Characterizing differentially expressed genes and their effect on action potential

Yufei Jin
yj3085@nyu.edu

Shaun Chua
sc4224@nyu.edu

May 8, 2025

Abstract

Neuronal action potentials vary widely across cell types despite being governed by the same core biophysical mechanisms, and the molecular foundation of this functional diversity remain largely unexplored. Here, we propose to leverage Patch-seq data from the Allen Institute's primary visual cortex, which pairs single-cell RNA-seq (more than 44,000 genes in 4,435 GABAergic interneurons) with detailed electrophysiological profiles to test the hypothesis that transcriptional variability of ion channels and their regulators drives distinct action-potential phenotypes. After rigorous quality control and alignment of sequencing reads, we will extract key AP metrics (threshold, peak amplitude, upstroke/downstroke slopes, latency) and normalize both transcriptomic and electrophysiological features. Dimensionality reduction techniques(PCA, UMAP), K nearest neighbors and K-means clustering will be applied to gene and feature space to define neuronal subpopulations, and DESeq2-based differential expression analysis will identify genes whose expression correlates with specific AP traits. Gene ontology enrichment and regression analyses will then pinpoint both well-known ion-channel subunits and new upstream/downstream regulators shaping excitability. By integrating high-resolution transcriptomic and functional data, this study aims to map gene expression signatures to biophysical parameters of the action potential, offering new targets for modulating neuronal firing in health and disease.

Keywords: Action potential · Differentially expressed genes· Patch-seq.

Contents

1 Research Question

Neuron action potentials are the fundamental units of information processing in the brain. Although the roles of sodium and potassium channels in the formation of action potentials are well established [1], the extent to which variability in action potential waveforms between neuronal subtypes can be explained by differences in gene expression remains unclear.

Recent advances in Patch-seq technology, which combines electrophysiological recording, transcriptomic profiling, and physical characterization of the same neuron, provide an opportunity to directly link gene expression patterns to functional properties at single cell resolution [2].

Research question
Investigating how variations in gene expression, particularly of ion channel genes and related regulatory pathways, influence specific electrophysiological features of the action potential, such as threshold, peak amplitude, and frequency.

2 Hypothesis

We hypothesize that the electrophysiological properties of neurons, particularly the characteristics of their action potentials, are fundamentally shaped by their gene expression profiles. Variations in gene expression, especially of ion channels and related regulatory gene traits, contribute to the diversity of action potential waveforms observed across neuronal subtypes.

3 Background and Motivation

Our project is set up to discover genes that have a direct relationship with action potential. Action potential is a very important concept in neuroscience, without action potential, neurons cannot work. It's generally accepted that the interplay between sodium and potassium channels drives this process. A key question then is: since various cell types exhibit distinct action potential patterns, to what extent can differences in gene expression account for these variations? We still don't know whether differences in ion-channel gene expression underlie these variations, or if changes in upstream or downstream genes are responsible, and until recently, we've lacked the tools to resolve this. Thanks to recent advances in biotechnology, we can now use patch-seq, a technique that simultaneously captures a cell's morphology, electrophysiological properties, and single-cell RNA-seq profiles, to deepen our understanding of action potentials.

4 Method

The dataset comprises single-cell transcriptomes and electrophysiological recordings from 4,435 interneurons in the mouse primary visual cortex, made available

by the Allen Institute[3]. For each cell, RNA-seq profiles quantify expression levels of 44,768 genes, and cells are annotated into five major GABAergic interneuron subtypes (Sst, Pvalb, Lamp5, Vip, Sncg). Complementing the gene-expression data are key action-potential features including spike threshold, upstroke/downstroke slope ratio, peak amplitude, trough voltage, and latency. These features were measured via patch-clamp electrophysiology. This enables direct mapping of transcriptional variation onto neuronal excitability phenotypes at single-cell resolution. We present our experimental pipeline as follows:

Quality Control and Alignment

- FASTQ sequences were attained from the NeMO Archive.
- Filter and remove adapter sequences with Trimmomatic v0.3.9[4].
- Reference mus musculus genome was attained from NIH's NCBI repository for sequence alignment.
- Align paired sequences using STAR v2.7.11[5].

Normalization

- Expression counts matrix was downloaded from the Allen Brain Map Repository.
- Filter out rows that contain NaN values in the metadata.
- Apply SCTransform[6] normalization to the data.
- Data is fit with a negative-binomial model, then the observations are centered and scaled by taking the Pearson residuals[7].
- Extreme residuals are clipped and remaining residuals are z-scored.

Dimensionality Reduction and Clustering

- Apply PCA and UMAP to gene expression and electrophysiology data.
- Build a K-nearest neighbors graph and perform community detection clustering with FindNeighbors and FindClusters
- Perform K-means clustering to identify subpopulations of neuron cell types and compare their results against FindClusters method.

Feature Extraction and Statistical Analysis

- Quantify electrophysiological properties (threshold, upstroke/downstroke ratio, etc.).
- Use ANOVA/histograms to compare action potential features across cell types.

Differential Expression and Enrichment

- Use Seurat[8] to identify genes that significantly differ with specific electrophysiological traits [9].
- Conduct GO-term enrichment and GSEA to reveal the biological pathways involved [10].

Gene-Feature Correlation

- Use regression techniques to pinpoint specific genes strongly correlated with individual electrophysiological features [11].

5 Results

5.1 Quality Control and Alignment

The dataset contains 4,435 cell samples from the mus musculus species. Due to the large number of samples that were provided for this study and the time constraints, it would be intractable for us to carry out quality control and sequence alignment on all 4,435 samples.

Instead we included bash scripts to do quality control, filtering of adapter sequences and to execute sequence alignment on 1 sample. This was to demonstrate that we know how to do sequence alignment with STAR. See our Github repository for more details. In the ideal scenario, we would perform sequence alignment with STAR on all cell samples and aggregate their expression counts into a matrix to be used for downstream analyses.

5.2 Dimensionality Reduction and Clustering

Dimensionality reduction and clustering was applied to a high-dimensional dataset of 3,600 single-cell samples and 44,768 genes. Our initial intention was to determine similarities between the different samples. This can be useful as a first-pass quality control step to ensure that the replicates of the same cell type have a higher similarity to each other than those of a different cell type.

Our plan was to carry out K-means clustering to the PCA results but realized that the Seurat package included their own community detection clustering technique.

We visualized both clustering results corresponding to the 5 cell types on a UMAP plot, their plots look identical, indicating that both clustering approaches recover the same underlying cellular structure. This increases our confidence that the dataset of expression profiles is well characterized to identify differentially expressed genes in a reproducible manner downstream.

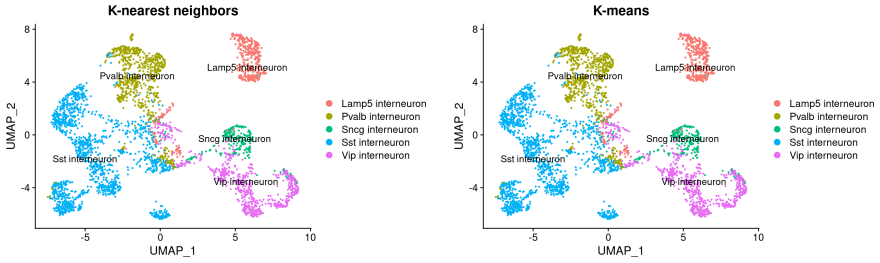

Figure 1: Dimensionality reduction based on community detection clustering and K-means clustering method

5.3 Feature Extraction

In order to characterize action potentials in different types of neuronal cells using the Patch-seq dataset, we focus on selecting electrophysiological features that not only represent key aspects of action potential dynamics, but also exhibit significant variation between cell types. Initially, six candidate features were considered based on their relevance to spike shape and excitability: Trough Voltage, F-I Curve Slope, Peak Voltage, Threshold Voltage, Upstroke/Downstroke Ratio, and Resting Membrane Potential.

To statistically assess whether these characteristics differ significantly between cell types, we performed a one-way ANOVA for each characteristic in five distinct cell types. This statistical test allows evaluation of whether the mean of at least one group differs from the others. For features showing significant ANOVA results, we conducted post hoc pairwise comparisons using Tukey's Honest Significant Difference (HSD) test[12]. This step identifies which specific cell type pairs show statistically significant differences in each electrophysiological property, providing a more granular view of how these features vary among cell classes.

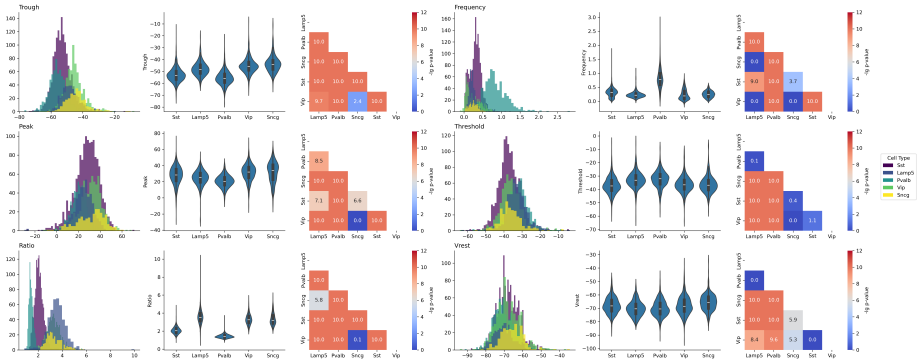

Figure 2: Feature distribution and pairwise T-test

The six selected electrophysiological characteristics exhibited statistically significant group-level differences across the five cell types, as determined by one-way ANOVA. However, pairwise comparisons revealed that Vip and Sncg cell types showed minimal to no significant differences across these features. This suggests that their action potential characteristics are highly similar.

This observation aligns with previous analyses based on transcriptomic data, specifically UMAP projections of gene expression profiles, which showed that Vip and Sncg cells cluster closely in the global structure. The convergence of electrophysiological and gene expression evidence supports the interpretation that these two cell types share both functional and molecular similarities. This consistency among data modalities strengthens the hypothesis that Vip and Sncg cells represent a closely related functional subtype within the inhibitory neuron population.

Among the five cell types analyzed, Pvalb cells exhibited the most pronounced differences across the six electrophysiological features, distinguishing them clearly from the other groups. This suggests that Pvalb neurons possess distinct biophysical properties that likely stem from unique underlying gene expression patterns related to ion channels, membrane excitability, or synaptic function. Given this divergence, we propose to use Pvalb cells as a reference group in subsequent analyses to better characterize how other cell types differ in both electrophysiological and transcriptomic dimensions. This approach may help uncover gene–function relationships that underlie cell-type-specific firing behaviors.

5.4 Differential Expression

These pairwise comparisons highlight distinct sets of differentially expressed genes in the four contrasts, reflecting the molecular diversity that underlies the functional specialization of these interneuron classes. Notably, each comparison reveals a unique pattern of upregulated and downregulated genes, suggesting

that Pvalb cells diverge from Sst, Sncg, Vip, and Lamp5 cells in both shared and cell type-specific transcriptional programs.

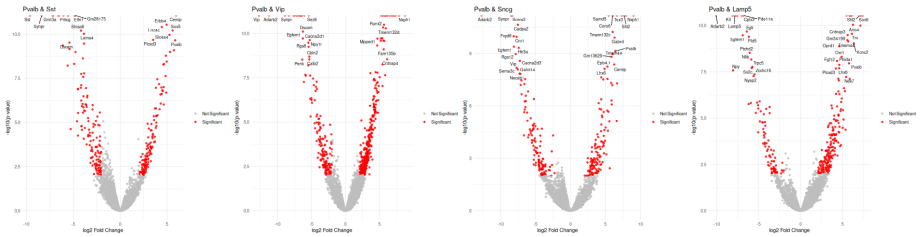

Figure 3: Differentially expressed genes

Building on this, we can further investigate how the expression of genes related to electrophysiological characteristics, such as ion channels, synaptic proteins, and signaling molecules, contributes to these differences. By integrating electrophysiologically relevant gene sets into the differential expression framework, we can assess whether changes in the expression of these genes account for or correlate with known functional differences between interneuron types, such as variations in action potential dynamics, firing patterns, and synaptic properties.

5.5 Electrophysiology-Gene correlation

To bridge the gap between molecular profiles and neuronal function, we next combined transcriptomic data with electrophysiological measurements. Specifically, we focus on identifying genes whose expression levels show a linear relationship with key electrophysiological characteristics between neurons. This approach allows us to directly link gene expression patterns to functional properties such as action potential threshold, amplitude and firing frequency.

To minimize noise and cell-to-cell variability, we first aggregated the data at the cell subtype level, grouping neurons by their assigned class (61 subtypes for 3600 cells). By summarizing the expression and electrophysiological characteristics within each subtype, we obtained more stable estimates that better reflect characteristic differences between groups.

We then applied linear regression analyzes to identify genes whose expression levels were significantly associated with specific electrophysiological parameters in these subtypes. This integrative approach allowed us to pinpoint candidate genes that may play mechanistic roles in shaping the distinct electrophysiological signatures of each type of interneuron.

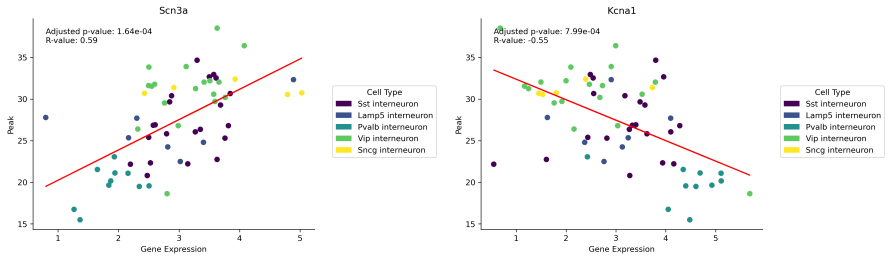

Figure 4: Sodium and potassium channel related to peak voltage

As an example of the combined analysis, we focused on the action potential peak amplitude, a key electrophysiological feature that varies across interneuron subtypes. By examining genes whose expression levels are linearly correlated with the peak amplitude, we identified several candidates with strong and biologically meaningful associations.

Notably, Scn3a, which encodes a voltage-gated sodium channel subunit, showed a positive linear relationship with peak amplitude across cell types. This is consistent with its known role in promoting sodium influx during the depolarization phase of the action potential, thereby contributing to higher peak voltages[13]. Conversely, Kcna1, which encodes a voltage-gated potassium channel (Kv1.1), displayed a negative linear relationship with peak amplitude, which is also consistent with the role of Kv1.1 channels in leading the potassium efflux to slow down the depolarization process[14].

These examples illustrate how integrating transcriptomic and electrophysiological data not only uncovers meaningful gene-feature relationships but also reinforces and extends mechanistic insights established by previous studies. Moving forward, we will combine these results with the previously identified differentially expressed genes to perform further analyses, allowing us to explore how molecular differences across cell types contribute to electrophysiological diversity at both the single-gene and pathway levels.

5.6 GO enrichment analysis

By applying GO enrichment separately to the gene sets linked to each electrophysiological feature, we can compare the shared and distinct biological processes that shape neuronal function. For example, we expect features like action potential peak amplitude to be enriched for ion channel activity, while other features such as threshold or frequency may reveal enrichment for metabolic regulation.

To facilitate clear and interpretable comparisons, we will focus our GO term enrichment analysis on genes identified from comparisons between Pvalb cells and each of the other four interneuron types (Sst, Sncg, Vip, Lamp5).

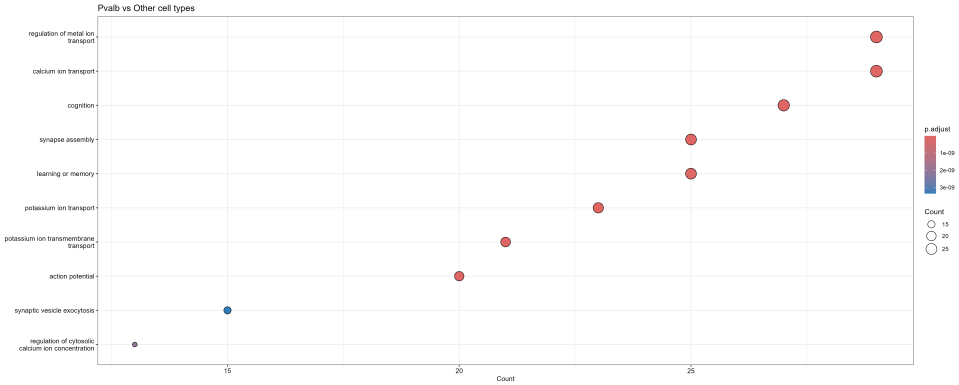

Figure 5: GO enrichment analysis

The GO enrichment analysis highlights the differences between Pvalb cells and other neuronal cell types. Pvalb cells are well known for their fast-spiking properties, and the enrichment results reflect this distinct physiological profile[15]. Specifically, we observe strong enrichment in pathways related to ion transport, which are essential for rapid firing and precise synaptic transmission. Additionally, some terms related to cognitive functions, such as learning and memory, also emerge, suggesting the involvement of Pvalb cells in higher-order neural processes.

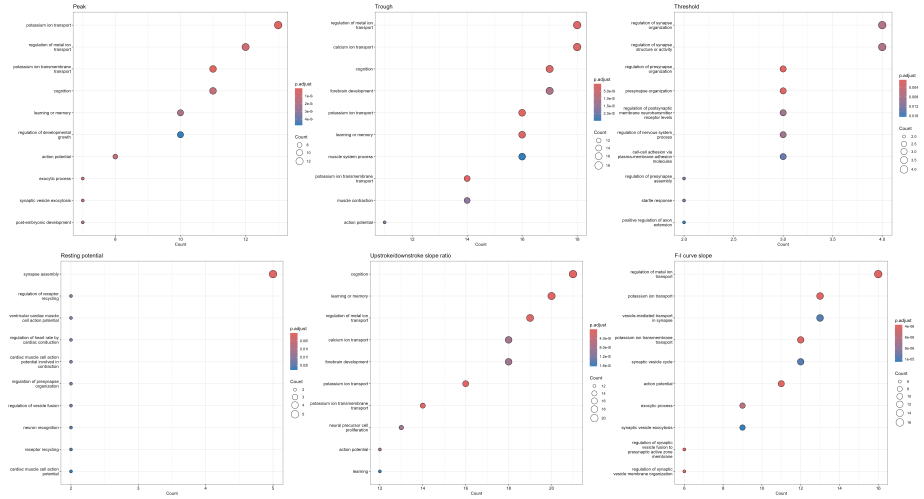

Figure 6: Go enrichment analysis combined with electrophysiological features

Next, we analyzed how integrating individual electrophysiological features influences the GO enrichment results, starting with the action potential peak.

We observed enrichment in terms related to potassium ion transport and potassium ion transmembrane transport, highlighting the critical role of potassium channels in shaping the peak voltage of action potentials[16]. Interestingly, we also found enrichment for terms related to cognitive functions, which was somewhat unexpected. Traditionally, action potentials are considered all-or-none events, suggesting that the peak amplitude may not carry substantial biological significance. However, our results suggest that even subtle variations in peak amplitude may contribute to broader cellular functions, offering new insights into the molecular underpinnings of neuronal activity.

For the trough, we observed enrichment of terms related to calcium ion channels, suggesting a key role for calcium dynamics in shaping the hyperpolarizing phase of the action potential. In contrast, the resting potential showed very few enriched GO terms, indicating that baseline membrane properties may be relatively conserved across these neuronal subtypes. The threshold revealed enrichment for regulatory processes, implying that the excitability of neurons is closely tied to the regulation of molecular and cellular mechanisms that control action potential initiation.

Lastly, we examined the F–I curve slope and the upstroke/downstroke slope ratio. The F–I curve slope, which reflects the relationship between firing rate and input current, was enriched for terms related to potassium channels and ion channel regulation, further emphasizing the importance of potassium conductances in controlling neuronal excitability. Interestingly, the upstroke/downstroke slope ratio was strongly associated with terms related to cognition, learning, and memory. This finding suggests that the fine-tuning of action potential dynamics may have important implications for higher-order brain functions, highlighting potential links between electrophysiological precision and cognitive processes.

6 Conclusion

In summary, this study focuses on how to analyze Patch-seq datasets by integrating transcriptomic and electrophysiological data, providing a framework to explore the molecular underpinnings of neuronal diversity. Our results not only reveal meaningful gene-feature relationships but also offer insights into the functional specialization of different neuronal subtypes. This analysis serves as a valuable starting point for future investigations into the complex interplay between gene expression and cellular activity in the nervous system.

7 Discussion

There are some limitations in the patch-seq dataset and our analysis. First, the connection between cellular activity and the transcriptome remains incomplete in Patch-seq datasets. Specifically, mRNA levels do not always reflect the abundance or functional state of the corresponding proteins, limiting our ability to directly link gene expression to electrophysiological properties. Second, while

this study focused on identifying linear relationships between genes and electrophysiological features, it is important to recognize that biological systems often exhibit nonlinear patterns, such as exponential or inverted U-shaped relationships. Exploring these complex patterns could yield deeper insights into the regulation of neuronal function.

8 Future work

In future studies, addressing the limitations of Patch-seq data will be essential to deepen our understanding of the relationship between gene expression and neuronal function. One important direction is to integrate proteomic data alongside transcriptomic and electrophysiological measurements. This would help bridge the gap between mRNA abundance and actual protein levels, providing a more accurate picture of how gene expression translates into cellular activity. Additionally, it will be valuable to explore nonlinear gene-feature relationships, as biological systems often follow exponential, sigmoidal, or inverted U-shaped patterns rather than strictly linear trends. Incorporating nonlinear modeling approaches could reveal hidden regulatory mechanisms and provide a more nuanced understanding of how genes shape the diverse functional properties of neurons.

References

[1] Hodgkin, A. L. and Huxley, A. F. *The Journal of physiology* **117**(4), 500 (1952).

[2] Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S., and Zhuang, X. *Science* **348**(6233), aaa6090 (2015).

[3] Gouwens, N. W., Sorensen, S. A., Baftizadeh, F., Budzillo, A., Lee, B. R., Jarsky, T., Alfiler, L., Baker, K., Barkan, E., Berry, K., et al. *Cell* **183**(4), 935–953 (2020).

[4] Bolger AM, Lohse M, U. B. *Bioinformatics* (2014).

[5] Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. *Bioinformatics* **29**(1), 15–21 (2013).

[6] Hafemeister, C. and Satija, R. *Genome biology* **20**(1), 296 (2019).

[7] Lause, J., Berens, P., and Kobak, D. *Genome biology* **22**, 1–20 (2021).

[8] Hao, Y., Stuart, T., Kowalski, M. H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., and Satija, R. *Nature Biotechnology* (2023).

[9] Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. *cell* **177**(7), 1888–1902 (2019).

[10] Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. *Omics: a journal of integrative biology* **16**(5), 284–287 (2012).

[11] Cadwell, C. R., Palasantza, A., Jiang, X., Berens, P., Deng, Q., Yilmaz, M., Reimer, J., Shen, S., Bethge, M., Tolias, K. F., et al. *Nature biotechnology* **34**(2), 199–203 (2016).

[12] Abdi, H. and Williams, L. J. *Encyclopedia of research design* **3**(1), 1–5 (2010).

[13] Catterall, W. A. *The Journal of physiology* **590**(11), 2577–2589 (2012).

[14] Jan, L. Y. and Jan, Y. N. *The Journal of physiology* **590**(11), 2591–2599 (2012).

[15] Hu, H., Gan, J., and Jonas, P. *Science* **345**(6196), 1255263 (2014).

[16] Noble, D. *Physiological Reviews* **46**(1), 1–50 (1966).