Differential Gene Expression in Brain Metastasis

Siddhu Bhimireddy
Western Washington University
USA

Andrew Cary
Western Washington University
USA

Aidan Clements
Western Washington University
USA

Carver Rasmussen
Western Washington University
USA

Abstract

Microglia are important macrophages that work to clean up dead cells, misfolded proteins, and fight cancerous growths inside the brain. Different extracellular matrix environments appear to have an effect on microglia's ability to accomplish these tasks. Dr. Snyder's project tested four different extracellular matrix environments on microglia and measured its ability to destroy cancer cells. We assisted on the RNA sequencing and gene feature count of the results of this experiment. Here, we study the differential gene expression between these extracellular environments to discover which signal pathways in microglia promote cancer defense.

1 Introduction

Dr. Snyder's project is on analyzing the extracellular matrix proteins that inhibit the microglia's ability to detect and destroy tumor cells within the brain. Microglia are the primary macrophage cells of the central nervous system, essentially the primary operators of the immune system in the brain and spine. They work to maintain homeostasis by keeping the brain 'clean' by removing dead cells, mis-folded proteins, and plaque buildup. For this reason, they are an important candidate for study to better understand brain health.

Different surrounding extracellular matrix (ECM) environments seem to induce different gene expression in the microglia. The levels of expression in a gene can deteremine what types of proteins are being built and what a cell's current function and priorities are. Discovering what gene pathways are being expressed can potentially explain what is inhibiting or enabling microglia's effective destruction of tumors in the brain. It is the purpose of this study to better understand what signaling pathways are engaged in microglia that have successfully eaten tumor cells and to characterize the ECM's effects on this process. In this study, all of the ECM's seemed to have a negative effect on phagocytosis, decreasing tumor intake by microglia. To better understand what caused this decrease is the main goal of Dr. Snyder's project.

2 Motivation

Along with the data uncovered in this project will come a better understanding of how to promote microglia's ability to "eat" cancer cells in the brain. Understanding what pathways are blocked when exposed to certain ECM's could explain what genes are expressed during phagocytosis. This knowledge can aid researchers in developing treatments that are more effective in promoting the body's immune system into destroying cancer cells and other neurodegenerative diseases. Better microglia therapies could potentially save lives and ease the suffering of many individuals with these neurodegenerateive diseases.

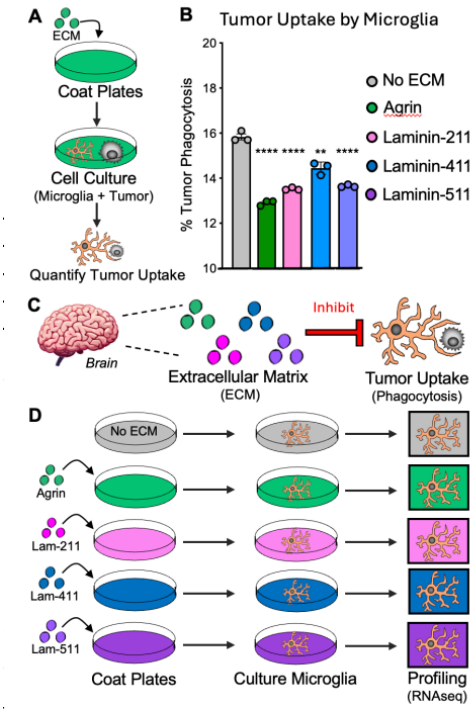

Figure 1: Initial findings of Dr. Snyder's experiment

3 Related Work

With the recent advancement in RNA sequencing technology, there is a growing interest in microglia among the scientific community as a means of promoting the brain's immune defense systems. While there is interest in microglia's ability to eat cancer, research is also being placed into its ability to prevent neurodegenerative diseases such as Alzheimers and Parkinsons [9]. Other research is in microglial replacement therapies as a way to repopulate and reinforce these macrophages in brains with neurodegenerative disease [12]. Any and all understanding of microglia's functions and

Bioinformatics'24, Dec 2024, Bellingham, WA

behavior in different environments can aid in these different areas of research.

4 Methods

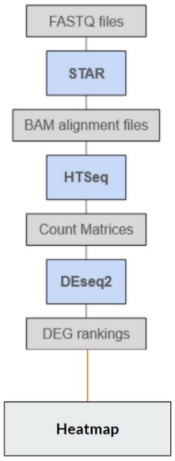

Figure 2: The four step process of our methodology

Our method of analysis was a 4 step process illustrated in Figure 2. We started with 46 separate FASTQ files that needed to be sequenced against the human genome. From there a unique feature count was taken in order to create the count matrices. Then, Deseq2 was used to determine the most influential genes and produce our heat maps.

4.1 Sequence Alignment

We started with FASTQ data files that contained the sequenced RNA of the microglia exposed to the different ECM's before and after measuring tumor uptake (a sample FASTQ file's contents are shown in Figure 3) [5, 6, 11]. These files need to be sequence-aligned against a reference human genome type in order draw any conclusions from the data. This was a computationally expensive task because of the size of the data needing to be aligned against the 46 fastq files. We used RNA STAR through the WWU HT-Condor cluster to align the data with the human genome[1, 3, 4]. The results were a SAM file for each FASTQ data file (sequence-aligned files).

```
Example FASTQ entries:

Read 1:
Header: @VH00126:363:2225GSKNX:1:1101:29365:1000 1:N:0:GACTGAGTAG+CACTATCAAC
Sequence length: 59 bases
Sequence: GCGCTGGGGTTGGCTCCATGACCAAGATCTATGGGGGACGTCAGAGAAACGGCGTCATG
Quality: IIIIIIIIIIIIIIIIIIIIIIIIIII9IIIIIIIIIIIIII9IIIII9II9IIIIII
--------------------------------------------------------------------------------

Read 2:
Header: @VH00126:363:2225GSKNX:1:1101:37185:1000 1:N:0:GACTGAGTAG+CACTATCAAC
Sequence length: 58 bases
Sequence: CTTCCGGTTCAATGCCAAGTATGCGGTGGAGCTGGAGGGGCAGCAGCCCATCAGCGTG
Quality: IIIIIIIIIII9IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
--------------------------------------------------------------------------------

Read 3:
Header: @VH00126:363:2225GSKNX:1:1101:40271:1000 1:N:0:GACTGAGTAG+CACTATCAAC
Sequence length: 57 bases
Sequence: CAGTCTCAACCCAGCCATCCCCCATCCCACACCAGCTGCAGCAACATTCTCTACACT
Quality: IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
--------------------------------------------------------------------------------

Total number of reads in file: 2,732,293
Each read is 57 bases long
Total bases in file: 155,740,701
```

Figure 3: Readings of the fastq files of sequenced RNA

4.2 File Conversion and Feature Counting

We converted the SAM files to BAM files, which are binary versions of the same data files. This shrinks the size of the files making them easier to process and less likely to produce errors. The next step was to run htSeq-count to generate the feature counts for the various gene types [8]. To do this, we needed to sort the BAM files. HtSeq-count runs significantly more efficiently and faster on sorted BAM files than unsorted BAM files. This is because sorted data allows htSeq-count to process reads mapped to similar regions in a single pass. HtSeq-count outputs .csv file for each control and each treatment (4 control .csv files and 2 treatment .csv files). These .csv files have columns corresponding to the replicates and rows corresponding to the gene being expressed. Each condition has 8 replicates (except for Agrin, which only has 6 due to an unfortunate contamination of data).

4.3 Data Analysis

We combine the .csv files into a large count matrix to run deSeq2 [2]. deSeq2 filters and processes the data; it outputs a table of useful statistics about the data. One of these is the dispersion. Essentially, deSeq2 calculates the variance for the expressed genes and fits a curve to it. Dispersion is calculated with the help of 'shrinkage', an estimation method that shares dispersion information between genes to improve the accuracy of the parameters for less expressed genes. Figure 4 shows a plot of the dispersion estimate. Our curve looks acceptable as it begins to plateau as there are more counts, indicating statistical consistency; had there been no inverse relationship, that would indicate an upstream issue with the data.

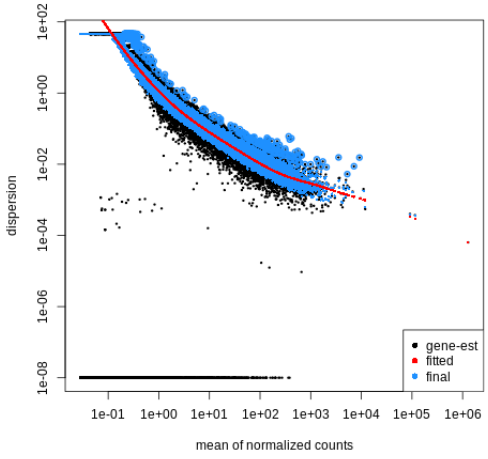

Figure 4: Dispersion estimates on the resulting data

4.4 Heatmap Generation

The final step was to plot the heatmaps. We want heat maps comparing the $\log_2$fold expression between each of the control/treatment pairs. This translates to 8 heat maps as we have 2 controls and 4 treatments. We first removed all gene expressions that had p-values

of > 0.05, deeming them statistically insignificant. Then we returned the 10 most differently expressed genes, and plotted them onto a heatmap.

5 Results

We've generated heatmaps that effectively show the levels of gene expression across Dr. Snyder's samples. There is much more work to be done building on the data presented here, but the heatmaps provide an effective visual framework to direct future inquiry into microglia growth. The following sections are separated by each individual ECM experiment.

5.1 Agrin Analysis

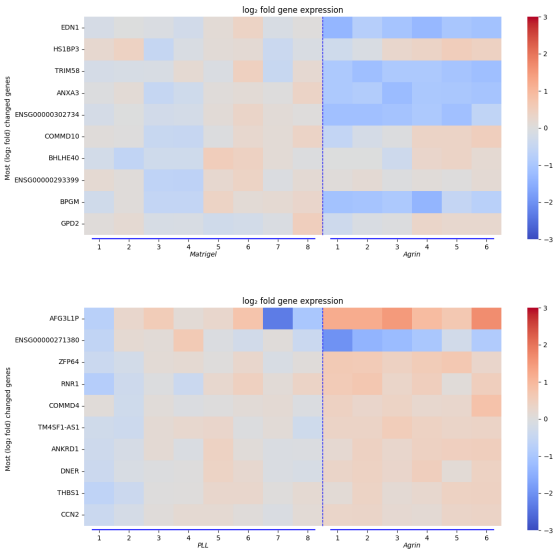

Figure 5: Agrin Analysis

No tremendous changes in gene expression across the samples. The highest differential expression is found in the AFG3L1P gene, and the novel ENSG00000271380 gene. AFG3L1P is implicated in mitochondrial protein processing. There is little connection between these changes and the microglia functions surrounding cancer intake.

5.2 Collagen I Analysis

Here we find THRA, a thyroid hormone regulation factor, is downregulated compared to the PLL control sample. STARD8 is also heavily downregulated, and is usually associated with upregulation in cancerous cells. SLC66A2, a trans lipid transport protein, is the only heavily differentially expressed gene compared to the matrigel control. There are no obvious standouts for pathways related to immune defense in this ECM either.

5.3 Pan-Laminin Analysis

Versus the PLL, Pan-Laminin shows the greatest overall level of differential expression, while showing the least relative to the matrigel

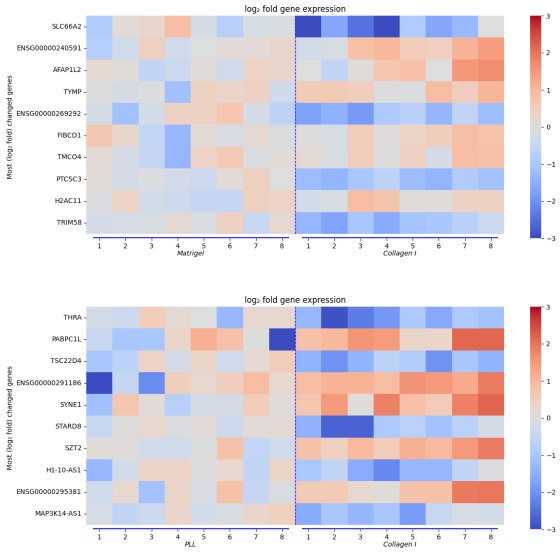

Figure 6: Collagen I Analysis

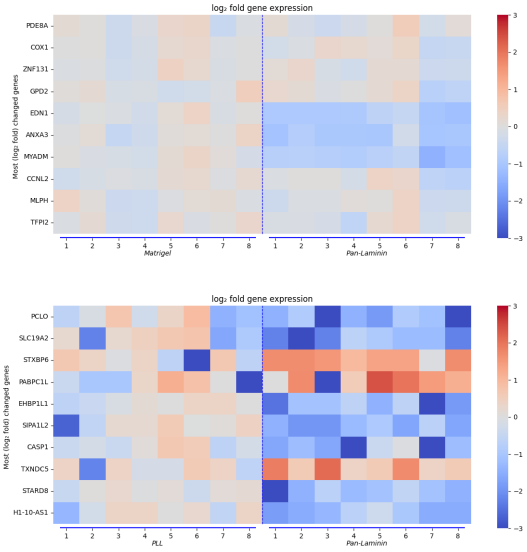

Figure 7: Pan-Laminin Analysis

control. CASP1 downregulation is the most immediately relevant factor so far, being a cleaving enzyme involved in cell apoptosis. SLC19A2 is involved in vitamin metabolism, PCLO is involved in establishing active synapse zones and both are heavily downregulated. STARD8 encodes a protein called Rho GTPase that is essential to cell movement and phagocytosis. STARD8 is decreased and this is consistent with the inhibited phagocytosis that occurred in the PLL experiment.

In the Matrigel experiment, very little changes in gene expression occurred. One gene that is decreased is MYADM and this gene is involved in plasma membrane reorganization during cell movement[10]. Cell movement is essential to phagocytosis and could be a key factor in the decreased tumor intake in this experiment.

5.4 Laminin Analysis

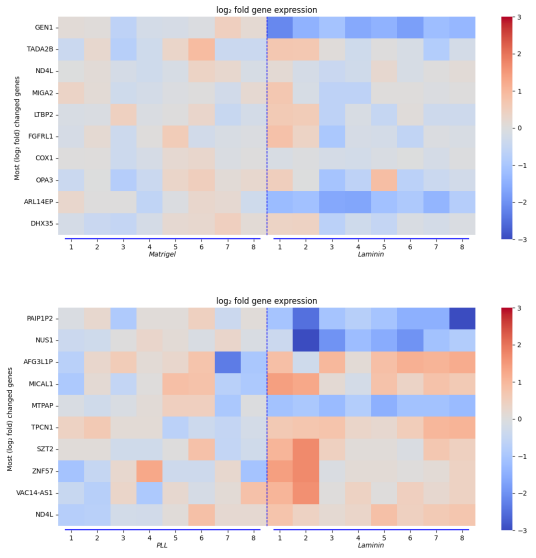

Figure 8: Laminin Analysis

PAIP1P2 and NUS1 are both downregulated. PAIP1P2 is a transcription inhibitor, while NUS1 is most well known for being linked to a specific form of genetically linked lung cancer. Compared to the matrigel, the GEN1 gene is downregulated. Gen1 is involved in the resolution of Holiday junctions, a specific 4-way structure found in DNA repair.

5.5 Final Results

First, the regulation of cytoskeletal components appears as a central theme, with MICAL1, MYADM, and STARD8 all showing expression patterns consistent with inhibited cell movement and phagocytosis[7]. Second, the modulation of immune response emerges as significant, with the decrease in ARL14EP and increase in MICAL1 suggesting a coordinated suppression of immune response capabilities[10]. Third, membrane dynamics show notable changes, as evidenced by MYADM's decrease in pan-laminin treatment, which aligns with reduced plasma membrane reorganization and supports the observed inhibited phagocytosis phenotype[10].

6 Discussion & Future Work

This team was expecting to see pathways related to inflammatory and anti-inflammatory responses expressed in ways relating to the tumor intake for the given ECM experiment. However, there were not any huge differences in many "classic" immune-related genes amongst the different ECM treatments. However there are some genes related to cell processes involved in healthy immune responses. These genes should be examined more closely to determine whether they have a relationship with the effectiveness of microglia in phagocytosis.

For future experiments, it would be helpful to test more unique ECM's to see if any stand out as especially effective. It would also be necessary to test and sequence the microglia at earlier and later times than this test did. This would ensure that the microglia have had time to process the ECM, attack tumors, and give researchers an accurate representation of the gene pathways that are expressed during those processes. We now have a subset of genes that are highly differentially expressed based on the composition of the extracellular matrix. Discovering exactly how the features these genes encode interact with tumors would be the next step in finding viable treatment options that involve the ECM and guide microglia growth.The high expression difference present in two novel gene transcripts likely warrants further research into these genes. Perhaps they hold the key to microglia's immune response triggers. There is a lot of analytical work to do that we were not able accomplish due to time restrictions, but we hope that this data processing will allow more experts to draw conclusions and discoveries from our work.

References

[1] Cornell University 2019. *STAR Usage Guide.* Cornell University. biohpc.cornell.edu/doc/RNA-Seq-2019-exercise1.html

[2] 2023. Differential gene expression analysis pipelines and bioinformatic tools for the identification of specific biomarkers. (2023). https://pmc.ncbi.nlm.nih.gov/articles/PMC10951429/

[3] Cornell University 2023. *STAR Manual.* Cornell University. https://physiology.med.cornell.edu/faculty/skrabanek/lab/angsd/lecture_notes/STARmanual.pdf

[4] HBC Training 2023. *STAR Training Guide.* HBC Training. https://hbctraining.github.io/main/

[5] NYU Gencore 2024. *FASTQ Format Guide.* NYU Gencore. https://learn.gencore.bio.nyu.edu/ngs-file-formats/fastq-format/

[6] 2024. *FASTQ Python Package.* https://pypi.org/project/fastq/

[7] Juan F Aranda, Natalia Reglero-Real, Leonor Kremer, Beatriz Marcos-Ramiro, Ana Ruiz-Sáenz, María Calvo, Carlos Enrich, Isabel Correas, Jaime Millán, and Miguel A Alonso. 2011. MYADM regulates Rac1 targeting to ordered membranes required for cell spreading and migration. *Molecular Biology of the Cell* 22, 8 (2011). https://doi.org/10.1091/mbc.e10-11-0910

[8] A Dobin et al. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29 (2013), 15–21.

[9] Chao Gao, Jingwen Jiang, Yuyan Tan, and Shengdi Chen. 2023. Microglia in neurodegenerative diseases: mechanism and potential therapeutic targets. *Signal Transduction and Targeted Therapy* 8, 1 (2023). https://www.nature.com/articles/s41392-023-01588-0

[10] Byung Cheon Lee, Zalan Peterfi, FuKun W Hoffmann, Richard E Moore, Alaattin Kaya, Andrei Avanesov, Lionel Tarrago, Yani Zhou, Eranthie Weerapana, Dmitri E Fomenko, Peter R Hoffmann, and Vadim N Gladyshev. 2013. MsrB1 and MICALs regulate actin assembly and macrophage function via reversible stereoselective methionine oxidation. *Molecular Cell* 51, 3 (2013), 397–404. https://doi.org/10.1016/j.molcel.2013.06.019

[11] M I Love, W Huber, and S Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15 (2014), 550.

[12] Lijuan Zhang, Yafei Wang, Taohui Liu, Ying Mao, and Bo Peng. 2023. Novel Microglia-based Therapeutic Approaches to Neurodegenerative Disorders. *Current Neuropharmacology* (2023). https://pmc.ncbi.nlm.nih.gov/articles/PMC10043109/