Neuronal Spike Shapes (NSS): a simple approach to study electrophysiological data for heterogeneity investigation.

Lorenzo Martini[a,*,1], Gianluca Amprimo[a,b,**,1], Stefano Di Carlo[a], Gabriella Olmo[a], Claudia Ferraris[b], Alessandro Savino[a] and Roberta Bardini[a,*]

[a]Politecnico di Torino - Control and Computer Engineering Department, Corso Duca degli Abruzzi, 24, Turin, 10129, Italy
[b]Institute of Electronics, Information Engineering and Telecommunications, National Research Council, Corso Duca degli Abruzzi, 24, Turin, 10029, Italy

ARTICLE INFO

Keywords:
action potential
patch-seq
multimodal single-cell
electrophysiology
data analysis
computational biology
bioinformatics
unsupervised learning

ABSTRACT

The mammalian brain exhibits a remarkable diversity of neurons, contributing to its intricate architecture and functional complexity. The analysis of multi-modal single-cell datasets enables the investigation of this heterogeneity. In this study, we introduce the Neuronal Spike Shapes (NSS), a straightforward approach for analyzing the electrophysiological profiles of cells based on their Action Potential (AP) waveforms. The NSS method holds the potential to effectively explore the heterogeneity of cell types and states by summarizing the AP waveform into a triangular representation complemented by a set of derived electrophysiological (EP) features. To support this hypothesis, we validate the proposed approach using two datasets of murine cortical interneurons. The validation process involves a combination of NSS-based clustering analysis, Differential Expression (DE), and Gene Ontology (GO) enrichment analysis. The results demonstrate that the NSS-based analysis captures distinct partitions of cells that possess biological relevance independent of cell subtype, suggesting them as potential cellular states. Furthermore, the analysis reveals the emergence of voltage-gated K^+ channels as transcriptomic markers associated with the identified electrophysiological partitions. This finding is particularly significant as the expression of voltage-gated K^+ channels regulates the hyperpolarization phase of the AP, providing further support for the biological significance of the NSS-based clustering approach.

1. Introduction

The mammalian brain displays many cell types, contributing to its complex structure and functions [26, 1, 63, 15]. Investigating cellular heterogeneity aims to understand how distinct cells specialize for specific functions [37, 47], and how they interact to form intricate neuronal networks during development or in adult organisms [18, 13, 64, 32, 33].

The multi-modal single-cell analysis provides multiple layers of data to explore neuronal states and their transcriptional profiles across cell types [40, 39, 24, 23, 67]. Indeed, comprehending cellular heterogeneity requires profiling thousands of cells in multiple dimensions to accurately identify and characterize neuronal types and subtypes [69, 53, 8]. A cell type is defined as a group of cells sharing a stable phenotype based on transcriptomic, morphological, and functional markers. However, these categories have significant uncertainties [73].

Furthermore, the dynamic nature of cell dynamics introduces additional layers of heterogeneity. A cell state refers to a *"transient or dynamically responsive property of a cell to a context"* [73], distinguished by complex molecular and functional properties [48]. In the mammalian brain, plasticity mechanisms play a crucial role in the variability of neuronal states, acting throughout the lifespan of the organism to balance stability and adaptation of neuronal networks, supporting brain functions and learning [66, 30]. These intricate processes arise from continuous changes in underlying cell dynamics, leading to various combinations of cellular states [11]. Therefore, investigating both cell types and cell states' heterogeneity is vital for unraveling the biological dynamics underlying complex brain functions. Specifically, studying cortical interneurons presents a unique challenge, as they do not conform to the conventional rules of neuroplasticity [43], making the interpretation of identified states non-trivial. The availability of single-cell electrophysiological (EP) data enables the systematic characterization of neuronal cells from a functional perspective, allowing for cross-validation and joint analysis of EP and transcriptional profiles. Among various EP features, the Action Potential (AP) waveform contains a wealth of information about neuronal types and states.

This work introduces Neuronal Spike Shapes (NSS), a straightforward approach for investigating heterogeneity in cell types and states by summarizing the AP waveform into a triangular representation, which describes a set of descriptive EP features. The proposed methodology is then

*Corresponding author
**Principal corresponding author
lorenzo.martini@polito.it (L. Martini); gianluca.amprimo@polito.it (G. Amprimo); roberta.bardini@polito.it (R. Bardini)
www.smilies.polito.it (L. Martini); https://www.polito.it/personale?p=gianluca.amprimo (G. Amprimo); www.smilies.polito.it (S. Di Carlo); https://www.sysbio.polito.it/analytics-technologies-health/ (G. Olmo); https://www.ieiit.cnr.it/people/Ferraris-Claudia (C. Ferraris); www.smilies.polito.it (A. Savino); www.smilies.polito.it (R. Bardini)
ORCID(s): 0000-0002-7794-7791 (L. Martini); 0000-0003-4061-8211 (G. Amprimo); 0000-0002-7512-5356 (S. Di Carlo); 0000-0002-3670-9412 (G. Olmo); 0000-0001-5381-4794 (C. Ferraris); 0000-0003-0529-7950 (A. Savino); 0000-0002-1809-3212 (R. Bardini)
[1]These authors share co-first authorship and contributed equally to this work.

validated using two datasets of murine cortical interneurons. The NSS-based analysis captures groups of cells that exhibit biologically relevant properties independent of cell subtypes, suggesting their candidacy as cellular states.

After providing an overview of the background and current research in Section 2, this manuscript presents the proposed methodology in Section 3, followed by a detailed presentation of the conducted analysis in Section 4. Section 5 discusses the results, draws conclusions, and suggests future directions to advance the proposed research.

2. Background

Contemporary methodologies for investigating the EP characteristics associated with cell types and states encompass the computational analysis of single-cell multimodal data and the formulation of mathematical models that describe the AP waveform.

2.1. Computational electrophysiological analysis

Computational investigations of cell heterogeneity aim to explore the distinct cell types within a single-cell dataset [44, 70, 28, 55]. Typically, this involves performing unsupervised clustering analysis to group cells and subsequently assigning identities to these clusters based on established knowledge [8]. For instance, in the context of single cell RNA sequencing (scRNA-seq) data, labeling is achieved by identifying known markers within the expression profiles of cells [29, 68], enabling the association of clustered cells with specific types or subtypes, or even the discovery of new cell subgroups. This marker-based approach prioritizes recognizing information about stable cell identities over detecting transient functional states [46, 28]. However, using multimodal datasets offers a valuable opportunity to enhance the study of cell heterogeneity and dynamics by integrating consistent information from multiple modalities derived from the same cells [23].

Patch-seq, a multi-modal single-cell technology, holds particular relevance in the study of neuronal cells as it combines transcriptomic profiling with electrophysiological measurements [42, 36, 10, 9]. The electrophysiological component of Patch-seq relies on the patch-clamp technique for the study of ion channels [50, 51, 25], enabling the characterization of cellular electrophysiological properties and providing a platform for both basic and pharmacological research [17, 35, 52]. Patch-seq analysis facilitates the simultaneous investigation of electrophysiological activity in neurons and their phenotypic identity, thereby supporting studies on heterogeneity [40, 39], including the examination of cellular responses to defined stimuli to analyze patterns of cellular activation as markers of transient functional states.

During a Patch-seq analysis, cells exhibit a range of electrophysiological responses, with the AP waveform being of particular interest due to its ability to convey crucial information about neuronal type and state. Indeed, the AP waveform consists of extrinsic components that arise from the interaction of neurons with other cells and the extracellular environment, as well as dynamic regulations of membrane composition driven by neuroplasticity mechanisms operating on different timescales [16, 43]. These factors transiently affect the cellular state at the time of stimulation.

2.2. Physiology of the action potential

In neurons, the APs mediate cell-cell communication, allowing signals to propagate along the axon and reach synaptic buttons at the axon terminals [3]. An AP is characterized by a rapid sequence of voltage changes occurring across the neuronal membrane [4]. Over time, the membrane potential is determined by the relative ratio of extracellular and intracellular ion concentrations and the membrane's permeability to each ion. These dynamics follow well-conserved phases described in Section 2.2.1.

2.2.1. Membrane dynamics underlying AP

The AP can be divided into four primary phases, as illustrated in Figure 1:

1. **Resting state:** In this phase, the membrane potential is close to the equilibrium potential of potassium ions ($E_{K^+} = -75mV$), and both sodium (Na^+) and potassium (K^+) voltage-gated channels are closed (Figure 1.1).
2. **Depolarization:** When the AP is triggered, depolarization of the membrane occurs, resulting in the opening of voltage-gated Na^+ and K^+ channels. This triggers a positive feedback loop, where depolarization opens more Na^+ channels, further increasing membrane depolarization. The membrane potential approaches a positive value that is closer to the equilibrium potential of sodium ions ($E_{Na^+} = +55mV$) (Figure 1.2).
3. **Repolarization:** As the membrane potential becomes positive, Na^+ channels begin to inactivate, while K^+ channels remain open. This halts the influx of Na^+ ions into the cell and allows the efflux of K^+ ions to prevail, leading to repolarization (Figure 1.3).
4. **Hyperpolarization:** During this phase, Na^+ channels remain inactivated, and the repolarization process causes additional K^+ channels to open. Furthermore, the AP triggers the influx of calcium ions (Ca^{2+}). As a result, the membrane potential briefly becomes more negative than the resting state (Figure 1.4).

Following the peak of the AP, the voltage-gated Na^+ channels enter an inactivated state, resulting in a refractory period where subsequent stimuli can not trigger APs. Over time, the Na^+ channels gradually reactivate, while the Na^+-K^+ ion pump actively restores the resting membrane potential.

2.3. Models of the action potential waveform

Although the AP waveform is well-conserved across species and within neuronal cells, variations in the composition of ion channel populations at the neuronal membrane

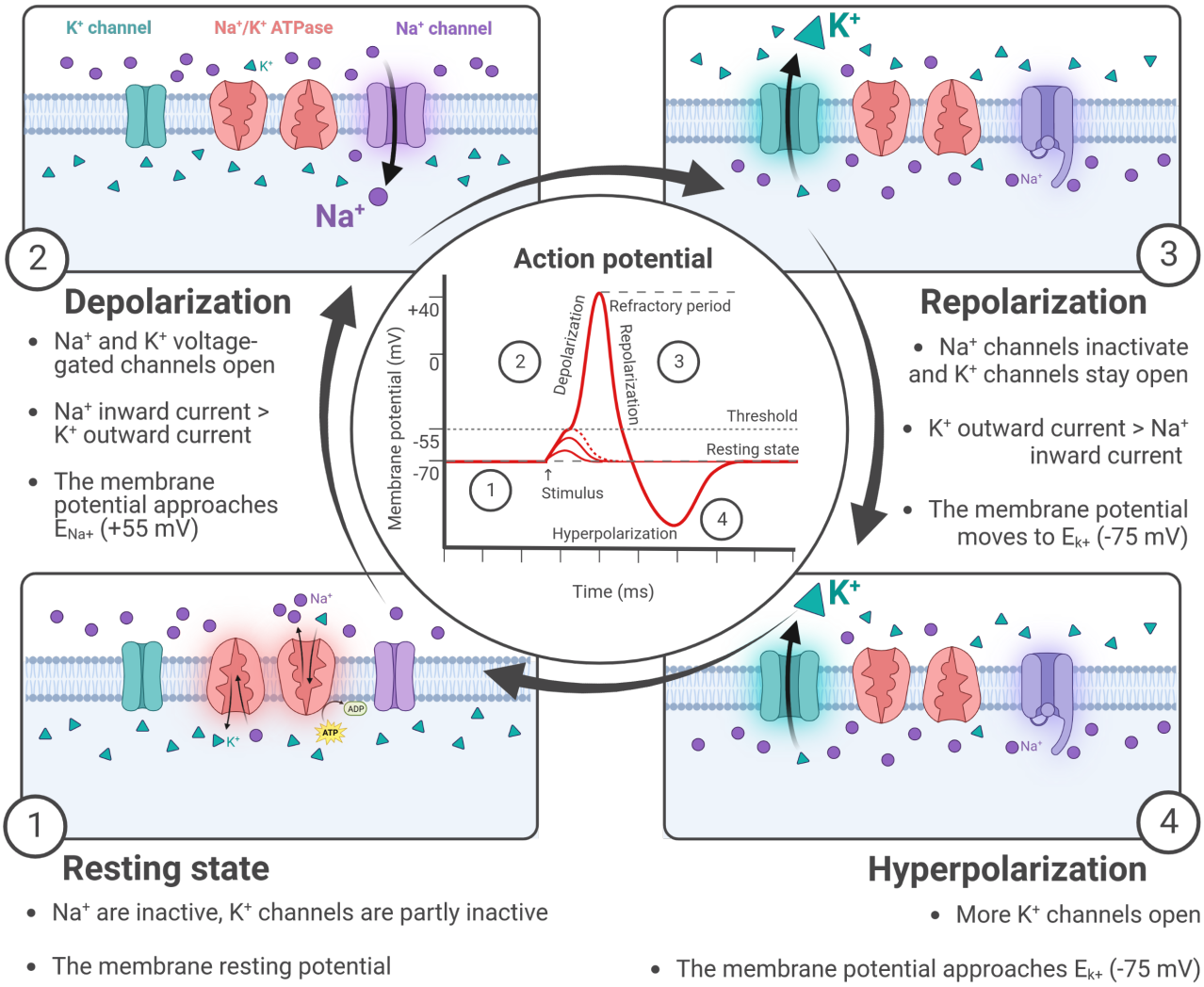

Figure 1: Membrane dynamics underlying AP. (1) *Resting state.* The membrane resting potential is close to the K^+ equilibrium potential ($E_{K^+} = -75mV$), and at this potential both Na^+ and K^+ voltage-gated channels are closed. (2) *Depolarization.* When the AP is triggered, membrane depolarization causes voltage-gated Na^+ and K^+ channels to open and starts a positive feedback where depolarization causes the opening of more Na^+ channels, which in turn further depolarizes the cell membrane, until reaching a positive value that approaches the Na^+ equilibrium potential ($E_{Na^+} = +55mV$) (3) *Repolarization.* As membrane potential gains positive values, Na^+ channels begin to inactivate, and K^+ channels stay open, stopping the Na^+ ions flux into the cell, and making the flow of K^+ ions outside the cell prevail (4) *Hyperpolarization.* Na^+ channels are still inactivated, and more K^+ channels open due to repolarization and the Ca^{2+} influx caused by the AP. Na^+ channels inactivation causes a refractory period where subsequent stimuli can not trigger APs. Then, the membrane slowly returns to its resting potential.

give rise to discernible differences in the shape of the AP. These variations also impact the overall electrophysiological responses of the cell.

The initial model of AP proposed by Hodgkin and Huxley represented its initiation and propagation in neuronal cells using a set of non-linear differential equations, which described a continuous-time dynamical system [27]. However, this model only accounted for sodium (Na^+) and potassium (K^+) conductances, each with a single type of voltage-dependent channel. This simplistic approach disregarded the complexity of voltage-dependent conductances in neuronal membranes and limited the exploration of the underlying biomolecular substrates of electrophysiological behavior. In reality, the neuronal membrane exhibits a wide diversity of conductances. All voltage-gated currents, including Na^+, calcium (Ca^{2+}), and K^+ currents, consist of at least two distinct components [4]. Additionally, other currents are activated by membrane hyperpolarization [6]. The complexity and diversity of the AP conductances are attributed to a wide range of types and isoforms of ion channels [31, 71]. The relative prevalence of different channel populations at the membrane contributes to the EP features of neuronal action potentials. The interplay and regulation of channel activation among these populations present a significant

challenge in understanding and modeling this phenomenon [4]. Alterations in the underlying biomolecular ion channel composition impact AP shape, firing rate, and the emergence of distinct higher-level patterns in the electrophysiological phenotype of different types of neurons [4].

The intrinsic components of the AP waveform generated by a particular neuronal type are regulated by its electrophysiological and morphological features, as well as its belonging to specific cell-type families (as demonstrated by the taxonomy described in [21]). In fact, the AP waveform exhibits substantial differences among various types of neurons in the mammalian brain [4]. Consequently, the analysis of AP waveforms alone facilitates the study of cellular heterogeneity across interneurons [12]. Mathematical models of AP shape have supported the construction of a coherent neuronal taxonomy that aligns with the transcriptomic profiles of cells [56].

The proposed approach, called Neuronal Spike Shapes (NSS) analysis, captures a wealth of information about neuronal cells and populations by examining AP waveforms. Given the intricate structural and functional interconnections among neurons within brain neuronal networks, each AP waveform property is intrinsic, dependent on the cell itself, and extrinsic, influenced by stimulation and cell connections during stimulation. Consequently, the AP waveform provides insights into cell types, cell states, and functional activations of neurons within their respective neuronal networks.

This study applies the NSS analysis to two datasets of murine cortical interneurons. The first dataset provides solely electrophysiological data [22], while the second dataset offers a comprehensive characterization of cells encompassing electrophysiological, transcriptomic, and morphological aspects [21]. Validation demonstrates that NSS-based analysis facilitates the detection of neuronal clusters that are independent of cell type and allows for the investigation of their biological relevance as neuronal states.

3. Methods

The NSS method utilizes a simplified triangular representation of the AP waveform (see Section 3.1), enabling a concise description of individual neuronal spikes. This triangular shape serves as a basis for computing mathematical electrophysiological (EP) features, which summarize the shape and duration of the AP generated by a particular neuronal cell.

3.1. Neuronal Spike Shape (NSS): a simple triangular approach

The AP waveform exhibits a well-conserved spike shape in the time-voltage domain, as depicted in Figure 2.

Despite variations observed across different cell types and states, all AP signals can be characterized by well-conserved key points contributing to the spike shape. Within the NSS framework, three points are considered particularly informative regarding the AP:

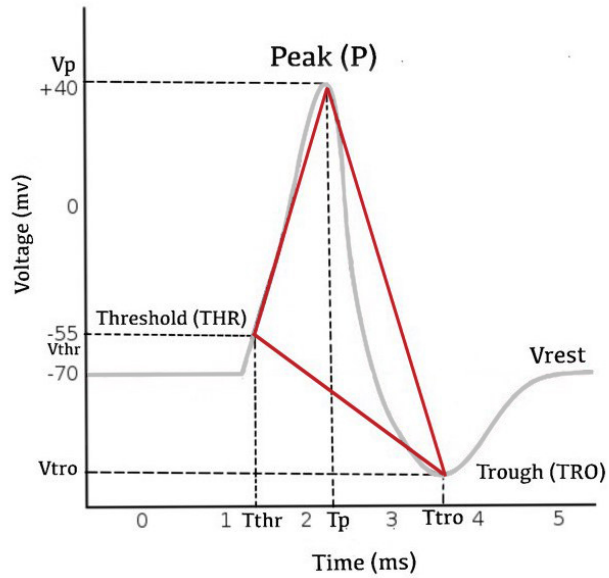

Figure 2: Example of neuronal AP. Threshold (THR), Peak (P), and Trough (TRO) points are the vertices of the NSS triangular representation (in red) overimposed to the AP waveform.

- **Threshold (THR)**: This point corresponds to the minimum depolarization of the cell membrane relative to the resting potential (see Figure 1.1) required to initiate an AP. The threshold dynamically adapts based on previous membrane activations [54] and varies depending on the cell type and state. For example, the voltage threshold for AP generation in cortical interneurons of the Fast-spiking (FS) and Non-Fast-spiking (NFS) types was found to be approximately -31.6 ± 5.0 mV and -39.6 ± 6.4 mV, respectively [20].

- **Peak (P)**: This point corresponds to the highest depolarized voltage attained during AP generation (i.e., the peak of the spike). At the peak, the membrane exhibits maximum permeability to sodium ions (Na^+), causing the membrane potential to approach the Na^+ equilibrium potential ($E_{Na^+} = +55mV$). Peak potential typically ranges from 0 to +40 mV, as reported in [5] (see Figure 1.1).

- **Trough (TRO)**: This value, as defined in [22] and [21], represents the hyperpolarization phase (see Figure 1.4). It can be further divided into two sub-points: the *Fast Trough (FTRO)* corresponding to the most negative membrane potential within 5 ms after the peak time, and the *Slow Trough (STRO)*, corresponding to the most negative membrane potential between 5 ms after the peak time and the subsequent action potential threshold time. During the hyperpolarization phase, the membrane potential approaches the potassium ion (K^+) equilibrium potential ($E_{K^+} =$

$-75mV$), with values typically ranging from -70 to $-75mV$, as reported in [5].

The key points P, THR, and TRO (specifically the FTRO) provide a concise yet highly informative representation of the shape of the AP. This representation is a simplified triangular structure, highlighted in red in Figure 2. The triangular representation serves as the fundamental framework of the NSS approach. The coordinates of its three vertices are determined by the voltage values (V_{THR}, V_P, V_{FTRO}) and time values (T_{THR}, T_P, T_{FTRO}) specific to each analyzed AP potential. Furthermore, the sides of the triangle summarize various functionally relevant geometric properties of the AP waveform. Specifically, the THR-P side approximates the depolarization phase (see Figure 1.2), the P-$FTRO$ side represents the repolarization phase (see Figure 1.3) and hyperpolarization phase (see Figure 1.4), while the THR-$FTRO$ side reflects the overall duration of the AP and the depth of the hyperpolarization.

3.2. NSS Features

By utilizing the abstract triangular simplification at the core of the NSS approach, it becomes possible to assess voltage and time differences, ratios between relevant triangle parameters, and slopes of lines intersecting its vertices. These measurements contribute to a concrete, multi-dimensional AP spike description. When the time (T_P, T_{THR}, T_{FTR}) and voltage (V_P, V_{THR}, V_{FTRO}) coordinates are known for all three vertices (P, THR, and FTRO), the necessary information for evaluating these features is available. These properties can be derived directly from raw EP signals obtained in patch-clamp experiments or as precomputed derived properties exposed in EP datasets.

In this work, the focus was on features generated solely from ramp stimulation. Ramps offer the advantage of rapidly and directly generating current-voltage relationships, making them suitable for analyzing cellular responses characterized by rapid activation or time independence [59]. This is particularly relevant for assessing the onset of the AP [59], as ramps help avoid artifacts arising from desensitization [72]. Additionally, the fast exploration of the entire voltage range in voltage-clamp stimulation enables more precise threshold detection. This is important, considering that the threshold for AP generation exhibits a dynamic range variation depending on the interneuron subtype [34] and activation-based neuroplasticity [16].

Table 1 presents a set of features that characterize the geometric properties of the triangular NSS representation. These properties offer a comprehensive overview of the depolarization phase (see Figure 1.2), as well as the repolarization and hyperpolarization phases (see Figure 1.3-4), providing a synthetic summary of the complete AP shape. It is worth noting that these features focus on modeling the AP shape rather than its position in the time-voltage plane. Therefore, they exclude absolute time and voltage values, emphasizing the relative relationships between absolute voltage and time values at the three key points, expressed as differences or ratios. This differential approach allows for a more generalized description of the AP shape, independent of absolute values. As shown, the NSS framework supports flexible analysis that can adapt to the specific characteristics of the data in different datasets.

Table 1
NSS features, along with description and unit of measurement.

Feature	Mathematical Description	Unit
$UpDown_{ratio}$	$\frac{S_{up}}{S_{down}}$	–
$Slope_{deep}$	$\frac{\lvert V_{FTRO}-V_{THR}\rvert}{\lvert T_{FTRO}-V_{THR}\rvert}$	V/s
$AP_{halfwidth}$	$\frac{T_P-T_{THR}}{2}$	μs
$Down_{width}$	$T_{FTRO}-T_P$	μs
$UpDown_{width}$	$T_{FTRO}-T_{THR}$	μs
$Width$	$\frac{T_P-T_{THR}}{2}-\frac{T_{FTRO}-T_P}{2}$	μs
$Height$	$\lvert V_P-V_{FTRO}\rvert$	V
ΔV_{deep}	$\lvert V_{FTRO}-V_{THR}\rvert$	V
ΔV_{THRP}	$\lvert V_{THR}-V_P\rvert$	V
ΔV_{ratio}	$\frac{\lvert V_P-V_{THR}\rvert}{\lvert V_P-V_{FTRO}\rvert}$	–

Finally, it is important to highlight the biunivocal mapping between each neuron and the multidimensional point described by the computed NSS features of its AP. This mapping enables the application of clustering and classification algorithms to investigate the heterogeneity of neuronal cells based on their EP profiles alone or in combination with multimodal analysis.

4. Results

4.1. Validation by multimodal analysis

This section utilizes multimodal analysis to validate the results of the NSS approach based on cell type labels and to identify cells with similar EP profiles across different transcriptional subtypes. The analysis is conducted on two murine cortical interneurons datasets, described in Section 4.2. The multimodal analysis results for these datasets are presented in Sections 4.3 and 4.4.

4.2. Datasets

The presented analysis relies on two datasets of murine visual cortical neurons. Table 2 illustrates their characteristics. *Gowens2020 (PatchSeqDataset)* [21] leverages the patch-seq technique, that combines patch-clamp with scRNA-seq to collect EP and transcriptomic profiles of

Dataset ID	Specimen	Cells	EP	Morph	scRNA-seq	References
PatchSeqDataset	Mouse visual cortex	GABAergic interneurons	4,200	517	3371	[21]
PatchClampDataset	Mouse visual cortex	Neurons	1,938	461	-	[22]

Table 2
Datasets for NSS analysis. For each dataset, the number of cells profiled under EP, morphological and transcriptomic profiles is provided. EP: Electrophysiological profile; Morph: morphological reconstructions; scRNA-seq: transcriptomic profile.

4,200 mouse visual cortical GABAergic interneurons, reconstructing the morphological conformation of 517 of them.

Gowens2019 (PatchClampDataset) [22] provides EP data from 1,938 neurons of adult mouse visual cortical neurons and morphological reconstructions for 461 of them. This dataset, accessible online via the *Allen Brain Atlas Cell Types Database* [2], provides a mixture of spiny and aspiny neurons. This work analyzed only cells labeled as *aspiny*, under the consensus that such a morphological label is a strong indication that the cell is an interneuron.

The main analysis presented in this study focuses on the *PatchSeqDataset*, as it provides both EP and transcriptomic data, enabling direct cross-validation of the NSS-based heterogeneity analysis with transcriptomic labels as a ground truth. The *PatchClampDataset* serves as a soft validation of the results, as it provides reliable EP characterization of individual cells. However, the phenotype profiling is based on coarse-grained labeling of cell types using transgenic lines (see Section 4.4).

Both *PatchClampDataset* and *PatchSeqDataset* provide comprehensive EP characterization of single cells, including voltage-clamp stimuli and responses collected for each sample. Each cell in the datasets is associated with multiple stimulation types, corresponding to standard patch-clamp protocols such as current ramps, short squares, and long squares. For each stimulation type, there are multiple *sweeps*, which are repeated trials where the parameters of the stimulation protocol vary, resulting in different EP responses. For example, the long square stimulation set includes a sweep for each step current amplitude applied as input to the neuron.

4.3. Validation on *PatchSeqDataset*

The validation of the NSS approach on *PatchSeqDataset* begins with a canonical unsupervised exploration of the features presented in Table 1. Figure 3 provides an overview of the sequence of actions performed during this validation process.

Following data cleaning, all the NSS features listed in Table 1 are evaluated to identify relevant trends in their distribution across the dataset. Subsequently, automatic clustering is performed using Hierarchical Agglomerative Clustering (HAC) on the standardized features obtained through z-score normalization. This clustering process based HAC on *z-score* standardized features aims to identify groups of inhibitory neurons within the dataset.

The final step involves analyzing the importance of the NSS features with respect to these identified partitions. This analysis is carried out using two approaches: Spearman's correlation and the *permutation feature importance* method based on an Random Forest (RF) classifier [7]. These approaches enable the detection of non-linear relationships between the NSS features and the identified groups of inhibitory neurons.

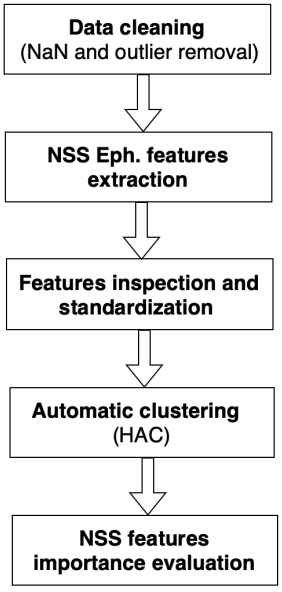

Figure 3: Pipeline of EP exploration using NSS on *PatchClampDataset* and *PatchSeqDataset*.

4.3.1. NSS features preliminary analysis

After data cleaning on *PatchSeqDataset*, 3,653 neurons are retained for further analysis. The data cleaning process involves removing cells that had *NaN* values for NSS features. Additionally, cells with apparent outliers in their feature values, potentially caused by errors in the patch-clamp experiment, are also excluded.

Subsequently, the distribution of NSS features among the selected cells is examined through histograms. Figure 4 displays the frequency distribution of certain NSS features, namely $UpDown_{ratio}$, ΔV_{deep}, ΔV_{ratio}, and $Slope_{deep}$. These histograms clearly reveal multiple peaks in the distribution of these features across cells.

Notably, the distribution of these features in the APs generated by inhibitory neurons suggests the presence of more than one modality in the data. Specifically, the prominent peaks observed in the distribution of ΔV_{deep} and ΔV_{ratio} indicate the existence of at least two modalities. There may even be three modalities if we consider the local maxima

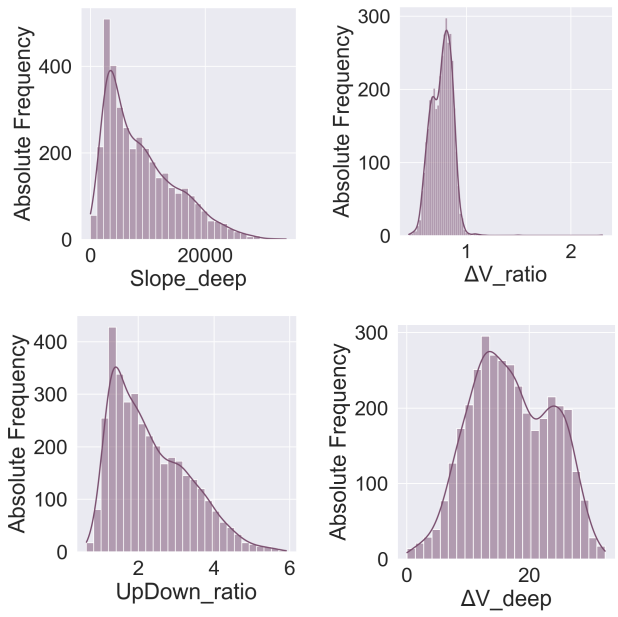

Figure 4: Histograms of $Slope_{deep}$ (top-left), ΔV_{ratio} (top-right), $UpDown_{ratio}$ (bottom-left), ΔV_{deep} (bottom-right), NSS features exhibiting multiple modalities in their distribution over *PatchSeqDataset*.

in the shape of $Slope_{deep}$. In contrast, the remaining features exhibit a single-modal distribution, providing limited insights into specific trends in the data (data not shown).

4.3.2. Clustering neurons using NSS features

The clustering process utilizes NSS features and applies HAC to identify significant groups of cells that potentially represent distinct cell types or transient states. The choice of using HAC offers several advantages: it is not sensitive to initialization conditions (unlike methods like K-Means) or assumptions about cluster shapes; it is robust against outliers in the data; and it allows for examining data groups at different levels of granularity. The algorithm requires selecting a distance measure to assess the proximity between individual elements and a merging approach for combining separate clusters during the iterative construction of the final clustering. This study chooses Euclidean distance and Ward linkage as they are commonly employed in the literature to identify compact clusters. The determination of the number of clusters, denoted as K, involves analyzing the dendrogram generated by the HAC algorithm and optimizing it a priori using the Silhouette score maximization [57]. The Silhouette metric s is calculated as follows:

$$s = \frac{b - a}{max(a, b)} \tag{1}$$

where the average intra-cluster distance a quantifies intra-cluster cohesion, and the average inter-cluster distance

Figure 5: Silhouette plot for identification of optimal number of clusters: $K = 3$ is the best configuration for HAC.

b measures inter-cluster separation. Unlike other metrics such as the Sum of Square Errors (SSE), the Silhouette metric is robust to non-spherical clusters.

Figure 5 presents the optimization results: the optimal solution, corresponding to the maximum Silhouette score, occurs when $K = 3$ clusters are formed. However, it is worth noting that $K = 2$ and $K = 3$ yield very similar results (i.e., 0.32 versus 0.34). Considering this similarity, the multimodal analysis includes and compares NSS clustering with $K = 2$ and $K = 3$. The case of $K = 4$ clusters might also be relevant since it aligns with the number of cell types identified through transcriptomic clustering. However, comparing the $K = 4$ clustering and transcriptomic labels does not yield significant results (see Section 4.3.4). Therefore, the multimodal analysis excludes the $K = 4$ clustering while favoring cases with higher Silhouette scores.

Figure 6 illustrates scatter plots of the data based on the three most relevant principal components (PCs) obtained from PCA analysis of NSS features. The dot colors represent the cluster labels from the clustering process with $K = 3$ (or $K = 2$), denoted as *EC0*, *EC1*, and *EC2*. The scatter plots demonstrate that using $K = 3$ divides the largest cluster from the $K = 2$ case (*EC1*) into two separate clusters, resulting in three clusters overall. However, the cluster *EC0* with $K = 2$ remains unchanged in the $K = 3$ case, indicating its compactness irrespective of the variation in cluster cardinality.

The clustering results based on NSS provide valuable characterizations of the neuronal population under investigation in terms of EP. Furthermore, the multimodal analysis (see Section 4.3.4) enhances the biological relevance of the NSS clusters by comparing and jointly analyzing them with transcriptomic clusters, as well as exploring their transcriptional associations.

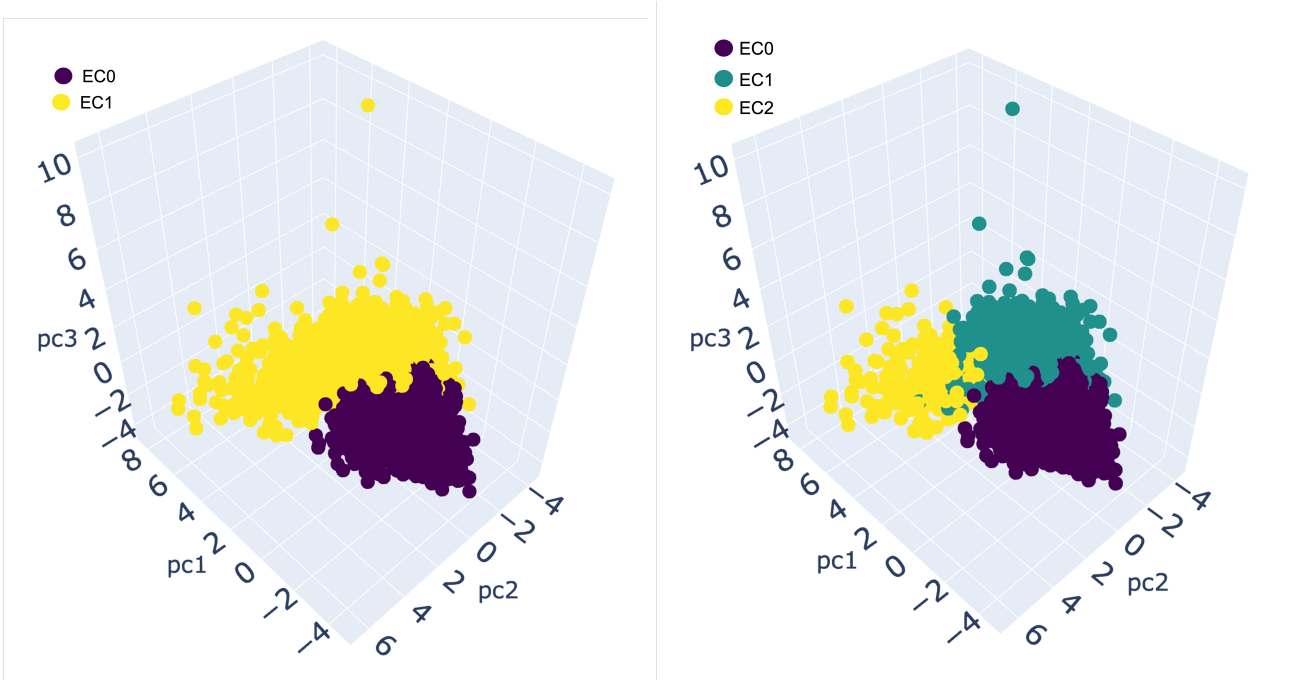

Figure 6: NSS clustering results. Tri-dimensional PCA plots along the three most relevant PCs for $K = 2$ (left) and $K = 3$ (right). *EC0* (the compact violet cluster on the right) remains unchanged between the $K = 2$ and the $K = 3$ cases, while using $K = 3$ causes the largest cluster of the $K = 2$ case (*EC1*) to split in two (*EC1* and *EC2*) in the $K = 3$ case. The color coding refers to clustering labels: in $K = 2$, violet for *EC0* and yellow for *EC1*; in $K = 3$, violet for *EC0*, green for *EC1* and yellow for *EC2*.

4.3.3. NSS features relevance

The relevance of NSS features in relation to the clustering labels was assessed by calculating Spearman's correlation coefficient (ρ), as shown in the upper triangular correlation matrices in Figure 7 (column *LABEL*).

In general, it can be observed that the NSS features associated with similar characteristics exhibit high correlations with each other, indicating a potential redundancy of information. Specifically, the NSS features ΔV_{deep}, ΔV_{ratio} and ΔV_{thrp} display consistent correlations in both the $K = 2$ and $K = 3$ clustering cases, albeit with a slightly lower ρ value in the latter case. Conversely, when considering $K = 2$, other features related to the width of the AP spike (e.g., $Down_{width}$, $UpDown_{width}$, $Width$) and the slope (e.g., $UpDown_{ratio}$, $Slope_{deep}$) become relevant for the clustering process. This behavior can be better understood by examining the Kernel Density Estimate (KDE) plots of the feature $Slope_{deep}$ presented in Figure 8. In the $K = 3$ clustering, the peaks of the distributions in *EC0* and *EC1* are close, and the two curves partially overlap, making it challenging to distinguish the clustering label. However, in the $K = 2$ case, *EC1* is merged with *EC2*, resulting in a more distinguishable distribution compared to *EC0*. Figure 8 also illustrates that when $K = 3$, the feature ΔV_{ratio} exhibits three distinct distributions, highlighting its importance in differentiating between APs of neurons in *EC1* and *EC2*.

To further explore the relevance of NSS features, Figure 9 presents the mean accuracy decrease in RF classification of the clustering label after randomly shuffling the values of each NSS feature. The results are provided for both the $K = 2$ and $K = 3$ clustering cases. In the $K = 2$ case, the feature with the most significant impact is $Slope_{deep}$, as altering its values results in a 10% decrease in accuracy. This substantial impact seems to overshadow the contributions of the other features. However, when transitioning to $K = 3$, the importance of $Slope_{deep}$ becomes comparable to that of the other features. In the $K = 3$ clustering, the permutation method highlights the importance not only of features already identified using correlation (such as ΔV_{ratio}), but also of those without a direct linear relationship with the clustering label, such as $UpDown_{width}$ and $UpDown_{ratio}$, which cause approximately a 4% and 2% decrease in mean accuracy, respectively.

To further validate the effectiveness of NSS features in characterizing the waveform of action potentials (APs), an analysis is conducted on the shape of the median APs for each cluster in the $K = 3$ case. This analysis specifically focuses on the triangular simplification proposed by the NSS approach. The results are presented in Figure 10, revealing that the NSS analysis successfully identifies three distinct APs in the dataset. The width of the curve and the repolarization phase of the waveform are found to be the primary factors explaining the variability across clusters, particularly in the region near the FTRO. On the other hand, the thresholds (THR) and the peak (P) are mostly consistent across the three clusters.

Specifically, the cluster labeled *EC0* exhibits a narrow spiking response with a deep FTRO, while *EC2* displays

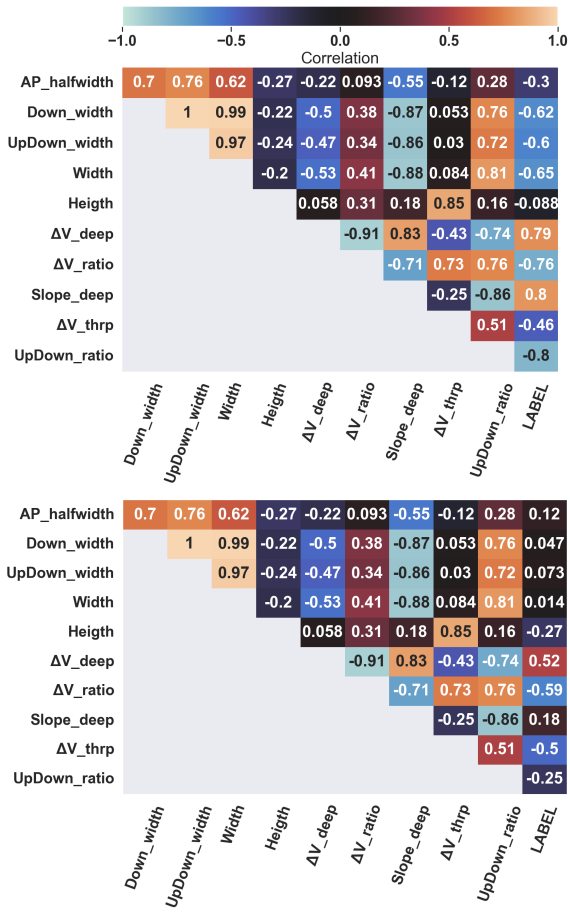

Figure 7: Spearman's correlation of NSS features to clustering label: correlation to $K = 2$ clusterization (top); correlation to $K = 3$ (bottom). An hot (high positive correlation) and cold (high negative correlation) diverging color map represent the correlation levels.

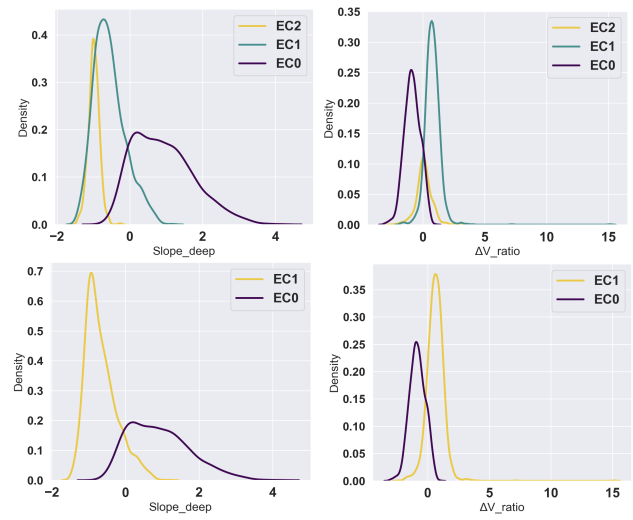

Figure 8: KDE plots of the distribution of features $Slope_{deep}$ and ΔV_{ratio}. On top, distributions in *EC0* (violet), *EC1* (green), and *EC2* (yellow) for $K = 3$ case; at the bottom, distributions in *EC0* (violet) and *EC1* (yellow) for $K = 2$ case.

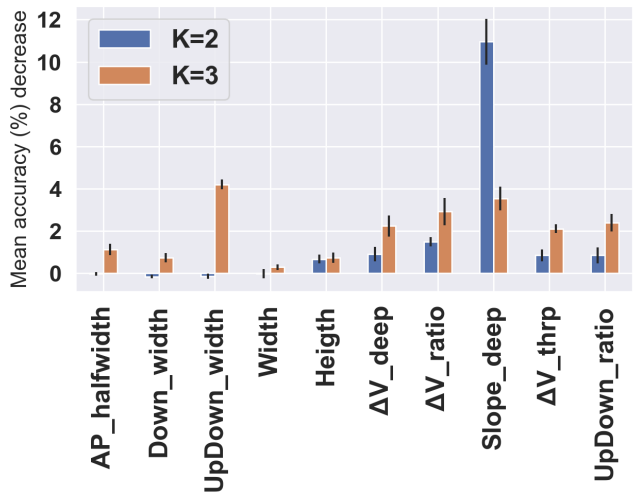

Figure 9: Mean accuracy decrease in the cluster label prediction done by RF model when shuffling values of NSS features: case with $K = 2$ clustering in blue; case with $K = 3$ clustering in orange.

the opposite behavior, characterized by a gradual repolarization phase and an almost flat FTRO. *EC1* occupies an intermediate position between these two extremes. Notably, this simplified model effectively distinguishes specific spike responses in the data. Furthermore, this result aligns with the investigation of feature importance (refer to Figure 7, 8, and 9), highlighting the significance of spike properties related to the repolarization and hyperpolarization phases (see Figure 1.3-4).

Finally, the fact that the THR and P points alone do not possess strong discriminatory properties supports the decision to prioritize difference- and ratio-based properties of the AP waveform over absolute voltage and time values in the context of NSS.

4.3.4. NSS and transcriptomic joint analysis

Given that the *PatchSeqDataset* includes scRNA-seq data, multimodal analysis allows for the examination of gene expression in neurons. Since the EP and transcriptomic profiles originate from the same cells in the dataset, it is straightforward to transfer NSS-based labels to the transcriptomic clustering results. This approach facilitates cross-validation and joint analysis of EP and transcriptomic findings.

The analysis workflow for scRNA-seq data is well-established in the research community [38]. In this study, the Seurat pipeline [23] was employed for transcriptomic analysis.

The dataset consists of 3,840 cells already filtered to remove low-quality cells. After excluding a few excitatory neurons, the remaining cells are 3,371 GABAergic neurons. Moreover, by considering only the cells that passed the EP filtering, the final dataset consists of 3,157 cells. The initial result of the transcriptomic analysis is the unsupervised

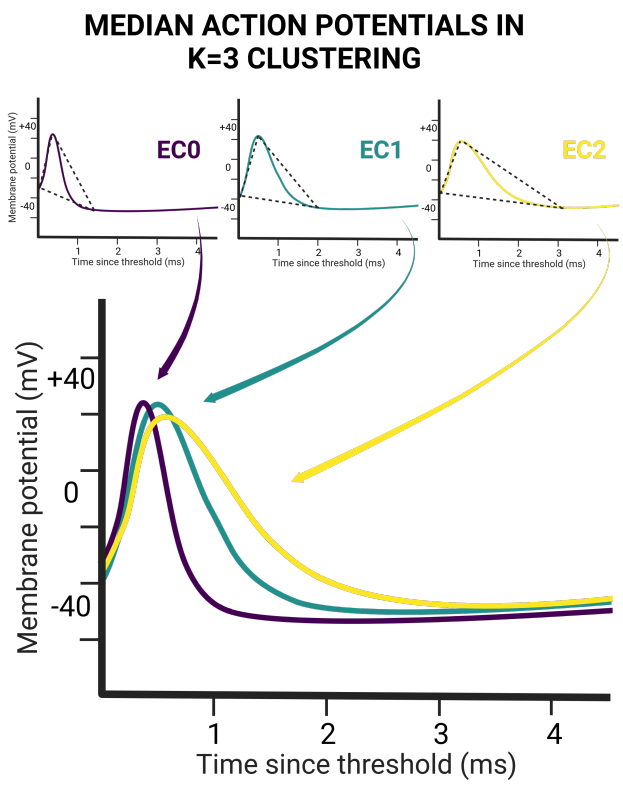

Figure 10: Median AP spikes for the $K = 3$ clustering and their superimposition. NSS backbones THR-P-FTRO triangles are reported as dashed line over the spiking graphs.

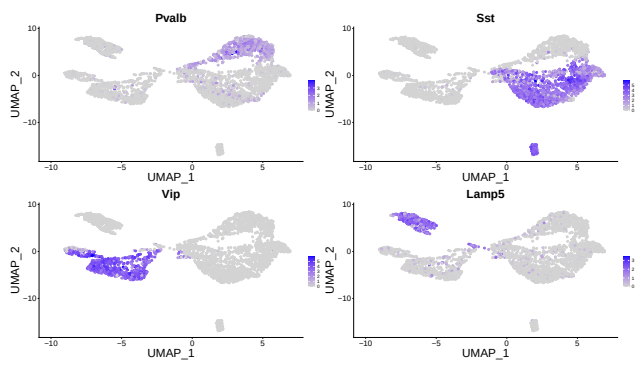

Figure 11: Cell type markers expression levels. Visualization of the four marker genes expression, namely Pvalb (top right), Sst (top left), Vip (bottom right), and Lamp5 (bottom left). Each marker denotes a specific cell population, coherently with the *PatchSeqDataset* metadata cell type labels in Figure 12. A color intensity scale marks the expression levels.

clustering of cells on the Uniform Manifold Approximation and Projection (UMAP) embedding, which enables a two-dimensional (2D) visualization. These transcriptomic clusters capture the overall heterogeneity of cell types, serving as the foundation for subsequent analyses.

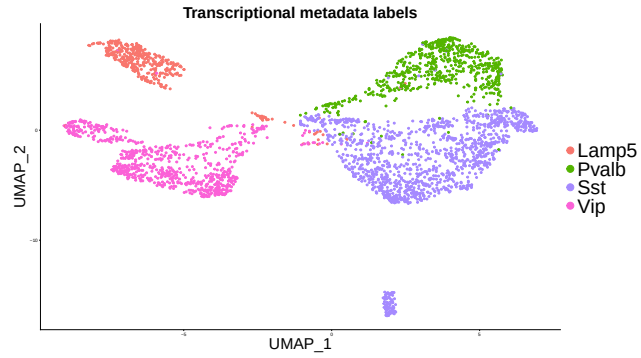

Figure 12: Cell type labels. Cells colored with the transcriptional labels provided by *PatchSeqDataset* metadata, where cells labeled as "Serpinf1", "Sncg" and "Meis2" are grouped under the broaded cell subtype "Vip". The four inhibitory types cover the four macro populations, coherently with the gene expression patterns from Figure 11. Colors refer to cell type labels: Pvalb (top right, green), Sst (bottom right, purple), Vip (bottom left, pink), and Lamp5 (top right, salmon).

Transcriptomic subtypes labeling Figure 11 and Figure 12 identify four distinct GABAergic neuronal subtypes by examining the expression of cell type marker genes and the distribution of cell type labels provided in the *PatchClampDataset* metadata [21]. Figure 11 depicts the expression levels of these marker genes across the cells, delineating the following GABAergic subtypes: Pvalb (top right, green), Sst (bottom right, purple), Vip (bottom left, pink), and Lamp5 (top right, salmon). These marker expression levels align with the cell type labels provided in the *PatchClampDataset* metadata, as shown in Figure 12.

NSS labels on transcriptomic embeddings Building upon this mapping, the present study examines the distribution of EP labels across transcriptomic subtypes by visualizing NSS cluster labels on the transcriptomic embedding. Figure 13 illustrates the clustering results for both $K = 2$ and $K = 3$. In both cases, the Pvalb area exhibits a clear predominance of *EC0*, while the Sst region displays a mixture of *EC0* and *EC1*.

In contrast, the Vip/Lamp5 regions differ between the $K = 2$ and $K = 3$ cases. With $K = 2$, the *EC1* cluster prevails in both regions. However, with $K = 3$, the *EC1* cluster prevails in the Vip region, while *EC2* prevails in the Lamp5 region.

Considering that GABAergic neurons consist of four cell subtypes, the analysis also included the $K = 4$ NSS clustering case, despite its slightly inferior Silhouette scores compared to the other two cases (see Section 4.3.2). This is done to investigate whether the NSS analysis was merely recapitulating transcriptomic clustering. However, the $K = 4$ case introduced only a further split within the *EC1* cluster, as revealed by the corresponding NSS labels on the transcriptomic embedding. This additional grouping neither recapitulates transcriptomic cell type regions nor provides

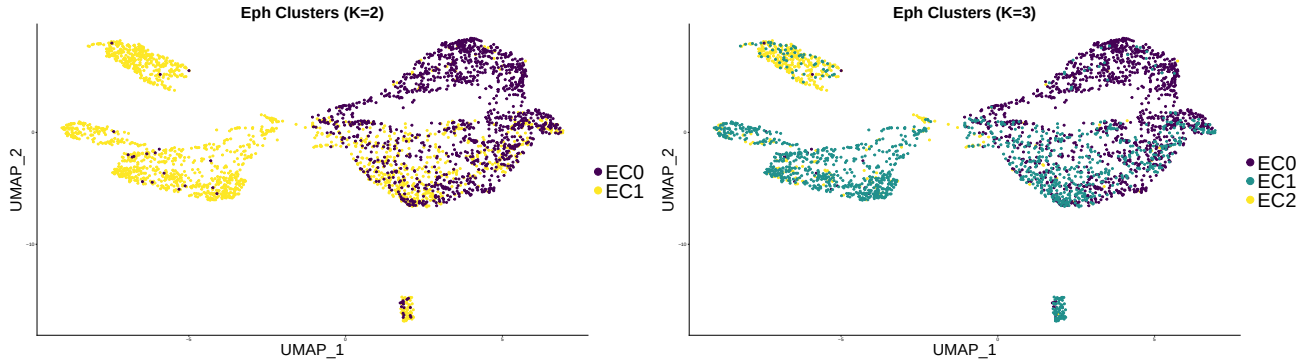

Figure 13: NSS labels on transcriptomic embeddings. Cells from the transcriptomic embedding, colored with the EP clusters, including both the $K = 2$ (left panel) and $K = 3$ (right panel) cases. In both cases, the *EC0* cluster (red) overlaps most of the Pvalb population and part of the Sst populations. In $K = 2$, the *EC1* cluster (blue) lies onto the remaining part of the Sst population, and on the Vip and Lamp5 populations. In $K = 3$, the *EC1* cluster (green) lies onto the remaining part of the Sst population, and on the Vip population mainly overlaps the Lamp5 cells.

more informative insights than the other cases (data not shown). Therefore, the analysis focuses solely on $K = 2$ and $K = 3$.

Based on these results, it can be concluded that NSS clustering effectively captures differences among GABAergic neuronal subtypes. Notably, the *EC0* cluster, which overlaps with the entire Pvalb region, exhibits the narrowest AP shape among the three clusters (see Figure 10). This finding aligns with the well-known characterization of Pvalb neurons as *fast-spiking* [49], a phenotype that is consistent with a narrow AP shape. However, it is worth noting that *EC0* does not exclusively overlap with the Pvalb region but also partially with the Sst region. This observation highlights that the NSS clusters conform to cell type-specific descriptions and extend beyond capturing cell type heterogeneity.

The transcriptomic analysis is further performed at a higher resolution to conduct a more in-depth investigation by dividing the cells into four subsets based on the metadata cell type label (Pvalb, Sst, Vip, and Lamp5). Each of these subsets undergoes the same clustering analysis as the entire dataset (see Section 4.3.4).

A focus on Sst cells The Sst subtype is of particular interest in this context. In both the $K = 2$ and $K = 3$ NSS clustering results, the examined cells predominantly display either *EC0* or *EC1* NSS clustering labels. This suggests that a subset of Sst cells exhibits NSS labels more similar to the Pvalb subtype, while the remaining cells possess an EP profile more akin to the Vip/Lamp5 subtype.

Furthermore, the mixed composition of Sst cells in terms of NSS labels persists even when considering Sst subtypes. Figure 14 illustrates the embedding of the Sst subset, with two sets of labels: the *PatchSeqDataset* metadata transcriptional cell sub-type labels, which highlight further divisions within the Sst subtype (left panel), and the NSS cluster labels (right panel). The figure demonstrates a very low coherence between Sst subtypes and NSS clusters, supporting the observation that NSS clusters exhibit cell-type-independent heterogeneity within the Sst group, regardless of the resolution of cell type labeling.

4.3.5. Multimodal data exploration

The multimodal nature of the *PatchSeqDataset* enables a more comprehensive exploration of the intricate relationships between the two biological levels. Transcriptomic data allow Differential Expression (DE) analysis between NSS clusters to investigate gene expression profiles that could shed light on the underlying biological mechanisms. Specifically, within the same transcriptomic subset, groups of cells with different NSS labels are compared to identify intra-type differences. The analysis provides a list of the top DE genes, highlighting the genes that characterize each group primarily and specifically compared to the other.

Analyzing heterogeneity in Sst cells' electrophysiology Due to the significant intra-subtype heterogeneity in electrophysiology observed in the Sst subset at various resolution levels, the analysis examines DE genes between the two NSS clusters within the Sst embedding.

To further investigate the resulting lists of DE genes systematically, Gene Ontology (GO) Enrichment Analysis (EA) [45] is carried out. This approach helps in generating hypotheses based on knowledge regarding the involvement of specific gene sets in biological processes. The GO is a curated and systematic nomenclature and annotation of biological processes, molecular functions, and cellular components associated with each gene. The GO EA of a gene list highlights the overrepresented processes and functions in the GO. Additionally, the analysis involves a manual investigation of the functional aspects of specific genes. Platforms such as GeneCards [62] or Uniprot [14] provide extensive and curated information about genes and their protein products to support these investigations. This work uses the GO platform supported by PANTHER [45], using the ShinyGO tool [19].

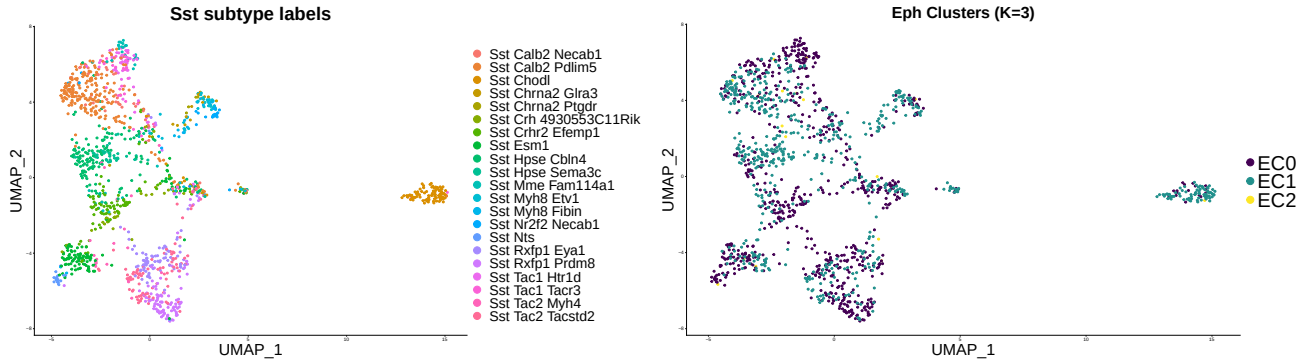

Figure 14: Sst cells labeled with the transcriptional cell sub-types provided by the metadata (on the left) and the EP clusters (on the right).

Specifically, the GO EA targets the 111 genes that are positively DE in *EC0* cells compared to *EC1* cells within the Sst subset, considering the *biological process*, *molecular function*, and *cellular component* ontologies.

As depicted in Figure 15, there is an enrichment of ontology elements related to potassium channels (K^+) (highlighted in the figure), particularly in the *molecular functions* ontology, where "Voltage-gated K^+" and the "K^+ channel activity" are enriched. The DE genes list contains numerous entries associated with K^+ channels, which are characterized by the "*Kcn*" naming pattern [58].

This raises the question of the significance of the enrichment of the K^+ channel in this multimodal analysis. As described in Section 2.2 and illustrated in Figure 1, K^+ currents and the underlying K^+ channels play a major role in the repolarization and hyperpolarization phases of the AP. Therefore, differences in the expression of K^+ channels, even through complex post-transcriptional regulations, manifest themselves as differences in the AP waveform. Specifically, a greater presence of K^+ channels at the membrane, facilitated by their increased gene expression, promotes a more pronounced K^+ current during repolarization, resulting in a steeper descent of the AP waveform after the peak.

This observation, which characterizes the *EC0* cluster and is derived from the GO EA of NSS labels within the Sst subtype, aligns with the NSS-based description of median AP shapes. The *EC0* cluster, characterized by a specifically higher expression of genes related to the K^+ channels, exhibits a narrower spike, particularly a steeper descent, as depicted in Figure 10.

Consequently, it can be concluded that, within the Sst population, the two NSS clusters primarily differ in the repolarization phase of the AP waveform at the EP biological level. They also exhibit significant and specific differences in the gene expression of underlying K^+ channels at the transcriptomic level. The NSS-based cell subsets are independent of cell type, as demonstrated by their distribution across regions corresponding to all subtypes in the Sst-specific embedding. These findings strongly support the hypothesis that NSS analysis effectively captures cell heterogeneity based solely on EP features. Furthermore, they suggest the identification of a cell-type-independent grouping of cells, which may indicate distinct cell states.

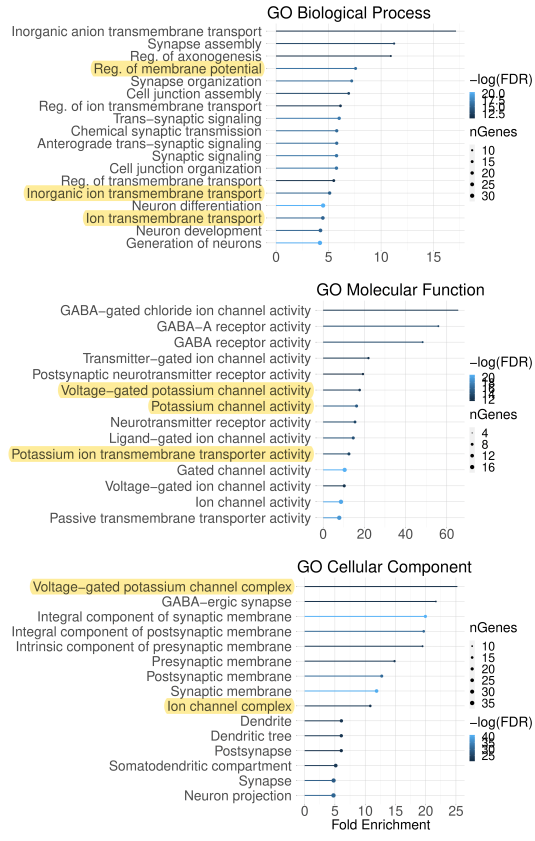

Figure 15: GO terms enrichment for Biological Process (Top), Molecular Function (Center), and Cellular Component (Bottom). The x axis represents the fold change, ranked from the highest. The highlighted terms are the ones directly correlated to K^+ channels.

Expanding the Analysis to the Entire Dataset The next step is to investigate whether the relationship between the AP waveform and specific gene expression patterns remains consistent across the entire dataset. To address this question, a DE analysis is performed on the complete dataset, first comparing the *EC0* and *EC1* clusters and then comparing the

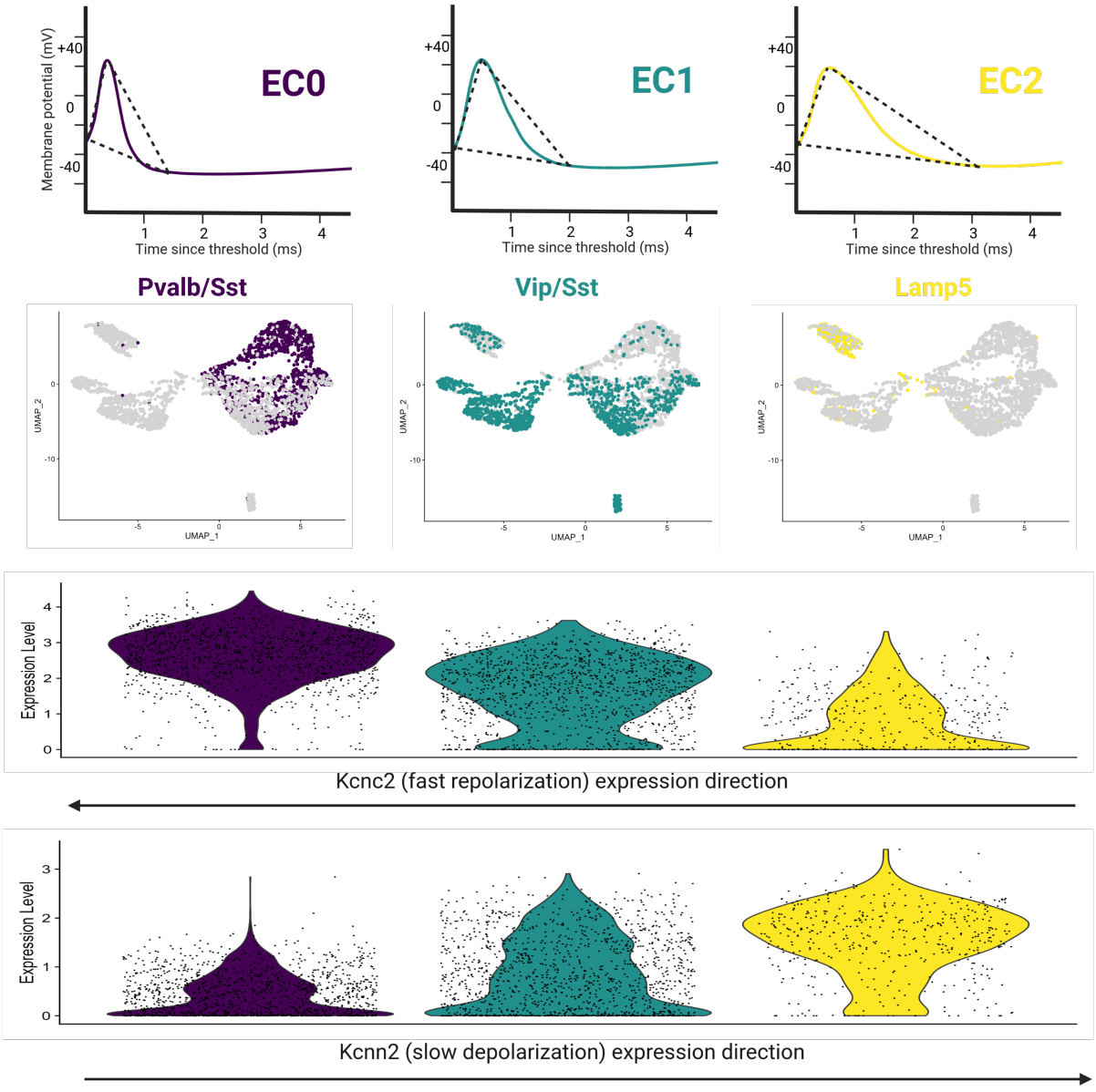

Figure 16: Relations between the median AP shapes with exemplary triangles (first row), the NSS clustering labels over the transcriptomic embedding, with an indication of the transcriptional GABAergic subtypes associated to each cluster (second row), and the expression levels of exemplary K^+-channels-related genes (*Kcnc2* and *Kcnn2*) across NSS clusters, represented through violin plots showing the gene expression levels distributions across cells in each cluster. The colors identify the NSS clusters for the $K = 3$ case, namely violet for *EC0*, green for *EC1*, and yellow for *EC2*.

EC1 and *EC2* clusters. Interestingly, the analysis reveals the presence of K^+ channel genes in both cases. Moreover, the expression patterns of these genes within each NSS cluster consistently align with the median AP shape of that cluster.

In summary, when the three clusters (*EC0*, *EC1*, and *EC2*) are ordered based on the shape of the AP waveform, ranging from steeper to slower repolarization phases, the expression of K^+ channel-related genes exhibits consistent trends from *EC0* to *EC1*, and finally to *EC2*. Figure 16 specifically highlights the expression levels of the *Kcnc2* gene, which is associated with "fast action potential repolarization" [60], and demonstrates a direct correlation with the steepness of AP repolarization. This finding strengthens the link between transcriptomic and electrophysiological biological levels. As discussed in Section 2.2.1, higher expression of K^+ channel genes may indicate a positive regulation of AP repolarization steepness, resulting in a narrower AP shape.

However, a few K^+ channel genes exhibit an opposite trend, namely *Kcnn2* and *Kcnq5*. These genes are more highly expressed in *EC0* than *EC1* and *EC0*, which contradicts the previous observations. Nevertheless, according to GeneCards [62], these genes contribute to the slow component of the synaptic after-hyperpolarization phase [61].

This phase primarily involves Ca^{++}-dependent and voltage-independent K^+ currents that can persist for several seconds following a series of APs. Hence, the inverse trend observed in the expression levels of *Kcnn2* does not contradict the direct correlation observed for *kcnc2*. Instead, this opposing trend could emphasize the increasing positive regulation of a slower and less steep repolarization phase from *EC2* to *EC1*, and finally to *EC0*, in complete alignment with the previous observation (Figure 16).

4.4. Validation on *PatchClampDataset*

Since *PatchClampDataset* consists solely of patch-clamp data, a more limited analysis is conducted using the same EP exploration approach described in Figure 3. The comparison between NSS-based clustering labels and cell subtype heterogeneity relies on the metadata from *PatchClampDataset*. These labels are based on the Cre lines of the transgenic mice used in the patch-clamp experiments, where conditional mutations enable the selective activation of fluorescent markers specific to certain cell types. Consequently, each patch-clamp experiment can be associated with a particular cell type. The dataset includes information on the specific transgenic mouse line for each cell, which allows for the identification of the corresponding cell type [22].

After the data cleaning process, the dataset comprises 770 valid GABAergic neurons. Clustering these cells based on NSS features results in two potential solutions: the optimal clustering is identified at $K = 2$ (Silhouette=0.40), but a very similar outcome is also achieved with $K = 3$ (Silhouette=0.38), which aligns with the observations made in Section 4.3 during the analysis of *PatchSeqDataset*.

Figure 17 presents the results of HAC clustering using scatter plots. In contrast to Section 4.3, which analyzes NSS labels over a transcriptomic embedding, due to the lack of single-cell transcriptomic data to replicate the same workflow, this Section analyzes both cell types and NSS labels over the EP embedding. The top scatter plot in Figure 17 displays the data color-coded according to the cell types derived from transgenic lines. In contrast, the bottom scatter plots show the data color-coded based on the EP clustering with different cardinalities. Once again, the NSS clustering effectively identifies the Pvalb group, whereas the Sst region contains cells from multiple clusters. In the case of $K = 3$, the tail of the Vip/Lamp region is further divided, resulting in an *EC1* cluster predominantly composed of Sst cells, while Lamp5 cells are likely assigned to cluster *EC2*. Quantitatively, *EC0* consists of 71% Pvalb, 10% Sst, and 3% Vip/Lamp5; *EC1* includes 5% Pvalb, 32% Sst, and 51% Vip/Lamp5; and *EC2* comprises 16% Pvalb, 3% Sst, and 67% Vip/Lamp5.

To conclude the analysis, Figure 18 displays the AP spikes of the median cells from the $K = 3$ clusters. Once again, the AP spikes from the median cells of the three clusters correspond to the three cell families (Pvalb, Sst, Vip/Lamp), and, similarly, the spike shapes observed in *PatchClampDataset* appear to be modeled by NSS features. Specifically, *EC0* is characterized by a rapid decline in the repolarization phase (the downslope), while *EC2* exhibits the opposite trend, with a large and slow spike response to stimulation. *EC1* displays an intermediate behavior. Moreover, higher variability is observed in the THR and P. However, no further conclusions can be drawn due to the absence of sequencing data.

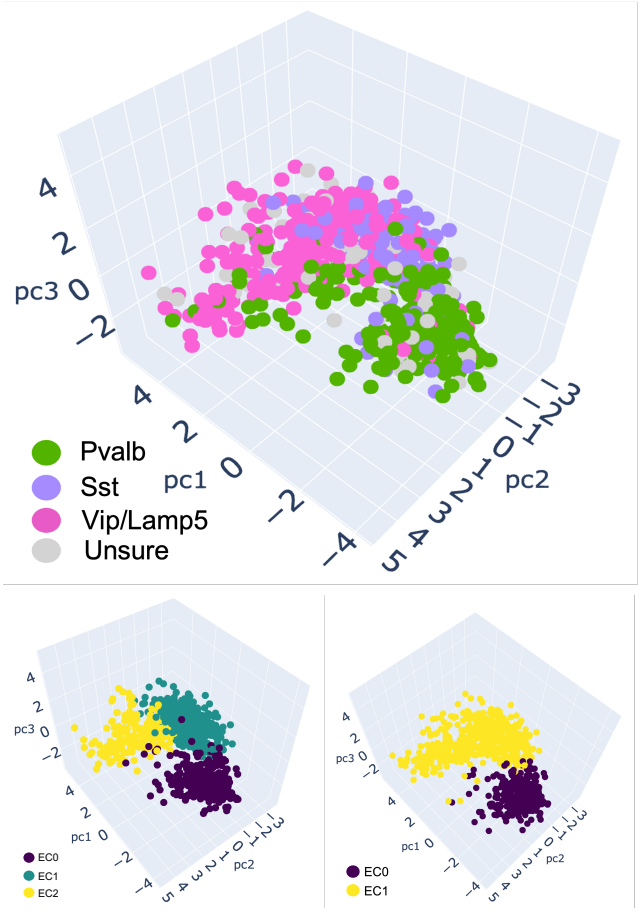

Figure 17: Scatter plots of *PatchClampDataset* NSS embedding with different labelings. On top, scatter with coloring concerning cell type based on Cre transgenic lines: Pvalb (green), Sst (violet), Vip/Lamp5 (pink); on the bottom-left, scatter with coloring with respect to the label of $K = 3$ NSS clustering: *EC0* (violet), *EC1* (green), *EC2* (yellow); on the bottom-right, scatter with coloring with respect to the label of $K = 2$ clustering: *EC0* (violet), *EC1* (yellow).

5. Discussion and conclusions

This study introduces the NSS approach, which offers a straightforward method to analyze AP waveforms for studying neuronal cell types and states. The paper presents the framework of NSS analysis, outlines its data requirements, and demonstrates its application using two datasets of murine neocortical neurons (*PatchClampDataset* and *PatchSeqDataset*, as described in [22] and [21]). The NSS-based EP feature analysis is performed to identify relevant

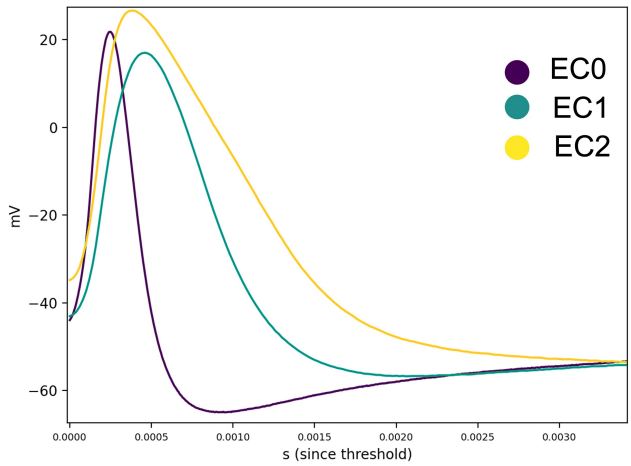

Figure 18: AP spikes of median cells identified using NSS on *PatchClampDataset* with coloring with respect to label of $K = 3$ NSS clustering: *EC0* (violet), *EC1* (green), *EC2* (yellow).

cell clusters, analyze the composition of the cellular population in relation to GABAergic cell subtypes, and explore the relationships between them. The results show that the EP partitions correspond to typical AP waveforms and show consistency with identified cell subtypes while highlighting cell-type independent groups at higher resolutions, as shown for Sst cells. This suggests that NSS approach may capture cell-type independent transient cell states. Interestingly, in *PatchSeqDataset*, the NSS-based partitions are associated with cellular processes related to K^+ channel metabolism and plasticity. This is consistent with K^+ channels play a crucial role in regulating AP waveforms, particularly in the repolarization and hyperpolarization phases, which explain most of the variability observed across the identified clusters. The EP analysis combined with cell line-based cell type labeling over *PatchClampDataset* provides additional validation of the NSS capability of capturing the heterogeneity of cells with respect to types and states that are consistent across the two datasets.

Future research aims to analyze more datasets containing different neuronal types and potentially from other species. This includes both patch-clamp and patch-seq datasets. NSS facilitates EP analysis and can be applied to single-cell patch-clamp datasets. Patch-seq datasets also enable multimodal analysis and biological validation of the NSS approach. The NSS-based analysis benefits from patch-seq datasets due to their inherent consistency between EP and transcriptomic cellular profiles. Unfortunately, patch-seq datasets currently face challenges in terms of data availability, and they often have low throughput due to technical limitations, resulting in a limited number of samples per dataset and hindering large-scale analysis [36, 41]. Nevertheless, recent advancements in semi-automatic patch-clamp setups allow for the generation and publication of large-scale patch-clamp datasets [42].

Many neuronal properties and behaviors lack a universally agreed-upon definition and comprehensive characterization in the existing literature. The NSS approach can contribute to the analytical study of these properties within patch-seq datasets, helping to establish consensus around essential and widely used concepts in neuronal physiology. An important future direction is to utilize the NSS approach to investigate the relationships between AP shape and crucial neuronal features such as *plasticity*, *intrinsic excitability* [16], and *firing rate set points* [65] in neuronal cells.

Future research will incorporate a more comprehensive characterization of gene expression profiles within NSS-based partitions regarding multimodal analysis. This will allow for the detection of complex expression patterns and expand the subsequent analysis to a broader range of genes. Such an approach will provide a more extensive exploration of the NSS biological significance, including considering alternative or integrative approaches to ontology-based analyses.

Code availability statement The code supporting this study is publicly available in GitHub: `https://github.com/smilies-polito/NSS`.

No competing interests statement The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Lorenzo Martini: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Gianluca Amprimo:** Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Stefano Di Carlo:** Conceptualization, Methodology, Resources, Writing - Review & Editing, Supervision, Funding acquisition. **Gabriella Olmo:** Conceptualization, Methodology, Resources, Writing - Review & Editing, Supervision, Funding acquisition. **Claudia Ferraris:** Conceptualization, Methodology, Resources, Writing - Review & Editing, Supervision, Funding acquisition. **Alessandro Savino:** Methodology, Resources, Supervision, Funding acquisition. **Roberta Bardini:** Conceptualization, Methodology, Writing - Original Draft, Writing - Review & Editing, Supervision, Project administration.

References

[1] Arlotta, P., Paşca, S.P., 2019. Cell diversity in the human cerebral cortex: from the embryo to brain organoids. Current opinion in neurobiology 56, 194–198.

[2] Atlas, A.B., 2023. Cell types database. URL: `https://celltypes.brain-map.org/`.

[3] Barnett, M.W., Larkman, P.M., 2007. The action potential. Practical neurology 7, 192–197.

[4] Bean, B.P., 2007. The action potential in mammalian central neurons. Nature Reviews Neuroscience 8, 451–465.

[5] Becker, W., Kleinsmith, L., Bertoni, G., et al., 2009. Signal transduction mechanisms: I. electrical and synaptic signaling in neurons, in: The world of the cell. Benjamin Cummings, p. .

[6] Biel, M., Wahl-Schott, C., Michalakis, S., Zong, X., 2009. Hyperpolarization-activated cation channels: from genes to function. Physiological reviews 89, 847–885.

[7] Breiman, L., 2001. Random forests. Machine learning 45, 5–32.

[8] Buettner, F., Natarajan, K.N., Casale, F.P., Proserpio, V., Scialdone, A., Theis, F.J., Teichmann, S.A., Marioni, J.C., Stegle, O., 2015. Computational analysis of cell-to-cell heterogeneity in single-cell rna-sequencing data reveals hidden subpopulations of cells. Nature biotechnology 33, 155–160.

[9] Cadwell, C.R., Palasantza, A., Jiang, X., Berens, P., Deng, Q., Yilmaz, M., Reimer, J., Shen, S., Bethge, M., Tolias, K.F., et al., 2016. Electrophysiological, transcriptomic and morphologic profiling of single neurons using patch-seq. Nature biotechnology 34, 199–203.

[10] Cadwell, C.R., Scala, F., Li, S., Livrizzi, G., Shen, S., Sandberg, R., Jiang, X., Tolias, A.S., 2017. Multimodal profiling of single-cell morphology, electrophysiology, and gene expression using patch-seq. Nature protocols 12, 2531–2553.

[11] Cardona-Alberich, A., Tourbez, M., Pearce, S.F., Sibley, C.R., 2021. Elucidating the cellular dynamics of the brain with single-cell rna sequencing. RNA biology 18, 1063–1084.

[12] Casale, A.E., Foust, A.J., Bal, T., McCormick, D.A., 2015. Cortical interneuron subtypes vary in their axonal action potential properties. Journal of Neuroscience 35, 15555–15567.

[13] Chamling, X., Kallman, A., Fang, W., Berlinicke, C.A., Mertz, J.L., Devkota, P., Pantoja, I.E.M., Smith, M.D., Ji, Z., Chang, C., et al., 2021. Single-cell transcriptomic reveals molecular diversity and developmental heterogeneity of human stem cell-derived oligodendrocyte lineage cells. Nature communications 12, 1–20.

[14] Consortium, T.U., 2022. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Research 51, D523–D531.

[15] Darmanis, S., Sloan, S.A., Zhang, Y., Enge, M., Caneda, C., Shuer, L.M., Hayden Gephart, M.G., Barres, B.A., Quake, S.R., 2015. A survey of human brain transcriptome diversity at the single cell level. Proceedings of the National Academy of Sciences 112, 7285–7290.

[16] Debanne, D., Inglebert, Y., Russier, M., 2019. Plasticity of intrinsic neuronal excitability. Current opinion in neurobiology 54, 73–82.

[17] Distasi, C., Dionisi, M., Ruffinatti, F.A., Gilardino, A., Bardini, R., Antoniotti, S., Catalano, F., Bassino, E., Munaron, L., Martra, G., et al., 2019. The interaction of sio2 nanoparticles with the neuronal cell membrane: activation of ionic channels and calcium influx. Nanomedicine 14, 575–594.

[18] Eze, U.C., Bhaduri, A., Haeussler, M., Nowakowski, T.J., Kriegstein, A.R., 2021. Single-cell atlas of early human brain development highlights heterogeneity of human neuroepithelial cells and early radial glia. Nature neuroscience 24, 584–594.

[19] Ge, S.X., Jung, D., Yao, R., 2020. Shinygo: a graphical gene-set enrichment tool for animals and plants. Bioinformatics 36, 2628–2629.

[20] Golomb, D., Donner, K., Shacham, L., Shlosberg, D., Amitai, Y., Hansel, D., 2007. Mechanisms of firing patterns in fast-spiking cortical interneurons. PLoS Computational Biology 3, e156.

[21] 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., 2020. Integrated morphoelectric and transcriptomic classification of cortical gabaergic cells. Cell 183, 935–953.

[22] Gouwens, N.W., Sorensen, S.A., Berg, J., Lee, C., Jarsky, T., Ting, J., Sunkin, S.M., Feng, D., Anastassiou, C.A., Barkan, E., et al., 2019. Classification of electrophysiological and morphological neuron types in the mouse visual cortex. Nature neuroscience 22, 1182–1195.

[23] Hao, Y., Hao, S., Andersen-Nissen, E., Mauck III, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al., 2021. Integrated analysis of multimodal single-cell data. Cell 184, 3573–3587.

[24] Herrera, A., Cheng, A., Mimitou, E.P., Seffens, A., George, D., Bar-Natan, M., Heguy, A., Ruggles, K.V., Scher, J.U., Hymes, K., et al., 2021. Multimodal single-cell analysis of cutaneous t-cell lymphoma reveals distinct subclonal tissue-dependent signatures. Blood, The Journal of the American Society of Hematology 138, 1456–1464.

[25] Hill, C.L., Stephens, G.J., 2021. An introduction to patch clamp recording. Patch Clamp Electrophysiology: Methods and Protocols , 1–19.

[26] Hodge, R.D., Bakken, T.E., Miller, J.A., Smith, K.A., Barkan, E.R., Graybuck, L.T., Close, J.L., Long, B., Johansen, N., Penn, O., et al., 2019. Conserved cell types with divergent features in human versus mouse cortex. Nature 573, 61–68.

[27] Hodgkin, A.L., Huxley, A.F., 1952. A quantitative description of membrane current and its application to conduction and excitation in nerve. The Journal of physiology 117, 500.

[28] Hwang, B., Lee, J.H., Bang, D., 2018. Single-cell rna sequencing technologies and bioinformatics pipelines. Experimental & molecular medicine 50, 1–14.

[29] Kim, H., Lee, J., Kang, K., Yoon, S., 2022. Markercount: A stable, count-based cell type identifier for single-cell rna-seq experiments. Computational and Structural Biotechnology Journal 20, 3120–3132.

[30] Kolb, B., Whishaw, I.Q., 1998. Brain plasticity and behavior. Annual review of psychology 49, 43–64.

[31] Kole, M.H., Letzkus, J.J., Stuart, G.J., 2007. Axon initial segment kv1 channels control axonal action potential waveform and synaptic efficacy. Neuron 55, 633–647.

[32] La Manno, G., Gyllborg, D., Codeluppi, S., Nishimura, K., Salto, C., Zeisel, A., Borm, L.E., Stott, S.R., Toledo, E.M., Villaescusa, J.C., et al., 2016. Molecular diversity of midbrain development in mouse, human, and stem cells. Cell 167, 566–580.

[33] Lake, B.B., Ai, R., Kaeser, G.E., Salathia, N.S., Yung, Y.C., Liu, R., Wildberg, A., Gao, D., Fung, H.L., Chen, S., et al., 2016. Neuronal subtypes and diversity revealed by single-nucleus rna sequencing of the human brain. Science 352, 1586–1590.

[34] Li, T., Tian, C., Scalmani, P., Frassoni, C., Mantegazza, M., Wang, Y., Yang, M., Wu, S., Shu, Y., 2014. Action potential initiation in neocortical inhibitory interneurons. PLoS biology 12, e1001944.

[35] Li, W., Luo, X., Ulbricht, Y., Wagner, M., Piorkowski, C., El-Armouche, A., Guan, K., 2019. Establishment of an automated patch-clamp platform for electrophysiological and pharmacological evaluation of hipsc-cms. Stem cell research 41, 101662.

[36] Lipovsek, M., Bardy, C., Cadwell, C.R., Hadley, K., Kobak, D., Tripathy, S.J., 2021. Patch-seq: Past, present, and future. Journal of Neuroscience 41, 937–946.

[37] Liu, H., Zhou, J., Tian, W., Luo, C., Bartlett, A., Aldridge, A., Lucero, J., Osteen, J.K., Nery, J.R., Chen, H., et al., 2021. Dna methylation atlas of the mouse brain at single-cell resolution. Nature 598, 120–128.

[38] Luecken, M.D., Theis, F.J., 2019. Current best practices in single-cell rna-seq analysis: a tutorial. Molecular systems biology 15, e8746.

[39] Martini, L., Bardini, R., Savino, A., Di Carlo, S., 2022a. Gagam: a genomic annotation-based enrichment of scatac-seq data for gene activity matrix. bioRxiv .

[40] Martini, L., Bardini, R., Savino, A., Di Carlo, S., 2022b. Gagam v1. 2: An improvement on peak labeling and genomic annotated gene activity matrix construction. Genes 14, 115.

[41] Martini, L., Bardini, R., Savino, A., Di Carlo, S., 2022c. High-resolution sample size enrichment of single-cell multi-modal low-throughput patch-seq datasets, in: 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE. pp. 2334–2341.

[42] Marx, V., 2022. Patch-seq takes neuroscience to a multimodal place. Nature Methods 19, 1340–1344.

[43] McFarlan, A.R., Chou, C.Y., Watanabe, A., Cherepacha, N., Haddad, M., Owens, H., Sjöström, P.J., 2022. The plasticitome of cortical interneurons. Nature Reviews Neuroscience , 1–18.

[44] Menon, S., Lui, V.C.H., Tam, P.K.H., 2021. Bioinformatics tools and methods to analyze single cell rna sequencing data. International

Journal of Innovative Science and Research Technology,(IJISRT) 6, 282–288.

[45] Mi, H., Muruganujan, A., other, 2018. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Research 47, D419–D426. doi:10.1093/nar/gky1038.

[46] Morris, S.A., 2019. The evolving concept of cell identity in the single cell era. Development 146, dev169748.

[47] Mu, Q., Chen, Y., Wang, J., 2019. Deciphering brain complexity using single-cell sequencing. Genomics, Proteomics & Bioinformatics 17, 344–366.

[48] Mulas, C., Chaigne, A., Smith, A., Chalut, K.J., 2021. Cell state transitions: definitions and challenges. Development 148, dev199950.

[49] Nahar, L., Delacroix, B.M., other, 2021. The role of parvalbumin interneurons in neurotransmitter balance and neurological disease. Front Psychiatry 12. doi:10.3389/fpsyt.2021.679960.

[50] Neher, E., Sakmann, B., 1976. Single-channel currents recorded from membrane of denervated frog muscle fibres. Nature 260, 799–802.

[51] Neher, E., Sakmann, B., 1992. The patch clamp technique. Scientific American 266, 44–51.

[52] Obergrussberger, A., Friis, S., Brüggemann, A., Fertig, N., 2021. Automated patch clamp in drug discovery: major breakthroughs and innovation in the last decade. Expert opinion on drug discovery 16, 1–5.

[53] Papalexi, E., Satija, R., 2018. Single-cell rna sequencing to explore immune cell heterogeneity. Nature Reviews Immunology 18, 35–45.

[54] Platkiewicz, J., Brette, R., 2010. A threshold equation for action potential initiation. PLoS computational biology 6, e1000850.

[55] Poirion, O.B., Zhu, X., Ching, T., Garmire, L., 2016. Single-cell transcriptomics bioinformatics and computational challenges. Frontiers in genetics 7, 163.

[56] Rodríguez-Collado, A., Rueda, C., 2021. Electrophysiological and transcriptomic features reveal a circular taxonomy of cortical neurons. Frontiers in Human Neuroscience 15, 684950.

[57] Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics 20, 53–65.

[58] Seal, R.L., et al., 2022. Genenames.org: the HGNC resources in 2023. Nucleic Acids Research 51, D1003–D1009. doi:10.1093/nar/gkac888.

[59] Spindler, A., Noble, S., Noble, D., 1999. Comparison of step and ramp voltage clamp on background currents in guinea-pig ventricular cells. Experimental physiology 84, 865–879.

[60] Stelzer, G., Rosen, N., Plaschkes, I., et al., a. Kcnn2 genecards. URL: https://www.genecards.org/cgi-bin/carddisp.pl?gene=KCNC2.

[61] Stelzer, G., Rosen, N., Plaschkes, I., et al., b. Kcnn2 genecards. URL: https://www.genecards.org/cgi-bin/carddisp.pl?gene=KCNN2.

[62] Stelzer, G., Rosen, N., Plaschkes, I., et al., 2016. The genecards suite: From gene data mining to disease genome sequence analyses. Current Protocols in Bioinformatics 54, 1.30.1–1.30.33. URL: https://currentprotocols.onlinelibrary.wiley.com/doi/abs/10.1002/cpbi.5, doi:https://doi.org/10.1002/cpbi.5.

[63] Sugino, K., Clark, E., Schulmann, A., Shima, Y., Wang, L., Hunt, D.L., Hooks, B.M., Tränkner, D., Chandrashekar, J., Picard, S., et al., 2019. Mapping the transcriptional diversity of genetically and anatomically defined cell populations in the mouse brain. Elife 8, e38619.

[64] Tepe, B., Hill, M.C., Pekarek, B.T., Hunt, P.J., Martin, T.J., Martin, J.F., Arenkiel, B.R., 2018. Single-cell rna-seq of mouse olfactory bulb reveals cellular heterogeneity and activity-dependent molecular census of adult-born neurons. Cell reports 25, 2689–2703.

[65] Trojanowski, N.F., Bottorff, J., Turrigiano, G.G., 2021. Activity labeling in vivo using campari2 reveals intrinsic and synaptic differences between neurons with high and low firing rate set points. Neuron 109, 663–676.

[66] Turrigiano, G.G., Nelson, S.B., 2000. Hebb and homeostasis in neuronal plasticity. Current opinion in neurobiology 10, 358–364.

[67] Van Den Hurk, M., Bardy, C., 2019. Single-cell multimodal transcriptomics to study neuronal diversity in human stem cell-derived brain tissue and organoid models. Journal of Neuroscience Methods 325, 108350.

[68] Wang, M., Song, W.m., Ming, C., Wang, Q., Zhou, X., Xu, P., Krek, A., Yoon, Y., Ho, L., Orr, M.E., et al., 2022. Guidelines for bioinformatics of single-cell sequencing data analysis in alzheimer's disease: review, recommendation, implementation and application. Molecular Neurodegeneration 17, 1–52.

[69] Welch, J.D., Kozareva, V., Ferreira, A., Vanderburg, C., Martin, C., Macosko, E.Z., 2019. Single-cell multi-omic integration compares and contrasts features of brain cell identity. Cell 177, 1873–1887.

[70] Wu, Y., Zhang, K., 2020. Tools for the analysis of high-dimensional single-cell rna sequencing data. Nature Reviews Nephrology 16, 408–421.

[71] Yang, J., Ye, M., Tian, C., Yang, M., Wang, Y., Shu, Y., 2013. Dopaminergic modulation of axonal potassium channels and action potential waveform in pyramidal neurons of prefrontal cortex. The Journal of physiology 591, 3233–3251.

[72] Yelhekar, T.D., Druzin, M., Karlsson, U., Blomqvist, E., Johansson, S., 2016. How to properly measure a current-voltage relation?—interpolation vs. ramp methods applied to studies of gabaa receptors. Frontiers in cellular neuroscience 10, 10.

[73] Zeng, H., 2022. What is a cell type and how to define it? Cell 185, 2739–2755.