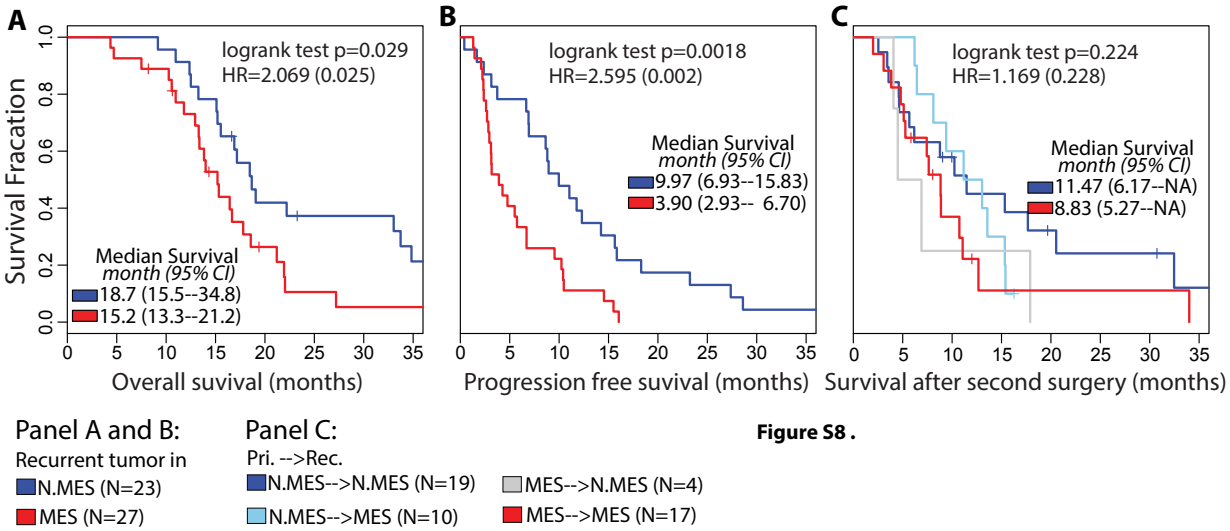


Figure S8 .

Figure S8. (A, B) Overall and progression free survival analysis between samples in different recurrent subtypes. **(C)** Survival after secondary surgery comparison between different transition types.

537 **Supplementary Methods and Experimental Procedures**

538 **Collection of pairs of primary and recurrent glioma samples**

539 U133A array profiles for 543 primary GBM, and RNA-Seq data for 166 primary and 13
540 recurrent GBM were obtained from the TCGA portal https://tcga-data.nci.nih.gov/tcga/.
541 Mutation calls and DNA copy number profiles were obtained for all samples, where
542 available. Tissues from 20 additional initial GBM and matched recurrent tumor were
543 obtained from Henry Ford Hospital (n = 9) in accordance with institutional policies and
544 all patients provided written consent, with approval from the Institutional Review Boards
545 (Henry Ford Hospital IRB protocol #402). All RNA samples tested were obtained from
546 frozen specimens. All of the recurrent GBMs had been previously treated with
547 chemotherapy and radiation. Three cases had a history of lower grade astrocytoma
548 prior to the first GBM (HF-2869/HF-3081/HF-3162). Tumors were selected solely on the
549 basis of availability. RNA-Seq libraries were generated using RNA Truseq reagents
550 (Illumina, San Diego, CA, USA) and paired-end sequenced using standard Illumina
551 protocols. Read length was 76 base pairs for cases sequenced by TCGA and from
552 Henry Ford (processed at MD Anderson). RNA-Seq data on frozen tissue from 44
553 patients with initial and recurrent GBM that received resection at Samsung Medical
554 Center and Seoul National University Hospital were provided by Dr. Nam's lab. Surgery
555 specimens were obtained in accordance to the Institutional Review Board (IRB) of the
556 Samsung Medical Center (No. 2010-04-004) and Seoul National University Hospital (No.
557 C-1404-056-572). Affymetrix CEL files of 39 pairs of initial and recurrent glioma were
558 retrieved from the Gene Expression Omnibus (GEO accession GSE4271, GSE42670,
559 GSE62153)[1-3]. The expression profiles of the 23 pairs from GSE4271 were determined

560 using Affymetrix HG-U133 GeneChips, the 1 pairs from GSE42670 were analyzed using
561 the Affmetrix HuGene-1-0-st platform, the 15 pairs from GSE62153 were analyzed
562 using Illumina Human HT-12 V4.0 expression BeadChip.

563 Genome wide DNA copy number profiling and exome sequencing on thirteen
564 TCGA tumor pairs and nine of ten Henry Ford tumor pairs were performed and data
565 was analyzed using standard protocols and pipelines as previously described [4].

566

567 **Data for multiplatform classification comparison**

568 RNA sequencing data was available for 162 primary GBMs[5] for which an Affymetrix HT-
569 U133A gene expression profile was also available. We observed a low Pearson
570 Correlation Coefficient (< 0.15) between RNA sequencing based reads per kilo base of
571 transcript per million reads (RPKM) and Affymetrix HT-U133A profiles in eighteen cases
572 and these were removed from further analysis. In summary, in order to assess the
573 concordance between classification results of the new 70-gene signatures and
574 previously published 210-gene signatures [6], 144 GBMs which were profiled in both RNA
575 sequencing and Affymetrix U133A platforms were used in our further analyses.

576

577 **Transcriptome Data processing**

578 The latest version custom CDF files (Version19, http://brainarray.mbni.med.umich.edu)
579 [7,8] were used to map probes from the Affymetrix HG-U133A and HuGene-1_0-st
580 GeneChip platforms to the Ensemble transcript database, combined in one probe set
581 per gene and normalized using the AROMA package with default parameters, resulting
582 in RMA normalized and log transformed gene expression values [9]. All RNA sequencing

583 data was processed by the PRADA pipeline [10]. Briefly, reads were aligned using BWA
584 against the genome and transcriptome. After initial mapping, the aligned reads were
585 filtered out if their best placements are only mapped to unique genomic coordinates.
586 Quality scores are recalibrated using the Genome Analysis Toolkit (GATK), and
587 duplicate reads are flagged using Picard. Mapped features were quantified and
588 normalized per kilo base of transcript per million reads (RPKM) and were converted to a
589 log2 scale to represent a gene expression level. RPKM values measuring the same
590 gene that mapped to the Ensemble transcript with longest size were selected to obtain
591 one expression value per gene and sample. RPKM values were converted to a log2
592 scale to represent gene expression level. The statistical environment R was used to
593 perform all the statistical analysis and graph plots.
594

595 **Deriving new gene signatures**

596 A pair-wise gene expression analysis identified 5,334 genes which are significant higher
597 expressed in glioma bulk samples compared to their derivative GSCs. These genes
598 were discarded from the gene list for developing tumor-specific molecular subtypes.
599 Consensus non-negative matrix factorization (CNMF) clustering method identified three
600 distinct subgroups among the 369 IDHwt primary GBMs. A set of 270 GBMs was
601 recognized as core samples based on a positive silhouette width. The gene expression
602 values of each subtype were compared with those from the other two subtypes
603 combined [6]. Signature genes per cluster were selected on the basis of differences in
604 gene expression level and were considered significant if they reached the cut-off value
605 with t-test p-value< 1E-3 for higher expressed in this class, while also showing a

606 significant lower expression with t-test p-value<1E-3 in the other two classes. In the
607 original gene signatures, genes could be either down-regulated or up-regulated, while
608 only up-regulated genes (n=70 per gene signature) were selected for revised gene
609 signatures. Only genes measured on both RNAseq and U133A platforms were
610 considered, and the U133A data from 162 GBM samples measured on both platforms
611 (which included the 144 cases used to compare U133A and RNAseq results) was used
612 in the final comparative analysis.
613
614 **Molecular classifications based on ssGSEA enrichment scores**
615 Single sample gene set enrichment analysis was performed as follows. For a given
616 GBM sample, gene expression values were rank-normalized and rank-ordered. The
617 Empirical Cumulative Distribution Functions (ECDF) of the signature genes and the
618 remaining genes were calculated. A statistic was calculated by integration of the
619 difference between the ECDFs, which is similar to the one used in GSEA but is based
620 on absolute expression rather than differential expression [11]. Since the ssGSEA test is
621 based on the ranking of genes by expression level, the uncentered and log-transformed
622 U133A and RPKM expression levels were used as input for ssGSEA. Since the scores
623 of the three signatures were not directly comparable, we performed a resampling
624 procedure to generate null distributions for each of the four subtypes. First we
625 generated a large number of virtual samples in which each gene obtains its expression
626 level by randomly selecting an expression value of the same gene in the remainder of
627 the samples. Then, the three ssGSEA scores for each signature were calculated.
628 Following this procedure we generated a large number (>1,000,000) of random ssGSEA

scores for each subtype, to build the null distribution and to give empirical p-values for the raw ssGSEA scores of each sample. By testing on multiple datasets with different sample sizes, we found the resampling generated distribution could be replaced with Student-T distribution (sample size>30) or Normal distribution (sample size>50) for getting very similar results. R-library with the code and expression matrices used is provided upon request.

Evaluate the simplicity of subtype activation

For a single sample, we decreased rank the empirical p-values for each subtype to generate order statistics as R_{N-1}, R_{N-2} ... R_1, R_0. In particular, R_0 equals to the minimum empirical p-value and points to the dominant subtype, i.e., the most significantly activated subtype. The accumulative distance to the dominant subtype (ADDS) was defined as:

$$ADDS = \sum_{i=1}^{N-1} (R_i - R_0)$$

Similarly, the accumulative distance between non-dominant subtypes (ADNS) as:

$$ADNS = \sum_{j>i>0} (R_j - R_i)$$

Obviously, the ADDS and ADNS are positive and negative correlated with single activation, respectively. Hence, we defined the simplicity score by combing ADDS and ADNS together and corrected with a constant $\frac{(R_{N-1}-R_0)}{N-1}$ as follows:

$$Simplicity\ score = [ADDS - ADNS] \times \frac{(R_{N-1} - R_0)}{N-1}$$

In a three subtypes case, the simplicity score of subtype activation can be simplified as:

647 $$Simplicity\ score = (\text{median-min}) \times (\text{max-min}).$$

648

649 **Tumor purity assessment**

650 The ESTIMATE package was used to evaluate tumor purity on the basis of the
651 expression level of marker genes in stromal and immune cells [12], where the fraction of
652 stromal cells and immune cells in each sample were represented by stromal score and
653 immune score respectively, and the mixed fraction of both stromal and immune cells
654 was represented by estimate scores. The ABSOLUTE package was used to confirm the
655 tumor purity on the basis of chromosome copy number and allele fraction ratios on
656 samples for which single nucleotide polymorphism array data were available [13].

657

658 **Sample collection and Neurosphere Cultures**

659 After obtaining approval from the institutional review board of The University of Texas
660 M.D. Anderson Cancer Center, glioblastoma tumor tissues were collected and named in
661 the order that they were acquired. Each tissue was enzymatically and mechanically
662 dissociated into single cells and grown in DMEM/F12 media supplemented with B27
663 (Invitrogen), EGF (20 ng/ml), and bFGF (20 ng/ml), resulting in neurosphere growth. All
664 cell lines were tested to exclude the presence of Mycoplasma infection. To minimize
665 any batch effect the downstream molecular analyses were performed on identical cell
666 culture batches. Total RNA from formalin fixed, paraffin embedded tumor tissues and
667 matching neurospheres was prepared using the Masterpure complete DNA and RNA
668 isolation kit (Epicenter) after proteinase K digestion per to the instructions from the
669 manufacturer. Paired-end Illumina HiSeq sequencing assays were performed resulting

670 in a medium number of 50 million 75bp paired end reads per sample. We employed the
671 PRADA pipeline to process the RNA sequencing data [10]. In short, Burroughs-Wheeler
672 alignment, Samtools, and Genome Analysis Toolkit were used to map short reads to the
673 human genome (hg19) and transcriptome (Ensembl 64) and RPKM gene expression
674 values were generated for each of the 135,994 transcripts of 21,165 protein coding
675 genes in Ensembl database.
676

677 **Western blotting**

678 Lysates were prepared from fresh frozen sections using RPPA lysis buffer (1% Triton X-
679 100 50mM HEPES pH 7.4, 150mM NaCl, 1.5mM MgCl2, 1mM EGTA, 100mM NaF,
680 10mM Na pyrophosphate, 1mM Na3VO4, 10% glycerol, plus protease and phosphatase
681 inhibitors cocktails from Roche Applied Science #05056489001 and 04906837001), with
682 sonication and clearing by centrifugation at 10,000g. Protein concentration was
683 measured using the BCA kit (Thermo Scientific - Pierce #23225). SDS-PAGE and
684 western blotting was performed using Midi gel system (Life Technologies - #WR0100)
685 and NuPage-Novex 4-12% Bis-Tris Midi (20-well) Protein Gels (Life Technologies -
686 #WG1402) using the following antibodies: ITGAM (CD11B) (Sigma Aldrich –
687 #HPA002274), IBA1 (AIF1) (Sigma Aldrich – #HPA049234), GFAP (Cell Signalling –
688 #3670), YKL40 (CHI3L1) YKL40 (CHI3L1, Santa Cruz Biotechnology - #sc-30465) a-
689 actinin (Sigma Aldrich A5044) and Tubulin (Sigma Aldrich T9026).
690

691 **Immunohistochemistry**

692 Formalin-fixed, paraffin-embedded tissue sections (4 μm thick) were collected on
693 Superfrost plus slides. Briefly, tissue sections were deparaffinized with xylene and
694 ethanol and re-hydrated with 95, 70 and 50% ethanol. Sections were antigen unmasked
695 using citrate buffer (Vector Labs #H-3300) and heating. Peroxidase block was
696 conducted with 3% H2O2 and blocking was with 5% goat serum (Vector Labs #S-1000).
697 Primary rabbit polyconal antibody against IBA1 (AIF1)(WAKO #016-20001) at 1:400
698 was used overnight. Secondary antibody was done using with the Rabbit-on-Rodent
699 HRP-Polymer (Biocare #RMR622L) for 1 hr at room temperature. The slides were
700 developed with Nova-red (Vector Labs #SK-4800) and counterstained with
701 haematoxylin, mounted and scanned with Pannoramic 250 slide scanner (Caliper Life
702 Sciences).
703

704 **Analysis of the Human Protein Atlas database**

705 We systematically queried the human protein atlas database (www.proteinatlas.org,
706 Version 12) for immunohistochemistry of specific markers of macrophage/microglial
707 lineage[14]. Typically, each antibody was used on a set of tissue microarrays covering
708 healthy controls and set of cancers of diverse sites, typically 10-15 individual cases for
709 each cancer. Only antibodies with high validation scores (concordant RNA seq/IHC
710 staining, western blot bands and antigen profile) were retained. ITGAM (Antibody HPA
711 002274), IBA1 (Antibody HPA 049234) and CD163 (Antibody HPA 002432) met these
712 criteria, as well as control protein vinculin (Antibody CAB002453).

713 ## Supplementary References

714 1 Phillips, H. S. *et al.* Molecular subclasses of high-grade glioma predict prognosis, delineate a
715 pattern of disease progression, and resemble stages in neurogenesis. *Cancer cell* **9**, 157-173,
716 doi:10.1016/j.ccr.2006.02.019 (2006).
717 2 Joo, K. M. *et al.* Patient-specific orthotopic glioblastoma xenograft models recapitulate the
718 histopathology and biology of human glioblastomas in situ. *Cell Rep* **3**, 260-273,
719 doi:10.1016/j.celrep.2012.12.013 (2013).
720 3 Kwon, S. M. *et al.* Recurrent Glioblastomas Reveal Molecular Subtypes Associated with
721 Mechanistic Implications of Drug-Resistance. *PLoS One* **10**, e0140528,
722 doi:10.1371/journal.pone.0140528 (2015).
723 4 Kim, H. *et al.* Whole-genome and multisector exome sequencing of primary and post-
724 treatment glioblastoma reveals patterns of tumor evolution. *Genome research* **25**, 316-327,
725 doi:10.1101/gr.180612.114 (2015).
726 5 Brennan, C. W. *et al.* The somatic genomic landscape of glioblastoma. *Cell* **155**, 462-477,
727 doi:10.1016/j.cell.2013.09.034 (2013).
728 6 Verhaak, R. G. *et al.* Integrated genomic analysis identifies clinically relevant subtypes of
729 glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer cell*
730 **17**, 98-110, doi:10.1016/j.ccr.2009.12.020 (2010).
731 7 Dai, M. *et al.* Evolving gene/transcript definitions significantly alter the interpretation of
732 GeneChip data. *Nucleic Acids Res* **33**, e175, doi:10.1093/nar/gni179 (2005).
733 8 Sandberg, R. & Larsson, O. Improved precision and accuracy for microarrays using updated
734 probe set definitions. *BMC Bioinformatics* **8**, 48, doi:10.1186/1471-2105-8-48 (2007).
735 9 Bengtsson, H., Ray, A., Spellman, P. & Speed, T. P. A single-sample method for normalizing
736 and combining full-resolution copy numbers from multiple platforms, labs and analysis
737 methods. *Bioinformatics* **25**, 861-867, doi:10.1093/bioinformatics/btp074 (2009).
738 10 Torres-Garcia, W. *et al.* PRADA: pipeline for RNA sequencing data analysis. *Bioinformatics*
739 **30**, 2224-2226, doi:10.1093/bioinformatics/btu169 (2014).
740 11 Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers
741 require TBK1. *Nature* **462**, 108-112, doi:10.1038/nature08460 (2009).
742 12 Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from
743 expression data. *Nature communications* **4**, 2612, doi:10.1038/ncomms3612 (2013).
744 13 Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer.
745 *Nature biotechnology* **30**, 413-421, doi:10.1038/nbt.2203 (2012).
746 14 Murray, P. J. & Wynn, T. A. Protective and pathogenic functions of macrophage subsets. *Nat*
747 *Rev Immunol* **11**, 723-737, doi:10.1038/nri3073 (2011).