BIOMA: Berkala Ilmiah Biologi

Available online: https://ejournal.undip.ac.id/index.php/bioma/index

Assessing glioblastoma cell population stability through bootstrap resampling of scRNA-seq data

Andi Rosilala[1], Rohmatul Fajriyah[1*], Linda Erlina[2], Nabilah Dwi Septiani[1]

[1]Bioinformatics Research Group, Master Program in Statistics, Universitas Islam Indonesia, Yogyakarta 55584, Indonesia

[2]Department of Medical Chemistry, Universitas Indonesia, Jakarta, Indonesia

A B S T R A C T

Glioblastoma (GBM) exhibits extreme cellular heterogeneity, comprising diverse tumor cell states and non-malignant microenvironment populations. Single-cell RNA-sequencing (scRNA-seq) enables resolution of this complexity, yet a critical unmet challenge persists: cluster reproducibility in GBM scRNA-seq studies is rarely validated, and standard clustering algorithms may generate artifactual partitions indistinguishable from biologically meaningful populations. To address this gap, we propose a cluster-wise bootstrap stability framework integrated with explicit tumor–microenvironment separation, an approach not previously applied systematically to GBM scRNA-seq data. We analyzed a public dataset (GSE131928; 10 tumors, 15,072 cells after quality control) and identified 14 clusters annotated via marker gene validation. Bootstrap resampling (100 iterations) with Jaccard coefficient quantification revealed that non-malignant populations (microglia/macrophage, oligodendrocytes) exhibited the highest stability (Jaccard >0.97). Among tumor states, MES-AC transitional and MES-like clusters were most stable (Jaccard 0.99 and 0.82), whereas NPC-like, AC-like, and rare populations showed low stability (Jaccard <0.5). Stability correlated positively with marker gene specificity, within-cluster homogeneity, and silhouette scores. These results demonstrate that cluster-wise bootstrap assessment provides a practical, quantitative criterion for distinguishing robust from unreliable cell populations, supporting more confident biological interpretation and therapeutic target prioritization in GBM.

Keywords: Bootstrap resampling; cluster stability; glioblastoma; Jaccard coefficient; single-cell RNA-seq

1. INTRODUCTION

Glioblastoma (GBM) is the most aggressive primary brain tumor in adults, with a median overall survival of only 14.6 months and a 5-year survival rate of 7.2% (Ostrom et al., 2022). Despite multimodal therapy including surgical resection, radiotherapy, and chemotherapy with temozolomide, patient prognosis remains poor. A major cause of treatment failure is high tumor heterogeneity, both at the inter-tumor and intra-tumor levels.

Single-cell RNA-sequencing (scRNA-seq) technology has revolutionized our understanding of tumor heterogeneity at single-cell resolution (Tirosh et al., 2016). Unlike bulk RNA-seq, which provides average gene expression profiles from millions of cells, scRNA-seq enables characterization of transcripts from individual cells, revealing rare cell populations, distinct cellular states, and cellular transition dynamics undetectable by conventional methods.

Several scRNA-seq studies on GBM have identified extraordinary heterogeneity. Patel (Patel et al., 2014) revealed that GBM cells with different TCGA molecular subtypes (The Cancer Genome Atlas Research Network, 2008; Verhaak et al., 2010) can coexist within individual tumors. Neftel (Neftel et al., 2019) constructed a cellular state model showing that each GBM tumor contains four main cellular states — neural progenitor-like (NPC-like), oligodendrocyte progenitor-like (OPC-like), astrocyte-like (AC-like), and mesenchymal-like (MES-like) — that can interconvert. Critically, GBM tumors also contain substantial non-malignant components including microglia, macrophages, T cells, oligodendrocytes, and endothelial cells that together constitute the tumor microenvironment (TME) (Darmanis et al., 2017), whose transcriptional identity must be distinguished from tumor cell states prior to biological interpretation.

However, scRNA-seq analysis faces significant methodological challenges, particularly in clustering and cluster validation. Clustering methods such as Louvain (Blondel et al., 2008) or Leiden (Traag et al., 2019) tend to produce clusters even in relatively homogeneous data, raising a critical question: do identified clusters represent true biological cell populations or artifacts of the analysis method? This question is particularly important in mixed tumor-TME datasets, where non-malignant cell types may confound cellular state annotation. Global quality

***Corresponding author:** rfajriyah@uii.ac.id

Received: 29 March 2026, **Accepted:** 24 June 2026

Cite this article as: Rosilala et al, (2026) *BIOMA: Berkala Ilmiah Biologi*, (28)1: 114-126

metrics such as the silhouette score (Rousseeuw, 1987) evaluate overall clustering quality but do not assess the reproducibility of individual clusters.

Hennig (Hennig, 2007) developed a cluster-wise approach to evaluate individual cluster stability using bootstrap resampling, based on the principle that valid and biologically meaningful clusters should remain consistent under non-essential data perturbations. The Jaccard coefficient (Jaccard, 1901) quantifies the overlap between original and bootstrapped clusters, providing a continuous stability measure per cluster rather than a global quality metric. This approach has been applied to microarray clustering (Dudoit and Fridlyand, 2003) and consensus clustering (Monti et al., 2003), but its application to scRNA-seq data with explicit tumor-TME separation has not been systematically explored in GBM.

Despite the growing number of GBM scRNA-seq studies, cluster validation remains an overlooked step. Most published analyses rely solely on global quality metrics such as silhouette scores or UMAP visual inspection, neither of which evaluates the reproducibility of individual clusters. Consensus clustering approaches have been applied to bulk microarray data, but do not account for the unique challenges of scRNA-seq, including dropout noise, ambient RNA contamination, and the co-occurrence of tumor and non-tumor populations within the same sample. Crucially, no existing GBM scRNA-seq study has systematically separated tumor cell states from microenvironment components before assessing cluster stability, creating a risk that stability estimates are confounded by transcriptionally distinct non-malignant populations.

To address this gap, the present study introduces a framework that combines cluster-wise bootstrap stability assessment with explicit tumor–TME annotation and separation. The novelty of this approach is twofold: (i) it applies Hennig's bootstrap stability method to scRNA-seq data from a complex tumor–microenvironment mixture, providing per-cluster reproducibility scores rather than a single global metric; and (ii) it performs rigorous marker gene-based cell type annotation to separate tumor cell states from microenvironment populations before evaluating stability, thereby ensuring that stability estimates reflect true biological coherence rather than transcriptional distinctness between cell lineages.

This study aims to evaluate cluster-wise bootstrap stability in a public GBM scRNA-seq dataset (GSE131928), with marker gene-validated cell type annotation that explicitly separates tumor from microenvironment populations, and to characterize the biological and methodological determinants of stable versus unstable clusters.

2. Material and Methods

2.1 Dataset and data source

We used a publicly available GBM scRNA-seq dataset from Gene Expression Omnibus (GEO), accession number GSE131928, published by Neftel (Neftel et al., 2019). The 10x Genomics platform subset was analyzed, comprising 10 tumor samples from GBM patients. Raw count matrices were downloaded directly from GEO and processed as described below. All analyses were performed using R (version 4.3) in a local computing environment.

2.2 Data preprocessing and quality control

Raw count matrices were processed using Seurat (Stuart et al., 2019). Quality control excluded cells with detected genes <200 or >6,000, UMI counts <500, and mitochondrial gene proportion >20%. Genes detected in fewer than three cells were excluded. After filtering, 15,072 cells were retained across 10 samples. Data were normalized using log-normalization with a scaling factor of 10,000. Highly variable genes (HVGs, n = 2,000) were identified using variance stabilizing transformation (Hafemeister and Satija, 2019). Data were scaled via z-score transformation for each gene, regressing out variation associated with UMI count per cell and mitochondrial gene proportion. No explicit batch-effect correction (e.g., Harmony or CCA) was applied across the 10 tumor samples, as the bootstrap stability assessment itself tests the robustness of clustering to data perturbation.

Moreover, the biological question focuses on intra-tumor heterogeneity within a single-platform dataset, reducing the need for cross-batch integration. This design choice is a limitation discussed further in the Results section.

2.3 Dimensionality reduction and clustering

Principal Component Analysis (PCA) was performed on HVGs, with the top 20 principal components (PCs) selected based on elbow plot analysis. Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018) was used for two-dimensional visualization based on the top 20 PCs. Graph-based clustering used the Louvain algorithm (Blondel et al., 2008) on a k-nearest neighbor (KNN) graph with k = 20. Resolution parameters 0.1–0.3 were systematically evaluated; resolution 0.1 was selected as optimal based on overall bootstrap stability across all clusters, yielding 14 clusters.

2.4 Marker-Based cell type annotation

Cluster marker genes were identified using the Wilcoxon rank-sum test implemented in presto::wilcoxauc (Korsunsky et al., 2019). Significance thresholds were: Benjamini-Hochberg adjusted $p < 0.05$, log2 fold change > 0.5, proportion of expressing cells within the cluster >25%, and AUC >0.6. Cell type annotation was assigned based on top marker genes cross-referenced against canonical markers from the literature (**Table 1**). This approach explicitly distinguished tumor cell states from non-malignant TME cell types and contaminating populations.

2.5 Tumor versus non-tumor cell separation

Following annotation, cells were classified as tumor or non-tumor. Tumor clusters comprised: MES-AC transitional (Cluster 1), NPC-like mature (Cluster 2), MES-like (Cluster 3), NPC-like (Cluster 4), AC-like (Cluster 6), and Cycling tumor (Cluster 7). Non-tumor clusters comprised microglia/macrophage (Clusters 0, 5, 12), T cells (Cluster 10), oligodendrocytes (Cluster 8), endothelial/astrocytes (Cluster 9), pericytes/stroma (Cluster 13), and erythrocytes (Cluster 11, ambient RNA contamination). Neftel cellular state annotation (Neftel et al., 2019) was subsequently applied only to tumor cells using module score-based assignment.

2.6 Cluster-wise stability assessment

Cluster-wise stability followed the method of Hennig (Hennig, 2007). For each bootstrap iteration ($B = 100$), the following steps were performed:

1. Bootstrap sample: n cells were sampled with replacement from the full dataset; duplicate cells arising from the with-replacement procedure were removed prior to re-clustering, retaining only unique cells. This step is necessary because graph-based clustering algorithms (Louvain/Leiden) operate on nearest-neighbor graphs in which cell identity must be unique; retaining duplicates would artificially inflate the density of sampled cells and distort neighborhood structure. The resulting bootstrap sample thus contains approximately 63.2% of the original cells (the expected unique fraction under Poisson sampling), effectively testing whether clusters remain identifiable when a random subset of cells is held out.
2. Re-clustering: Louvain clustering was performed on the bootstrap cell set using identical parameters ($k = 20$, resolution = 0.1, top 20 PCs).
3. Cluster matching: for each original cluster i, the most similar bootstrap cluster was identified as arg $max_j\ J(C_i, C_j^b)$.
4. Aggregation: stability of cluster i was calculated as the mean Jaccard coefficient across all B iterations.

Formally, cluster stability is defined as:

$$Stability(i) = \frac{1}{B}\sum_{b=1}^{B} max_j\ J(C_i, C_j^b) \quad (1)$$

Where C_i is cluster i in the original clustering, C_j^b is the cluster j in bootstrap clustering b, and J is the Jaccard coefficient:

$$J(A,B) = \frac{|A\ \cap\ B|}{|A\ \cup\ B|} \quad (2)$$

Stability categories followed Hennig (Hennig, 2007): highly stable (> 0.85), stable (0.75–0.85), moderately stable (0.60–0.75), and unstable (< 0.60). Dissolution frequency was defined as the proportion of bootstrap iterations in which no match (Jaccard > 0) was found for a given original cluster.

2.7 Biological characterization

Marker gene specificity was calculated as the ratio of mean expression within versus outside each cluster. Within-cluster transcriptomic homogeneity was measured as the mean pairwise Pearson correlation among all cells within a cluster, computed on the top 2,000 HVGs. Silhouette scores (Rousseeuw, 1987) were calculated in PCA space (top 20 PCs). Gene Ontology (GO) biological process enrichment and GSEA with Hallmark gene sets (Liberzon et al., 2015) were performed using clusterProfiler (Yu et al., 2012).

2.8 Neftel cellular state assignment

Neftel cellular state scores (MES1, MES2, AC, OPC, NPC1, NPC2) were calculated on tumor cells only using Seurat::AddModuleScore with gene signatures from Neftel (Neftel et al., 2019). Each cell was assigned the state with the highest module score. Cells for which the margin between the highest and second-highest score was <0.1 were classified as 'Mixed'. Per-sample state proportions were calculated for samples with ≥50 tumor cells.

2.9 Resampling method comparison

To assess methodological robustness, stability was additionally evaluated using: (i) subsampling (80% of cells sampled without replacement, 100 iterations), and (ii) Gaussian noise injection (additive Gaussian noise with SD = 0.1 x the per-gene SD of PCA embeddings, 100 iterations). Spearman rank correlations between bootstrap stability and each alternative method were computed to quantify cross-method concordance.

3. Results and Discussion

3.1 Dataset characteristics and cell type annotation

After quality control, 15,072 cells were retained from 10 GBM samples. Louvain clustering at resolution 0.1 identified 14 clusters. All 100 bootstrap iterations completed without error (total runtime: 2.4 minutes on a standard workstation). Marker-based annotation resolved 14 distinct cell populations spanning both tumor and TME compartments (Table 1; Figure 1). The dataset comprised approximately 53% tumor cells (n ≈ 8,585) and 47% non-tumor cells. Among tumor populations, MES-AC transitional was the largest (n = 2,176), followed by NPC-like mature (n = 1,881), MES-like (n = 1,590), and NPC-like (n = 1,548). Figure 2 shows the binary tumor versus TME separation, supporting a clear transcriptional distinction between malignant and non-malignant compartments.

Table 1. Marker-based cell type annotation of the 14 clusters identified by Louvain clustering at resolution 0.1. Key markers represent the top differentially expressed genes per cluster (AUC > 0.6, adjusted p < 0.05). Cell type classification is based on marker gene cross-referencing with published GBM literature.

Cluster	Annotation	Key Markers	Type
0	Microglia/Macrophage	*TYROBP, C1QC, CD74, LAPTM5*	TME
1	MES-AC Transitional	*VIM, SOX9, FABP7, NES*	Tumor
2	NPC-like (mature)	*MAP1B, TUBA1A, TUBB2B, UCHL1*	Tumor
3	MES-like	*CHI3L1, TIMP1, CLU*	Tumor
4	NPC-like	*SOX4, STMN1, NOVA1*	Tumor
5	Macrophage (complement)	*C1QA, C1QB, TYROBP, CD74*	TME
6	AC-like	*TTYH1, CST3, CKB, EGFR*	Tumor
7	Cycling tumor	*CDK4, MDM2, NUP107*	Tumor
8	Oligodendrocyte	*PLP1, MBP, CNP, MAG*	Normal
9	Endothelial/Astrocyte	*AQP1, SPARCL1, GJA1*	TME
10	T cell	*CD3D, CD52, CXCR4, IL32*	TME
11	Erythrocyte	*HBA1, HBA2, HBB*	Contam.
12	Macrophage (LAM)	*CTSB, CTSD, FTL, FTH1*	TME
13	Pericyte/Stromal	*SPARC, SFRP4, ANXA1, PSAP*	TME

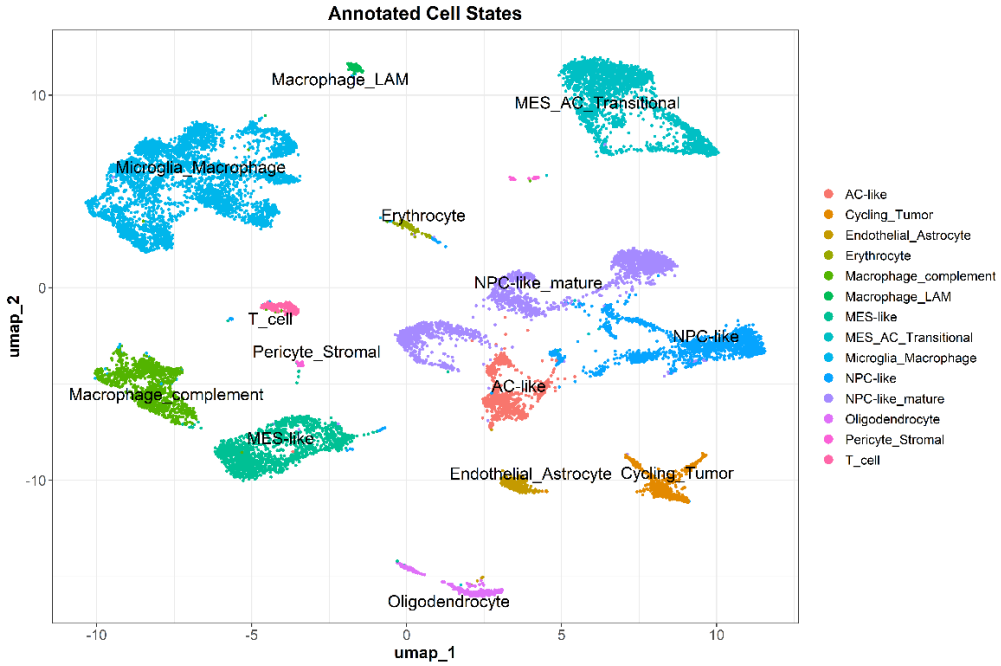

Figure 1. UMAP visualization of marker gene-validated cell type annotation across 14 clusters.
UMAP projection of all 15,072 cells colored by marker gene-validated cell type annotation. The 14 clusters are labeled and color-coded by cell type. Non-tumor TME populations are spatially separated from tumor cell state clusters in reduced-dimensional space.

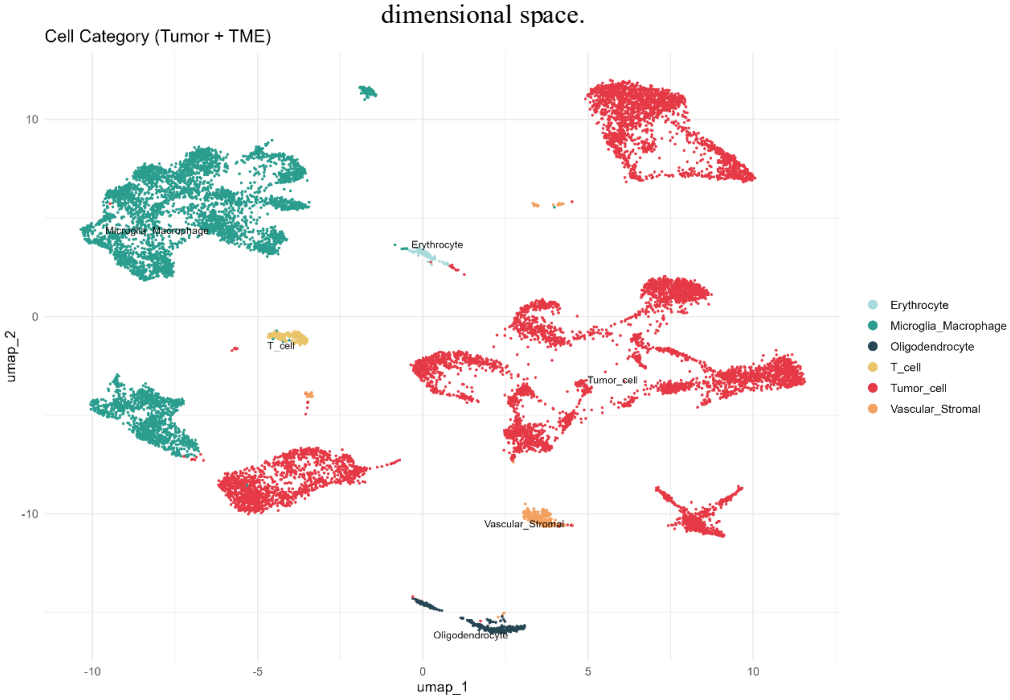

Figure 2. Tumor versus non-tumor cell classification projected onto UMAP space.
UMAP projection colored by broad cell category: tumor cells (red) versus non-tumor microenvironment components (blue/grey). Approximately 53% of cells were classified as tumor (n ≈ 8,585) and 47% as non-tumor.

3.2 Cluster stability results

Bootstrap stability assessment revealed substantial heterogeneity across the 14 clusters (Table 2; Figure 3). Of 14 clusters, 4 were highly stable (>0.85), 1 stable (0.75-0.85), 3 moderately stable (0.60-0.75), and 6 unstable (<0.60). Among non-malignant populations, microglia/macrophage (Cluster 0: mean Jaccard = 0.974), oligodendrocytes (Cluster 8: 0.983), and MES-AC transitional (Cluster 1: 0.989) showed the highest stability, consistent with their transcriptionally well-defined identities. Macrophage complement (Cluster 5: 0.891) also reached the highly stable category, though with a wider confidence interval (SD = 0.134), reflecting moderate inter-sample variation in complement gene expression. Among confirmed tumor cell clusters, MES-like (Cluster

3) was the most stable tumor-intrinsic state (mean Jaccard = 0.822; Stable category). Cycling tumor cells (Cluster 7: 0.728) were moderately stable, consistent with the transcriptionally coherent cell-cycle program defining this cluster. In contrast, NPC-like (Cluster 4: 0.476), NPC-like mature (Cluster 2: 0.383), and AC-like (Cluster 6: 0.292) were unstable, with dissolution frequencies of 54%, 98%, and 91%, respectively. Erythrocytes (Cluster 11: 0.042) and pericytes/stroma (Cluster 13: 0.034) showed near-zero stability, suggesting potential contamination or low-confidence clustering artifacts.

Table 2. Cluster-wise stability metrics from 100 bootstrap iterations. Clusters ordered by mean Jaccard coefficient. Dissolution frequency is the proportion of iterations in which no matching cluster was found. Stability categories follow Hennig (2007).

Cl.	Cell Type	n	Jaccard	Diss.%	Category
1	MES-AC Transitional	2,176	0.989	0	Highly Stable
8	Oligodendrocyte	420	0.983	0	Highly Stable
0	Microglia/Macrophage	3,997	0.974	0	Highly Stable
5	Macrophage (compl.)	1,228	0.891	6	Highly Stable
3	MES-like	1,590	0.822	14	Stable
7	Cycling Tumor	641	0.728	0	Mod. Stable
10	T cell	207	0.699	34	Mod. Stable
12	Macrophage (LAM)	125	0.602	40	Mod. Stable
4	NPC-like	1,548	0.476	54	Unstable
2	NPC-like (mature)	1,881	0.383	98	Unstable
6	AC-like	749	0.292	91	Unstable
9	Endothelial/Astrocyte	252	0.276	100	Unstable
11	Erythrocyte	148	0.042	100	Unstable
13	Pericyte/Stromal	110	0.034	99	Unstable

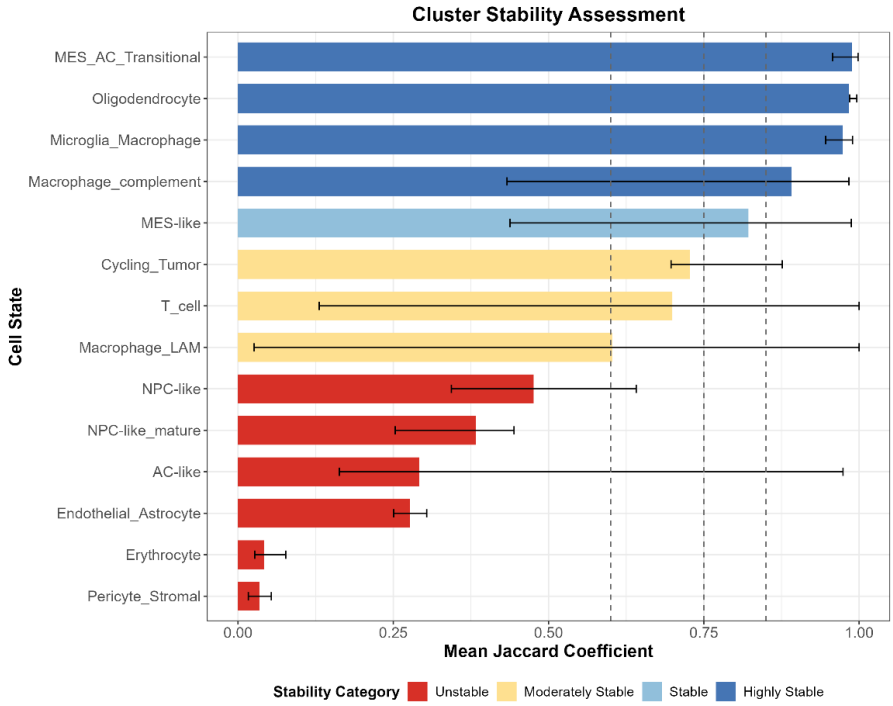

Figure 3. Mean Jaccard coefficient with 95% CI for each cluster across 100 bootstrap iterations. Cluster stability barplot showing mean Jaccard coefficient ± 95% CI for each of the 14 clusters across 100 bootstrap iterations, ordered from highest to lowest stability. Bars are colored by stability category: Highly Stable (>0.85), Stable (0.75-0.85), Moderately Stable (0.60-0.75), and Unstable (<0.60). Dashed horizontal lines indicate category thresholds.

The ridgeplot (Figure 4) illustrates the full distribution of Jaccard coefficients across 100 bootstrap iterations per cluster. Notably, several clusters — including MES-like (Cluster 3), T cell (Cluster 10), and Macrophage LAM (Cluster 12) — exhibited bimodal distributions with high standard deviations (SD = 0.22, 0.40, and 0.47, respectively). This pattern is consistent with populations near the minimum cell-number detection threshold at the current clustering resolution, rather than reflecting algorithmic instability per se.

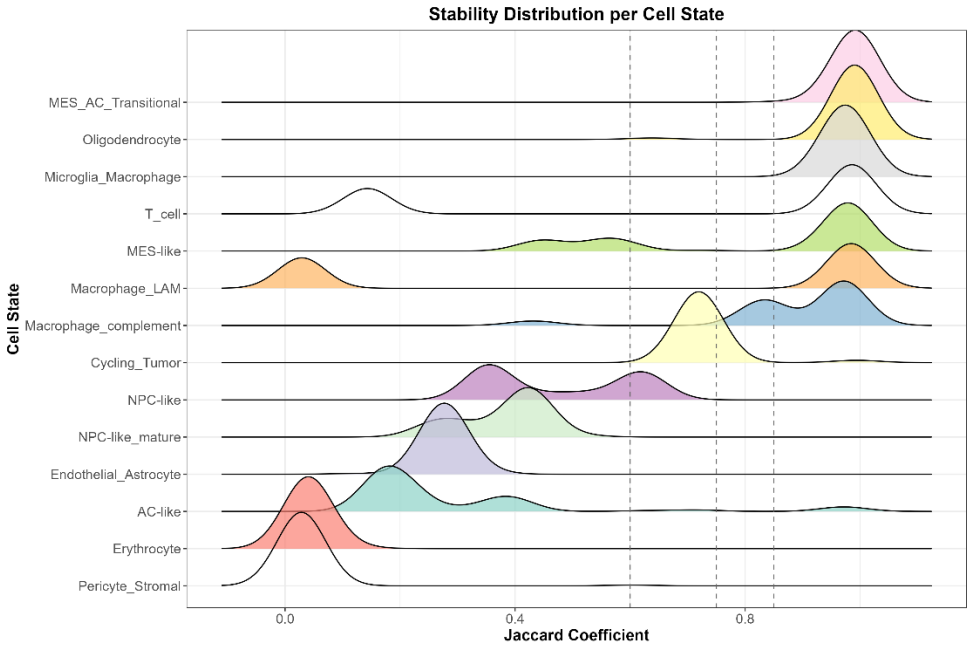

Figure 4. Distribution of per-iteration Jaccard coefficients for each cluster (ridgeplot).
Ridgeplot of Jaccard coefficient distributions across 100 bootstrap iterations for each cluster, ordered by median stability. Bimodal distributions (e.g., MES-like, T cell, Macrophage LAM) reflect clusters that occasionally dissolve completely in bootstrap resampling.

3.3 Biological characteristics and stability

Positive correlations were observed between cluster stability and biological characteristics. Marker gene specificity (ratio of within-cluster to outside-cluster expression) correlated positively with Jaccard stability (Spearman r = 0.86, p = 0.007), with MES-AC transitional and MES-like showing the highest specificity scores. Within-cluster transcriptomic homogeneity (mean pairwise Pearson correlation) also correlated with stability (r = 0.79, p = 0.02), as did silhouette scores (r = 0.84, p = 0.009). Non-malignant and MES-state clusters consistently occupied the highest values across all three biological metrics. Pathway enrichment analysis confirmed the biological coherence of stable clusters. MES-like cells were enriched for inflammatory response, TNF-alpha signaling via NF-kappaB, and epithelial-mesenchymal transition — consistent with established MES-state biology (Neftel et al., 2019). Cycling tumor cells showed enrichment for G2/M checkpoint and E2F targets. NPC-like clusters, despite their unstable Jaccard values, showed enrichment for neurogenesis-related biological processes, suggesting that a genuine biological signal exists but is distributed across a transcriptionally heterogeneous cell population.

3.4 Neftel state distribution in tumor cells

Among the 8,585 tumor cells, module score-based assignment revealed the distribution of MES1/MES2, AC, OPC, NPC1, NPC2, and Mixed states. Samples 114 (n = 39) and 126 (n = 19) were excluded from per-sample proportion analysis due to insufficient tumor cell count (<50 cells). Figure 5 shows the UMAP of Neftel state assignments for tumor cells only. Figure 6 shows state proportions across the 8 qualifying samples, revealing inter-sample heterogeneity in the relative abundance of MES, AC, OPC, and NPC states consistent with the pan-GBM model of Neftel (Neftel et al., 2019).

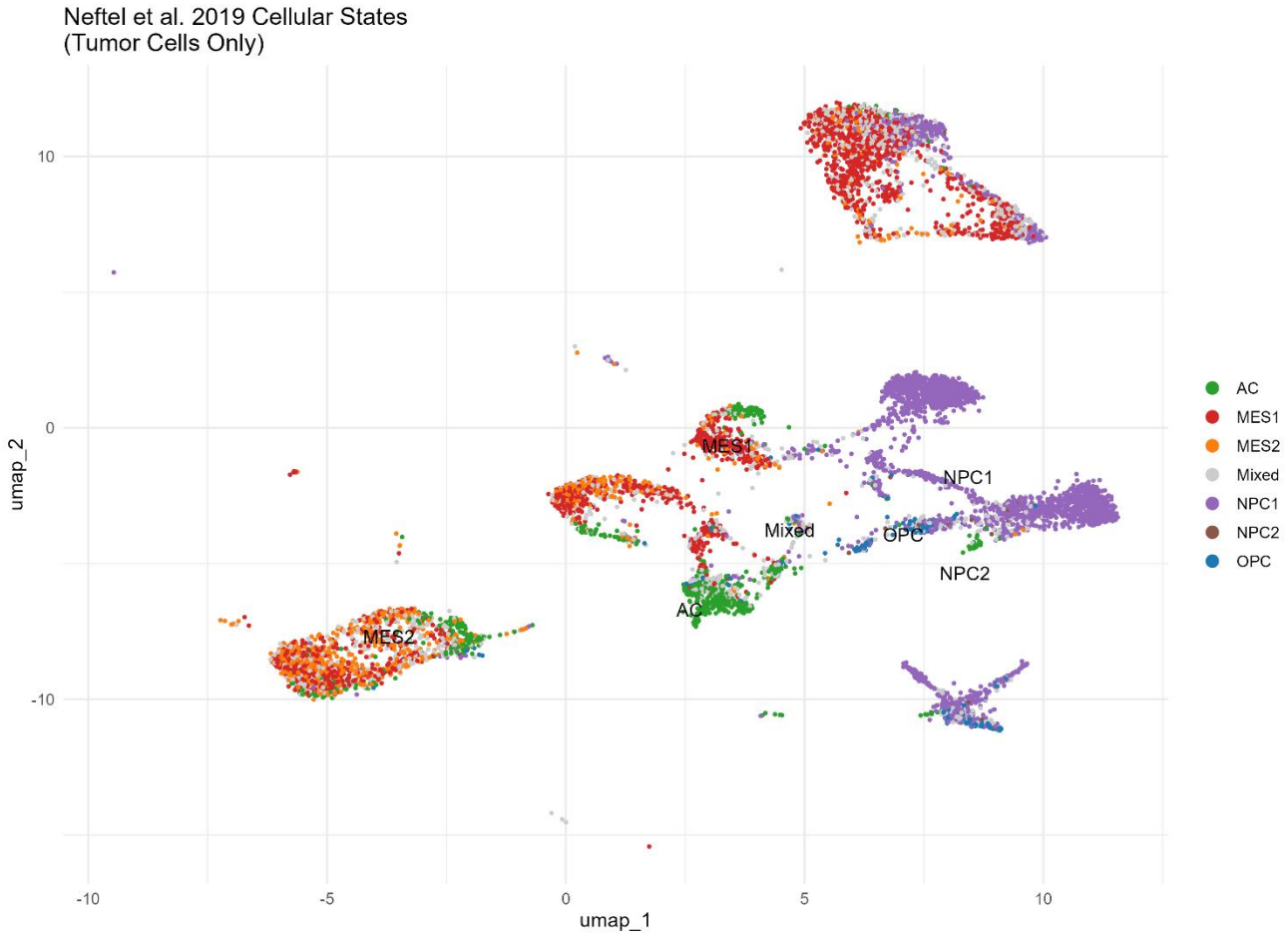

Figure 5. Neftel cellular state assignment for tumor cells projected onto UMAP space.
UMAP projection of tumor cells only (n ≈ 8,585), colored using the cellular state model of Neftel (Neftel et al., 2019) (MES1, MES2, AC, OPC, NPC1, NPC2, Mixed). Cells with a margin between highest and second-highest score <0.1 were classified as Mixed.

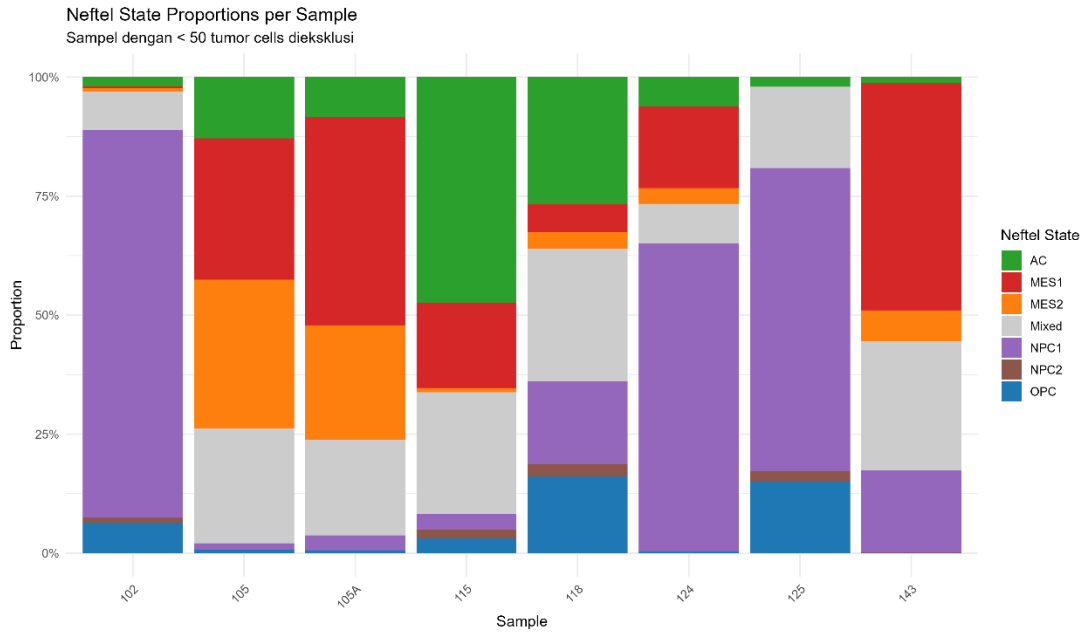

Figure 6. Proportional composition of Neftel cellular states across qualifying tumor samples.
Proportional stacked bar chart of Neftel cellular state composition per tumor sample. Eight of 10 samples with ≥50 tumor cells are shown; Samples 114 (n = 39) and 126 (n = 19) were excluded due to insufficient cell count.

3.5 Comparison across resampling methods

To assess the robustness of bootstrap-derived stability estimates, we compared results across three perturbation strategies (Table 3; Figure 7). Spearman rank correlation between bootstrap and subsampling stability was rho = 0.982 ($p < 0.001$), and between bootstrap and noise injection rho = 0.965 ($p < 0.001$), indicating that the cluster stability ranking is highly consistent regardless of perturbation type. Critically, the categorical classification of all 14 clusters was identical across all three methods, validating the robustness of our conclusions.

In absolute terms, subsampling generally yielded slightly higher estimates than bootstrap (mean difference +0.023 across clusters). The most notable discrepancy occurred in the Pericyte/Stromal cluster (Cluster 13), where noise injection yielded a markedly higher value (0.292) compared to bootstrap (0.034) and subsampling (0.102). This reflects the sensitivity of small, transcriptionally heterogeneous clusters to perturbation type: noise injection perturbs gene expression magnitudes without removing cells, paradoxically preserving a cluster that would otherwise dissolve under cell removal. As this cluster remained near-zero or unstable across all methods, the biological conclusion is unaffected.

For MES-like (Cluster 3), noise injection produced a lower estimate (0.706) compared to bootstrap (0.822) and subsampling (0.884), suggesting the MES-like transcriptional program is more sensitive to expression-level perturbation than to cell composition changes — possibly because its defining markers (CHI3L1, TIMP1) exhibit more finely graded expression than canonical immune or myelination gene programs.

Table 3. Mean Jaccard stability estimates across three resampling methods for all 14 clusters (100 iterations each). Clusters ordered by bootstrap stability. Spearman rank correlations (rho) between bootstrap and each alternative method are shown in the final row.

Cl.	Cell Type	Bootstrap	Subsampling	Noise Inj.
1	MES-AC Transitional	0.989	0.994	0.995
8	Oligodendrocyte	0.983	0.991	0.992
0	Microglia/Macrophage	0.974	0.970	0.969
5	Macrophage (compl.)	0.891	0.872	0.839
3	MES-like	0.822	0.884	0.706
7	Cycling Tumor	0.728	0.720	0.719
10	T cell	0.699	0.649	0.741
12	Macrophage (LAM)	0.602	0.688	0.582
4	NPC-like	0.476	0.483	0.541
2	NPC-like (mature)	0.383	0.418	0.363
6	AC-like	0.292	0.419	0.374
9	Endothelial/Astrocyte	0.276	0.276	0.276
11	Erythrocyte	0.042	0.041	0.046
13	Pericyte/Stromal	0.034	0.102	0.292
14	*Spearman rho vs. bootstrap*	—	0.982	0.965

Figure 7. Comparison of stability estimates across bootstrap, subsampling, and noise injection methods. Grouped barplot comparing mean Jaccard stability estimates across three resampling methods (bootstrap, subsampling, noise injection) for all 14 clusters, ordered by bootstrap stability. The near-identical ordering of clusters across methods (Spearman rho = 0.982 for bootstrap vs. subsampling; rho = 0.965 for bootstrap vs. noise injection) supports methodological robustness.

3.6 Stability of non-malignant versus tumor cell populations

The high stability of non-malignant cell types — microglia/macrophage (Jaccard = 0.974) and oligodendrocytes (0.983) — reflects the well-defined, lineage-committed transcriptional identity of these cell types. Their correct identification as non-tumor cells prior to stability analysis is critical: without explicit annotation and separation, macrophage clusters sharing mesenchymal gene signatures (e.g., VIM, CD44) could be misclassified as MES-like tumor states, inflating the apparent stability and size of the MES compartment. The MES-AC transitional cluster showed the highest stability among all cell types (Jaccard = 0.989). This cluster is characterized by co-expression of MES markers (VIM) and AC markers (SOX9, FABP7), consistent with documented intermediate states along the MES-AC axis in GBM (Neftel et al., 2019). Its near-perfect bootstrap stability indicates that this transitional phenotype is a genuinely committed and reproducible cellular state. The canonical MES-like cluster (CHI3L1, TIMP1) also showed robust stability (0.822), consistent with the established view that the mesenchymal state is the most committed and therapy-resistant GBM phenotype (Neftel et al., 2019).

3.7 Biological interpretation of unstable tumor clusters

The low stability of NPC-like (0.476) and NPC-like mature (0.383) clusters is biologically interpretable as reflecting the intrinsic plasticity of progenitor-like states. NPC states are known to be transitional and situated at the apex of the GBM cellular hierarchy (Couturier et al., 2020); their instability quantifies this heterogeneity rigorously. Importantly, low stability does not imply biological irrelevance — NPC-like clusters showed GO enrichment for neurogenesis-related processes despite their near-100% dissolution frequency — but rather suggests that the NPC transcriptional program is distributed across a heterogeneous population whose cluster boundaries are not reproducibly defined at the current resolution.

The near-zero stability of erythrocytes (0.042) and pericytes/stroma (0.034) is consistent with contamination artifacts or extreme minority populations, supporting their exclusion from all downstream analyses. The bimodal stability distributions of MES-like, T cell, and LAM macrophage clusters are consistent with populations near the minimum cell-number detection threshold at the current resolution (n = 1,590, 207, and 125, respectively). Increasing bootstrap iterations (B >= 500) would provide tighter confidence intervals for these clusters.

3.8 Implications for therapeutic targeting

Cluster stability provides a principled criterion for prioritizing therapeutic targets. MES-like and MES-AC transitional tumor cells, with their high stability and established association with therapy resistance, invasiveness, and poor prognosis (Neftel et al., 2019), represent the most robustly defined therapeutic targets from this analysis. The consistent pathway enrichment for TNF-alpha/NF-kappaB signaling and epithelial-mesenchymal transition in the MES-like cluster supports therapeutic approaches targeting the mesenchymal niche.

Unstable tumor clusters (NPC-like, AC-like) should be interpreted with caution. While these states are biologically real in GBM, their low cluster stability indicates that current data and resolution do not provide sufficiently robust boundaries for confident marker gene or target identification. Independent validation with additional datasets, higher-resolution sub-clustering, or single-cell multi-omics would be required before therapeutic conclusions are drawn for these populations.

3.9 Methodological robustness and limitations

The high Spearman rank correlations across all three resampling methods (rho = 0.982 and 0.965) provide important methodological validation. Bootstrap, subsampling, and noise injection each test a different axis of cluster robustness: cell composition variability, data completeness, and measurement noise, respectively. Their convergence on identical categorical classifications for all 14 clusters strongly supports the conclusion that our stability results reflect genuine biological properties rather than artifacts of the specific perturbation strategy. This study has several limitations. First, bootstrap iterations were limited to B = 100; higher iterations (500-1,000) would provide tighter confidence intervals, particularly for bimodal-distribution clusters. Second, the dataset encompasses only 10 samples from the 10x Genomics platform, limiting generalizability. Third, the study relies on a single public dataset (GSE131928) without external validation; replication using independent GBM scRNA-seq cohorts (e.g., from the Human Cell Atlas or other GEO datasets) would strengthen the generalizability of our findings. Fourth, scRNA-seq cannot preserve spatial information; integration with spatial transcriptomics

4. CONCLUSION

This study demonstrates that cluster-wise bootstrap stability assessment, combined with explicit tumor–TME separation, provides a quantitative and reproducible framework for evaluating the reliability of cell population assignments in GBM scRNA-seq data. Non-malignant TME populations and committed mesenchymal tumor states (MES-AC transitional, MES-like) exhibited the highest stability, whereas progenitor-like and rare populations were unstable, reflecting biological plasticity or potential artifacts. Stability correlated with marker gene specificity, within-cluster homogeneity, and silhouette scores, reinforcing its utility as a biologically interpretable validation criterion. Future work should validate these findings using independent GBM scRNA-seq cohorts and integrate complementary modalities such as spatial transcriptomics to assess whether spatially co-localized cells correspond to the stable clusters identified here. Multi-omics integration (e.g., combining scRNA-seq with scATAC-seq or proteomic profiling) may further resolve whether transcriptionally unstable clusters harbor epigenetically or functionally distinct subpopulations. Application of this framework to longitudinal datasets spanning treatment timepoints could also clarify whether cluster stability shifts in response to therapy.

ACKNOWLEDGMENT

We acknowledge the authors of GSE131928 for making their data publicly available. We thank the members of the Master Program in Statistics, Universitas Islam Indonesia, for constructive discussions. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008, P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Couturier, C.P., Ayyadhury, S., Le, P.U., Nadaf, J., Monlong, J., Riva, G., Allache, R., Baig, S., Yan, X., Bourgey, M., Lee, C., Wang, Y.C.D., Wee Yong, V., Guiot, M.-C., Najafabadi, H., Misic, B., Antel, J., Bourque, G., Ragoussis, J.,

Petrecca, K., 2020. Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy. Nat. Commun. 11, 3406. https://doi.org/10.1038/s41467-020-17186-5

Darmanis, S., Sloan, S.A., Croote, D., Mignardi, M., Chernikova, S., Samghababi, P., Zhang, Y., Neff, N., Kowarsky, M., Caneda, C., Li, G., Chang, S.D., Connolly, I.D., Li, Y., Barres, B.A., Gephart, M.H., Quake, S.R., 2017. Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma. Cell Rep. 21, 1399–1410. https://doi.org/10.1016/j.celrep.2017.10.030

Dudoit, S., Fridlyand, J., 2003. Bagging to improve the accuracy of a clustering procedure. Bioinformatics 19, 1090–1099. https://doi.org/10.1093/bioinformatics/btg038

Hafemeister, C., Satija, R., 2019. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 20, 296. https://doi.org/10.1186/s13059-019-1874-1

Hennig, C., 2007. Cluster-wise assessment of cluster stability. Comput. Stat. Data Anal. 52, 258–271. https://doi.org/10.1016/j.csda.2006.11.025

Jaccard, P., 1901. Distribution de la Flore Alpine dans le Bassin des Dranses et dans quelques régions voisines. Bulletin de la Societe Vaudoise des Sciences Naturelles 37, 241–72.

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P., Raychaudhuri, S., 2019. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods 16, 1289–1296. https://doi.org/10.1038/s41592-019-0619-0

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., Tamayo, P., 2015. The Molecular Signatures Database Hallmark Gene Set Collection. Cell Syst. 1, 417–425. https://doi.org/10.1016/j.cels.2015.12.004

McInnes, L., Healy, J., Melville, J., 2018. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. ArXiv.

Monti, S., Tamayo, P., Mesirov, J., Golub, T., 2003. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. Mach. Learn. 52, 91–118. https://doi.org/10.1023/A:1023949509487

Neftel, C., Laffy, J., Filbin, M.G., Hara, T., Shore, M.E., Rahme, G.J., Richman, A.R., Silverbush, D., Shaw, M.L., Hebert, C.M., Dewitt, J., Gritsch, S., Perez, E.M., Gonzalez Castro, L.N., Lan, X., Druck, N., Rodman, C., Dionne, D., Kaplan, A., Bertalan, M.S., Small, J., Pelton, K., Becker, S., Bonal, D., Nguyen, Q.-D., Servis, R.L., Fung, J.M., Mylvaganam, R., Mayr, L., Gojo, J., Haberler, C., Geyeregger, R., Czech, T., Slavc, I., Nahed, B. V., Curry, W.T., Carter, B.S., Wakimoto, H., Brastianos, P.K., Batchelor, T.T., Stemmer-Rachamimov, A., Martinez-Lage, M., Frosch, M.P., Stamenkovic, I., Riggi, N., Rheinbay, E., Monje, M., Rozenblatt-Rosen, O., Cahill, D.P., Patel, A.P., Hunter, T., Verma, I.M., Ligon, K.L., Louis, D.N., Regev, A., Bernstein, B.E., Tirosh, I., Suvà, M.L., 2019. An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. Cell 178, 835-849.e21. https://doi.org/10.1016/j.cell.2019.06.024

Ostrom, Q.T., Price, M., Neff, C., Cioffi, G., Waite, K.A., Kruchko, C., Barnholtz-Sloan, J.S., 2022. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2015–2019. Neuro. Oncol. 24, v1–v95. https://doi.org/10.1093/neuonc/noac202

Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B. V., Curry, W.T., Martuza, R.L., Louis, D.N., Rozenblatt-Rosen, O., Suvà, M.L., Regev, A., Bernstein, B.E., 2014. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science (1979). 344, 1396–1401. https://doi.org/10.1126/science.1254257

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., Satija, R., 2019. Comprehensive Integration of Single-Cell Data. Cell 177, 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031

The Cancer Genome Atlas Research Network, 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455, 1061–1068. https://doi.org/10.1038/nature07385

Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., Fallahi-Sichani, M., Dutton-Regester, K., Lin, J.-R., Cohen, O., Shah, P., Lu, D., Genshaft, A.S., Hughes, T.K., Ziegler, C.G.K., Kazer, S.W., Gaillard, A., Kolb, K.E., Villani, A.-C., Johannessen, C.M., Andreev, A.Y., Van Allen, E.M., Bertagnolli, M., Sorger, P.K., Sullivan, R.J., Flaherty, K.T., Frederick, D.T., Jané-Valbuena, J., Yoon, C.H., Rozenblatt-Rosen, O., Shalek, A.K., Regev, A., Garraway, L.A., 2016. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science (1979). 352, 189–196. https://doi.org/10.1126/science.aad0501

Traag, V.A., Waltman, L., van Eck, N.J., 2019. From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. 9, 5233. https://doi.org/10.1038/s41598-019-41695-z

Verhaak, R.G.W., Hoadley, K.A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M.D., Miller, C.R., Ding, L., Golub, T., Mesirov, J.P., Alexe, G., Lawrence, M., O'Kelly, M., Tamayo, P., Weir, B.A., Gabriel, S., Winckler, W., Gupta, S., Jakkula, L., Feiler, H.S., Hodgson, J.G., James, C.D., Sarkaria, J.N., Brennan, C., Kahn, A., Spellman, P.T., Wilson, R.K., Speed, T.P., Gray, J.W., Meyerson, M., Getz, G., Perou, C.M., Hayes, D.N., 2010. Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell 17, 98–110. https://doi.org/10.1016/j.ccr.2009.12.020

Yu, G., Wang, L.-G., Han, Y., He, Q.-Y., 2012. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS 16, 284–287. https://doi.org/10.1089/omi.2011.0118