Analysis of scRNA-seq data of mouse myeloid cells in tumor-bearing cerebella ± radiation treatment to identify differential gene expression

Kaitlyn Terrell & Kristina Ivanov

Abstract:

This project explores the interplay between myeloid cells in both healthy and tumor microenvironments, focusing on medulloblastoma. Myeloid cells, composed of monocytes, macrophages, dendritic cells, and others, hold important roles in maintaining homeostasis and supporting tumor growth. Tumors take advantage of myeloid cells, especially within the context of the immunosuppressive microenvironment. During the analysis, the primary focus was to investigate potential differentially expressed genes among distinct sample groups, aiming to uncover potential therapeutic targets for disrupting the tumor microenvironment. Using the Seurat analysis platform and Gene Ontology term analysis (GO-term), we examine differentially expressed genes to find genes that might contribute to medulloblastoma and other genes that promote myeloid cells that foster a healthy environment while undermining the tumor microenvironment. After analyzing the data, there are potential opportunities for therapeutic targets and offer insights into how myeloid cells may affect medulloblastoma.

Introduction:

During this project, the primary focus was to understand the function of myeloid cells in a healthy environment versus myeloid cells in a tumor microenvironment. Myeloid cells are composed of immune cells like monocytes, macrophages, myeloid dendritic cells, and cells that originate from a common myeloid progenitor in the bone marrow. Myeloid cells play a vital role in a healthy functioning environment, but also play a vital role in tumor microenvironments. They are most abundantly available in tumor microenvironments. Tumors are able to recruit the myeloid cells to tumor-associated macrophages, dendritic cells, and neutrophils to sustain an immunosuppressive environment (1). Throughout the project, we will be focusing on cancerous tumors specifically in the cerebellum, which is known as medulloblastoma. Medulloblastoma can

occur at any age, but it develops most often in young children (2). The symptoms that people may experience when living with medulloblastoma are tiredness, nausea, poor coordination, unsteady walking, headaches, etc. There are some treatments for medulloblastoma which include surgery to remove the tumor, radiation therapy, and chemotherapy (2). It would be interesting to understand if there was a possibility to increase myeloid cells that promote a healthy environment in order to combat the tumor microenvironment. In this project, the main goal is to understand if there are differentially expressed genes between the different sample groups. It would be interesting to see if these genes could be a potential therapeutic target for degrading the microtumor environments in medulloblastoma.

Methods:

Our dataset consists of six total samples harvested from Male Mus musculus of the Ptch1+/-p53-/- genotype where sonic hedgehog subgroup of medulloblastoma (SHH-MB) are known to develop in the cerebellum by 8 weeks of age (3). Our control sample (GSM5077857) is pooled CD11+ myeloid cells harvested from 3, 2-week old mouse cerebella, prior to SHH-MB development. We have two biological replicates of CD11+ myeloid cells harvested from 8-week old untreated mouse tumor-bearing cerebella (GSM5077859, GSM5077860). And three biological replicates harvested from 8-week old mouse tumor-bearing cerebella treated with radiation (GSM5077863, GSM5077864, GSM5077866) (3).

During the analysis, the main package used to analyze the data is the Seurat package (v.4.3.0.1) (4). The first step of the analysis is running quality control (QC) on the dataset to make sure that we remove any unwanted samples that have high mitochondrial contamination. Using the PercentageFeautureSet() function that is a part of the Seurat package which is able to calculate the percentage of counts that have a set of features that we defined. The pattern that is chosen in the function is “-MT”, which indicates that any genes starting with this pattern are mitochondrial genes, which we do not want in our final dataset. To visualize how the QC metrics

look, we used the VlnPlot() function to see the distribution of the features. We can see that there are no mitochondrial genes in any of our samples (Figure 1).

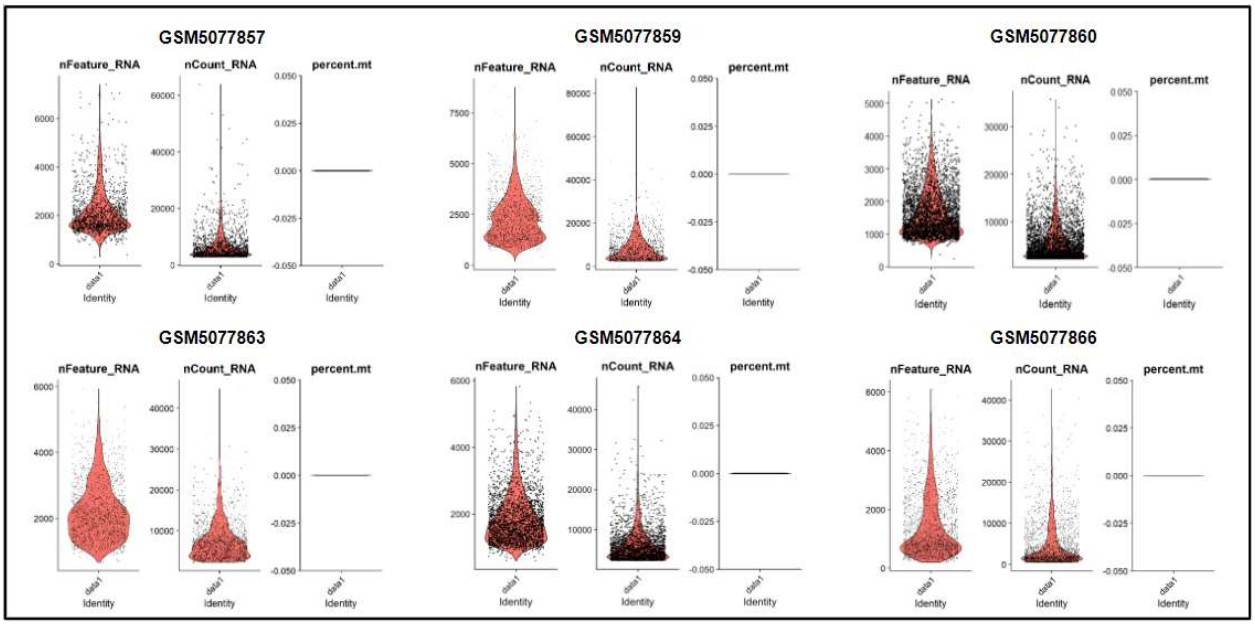

Figure 1. Results of Quality Control (QC) to remove mitochondrial genes in the dataset. GSM5077857 is the control Pooled Cd11+ myeloid cells from 3 mouse cerebella, MG, 2 week old Mus Musculus. GSM5077859 and GSM5077860 are biological replicates of CD11+ myeloid cells from tumor-bearing cerebellum untreated in 8 week old Mus Musculus. GSM5077863, GSM5077864, and GSM5077866 are 3 biological replicates of CD11+ myeloid cells from tumor-bearing cerebellum treated with radiation in 8 week old Mus Musculus.

After, using the subset() feature we filtered through the data with different criteria to make sure that our data is clean moving forward. The data was normalized by using the normalization method “LogNormalize" with a scale factor of 10000. This allowed us to calculate the gene variation between the groups. Next, we used principal component analysis (PCA), a procedure that allows for summarization of the dataset information by a smaller set of indices, so that it can be more easily visualized and analyzed (5). In PCA analysis UMAP, the control is clustered mainly to the left of PC_1 whereas the treated and untreated and mainly clustered on the opposite side of PC_1(Figure 2). This shows that there is significance in the PC_1 values when comparing the untreated and treated to the control. An interesting note is that one of the

samples that had radiation treatment (GSM5077864), is starting to have a PC_1 trend more similar to the control where there is a cluster by the 0 on PC_1 (Figure 2). Finally, we performed a GO-term enrichment analysis using hypergeometric tests on the differentially expressed genes of each sample to better understand the biological significance of the data found. The p-values of the enriched GO-terms were compared between all six samples and visualized in a barplot (Figure 5).

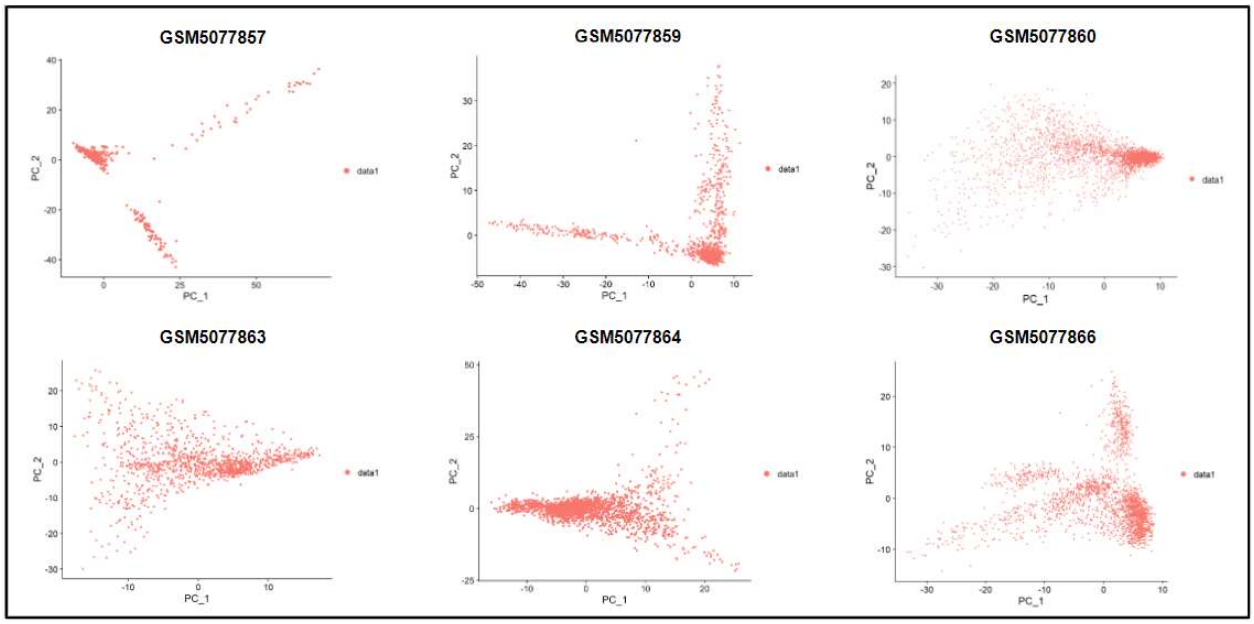

Figure 2. Results of PCA Analysis Dim plot. GSM5077857 is the control, GSM5077859 and GSM5077860 are biological replicates of untreated tumor-bearing, GSM5077863, GSM5077864, and GSM5077866 are biological replicates of radiation treated tumor-bearing.

Results:

The normalized data was clustered and visualized in a UMAP to show the clusters, the number of which vary between samples (Figure 3). From the clustered data, we found the log2foldchange for each datapoint and considered genes with two highest p-values for each sample, producing a dataframe for differentially expressed genes in each sample. The variations observed in UMAP results between different treatment groups can be a result of a combination of biological, technical, and experimental factors. These differences may come from distinct

biological responses induced by the treatments, potentially reflecting heterogeneous cellular compositions within the tumors.

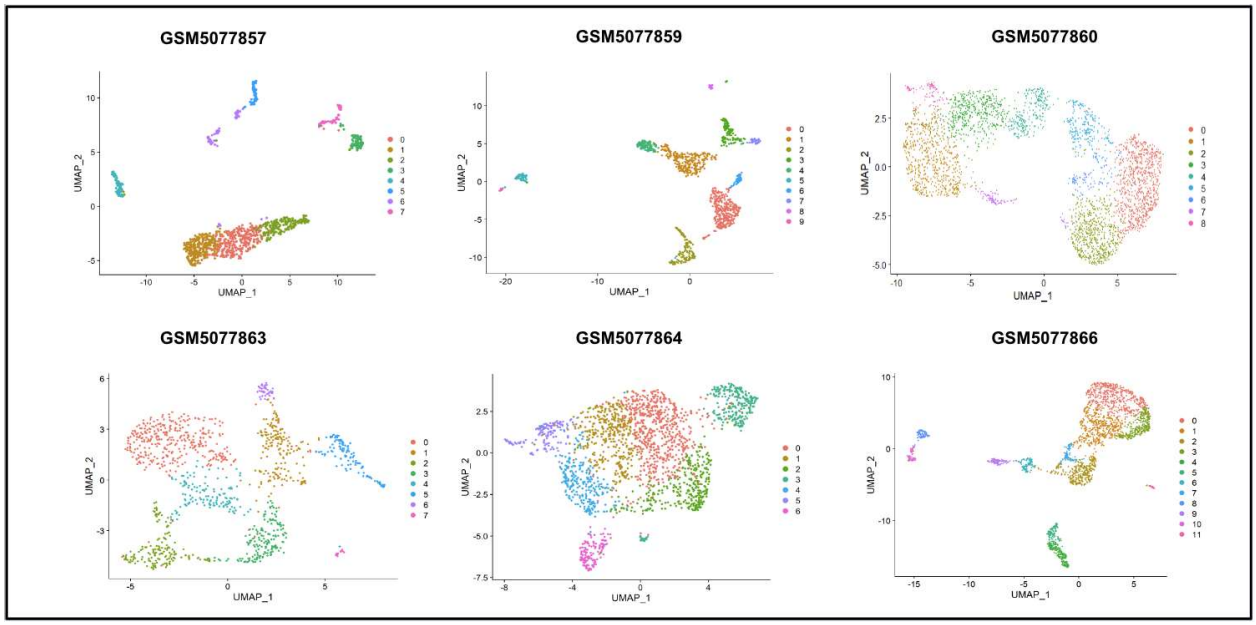

Figure 3. Results of UMAP Clustering. GSM5077857 is the control, GSM5077859 and GSM5077860 are biological replicates of untreated tumor-bearing, GSM5077863, GSM5077864, and GSM5077866 are biological replicates of radiation treated tumor-bearing.

From the lists of differentially expressed genes, we were able to produce venn diagrams and upset plots comparing DEGs between samples. First, we considered the common genes between biological replicates, and then the most common genes were then compared between treated, untreated, and control samples.

When comparing differentially expressed genes between biological replicates of radiation treated tumor-bearing samples, we found that there was only one shared gene between all three samples. However, there were 9 shared genes between samples GSM5077863 and GSM5077864 (Figure 4B). We were able to take this list of 9 genes and compare them against significant DEGs found in the control and untreated samples to see if there is upregulation or downregulation of certain genes.

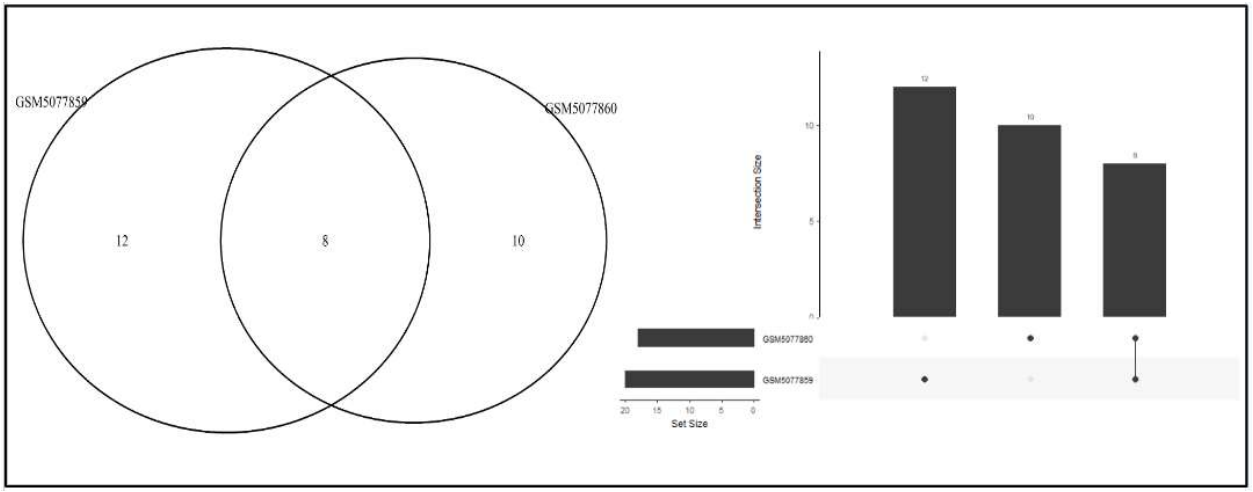

Figure 4A. Venn diagram and upset plot of similar differentially expressed genes between biological replicates of radiation treated tumor-bearing samples GSM5077859 and GSM5077860. Differentially expressed genes shared between GSM5077859 and GSM5077860 include: Cst7, Apoc1, Ccl4, Plac8, H2-Eb1, S100a8, S100a9, Tuba1b.

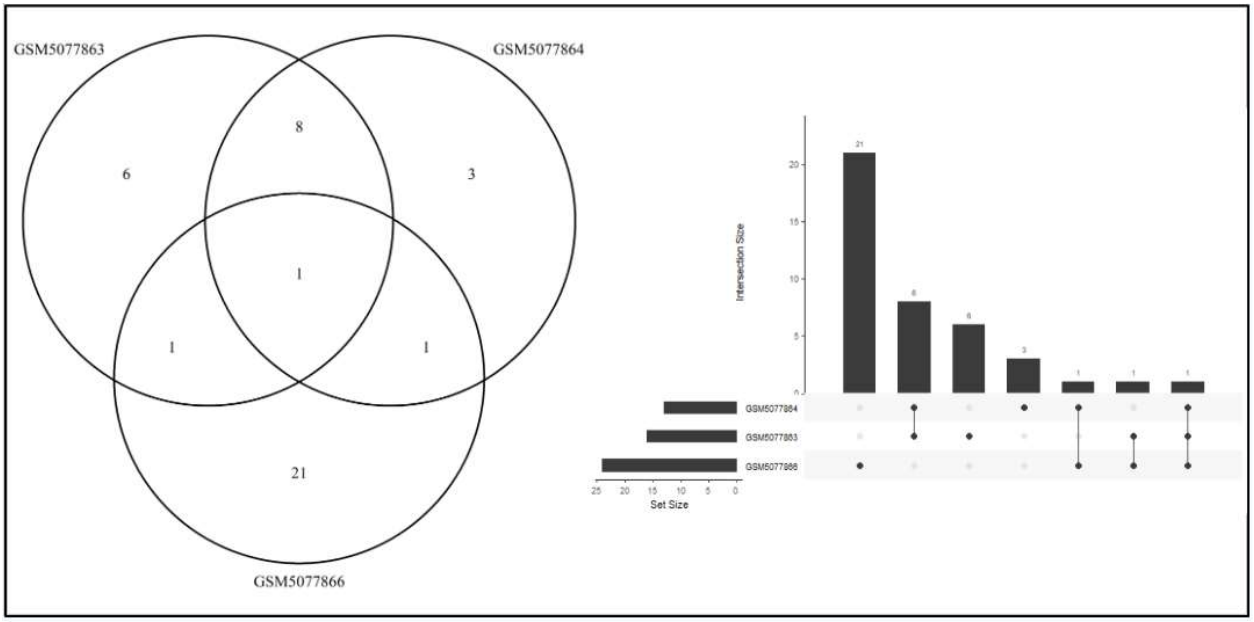

Figure 4B. Venn diagram and upset plot of similar differentially expressed genes between biological replicates of radiation treated tumor-bearing samples GSM5077863, GSM5077864, and GSM5077866. Differentially expressed genes shared between GSM5077863 and GSM5077864 include: Chil3, Apoc1, Cd81, S100a8, S100a9, Spp1, Gpnmb, Ccl5, H2-Eb1.

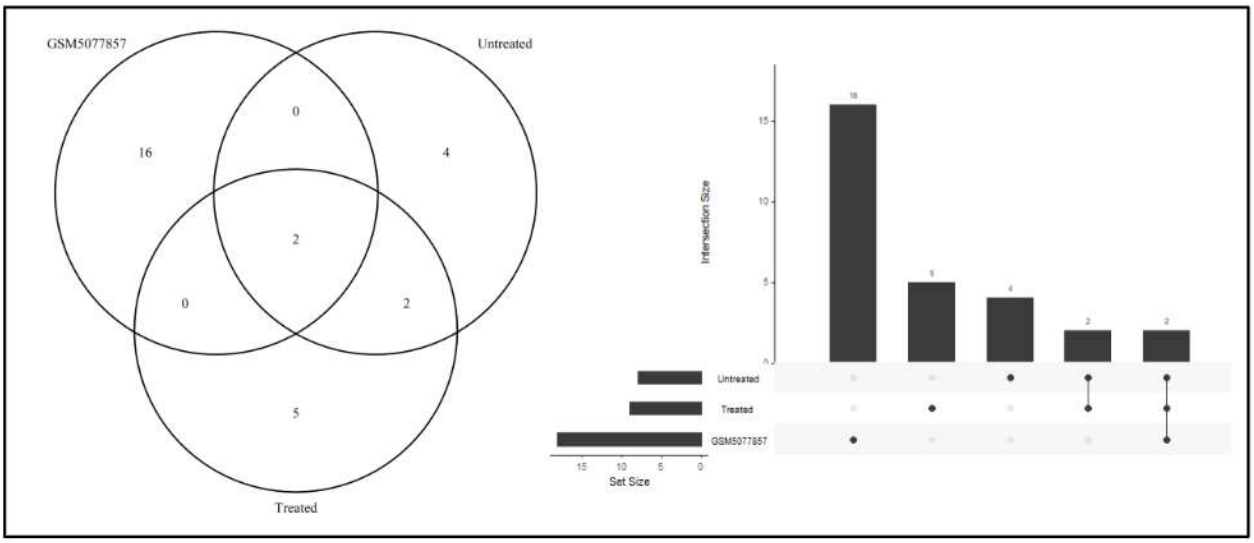

Figure 4C. Venn diagram and upset plot of similar differentially expressed genes between radiation Treated (from GSM5077863 and GSM5077864 shared DEGs), Untreated (from GSM5077859 and GSM5077860 shared DEGs), and Control (GSM5077857 samples. DEGs between all three include: S100a8, s100a9. DEGs shared between treated and untreated include: Apoc1, H2-Eb1.

Analyzing the comparison among untreated, treated, and control samples depicted in Figure 4C, we observe from the upset plot that there are two samples exhibiting overlap with both control and treated/untreated samples, along with two genes exclusively shared between treated and untreated groups. The two differentially expressed genes shared by the control and other samples are S100a8 and s100a9, which for the project's purposes may be ignored. The two remaining genes shared between treated and untreated samples, H2-Eb1 and Apoc1, may be used to determine which genes are differentially expressed and not shared between the two.

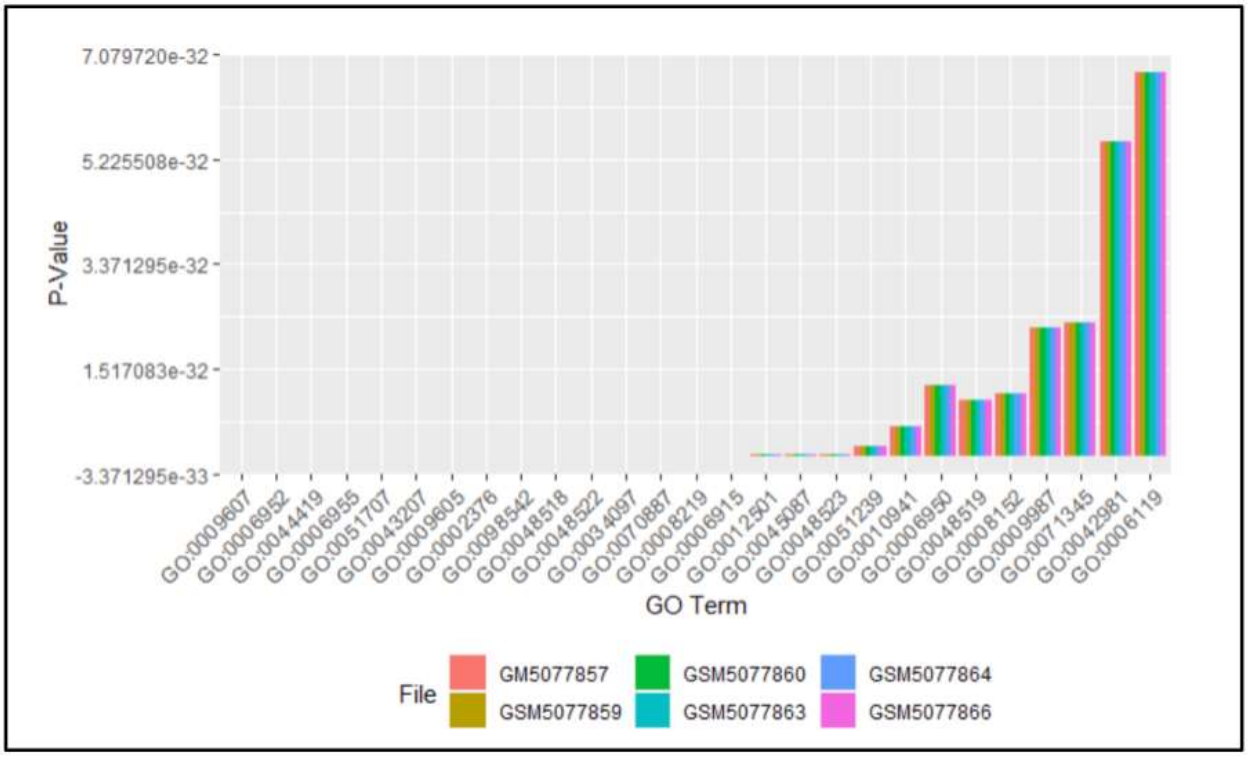

Figure 5. Gene Ontology Enrichment (GO-Term Enrichment) plot. Showing the top 30 Go-terms within the dataset across all six samples.

When comparing the enriched GO Terms between the six different sample sets, we found that the samples with the treatment and tumor have minimal impact on the gene functions (Figure 5).This result could also be an indication that the dataset? was too small, which would make the enrichment analysis difficult to identify unique GO Terms for each group. The similarity of the GO Terms can also indicate that the samples might have biological processes or molecular functions in common or consistent among the different groups. This would indicate that the different groups have similar underlying biological functions. The enrichment levels were high in these GO terms compared to the rest of the GO-terms that ranged anywhere between 9E-04 to 2E-32. We attempted to produce pie charts to visualize the results of the GO-term enrichment analysis but there were too many terms to express in one chart.

Discussion:

After performing a final comparison between the treated and untreated samples, we found four differentially expressed genes Cst7, Ccl4, Plac8, Tuba1b, only in untreated tumor-bearing samples in two replicates. We found five Differentially expressed genes Gpnmb, Spp1, Chil3, Cd81, Ccl5, only in radiation treated tumor-bearing samples in two replicates. Additionally, there were 16 differentially expressed genes found only in controls. The identification of differentially expressed genes (DEGs) in radiation-treated tumor-bearing samples presents a potential target for understanding the dynamics between gene expression and tumor response to radiation therapy. The presence of Gpnmb, Spp1, Chil3, Cd81, and Ccl5 in these samples suggests their potential involvement in radiation-induced tumor regression or therapeutic response. We also found that there are 16 DEGs found only in the controls. Investigating why these genes are missing from the tumor bearing samples could reveal whether the genes act as tumor suppressors or if they play a role in starting tumors. Exploring the reasons behind their absence might help us understand important pathways that could stop tumors from forming.

Conclusion:

From our found differentially expressed genes, it appears that none of the genes are expressed in only treated or untreated samples. This suggests that if genes significant in pre or post-radiation treatment their expression levels are lower than the p-value cutoff we set in our analysis, or that they were not included in the limited list of DEGs we used in our comparison (since we only took two genes per cluster with the highest p-values. If we were to consider more or all genes in our analysis we may have been able to quantify more genes for the purposes of our analysis.

The DEGs that were identified from each group might offer new ways to find targets or methods to stop tumors from growing. By learning how certain genes change after radiation, we can gather important information about what happens inside cells during radiation treatment,

like fixing DNA, cell death, or adjusting the immune system. The differences in gene activity between treated and untreated conditions give us valuable hints about how genes react to radiation and other outside factors. Exploring the roles of these DEGs could help us learn more about how tumors work, improve radiation treatment, and develop new ways to treat medulloblastoma.

After analyzing the scRNA-seq data and finding genes that are differentially expressed, there are several possible directions to explore. One interesting direction involves creating gene network maps, which would show how genes work together and unveiling connected functions and processes. This would include calculating how genes cooperate and building maps that show how they team up. Another interesting path to go is to combine the scRNA-seq data analysis with proteomics data of the sample set. This would provide a full understanding by connecting gene expression and protein levels. The differentially expressed genes found in untreated samples and those treated with radiation could be further explored through proteomics research to detect structure, function, and potential interacting partners of our genes of interest.

References:

1. Diab M, El-Rayes BF. 2022. The heterogeneity of CAFs and immune cell populations in the tumor microenvironment of pancreatic adenocarcinoma. Journal of Cancer Metastasis and Treatment. 8:42. doi:https://doi.org/10.20517/2394-4722.2022.60.
2. Mayo Clinic Staff. 2023. Medulloblastoma. Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/medulloblastoma/cdc-20363524#:~:text=Medulloblastoma%20is%20a%20type%20of,back%20part%20of%20the%20brain..
3. Dang M, Gonzalez M, Mafra F. 2021 Feb 13. Macrophages in SHH subgroup medulloblastoma display dynamic heterogeneity that varies with treatment modality. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE166691. AccessionGSE166691.
4. Hao and Hao et al. Integrated analysis of multimodal single-cell data. Cell (2021) [Seurat V4]
5. What Is Principal Component Analysis (PCA) and How It Is Used? 2020. Sartorius. https://www.sartorius.com/en/knowledge/science-snippets/what-is-principal-component-analysis-pca-and-how-it-is-used-507186#:~:text=Principal%20component%20analysis%2C%20or%20PCA,more%20easily%20visualized%20and%20analyzed.