Single-Cell RNA Sequencing (scRNAseq) Analysis of Murine Glioma to Investigate the Potential of Astrocytes as a Therapeutic Target

Abstract

BACKGROUND Glioblastoma is an aggressive brain tumor in which only 5% of patients reach the 5-year survival mark. Improvement of immunotherapies for glioblastoma requires a thorough understanding of the shifting cell dynamics in response to different treatment modalities. This can provide insight into the mechanism behind therapeutic success and explanations for failed trials. Cell dynamics is well captured through single-cell RNA sequencing (scRNAseq), a technique in which the genetic transcript of each cell from a tumor sample is identified. The role of astrocytes in tumor progression is largely unknown but has shown potential as a therapeutic target. **METHODS** scRNAseq data from glioma tumor samples on murine models treated with 9 Gy radiation, TOFU-ACT immunotherapy, and treatment-naïve models was analyzed using Seurat. Uniform Manifold Approximation and Project (UMAP) plots were constructed. UMAPs were visually inspected and fold-change increases in cell type counts of interest were investigated. The scRNAseq analysis pipeline was repeated on astrocytes. **RESULTS** A decrease in tumor cells validated the success of the therapies. The 3-fold increase in astrocytes highlights a potential mechanism behind therapeutic success. Astrocytes could not be further classified into cell subtypes, either due to cell homogeneity or the need for a more complex reference data set.

Introduction

Glioblastoma (GBM) is an aggressive and lethal brain tumor for which there is no known cure. Only 5% of patients meet the 5-year survival mark, as identified by Delgado-López (2016). The cell types recruited to the brain during tumor formation and growth are constantly modulated; furthermore, these cells may even change in function. Preclinical studies require a thorough understanding of the cell dynamics in the tumor microenvironment in response to brain malignancies and therapies.

Cell dynamics can be investigated through single-cell RNA sequencing (scRNAseq), which records the genetic transcript of the individual cells in a sample. Samples are taken from the tumor microenvironment, sequenced, and analyzed. This study validates the scRNAseq analysis of Trivedi et al. and builds upon their work to investigate the role of astrocytes in tumor growth and response to therapy. Astrocytes are glial cells whose role in tumor progression is largely unknown. Roemeling (2023) demonstrates that these cells show promise as a therapeutic target due to its ability to recruit cells to the tumor microenvironment through chemokine secretion, tuning the tumor landscape to one that either promotes or prevents tumor growth.

Methods & Interpretation

Sample Collection and Sequencing Tumor samples were collected from mice that were either treatment naïve or had undergone 9 Gy radiation or TOFU-ACT immunotherapy. The samples were sequenced using Illumina Novaseq 6000 and demultiplexed by Trivedi (2024) using the Cellranger 7.0 pipeline and Seurat 4.0 HTOdemux.

Data was collected by Trivedi (2024) and is available at the NCBI Gene Expression Omnibus with accession numbers GSE251798, GSE251799, and GSE251800. All original analysis of single-cell data was conducted in R 4.3.3 using Seurat 5.0.3, which was developed by Hao (2023).

Quality Control and Data Transformation The goal of QC is to preserve only live and representative cells. Only genes expressed in at least 3 cells were analyzed. High mitochondrial genetic content is indicative of dying cells, so cells with mitochondrial content values greater than 5% were filtered out. Cells which express fewer than 250 genes and have fewer than 500 molecules are filtered out in an effort to remove poor reads and empty samples. Remaining cells were downsampled to retain 1000 cells per treatment group.

All functions used to transform data are available in the Seurat package developed by Hao (2023). Expression per cell is normalized based on the expression of that gene across all detected cells before performing a log transformation. This was done using NormalizeData(). Next, FindVariableFeatures() was used to find the most variable genes. Data was then linearly transformed such that gene expression had $\mu = 0, \sigma = 1$ using ScaleData(). This avoids placing undue importance in downstream analyses on genes that are more highly expressed. PCA was performed on the variable genes using RunPCA().

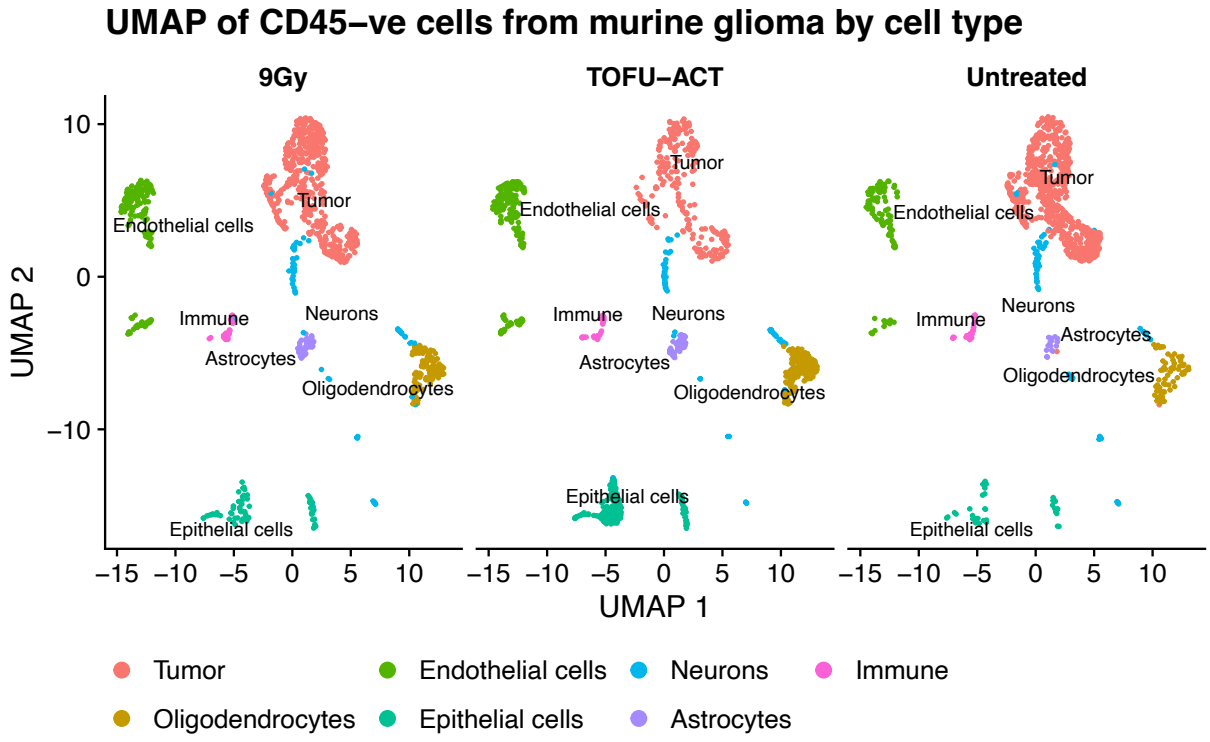
The dimensionality of the dataset determines the principal components with which UMAPs are constructed. Here, we use an elbow plot (Figure 2) to determine dimensionality as suggested by Seurat developers. This orders principal components by the proportion of total variance it accounts for. 20 principal components was taken to be the ideal parameter for returning 7-9 clusters, which is consistent with the "elbow" landmark in the plot.

Clustering and Dimensionality Reduction A K-nearest neighbors graph is constructed based on the distance between principal components using FindNeighbors() based on the first 20 principal components as determined in the previous section. Cells are grouped together with FindClusters(), which uses the Louvain algorithm. The resolution argument was chosen to be 0.4. Typical values for this parameter are greater than 0.6, but can be lower for smaller datasets like this one (n=3000 after downsampling). Additionally, this resolution value was chosen to optimize the number of clusters as requested by the lead computational biologist (7-9 clusters). 13 clusters were found, some of which were combined (see Cluster Annotation) to produce 7 total clusters.

A Uniform Manifold Approximation and Projection (UMAP) plot was constructed for each treatment group using Seurat::DimPlot using the integrated reduction (Figure 3).

Cluster Annotation Cells were annotated (Figure 1) using the reference murine gene set Immgen provided by SingleR. SingleR uses the Wilcoxon ranked sum test to make pairwise comparisons between cell types.

Figure 1: Annotated UMAP reveals 7 unique cell clusters



Feature plots were constructed to visualize the expression of luciferase and the PTPRC gene across all clusters (Figure 4). These features are indicative of tumor cells and immune cells, respectively. CD45 is a marker of immune cells, so in the context of performing scRNAseq on CD45-ve cells, cells expressing PTPRC are considered contamination. While the immune population is small enough to allow accurate interpretation of the UMAP, efforts should still be taken in the future to improve the biological protocol such that contamination is minimized.

Visual inspection of the feature plots (Figure 4) reveals clusters 0, 2, 4, and 7 highly express luciferase. These clusters were manually annotated as tumor cells. Similarly, cluster 10 highly expressed the PTPRC gene and was called immune cells.

Immune cell (contamination) populations are small and constant throughout all treatment groups, meaning annotated UMAPS (Figure 1) can be interpreted as is.

Investigate Astrocyte Subtypes A reduction in tumor cell count confirms the success of radiation and the novel immunotherapy TOFU-ACT. Murine models that underwent each therapy experienced more than a 3-fold increase in astrocytes, a cell type whose role in modulating the immune system is still an active area of investigation. Astroyctes have been shown to release immunosuppresive chemokines in some glioma studies, but have also been shown by Roemeling (2023) to have potential for being infected by virotherapies

to kickstart anti-tumor action. This demonstrates the need to further classify astrocytes in the context of glioblastoma.

In an attempt to uncover the role of astrocytes in glioma and their response to immunotherapy, the astrocyte cluster is further investigated. Cells annotated as astrocytes are subsetted before performing the same anlaytical workflow that was applied to the entire data set. The elbow plot (Figure 5) reveals the data has dimensionality of 5.

UMAPs were constructed (Figure 6) and annotated (Figure 7).

Discussion

Astrocytes as a Therapeutic Target Cell clusters in the astrocyte UMAPs (Figure 6) seem relatively evenly distributed across PCA space for both therapeutic modalities; subtypes of the astrocyte cells are not differentiated. Astrocytes in the untreated murine models have less variance. It is possible that there are too few detected astrocytes for an accurate investigation. Another possibility is that astrocytes in this tumor model are homogenous: astrocytes are not differentiated into different activation states or phenotypes. The lack of distinct clustering may also be due to a lack of reference genes provided by the Immgen data set. Literature review should be conducted to find genes unique to different activation states of astrocytes, and a new reference gene set should be constructed.

If astrocytes play an immunosuppressive role in this case, it is possible that their multiplication may be a consequence of the need for immune regulation in response to the therapies' increase in immune response. Abberant inflammation in the brain can be fatal, so regulation and immune-suppressive feedback systems are prevalent in this tumor model.

However, it is also possible that the therapies have some role in shifting the astrocytes' phenotype towards one that is inflammatory and rejects tumors. This study highlights the need for further investigation of astrocytes in response to brain malignancies and therapies.

Relevance of UMAPs to Tumor Studies Constructed UMAPs validate therapeutic efficacy against glioma in murine models and visualize the changing non-immune cell subsets. This technique of single-cell RNA sequencing analysis is very useful as it reduces high-dimensional data to a 2 dimensional plot that can be easily interpreted. The technique is only gaining more significance and is often thought of as the "gold standard" for modeling cell dynamics in the cancer field.

Despite its ease of interpretability and simple visualization of high dimensional data, biological conclusions should not be drawn from these plots alone because it is impossible to preserve all complexities of the data with a simple model. Furthermore, this represents predicted cell dynamics based on genetic transcripts, all of which are not translated into functional protein. While genetic transcripts may be indicative of cell type and function, transcripts are still modified and can be excluded entirely before being translated into protein. Even proteins are often modified to produce a host of different functions.

Conclusion

The UMAP plots produced by scRNAseq analysis in Trivedi et al. were validated. The high fold-change increase in astrocyte counts in response to both radiation and immunotherapy highlights its potential as a therapeutic target. Astrocytes were subsetted and further analyzed using the same scRNAseq analysis pipeline. Analysis failed to differentiate astrocytes intro further cell subtypes, either due to cell homogeneity or need for a more complex reference data set. Future work should involve building an improved reference gene data set based on known markers of astrocytes in different activation states and applying the analysis pipeline to other data sets where a greater number of astrocytes were detected.

References

Delgado-López, P. D. 2016. "Survival in Glioblastoma: A Review on the Impact of Treatment Modalities." *Clinical and Translational Oncology*.

Hao, Yuhan. 2023. "Dictionary Learning for Integrative, Multimodal and Scalable Single-Cell Analysis." *Nature Biotechnology*.

Roemeling, Christina von. 2023. "CXCL9 Recombinant Adeno-Associated Virus (AAV) Virotherapy Sensitizes Glioblastoma (GBM) to Anti-PD-1 Immune Checkpoint Blockade." *Research Square*.

Trivedi, Vrunda. 2024. "mRNA-Based Precision Targeting of Neoantigens and Tumor-Associated Antigens in Malignant Brain Tumors." *Genome Medicine* 16 (17).

Appendix

Figure 2: Principal components of tumor data ranked by proportion of total variance accounted for reveals dimensionality = 20

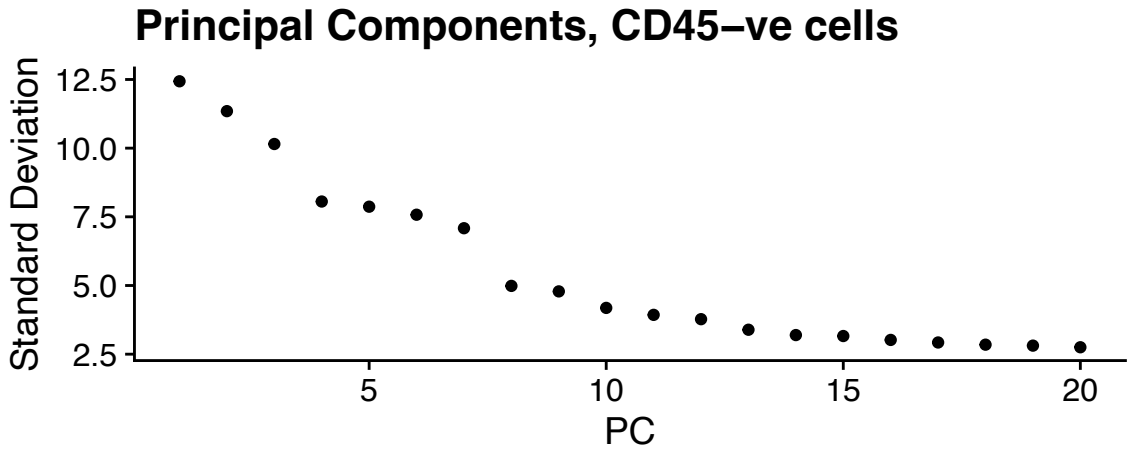


Figure 3: PCA and clustering reveals 14 unique clusters across treatment groups 9 Gy radiation, TOFU-ACT immunotherapy, and treatment naïve mice

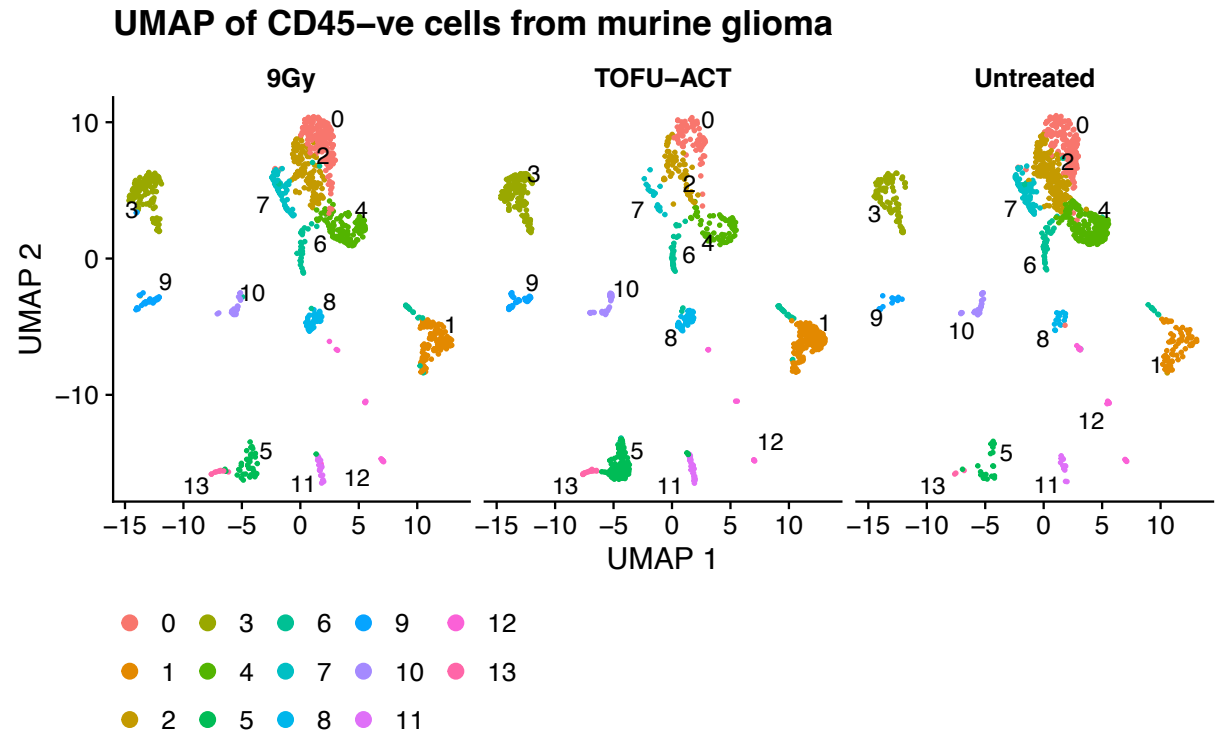


Figure 4: Feature plots show strength of luciferase expression (row 1) and PTPRC gene expression (row 2). Gene expression increases as color approaches blue

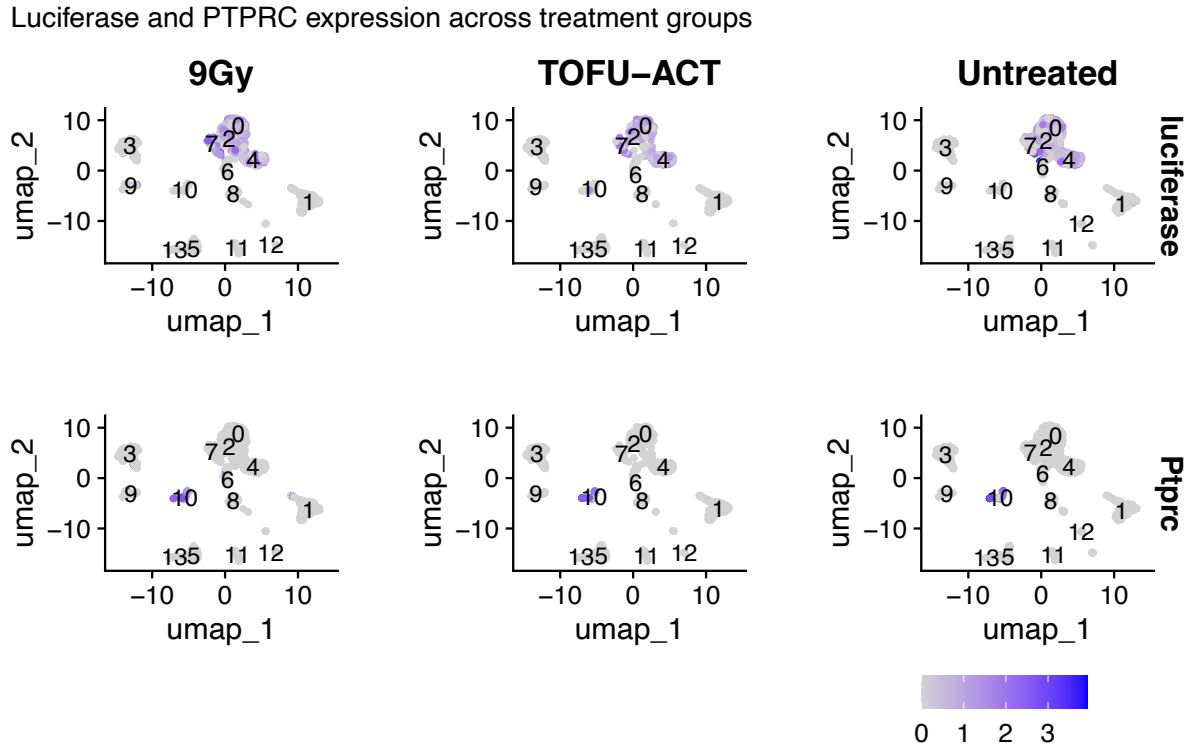


Table 1: Cell type distribution (as percentage) across treatment types

	9Gy	TOFU-ACT	Untreated
Tumor	48.0	21.3	66.2
Oligodendrocytes	13.2	20.7	7.3
Endothelial cells	17.3	21.7	9.4
Epithelial cells	7.8	21.2	3.7
Neurons	6.9	6.7	9.2
Astrocytes	4.3	5.5	1.5
Immune	2.5	2.9	2.7

Figure 5: Principal components of astrocyte data ranked by proportion of total variance accounted for reveals dimensionality = 5

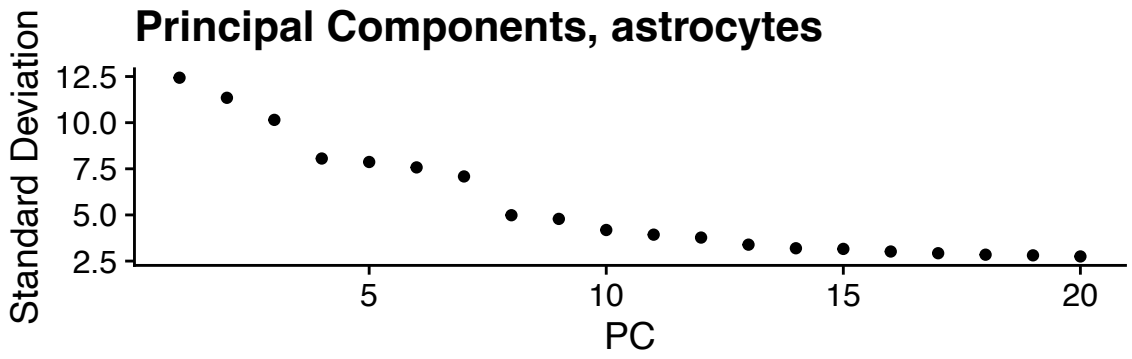


Figure 6: Clustering of astrocyte data reveals 3 unique clusters with large variance

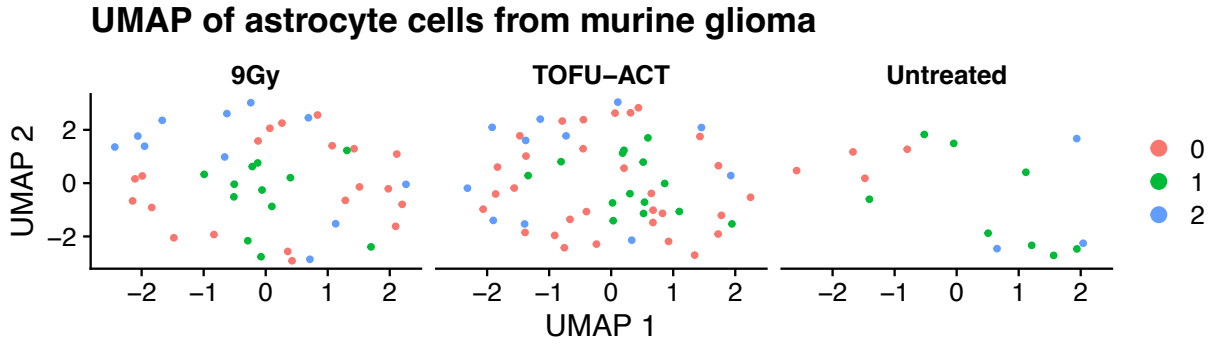


Figure 7: Annotation of astrocyte data does not uncover astrocyte subtypes