Gabriel Fonseca, Sumitra Lele

Genomics and Epigenetics of the Brain

Fall 2022

PSC ID: gfonseca, slele

Abstract:

Prions are of crucial significance to a multitude of neurodegenerative disorders, including Huntington's, Parkinson's and Alzheimer's. The quantification of such pathogenic protein accumulation is a relatively new field, highlighting the importance of computational methods to model underlying molecular mechanisms. Cellular processes in neuronal, microglial and astrocyte cell types are disrupted by prion propagation. The epigenetic activity of brain cell types in relation to inflammatory response or oxidative stress is of importance in potentially treating proteinopathy. In this analysis, we will utilize mouse single-cell RNA seq data to quantify epigenetic changes in these cell types. Using dimensionality reduction, linear regression, and multi-dimensional hypothesis testing, we hope to discover reads significantly expressed in prion infection across cell types. Ultimately, the identification of specific epigenetic markers for prion accumulation will be a step forward in producing treatment for this fatal class of diseases.

Background:

Prions are misfolded proteins that are implicated in a number of neurological diseases. Prions are able to self-propagate and transmit within organisms of the same species, and sometimes within other species. For the vast majority of cases, these diseases are caused by host proteins that cannot function properly, often forming fibrous structures that interrupt regular function. (3) There are multiple classes of prions, leading to a broad classification of these proteins. Among disease types, the onset can be either sporadic, environmental or genetic (2). For example, in Huntington's disease, prions are present since early development, but neurological effects only manifest later in life (3). In addition to humans, prions also affect other mammals such as cows (in the case of "mad cow disease"), as the implicated proteins are conserved across several species (1). Because of this, studying the epigenetic changes induced by prions will better help to develop therapeutic strategies against them.

The idea that prion accumulation can develop into more complex neurological disorders is relatively new. In 2006, it was shown for the first time that prion fiber accumulation leads to inflammation in the brain, which is a known cause of Alzheimer's disease (3). Parkison's disease is another condition that can be induced by accumulation of prions (alpha-synuclein proteins). One study demonstrated that alpha-synuclein can transfer between microglial cells in mouse brains, and that this contributes to the onset of Parkinson's (4). Other cell types, such as oligodendrocytes and neurons, can also be affected by the buildup of prions. In multiple system

atrophy, of which Parkinson's may be a symptom, several cellular processes are affected. The accumulation of alpha-synuclein leads to a rapidly progressing disease due to organelle failure in cells, as well as inflammation in the brain (6). This disease also has both genetic and environmental triggers, suggesting that quantifying gene expression in related prion diseases may provide a basis to understand underlying causal mechanisms.

Dataset:

The dataset contains single cell RNA-Sequencing data. Within the dataset there are normal samples and samples having the Prion disease. This will help us compare the differentially expressed genes. The data contains reads from the mouse genome. The dataset contains the matrix of scRNA data, which has the gene and cell ID. Along with the .mtx file, the dataset also contains the 'genes.tsv' and 'barcodes.tsv' files, which store the gene and cell names. The metadata file contains information about the disease group ('normal', 'Prion Disease'), species_id, biosample_id, cell type information and gene ID. After mapping the read counts to the metadata, we can perform the differential expression analysis.

The dataset can be found at https://singlecell.broadinstitute.org/single_cell/study/SCP1962/dysregulation-of-neuroprotective-astrocytes-a-spectrum-of-microglial-activation-states-and-altered-hippocampal-neurogenesis-are-revealed-by-single-cell-rna-sequencing-in-prion-disease#/

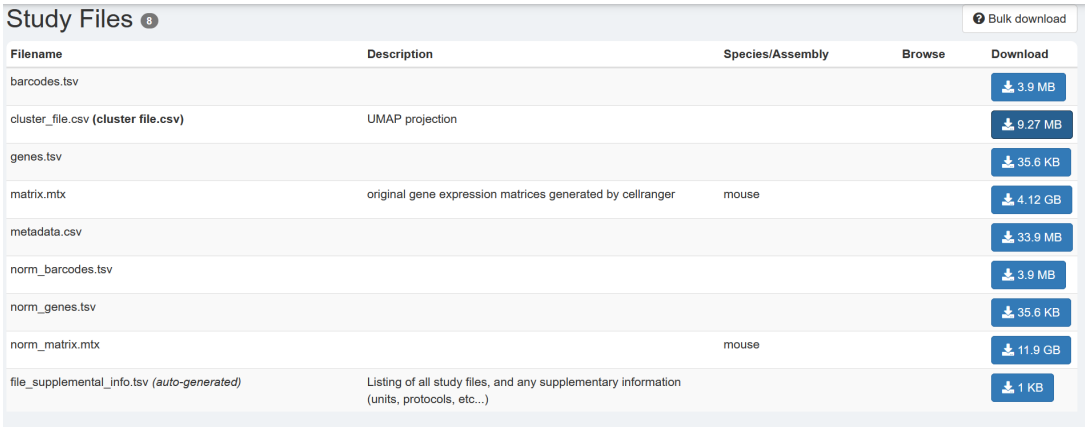


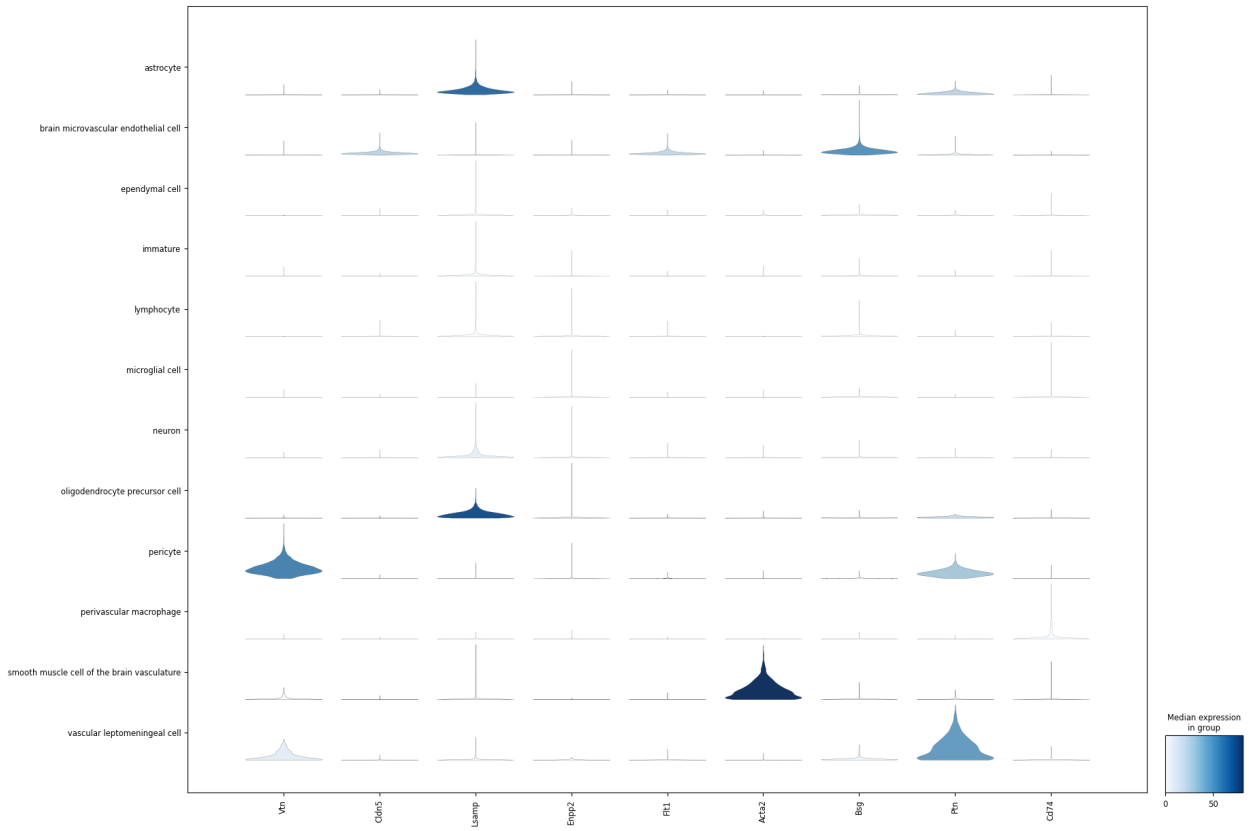
Fig 1. All the study files - metadata, scRNA reads matrix, gene and cell type data

```
%%MatrixMarket matrix coordinate real general
2868 147536 151919585
8 1 .3357474155861516
16 1 74.24752800616045
22 1 .0020565110421778554
42 1 0
45 1 0
49 1 11.861928822323684
52 1 36.60996416356757
65 1 0
77 1 .30826853756451444
78 1 25.92652485392351
80 1 118.09655970941886
84 1 .7860215823620027
88 1 0
89 1 40.85729695797993
91 1 46.69050483808583
94 1 21.0307210881287
95 1 34.709352726447285
97 1 2.6057137352052955
101 1 5.799854134665223
102 1 31.911799734044337
109 1 20.006646870671922
```

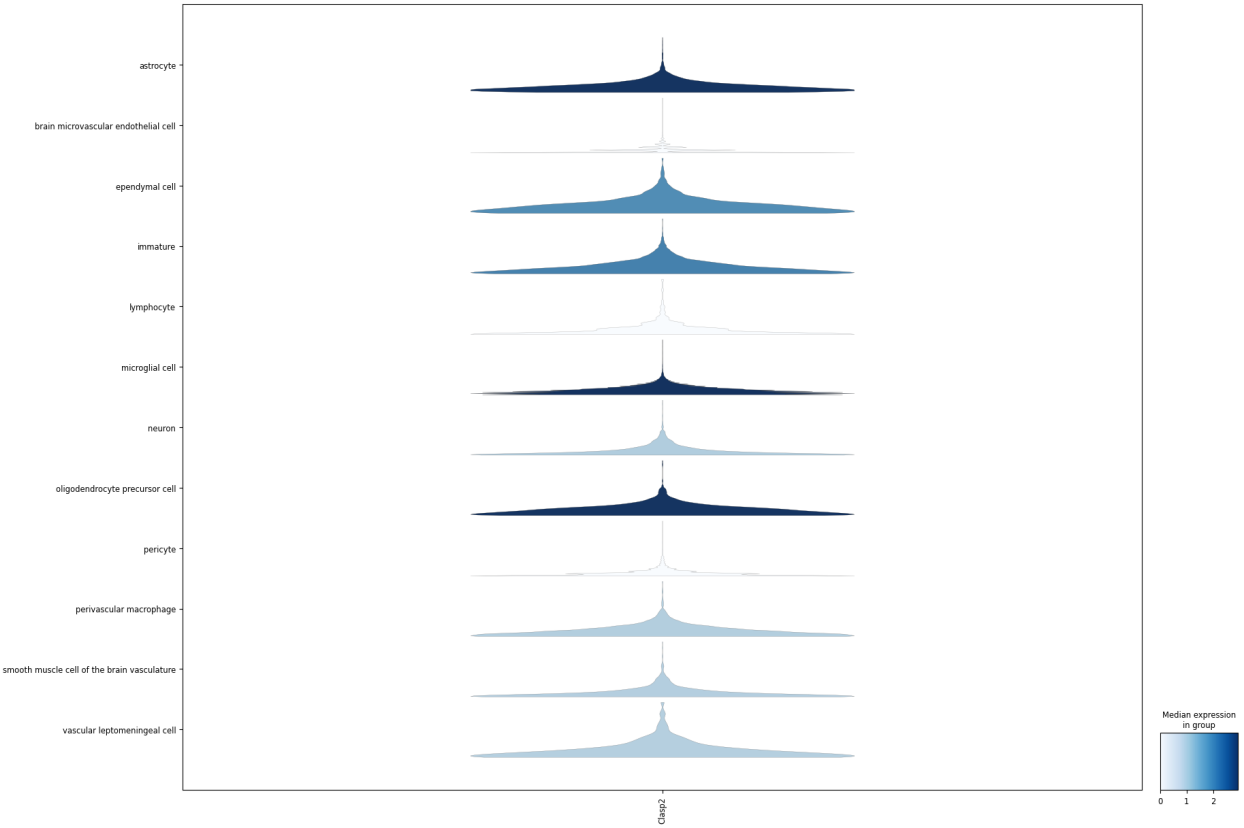
Fig 2. Loading the data in the terminal

Results:

We performed a number of analyses using the scanpy package in Python 3.0. All code was run on a Jupyter notebook. Libraries pandas and anndata were also used. After loading the dataset as an AnnData object, we performed t-tests, logistic regression and wilcoxon rank tests to display a number of violin plots for some genes in the dataset, including each cell type. Of interest, we were able to identify Clasp2 in the gene dataset, and validated the results from the previous analysis on brain size residual (it is differentially expressed in Oligodendrocytes, Astrocytes and Microglia).

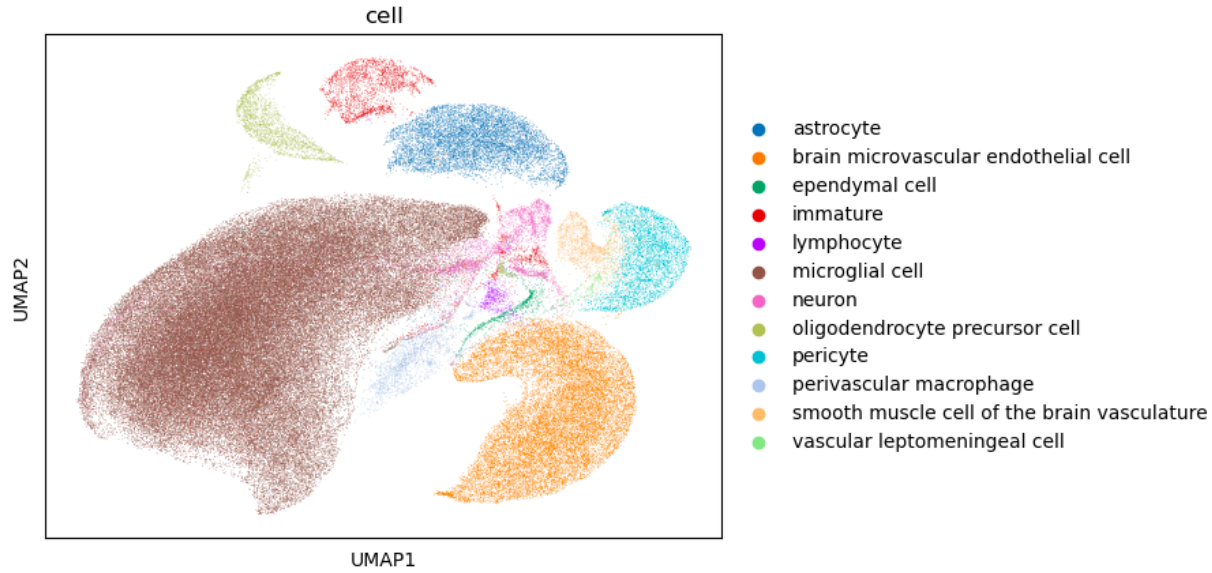


A sample stacked violin plot showing the expression of 10 genes across each cell type

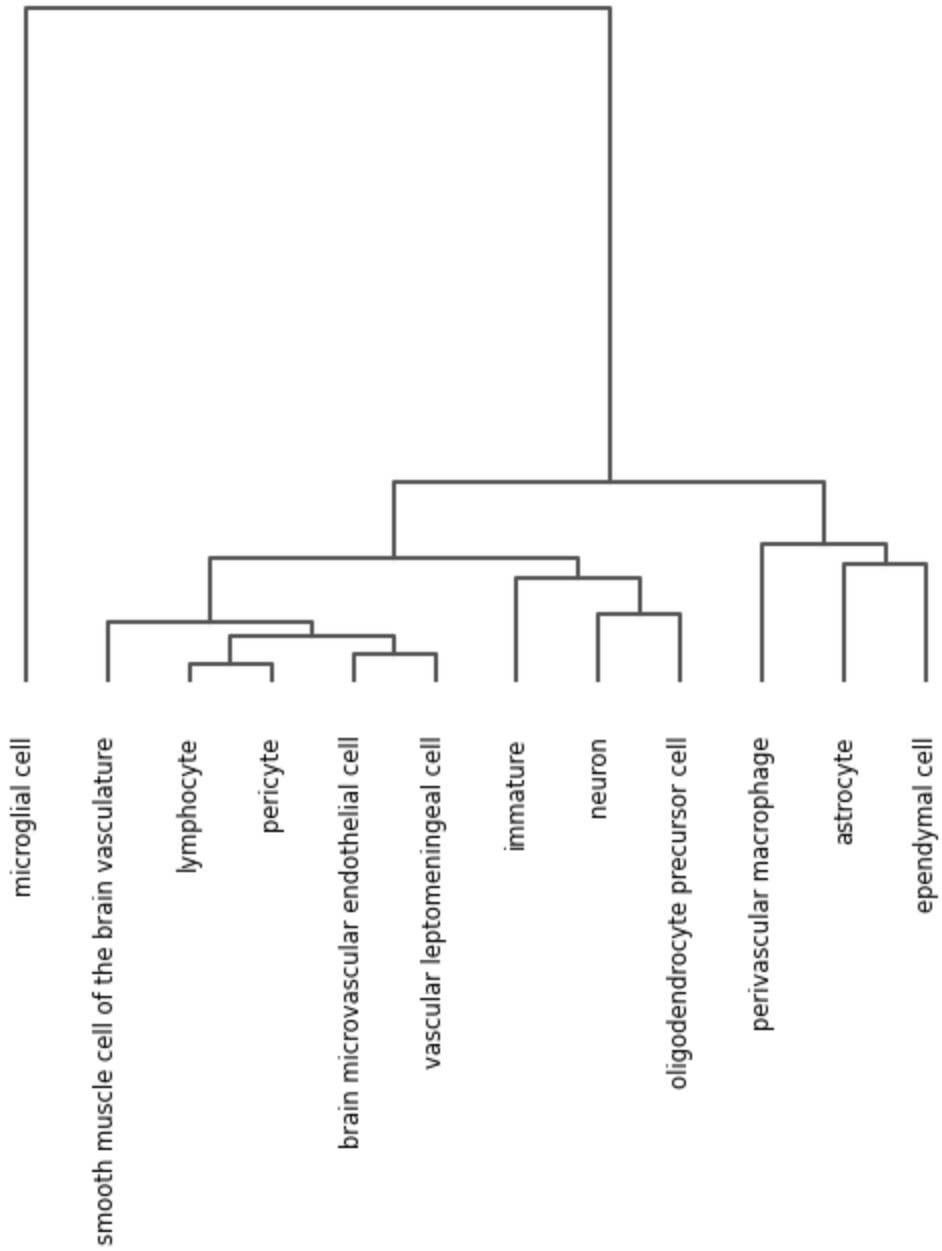


Stacked violin plot showing the CLASP2 expression across each cell

We performed two different types of clustering analyses - UMAP and PCA-based dendrogram clustering. This allowed us to visualize the different cell types within the dataset and their relationships with each other.

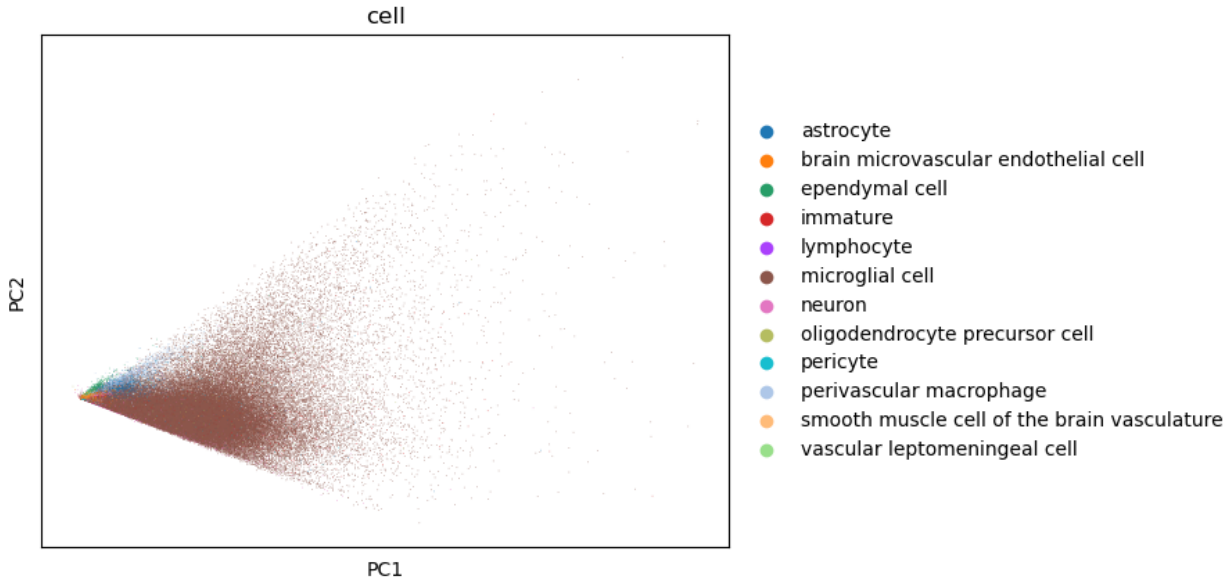


UMAP clustering for each cell type in the data



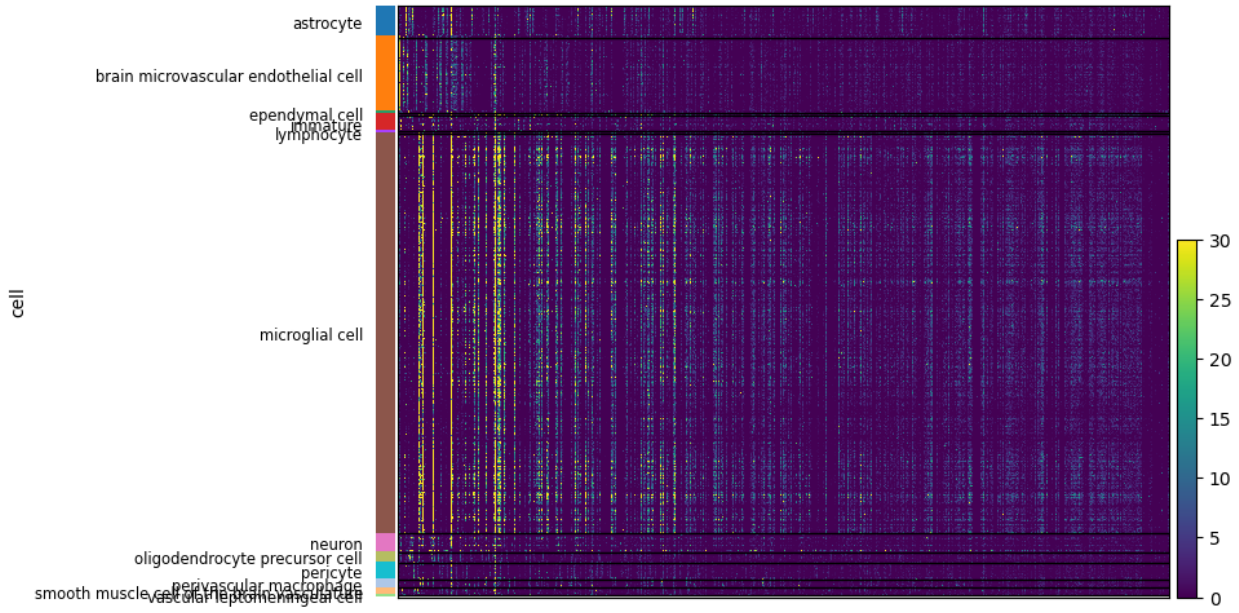
Dendrogram for each cell type

We saw a wide range of data differences in terms of the principal components in microglial cells specifically. This makes sense as these cells are likely involved in the inflammatory response associated with prion disease. There may be some bias in this due to microglial cells composing a large portion of the data, as seen in the UMAP. The PCA plot is shown below:

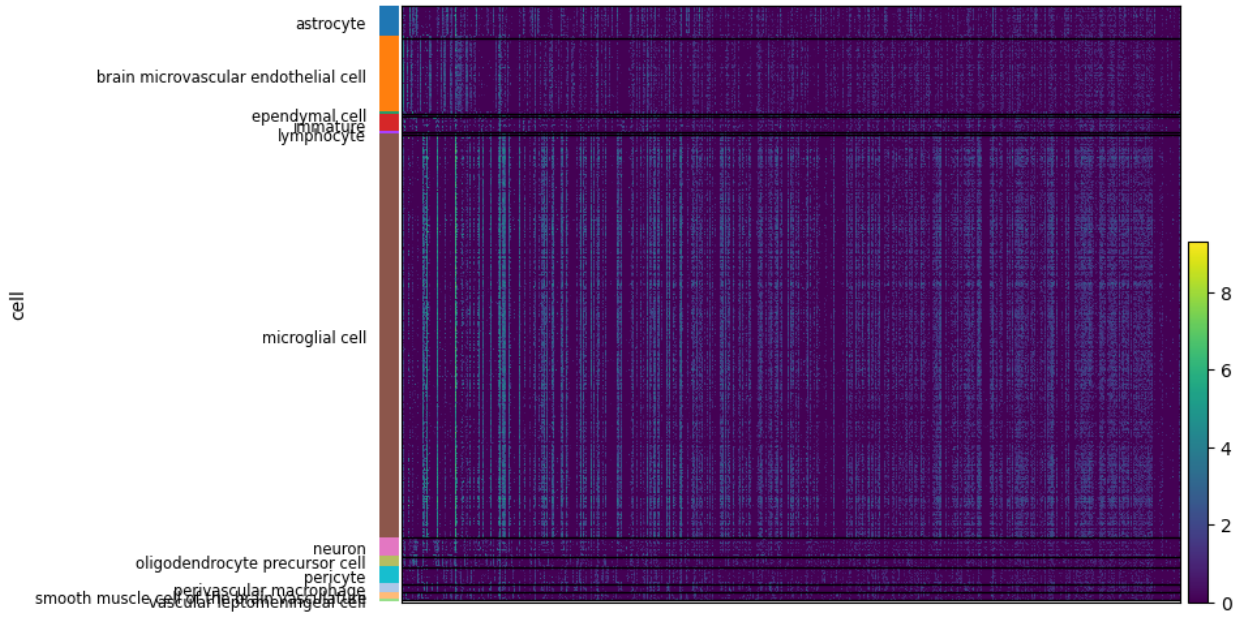


PCA plot for each cell type

Finally, we analyzed the differential expression of each gene across every cell type using a heatmap. The log-transformed heatmap and the raw heat map are shown. Once again, microglial cells appear to clearly differentiate from other cell types in a few genes. The yellow bands represent these genes over-expressed in microglia. This suggests that microglia have a potentially impactful role within the context of prion disease, and that prion disease may in turn affect the epigenetic landscape of these cells. The logarithmic heatmap shows similar results, with a number of genes having around a 6 log-fold change compared to others.

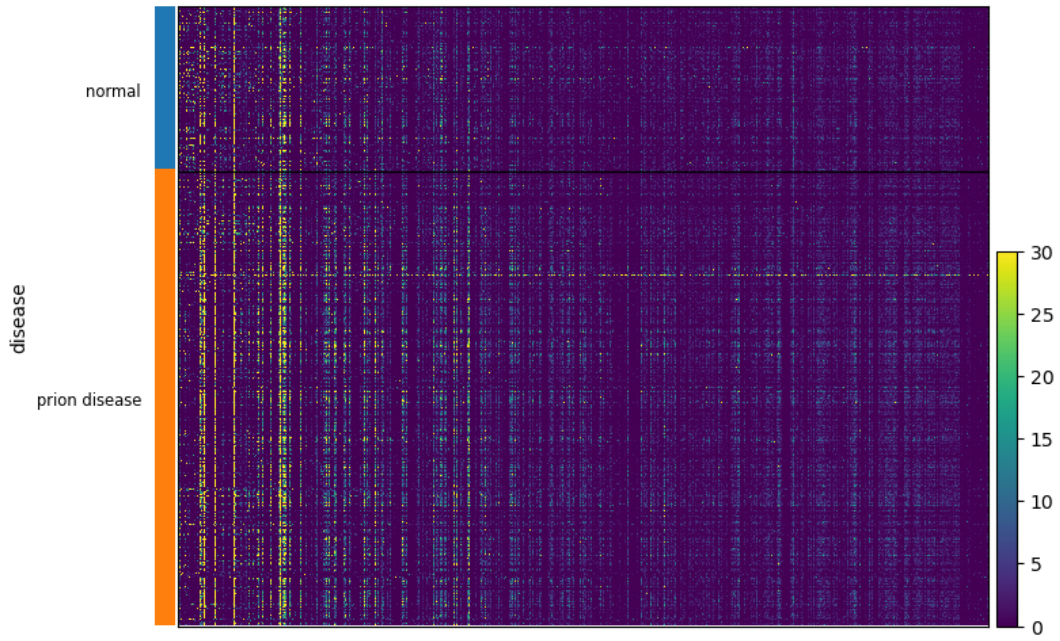


The raw heatmap for each cell type

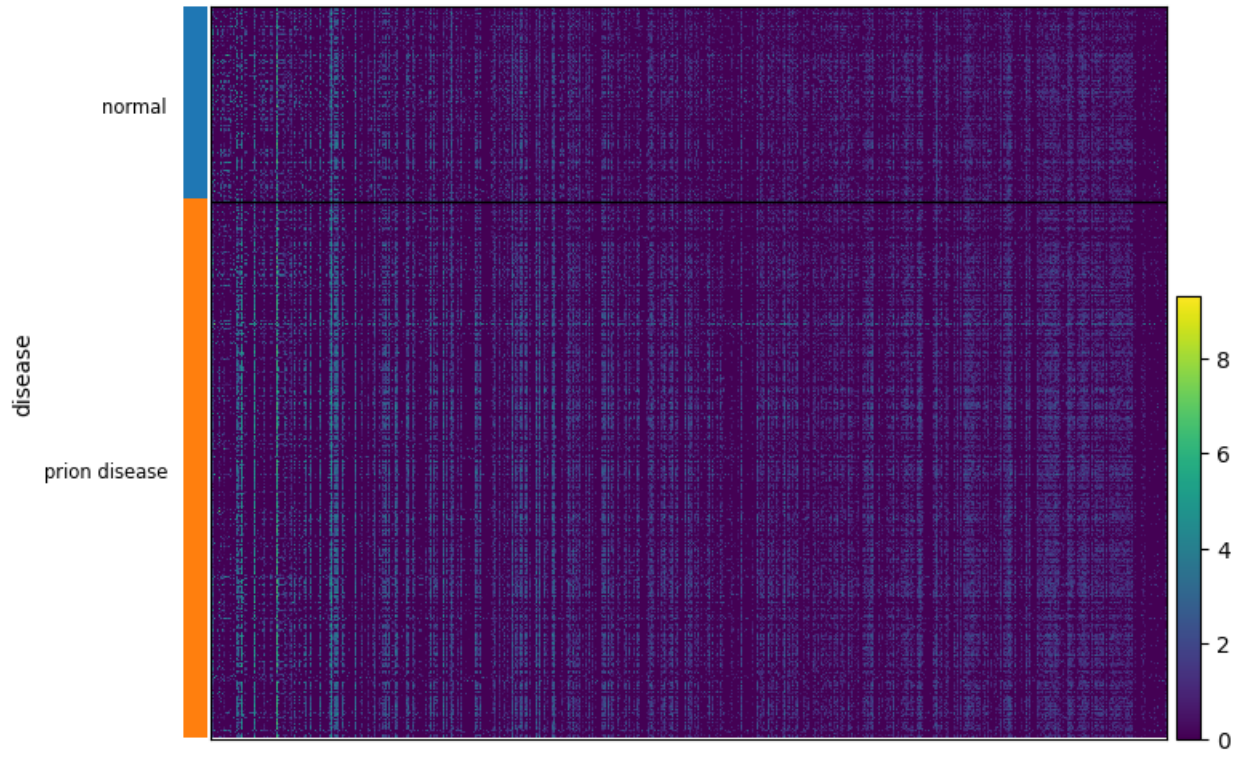


The log heatmap for each cell type

In order to account for disease states, we next moved on to identifying differential expression among the normal conditions and the prionic conditions. The disease vs non-diseased heatmaps are shown:

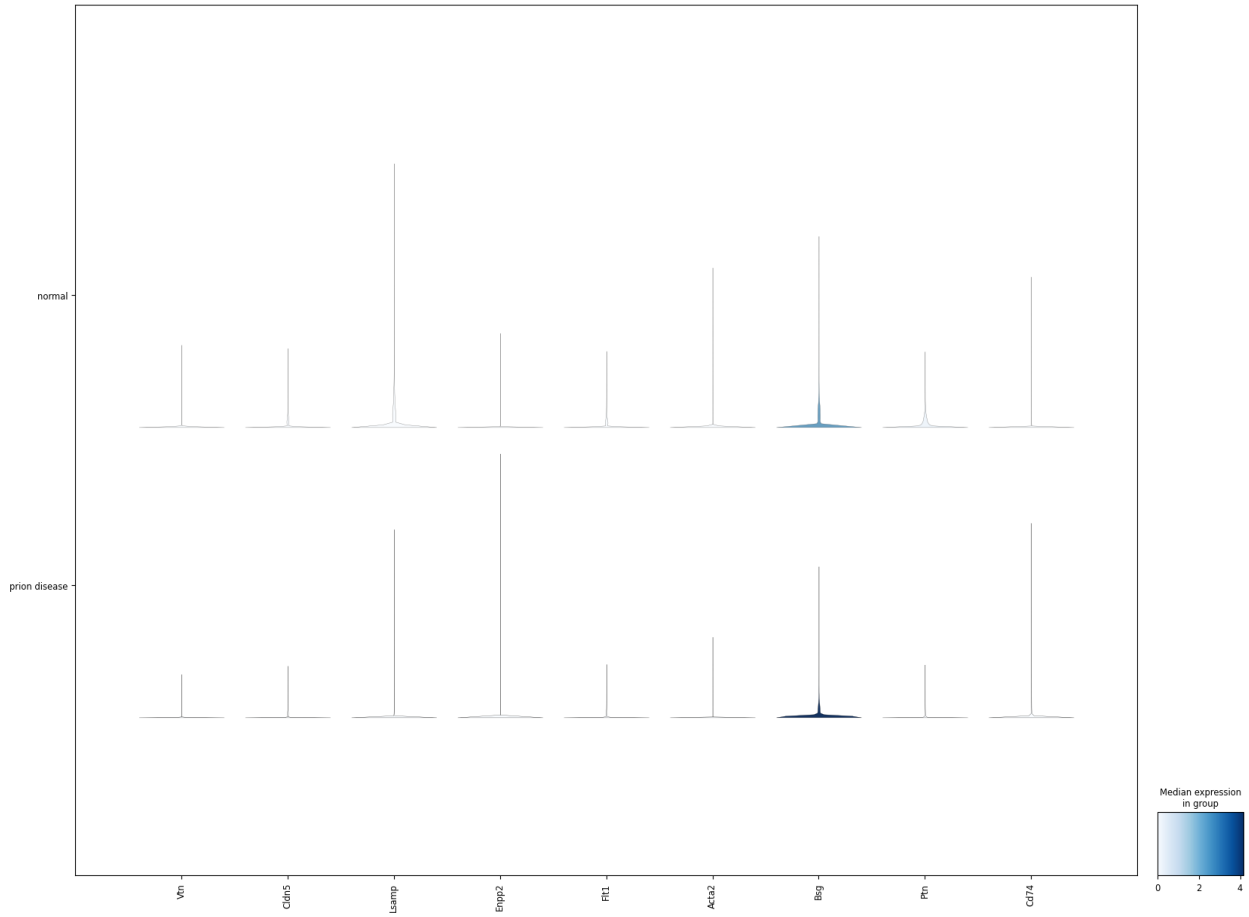


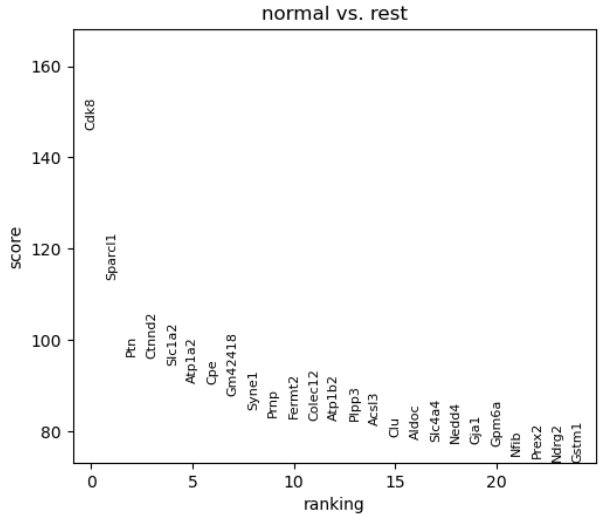
For the non-log transformed, some genes appear to be expressed only in one condition (horizontal lines)

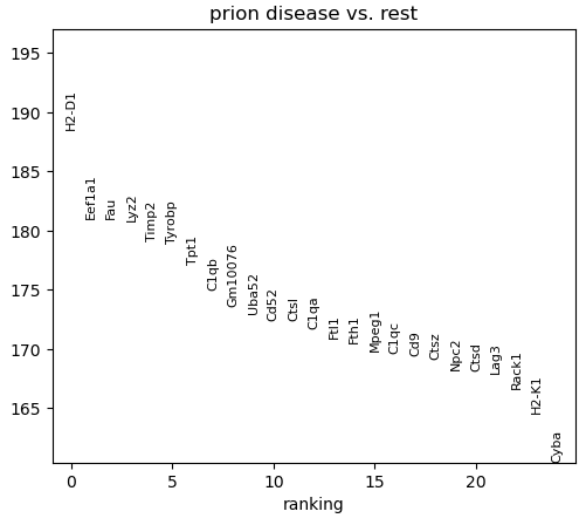


The heatmap for the log-fold change across the two conditions

For the stacked violin plot, we once again looked at 10 genes for a visualization. We then performed a ranked sum analysis to find the most associated genes with prion disease:







For the wilcoxon test, the most influential genes appear to be Cdk8 and H2-D1. Cdk8 is a gene known to be involved in cell cycle signaling and as such has been linked to cancer progression (7). There is not a clear link to this gene and prion disease in the literature, but it is possible that this infers a novel function as well. H2-D1 appears to have a more direct relation to the brain. Its expression tends to be higher in a number of brain regions, and has been linked to aging (8). This may suggest a link between this gene and neurodegenerative disorders that incur with age, such as Alzheimer's, that can be induced by prions.

The gene specificity of prion disease is overall difficult to pinpoint. There are many candidate genes across a number of cell types. Perhaps cell type specific analysis would give better insight as to the function of candidate genes in relation to prion disease. Nevertheless, given the importance of prions when considering neurodegenerative disorders, the ability to find possible gene targets for treatment is a step in the right direction.

Citations:

1. Collinge, John. "Mammalian Prions and Their Wider Relevance in Neurodegenerative Diseases." *Nature*, vol. 539, no. 7628, 9 Nov. 2016, pp. 217–226., https://doi.org/10.1038/nature20415.
2. Saá, Paula, et al. "Mechanisms of Prion-Induced Neurodegeneration." *Expert Reviews in Molecular Medicine*, vol. 18, 8 Apr. 2016, https://doi.org/10.1017/erm.2016.8.
3. Prusiner, Stanley B. "A Unifying Role for Prions in Neurodegenerative Diseases." *Science*, vol. 336, no. 6088, 22 June 2012, pp. 1511–1513., https://doi.org/10.1126/science.1222951.
4. George, Sonia, et al. "Microglia Affect α-Synuclein Cell-to-Cell Transfer in a Mouse Model of Parkinson's Disease." *Molecular Neurodegeneration*, vol. 14, no. 1, 16 Aug. 2019, https://doi.org/10.1186/s13024-019-0335-3.
5. Hao Y, Hao S, Andersen-Nissen E, III WMM, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zagar M, Hoffman P, Stoeckius M, Papalexi E, Mimitou EP, Jain J, Srivastava A, Stuart T, Fleming LB, Yeung B, Rogers AJ, McElrath JM, Blish CA, Gottardo R,

Smibert P, Satija R (2021). “Integrated analysis of multimodal single-cell data.” *Cell*. doi:10.1016/j.cell.2021.04.048, https://doi.org/10.1016/j.cell.2021.04.048.

6. Jellinger, Kurt A. “Multiple System Atrophy: An Oligodendroglioneural synucleinopathy1.” *Journal of Alzheimer's Disease*, vol. 62, no. 3, 13 Mar. 2018, pp. 1141–1179., https://doi.org/10.3233/jad-170397.
7. Menzl, Ingeborg, et al. “Cdk8-Novel Therapeutic Opportunities.” *Pharmaceuticals*, vol. 12, no. 2, 2019, p. 92., https://doi.org/10.3390/ph12020092.
8. Mangold, Colleen A., et al. “CNS-Wide Sexually Dimorphic Induction of the Major Histocompatibility Complex 1 Pathway with Aging.” *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 72, no. 1, 2016, pp. 16–29., https://doi.org/10.1093/gerona/glv232.