Computational Analysis of Single-Cell Gene Expression in the Tumor Microenvironment of Metastatic Cancer Cells in the Brain

Tara Roy Choudhury

Abstract

Brain metastases are the most common type of brain tumor, and around 10% to 26% of patients who die from cancer developed brain metastases [1]. Therefore, there is a great need to investigate the heterogeneity of these tumors so that people can identify and treat them. The developments of single cell RNA sequencing (scRNAseq) allow one to explore the TME (tumor microenvironment) and show how different genes can be expressed and stored in a computational library. In this study, I investigated published scRNAseq datasets of tumors from human patients with brain metastases. I chose to look at endothelial cells, and noticed that they participated in active communication with fibroblasts within the tumor microenvironment. Moreover, in another source that studied fibroblasts, I noticed that tumors are similar to endothelial cells, which may be an important aspect of the TME to target in therapeutics. This paper is a review and analysis of published scRNAseq datasets.

Introduction

Not many people would know what cancer is, how it originates, or how deadly it is. Cancer is a disease when cells in an organ have a gene or multiple genes that mutate, so that the cells uncontrollably divide and interfere with the functioning of healthy tissues and organs [2]. Cancerous cells can originate from specific sites in the body or from multiple organs. When a cancer is metastatic, malignant cells can travel to other parts of the body and invade healthy organs [3]. These cancers travel through the bloodstream and can invade places like the brain [4]. Until very recently, diseases, like cancer, were not being able to be treated due to the lack of understanding of the mutations that had occurred, and which cells were involved. A new gene expression technology being used to study features of individual cells from a tissue, e.g. a brain metastasis, and to decode the different genes that are activated in a particular cell. This is known as single cell technologies, specifically single cell RNA sequences [5]. We can now use single cell technologies to take apart complex tumors and analyze how specific cell types build up a unique tumor microenvironment (TME). This information will be crucial to develop therapeutics that target the correct cell types and potentially the correct genes of interest.

Background

Single cell RNA sequencing (scRNAseq) works by separating single cells that originate from tissue samples, e.g. tumorous tissues, into a tube, which are then molecularly barcoded for identification use later on, and the genes are amplified for sequencing. Then, the cells go through sequencing machines that take in single cells

suspended in a droplet [5]. This is to identify all the genes and their functions and report them back in bioinformatic files. Finally, all the data is stored in a computational library. Single cell RNA sequencing reveals the composition of different cell marker genes which identify the different types of cells and their functions in organisms [6]. We can also use these data files to identify any anomalies in gene expression and compare cell types to each other.

Brain metastasis is a form of cancer that originates from another part of the body, but then travels to the brain through the bloodstream. A metastatic tumor needs to establish a habitable space within another organ. A tumor consists of not just the metastatic cancerous cells, but also additional cell types. A tumor can send signals to the cells around it to promote the tumor's growth and survival. A tumor microenvironment is the tumor and the cells around it that stimulate and interact with each other. Cells from a tumor in an organ can dislodge, travel through the bloodstream to another organ like the brain, and then use the resources of its surroundings to continue to divide and grow.

Recently, work has been done to characterize the gene expression profiles of individual cells coming from metastatic tumors. There are many datasets, which are publicly available, that provide resources to study these cells and their genes further. One can look into what genes distinguish tumor cells from healthy cells, which genes are required to form a tumor's microenvironment, or look into specific cells in the TME like the endothelial cells, which are cells found in the lining of blood vessels. In this paper, I have analyzed some published data and learned that there are many aspects to a tumor's microenvironment, and that specific unexpected cells like myeloid cells or fibroblasts have a lot of impactful interactions with tumor cells, and sometimes even act like tumor cells.

This paper is a literature review which summarizes complicated works in a digestible format for everyone to read. It also analyzes the full set of genes that are involved, and asks whether the TME cells are completely like the cell type they are described as, or if they express some abnormal genes. It also will bring light to any cells that are not discussed or unknown in previous works.

Methods

The RStudio model R-4.4.1 for Windows was downloaded and installed on my computer. Seurat and Seurat Object packages were downloaded. Data was pulled from the Gene Expression Omnibus. Various files were then imported into RStudio. The datasets in the paper I analyzed are found under GEO number GSE234832 [7]. All the analyses were enabled by the Seurat package. A website about Seurat [8] and a YouTube tutorial were used to help me with the code [9]. A sample of the code can be found in Figure 1.

Figure 1:

```
library(Seurat)
filename = 'C:/Users/bholetara/Desktop/Polygence_Data_Sets/July7_Dataset_1/'
BRBMET2.data <- Read10X(data.dir = filename)

BRBMET2 <- CreateSeuratObject(counts = BRBMET2.data,project = "BRBMET2")

BRBMET2 <- NormalizeData(object = BRBMET2)

BRBMET2 <- FindVariableFeatures(object = BRBMET2)

BRBMET2 <- ScaleData(object = BRBMET2)

BRBMET2 <- RunPCA(object = BRBMET2)

ElbowPlot(BRBMET2)

BRBMET2 <- RunUMAP(object = BRBMET2,reduction = "pca",dims = 1:8)

BRBMET2 <- FindNeighbors(BRBMET2)

BRBMET2 <- FindClusters(BRBMET2)

DimHeatmap(BRBMET2, dims = 1:15, cells = 500, balanced = TRUE)

DimPlot(object = BRBMET2, reduction = "umap")

VlnPlot(BRBMET2, features = c("KRT19", "COL1A2"))

#3 is Myeloid ; Tumor most prominet in 0,1,2,4,5,6,7 ; 10 is T-Cell ; 4, 5, 6 are NSC
#9 is Oligo ; 11 is Fibroblast ; 11 high in Endothelial

FeaturePlot(BRBMET2, features = c("TREM2", "SFN", "TRAT1", "UBE2C", "GAL3ST1", "COL1A2", "CDH5"))

BRBMET2 <- RunTSNE(BRBMET2, dims = 1:10)

DimPlot(BRBMET2, reduction = "tsne")

current.cluster.ids = c(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11)

new.cluster.ids = c("Tumor", "Tumor", "Unknown", "Myeloid", "NSC", "Tumor", "Tumor", "Tumor", "Tumor", "Oligo", "T-Cell",

names(x = new.cluster.ids) <- levels(x = BRBMET2)

BRBMET2 <- RenameIdents(object = BRBMET2, new.cluster.ids)

DimPlot(object = BRBMET2, reduction = "tsne", label = TRUE, pt.size = 0.5)

TSNEPlot(BRBMET2, label = TRUE, pt.size = 1, label.size = 4)
```

Figure 1 is a piece of the code used to generate a tSNE plot using Seurat in RStudio. This code plotted a scRNAseq dataset with labels depending on the genes found in the different cells.

Results

One dataset that took tumor samples from five patients was evaluated. These patients had metastatic tumors in the brain that originated from either the breast or the lung [10]. The cells were run through single cell RNA sequencing to reveal specific characteristics relating to the tumor and its TME. These datasets came from tumors that have been established because of metastasis, not from primary tumors. They are rare and hard to come by because these clinical samples are not readily available. This work lays the foundation for future work that can be done when more samples are collected and datasets become available.

In the first dataset, the single cell technology identified a type of cell called the fibroblast that was important for establishing the tumor microenvironment. It also identified a set of genes encoding type I collagen and collagen signaling pathways that were highly expressed. Although the first study found an important role for fibroblasts, there were other cell types that I wanted to explore further. There were also other open

questions that I thought of that remained to be answered. For example, are the fibroblasts expressing non-fibroblast genes as well? Are they completely normal fibroblasts? One can ask these same questions for other cell types populating the TME like endothelial cells and myeloid cells.

Figure 2: A tSNE plot shows the various cell types in the first TME scRNAseq dataset. Seurat was used to plot the scRNAseq dataset, where each dot represents the gene expression state of a single cell. Dots that are close to each other and have the same color are cells of the same type. We can study the distribution of points on this plot to make conclusions about how similar or different cells in the TME are.

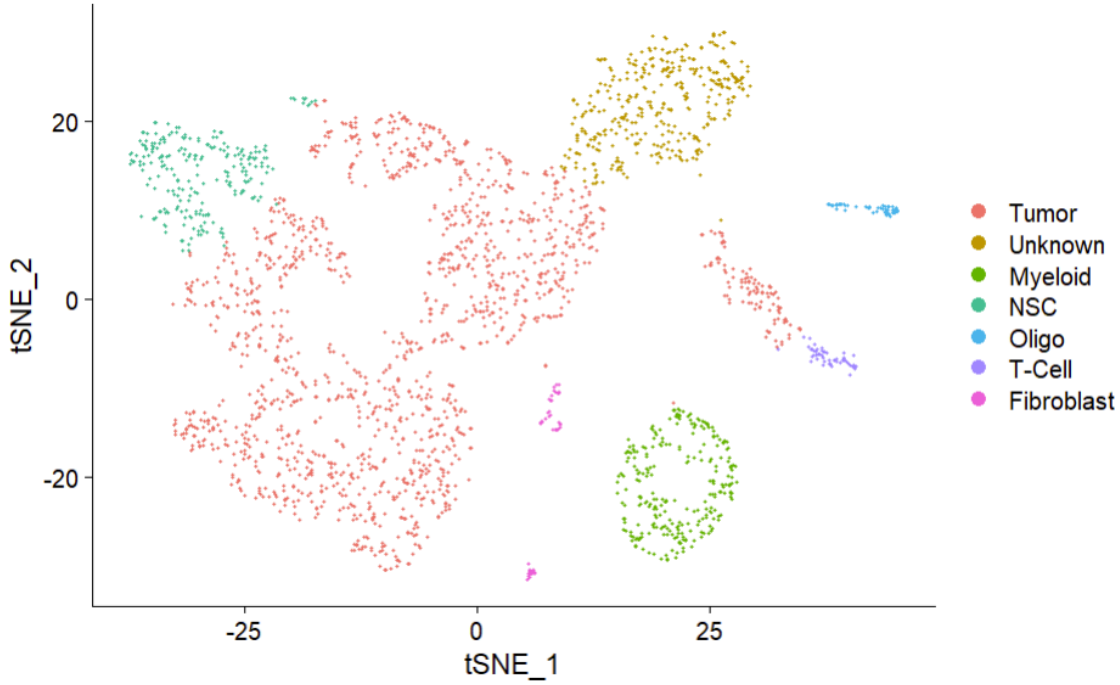


Figure 2 shows all the different cell types clustered under labels of the type of cell they are. This is dependent on the genes they express. I found it interesting how there was a big unknown region on the figure, and that it had a small amount of tumor cells in it. However, there was no cell type that specifically claimed to be a part of the unknown region. The majority of the cells were tumor cells. The Myeloid cells like to stay secluded in a circular shape away from the tumor. The NSC cells and the T-Cells reside very closely to the tumor cells. The Fibroblasts and the Endothelial cells also appear to be very similar considering they both have high concentrations in the same areas.

Figure 3: A tSNE plot portrays the various cell types in the second individual's TME scRNAseq dataset. Seurat was used to plot the scRNAseq data. It is the same type of plot as in Figures 2 and 4.

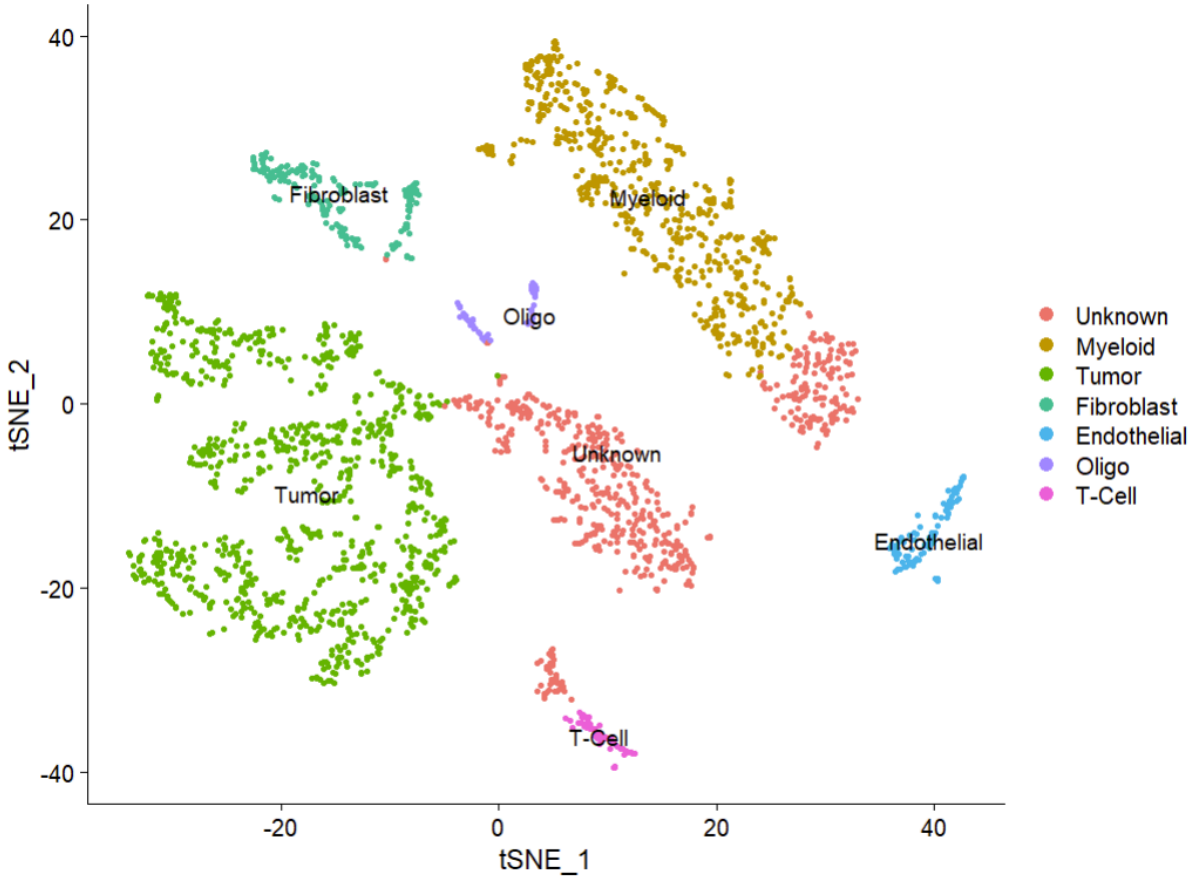
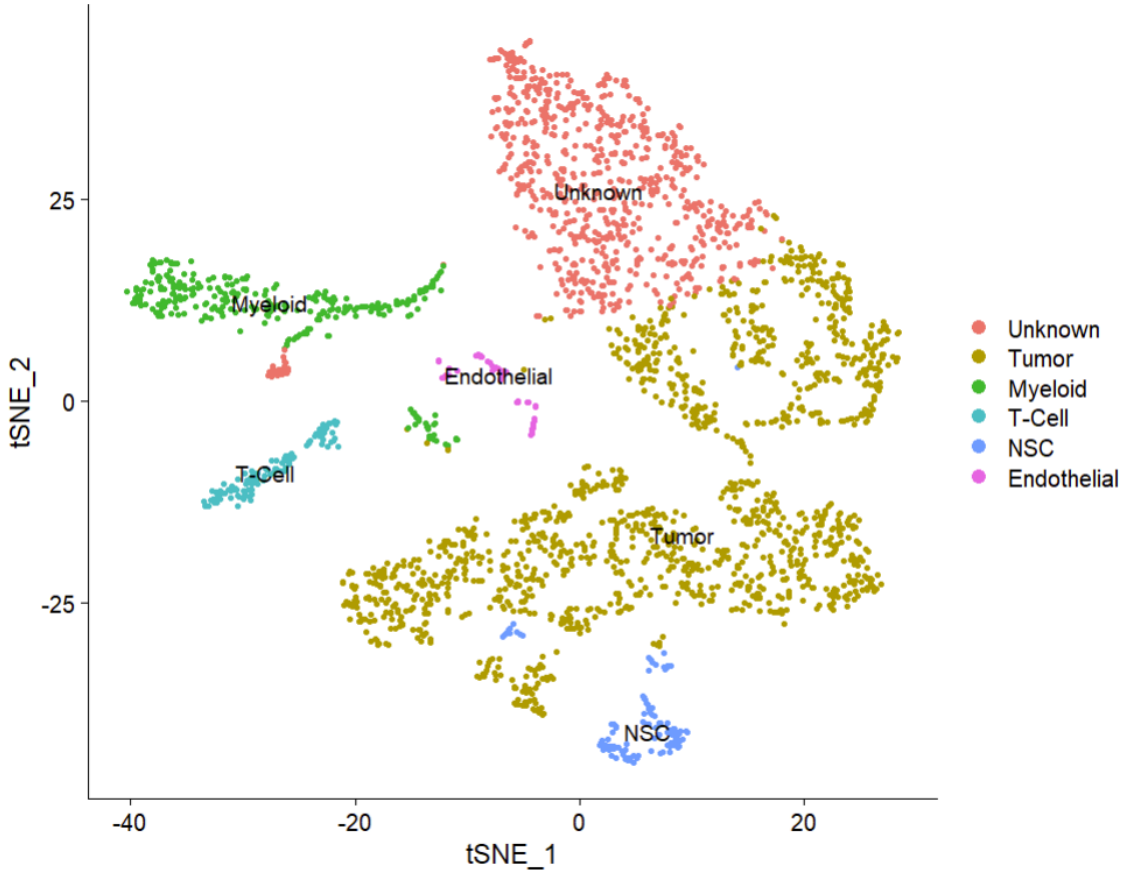


Figure 3 was interesting because it also had a relatively large tumor aspect to it, just like Figure 2. The tumor cells were located in the same vicinity as the second figure. Moreover, there was a big unknown region on the figure, and it had a small amount of tumor cells in it. Again, there were no cell types that specifically claimed to be a part of the unknown region. In Figure 2, the Myeloid cells were located separately from the tumor cells and excluded from the rest of the cell types. While the Myeloid cells stayed away from the tumor cells in Figure 3 as well, they were similar to some of the unknown cells. The T-Cells resided very closely to the tumor cells in both of the figures too. The Fibroblasts and the Endothelial cells also both seem to be appearing in the same area; however, I only marked it as the Endothelial cells on the figure.

Figure 4: Another tSNE plot which is very similar to Figures 2 and 3. This figure, however, has a different TME.



Just like both Figures 2 and 3 above, there were a large portion of tumor cells that resided in the tSNE plot. There were still a large portion of unknown cells that were closely related to the tumor cells, and a few of the unknown cells were related to the Myeloid cells just like in Figure 3.

In all three datasets, the Fibroblasts and the Endothelial cells appeared to be very similar, and were in the same areas in all the figures. It is interesting to see how all three of the datasets' tumor cells are located in different parts of the figures, meaning that they are not all in one specific area. One of the questions I asked in the beginning was if the fibroblasts were expressing non-fibroblast genes as well? They could possibly be expressing endothelial cell genes, because I noticed that the endothelial and fibroblast cells were producing similar enough genes to be classified as the same type of cell. Another question raised was if the identified fibroblasts were completely normal fibroblasts? One may never know what a normal fibroblast cell is due to the ever-changing nature of biology. Nonetheless, they were classified as fibroblast cells, so they must be producing the genes of a fibroblast cell. Many of the important gene markers for fibroblast cells were present in these datasets. However, we would need to directly compare TME fibroblasts to healthy non-TME associated fibroblast cells to definitively tell whether they are the exact same cell type. They overall do not share many characteristics with tumor cells, as they are found in their own cluster in the tSNE plots.

Discussion

I observed significant heterogeneity of tumor cells in terms of their gene expression analyzed by scRNAseq. This might mean that tumor cells can have many different identities depending on where they are in the tumor microenvironment. I even observed some tumor cells that were similar to T cells, which were overall very different from the majority of the population of tumor cells. The unknown regions in my figures were very close to the tumor cells. Some were mixed with the tumor cells. I hypothesize that there could be a different cell type altogether that has changed into a gene expression state that we have not fully characterized. Alternatively, the unknown regions are the same tumor cells but are expressing different genes that Seurat was not able to identify.

I hypothesize that tumor cells cannot be targeted by therapeutics by considering one particular set of genes that are expressed, but rather therapeutics need to consider the entire gene expression profile. This includes genes that are unexpected, like those found in the unknown cell type, or genes from cells in the TME that are not tumors, like fibroblasts. I further hypothesize that the unknown cell type is some form of tumor that possesses genes that we do not already know about. More experiments will need to be done to identify the unknown cell type that I have found in these tumor microenvironment data. One could study which genes are highly expressed in the unknown cell type, and then compare the data to all known cell types in the body, not just TME cells. This analysis would identify whether the unknown cell type is a population of healthy cells that have invaded the TME, or a population of abnormal cells that we would need to identify in order to target.

There is much to do, and one needs to realize what the limitations could be. Single cell technologies are useful, but can potentially miss some cell types or be confounded by noise that does not represent a real biological signal. This will result in technical error that we can misinterpret. One way we can counter this is by using controlled replicates. We could do the exact process of scRNAseq in the same way, using the same technology, incorporating multiple different tissue samples or different parts of the same tissue sample in the experiment. If the unknown cell type appears consistently, it may be a significant feature of a TME that we need to characterize. If it shows up sporadically, it may be an error that we need to verify. It could also very well be a really spontaneous feature of a tumor.

Moreover, this study has considered patient samples that can be of different quality. The samples could have been processed at different times, from different labs, or by different technicians. There are many variables that can contribute to inconsistency across samples. One can counter this by designing a study in which samples are processed in an identical way and at the same time.

I have learned the utility of single cell approaches for studying heterogeneity of gene expression in cells that make up a TME. Heterogeneity means the full spectrum or consisting of diverse elements. However, these datasets are just a snapshot of the end stage of cancer progression. To really target cancer progression, we need to understand how tumor gene expression profiles develop and change over time. A dream experiment would be to perform scRNAseq experiments on primary tumors where they originated, before the metastasis is created, and generate a tumor microenvironment. The data I have analyzed suggest that targeting cell types that are

similar to NSCs and T-cells could be most useful for interacting with the tumor cells given the distribution gene expression states.

I also looked at a second dataset that consisted of tumor samples from 14 patients who had metastatic tumors that originated from various other parts of the body [11]. The dataset is made up of four gliomas, and 10 lung-to-brain metastases. A future direction will be to apply the same approaches outlined in this paper to these datasets.

Conclusion

In summation, scRNAseq is needed to study the gene heterogeneity of cells, tissues, and at other biological levels. The tumor microenvironments of tissues are diverse and scRNAseq captures multiple cell types that appear in the vicinity. Some of those cell types are not annotated. There may be a sporadic, or a not yet known feature of the TME that requires further study and to be discovered.

References

[1] Amsbaugh, M. J., & Kim, C. S. (2023). Brain Metastasis. In StatPearls. StatPearls Publishing.
https://pubmed.ncbi.nlm.nih.gov/29262201/

[2] What Is Cancer? - NCI. (2021, October 11). National Cancer Institute. Retrieved November 22, 2024.
https://www.cancer.gov/about-cancer/understanding/what-is-cancer

[3] Metastatic Cancer: When Cancer Spreads. (2020, November 10). National Cancer Institute. Retrieved November 22, 2024.
https://www.cancer.gov/types/metastatic-cancer

[4] Nicolson, G. L. (1979). Cancer Metastasis. *Scientific American*, *240*(3), 66–77.
http://www.jstor.org/stable/24965152

[5] Single-cell RNA sequencing technologies and applications: A brief overview. (2021, December 20). 10.1002/ctm2.694.
https://pubmed.ncbi.nlm.nih.gov/35352511/

[6] Xinxin Zhang, Yujia Lan, Jinyuan Xu, Fei Quan, Erjie Zhao, Chunyu Deng, Tao Luo, Liwen Xu, Gaoming Liao, Min Yan, Yanyan Ping, Feng Li, Aiai Shi, Jing Bai, Tingting Zhao, Xia Li, Yun Xiao, CellMarker: a manually curated resource of cell markers in human and mouse, Nucleic Acids Research, Volume 47, Issue D1, 08 January 2019, Pages D721–D728.
https://academic.oup.com/nar/article/47/D1/D721/5115823

[7] Song Q, Ruiz J, Xing F, Lo HW et al. Single-cell sequencing reveals the landscape of the human brain metastatic microenvironment. Commun Biol 2023 Jul 21;6(1):760. PMID: 37479733.
https://pubmed.ncbi.nlm.nih.gov/37479733/

[8] Waltman, L., van Eck, N. J., & Satija, R. (2023, October 31). Analysis, visualization, and integration of Visium HD spatial datasets with Seurat • Seurat. Analysis, visualization, and integration of Visium HD spatial datasets with Seurat • Seurat. Retrieved November 22, 2024. https://satijalab.org/seurat/articles/pbmc3k_tutorial.html#setup-the-seurat-objectAnalysis,%20visualization,%20and%20integration%20of%20Visium%20HD%20spatial%20datasets%20with%20Seurat%20%E2%80%A2%20Seurat

[9] math et al. (2019, August 28). t-SNE Cluster Labeling | scRNA-Seq Analysis in Seurat [Video]. YouTube. https://youtu.be/UbOlhhpNj-A?si=ANMqX4FWKxpsWRIw

[10] Song, Q., Ruiz, J., Xing, F. et al. Single-cell sequencing reveals the landscape of the human brain metastatic microenvironment. Commun Biol 6, 760 (2023). https://doi.org/10.1038/s42003-023-05124-2

[11] Sun HF, Li LD, Lao IW, Li X et al. Single-cell RNA sequencing reveals cellular and molecular reprogramming landscape of gliomas and lung cancer brain metastases. Clin Transl Med 2022 Nov;12(11):e1101. PMID: 36336787. https://pubmed.ncbi.nlm.nih.gov/36336787/