Research Proposal: Quantifying Functional Phenotypes of Microglia

Mike Schachter
March 15, 2021

Abstract

Brain cancers comprised 1.3% of all cancer cases in 2020 but 3% of all cancer deaths. In the pediatric population, brain tumors comprise 15% of childhood cancers. The most deadly form of brain cancer, Glioblastoma, has a median survival time of 12-18 months. Immunotherapies have had limited success, due to immuno-suppressive effects by tumor cells, macrophages and microglia in the tumor microenvironment. We aim to study the functional phenotypes of microglia cells in healthy and disease environments to advance understanding of their role.

Introduction

The blood-brain barrier

The brain is separated from peripheral blood flow by the blood-brain-barrier (BBB). The BBB is a system of interconnected endothelial cells and astrocytes that separate peripheral blood from the cerebro-spinal fluid of the brain, and allow exchange of a tightly-controlled set of nutrients and cell types (Fig. 1). The BBB also restricts passage of many pathogens that circulate in the blood.

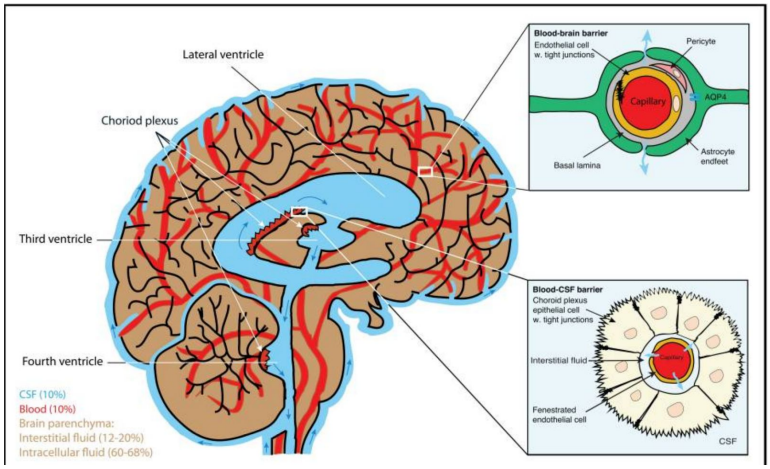

Figure 1: The blood-brain barrier. From Jenssen et. al 2015.

Microglia: resident immune cells of the brain

When cellular debris accumulates in the brain from normal or degenerative function, or pathogens manage to enter the brain, the first line of defense is not peripheral immune cells such as macrophages and neutrophils, it is an immune cell unique to the brain called a microglial cell. Microglia have a diverse array of receptors that detect extracellular ATP, neurotransmitter, glucose, pathogen-associated molecular patterns, and neuronal computation associated ions such as sodium and potassium. They can phagocytize debris and pathogens, can migrate through the brain, and can change their morphology to suit their function (Fig. 2). Microglia are implicated in brain tissue repair, neural homeostasis, antigen presentation to peripheral T cells, and angiogenesis.

Microglia are similar to peripheral macrophages, for example they can both phagocytose, and both serve as antigen-presenting cells. However they originate from different places during development - macrophages come from self-renewing stem cells in the bone marrow, while microglia migrate to the brain from the yolk sack at an early embryonic stage, and continually self-renew in the brain throughout a lifespan.

Although their functional immune phenotypes can overlap at the level of gene expression, it's likely that microglia have a more diverse phenotypic repertoire than macrophages. Peripheral macrophages are often classified as having one of two phenotypes: an M1 phenotype which is pro-inflammatory, and an anti-inflammatory M2 phenotype which promotes tissue repair. The M1/M2 phenotyping scheme has also been applied to microglia, however there is significant evidence that microglia can exhibit a more diverse set of phenotypes than a macrophage. For example, M1 vs M2 gene expression profiles can overlap in the same microglia (Ransohoff 2016).

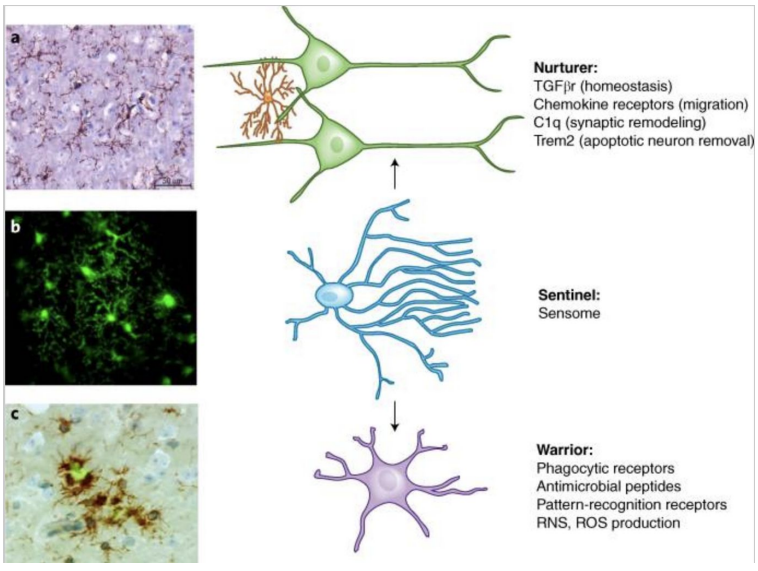

Figure 2: Functional roles of microglia. From Hickman et. al 2018.

Glioblastoma

Glioblastoma is a deadly and nearly intractable form of brain cancer, and accounts for almost half of all malignant brain tumors (brain-tumor.org). Patients have a median survival time of 12-18 months.

One feature of glioblastoma that makes it very difficult to treat is the heterogeneity of its cells. Wang et. al (2019) studied neoplastic cell heterogeneity using a variety of techniques. They applied scRNA-seq to roughly 30,000 cells extracted from 19 tumor samples. Through clustering, cell markers, and identification of mutations, they found roughly two-thirds of those cells to be neoplastic.

By quantifying RNA velocity, analyzing scATAC-seq and DNAseq data for some tumors, and applying lineage tracing, Wang et. al (2019) found a hierarchical axis of heterogeneity. At one extreme of the axis were mesenchymal-like cells that were mostly quiescent, present predominantly in the tumor core, and infiltrated with immune cells. At the other end of the axis were pro-neural cells that were actively cycling and present more often at the tumor's leading edge. These pro-neural neoplastic cells were more like oligodendrocytes, while the mesenchymal neoplastic cells expressed markers for astrocytes. In the middle of the axis were cells that exhibited markers of both types, but RNA velocity analysis showed their expression to gravitate from mesenchymal to pro-neural, not in the other direction. Further work by Couturier et. al (2020) supported these findings, and investigated the role of glioma stem cells in tumors and their relationship to embryonic neural development.

Microglia and glioblastoma

Glioblastoma is thought to be an immunologically "cold" tumor; the number of infiltrating cytotoxic T lymphocytes is low and treatments with immunotherapies have not been very successful (Fig. 3, Lim et. al 2018). However, an scRNA-seq study from several tumors has shown that almost half of the necrotic core is composed of non-mutated myeloid immune cells, both peripheral macrophages and native microglia (Darmanis et. al 2017).

Investigations into why resident microglia and infiltrating peripheral macrophages fail to destroy gliomas reveal some common patterns. Reviews by Poon et. al (2017) and Matias et. al (2018) break immuno-suppressive properties of glioblastoma into four main factors. First, microglia and macrophages are recruited to the tumor by chemokines such as CCL2. Second, they are manipulated to express a pro-growth, anti-inflammatory phenotype by cytokines such as IL-6 and TGF-beta. Third, their pro-growth phenotypes promote angiogenesis by producing growth factors such as VEGF. Last, their debris-removing capabilities are harnessed when microglia release matrix-metalloproteases (MMPs) that destroy the surrounding extracellular matrix and allow the tumor cells to proliferate and spread.

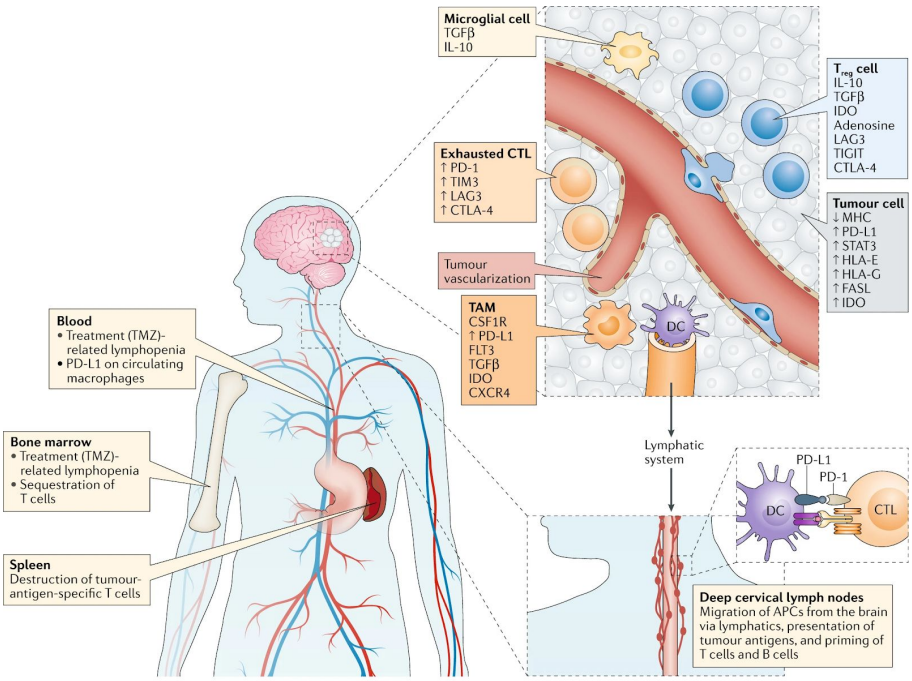

Figure 3: Immunosuppression by glioblastoma. From Lim et. al (2018).

Aims

Many of the works referenced here have sought to characterize the characteristics of neoplastic cells in glioblastoma through transcriptome analysis. In that process they have also identified other cell types in their samples, including neurons, astrocytes, and immune cells, and performed identification and basic analysis of the transcriptomes of microglia cells. Darmanis et. al 2015 & 2017 have released their datasets into the public domain. In Lau et. al (2020), not discussed, they analyze glial cells and release a large open access transcriptome dataset that contains healthy and disease cells for patients with neuro-degenerative diseases.

This work aims to extend analysis and go deeper into the transcriptomes of healthy microglia vs those in tumor microenvironments, and tangentially, neuro-degenerative diseases. The goal is to quantitatively define functional phenotypes of microglia across these datasets. A functional phenotype of microglia should encompass several domains:

- Motility: microglia can migrate.
- Morphology: microglia can change shape and extend processes.
- Sensory receptor expression
- Cytokine/chemokine production
- Phagocytosis and Antigen presentation
- Metabolism

The following set of aims is proposed to quantitatively describe functional phenotypes of microglia in a human-understandable way.

Aim 1: Robust algorithm to identify microglia from scRNA-seq data

Three datasets will be utilized in the analysis for this project:

1. Darmanis 2015: single-cell RNAseq on 466 cells, from fetal and healthy human brains.
2. Darmanis 2017: single-cell RNAseq on 3589 tumor cells from four subjects with glioblastoma.
3. Lau 2020: single-nucleus RNAseq on 169,496 cells from pre-frontal cortex of Alzhiemers patients and healthy controls.

Each of these datasets claims to contain microglia samples. In this aim, unsupervised dimensionality reduction, for example PCA, tSNE, or UMAP, will be applied to the datasets. There are several packages out there, and default parameters often produce good results (Raimundo et. al 2020). Once the data is dimensionally reduced, an unsupervised clustering algorithm will be applied. Visualization and heuristics will be used to determine hyperparameters of the clustering, such as the number of clusters.

Similar to what is done in the literature, known marker genes of microglia will then be used to identify putative microglial transcriptomes. Lau et. al (2020) described three - CD3, LRMDA, and DOCK8. Darmanis et. al (2015) used an unbiased approach and enrichment analysis to define the following cluster of additional microglia markers: CD53, ITGAX, OLR1, FCGR3A, IL1B, CCL3, CCL4, LILRB4, C3AR1, LAPTM5, DHRS9, PLEK, ALOX5AP, CD74, MSR1, PTPRC, GPR183, RGS10, C1QA, LCP1. See Fig. 4 for an example of unsupervised clustering.

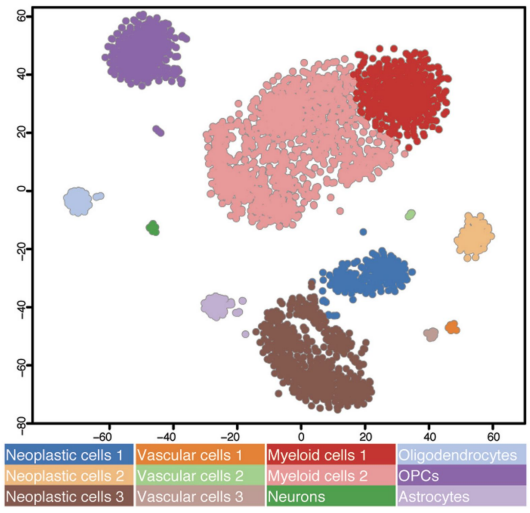

Figure 4: 2D tSNE and k-means clustering on glioblastoma data. From Darmanis et. al 2017.

Aim 2: Identify regulatory networks within each microglial cell

The expression of protein-coding genes is often controlled by one or more transcription factors, proteins that have DNA binding domains and whose purpose is to promote transcription of their downstream targets. A gene regulatory network is a set of transcription factors that work together to promote or inhibit sets of genes.

There are potential benefits to analyzing a gene regulatory network instead of directly analyzing the full gene expression vector. First, it reduces the dimensionality of the transcriptome, which reduces complexity and noise during computation. Second, understanding a cell at the level of it's transcriptional programs rather than enumerating sets of genes is potentially more powerful for gaining intuition about the cell's functional phenotype.

Although there are several software packages available for detecting regulatory networks, there appear to be some potential difficulties identifying networks from transcriptome data. The count data may be noisy, and evidence has accumulated that cells exhibit "transcriptional bursting", which means transcription of genes does not occur at a steady state, but in bursts. The counts of transcription factors and their influenced genes may not be in sync when the transcriptome of a cell is sequenced. A recent review of methods for gene regulatory network identification found that they did not perform very well with respect to ground truth data (Chen and Mar, 2018).

Nevertheless, gene regulatory network identification algorithms may still produce useful and meaningful output. To provide an example of how they work, an algorithm called SCENIC can be described (Aibar et. al 2017). The SCENIC algorithm produces a correlation matrix of gene expression across a group of transcriptomes by predicting the counts of one gene from the other genes. The correlation matrix is then segmented to produce groups of co-expression networks. Next, in a filtering step, the binding pattern of a transcription factor is compared with upstream regions of the genes within it's co-expression network. This step is used to remove edges in co-expression graphs that are less-likely to be real. There is then a subsequent cell scoring step and the regulatory networks can be clustered.

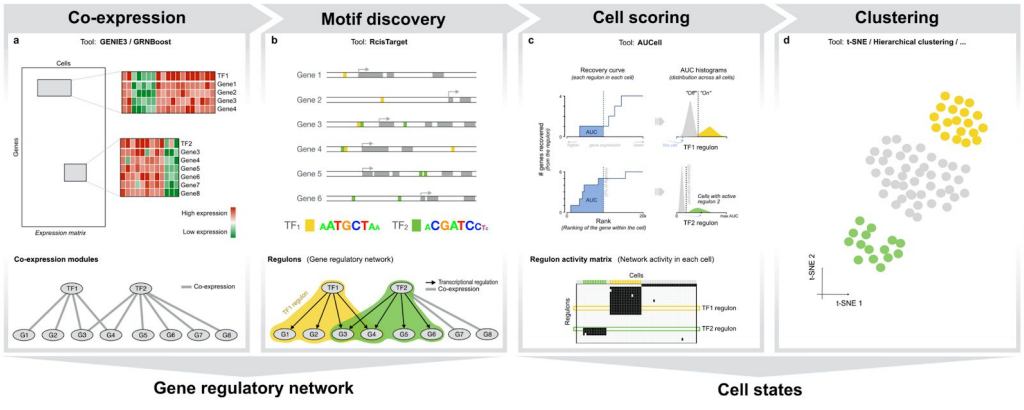

Figure 5: Schematic of SCENIC for identifying gene regulatory networks. From Aibar (2017).

Aim 3: Quantify ontologies from regulatory networks and cluster

Annotations of functional roles for genes by experts is an essential step in understanding the phenotypes of cells. The GeneOntology (geneontology.org) is a structured database that can provide a hierarchical graph of functional annotations for a given set of genes (Fig 6.). It utilizes statistical tests to compare the frequency of occurrences for elements in the hierarchy so that functional roles for sets of genes can be identified with some confidence.

GeneOntology provides three levels of granularity (Hill et. al 2008). The first is cellular component - where in the cell a gene affects function. The second is molecular function - what is the molecular mechanism of the gene's product. The third level is biological process - what high level function is being performed by the gene products. The level of biological process is the level at which microglia functional phenotypes will be evaluated.

Programmatic access to the GeneOntology is provided by software such as GOATOOLS (Klopfenstein et. al 2018), and refinement of results to remove redundancy can be achieved by using Revigo (Supek et. al 2011).

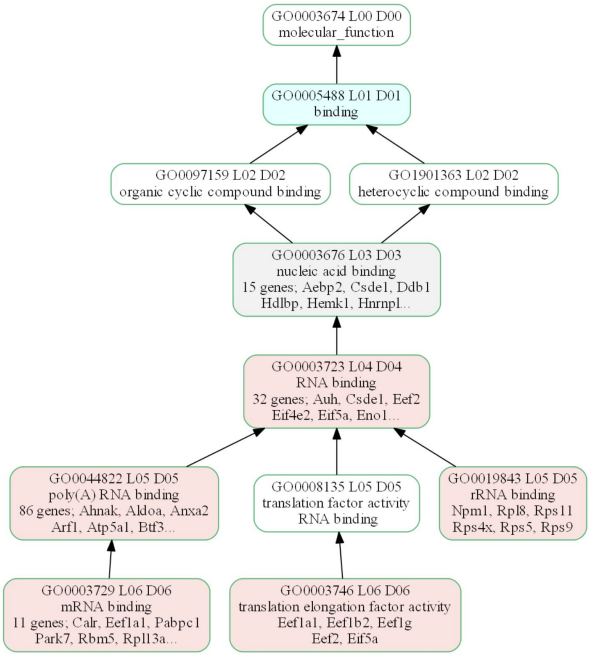

Figure 6: An example gene ontology graph. From GOATOOLS.

The final aspect of this analysis will be to utilize the functional annotations retrieved from regulatory networks obtained in aim 2, for the purposes of clustering functional phenotypes. This aspect might be challenging and require natural language tools if there are many annotations. One particular NLP tool that may be of interest is topic modeling, which traditionally is used to model each documents in a large corpus with a small set of descriptive terms.

References

Aibar, S., González-Blas, C. B., Moerman, T., Imrichova, H., Hulselmans, G., Rambow, F., ... & Aerts, S. (2017). SCENIC: single-cell regulatory network inference and clustering. *Nature methods*, *14*(11), 1083-1086.

Chen, S., & Mar, J. C. (2018). Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. *BMC bioinformatics*, *19*(1), 1-21.

Couturier, C. P., Ayyadhury, S., Le, P. U., Nadaf, J., Monlong, J., Riva, G., ... & Petrecca, K. (2020). Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy. *Nature communications*, *11*(1), 1-19.

Darmanis, S., Sloan, S. A., Zhang, Y., Enge, M., Caneda, C., Shuer, L. M., ... & Quake, S. R. (2015). A survey of human brain transcriptome diversity at the single cell level. *Proceedings of the National Academy of Sciences*, *112*(23), 7285-7290.

Darmanis, S., Sloan, S. A., Croote, D., Mignardi, M., Chernikova, S., Samghababi, P., ... & Quake, S. R. (2017). Single-cell RNA-seq analysis of infiltrating neoplastic cells at the migrating front of human glioblastoma. *Cell reports*, *21*(5), 1399-1410.

Hickman, S., Izzy, S., Sen, P., Morsett, L., & El Khoury, J. (2018). Microglia in neurodegeneration. *Nature neuroscience*, *21*(10), 1359-1369.

Hill, D. P., Smith, B., McAndrews-Hill, M. S., & Blake, J. A. (2008, April). Gene Ontology annotations: what they mean and where they come from. In *BMC bioinformatics* (Vol. 9, No. 5, pp. 1-9). BioMed Central.

Jessen, N. A., Munk, A. S. F., Lundgaard, I., & Nedergaard, M. (2015). The glymphatic system: a beginner's guide. *Neurochemical research*, *40*(12), 2583-2599.

Klopfenstein, D. V., Zhang, L., Pedersen, B. S., Ramírez, F., Vesztrocy, A. W., Naldi, A., ... & Tang, H. (2018). GOATOOLS: A Python library for Gene Ontology analyses. *Scientific reports*, *8*(1), 1-17.

Lau, S. F., Cao, H., Fu, A. K., & Ip, N. Y. (2020). Single-nucleus transcriptome analysis reveals dysregulation of angiogenic endothelial cells and neuroprotective glia in Alzheimer's disease. *Proceedings of the National Academy of Sciences*, *117*(41), 25800-25809.

Lim, M., Xia, Y., Bettegowda, C., & Weller, M. (2018). Current state of immunotherapy for glioblastoma. *Nature reviews Clinical oncology*, *15*(7), 422-442.

Matias, D., Balça-Silva, J., da Graça, G. C., Wanjiru, C. M., Macharia, L. W., Nascimento, C. P., ... & Moura-Neto, V. (2018). Microglia/astrocytes–glioblastoma crosstalk: crucial molecular mechanisms and microenvironmental factors. *Frontiers in cellular neuroscience*, *12*, 235.

Poon, C. C., Sarkar, S., Yong, V. W., & Kelly, J. J. (2017). Glioblastoma-associated microglia and macrophages: targets for therapies to improve prognosis. *Brain*, *140*(6), 1548-1560.

Raimundo, F., Vallot, C., & Vert, J. P. (2020). Tuning parameters of dimensionality reduction methods for single-cell RNA-seq analysis. *Genome biology*, *21*(1), 1-17.

Ransohoff, R. M. (2016). A polarizing question: do M1 and M2 microglia exist?. *Nature neuroscience*, *19*(8), 987.

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS one*, *6*(7), e21800.

Wang, L., Babikir, H., Müller, S., Yagnik, G., Shamardani, K., Catalan, F., ... & Diaz, A. A. (2019). The phenotypes of proliferating glioblastoma cells reside on a single axis of variation. *Cancer discovery*, *9*(12), 1708-1719.