

University code: 10246
Student ID: 16211520021



Master's thesis
(Academic degree)

**Classification of Electrophysiological Types of Cortical
Interneurons Using Ward's Hierarchical Clustering and
Iterative PCA Clustering**

Zhou Ying

Master of Neurobiology
Institutes of Brain Science
Supervisor: Prof. Yu Yongchun
27Th Sep. 2019

Contents

List of Figures	iv
List of Tables	v
Abstract	vi
Chapter 1 Review of GABAergic Interneuron	
Classification	1
1.1 Introduction	1
1.2 GABA interneurons are diverse	2
1.2.1 Electrophysiological characteristics	3
1.2.2 Morphological characteristics	4
1.2.3 Molecular characteristics	5
1.3 Purpose of interneuron classification	6
1.3.1 Reproducibility of Research	6
1.3.2 Discovery of New Types and Alternative Marker Genes	7
1.3.3 Understanding Development and Evolutionary Processes	7
1.3.4 Induction of Units Constituting the Nervous System (Dimension Reduction)	7
1.3.5 Disease Research	7
1.4 Limitations of qualitative classification	8
1.5 Interneuron classification requires quantitative classification	9
1.6 Quantitative classification	10
1.6.1 Supervised classification and unsupervised classification	10
1.6.2 Electrophysiological Quantitative Classification	10
1.6.3 Morphological Quantitative Classification	11
1.7 Hierarchical classification system	12
1.8 Further demand for interneuron classification in both dry and wet experiments	13

Chapter 2 Introduction	15
2.1 Electrophysiological classification	16
2.2 Clustering analysis	17
2.3 Principle Component Analysis	18
2.4 Iterative principal component analysis	18
2.5 Significance of this study	19
 Chapter 3 Materials and Methods	 20
3.1 environment, software and modules	20
3.2 Data resources	20
3.2.1 Datasets selection	20
3.2.2 Data download	21
3.3 Data format conversion	22
3.3.1 Convert <i>NWB</i> format to <i>txt</i> format	22
3.3.2 <i>data</i> format to <i>txt</i> format	22
3.4 Extraction of electrophysiological parameters	23
3.4.1 <i>eFEL</i> toolkit installation and invocation	23
3.4.2 Electrophysiological Parameter Extraction Algorithm	24
3.4.3 Statistical integration and construction of analysis matrix for electrophysiological parameters	25
3.5 Selection of parameter	26
3.6 Ward's hierarchical clustering	27
3.6.1 Algorithm Principles	27
3.6.2 Performing in R	28
3.6.3 Deciding on cut tree threshold	28
3.7 iPCA	29
3.7.1 PCA Analysis in R language	29
3.7.2 Determine whether the classified subsets exhibit significant clustering	29
3.7.3 Integrating PCA and Ward's hierarchical clustering to build iPCA	30
3.8 Code accessibility	30
3.9 Statistics	30
 Chapter 4 Results	 32
4.1 Quantification of electrophysiological characteristics	32
4.2 Further parameter filtering	35

4.2.1	PCA analysis	35
4.2.2	Parameter Correlation Analysis	36
4.3	Ward's hierarchical clustering	37
4.4	iPCA	40
4.4.1	iPCA can visualize the intermediate classification process . . .	40
4.4.2	iPCA can adjust parameter weights based on different subclasses	40
4.4.3	iPCA clustering 8 electrophysiological categories	40
4.4.4	iPCA is able to flexibly integrate classification information from other aspects	44
4.5	Comparing Ward's hierarchical clustering with iPCA analysis results	44
Chapter 5	Disscusion	49
5.1	Quantitative parameters	49
5.2	Hierarchical Unsupervised Classification	50
5.3	Supervised Classification	51
5.4	Electrophysiological type and molecular type	51
Bibliography		53
Acknowledgements		61

List of Figures

1-1	Classification of cortical GABAergic interneurons	2
1-2	Petilla's nomenclature of Various Discharge pattern	3
1-3	Diversity of Cortical Neuron Morphology	4
1-4	Molecular diversity of cortical neurons	6
1-5	Neuronal multidimensional characteristic data	9
1-6	Classification Methods and Their Data Structures	10
1-7	Analytical Process of Morphology	11
1-8	Quantitative parameters of morphology	12
2-1	Types of Interneuron Electrophysiology	16
2-2	iPCA Process	19
3-1	Hierarchical Clustering Process Diagram	27
4-1	Schematic Diagram of Main Electrophysiological Parameters	32
4-2	Kernel density distribution map of electrophysiological characteristic parameters	34
4-3	PCA Analysis Results	36
4-4	Hot map displays the correlation between various parameters	37
4-5	Situation of Ward hierarchical clustering	38
4-6	Classification results of Ward hierarchical clustering for each cell	38
4-7	Classification of Ward hierarchical clustering yielded 7 types of electrophysiological discharge patterns	39
4-8	Visualization of the iPCA analysis process	41
4-9	The different weights of parameters on PC1/PC2 in each iteration	42
4-10	iPCA is divided into representative figures of various electrophysiological types	43
4-11	Molecular type and layer distribution binary tree	45
4-12	Comparison of iPCA Method and Ward's Method	48
5-1	Diagram illustrating hierarchical relationship of classification results . . .	52

List of Tables

3-1	Software and tools	20
3-2	Dataset composition	21
4-1	Brief description of each feature.	33
4-2	Summary table of ANOVA test results among 7 molecular types	35
4-3	All parameters of clusters in iPCA.	46
4-4	All parameters Tukey test between iPCA clusters.	47
4-5	All parameters ANOVA test between iPCA cluster 2,3,4,5.	47

Abstract

Cortical GABAergic interneurons exhibit high diversity in electrophysiological properties, which allows them to dynamically sculpt neuronal activity and endow neural circuits with remarkable computational power. Although different strategies have been employed to characterize this diversity, it still remains largely unclear how many electrophysiological types (e-type) of cortical interneuron actually exist. Here, we quantitatively analysis electrophysiological properties of 569 interneurons in mouse visual cortex. All cells are fluorescent protein-positive and originate from 7 transgenic mouse lines. For each cell, 30 electrophysiological parameters are automatically extracted from responses to current stimuli, and systematically analyzed their distribution in 7 transcriptomic cell types and contribution in e-types clustering. To identify e-types of cortical interneurons, we used and compared two different unsupervised clustering algorithms. We found that Ward's method and Iteration Principal Component Analysis (iPCA) method revealed respectively 7 and 8 clusters within the data. Although two clustering methods revealed roughly similar clusters, distinct clusters 2, 3, 4 obtained from iPCA method are merged into one cluster (Ward's method cluster 2) in Ward's method. We found that iPCA is a better analytical framework to the big and complex interneuron dataset. The diversity of interneurons in cerebral cortex is an important issue in the research of neuroscience. With the further study of interneurons, a large number of interneuron data are generated in different laboratories. How to unify and organize these data is very important. It can be optimized furtherly in future to perform in multilevel classify interneuron study, as a first step to elucidate their function within the cortical circuit. It lays a foundation for data integration and comparison between different laboratories in the future.

Keywords: Interneuron; electrophysiological feature; clustering; cortex

CLC code: Q42

Chapter 1

Review of GABAergic Interneuron Classification

Abstract: The dynamic microcircuits composed of excitatory neurons with glutamatergic capabilities and inhibitory neurons with GABAergic capabilities endow the nervous system with various powerful functions. The dynamic balance of these microcircuits relies on the regulation of various types of interneurons. Although interneurons in the cerebral cortex account for only 20%-30% of cortical neurons, their high diversity plays a crucial role in regulating neural activity, neural circuit plasticity, and neural network synchronization. The complex diversity of interneurons is currently a highly researched topic. Early scientists primarily described and classified them using qualitative methods. Nowadays, with the development of high-throughput techniques, the demands of research deepening, the accumulation of large amounts of data, and the advancement of big data analysis, quantitative classification studies are increasingly used for interneuron classification research. This review mainly discusses the current status of interneuron classification research, the development from qualitative to quantitative classification, and the issues and demands of quantitative research on interneurons.

Keywords: Interneuron, Quantitative classification of interneurons, Electrophysiological quantification, morphological quantification

1.1 Introduction

The powerful computational capability of the nervous system mainly relies on the highly connected neurons and the dynamic microcircuits composed of neurons. The neurons constituting the microcircuits are mainly of two types: Glutamatergic excitatory neurons and GABAergic inhibitory neurons, also known as interneurons. The dynamic balance between excitation and inhibition in the cerebral cortex relies on the regulation of various types of interneurons. Interneurons in the cortex exhibit high diversity in electrophysiological, morphological, molecular, input-output connectivity patterns, and other aspects^[1-9]. Due to their high diversity, although they represent only 20%-30% of cortical neurons, they play a crucial role in regulating neuronal activity, neural circuit plasticity, and synchronization of neural networks^[1-3,5-7]. The dysregulation of interneurons can lead to various neurological disorders such as epilepsy, schizophrenia, anxiety disorders, and autism^[10-13]. Therefore, studying interneurons is crucial for understanding neural network regulation, nervous system function, and neurological disorders.

However, the diversity of interneurons and their low proportion in the entire nervous system pose the biggest challenge in studying them. Currently, with the development of molecular genetic manipulation techniques and high-throughput technologies such as high-throughput sequencing, it is possible to label and observe different types

of interneurons on a large scale and manipulate their activities to study their circuit functions^[3,14–15]. To some extent, the low proportion no longer hinders the study of interneurons, and their diversity becomes the main issue. Although in recent years, through high-throughput experimental techniques and the collaborative efforts of many laboratories, we have gained some understanding of interneuron diversity, organizations like The Petilla Interneuron Nomenclature Group (PING) have also attempted to standardize the Nomenclature and description of interneurons^[8–9]. But we still do not know exactly how many types of interneurons there are, the similarities and differences between different types, and the distribution and organizational structure of different types of interneurons^[1–3].

Qualitative descriptions in early studies of the nervous system were sufficient, but with the development of high-throughput technologies, the generation of large amounts of high-dimensional data, the increasing of research refinement and the known complexity of neurons, qualitative approaches are no longer sufficient for current research on the nervous system^[3,8–9,15–16]. Due to the complex diversity of interneurons, there is an increasing demand for quantitative classification^[9,16]. With the development of fields such as machine learning and data mining, the emergence of more robust algorithms enables the objective and detailed classification of neuron types using quantitatively described neuronal characteristics^[16–21]. This review mainly discusses the current research status of interneuron classification, the transition from qualitative to quantitative classification, and the issues and requirements of quantitative research on interneurons.

1.2 GABA interneurons are diverse

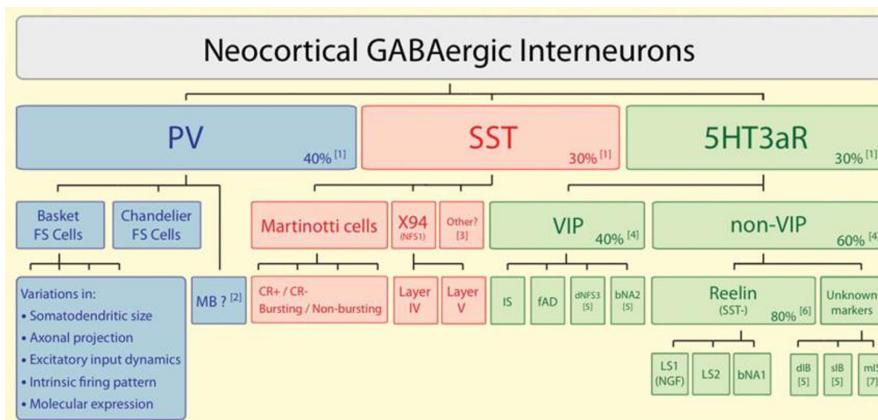


Figure 1-1 **Classification of cortical GABAergic interneurons**^[5]. Close to 100% of cortical GABAergic interneurons belong to three major classes: PV (parvalbumin), SST (somatostatin), and 5HT3aR (the ionotropic serotonin receptor 5HT3a). Different types of interneurons exhibit distinct morphological, electrophysiological, and molecular characteristics.

With the advancement of molecular genetic technology, many transgenic mice have been generated under the control of different molecular marker promoters, connecting fluorescent protein genes or Cre or Flp recombinase protein genes, allowing to label different types of interneurons^[14]. Through these transgenic mice, analysis of specific types of interneurons including morphology, electrophysiology, and circuit connections revealed that nearly 100% of cortical GABAergic interneurons belong to three major classes: parvalbumin (PV), somatostatin (SST), and the ionotropic serotonin receptor 5HT3a (5HT3aR)^[2,5]. Recent research on these three major classes of interneurons

has revealed distinct morphological, electrophysiological, and molecular characteristics among them (Figure 1-1)^[3,15,22–24]. These studies further demonstrate that classifying PV, SST, and 5HT3aR into the three major categories is a reliable first-level classification.

1.2.1 Electrophysiological characteristics

Initially, all interneurons were described as fast spiking (FS)^[8,25–27], and subsequent recording and analysis of discharge patterns revealed another discharge pattern known as low-threshold-spiking (LTS), also referred to as burst-spiking non-pyramidal (BSNP)^[28–29]. Subsequently, a discharge type with a significant delay after stimulation was discovered and considered as Late-spiking (LS) type^[25,30], and discharge patterns with irregular amplitudes and irregular intervals between action potentials were termed Irregular-spiking (IS)^[31–32].

Early descriptions and nomenclature of these electrophysiological discharge patterns were quite ambiguous, with different laboratories having different descriptions and names. It wasn't until the Petilla Naming Committee's analysis in 2008 that a more systematic naming of interneuron characteristics was carried out^[9]. Electrophysiology was categorized and described based on transient-state patterns (before reaching maximum discharge frequency) and steady-state patterns (after reaching maximum discharge frequency), classified and described electrophysiology based on sweep characteristics in these two states (Figure 1-2).

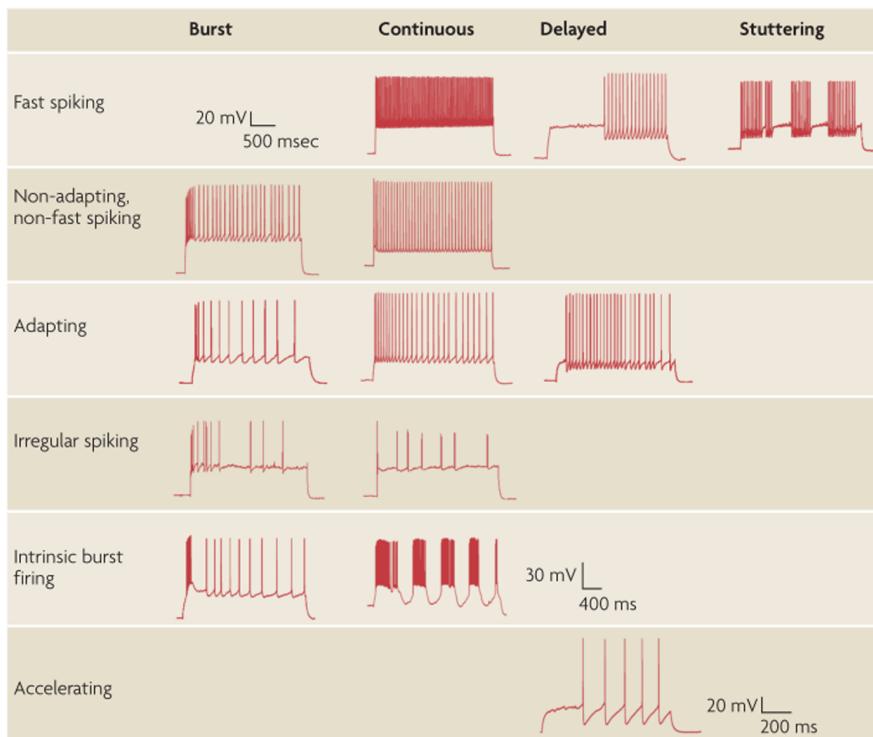


Figure 1-2 Petilla's nomenclature of Various Discharge pattern^[9].

In Figure 1-2, each column corresponds to the discharge mode of sweep under instantaneous state, represented by lowercase letters. The main characteristics of Burst (b) are that at the beginning of stimulation, there are several action potentials with smaller intervals than under stable conditions; Delay (d), also known as LS mentioned earlier,

has a large time delay at the beginning of stimulation before action potentials occur; Stuttering (s) is mainly characterized by longer periods of silence between a series of short intervals of discharge; Continuous (c) is defined as a discharge type other than the first three types under instantaneous state.

Each row in Figure 1-2 corresponds to the discharge mode of sweep under stable state, represented by uppercase letters. Fast Spiking (FS) is similar to the earliest recognized FS discharge type, with a higher maximum discharge frequency (generally considered to be greater than 50Hz); Adapting (AD) is characterized by increasing interval times between action potentials as stimulation time lengthens; Irregular Spiking (IS) is consistent with previous IS type characteristics, having irregular intervals and action potential amplitudes; non-adapting, non-fast spiking (NA) is a type that lacks adapting features and is not FS.

Describing electrophysiological characteristics through the combination of discharge modes under two states, such as the type in the first row and second column being referred to as c-FS. According to the unified nomenclature description of neuronal electrophysiological characteristics in 2008, 13 types of intermediate neuron electrophysiological characteristics were classified, and subsequently, new types have been described, indicating the diversity of intermediate neurons in electrophysiology^[9].

1.2.2 Morphological characteristics

The interneurons not only exhibit diverse electrophysiological characteristics but also possess various morphologies, with the main morphologies described in Figure 1-3 of current research.

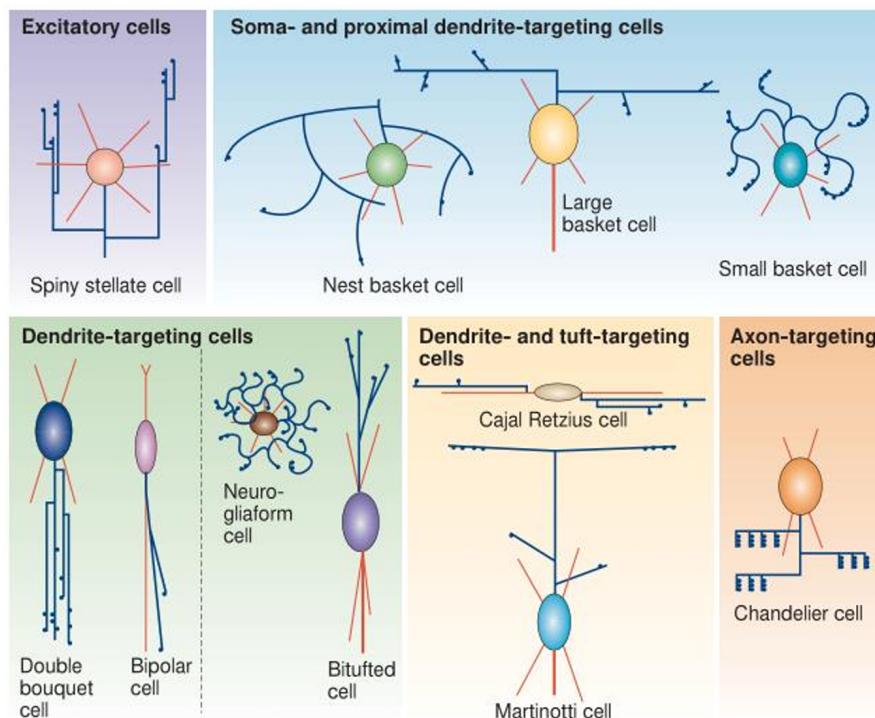


Figure 1-3 Diversity of Cortical Neuron Morphology.^[8]

Basket cells. Approximately 50% of interneurons are Basket cells, with axons forming synaptic connections predominantly with proximal dendrites and soma. The name

”Basket” derives from the basket-like appearance of these cells, which surround pyramidal neurons, with multiple Basket cells forming convergent neural dominance. Based on the differing morphologies of their axons and dendrites, this type can be further subdivided into three subtypes: large basket cells (LBCs), small basket cells (SBCs), and nest basket cells (NBCs)^[33–34].

- Large basket cells (LBCs): LBCs are the classic Basket cells, characterized by thick, aspiny, multi-level dendrites and extensively branching axons, capable of inhibiting excitations from above and within a certain range of functional columns^[33,35].
- Small basket cells (SBCs): SBCs are aspiny interneurons that form synaptic connections with nearby dendrites and soma. They have densely branched axons with short-range projections and typically do not project across layers or functional columns^[33,36].
- Nest basket cells (NBCs): NBCs were previously described as irregularly branching interneurons^[37–39]. They are now considered a type projecting onto soma, intermediate between LBCs and SBCs, resembling a bird’s nest in appearance. They have clustered axons around them like SBCs, as well as longer, less branched axons with low-density boutons, which resemble LBCs^[33–34].

Chandelier cells (ChCs). ChCs are interneurons forming synaptic connections with axons. This characteristic allows ChCs to rewrite almost all complex dendritic integration and soma gain by regulating axonal signal output^[40–42]. ChCs can be multi-level or bipolar, with axons typically clustering around forming multiple branches and high-density boutons distribution. Their most distinctive feature is the vertical arrangement of boutons at the axon terminals, like chandeliers^[43–44].

Martinotti cells (MCs). MCs are mostly distributed in L2-4, projecting their axons to inhibit pyramidal neuron dendritic clusters in L1 and extending horizontally for several millimeters along L1 to inhibit neighboring dendritic clusters of functional columns^[45]. However, they not only inhibit distant dendrites but also inhibit nearby dendrites, which around the soma, and the soma itself^[45].

Bipolar cells (BPCs). BPCs have small, oval or spindle-shaped somas, with bipolar dendrites extending vertically up to L1 and downward to L5^[46–47].

Double bouquet cells (DBCs). DBCs have dendrites in a double bouquet formation, forming tightly bundled circular formations resembling a ”ponytail”^[48–49].

Bitufted cells (BTCs). BTCs like BPCs and DBCs, have oval-shaped somas. They branch in opposite directions from the main dendrite, with axons differing from BPCs and DBCs in horizontally extending widely to inhibit across functional columns^[8,50].

Neurogliaform cells (NGCs). NGCs have small soma, with many short, aspiny, bead-like, sparsely branched radiating dendrites forming symmetrical spherical dendritic domains. Axons may originate from anywhere on the soma or dendritic base, but after a short distance, they branch extensively, forming a highly intertwined structure^[8].

In addition to the typical types mentioned above, interneurons have extremely complex and intricate structures that are currently very difficult to describe accurately, and the description of their morphology is still rough and vague.

1.2.3 Molecular characteristics

The molecular differences of interneurons predominantly in the form of ion channels, ionotropic, and metabotropic receptors on the cell membrane. Within the cytoplasm,

differences have been observed in various calcium-binding proteins^[2–5]. Additionally, interneurons are primarily regulated by GABAergic neuromodulatory peptides such as Somatostatin (SST), Vasoactive Intestinal Polypeptide (VIP), Cholecystokinin (CCK), Neuropeptide Y (NPY), etc^[2–5,8–9,22]. Several molecular markers within one interneuron lead to diverse molecular composition^[2–5,22].

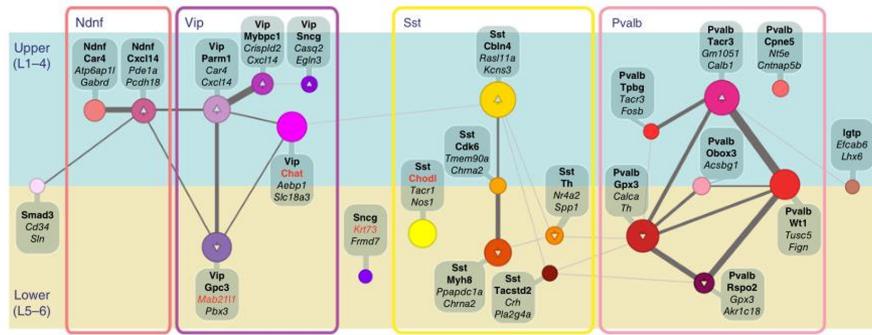


Figure 1-4 Molecular diversity of cortical neurons.^[22]

In recent years, the rapid development of single-cell sequencing technology has greatly propelled the research on molecular diversity of interneurons. Molecular characteristics of interneurons from different brain regions, species, and developmental time points are being investigated successively, revealing the diversity of interneuron molecules^[3,15,22–23,51] (Figure 1-4). While a large amount of molecular data has been generated, the relationships between various markers, their functions, and the number of types remain unclear. As molecular expression is a continuous process and can be influenced by factors such as development and environment, it is necessary to repeatedly compare and verify the differences between the transient and the stable states among interneurons. Stable differences are used as markers for classifying molecular types and continuously adding correlation information with phenotypes^[22].

1.3 Purpose of interneuron classification

The purpose of classification should not be merely for the sake of categorization itself, but rather to facilitate our understanding of the mechanisms of brain function and the causes of diseases through classification. Only by clearly identifying our purposes for classification can we better formulate classification schemes.

1.3.1 Reproducibility of Research

Ensuring the reproducibility of research is the most important goal of interneurons classification. The lack of uniform description and terms of interneurons across different laboratories has led to significant additional workload. As Crick discussed the application of molecular biology in neuroscience, the complexity of neuronal types and the different descriptions by different experimenters have turned research into a collection or philately rather than a scientific endeavor^[52]. This necessitates a comprehensive classification strategy to address such controversies.

1.3.2 Discovery of New Types and Alternative Marker Genes

In the process of classifying known types, comparing certain features and defining standards will lead to the direct discovery of new types. Additionally, during this process, genes related to phenotypes or cell types can be discovered as alternative marker genes, which can be used to label specific types of interneurons or study specific phenotypes and functions through genetic manipulation^[3].

1.3.3 Understanding Development and Evolutionary Processes

In developmental studies, describing neuronal activities and functions has been challenging. With the development of modern molecular techniques, the use of molecular markers for immunohistochemistry and molecular genetic manipulation has had a significant impact on the field of development. Molecular type classification, along with other classification information, can lead to the identification of distinctive molecular markers for studying developmental activities^[3].

Meanwhile, the diversity of interneurons greatly contributes to the advanced functions of neural networks, with different species possessing different types of interneurons, imparting various functions to the nervous systems of different organisms. Cell types are defined from an evolutionary perspective as a group of cells that have co-evolved, are partially distinct from other cells, and are evolutionarily closer to each other within the group but farther from other cells^[53]. Comparing cell types between species can infer evolutionary relationships between species.

1.3.4 Induction of Units Constituting the Nervous System (Dimension Reduction)

The nervous system is a highly complex system, with neurons as the primary units that generate biological functions through different network structures. Due to the diversity of neurons, especially interneurons, understanding the nervous system as a single component is difficult. Considering neurons as types can reduce this complexity and serve as a strategy to reduce dimensions^[3].

1.3.5 Disease Research

Abnormalities in interneurons are associated with various neurological disorders. Defining the characteristics of different types of interneurons is necessary for disease classification, and observing specific abnormal phenotypes in disease samples can contribute to a deeper understanding of disease mechanisms.

In summary, the purpose of classification is clearly significant for evolution, development, disease, and other related direction, particularly, discovery of molecular markers. Markers are not just gene symbols but correlations with phenotypes (morphology, electrophysiology, circuitry, etc.), which contain a wealth of information. Therefore, during the classification process, particular attention should be paid to the correlation of various classification information.

1.4 Limitations of qualitative classification

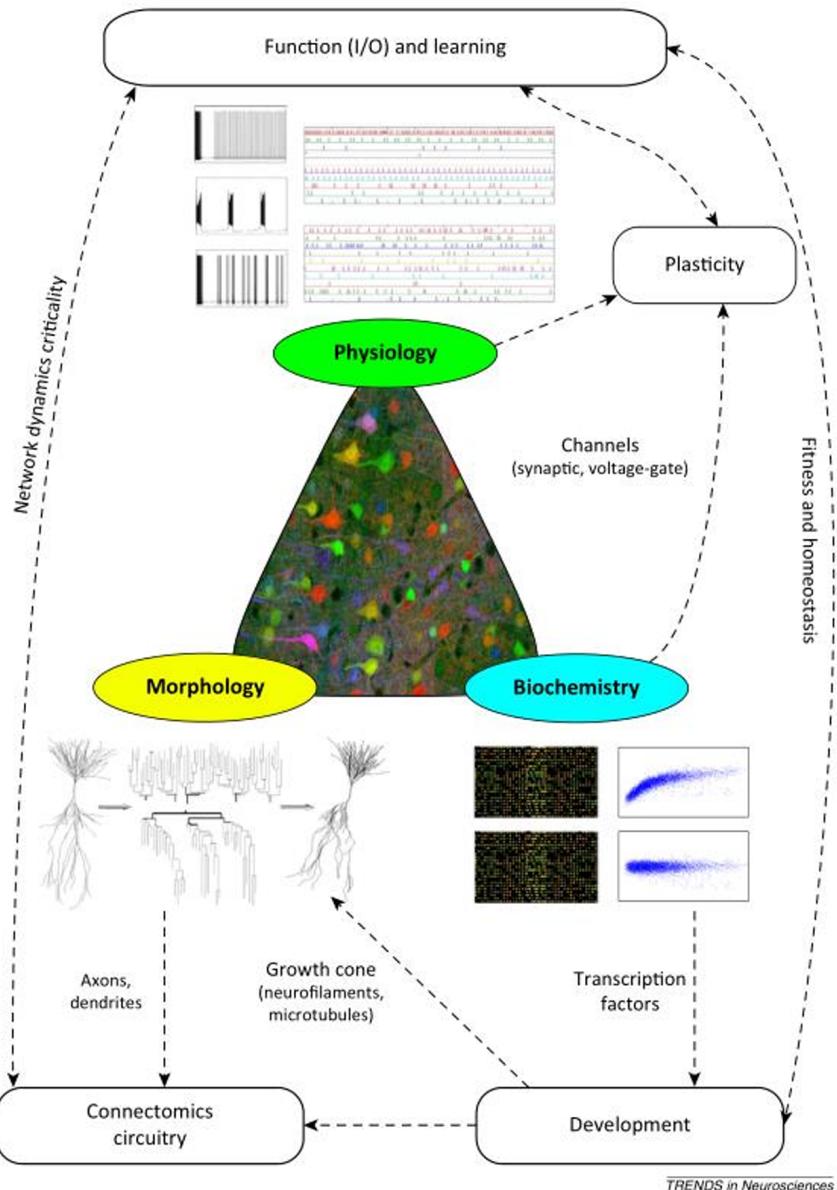
Currently, electrophysiology classification primarily emphasizes voltage changes in action potential waveforms induced by specific current stimuli^[8–9,54–56]. However, while qualitative descriptions can differentiate types during stable maximum discharge frequency with regular patterns, transient or irregular discharges pose challenges due to inherent randomness^[9,31,54,57]. Distinguishing qualitative descriptions can be challenging. For instance, the IS type was employed to depict different patterns, leading to confusion when reading papers^[8–9,31,58–60]. Moreover, electrophysiological recordings entail the application of increasing current gradients to elicit voltage responses across various stimuli. Consequently, the raw data encompass a series of discharge waveforms, each comprising numerous action potentials, thus constituting high-dimensional data. Traditional qualitative analyses in electrophysiology typically focus on describing only 1-2 sweeps, which inadequately harnesses the wealth of information available and potentially introduces biases in sweep selection^[54,56,61–62].

Currently, morphological qualitative classifications are mostly descriptive and highly subjective, as mentioned earlier, very rudimentary. Different individuals may have different descriptions and interpretations, leading to different experimental results. Morphological data itself consists of image data, and obtaining descriptive data through manual descriptions only allows for simple analysis through manual counting. It is challenging to use statistical methods to find relationships with other levels of characteristics such as electrophysiology and molecular data^[8–9,60,63–64], and the described relationships are highly subjective with poor repeatability.

In the past, semi-quantitative RT-PCR was used for qualitative analysis^[60], but with the development of high-throughput technologies, molecular data is primarily quantitative. Qualitative analysis is now mostly used for verifying the presence or absence of a few molecular expressions and is rarely used for qualitative classification^[15,22–23,51].

The development of single-cell sequencing technology has led to increasing sequencing throughput and efficiency. Large amounts of single-cell sequencing data for interneurons have been generated^[15,22,65–66], with thousands of genes detected in a single cell. High-throughput single-cell sequencing platforms, like 10X, can sequence thousands of cells at once, resulting in massive data generation. Such data must be analyzed using statistical methods, as qualitative analysis alone is insufficient. Additionally, as research progresses, there has been an accumulation of electrophysiological and morphological data. However, due to the complexity of interneurons and the large amount of data required for classification, relying solely on qualitative analysis by humans is inefficient and cannot meet current research demands.

The data of interneurons is depicted in Figure 1-5, described from various aspects including molecular, morphological, electrophysiological, and circuit connections. Moreover, as mentioned above, both morphological, electrophysiological, and molecular data itself possess high dimensions, making qualitative analysis inefficient in covering the information comprehensively. It is challenging to provide an objective and accurate judgment on data of such high dimensions.

Figure 1-5 Neuronal multidimensional characteristic data^[16].

1.5 Interneuron classification requires quantitative classification

As research progresses, an increasing number of characteristics of intermediate neurons are being revealed, demanding stable and efficient methods for information extraction and integration. In order to achieve reproducibility in classification, clear and quantitative classification criteria are needed. Molecular data types are mostly generated by high-throughput sequencing technologies, which inherently possess quantitative features. Combined with various statistical analysis methods, other characteristic data inevitably require quantitative analysis when jointly analyzed with molecular data. Research in the molecular field has propelled the development of bioinformatics, and through the integration of machine learning, it is hopeful to combine molecular, morphological, electrophysiological, and other multidimensional data^[16].

1.6 Quantitative classification

1.6.1 Supervised classification and unsupervised classification

The term "classification" has two meanings for classifying interneurons. The first involves categorizing the research subject into a known type, such as determining whether interneurons are PV, SST, or 5Ht3aR types through gene expression detection using RT-PCR. The second, more general meaning, involves defining and distinguishing neuron types and categorizing them. The difference lies in whether the classification types are known, corresponding to two major analysis techniques in data mining: supervised classification (also directly referred to as classification), which requires known labels for data and assigns labels to new data; unsupervised classification, commonly known as clustering analysis, where the number of classes into which data can be divided is unknown, and natural clusters are formed based on the "distance" between data points, thereby classifying the data^[16].

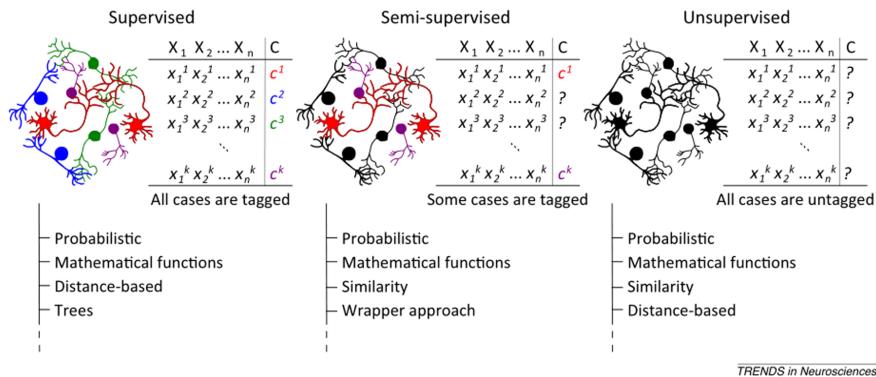


Figure 1-6 Classification Methods and Their Data Structures^[16].

As shown in Figure 1-6, assuming a dataset D to be classified, which records k observed neurons, each neuron described by n+1 descriptive parameters. The first n parameters represent observed parameters $X_1 \ X_n$ (which can be, electrophysiological frequency, rheobase current strength, coefficient of variation of interspike interval (ISI CV), etc.; morphological parameters such as cell body size, axonal branching number, molecular expression levels, etc.). The last parameter records the classification label C of the neuron. In supervised classification, all k neurons recorded have corresponding classification labels C for all n parameters. In semi-supervised classification, some records do not have corresponding classification labels C, while in unsupervised classification, none of the records have classification labels C^[16].

1.6.2 Electrophysiological Quantitative Classification

The data itself of electrophysiology has certain quantitative characteristics. At each sampling point, a membrane potential value is recorded according to the sampling frequency, and all the sampling points recorded by a single stimulus constitute a discharge waveform, referred to as a "sweep". Starting from a subthreshold stimulus, each subsequent stimulus incrementally increases until the discharge frequency reaches a stable state. All recorded sweeps constitute a set of experimental data. The quantitative classification of electrophysiological data has evolved from early descriptions^[8-9,16,54]. The primary parameters for describing physiological characteristics are the shape and firing

time of action potentials, such as peak width, amplitude, interspike interval (ISI), Latency, etc [9, 56, 62]. Unsupervised classification methods such as Ward hierarchical clustering and k-means are widely used [56, 60, 62–64, 67–70]. Some studies analyze morphology and electrophysiology separately using hierarchical clustering [67–70]. Some studies combine morphology and electrophysiological quantitative data for cluster analysis [62]. There are even studies on clustering analysis of the distribution patterns of ISIs within a single sweep of action potentials.

Single hierarchical clustering or k-means cannot adapt to the analysis of high-dimensional large datasets. Some studies extract 38 parameters based on the main electrophysiological characteristics proposed by the Petilla Committee [9], and apply unsupervised classification methods such as Objective Nested Clustering Analysis (ONCA) to analyze the parameters extracted from sweeps recorded under instantaneous and stable states. They also combine supervised classification methods like DF Analysis with manual classification for comparison, constructing a hierarchical classification framework to adapt to large and complex datasets [54].

In summary, regardless of the classification method, electrophysiological quantitative research requires parameter extraction from data. Too many parameters lead to high-dimensional datasets, which are not conducive to analysis, while too few parameters may omit information and fail to accurately describe the data. Therefore, parameters that can comprehensively capture information more accurately still need to be explored. Additionally, the diversity of interneurons determines the demand for large datasets. Thus, classification methods must be able to adapt to large datasets, making a robust analytical system a necessary solution.

1.6.3 Morphological Quantitative Classification

The original data of morphology is image data, which is the most difficult part to quantify among several aspects of interneurons. Its quantitative study has always been an important issue in morphology research. The basic analysis process is shown in Figure 1-7, mainly with two challenges: firstly, efficient image processing algorithms are required; secondly, quantitative description (parameter extraction) [71].

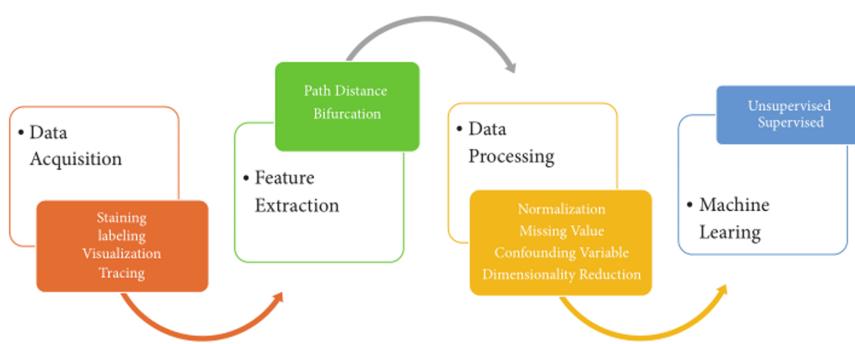


Figure 1-7 Analytical Process of Morphology.

Currently, quantitative analysis of morphology is primarily conducted through the reconstruction software applied to acquired neuron images, followed by extraction of descriptive parameters from the reconstructed images to obtain morphological quantitative information [71]. Interneurons possess complex dendritic structures, making manual reconstruction time-consuming and subjective. Some commercial software such as

Amira, Neurolucida, Imaris, as well as open-source software like Vaa3d, ImageJ, have gradually begun research on automated tracing^[72–77]. Automated tracing significantly improves reconstruction efficiency, avoids errors caused by subjective judgment, but demands high image quality and faces challenges in biological specimen background noise. Hence, noise reduction algorithms have become the main research focus in automated neuron tracing^[16,71,73]. Neuron image data are influenced by various factors such as biological organisms and specimen preparation, typically requiring complex image processing workflows that need adjustments based on experimental data. To better analyze neuron morphology data, many laboratories have developed morphology processing pipelines, with analysis processes like NBLAST proposed^[78]. Additionally, some scientists have established online forums and communities like BigNeuron to share neuron morphology analysis workflows, offering researchers choices based on their analytical goals^[79]. However, although numerous analysis processes and algorithms have been proposed, the high requirements in computer technology make it challenging for experimental neuroscientists to utilize them, necessitating more experimental testing to find tools with higher adaptability and user-friendliness.

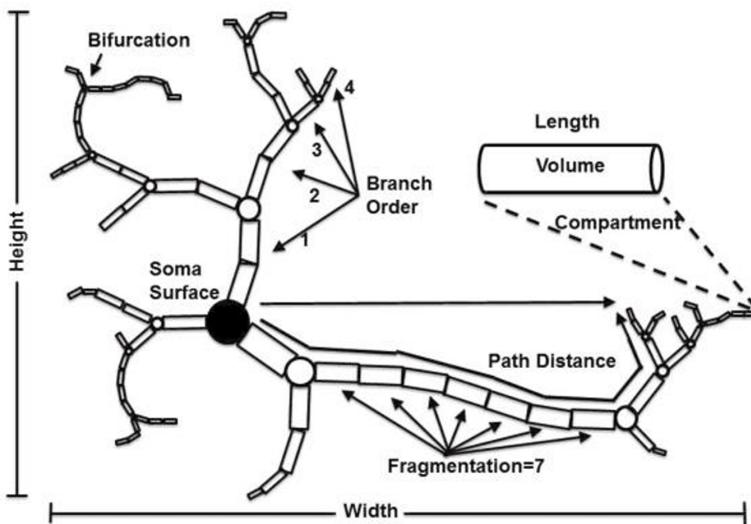


Figure 1-8 Quantitative parameters of morphology.

After obtaining reconstructed data, it is necessary to extract descriptive parameters from it for the application of various analytical methods. Currently, commonly used morphological descriptions, as shown in Figure 1-8, mainly include descriptions based on soma, branching projection distance, and coverage area^[8–9,71]. L-measure is a morphological calculation software developed in recent years, capable of extracting over 30 morphological parameters from reconstructed neurons^[80]. Once the parameters are extracted, they can be analyzed using various data analysis methods. Currently, popular methods for analyzing neuronal morphology include Ward's hierarchical clustering, k-means clustering^[59–60,63–64,68,81], and some studies have also utilized tSNE (t-Distributed Stochastic Neighbor Embedding) for analysis^[24].

1.7 Hierarchical classification system

The classification of interneurons and the challenges faced at the species level can be referenced in strategies for researching the classification of interneurons^[?]. There are pri-

marily two strategies for defining types of cells: 1) Based on phylogenetic relationships, where cells that undergo simultaneous changes are artificially considered the same type from an evolutionary perspective^[53]; 2) Using typological methods to group genes or phenotypic similarities into the same type. The first strategy is more ideal, as both development and evolution are complex processes making it difficult to prove sameness. It is not necessarily the case that the same types of neurons are invariably linked in development and evolution, nor are different types^[3]. The second strategy better suits the needs of interneuron classification and is currently the preferred approach in this research field. It aims to group neurons with similar molecular types, morphology, and electrophysiology into one category^[3,8–9,16,82].

Inspired by typology, Hongkui Zeng and Joshua R. Sanes propose that neuron classification should^[3]:

1. Define neuron types through multiple criteria, with objectivity rather than subjective preferences of researchers;
2. Classify neurons based on rules, clear, and quantifiable standards;
3. The classification system should be hierarchical;
4. The results of classification should be viewed as a hypothesis to be tested, rather than rigid criteria;
5. Classification should focus on inter-group discrete variables rather than continuous variables.

1.8 Further demand for interneuron classification in both dry and wet experiments

With the rise of bioinformatics and the deepening research in data analysis, traditional biology research based on biological experiments has been transformed. The concept of experiments has expanded, and current biological research experiments are divided into: wet experiments, which are traditional biology experiments using biological materials as research materials; dry experiments, which primarily use data as research materials and explore and extract information through data analysis. In summary, intermediate neurons exhibit complex types in morphology, electrophysiology, molecular biology, and other aspects. As research deepens and a large amount of experimental data accumulates, the classification of interneurons needs to be quantified. Analytical methods using data mining integrate information from various aspects to classify them. Currently, the demand for data analysis in the classification of interneurons is increasing. In order to better analyze a large amount of classification data, both wet and dry experiments have different requirements, allowing them to better cooperate and extract meaningful information.

In traditional biological experiments, different laboratories have their own experimental procedures. Due to the complexity of interneurons, relying solely on data from one laboratory is far from enough. Many data analysis researchers need a large amount of data as research material. The lack of a unified standardized data collection process is very detrimental to dry experiment research^[3,16,54]. The inability to integrate various experimental data makes it difficult to obtain reliable classification results from small

sample data, which also leads to data waste and increased workload. Currently, many laboratories have realized this problem and are starting to standardize the data collection of wet experiments^[22,24].

Integrating data from various laboratories and levels requires not only standardization of wet lab data collection processes but also the establishment of a common platform database for integrating data from different laboratories, similar to public databases in molecular biology such as NCBI. Additionally, it is necessary to establish a sharing system, akin to sequencing articles being published in public databases like GEO, with standardized experimental descriptions and metadata.

Since the classification data of interneurons will be very large and high-dimensional, and classification requires a multi-level system, a robust analysis system and analysis process are needed. It is necessary to transition from using a single classification method to building a hierarchical classification system using simple classification methods, continuously attempting dry experiments, constructing a reasonable analysis system, and extracting meaningful information as much as possible during this process.

Chapter 2

Introduction

The powerful computational capacity of the nervous system primarily relies on the highly connected neural networks and the dynamic microcircuits composed of neurons. The neurons constituting microcircuits are mainly of two types: excitatory neurons, which are glutamatergic, and inhibitory neurons, which are GABAergic (gamma-aminobutyric acidergic) also known as interneurons^[1,3,6,8–9]. The maintenance of dynamic balance between excitation and inhibition in the cortical circuits relies on the regulation of various types of interneurons. Interneurons in the cortex exhibit high diversity in terms of electrophysiological, morphological, and molecular features, input-output connectivity patterns, etc. Despite comprising only 20%-30% of cortical neurons, their high diversity plays an extremely important role in regulating neuronal activity, neural circuit plasticity, and synchronization of neural networks^[22–24,51,83–84].

The crucial regulatory functions of interneurons correspond to disruptions in neural system function, leading to various neurological disorders such as epilepsy, schizophrenia, anxiety disorders, and autism^[10–13]. Therefore, research on interneurons is of significant importance for understanding neural network regulation, neural system function, and neurological disorders.

However, the diversity of interneurons and their low proportion in the entire nervous system pose the greatest challenges in studying them. Currently, with the development of molecular genetic manipulation techniques and high-throughput technologies such as next generation sequencing, it is possible to label and observe different types of interneurons on a large scale and manipulate their activities to study their circuit functions^[1,14–15,22,51,83]. To some extent, the low proportion no longer hinders the study of interneurons, and their diversity has become the main issue. Although in recent years, through high-throughput experimental techniques and the collective efforts of many laboratories, we have gained some understanding of interneuron diversity, organizations such as PING (The Petilla Interneuron Nomenclature Group) have also attempted to standardize the Nomenclature and description of interneurons. However, there is still much we do not understand about interneurons—such as the various types that exist, their similarities and differences, and the distribution and organizational structure of these different types^[1,3,8–9,22,24,51,82].

Qualitative descriptions in early studies of the nervous system were sufficient, but with the development of high-throughput technologies, the generation of large amounts of high-dimensional data, the increasing complexity of discovered neurons, and deepening understanding, qualitative methods are no longer adequate for current research on the nervous system^[3,8–9]. Due to the complex diversity of interneurons, there is an increasing demand for quantitative classification^[3,9,16,19]. With the development of fields such as machine learning and data mining, more robust algorithms are emerging, allowing us to objectively and finely classify neuron types using quantitatively described

neuronal characteristic data^[16].

Currently, the classification of interneurons mainly focuses on electrophysiological, morphological, and molecular aspects^[3,8–9,15–16,82]. This study focuses on the quantitative classification of interneurons based on electrophysiology.

2.1 Electrophysiological classification

The electrophysiological data itself has certain quantitative characteristics. According to the sampling frequency, each sampling point records a membrane potential value. All sampling points recorded during a single stimulus constitute a discharge waveform, referred to as a "sweep." Starting from a sub-threshold stimulus, the stimulus intensity increases gradually until the discharge frequency reaches a stable state. All recorded sweeps constitute one experimental data record^[8–9,54,81].

Early descriptions and nomenclature of these electrophysiological discharge patterns were very vague, with different laboratories using different descriptions and names. It wasn't until 2008 when the Petilla provided a more systematic naming of interneuronal characteristics before the description of these neurons^[9]. Electrophysiology is categorized into transient-state patterns and steady-state patterns, based on whether it has not yet reached maximum firing frequency or has reached maximum firing frequency, respectively. Classification and description of electrophysiology are performed based on sweep characteristics under these two states (Figure 2-1).

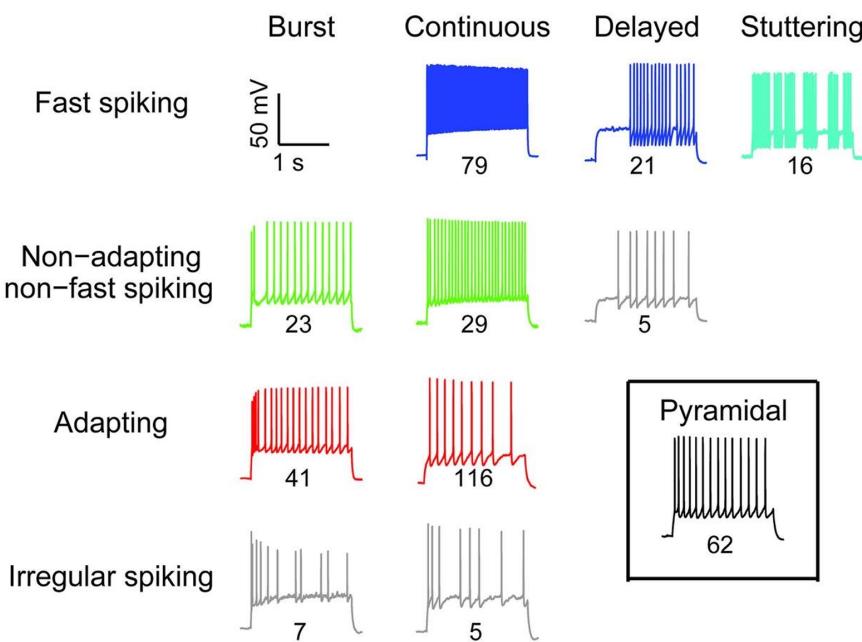


Figure 2-1 **Types of Interneuron Electrophysiology.** Each row corresponds to the discharge characteristics of sweep under stable state, and each column corresponds to the discharge characteristics of sweep under transient state. The numbers represent the number of cells classified in this literature^[9].

Quantitative classification of electrophysiological data has evolved based on early descriptions^[8–9,16]. The main parameters for describing physiological characteristics are the shape and firing time of action potentials, such as peak width, amplitude, ISI, Latency, etc.^[8–9,54,56]. Unsupervised classification methods such as Ward hierarchical clustering and k-means are widely used^[56,61–62,85], with studies using hierarchical

clustering for morphological and electrophysiological analysis separately^[56,68–69,85–86]; there are also studies combining morphological and electrophysiological quantitative data for cluster analysis^[62]; and even studies clustering analysis of intersweep interval (ISI) distribution patterns of a single sweep action potential^[61].

Single hierarchical clustering or k-means cannot adapt to the analysis of high-dimensional large datasets. Some studies have proposed the extraction of main electrophysiological type characteristics according to the Petilla, extracting 38 parameters from sweeps under instantaneous and stable states, respectively, applying unsupervised clustering using Objective Nested Clustering Analysis (ONCA) and supervised classification with DF Analysis, comparing with manual classification to construct a hierarchical classification framework to adapt to large and complex datasets^[54].

In summary, regardless of the classification method, electrophysiological quantitative classification studies require parameter extraction from data. Too many parameters lead to high-dimensional datasets, which are not conducive to analysis, while too few parameters may omit information and fail to accurately describe it. Therefore, parameters that can comprehensively summarize information and are more precise and suitable need further exploration. Meanwhile, the diversity of interneurons determines the demand for large datasets, so classification methods must be able to adapt to large datasets, making a robust analytical system a necessary consideration.

This study describes interneuronal electrophysiological data by extracting quantitative parameters and conducts quantitative classification research on interneuronal electrophysiology using different clustering methods.

2.2 Clustering analysis

Two major analytical techniques in Data Mining: supervised classification (also directly referred to as classification), which requires known data labels for classifying new data; unsupervised classification, commonly known as clustering analysis, where the number of classes into which data can be divided is unknown, and natural clusters are formed based on the "distance" between data points, thus classifying the data^[16].

Cluster analysis is an important means of exploratory cluster analysis and is now widely used in many fields such as machine learning, image analysis, and bioinformatics^[64,67,87–89]. Currently, there are many clustering algorithms (cluster models), which can be broadly summarized as follows:

1. **Centroid-based clustering models:** Algorithms designed from the perspective of repeatedly searching for class centroids. For example, K-means clustering, also known as fast clustering, uses centroids as the core to group multiple observations that are close to the centroid into one class. The resulting distance results are generally determined and do not have a hierarchical structure relationship, and the number of clusters K needs to be determined in advance. K-means is currently a widely used clustering analysis method in neuronal classification^[56,62].
2. **Connectivity-based clustering methods:** Algorithms designed from the perspective of distance and connectivity. These algorithms regard multiple observations that are close in space as one class, and the resulting clustering results are generally deterministic and hierarchical. The core idea of hierarchical clustering is to gradually merge individual observations into small clusters, and then merge

small clusters into larger ones. Different algorithms for calculating distances between classes produce different hierarchical distance methods, with main distance calculation methods including: nearest neighbor method, complete linkage method, average linkage method, centroid method, and Ward hierarchical clustering (also known as the sum of squares method). Ward hierarchical clustering is the most commonly used classification strategy in this category of algorithms, often used together with K-means in quantitative studies of neuronal classification^[56,60,62,81].

In addition to the above clustering methods, there are also clustering models based on statistical distributions, density-based clustering models, graph-based clustering models, etc.

Currently, in neuronal classification studies, K-means and Ward hierarchical clustering are typically used together for classification. There is also evidence suggesting that K-means is more efficient than Ward hierarchical clustering. However, since K-means requires setting the number of clusters k and lacks a hierarchical structure, it is highly sensitive to parameters with strong correlations. Neurons themselves exhibit heterogeneity and tend towards continuity, thus having intermediate states. A classification method with a hierarchical structure can more flexibly obtain information about the relationships between different types of cells and is more conducive to research on interneuron classification.

2.3 Principle Component Analysis

Principal Component Analysis (PCA) is a method of simplifying a dataset consisting of correlated multivariate data points in a multidimensional space into a reduced set of mutually independent composite indicators. In practical research, dealing with multivariate data is a common challenge, where variables often exhibit certain degrees of correlation, leading to overlapping information and increasing the complexity of the problem. PCA, inspired by Singular Value Decomposition (SVD) in mathematics, transforms correlated multivariate data into a set of mutually independent and unrelated variables. By selecting new variables that contain the most information, PCA aims to retain as much of the original information as possible while discarding irrelevant details. With the advancement of high-throughput sequencing driving the field of bioinformatics, PCA finds extensive applications in the analysis of molecular sequencing data^[22,90–91].

2.4 Iterative principal component analysis

With the rapid development of big data analysis in recent years, classical single clustering analysis methods have been commonly applied, driving the advancement of bioinformatics, especially in the field of molecular biology research^[87–89]. However, classical single clustering analysis methods struggle to adapt to high-dimensional massive datasets. PCA effectively extracts the main information from the dataset by dimensionality reduction. By iteratively integrating PCA and clustering methods, a more stable analysis system is constructed, enabling better analysis of heterogeneous big data^[22,92–93], which has started to be applied in the molecular domain^[10,22].

The figure shown in Figure 2-2 illustrates the iterative process of Iteration Principle Component Analysis (iPCA). Each column represents one iteration, during which a

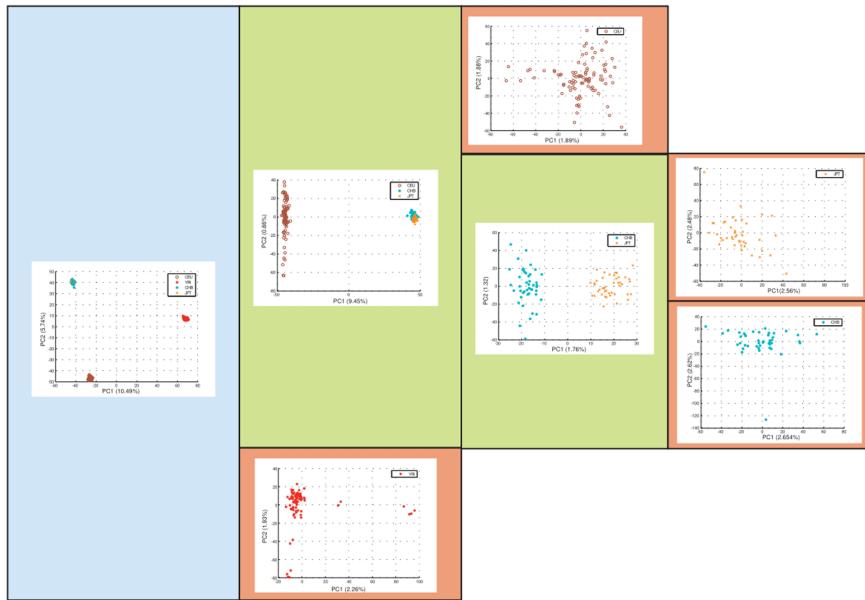


Figure 2-2 iPCA Process. Each column represents one iteration, with each subclass undergoing PCA computation once. Scatter plots depict data in the PC1/PC2 coordinate system; different colored rectangles represent the subcategories involved in each iteration, while differently colored dots represent distinct subclasses obtained from PCA calculations in each iteration^[93].

PCA calculation is performed to extract the most important information from the current subclass. Clustering algorithms are then applied to cluster the data into two subclasses for the next iteration^[93].

2.5 Significance of this study

The aim of this study is to quantify the electrophysiological data of interneurons, using clustering methods to attempt classification of interneuron electrophysiology, and to construct a quantified classification system for interneuron electrophysiology, thus avoiding the interference of artificial classification. This system will be used for integrating electrophysiological, morphological, and molecular classification data in future interneuron research.

This study selected 569 interneuron electrophysiological data from the Allen Cell Type public database, focusing primarily on statistical parameters of variation between different action potentials (APs). Thirty different electrophysiological characteristic parameters were extracted to construct the quantified classification system for interneuron electrophysiology, aimed at future interneuron classification research.

Ward hierarchical clustering and iPCA methods were employed for cluster analysis, resulting in 7 and 8 subclasses, respectively. A comparison between the two methods revealed that iPCA performed better for clustering large datasets compared to Ward hierarchical clustering. Additionally, iPCA was found to more flexibly demonstrate hierarchical relationships among various types during the analysis process.

Chapter 3

Materials and Methods

3.1 environment, software and modules

Intel® Core™ i7-7700K processor, 64GB memory, x86_64-w64-mingw32/x64 (64-bit), running Ubuntu for Windows 4.4.0-17134-Microsoft environment, with Python 3.6.4 and R 3.5.0 as the primary analysis tools.

Name	Version	Description	Type
Cmder	1.3.10.811	Linux command console	Software
Xming	6.9.0.31	Linux subsystem graphical interface	Software
Vim	7.4	Script writing under Linux subsystem	Software
R	3.5.0	R language compiler	Software
RStudio	1.1.463	R language visual programming env.	Software
Python	3.6.4	Python compiler	Software
pandas	-	Python data analysis package	Python package
numpy	-	Matrix processing under Python	Python extension library
matplotlib	-	2D plotting library for Python	Python extension library
collections	-	Data structure containers in Python	Python module
allensdk	-	Allen Cell Type database tool	Python module
eFEL	-	Blue Brain electrophysiology analysis library	Python package
Igor	6.2.2.2	Electrophysiology data analysis	Software
ggplot2	-	Statistical plot drawing in R	R package
ggbiplot	-	Biplots drawing in R	R package
ggpubr	-	Statistical plot formatting in R	R package
pheatmap	-	Heatmap drawing in R	R package
CorelDraw	16.0.0	Vector graphics processing, schematic drawing Software	

Table 3-1 Software and tools

3.2 Data resources

3.2.1 Datasets selection

This study selected 569 interneuron electrophysiological data from [the Allen Cell Type public database](#), from the mouse visual cortex. The data is categorized as Table 3-2:

All the above data were recorded using whole-cell electrophysiology under 1s current stimulation, with sampling frequencies of 50Hz and 20Hz. For detailed experimental recording methods, please refer to the documentation of [the Allen Cell Type database](#).

Mouse Strain	Number of Cells
Chat-IRES-Cre-neo	38
Nos1-CreERT2 Sst-IRES-FlpO	26
Htr3a-Cre-NO152	127
Pvalb-IRES-Cre	139
Vip-IRES-Cre	70
Ndnf-IRES2-dgCre	68
Sst-IRES-Cre	101

Table 3-2 Dataset composition

3.2.2 Data download

The Allen Cell Type database can be accessed and utilized for data selection and downloading using the *Python* toolkit provided by this database, called *allensdk*, through its Application Programming Interface (API). The specific steps for data selection and downloading are as follows:

1. Installation of *allensdk*

Ensure *Python* version 3.6.4 is installed and pip is available, then install *allensdk* using pip:

```
pip install allensdk
```

2. Building *CellTypesCache*

The *CellTypesCache* class provides a *Python* interface for downloading data from the Allen Cell Types database to a specified location. The usage is as follows:

```
from allensdk.core.cell_types_cache import CellTypesCache
# Import the CellTypesCache module through import
ctc = CellTypesCache(manifest_file='PATH/manifest.json')
```

Create a *manifest.json* file, where *PATH* is the file path. Upon successful creation, a *manifest.json* file will be generated under this path, and a *CellTypesCache* class named *ctc* will be created as the database interface.

3. Downloading metadata of cell records from the database Store it in CSV format for easy browsing and filtering in Excel.

```
cells = ctc.get_cells()
#Obtain a list named cells that records metadata information.
```

4. After filtering metadata in Excel, extract the IDs of the required data to build a CSV file for batch downloading.

5. Downloading electrophysiological data from the database.

```
cdata_set = ctc.get_ephys_data(Id)
```

Fill in the data Id in the parentheses to download the corresponding Id. The downloaded file is an *NWB* binary file, and the download path is: 'PATH/Id/ephys.nwb'. PATH is the path where the *manifest.json* file is located.

3.3 Data format conversion

The raw electrophysiological data recorded in the Allen Cell Type Database are stored as *NWB* binary files, from which specific electrophysiological parameters can be extracted directly using the *SDK*. However, considering that the electrophysiological data recorded in the laboratory are in *data* format, to facilitate subsequent analysis and comparison of the experimentally recorded electrophysiological data, a compromise is made to convert all data into text format. In this format, even-numbered columns record stimulus currents, while odd-numbered columns record stimulus voltages.

3.3.1 Convert NWB format to *txt* format

1. Calling the *NwbDataSet* module of the *allenSDK*

```
from allensdk.core.nwb_data_set import NwbDataSet
```

2. Obtain the sweep ID for the 'Long Square' recording mode

```
# Obtain corresponding sweep ID records
sweeps = ctc.get_ephys_sweeps(file_list.loc[i, 0])
# Iterate through sweeps
for sweep in sweeps:
    sweep_numbers[sweep['stimulus_name']].append(
        sweep['sweep_number'])
# Get sweep IDs for the "Long Square" recording mode
sweep_numbers = sweep_numbers['Long Square']
```

3. Loading downloaded *NWB* data

```
data_set = NwbDataSet(filename)
```

4. Extract current stimulation information and discharge reaction data using sweep ID

```
sweep_data = data_set.get_sweep(sweep_number)
# Extract data corresponding to sweep ID
stim = sweep_data['stimulus'] * 1000000000000
# Extract stimulus information, unit in pA
voltage = sweep_data['response'] * 1000
# Extract discharge response information, unit in mV
```

(The above are all single sweep methods, which need to be combined through data frame operations. For detailed methods, please refer to the complete code.)

3.3.2 *data* format to *txt* format

The electrophysiological data recorded in the laboratory are in *data* format. After being read by *Igor*, the information of voltage and current is imported. It is then structured into a table with even-numbered columns recording stimulus current and odd-numbered columns recording stimulus voltage, and saved as a text file.

3.4 Extraction of electrophysiological parameters

3.4.1 eFEL toolkit installation and invocation

eFEL (The Electophys Feature Extract Library) is an algorithm library developed by the Blue Brain Project, provided free and open-source for neurobiologists to extract electrophysiological data characteristics from whole-cell patch-clamp recordings. There are two versions written in *C++* and *Python*. This study utilizes the *Python* version.

1. eFEL installation

eFEL requires *Python 2.7+* or *Python 3.4+* versions. *Pip* is required for installing *eFEL*, and *Numpy* needs to be installed simultaneously. Installation should be performed on Linux/UNIX/MacOSX/Cygwin systems. To install using *pip*:

```
pip install git + git://github.com/BlueBrain/eFEL
```

If installation fails, install directly in the *home* folder:

```
pip install --user git + git://github.com/BlueBrain/eFEL
```

2. Basic Process of eFEL Invocation

Firstly, load the *eFEL* and *Numpy* modules:

```
import efel
import numpy
# Importing data using \textit{Numpy}
# (data already converted to \textit{txt} format):
data = numpy.loadtxt("ephys.txt")
```

eFEL recognizes data structures of the list class, capable of recording multiple sweeps for parameter extraction. Each sweep's data should be of the dictionary class, containing four key names: 'T', 'V', 'stim_start', and 'stim_end', corresponding to key values: an array for recording time, an array for recording corresponding voltages, the start time of stimulation (a list type containing only one element), and the end time of stimulation (a list type containing only one element). Below is the method for constructing a list of single sweeps; detailed analysis for multiple sweeps is provided in the analysis code.

Construct time variable based on total recording duration:

```
time = np.arange(0, 1200, float(1200) / len(data))
#Recording duration is 1200
```

Extract voltage information for the first sweep from data:

```
Voltage equals data[:, 1].
# Assigning the first column of data to the variable 'voltage'
```

Construct a dictionary class for a single sweep:

```
sweep1 = {} # Initialize a dictionary to record a single sweep
sweep1['T'] = time # Assign the variable 'time' to 'T'
sweep1['V'] = voltage # Assign the variable 'voltage' to 'V'
sweep1['stim_start'] = [200] # Stimulation starts at time 200ms
sweep1['stim_end'] = [1000] # Stimulation ends at time 1000ms
# Constructing a list containing the single sweep:
sweep = [sweep1]
```

The constructed sweep data can utilize the `efel.getFeatureValue()` function to call algorithms from the library for parameter extraction. The following code calculates the action potential amplitude and baseline:

```
result = efel.getFeatureValue(sweep, ['AP_amplitude', 'voltage_base'])
# Parameter names should also be in list format,
# allowing extraction of multiple parameter names.
```

The return value is also assigned to a variable `result`, which is also of the list class, where each element corresponds to a record of each sweep in the sweep list. Key names correspond to parameter names, and key values represent parameter values.

3.4.2 Electrophysiological Parameter Extraction Algorithm

1. Fi fit slope

The discharge frequency and corresponding stimulus current are fitted with a linear curve, and the slope of the fitting curve is termed as "fi fit slope". Using *eFEL*, the number of action potentials (APs) produced in all sweeps is calculated as n divided by the stimulus time (1s) to obtain the discharge frequency of each sweep. The fi fit slope is calculated from the fitted current and frequency.

2. AP interval time correlation

AP interval time correlation mainly includes:

ISI (Interspike interval): the time interval between adjacent peaks, where a sweep with n APs has $n - 1$ ISIs, denoted as $ISI_1, ISI_2, \dots, ISI_n$.

Latency: the time from the start of the stimulus to the peak of the first AP.

Last ISI end: the time from the peak of the last AP to the end of the stimulus.

These parameters serve as the basic parameters for AP variations within the sweep, while other related parameters are ISI statistical parameters. The algorithmic approach to calculating these basic parameters is as follows:

Using *eFEL*, obtain the peak position index of each AP on each sweep. The first AP to the last AP is recorded as x_1, x_2, \dots, x_n . Calculate the time corresponding to each index based on the sampling frequency to obtain the time at each AP peak, then compute the ISIs between adjacent two AP peaks. Latency is obtained by subtracting `start_time` from the time of the first AP peak, and last ISI end is obtained by subtracting `end_time` from the time of the last AP peak.

3. AP shape-related parameters

Parameters describing AP shape include amplitude and threshold, which are commonly used electrophysiological parameters and are directly obtained from *eFEL*. The threshold is defined as the index where the first derivative exceeds $12V/s$ at the onset of the action potential (AP begin index), with the corresponding voltage value being the threshold. The time difference at the AP peak is the rise time. The amplitude is the corresponding voltage value at the AP peak.

4. AHP-related parameters

These primarily include the amplitude and latency of afterhyperpolarization (AHP), which are respectively the voltage difference from the threshold position to the lowest point of the AHP and the time difference between the threshold position and the time of the AHP's lowest point. The index of the AHP's lowest point is obtained from *eFEL*.

5. Hyperpolarization stimulus-related parameters

These mainly concern changes in membrane potential under hyperpolarization conditions, including:

Input Resistance (R_{in}): The voltage at the stable position under hyperpolarization (voltage value before the end of the stimulus) is obtained by calling *eFEL*. The ratio to the corresponding voltage is the resistance of the sweep, with the average resistance of all hyperpolarization sweeps representing R_{in} .

Time constant (tau), Sag value, and Baseline V are directly obtained from *eFEL*.

3.4.3 Statistical integration and construction of analysis matrix for electrophysiological parameters

1. Coefficient of variation

Describing the changes in APs within a sweep in the future, using the coefficient of variation (CV) for parameters such as ISIs, amplitudes, thresholds, etc.

2. Adapting Index

There is a gradual increase in the interval between action potentials of adapting electrophysiological type from the beginning of stimulation. Introducing the Adaption Index parameter:

$$\frac{1}{N - 1} \sum_{n=1}^{N-1} \frac{ISI_{n+1} - ISI_n}{ISI_{n+1} + ISI_n} \quad (3.1)$$

3. Parameters of Stable State

When calculating parameters, all sweep parameters are calculated, but for the description of electrophysiological characteristics, it is not possible to describe a cell with all sweep parameters, which would lead to excessive data dimensions and redundancy. According to the definition of the PING conference, the characteristic types are divided into transient state and stable state. Parameters describing non-transient state characteristics in this study are described using parameters calculated from stable state sweeps.

4. Parameters of Transient State

The characteristic of Delay type is that the first few sweeps that generate action potentials have a relatively large latency time, so the first sweep with an action potential is taken to describe the latency of that sweep, calculated as 1st latency; Since $ISI\ CV$ describes the variation in time intervals between APs within a sweep, considering that both transient and stable states have variations, in order to reduce

the number of parameters while covering overall information, this parameter is calculated as the mean of the coefficients of variation of all sweep *ISIs*, as the final analysis parameter. Similarly, First two *ISI* ratio and delay ratio are also considered, and the mean of all sweeps is calculated as the final analysis parameter.

3.5 Selection of parameter

1. Kernel Density Estimation

Extract all parameters extracted by the statistics department, and analyze their nuclear density distribution in different transgenic mouse strains (Kernel Density Distribution). Conduct ANOVA hypothesis testing on various parameters across different mouse strains, selecting those parameters with significant differences ($p < 0.05$).

2. PCA

Using the `scale` function in R language to perform z-score transformation on the constructed analysis matrix using standard deviation (SD) for standardization:

```
scale (x, center = TRUE, scale = TRUE)
```

- *x* is the parameter matrix.
- *center* describes whether the input matrix is numeric, set to TRUE.
- *scale* describes the form of the output standardized matrix, set to TRUE to output a numeric matrix consistent with the input matrix in terms of column numbers.

Using the `princomp` function in R language to perform PCA calculation on the standardized matrix:

```
princom (x, cor = FALSE)
```

- *x* is the matrix standardized through z-score.
- *cor* describes whether to use a correlation matrix or covariance matrix for principal component analysis, set to FALSE to use covariance matrix.

Using the R package *ggbiplots* to visualize the PCA analysis results through a biplot, analyzing the contribution of each parameter.

3. Correlation Analysis

Using the `cor` function in R language to compute the correlation coefficient matrix for the standardized matrix. Correlation coefficients greater than 0.6 are considered significant.

Using the R package *pheatmap* to draw a heatmap.

3.6 Ward's hierarchical clustering

3.6.1 Algorithm Principles

1. Hierarchical Clustering Basic Calculation Process:

- Initially, each data record forms its own cluster.
- Then, calculate the distances between all data points and cluster the closest ones into a smaller cluster, forming n-1 clusters.
- Next, measure the distances between the remaining data points and clusters and merge the data point or cluster that is closest in distance to form a new cluster.
- Repeat the above process, continually aggregating all observation points and clusters into larger and larger clusters until all data points are merged into one cluster.

The final classification results show hierarchical relationships between clusters, as shown in Figure 3-1.

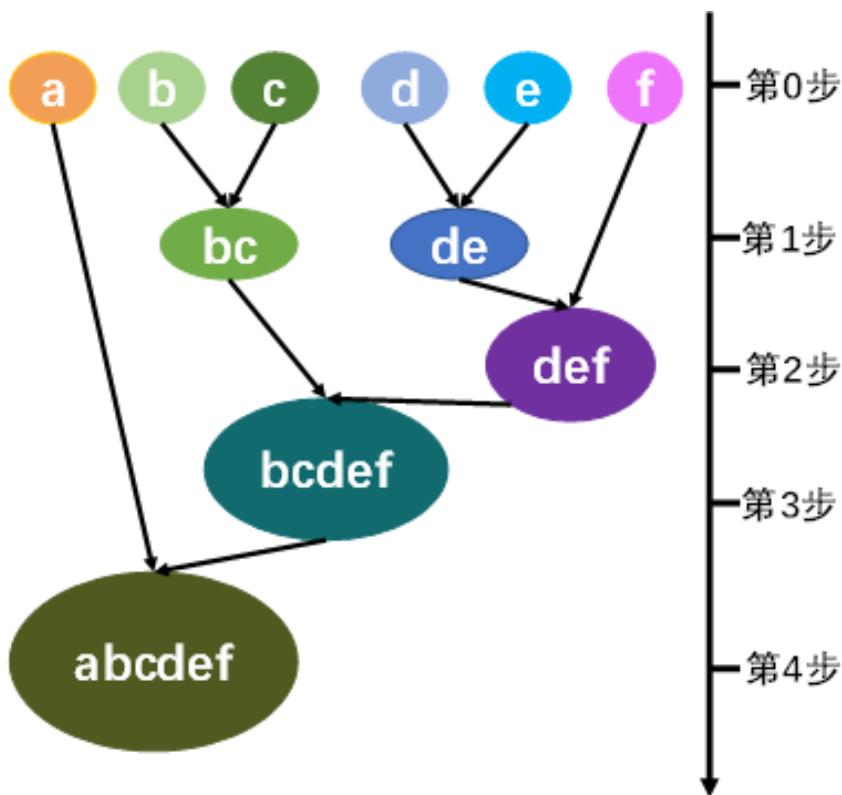


Figure 3-1 Hierarchical Clustering Process Diagram. At the initial stage, a, b, c, d, e, f form distinct groups. In the first step, b and c, d and e, which are closest to each other, are respectively merged into the groups bc and de. In the second step, distances between a, bc, de, and f are calculated, and f is closest to de, merging into the group def. In the third step, distances between a, bc, and def are calculated, and bc is closest to def, merging into the group bcdef. In the fourth step, a and bcdef are merged into one large group, and the calculation stops. It can be observed that there exists a hierarchical containment relationship among these groups.

2. Ward Algorithm:

The Ward method, also known as the minimum variance method, is used to calculate the distances between data points or clusters, known as Ward hierarchical clustering. The Ward method posits that if the classification is correct, the within-cluster sum of squares should be small, while the sum of squares between clusters should be relatively large. The formula for the sum of squares is as follows:

Assuming n samples are divided into k clusters G_1, G_2, \dots, G_k , with X_{it} representing the variable value of the i th sample in cluster G_t , n_t representing the number of samples in cluster G_t , and X_t representing the centroid of cluster G_t , the sum of squares for samples in cluster G_t is given by:

$$S_t = \sum_{i=1}^{n_t} (X_{it} - \bar{X}_t)(X_{(it)} - \bar{X}_t) \quad (3.2)$$

The total within-cluster sum of squares is given by:

$$S = \sum_{t=1}^k S_t \quad (3.3)$$

When n samples form a single cluster, $S = 0$. As samples are merged, S begins to increase. Each time, the two clusters with the smallest incremental increase in S are merged. With the Ward method, Euclidean distance is used to calculate the distances between samples.

3.6.2 Performing in R

1. Using the R language `dist` function to calculate the Euclidean distances between standardized data:

```
dist(x, method = "euclidean")
```

`-x` is the standardized matrix

`-method` is the distance calculation method, chosen as Euclidean distance.

2. Using the R language `hclust` function to perform Ward hierarchical clustering distance calculation:

```
hclust(d, method = "ward")
```

`-d` is the distance matrix obtained from calculating Euclidean distances

`-method` is the distance calculation method, chosen as Ward's method for hierarchical clustering.

3.6.3 Deciding on cut tree threshold

Using *heights (a.u)* values to describe the differences between two connected clusters. A higher *heights* value indicates that a group can be divided into two subgroups. The

optimal pruning test and upper-tailed t -test can be used to find the optimal number of clusters, k . The formula for the optimal pruning test is as follows:

$$\text{bestcut} = \text{mean}(\text{fusions}) + 1.96 \times \text{fusions} \quad (3.4)$$

Where fusions are all the heights sets in the Ward's clustering dendrogram, details can be found in the publication by McGarry et al. [56,70].

3.7 iPCA

3.7.1 PCA Analysis in R language

Principal Component Analysis (PCA) is a commonly used method in multivariate analysis. It involves mapping the dataset of multivariables into a space of mutually independent and unrelated variables using a mathematical algorithm called singular value decomposition, aiming to simplify the description of variables while retaining as much of the main information as possible [94].

1. After selecting parameters, the matrix is standardized using the *scale* function in R language to obtain a standardized matrix for further calculation. The parameters are set as follows:

```
scale (x, center = TRUE, scale = TRUE)
```

-x is the parameter matrix;

-center describes whether the input matrix is numerical, set to *TRUE*;

-scale describes the form of the standardized matrix output, set to *TRUE* to output a numerical matrix consistent with the input matrix columns.

2. In R language, the *princomp* function can perform PCA calculation. Its calling format is as follows:

```
princom (x, cor = FALSE)
```

-x is the matrix standardized through *z-score*;

-cor describes whether to perform principal component analysis using a correlation matrix or a covariance matrix, set to *FALSE* to use a covariance matrix.

3.7.2 Determine whether the classified subsets exhibit significant clustering

Statistical Significance of Clustering method (SigClust) is used to determine whether a dataset exhibits significant clustering, indicating the ability to be further divided into two or more subsets, suitable for iterative structural analysis. This analysis method sets the hypotheses as follows:

H0: The data come from a single Gaussian distribution.

H1: The data do not come from a single Gaussian distribution.

By hypothesis testing, a p -value is calculated, where $p < 0.05$ rejects the null hypothesis H0 and accepts H1, indicating that the dataset can be further partitioned; otherwise, H0 is accepted.

Using the `sigclust` function in the *R* package *SigClust*, the parameters are set as follows:

```
sigclust (x, nsim=50, label, icovest=2)
```

-*x* the subset data matrix to be tested;
 -*nsim* the number of simulated Gaussian samples, set to 50;
 -*label* the classification label vector;
 -*icovest* the covariance estimation type, set to 2, using sample covariance for estimation.

3.7.3 Integrating PCA and Ward's hierarchical clustering to build iPCA

iPCA utilizes an iterative process with unsupervised clustering methods to enhance the resolution of substructures. In each iteration, the dataset is recursively subdivided into two groups until a termination condition is met^[92–93]. We integrate PCA with Ward's hierarchical clustering method through an iterative process, where at each iteration, each group is further divided into two subgroups based on PCs until a termination point (determined by *SigClust* analysis $p > 0.05$ or when the subset data becomes too small). Each iteration proceeds as follows:

1. Perform PCA analysis on each subset data (or standardized matrix) using the *princomp* function, selecting principal components (PCs) with eigenvalues greater than the mean plus two times the standard deviation to describe dataset features.
2. Calculate the Euclidean distance matrix of PCs using the *dist* function, input it into the hierarchical clustering function *hclust*, and perform hierarchical clustering analysis using the Ward method.
3. Utilize the *cutree* function in *R* language with parameter *k* set to 2 to partition the dataset into two subsets and assign classification labels to the standardized matrix.
4. Input the labeled dataset into the *sigclust* function to compute *p*-values and determine if the dataset can be divided into two subsets.
5. Repeat steps 1-4 until no further subdivision is possible for all subset datasets.
6. Integrate all information from the sub-datasets obtained during the iPCA classification process into a binary tree structure and analyze the classification relationships among different subsets. See analysis code for details.

3.8 Code accessibility

All code on [my GitHub](#).

3.9 Statistics

The data are presented as *mean±se*. Student's *t*-test is used for comparisons between two groups, while analysis of variance (ANOVA) is employed for comparisons involving

more than two groups. Differences between multiple groups are assessed using the Tukey-Kramer post hoc test following ANOVA. A $p < 0.05$ is considered statistically significant.

Chapter 4

Results

4.1 Quantification of electrophysiological characteristics

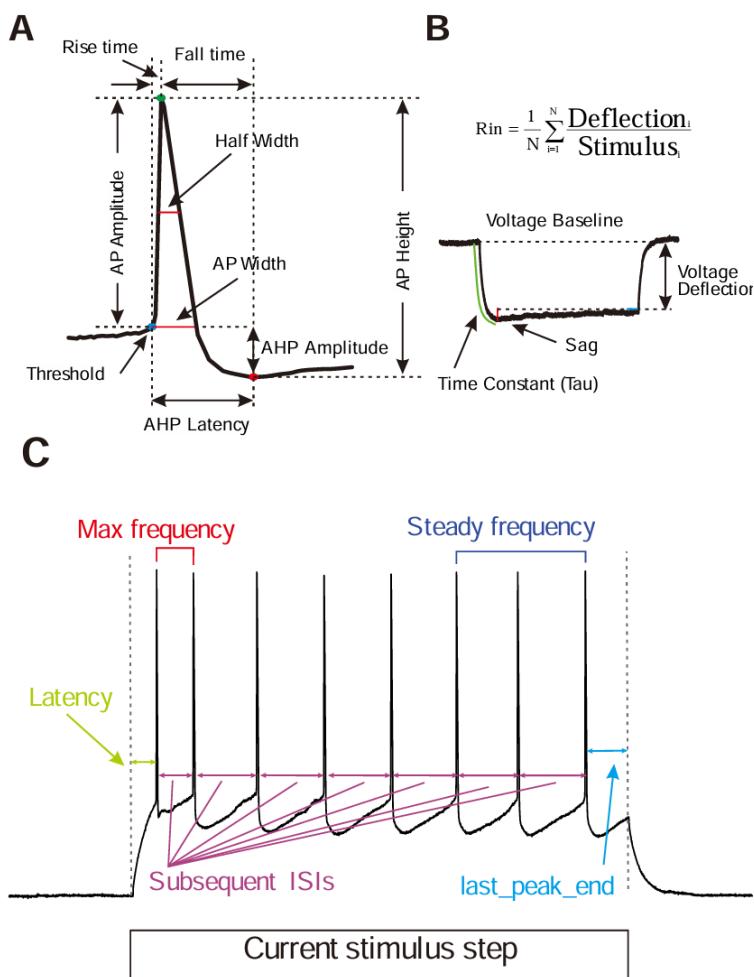


Figure 4-1 **Schematic Diagram of Main Electrophysiological Parameters.** **A:** Description parameters of action potential shape; **B:** Description parameters of electrical properties under hyperpolarizing stimulation; **C:** Description parameters of electrophysiological properties under field stimulation sweeps; ISI (Interspike interval): peak-to-peak distance of action potentials; AP (action potential): action potential; AHP (after-hyperpolarization potential): potential after hyperpolarization.

According to PING's description of the electrophysiological types of interneurons in 2008, the main classification of electrophysiology is divided into steady-state (steady state) fast spiking [FS], non-adapting, non-fast spiking [NA], adapting [AD], irregular

spiking [IS]. At the same time, according to the discharge characteristics of the transient state, it is divided into four discharge characteristics: burst [b], continuous [c], delayed [d], and stuttering [s]. The combination of discharge characteristics under the two states divides out a total of 10 main discharge types (Figure 2-1)^[9]. It can be seen that the main distinction of these discharge types is mainly caused by action potential firing induced by a relatively long domain stimulation, and the distinction description is based on the change between AP locations according to the discharge frequency.

In order to quantify based on this recognized electrophysiological nomenclature, in addition to parameters at the cellular level (such as tau, sag, and other common parameters, Fig. 2-1B), and parameters at the single action potential level (such as amplitude, threshold, and other common parameters, Fig. 2-1A), we focus on describing the variability between different AP under the same stimulus. We aim to differentiate electrophysiological types based on the different numerical ranges of quantified parameters and obtain a classification model with a specific value range. Therefore, we introduce the coefficient of variation (CV) to statistically describe the variability of interspike intervals (ISIs) under the same stimulus, as well as the variability of amplitude, threshold, AHP, up/down stroke, and corresponding CV values as quantified parameters for unsupervised clustering analysis. Table 4-1 provides a brief description of 30 electrophysiological parameters, with detailed parameter definitions and calculations provided in the Materials and Methods.

Feature Name	Feature Description
fi fit slope	Slope of f-i curve
Input Resistance (mΩ)	The slope of a line fit to subthreshold voltage responses to current input
Rheobase i (pA)	The minimum current input over a 1-s stimulus that generates an AP
Sag (mV)	The difference between the minimal voltage and the steady state at hyperpolarized state
tau (ms)	The membrane time constant
V baseline (mV)	The average voltage before the start of the stimulus
last ISI end (ms)	Time from the last peak to the end of the stimulus
adaption index	Normalized average difference of two consecutive ISIs
adaption	Difference of the max frequency and steady frequency divided by the max frequency
CV amp	The coefficient of variation of the amplitude
CV threshold	The coefficient of variation of the threshold
CV AP width	The coefficient of variation of the AP width
CV half width	The coefficient of variation of the half width
CV AHP latency	The coefficient of variation of the AHP latency
CV AHP amp	The coefficient of variation of the AHP amplitude
CV AP rise time	The coefficient of variation of the AP rise time
CV AP fall time	The coefficient of variation of the AP fall time
1st latency (ms)	Time from the start of the stimulus to the maximum of the first peak at first sweep
Amplitude (mV)	Difference of the peak and threshold voltage
Threshold (mV)	Voltage value at the onset of each action potential
AHP Amplitude (mV)	Difference voltage of the after-hyperpolarization and threshold
AHP Latency (ms)	Difference time of the after-hyperpolarization and threshold
AP Width (ms)	Width of each peak at the value of threshold
Half Width (ms)	Width of each peak at the value of half AP height
Rise Time (ms)	Time from action potential onset to the maximum
Fall Time (ms)	Time from action potential maximum to the offset
ISI cv	The coefficient of variation of the ISIs
First two ISI ratio	The ratio of the average ISIs and the average time of the first two ISI
delay ratio	The ratio of the average ISIs and latency at steady state response
Up down stroke ratio	The ratio of the voltage change rate during the falling phase and the rising phase

Table 4-1 Brief description of each feature.

We used the publicly available database Allen Cell Type, consisting of 569 transgenic mouse strains expressing different neuronal markers including Chat-IRES-Cre-

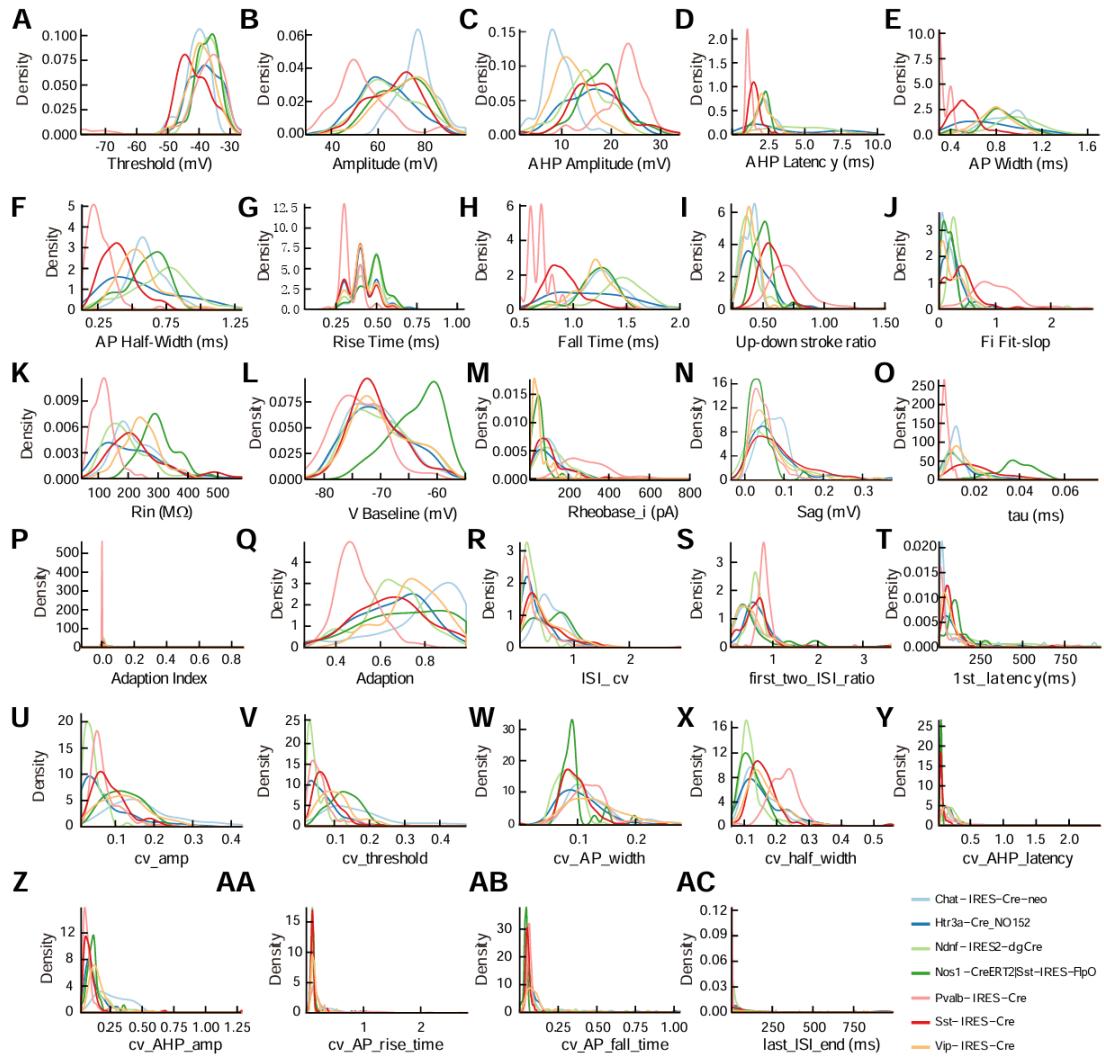


Figure 4-2 Kernel density distribution map of electrophysiological characteristic parameters.
A-AC: Presents the kernel density estimation values of 30 electrophysiological features of 659 intermediate neurons in the visual cortex according to different molecular types.

neo, Nos1-CreERT2|Sst-IRES-FlpO, Htr3a-Cre-NO152, Pvalb-IRES-Cre, Vip-IRES-Cre, Ndnf-IRES2-dgCre, and Sst-IRES-Cre, to extract electrophysiological data from visual cortex cells. Thirty electrophysiological parameters were extracted and respective kernel density distribution plots were generated to describe the distribution of each parameter among different types of interneurons (Figure 4-2). Additionally, ANOVA tests were performed on the parameters among interneurons with different molecular characteristics (Table 4-2). Except for three parameters: AHP Latency, Rise Time, and Up/Down Stroke Ratio, all other parameters exhibited significant differences among interneurons of different molecular types ($p<0.05$). This indicates that the extracted parameters are meaningful for classifying interneurons in this dataset. Subsequently, a $n \times m$ matrix ($n=569$, $m=30$) was constructed for multivariate analysis.

Feature Name	<i>p</i> value	Feature Name	<i>p</i> value
fi fit slope	7.29E-103	input resistance	2.24E-43
rheobase i	9.47E-77	sag	3.09E-08
tau	7.80E-78	v baseline	3.28E-17
last ISI end	8.46E-10	adaption index	2.78E-19
adaption	1.43E-41	cv amp	1.19E-31
cv threshold	9.50E-42	cv AP width	2.37E-17
cv half width	8.68E-39	cv AHP latency	5.79E-09
cv AHP amp	5.87E-23	cv AP rise time	5.19E-17
cv AP fall time	2.6E-09	1st latency	6.98E-08
Amplitude	1.06E-34	Threshold	0.000254
AHP Amplitude	2.27E-67	AHP Latency	ns
AP Width	6.30E-94	Half Width	1.32E-65
Rise Time	ns	Fall Time	1.85E-93
ISI cv	2.44E-09	first two ISI ratio	4.63E-09
delay ratio	2.02E-33	up down stroke ratio	ns

Table 4-2 Summary table of ANOVA test results among 7 molecular types. Analysis of Variance (ANOVA) tests the differences among 7 molecular types for each parameter. Time constant (τ); coefficient of variation (CV); afterhyperpolarization potential (AHP); action potential (AP); amplitude (amp); No significantly difference (ns) ($p > 0.05$). Features are provided in Table 4-1.

4.2 Further parameter filtering

4.2.1 PCA analysis

To analyze the weights of 30 parameters in this dataset, PCA analysis was performed, and a biplot visualization of the analysis results was drawn as shown in Figure 4-3. All parameters are mapped in the biplot (Figure 4-3A and 4-3C) and PC1/PC3 (Figure 4-3B and 4-3D) coordinates. In Figures 4-3A-B, the arrows represent the relationship between each parameter and the corresponding Principal Component (PC), with the length indicating the weight contribution of each parameter in this space, and the direction representing the relationship between parameters. Smaller angles between directions indicate stronger positive correlations, while opposite directions indicate stronger negative correlations, and perpendicular directions indicate no correlation between parameters. Figures 4-3C-D show the distribution of all cells in the space of two groups of PCs, with different colors representing different molecular types of interneurons. It is evident that interneurons labeled with Pvalb-Cre naturally cluster together, far from cells of other types, indicating significant differences in electrophysiological characteristics compared to other types. Furthermore, the electrophysiological characteristics of PV-type interneurons are relatively uniform compared to other types. From the direction of the arrows, PV-type interneurons have a larger slope of the f-i curve compared to other types, indicating larger changes in discharge frequency, higher threshold currents, larger hyperpolarized potentials, and larger amplitudes, consistent with previous reports [24,56].

Based on the overlap of confidence intervals represented by ellipses for different cell types, it can be seen that the electrophysiological characteristics of Htr3a-type interneurons are the most mixed, overlapping with all types except PV. Htr3a-type interneurons also express Vip, Ndnf, and Chat [22,84], hence the overlap with these three indicates potential co-expression types with Htr3a. It is also possible that due to the inherent

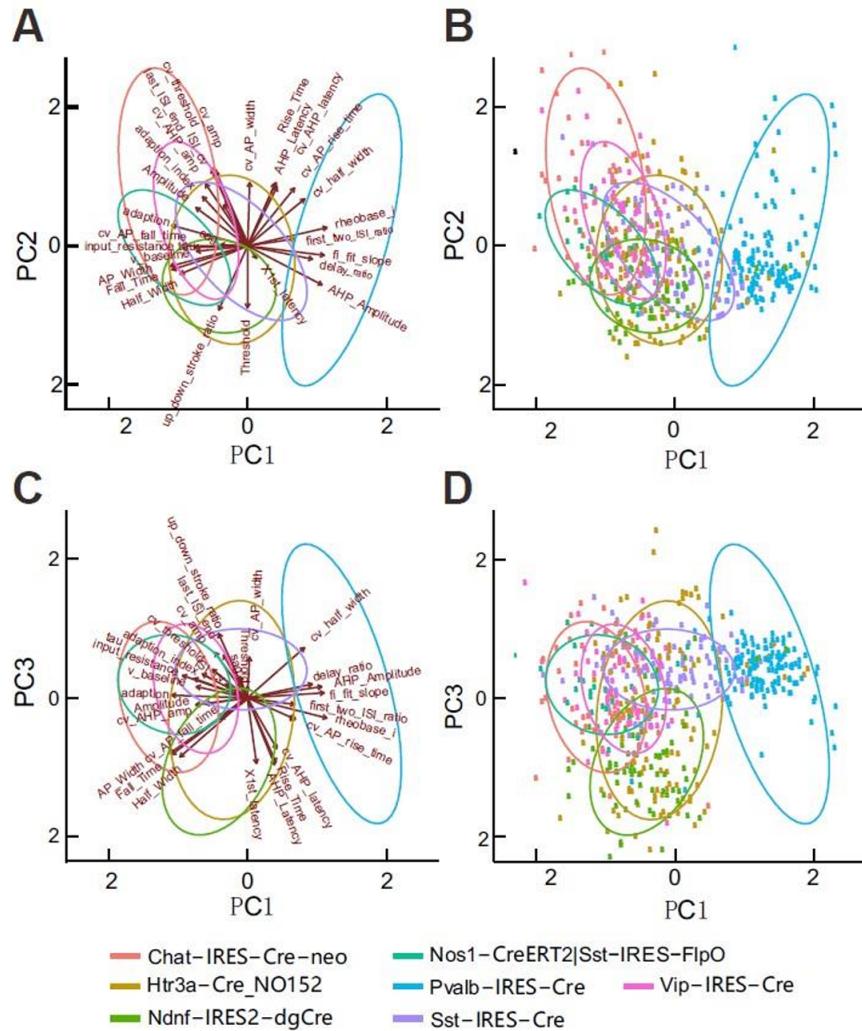


Figure 4-3 PCA Analysis Results. **A-B:** Represent the weights between parameters and the preferences between different molecular types in PC1/PC2 and PC1/PC3 spaces, respectively. The relationships between parameters and principal components are represented by spatial vectors, where the length indicates the weight and the direction represents the relationship between parameters. Different molecular types are depicted by multicolored ellipses showing the 95% confidence interval distribution. **C-D:** Represent the distribution of all cells in PC1/PC2 and PC1/PC3 spaces corresponding to A-B. The first three principal components account for 22.5%, 15.2%, and 11.3% of the variability, respectively. PC stands for Principal Component.

limitations of PCA analysis, high-dimensional data may not be fully represented in a two-dimensional space, leading to overlaps that are not indicative of actual data overlaps [54].

4.2.2 Parameter Correlation Analysis

Using a standardized analysis matrix, the correlation coefficients between parameters were calculated to analyze their relationships, and Figure 4-4 shows a visual heatmap. Red indicates a correlation coefficient close to 1, indicating a strong positive correlation between parameters. From the figure, it can be observed that there are 5 groups of parameters with high correlations (correlation coefficient >0.6). To minimize the correlation between parameters and reduce data dimensionality as much as possible, com-

bined with the PCA analysis results mentioned above, the parameters with the highest weights within these five groups were retained, and other parameters within the groups were removed. The retained parameters are highlighted with red arrows in Figure 4-4, leaving 24 parameters for further analysis.

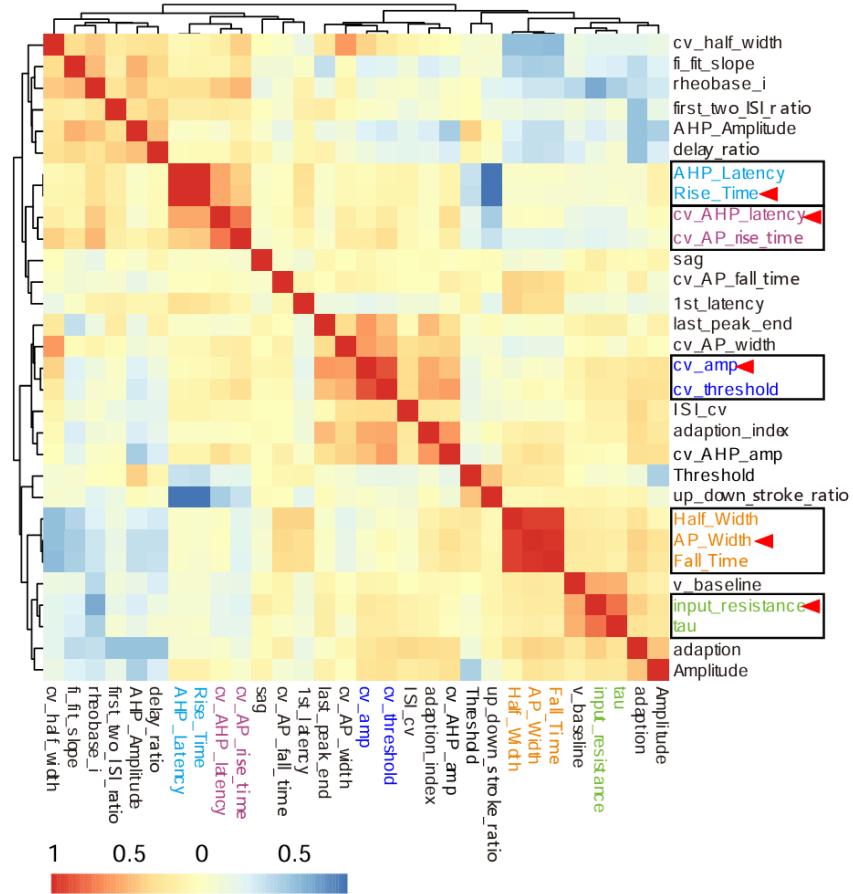


Figure 4-4 Hot map displays the correlation between various parameters. Values range from -1 to 1, where red indicates positive correlation and blue indicates negative correlation. Related parameter groups are represented by rectangles, and selected parameters are indicated by red arrows.

4.3 Ward's hierarchical clustering

First, use the Ward hierarchical clustering method to classify the electrophysiological types of interneurons. The relationship between different numbers of classifications, denoted as k , and height is shown in Figure 4-5A. A larger difference in height values indicates greater differences between the subtypes; height is measured in arbitrary units (a.u.), quantifying the distance between each type. The optimal pruning height is determined to ascertain the number of clusters, k , through optimal pruning testing [56,70]. The optimal pruning height is calculated to be 54.9 (a.u.). Figure 4-5B illustrates the results of the Ward's hierarchical clustering through a dendrogram, ultimately resulting in 7 clusters. PCA analysis is conducted on the data, and the distribution of interneurons classified into 7 clusters by the Ward's hierarchical clustering is shown using the first three PCs (Figure 4-6A-B). It is evident that the 7th cluster deviates significantly from the overall distribution, with only 3 cells, likely indicating outliers, suggesting that

the Ward's clustering method is highly sensitive to outliers, indicating relatively poor stability.

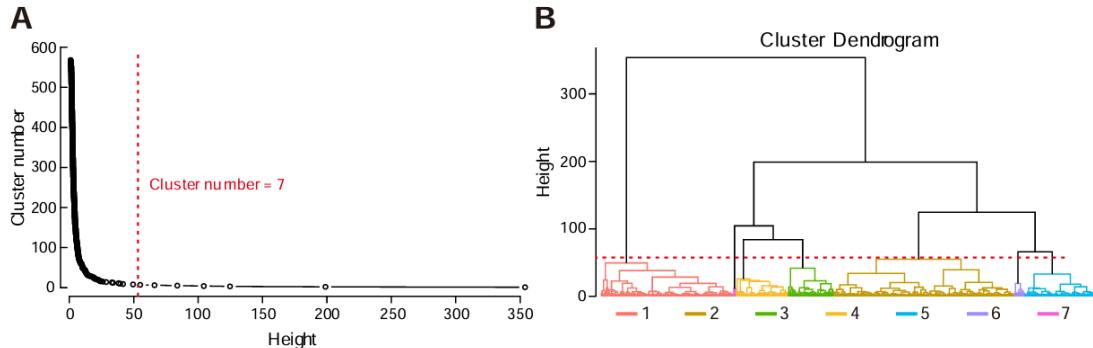


Figure 4-5 Situation of Ward hierarchical clustering. **A:** Relationship between the number of classifications and Height, the optimal pruning height value is calculated as 54.9 a.u. according to the optimal pruning test, marked with a red dashed line in the figure, and the optimal number of classifications is 7; **B:** Tree diagram of Ward hierarchical clustering results. The tree diagram is constructed using the height value (Height) of cell clusters merging. Branch points are formed by merging two smaller clusters into a new cluster. All types are merged at 353.8 (a.u.), and the red dashed line represents the optimal pruning height value of 54.9 a.u. Different colors represent seven different clusters obtained by clustering.

Observing the discharge patterns in different clusters (Figure 4-7), the first cluster comprises 152 cells, corresponding to the Fast Spiking (FS) discharge type (Figure 4-7A), with relatively homogeneous discharge patterns. The third cluster ($n = 60$) and the fifth cluster ($n = 81$) correspond to the Fast Adapting (fAD) discharge type and the Delay discharge type (delay), respectively (Figure 4-7C and E). Although the typical characteristic of the latter is delay, there are still variations in the degree of delay and discharge frequency within this type, suggesting that further subdivision may be possible.

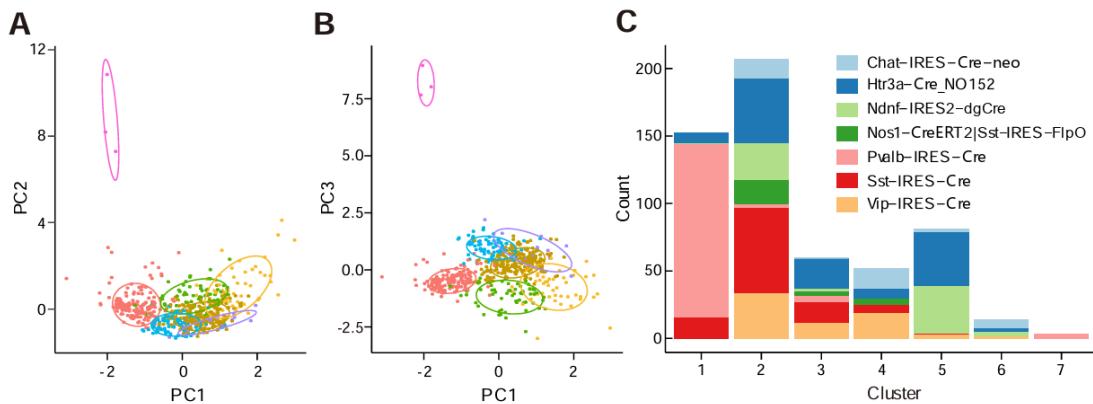


Figure 4-6 . **A-B:** Represents the distribution of intermediate neurons in each type obtained by Ward clustering in the PC1/PC2 and PC1/PC3 spaces respectively; **C:** Represents the number of subtypes corresponding to each type in the Ward hierarchical clustering.

The fourth cluster ($n = 52$) and the sixth cluster ($n = 14$) exhibit mixed discharge patterns similar to other clusters. The fourth cluster (Figure 4-7D) mainly comprises irregular spiking and stuttering, with the distribution of action potentials appearing relatively random. As reported, under the same stimulus, action potentials of irregular

spiking occur at different positions, appearing to be randomly distributed^[95]. Additionally, irregular and stuttering are considered a generalized form of irregular, differing only in the degree of ISI variation^[57]. Hence, it is possible that irregular and stuttering cannot be separated within the selected parameters or that the Ward clustering method is easily influenced by their randomness. Furthermore, the second cluster ($n = 207$) exhibits various discharge characteristics such as non-fast, non-adapting spiking, burst spiking, stuttering, and adapting (Figure 4-7B), without further subdivision by Ward clustering.

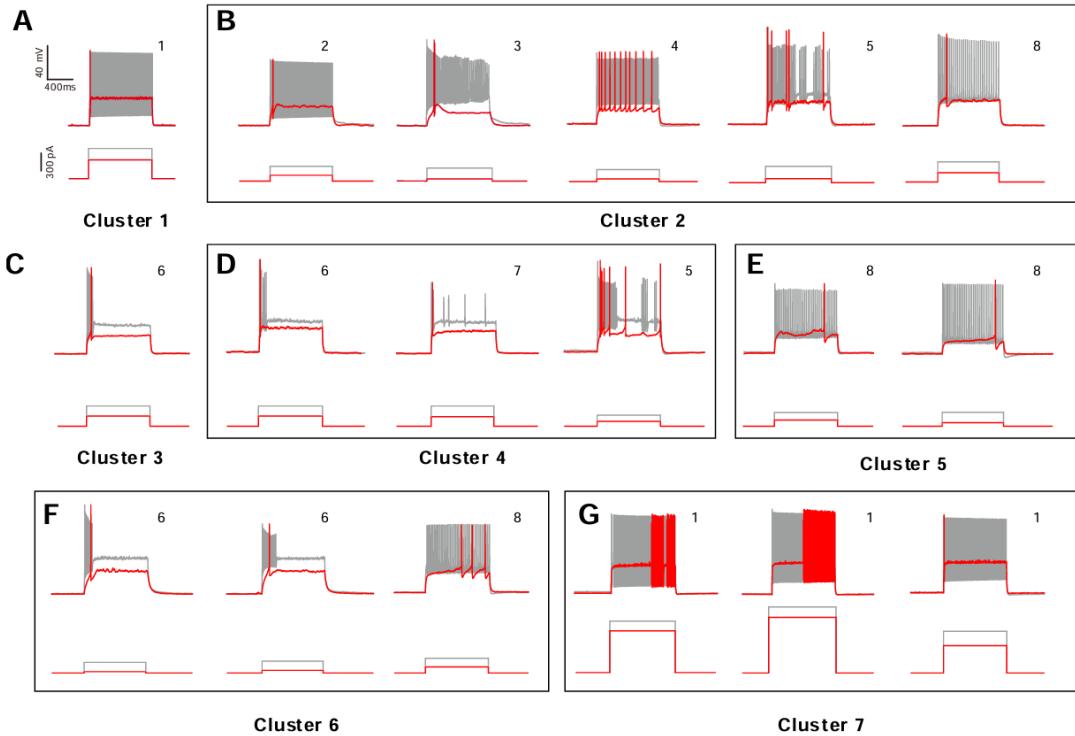


Figure 4-7 Classification of Ward hierarchical clustering yielded 7 types of electrophysiological discharge patterns. A-G: The upper figures show discharge patterns, while the lower figures represent 1-s depolarization current stimulation. Red indicates discharge patterns and stimulation obtained during basic intensity current stimulation, while gray indicates discharge patterns and stimulation when reaching a stable state. The discharge pattern data within the rectangles are from the same class; the numbers represent the corresponding types obtained from iPCA.

The seventh cluster ($n = 3$) (Figure 4-7G) clearly shows that all cells have relatively high rheobase currents, with two cells exhibiting typical delayed fast spiking (d-FS). The results suggest that the Ward clustering method is susceptible to the influence of several large parameter values, overlooking major characteristics, implying that transient state features may be masked by stable state features.

Furthermore, we investigated the correspondence between the types obtained by the Ward clustering method and molecular types (Figure 4-6C). We observed that 84.2% of cells in the first group and 100% in the seventh group are labeled as pvalb-cre interneurons, indicating that PV-type interneurons have relatively homogeneous electrophysiological types, belonging to FS. The fifth cluster mainly contains 5Htr3a and Ndnf cells (92.6%) (Figure 4-6C), suggesting that cells expressing Ndnf and 5Htr3a have similar discharge types, with a long delay in action potential firing after a few sweeps of AP generation, possibly indicating a molecular type expressing Reelin in the 5Htr3a type.

Other types exhibit highly mixed molecular characteristics.

4.4 iPCA

In current quantitative classifications, the k-means and Ward's methods are often used together for classification. There is also an evaluation that suggests the efficiency of k-means is higher compared to the Ward's method. However, due to the need to set the number of clusters (k) in k-means, and its lack of hierarchical structure, it is highly sensitive to parameters with strong correlations. Neurons themselves exhibit a tendency towards continuity, implying the existence of intermediate states. A hierarchical classification method would be more flexible in capturing information about relationships between different types of cells, which would be more advantageous for the study of interneuron classification. Therefore, we attempt to combine PCA and the Ward's clustering method iteratively. PCA is used to select subsets and assign weights to parameters, while the Ward's clustering method is employed for binary partitioning. Through iterative computation, we aim to construct a more stable classification system.

4.4.1 iPCA can visualize the intermediate classification process

Performing Iterative Principal Component Analysis (iPCA), we conduct PCA analysis on each sub-dataset, extracting the top n PCs that are greater than the mean plus two times the standard deviation. We then compute the distance matrix, followed by Ward's clustering analysis. The top two classes at the highest level are selected as the first-level subclasses. We iterate this process by applying PCA to the obtained subclasses, followed by Ward's clustering analysis, until further division is not possible ($n < 35$ or $p > 0.05$). During the classification process, a biplot can be visualized through PCA to illustrate the primary influential factors at different splitting nodes, as shown in Figure 4-8. Weightings on PC1 and PC2 reveal different subclass parameter weights. Therefore, a singular PCA analysis mapping data onto a two-dimensional coordinate axis does not adequately cluster the data. Enriching information into the most significant principal components helps extract primary information while avoiding interference from minor information. By iteratively combining Ward's clustering analysis, the entire classification system maintains hierarchical flexibility while enhancing stability.

4.4.2 iPCA can adjust parameter weights based on different subclasses

Due to the fact that iPCA conducts a PCA calculation in each iteration, extracting the top n PCs that are greater than the mean plus 2 times the standard deviation for computation, subclass information is enriched into the first few principal components through PCA. Different weighting is applied to each parameter at each split, as shown in Figure 4-9. Consequently, each hierarchical classification automatically undergoes more refined calculations, making the system more stable.

4.4.3 iPCA clustering 8 electrophysiological categories

iPCA ultimately divided into 8 classes, although some classes, due to limitations in data volume, were not further classified, but essentially each class obtained has unique electrophysiological characteristics.

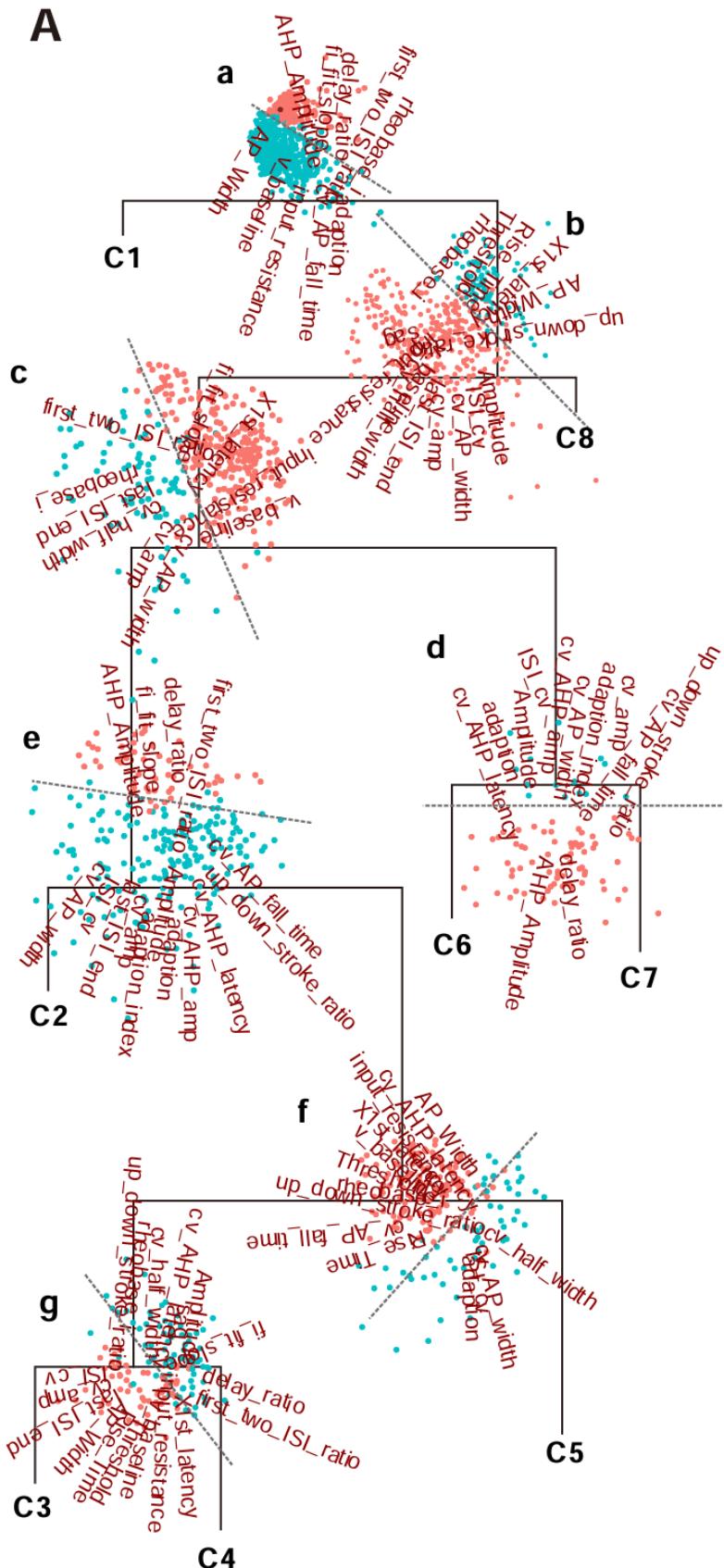


Figure 4-8 Visualization of the iPCA analysis process. Visualization of the iterative process of binary tree iteration. Lowercase letters represent each iteration split point corresponding to each biplot in PC1/PC2 space. Red scatter points represent subclasses in the left branch, while blue scatter points represent the right branch. Major effective parameters are displayed in each biplot. The weight values for each iteration are detailed in Figure 4-9. Leaf nodes C1-C8 represent the final classified classes.

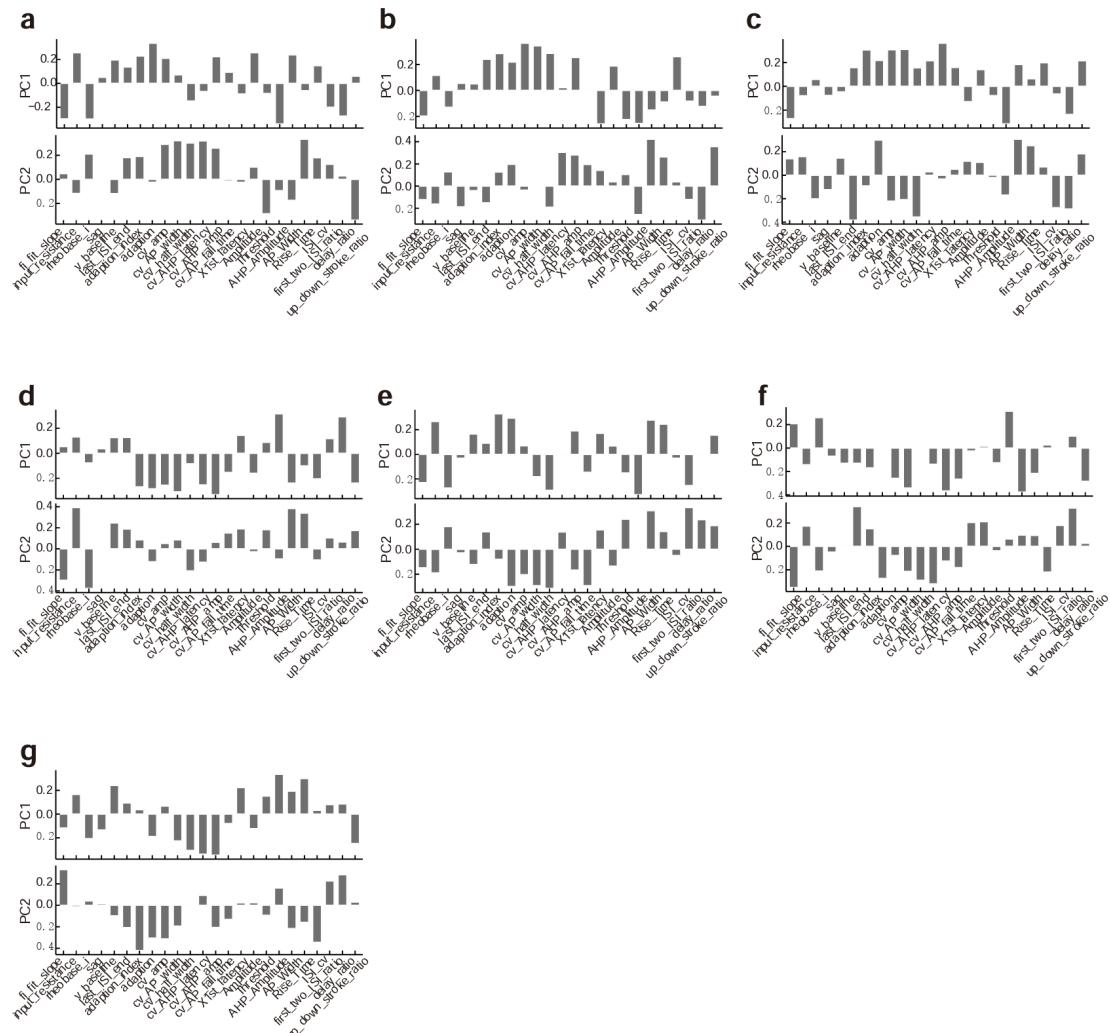


Figure 4-9 The different weights of parameters on PC1/PC2 in each iteration. At each iteration split point, the different weights of each feature parameter on PC1/PC2 correspond to the lowercase letter labels in Figure 4-8.

Class 1 ($n = 146$) is similar to FS type (Figure 4-10A). Overall, compared to other groups, neurons in Class 1 have significantly higher I-F curve slopes, AHP amplitude, and Rheobase i (Anova and Turkey tests, $p < 0.05$) (Tables 4-3 and 4-4). However, some d-FS and stuttering fast spiking (s-FS) within Class 1 suggest that if there were more samples in Group 1, they could be further subdivided in transient states.

Class 2 ($n = 50$) is similar to NA, with a frequency lower than FS, and no adapting (Figure 4-10B). Electrophysiological characteristics of Class 2 show significantly lower adaption index values compared to Classes 3-8, with adaption values lower than Classes 3, 4, 5, and 7 (ANOVA followed by post hoc Tukey tests, $p < 0.05$; Tables 4-3 and 4-4).

Additionally, the slope of the I-F curve in Class 2 was significantly higher than Classes 3-8, but lower than Class 1 (ANOVA followed by post hoc Tukey tests, $p < 0.05$; Tables 4-3 and 4-4). Meanwhile, the amplitude of the AHP in Class 2 was significantly higher than Classes 3-8 (Tables 4-3 and 4-4). However, Class 2 also exhibited different delay and burst characteristics compared to Classes 3 and 8, albeit to a lesser extent (Figure 4-10B).

Class 3 ($n = 59$) exhibited a typical irregular spike discharge pattern (Figure 4-10C).

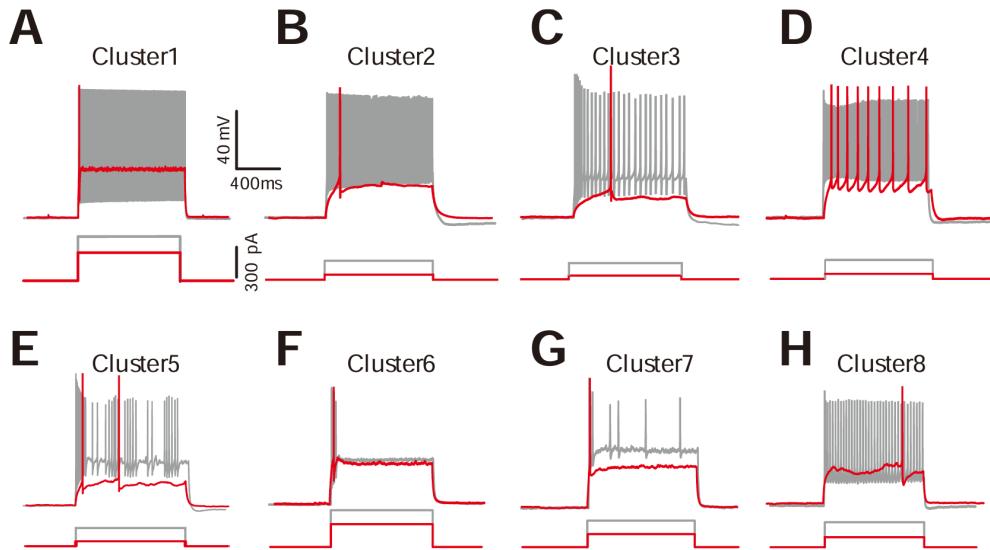


Figure 4-10 iPCA is divided into representative figures of various electrophysiological types. The upper figure represents the discharge mode, and the lower figure represents 1s depolarization current stimulation. Red indicates discharge mode and current stimulation under baseline current intensity stimulation, while gray indicates stable-state discharge mode and stimulation current.

The first two ISI ratio values in Class 3 were significantly lower than Classes 1, 2, 4, 5, 6, and 8 (ANOVA followed by post hoc Tukey tests, $p < 0.05$; Tables 4-3 and 4-4). Furthermore, the coefficient of variation (ISI cv) of ISIs in Class 3 was significantly higher than Classes 1, 2, 4, and 8, but lower than Classes 5 and 7 (ANOVA followed by post hoc Tukey tests, $p < 0.05$; Tables 4-3 and 4-4). Simultaneously, the coefficient of variation (cv amp) of amplitudes in Class 3 was significantly higher than Classes 1, 2, 4, and 8 (ANOVA followed by post hoc Tukey tests, $p < 0.05$; Tables 4-3 and 4-4). Class 3 also exhibited varying degrees of bursts.

Class 4 ($n = 75$) resembled an adapting discharge type (Figure 4-10D). The adaptation index value in Class 4 was significantly higher than Classes 5, 6, and 7, and the adaptation value was significantly higher than Classes 1, 2, 3, 5, 6, and 8 (ANOVA followed by post hoc Tukey tests, $p < 0.05$; Tables 4-3 and 4-4).

Class 5 ($n = 58$) exhibited characteristics similar to stuttering discharge type (Figure 4-10E). The ISI CV in Class 5 was significantly higher than Classes 1, 2, 3, 4, 5, 6, and 8, and the CV amp was significantly higher than Classes 1, 2, 4, 6, 7, and 8 (ANOVA followed by post hoc Tukey tests, $p < 0.05$; Tables 4-3 and 4-4).

Class 6 ($n = 72$) and Class 7 ($n = 14$) both exhibited features of irregular spiking (Figure 4-10F-G). The slope of the I-F curve in these two classes was significantly lower than in other classes, and the last ISI end and CV amp were higher than other classes (ANOVA followed by post hoc Tukey tests, $p < 0.05$; Tables 4-3 and 4-4). However, compared to Class 6, the CV amp in Class 7 was significantly higher, and the last ISI end was smaller (ANOVA followed by post hoc Tukey tests, $p < 0.05$; Tables 4-3 and 4-4). Most cells in Class 6 were also considered as fAD in previous studies. This class may further be subdivided into fAD and a type of irregular spiking.

Class 8 ($n = 95$) exhibited typical delay characteristics (Figure 4-10H). The 1st latency value in Class 8 neurons was significantly greater than other clusters (ANOVA followed by post hoc Tukey tests, $p < 0.05$; Tables 4-3 and 4-4). Class 8 also exhibited varying degrees of delay that could be further subdivided. Therefore, some cells with

lower degrees of delay were grouped into Class 2 of the Ward clustering (Figure 4-7B).

4.4.4 iPCA is able to flexibly integrate classification information from other aspects

Due to the complex characteristics of interneurons in morphology, electrophysiology, and molecular aspects, the classification of intermediate neurons requires the integration of multi-dimensional information. We attempted to integrate molecular and laminar distribution information during the classification process through iPCA, as shown in Figure 4-11. In terms of the distribution of molecular types in each class, we observed that 87% of the first class cells and 72% of the second class cells were respectively PV neurons and SST neurons (Figure 4-11A). The eighth class mainly included 5Htr3a and Ndnp molecular types, while other classes had a more mixed molecular background (Figure 4-11A). Figure 4-11B shows the laminar distribution of each type, and we found that each electrophysiological type is widely distributed across cortical layers. In summary, iPCA can flexibly integrate various aspects of information and effectively demonstrate the information flow throughout the entire classification process, making it suitable for the classification needs of interneurons.

4.5 Comparing Ward's hierarchical clustering with iPCA analysis results

We further compared the clustering results of iPCA and Ward clustering methods (Figure 4-12). We found that in the first class of iPCA, 90.1% of neurons (137 out of 152 neurons) corresponded to the first class of Ward's clustering; the first class of Ward's clustering corresponded to 93.8% (137 out of 146 neurons) of the first class of iPCA. Similarly, 90.1% of cells in the eighth class of iPCA (73 out of 81) were consistent with the fifth class of cells in Ward's clustering, while 76.8% of cells in the fifth class of Ward's clustering (73 out of 95) were consistent with the eighth class of iPCA. At the same time, 90% of neurons in the sixth group of iPCA (54 out of 60) and 75% of neurons in the third class of Ward's clustering (54 out of 72) matched. These results indicate that the results of the two methods are basically consistent. However, classes 2, 3, 4, and 5 in iPCA are merged into one class in the results of the Ward's clustering method (class 2, Figure 4-12).

We examined the iPCA classes 2, 3, 4, and 5 that were merged in Ward's clustering class 2, and all parameters showed significant differences (ANOVA test $p < 0.05$, Table 4-5). In addition, the fourth class of cells in Ward's clustering exhibited highly consistent discharge characteristics. These results suggest that the iPCA method is more suitable for the classification of electrophysiological characteristics of neurons in the cortical interstitium compared to the Ward's method.

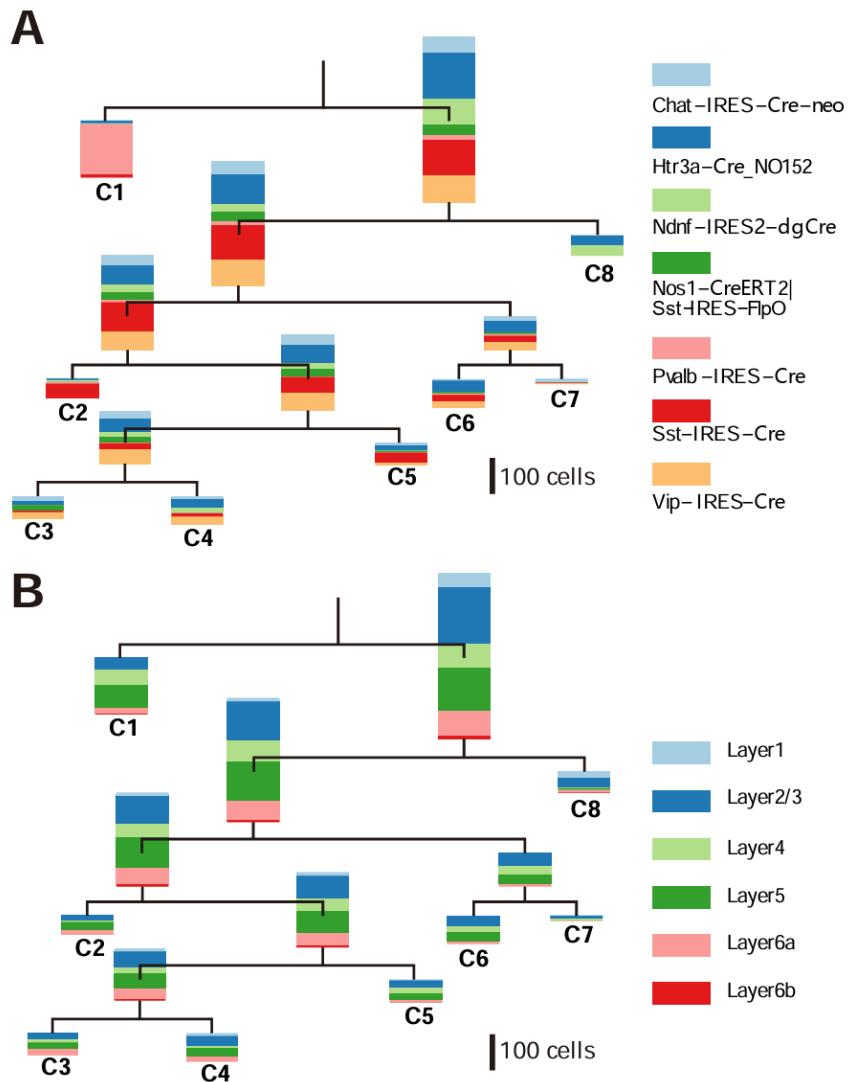


Figure 4-11 Molecular type and layer distribution binary tree. **A:** Represents the molecular type of subclasses in each iteration; **B:** Represents the layer distribution of subclasses in each iteration.

		Cluster								<i>p</i> value
		1	2	3	4	5	6	7	8	
fit fit slope	0.9244±0.0311	0.4987±0.0235	0.2497±0.0195	0.3431±0.022	0.2127±0.0239	0.0706±0.0099	0.0399±0.0089	0.2441±0.0157	<2e-16	
input resistance	11.3.8281±3.1741	241.7538±11.6895	292.771±11.3979	218.966±8.3685	226.0306±10.4713	232.5605±10.1471	192.3196±18.6116	144.2633±5.1967	<2e-16	
rheobase.i	268.7413±8.5404	80.6±5.6814	48.0508±3.1798	85.8667±6.931	101.2069±8.8787	96.3889±8.3739	126.4286±23.9873	162±7.8016	<2e-16	
sag	0.0494±0.0032	0.1129±0.0099	0.0569±0.0059	0.0714±0.0056	0.0704±0.0069	0.0799±0.0091	0.0626±0.0071	0.0451±0.0044	1.8e-13	
tau	0.0079±0.0003	0.02±0.0012	0.0254±0.0017	0.0149±0.0009	0.019±0.0012	0.0194±0.0015	0.0117±0.0008	0.0111±0.0005	<2e-16	
v baseline	-73.7508±0.3571	-69.4628±0.6579	-65.6425±0.6174	-70.6121±0.5528	-70.5638±0.621	-68.9236±0.621	-73.2523±1.1838	-71.8772±0.5414	<2e-16	
last ISI end	34.3.147±12.3656	15.596±4.4935	55.7356±12.4311	21.4213±4.8337	160.3.172±24.6142	782.6847±33.1484	562.6429±87.2051	17.3937±1.3428	<2e-16	
adaption index	0.0021±0.001	0.0053±0.0004	0.0241±0.0027	0.0166±0.0013	0.0812±0.0134	0.0721±0.0083	0.3248±0.0702	0.0129±0.0011	<2e-16	
adaption	0.4825±0.007	0.5832±0.0118	0.761±0.0149	0.77±0.0105	0.8335±0.0146	0.5878±0.0179	0.8489±0.0355	0.621±0.0135	<2e-16	
cv amp	0.0679±0.0028	0.0553±0.0032	0.1047±0.0043	0.0611±0.0039	0.1246±0.0063	0.1598±0.0065	0.2622±0.0182	0.0349±0.0026	<2e-16	
cv threshold	0.0611±0.0033	0.055±0.0028	0.0995±0.0039	0.0639±0.0036	0.11±0.0077	0.121±0.0051	0.2632±0.0264	0.0394±0.0027	<2e-16	
cv AP width	0.1161±0.0026	0.0858±0.0022	0.1001±0.0032	0.0973±0.003	0.1184±0.004	0.1356±0.005	0.2057±0.0144	0.0777±0.0019	<2e-16	
cv half width	0.2247±0.0036	0.1414±0.0043	0.1213±0.003	0.1301±0.0035	0.1716±0.0046	0.2141±0.0093	0.2245±0.0156	0.1071±0.0019	<2e-16	
cv AHP latency	0.2037±0.0246	0.0739±0.0065	0.1019±0.0085	0.217±0.013	0.0987±0.008	0.1308±0.0088	0.2786±0.0505	0.2159±0.0095	1.5e-08	
cv AHP amp	0.0745±0.0056	0.0575±0.0023	0.1217±0.0043	0.1435±0.0082	0.1465±0.0154	0.1566±0.0111	0.4411±0.0725	0.1333±0.0085	<2e-16	
cv AP rise time	0.3483±0.0366	0.1118±0.0081	0.1168±0.0045	0.1244±0.01	0.1464±0.0145	0.1418±0.013	0.3197±0.0831	0.1207±0.013	<2e-16	
cv AP fall time	0.0612±0.0011	0.0494±0.0016	0.1176±0.0178	0.0788±0.0055	0.066±0.0052	0.0877±0.0077	0.2138±0.0708	0.1371±0.0213	4.3e-10	
1st latency	84.2923±14.17	103.2545±5.7761	109.6271±14.4106	62.6933±5.0298	48.5586±4.4051	49.0431±4.0948	23.4429±2.4065	314.5095±26.0252	<2e-16	
Amplitude	52.4729±0.913	63.6069±1.4144	68.7813±1.5335	73.4279±1.2438	71.875±1.2048	65.4067±1.0751	77.3013±2.7774	61.0503±1.1529	<2e-16	
Threshold	-37.939±0.6472	-38.8225±0.5751	-38.5885±0.512	-40.8242±0.4474	-42.5016±0.5447	-39.5694±0.4679	-41.5603±1.281	-35.4717±0.367	4.7e-11	
AHP Amplitude	23.1215±0.3205	20.2275±0.5827	13.2913±0.4976	11.7908±0.4542	13.7742±0.5993	14.9674±0.5025	8.8839±0.8678	16.3322±0.5321	<2e-16	
AHP Latency	1.4231±0.1239	1.82±0.1369	2.5017±0.0951	2.6493±0.1688	1.6931±0.0684	1.8611±0.0864	2.3786±0.1688	6.6537±0.2476	ns	
AP Width	0.3388±0.0047	0.544±0.0183	0.8729±0.0216	0.7547±0.0155	0.6±0.0266	0.6069±0.0201	0.8571±0.0416	1.00653±0.0201	<2e-16	
Half Width	0.2613±0.0063	0.4024±0.017	0.6269±0.0185	0.5247±0.0148	0.4348±0.0181	0.433±0.0181	0.5886±0.041	0.8315±0.0192	<2e-16	
Rise Time	0.6084±0.1244	0.394±0.0092	0.4695±0.0085	0.4173±0.0092	0.4121±0.0107	0.3903±0.0082	0.4286±0.0163	0.4789±0.0082	ns	
Fall Time	0.6727±0.0074	0.918±0.0255	1.2729±0.0229	1.1773±0.0178	0.969±0.0284	0.9819±0.0247	1.2714±0.0507	1.5284±0.0194	<2e-16	
ISI cv	0.3039±0.0303	0.2438±0.0277	0.5454±0.0386	0.3691±0.0244	0.7817±0.0426	0.3946±0.0341	0.6697±0.0479	0.1843±0.0132	<2e-16	
first two ISI ratio	0.8605±0.0256	0.6759±0.0149	0.4211±0.0223	0.4783±0.0249	0.3017±0.0221	0.9159±0.0643	0.686±0.1496	0.6887±0.0206	<2e-16	
delay ratio	0.7422±0.0235	0.6605±0.0304	0.3273±0.0209	0.354±0.0208	0.2903±0.0275	0.6542±0.037	0.1365±0.0274	0.4235±0.0167	<2e-16	
up down stroke ratio	1.1024±0.1682	0.5698±0.0147	0.4344±0.0096	0.4037±0.0086	0.5157±0.0151	0.4951±0.0131	0.3724±0.0147	0.3589±0.0061	ns	

Table 4-3 All parameters of clusters in iPCA.

Feature Name	Tukey test
fi fit slope	1 vs 2-8; 2 vs 3-8; 3 vs 6; 4 vs 5,6,7; 5 vs 6; 8 vs 6,7
Input Resistance (mΩ)	1 vs 2-8; 2 vs 3,8; 3 vs 4-8; 4 vs 8; 5,6 vs 8
Rheobase i (pA)	1 vs 2-8; 2 vs 8; 3 vs 5-8; 4,5,6 vs 8
Sag (mV)	1 vs 2,6; 2 vs 3-8; 4,6 vs 8
tau (ms)	1 vs 2-6; 2 vs 3,4,7,8; 3 vs 4-8; 4 vs 6; 5 vs 7,8; 6 vs 7,8
V baseline (mV)	1 vs 2-6,8; 3 vs 4-8; 6 vs 7,8
last ISI end (ms)	(1-4,8) vs (5-7); 5 vs 6,7; 6 vs 7;
adaption index	(1-4,8) vs (5-7); (5,6) vs 7;
adaption	1 vs 2-8; 2 vs 3-5,7; (3,4) vs 5,6,8; 5 vs 6,8; 6 vs 7; 7 vs 8
CV amp	(1,2,4) vs (3,5),6-8; 3 vs 8; 6 vs 7,8; 7 vs 8
CV threshold	(1,2,4) vs (3,5),6-8; 6 vs 7,8; 7 vs 8
CV AP width	(1,5) vs (2,3,4),6-8; (2,3,4) vs 6-8; 6 vs 7,8; 7 vs 8
CV half width	(1,6,7) vs (2,3,4),5,8; 5 vs 8
CV AHP latency	1 vs 2,3,5,6; 2 vs 2,4,7,8; 3 vs 4,8; 4 vs 5; 5 vs 8
CV AHP amp	(1,2) vs (3-6,8),7; (3-6,8) vs 7
CV AP rise time	1 vs (2-6,8)
CV AP fall time	(1,2) vs 3,(7,8); (4-6) vs (7,8)
1st latency (ms)	8 vs (1-7)
Amplitude (mV)	1 vs (3,4,5,7) ,(2,8); (2,8) vs (4,5,7)
Threshold (mV)	(1-3),(4-7) vs 8, 5 vs (1-3);
AHP Amplitude (mV)	(1,2) vs 3-8; (3,-6) vs 7,8; 7 vs 8
AHP Latency (ms)	ns
AP Width (ms)	1 vs 2-8; (2,5,6) vs 3,4,7,8; 3 vs 4,8; 4 vs 8; 7 vs 8
Half Width (ms)	1 vs 2-8; (2,5,6) vs 3,4,7,8; 3 vs 4,8; 4 vs 8; 7 vs 8
Rise Time (ms)	ns
Fall Time (ms)	1 vs 2-8; (2,5,6) vs 3,4,7,8; 3 vs 4,8; 4 vs 8; 7 vs 8
ISI cv	1 vs 3,5,7,8; 2 vs 3,5,7; 3 vs 4,5,8; 4 vs 5,7,8; 5 vs 6,8; 6 vs 7,8; 7 vs 8
First two ISI ratio	1 vs 2-5,8; 2 vs 3-6; 3 vs 6,8; 4 vs 5,6,8; 5 vs 6-8; 6 vs 8
delay ratio	1 vs 3-5,7,8; 2 vs 3-5,7,8 3 vs 6; 4 vs 6,7; 5 vs 6,8; 6 vs 7, 8; 7 vs 8
Up down stroke ratio	ns

Table 4-4 All parameters Tukey test between iPCA clusters.

Feature Name	p Value		
fi fit slope	8.2e-08	Input Resistance (mΩ)	1.2e-06
Rheobase i (pA)	6.8e-07	Sag (mV)	8.1e-08
tau (ms)	6.6e-10	V baseline (mV)	5.5e-08
adaption index	5.9e-08	adaption	<2.2e-16
CV amp	8.9e-15	CV AP width	0.00017
CV half width	2e-12	CV AHP latency	2.2e-16
CV AHP amp	8.6e-10	CV AP fall time	0.00052
1st latency (ms)	1.3e-07	Amplitude (mV)	0.0017
Threshold (mV)	4.1e-06	AHP Amplitude (mV)	< 2.2e-16
AP Width (ms)	< 2.2e-16	Rise Time (ms)	9.3e-06
ISI cv	4e-12	First two ISI ratio	1.4e-11
Delay ratio	9.9e-12	up down stroke ratio	< 2.2e-16

Table 4-5 All parameters ANOVA test between iPCA cluster 2,3,4,5. ANOVA test determined significant differences in parameters of iPCA among classes 2, 3, 4, and 5. Time constant τ; coefficient of variation CV; AHP, afterhyperpolarization potential; AP, action potential; amp, amplitude; ns, no statistically significant difference ($p>0.05$). Units are described in Table 4-1.

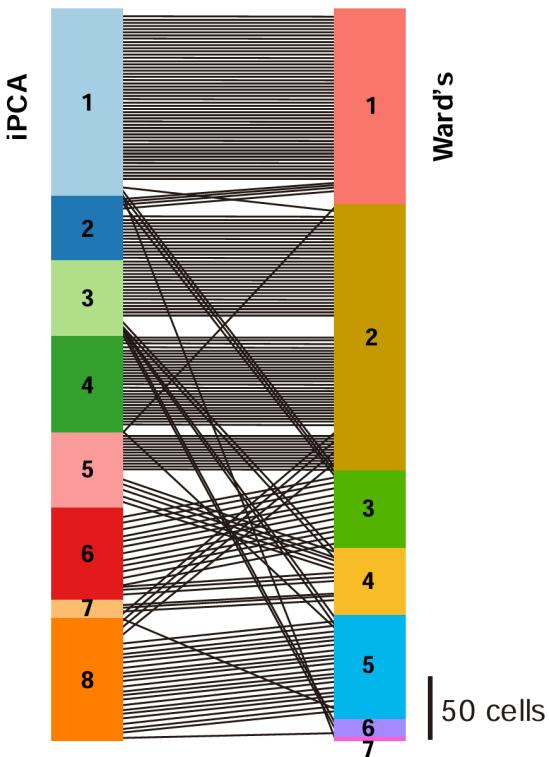


Figure 4-12 Comparison of iPCA Method and Ward's Method. Distribution plots of iPCA and Ward's methods. The bars on the left represent the classification results of iPCA, while the bars on the right represent the classification results of Ward. The solid lines match the classification results of iPCA and Ward's clustering. The probability of complete match between iPCA and Ward clustering is 58%. In iPCA, 94.5% of cells in Class 1 correspond to cells in Class 1 of Ward's clustering; 75% of cells in Class 6 of iPCA correspond to cells in Class 3 of Ward's clustering; 76.8% of cells in Class 8 of iPCA correspond to cells in Class 5 of Ward's clustering. Additionally, 76.9% of cells in Class 2-5 of iPCA belong to Class 2 of Ward's clustering. Clusters 4, 6, and 7 of Ward's clustering do not correspond to iPCA.

Chapter 5

Disscusion

The interneurons in the mammalian cerebral cortex exhibit a variety of electrophysiological characteristics. The division of these neurons into different electrophysiological types provides a useful abstract concept that can facilitate the study of the interactions of neurons in cortical circuits^[1–3,6–7]. In this study, we conducted a classification study of mouse visual cortex neurons using unsupervised classification methods based on electrophysiological characteristics.

By selecting appropriate quantification parameters and using two unsupervised classification analysis methods, we ultimately obtained 7 classes using the Ward's clustering method and 8 classes using the iPCA method. Through comparison of the two methods, we believe that iPCA is more suitable for large datasets, yielding stable classification results. The electrophysiological types identified in our classification under stable conditions can be quantitatively described, while those under transient conditions can be roughly distinguished. This indicates that the system naming proposed by the PING conference can be quantified through appropriate parameters and the iPCA method, with the potential for further refinement.

Our final classification results did not include the d-FS type of electrophysiology, primarily due to limitations in dataset size. Establishing a stable type is believed to require more than 10 neurons; hence, we set a splitting limit condition of more than 10 cells per type. Thus, Class 1 did not proceed to further computation to generate subclasses. However, upon examining the corresponding discharge waveforms, we found that there were a few cells resembling the d-FS type, although in relatively small numbers. This also indicates that different electrophysiological types, occupying a larger number of cells, are separated first, and the features of stable state sweeps are more stable and easier to quantify.

Additionally, due to the last ISI end parameter set to describe the time from the last AP to the end of the stimulus, we obtained Class 6, where very few APs were fired, concentrated at the beginning of the stimulus. Because some data in this type also had one or two APs fired during the stimulus, it bore some resemblance to the irregular type of Class 7. We believe it may be a special type of irregular. According to existing literature, some studies describe types that do not fire APs for more than 600ms before the end of the stimulus as fast adapting (fAD); it is possible that Class 6 includes fAD and a subtype of irregular. Given a sufficient amount of data, further differentiation should be feasible.

5.1 Quantitative parameters

Our study quantifies the system naming proposed by the PING conference, yielding corresponding classification results, indicating that selecting quantification parameters

based on dataset characteristics can produce corresponding results. However, such results are still somewhat crude. To achieve finer divisions, besides requiring more experimental data, more precise and comprehensive quantification parameters are also needed. The classification of intermediate neuron electrophysiology faces a major challenge in selecting quantification parameters because the raw data of electrophysiology are multi-layered and high-dimensional. We need to statistically describe them and reduce data dimensions through statistical methods. Quantification descriptions have two main requirements: 1) Extract as much information as possible, and 2) Reduce dimensions as much as possible. Clearly, these two requirements are contradictory, as more information inevitably leads to higher dimensions. Thus, different datasets require different quantification studies to balance these two aspects.

Currently, many laboratories have done extensive work on quantitative studies of electrophysiological features, primarily focusing on adaptation, burst, stuttering, and irregular types. Especially for stuttering and irregular types, which exhibit significant variation^[57,95], there is much debate. The same cell can also produce different waveforms under the same size of current stimulation^[95]. Stuttering types also exhibit certain variations and are considered a special type of irregular type^[57]. Many subclasses within the irregular type are difficult to unify^[8–9,59–60,81,84]. Our results also indicate that the stuttering and irregular types are closely related, as Class 5 also includes some cells with frequencies similar to those of irregular types in Classes 6 and 7. Thus, further research should focus on the quantitative study of irregular types and select appropriate quantification descriptions to integrate into the classification system to optimize the classification of irregular types.

We conducted our study iteratively, computing PCA multiple times to change the parameter weights of each subclass to adapt to larger datasets. However, this only biases the parameters, and when there are many parameters, information may be dispersed among PCs, making it difficult to concentrate all the information onto a few PCs. The ideal approach would be to actively select parameters for each subset, such as binding some parameters to types.

5.2 Hierarchical Unsupervised Classification

One of the most important questions in interneuron classification is "How finely should we classify?" Whether in species classification or cell classification, our goal is the same: to better understand and study. Thus, the "fineness" depends on the resolution we wish to achieve in our study. We can choose different scales based on different research purposes while still unifying within a standard. Classifying through a hierarchical classification system and selecting different levels of classification according to research purposes is more meaningful. iPCA optimizes such a hierarchical structure. We found through comparison between the Ward's clustering method and iPCA. iPCA is more robust and flexible, outperforming the Ward's clustering method. iPCA provides a hierarchical structure with more information, effectively demonstrating the relationships between different types. Our classification results indicate that the FS type differs significantly from other types on multiple parameters and is separated first. Following are the delay type, irregular and NA types, burst, adapting, and stuttering types, which are closely related. As shown in Figure 5-1, the hierarchical relationship of the classification results can be observed, where the AD and burst types are closely related, as are the fAD and IS types, while FS differs the most from other types.

We also found that types with a larger number of cells and stable features are easier to separate, while those with fewer cells are easily obscured. Further divisions would require more precise parameters, which would affect the classification of types with stable features. Systems like iPCA, with hierarchical structures, can optimize this issue by selectively choosing parameters at different levels. By selecting different parameters for different nodes at different levels, a more refined classification can be achieved to adapt to larger datasets.

5.3 Supervised Classification

This study employed unsupervised classification methods for classification. Unsupervised classification naturally clusters data based on its characteristics. However, unsupervised classification relies solely on computational data and makes it difficult to detect classification errors if there is significant noise or instability in the classification system. Supervised classification can set standards based on known knowledge, continuously train and calibrate the classification system, and combine manual classification with automatic classification, making the classification process more controllable^[54]. In single-cell sequencing, for example, combining Random Forest with iPCA for classification, using iPCA results to train Random Forest classifiers, and then validating iPCA results with trained Random Forest classifiers allow for mutual calibration and validation between supervised and unsupervised classifications, constructing a more robust analysis system^[22].

In summary, whether supervised or unsupervised, in the face of a complex group like interneurons, building a stable system is crucial. Single analytical methods should be used to construct an analysis hierarchy, making the analysis system more stable.

5.4 Electrophysiological type and molecular type

In the classification process of iPCA, we integrate the molecular types and laminar distribution information of cells. Currently, there is no significant difference in classification on the layers. PV cells are mainly FS, Ndnf (reelin) preferring to have a larger latency, Sst cells are more mixed than PV but purer than 5Htr3a. The main complex discharge characteristics of 5Htr3a correspond to the type of Vip. Mainly corresponding to burst, AD, and irregular/fAD, which are basically consistent with the results of current reports^[5,59,84]. This indicates that iPCA can integrate well with other characteristics of interneurons and can be extended to the classification research of other characteristics.

This study quantitatively analyzed the electrophysiological data of interneurons and classified interneurons using clustering methods, preliminarily constructing a quantitative analysis system for interneurons. This lays the foundation for future application of multilevel data integration such as electrophysiology, morphology, and molecular biology of interneurons, avoiding artificial interference. It is expected to compare and integrate the classification of interneuron data published by different laboratories.

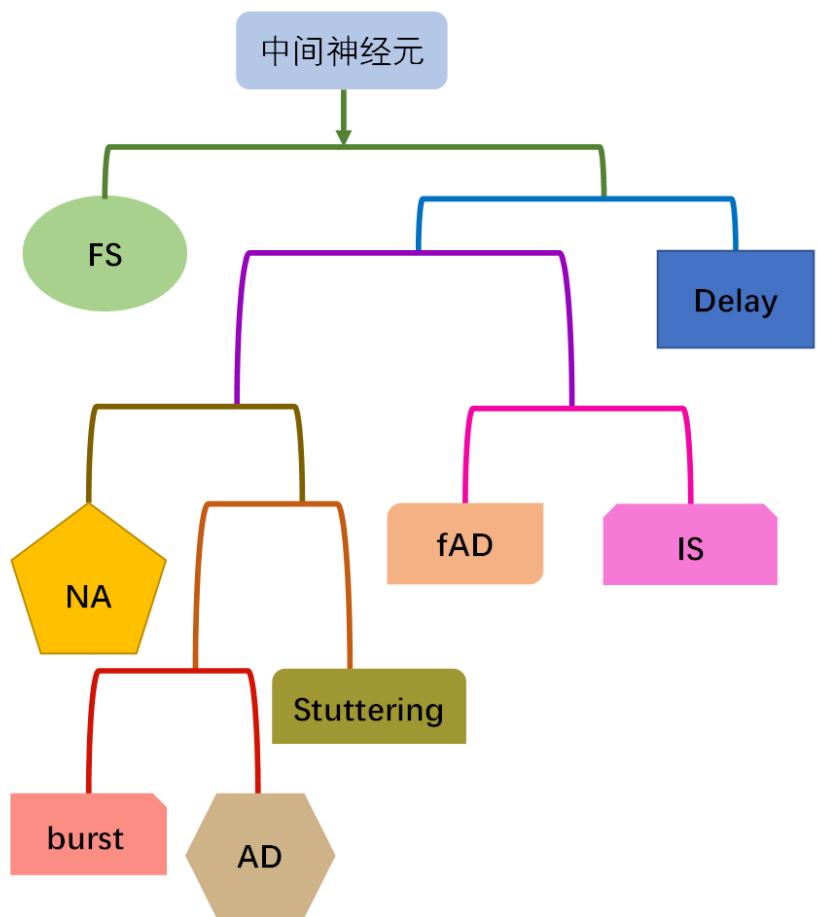


Figure 5-1 Diagram illustrating hierarchical relationship of classification results. FS: Fast Spiking; fAD: fast adaptation; IS: Irregular Spiking; NA: non-fast, non-spiking; AD: adapting.

Bibliography

- [1] LIM L, MI D, LLORCA A, et al. Development and Functional Diversification of Cortical Interneurons[J/OL]. *Neuron*, 2018, 100(2): 294-313. DOI: [10.1016/j.neuron.2018.10.009](https://doi.org/10.1016/j.neuron.2018.10.009).
- [2] BLATOW M, CAPUTI A, MONYER H. Molecular diversity of neocortical GABAergic interneurones[J/OL]. *The Journal of Physiology*, 2005, 562(1): 99-105. DOI: [10.1113/jphysiol.2004.078584](https://doi.org/10.1113/jphysiol.2004.078584).
- [3] ZENG H, SANES J R. Neuronal cell-type classification: Challenges, opportunities and the path forward[J/OL]. *Nature Reviews Neuroscience*, 2017, 18(9): 530-546. DOI: [10.1038/nrn.2017.85](https://doi.org/10.1038/nrn.2017.85).
- [4] BERNARD A, SORENSEN S A, LEIN E S. Shifting the paradigm: New approaches for characterizing and classifying neurons[J/OL]. *Current Opinion in Neurobiology*, 2009, 19(5): 530-536. DOI: [10.1016/j.conb.2009.09.010](https://doi.org/10.1016/j.conb.2009.09.010).
- [5] RUDY B, FISHELL G, LEE S, et al. Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons[J/OL]. *Developmental Neurobiology*, 2011, 71(1): 45-61. DOI: [10.1002/dneu.20853](https://doi.org/10.1002/dneu.20853).
- [6] TREMBLAY R, LEE S, RUDY B. GABAergic Interneurons in the Neocortex: From Cellular Properties to Circuits[J/OL]. *Neuron*, 2016, 91(2): 260-292. DOI: [10.1016/j.neuron.2016.06.033](https://doi.org/10.1016/j.neuron.2016.06.033).
- [7] KEPECS A, FISHELL G. Interneuron cell types are fit to function[J/OL]. *Nature*, 2014, 505(7483): 318-326. DOI: [10.1038/nature12983](https://doi.org/10.1038/nature12983).
- [8] MARKRAM H, Toledo-Rodriguez M, WANG Y, et al. Interneurons of the neocortical inhibitory system[J/OL]. *Nature Reviews Neuroscience*, 2004, 5(10): 793-807. DOI: [10.1038/nrn1519](https://doi.org/10.1038/nrn1519).
- [9] The Petilla Interneuron Nomenclature Group (PING), ASCOLI G A, Alonso-Nanclares L, et al. Petilla terminology: Nomenclature of features of GABAergic interneurons of the cerebral cortex[J/OL]. *Nature Reviews Neuroscience*, 2008, 9 (7): 557-568. DOI: [10.1038/nrn2402](https://doi.org/10.1038/nrn2402).
- [10] COSSART R, DINOCOURT C, HIRSCH J C, et al. Dendritic but not somatic GABAergic inhibition is decreased in experimental epilepsy[J/OL]. *Nature Neuroscience*, 2001, 4(1): 52-62. DOI: [10.1038/82900](https://doi.org/10.1038/82900).
- [11] Gonzalez-Burgos G, LEWIS D A. GABA Neurons and the Mechanisms of Network Oscillations: Implications for Understanding Cortical Dysfunction in Schizophrenia[J/OL]. *Schizophrenia Bulletin*, 2008, 34(5): 944-961. DOI: [10.1093/schbul/sbn070](https://doi.org/10.1093/schbul/sbn070).

- [12] COBOS I, CALCAGNOTTO M E, VILAYTHONG A J, et al. Mice lacking *Dlx1* show subtype-specific loss of interneurons, reduced inhibition and epilepsy[J/OL]. *Nature Neuroscience*, 2005, 8(8): 1059-1068. DOI: [10.1038/nn1499](https://doi.org/10.1038/nn1499).
- [13] LEVITT P, EAGLESON K L, POWELL E M. Regulation of neocortical interneuron development and the implications for neurodevelopmental disorders[J/OL]. *Trends in Neurosciences*, 2004, 27(7): 400-406. DOI: [10.1016/j.tins.2004.05.008](https://doi.org/10.1016/j.tins.2004.05.008).
- [14] TANIGUCHI H, HE M, WU P, et al. A Resource of Cre Driver Lines for Genetic Targeting of GABAergic Neurons in Cerebral Cortex[J/OL]. *Neuron*, 2011, 71(6): 995-1013. DOI: [10.1016/j.neuron.2011.07.026](https://doi.org/10.1016/j.neuron.2011.07.026).
- [15] Muñoz-Manchado A B, BENGTSSON GONZALES C, ZEISEL A, et al. Diversity of Interneurons in the Dorsal Striatum Revealed by Single-Cell RNA Sequencing and PatchSeq[J/OL]. *Cell Reports*, 2018, 24(8): 2179-2190.e7. DOI: [10.1016/j.celrep.2018.07.053](https://doi.org/10.1016/j.celrep.2018.07.053).
- [16] ARMAÑANZAS R, ASCOLI G A. Towards the automatic classification of neurons[J/OL]. *Trends in Neurosciences*, 2015, 38(5): 307-318. DOI: [10.1016/j.tins.2015.02.004](https://doi.org/10.1016/j.tins.2015.02.004).
- [17] GLASER J I, KORDING K P. The Development and Analysis of Integrated Neuroscience Data[J/OL]. *Front. Comput. Neurosci.*, 2016, 10. DOI: [10.3389/fncom.2016.00011](https://doi.org/10.3389/fncom.2016.00011).
- [18] SHARPEE T O. Toward Functional Classification of Neuronal Types[J/OL]. *Neuron*, 2014, 83(6): 1329-1334. DOI: [10.1016/j.neuron.2014.08.040](https://doi.org/10.1016/j.neuron.2014.08.040).
- [19] DEFELIPE J, López-Cruz P L, Benavides-Piccione R, et al. New insights into the classification and nomenclature of cortical GABAergic interneurons[J/OL]. *Nature Reviews Neuroscience*, 2013, 14(3): 202-216. DOI: [10.1038/nrn3444](https://doi.org/10.1038/nrn3444).
- [20] ASCOLI G A. Mobilizing the base of neuroscience data: The case of neuronal morphologies[J/OL]. *Nature Reviews Neuroscience*, 2006, 7(4): 318-324. DOI: [10.1038/nrn1885](https://doi.org/10.1038/nrn1885).
- [21] BORNSTEIN J, FURNESS J, KUNZE W. Electrophysiological characterization of myenteric neurons: How do classification schemes relate?[J/OL]. *Journal of the Autonomic Nervous System*, 1994, 48(1): 1-15. DOI: [10.1016/0165-1838\(94\)90155-4](https://doi.org/10.1016/0165-1838(94)90155-4).
- [22] TASIC B, MENON V, NGUYEN T N, et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics[J/OL]. *Nature Neuroscience*, 2016, 19(2): 335-346. DOI: [10.1038/nn.4216](https://doi.org/10.1038/nn.4216).
- [23] ABRAHAO K P, LOVINGER D M. Classification of GABAergic neuron subtypes from the globus pallidus using wild-type and transgenic mice[J/OL]. *The Journal of Physiology*, 2018, 596(17): 4219-4235. DOI: [10.1113/JP276079](https://doi.org/10.1113/JP276079).
- [24] GOUWENS N W, SORENSEN S A, BERG J, et al. Classification of electrophysiological and morphological neuron types in the mouse visual cortex[J/OL]. *Nature Neuroscience*, 2019, 22(7): 1182-1195. DOI: [10.1038/s41593-019-0417-0](https://doi.org/10.1038/s41593-019-0417-0).

- [25] BACCI A, HUGUENARD J R, PRINCE D A. Modulation of neocortical interneurons: Extrinsic influences and exercises in self-control[J/OL]. Trends in Neurosciences, 2005, 28(11): 602-610. DOI: [10.1016/j.tins.2005.08.007](https://doi.org/10.1016/j.tins.2005.08.007).
- [26] CONNORS B W, GUTNICK M J. Intrinsic firing patterns of diverse neocortical neurons[J/OL]. Trends in Neurosciences, 1990, 13(3): 99-104. DOI: [10.1016/0166-2236\(90\)90185-D](https://doi.org/10.1016/0166-2236(90)90185-D).
- [27] Anderson John C., Martin Kevan A. C., Picanco-Diniz Cristovam W. The neurons in layer 1 of cat visual cortex[J/OL]. Proceedings of the Royal Society of London. Series B: Biological Sciences, 1992, 248(1321): 27-33. DOI: [10.1098/rspb.1992.0038](https://doi.org/10.1098/rspb.1992.0038).
- [28] KAWAGUCHI Y, KUBOTA Y. GABAergic cell subtypes and their synaptic connections in rat frontal cortex[J]. Cerebral Cortex (New York, N.Y.: 1991), 1997, 7(6): 476-486.
- [29] KAWAGUCHI Y. Groupings of nonpyramidal and pyramidal cells with specific physiological and morphological characteristics in rat frontal cortex[J/OL]. Journal of Neurophysiology, 1993, 69(2): 416-431. DOI: [10.1152/jn.1993.69.2.416](https://doi.org/10.1152/jn.1993.69.2.416).
- [30] KAWAGUCHI Y, KUBOTA Y. Correlation of physiological subgroupings of nonpyramidal cells with parvalbumin- and calbindinD28k-immunoreactive neurons in layer V of rat frontal cortex[J/OL]. Journal of Neurophysiology, 1993, 70 (1): 387-396. DOI: [10.1152/jn.1993.70.1.387](https://doi.org/10.1152/jn.1993.70.1.387).
- [31] MENDONÇA P R, Vargas-Caballero M, ERDÉLYI F, et al. Stochastic and deterministic dynamics of intrinsically irregular firing in cortical inhibitory interneurons[J/OL]. eLife, 2016, 5: e16475. DOI: [10.7554/eLife.16475](https://doi.org/10.7554/eLife.16475).
- [32] PORTER J T, CAULI B, STAIGER J F, et al. Properties of bipolar VIPergic interneurons and their excitation by pyramidal neurons in the rat neocortex[J/OL]. The European Journal of Neuroscience, 1998, 10(12): 3617-3628. DOI: [10.1046/j.1460-9568.1998.00367.x](https://doi.org/10.1046/j.1460-9568.1998.00367.x).
- [33] WANG Y, GUPTA A, TOLEDO-RODRIGUEZ M, et al. Anatomical, physiological, molecular and circuit properties of nest basket cells in the developing somatosensory cortex[J/OL]. Cerebral Cortex (New York, N.Y.: 1991), 2002, 12(4): 395-410. DOI: [10.1093/cercor/12.4.395](https://doi.org/10.1093/cercor/12.4.395).
- [34] MARIN-PADILLA M. Origin of the pericellular baskets of the pyramidal cells of the human motor cortex: A Golgi study[J/OL]. Brain Research, 1969, 14(3): 633-646. DOI: [10.1016/0006-8993\(69\)90204-2](https://doi.org/10.1016/0006-8993(69)90204-2).
- [35] KISVÁRDY Z F, EYSEL U T. Cellular organization of reciprocal patchy networks in layer III of cat visual cortex (area 17)[J/OL]. Neuroscience, 1992, 46(2): 275-286. DOI: [10.1016/0306-4522\(92\)90050-c](https://doi.org/10.1016/0306-4522(92)90050-c).
- [36] KISVÁRDY Z F, MARTIN K A, WHITTERIDGE D, et al. Synaptic connections of intracellularly filled clutch cells: A type of small basket cell in the visual cortex of the cat[J/OL]. The Journal of Comparative Neurology, 1985, 241(2): 111-137. DOI: [10.1002/cne.902410202](https://doi.org/10.1002/cne.902410202).

- [37] JONES E G. Varieties and distribution of non-pyramidal cells in the somatic sensory cortex of the squirrel monkey[J/OL]. *The Journal of Comparative Neurology*, 1975, 160(2): 205-267. DOI: [10.1002/cne.901600204](https://doi.org/10.1002/cne.901600204).
- [38] DEFELIPE J, FAIRÉN A. A type of basket cell in superficial layers of the cat visual cortex. A Golgi-electron microscope study[J/OL]. *Brain Research*, 1982, 244(1): 9-16. DOI: [10.1016/0006-8993\(82\)90898-8](https://doi.org/10.1016/0006-8993(82)90898-8).
- [39] LUND J S, LEWIS D A. Local circuit neurons of developing and mature macaque prefrontal cortex: Golgi and immunocytochemical characteristics[J/OL]. *The Journal of Comparative Neurology*, 1993, 328(2): 282-312. DOI: [10.1002/cne.903280209](https://doi.org/10.1002/cne.903280209).
- [40] BUHL E H, HALASY K, SOMOGYI P. Diverse sources of hippocampal unitary inhibitory postsynaptic potentials and the number of synaptic release sites[J/OL]. *Nature*, 1994, 368(6474): 823-828. DOI: [10.1038/368823a0](https://doi.org/10.1038/368823a0).
- [41] MILES R, TÓTH K, GULYÁS A I, et al. Differences between somatic and dendritic inhibition in the hippocampus[J/OL]. *Neuron*, 1996, 16(4): 815-823. DOI: [10.1016/s0896-6273\(00\)80101-4](https://doi.org/10.1016/s0896-6273(00)80101-4).
- [42] ZHU Y, STORNETTA R L, ZHU J J. Chandelier cells control excessive cortical excitation: Characteristics of whisker-evoked synaptic responses of layer 2/3 nonpyramidal and pyramidal neurons[J/OL]. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 2004, 24(22): 5101-5108. DOI: [10.1523/JNEUROSCI.0544-04.2004](https://doi.org/10.1523/JNEUROSCI.0544-04.2004).
- [43] SOMOGYI P. A specific 'axo-axonal' interneuron in the visual cortex of the rat [J/OL]. *Brain Research*, 1977, 136(2): 345-350. DOI: [10.1016/0006-8993\(77\)90808-3](https://doi.org/10.1016/0006-8993(77)90808-3).
- [44] SOMOGYI P, FREUND T F, COWEY A. The axo-axonic interneuron in the cerebral cortex of the rat, cat and monkey[J/OL]. *Neuroscience*, 1982, 7(11): 2577-2607. DOI: [10.1016/0306-4522\(82\)90086-0](https://doi.org/10.1016/0306-4522(82)90086-0).
- [45] WANG Y, Toledo-Rodriguez M, GUPTA A, et al. Anatomical, physiological and molecular properties of Martinotti cells in the somatosensory cortex of the juvenile rat[J/OL]. *The Journal of Physiology*, 2004, 561(Pt 1): 65-90. DOI: [10.1113/jphysiol.2004.073353](https://doi.org/10.1113/jphysiol.2004.073353).
- [46] PETERS A, HARRIMAN K M. Enigmatic bipolar cell of rat visual cortex[J/OL]. *The Journal of Comparative Neurology*, 1988, 267(3): 409-432. DOI: [10.1002/cne.902670310](https://doi.org/10.1002/cne.902670310).
- [47] PETERS A. The axon terminals of vasoactive intestinal polypeptide (VIP)-containing bipolar cells in rat visual cortex[J]. *Journal of Neurocytology*, 1990, 19(5): 672-685.
- [48] SOMOGYI P, COWEY A. Combined Golgi and electron microscopic study on the synapses formed by double bouquet cells in the visual cortex of the cat and monkey[J/OL]. *The Journal of Comparative Neurology*, 1981, 195(4): 547-566. DOI: [10.1002/cne.901950402](https://doi.org/10.1002/cne.901950402).

- [49] DEFELIPE J, HENDRY S H, HASHIKAWA T, et al. A microcolumnar structure of monkey cerebral cortex revealed by immunocytochemical studies of double bouquet cell axons[J/OL]. *Neuroscience*, 1990, 37(3): 655-673. DOI: [10.1016/0306-4522\(90\)90097-n](https://doi.org/10.1016/0306-4522(90)90097-n).
- [50] SOMOGYI P, TAMÁS G, LUJAN R, et al. Salient features of synaptic organisation in the cerebral cortex[J/OL]. *Brain Research. Brain Research Reviews*, 1998, 26(2-3): 113-135. DOI: [10.1016/s0165-0173\(97\)00061-1](https://doi.org/10.1016/s0165-0173(97)00061-1).
- [51] BOLDOG E, BAKKEN T E, HODGE R D, et al. Transcriptomic and morphophysiological evidence for a specialized human cortical GABAergic cell type[J/OL]. *Nature Neuroscience*, 2018, 21(9): 1185. DOI: [10.1038/s41593-018-0205-2](https://doi.org/10.1038/s41593-018-0205-2).
- [52] Zeki S., Crick Francis. The impact of molecular biology on neuroscience[J/OL]. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 1999, 354(1392): 2021-2025. DOI: [10.1098/rstb.1999.0541](https://doi.org/10.1098/rstb.1999.0541).
- [53] ARENDT D, MUSSER J M, BAKER C V H, et al. The origin and evolution of cell types[J/OL]. *Nature Reviews Genetics*, 2016, 17(12): 744-757. DOI: [10.1038/nrg.2016.127](https://doi.org/10.1038/nrg.2016.127).
- [54] DRUCKMANN S, HILL S, SCHÜRMANN F, et al. A Hierarchical Structure of Cortical Interneuron Electrical Diversity Revealed by Automated Statistical Analysis[J/OL]. *Cereb Cortex*, 2013, 23(12): 2994-3006. DOI: [10.1093/cercor/bhs290](https://doi.org/10.1093/cercor/bhs290).
- [55] LI T, TIAN C, SCALMANI P, et al. Action Potential Initiation in Neocortical Inhibitory Interneurons[J/OL]. *PLOS Biology*, 2014, 12(9): e1001944. DOI: [10.1371/journal.pbio.1001944](https://doi.org/10.1371/journal.pbio.1001944).
- [56] HELM J, AKGUL G, WOLLMUTH L P. Subgroups of parvalbumin-expressing interneurons in layers 2/3 of the visual cortex[J/OL]. *Journal of Neurophysiology*, 2012, 109(6): 1600-1613. DOI: [10.1152/jn.00782.2012](https://doi.org/10.1152/jn.00782.2012).
- [57] STIEFEL K M, ENGLITZ B, SEJNOWSKI T J. Origin of intrinsic irregular firing in cortical interneurons[J/OL]. *PNAS*, 2013, 110(19): 7886-7891. DOI: [10.1073/pnas.1305219110](https://doi.org/10.1073/pnas.1305219110).
- [58] POHLKAMP T, DÁVID C, CAULI B, et al. Characterization and Distribution of Reelin-Positive Interneuron Subtypes in the Rat Barrel Cortex[J/OL]. *Cereb Cortex*, 2014, 24(11): 3046-3058. DOI: [10.1093/cercor/bht161](https://doi.org/10.1093/cercor/bht161).
- [59] NASSAR M, SIMONNET J, LOFREDI R, et al. Diversity and overlap of Parvalbumin and Somatostatin expressing interneurons in mouse presubiculum[J/OL]. *Front. Neural Circuits*, 2015, 9. DOI: [10.3389/fncir.2015.00020](https://doi.org/10.3389/fncir.2015.00020).
- [60] KARAGIANNIS A, GALLOPIN T, DÁVID C, et al. Classification of NPY-Expressing Neocortical Interneurons[J/OL]. *J. Neurosci.*, 2009, 29(11): 3642-3659. DOI: [10.1523/JNEUROSCI.0058-09.2009](https://doi.org/10.1523/JNEUROSCI.0058-09.2009).
- [61] BOEHLEN A, HEINEMANN U, HENNEBERGER C. Hierarchical spike clustering analysis for investigation of interneuron heterogeneity[J/OL]. *Neuroscience Letters*, 2016, 619: 86-91. DOI: [10.1016/j.neulet.2016.03.024](https://doi.org/10.1016/j.neulet.2016.03.024).

- [62] MARTÍNEZ J J, RAHSEPAR B, WHITE J A. Anatomical and Electrophysiological Clustering of Superficial Medial Entorhinal Cortex Interneurons[J/OL]. *eNeuro*, 2017, 4(5): ENEURO.0263-16.2017. DOI: [10.1523/ENEURO.0263-16.2017](https://doi.org/10.1523/ENEURO.0263-16.2017).
- [63] SANTANA R, MCGARRY L, BIELZA C, et al. Classification of neocortical interneurons using affinity propagation[J/OL]. *Front. Neural Circuits*, 2013, 7. DOI: [10.3389/fncir.2013.00185](https://doi.org/10.3389/fncir.2013.00185).
- [64] COSSART R, PETANJEK Z, DUMITRIU D, et al. Interneurons targeting similar layers receive synaptic inputs with similar kinetics[J/OL]. *Hippocampus*, 2006, 16(4): 408-420. DOI: [10.1002/hipo.20169](https://doi.org/10.1002/hipo.20169).
- [65] POLLEN A A, NOWAKOWSKI T J, CHEN J, et al. Molecular Identity of Human Outer Radial Glia during Cortical Development[J/OL]. *Cell*, 2015, 163(1): 55-67. DOI: [10.1016/j.cell.2015.09.004](https://doi.org/10.1016/j.cell.2015.09.004).
- [66] ZHANG J, TANG F, WANG X. The Primate-Specific Gene TMEM14B Marks Outer Radial Glia Cells and Promotes Cortical Expansion and Folding Article The Primate-Specific Gene TMEM14B Marks Outer Radial Glia Cells and Promotes Cortical Expansion and Folding[J/OL]. *Stem Cell*, 2017: 1-15. DOI: [10.1016/j.stem.2017.08.013](https://doi.org/10.1016/j.stem.2017.08.013).
- [67] HOSP J A, STRÜBER M, YANAGAWA Y, et al. Morpho-physiological criteria divide dentate gyrus interneurons into classes[J/OL]. *Hippocampus*, 2014, 24(2): 189-203. DOI: [10.1002/hipo.22214](https://doi.org/10.1002/hipo.22214).
- [68] HAMAM B N, AMARAL D G, ALONSO A A. Morphological and electrophysiological characteristics of layer V neurons of the rat lateral entorhinal cortex[J/OL]. *Journal of Comparative Neurology*, 2002, 451(1): 45-61. DOI: [10.1002/cne.10335](https://doi.org/10.1002/cne.10335).
- [69] EMMENEGGER V, QI G, WANG H, et al. Morphological and Functional Characterization of Non-fast-Spiking GABAergic Interneurons in Layer 4 Microcircuitry of Rat Barrel Cortex[J/OL]. *Cereb Cortex*, 2018, 28(4): 1439-1457. DOI: [10.1093/cercor/bhx352](https://doi.org/10.1093/cercor/bhx352).
- [70] MCGARRY L M, PACKER A M, FINO E, et al. Quantitative classification of somatostatin-positive neocortical interneurons identifies three interneuron subtypes[J/OL]. *Front. Neural Circuits*, 2010, 4. DOI: [10.3389/fncir.2010.00012](https://doi.org/10.3389/fncir.2010.00012).
- [71] AGHILI M, FANG R. Mining Big Neuron Morphological Data[J/OL]. *Computational Intelligence and Neuroscience*, 2018. DOI: [10.1155/2018/8234734](https://doi.org/10.1155/2018/8234734).
- [72] HALAVI M, HAMILTON K A, PAREKH R, et al. Digital reconstructions of neuronal morphology: Three decades of research trends[J/OL]. *Frontiers in Neuroscience*, 2012, 6: 49. DOI: [10.3389/fnins.2012.00049](https://doi.org/10.3389/fnins.2012.00049).
- [73] MEIJERING E. Neuron tracing in perspective[J/OL]. *Cytometry. Part A: The Journal of the International Society for Analytical Cytology*, 2010, 77(7): 693-704. DOI: [10.1002/cyto.a.20895](https://doi.org/10.1002/cyto.a.20895).

- [74] LONGAIR M H, BAKER D A, ARMSTRONG J D. Simple Neurite Tracer: Open source software for reconstruction, visualization and analysis of neuronal processes[J/OL]. *Bioinformatics* (Oxford, England), 2011, 27(17): 2453-2454. DOI: [10.1093/bioinformatics/btr390](https://doi.org/10.1093/bioinformatics/btr390).
- [75] FERREIRA T A, BLACKMAN A V, OYRER J, et al. Neuronal morphometry directly from bitmap images[J/OL]. *Nature Methods*, 2014, 11(10): 982-984. DOI: [10.1038/nmeth.3125](https://doi.org/10.1038/nmeth.3125).
- [76] PENG H, RUAN Z, LONG F, et al. V3D enables real-time 3D visualization and quantitative analysis of large-scale biological image data sets[J/OL]. *Nature Biotechnology*, 2010, 28(4): 348-353. DOI: [10.1038/nbt.1612](https://doi.org/10.1038/nbt.1612).
- [77] LUISI J, NARAYANASWAMY A, GALBREATH Z, et al. The FARSIIGHT trace editor: An open source tool for 3-D inspection and efficient pattern analysis aided editing of automated neuronal reconstructions[J/OL]. *Neuroinformatics*, 2011, 9 (2-3): 305-315. DOI: [10.1007/s12021-011-9115-0](https://doi.org/10.1007/s12021-011-9115-0).
- [78] COSTA M, MANTON J D, OSTROVSKY A D, et al. NBLAST: Rapid, Sensitive Comparison of Neuronal Structure and Construction of Neuron Family Databases [J/OL]. *Neuron*, 2016, 91(2): 293-311. DOI: [10.1016/j.neuron.2016.06.012](https://doi.org/10.1016/j.neuron.2016.06.012).
- [79] PENG H, HAWRYLYCZ M, ROSKAMS J, et al. BigNeuron: Large-Scale 3D Neuron Reconstruction from Optical Microscopy Images[J/OL]. *Neuron*, 2015, 87(2): 252-256. DOI: [10.1016/j.neuron.2015.06.036](https://doi.org/10.1016/j.neuron.2015.06.036).
- [80] SCORCIONI R, POLAVARAM S, ASCOLI G A. L-Measure: A web-accessible tool for the analysis, comparison and search of digital reconstructions of neuronal morphologies[J/OL]. *Nature Protocols*, 2008, 3(5): 866-876. DOI: [10.1038/nprot.2008.51](https://doi.org/10.1038/nprot.2008.51).
- [81] POHLKAMP T, DÁVID C, CAULI B, et al. Characterization and Distribution of Reelin-Positive Interneuron Subtypes in the Rat Barrel Cortex[J/OL]. *Cerebral Cortex*, 2014, 24(11): 3046-3058. DOI: [10.1093/cercor/bht161](https://doi.org/10.1093/cercor/bht161).
- [82] ZAITSEV A V. Classification and function of GABAergic interneurons of the mammalian cerebral cortex[J/OL]. *Biochemistry (Moscow) Supplement Series A: Membrane and Cell Biology*, 2013, 7(4): 245-259. DOI: [10.1134/S1990747813040090](https://doi.org/10.1134/S1990747813040090).
- [83] HE M, TUCCiarone J, LEE S, et al. Strategies and Tools for Combinatorial Targeting of GABAergic Neurons in Mouse Cerebral Cortex[J/OL]. *Neuron*, 2016, 91(6): 1228-1243. DOI: [10.1016/j.neuron.2016.08.021](https://doi.org/10.1016/j.neuron.2016.08.021).
- [84] NIQUILLE M, LIMONI G, MARKOPOULOS F, et al. Neurogliaform cortical interneurons derive from cells in the preoptic area[J/OL]. *eLife*, 2018, 7: e32017. DOI: [10.7554/eLife.32017](https://doi.org/10.7554/eLife.32017).
- [85] ZAITSEV A V, POVYSHEVA N V, Gonzalez-Burgos G, et al. Electrophysiological classes of layer 2/3 pyramidal cells in monkey prefrontal cortex[J/OL]. *Journal of Neurophysiology*, 2012, 108(2): 595-609. DOI: [10.1152/jn.00859.2011](https://doi.org/10.1152/jn.00859.2011).

- [86] LI W, DOYON W M, DANI J A. Quantitative unit classification of ventral tegmental area neurons in vivo[J/OL]. *Journal of Neurophysiology*, 2012, 107(10): 2808-2820. DOI: [10.1152/jn.00575.2011](https://doi.org/10.1152/jn.00575.2011).
- [87] PAGNUCO I A, PASTORE J I, ABRAS G, et al. Analysis of genetic association using hierarchical clustering and cluster validation indices[J/OL]. *Genomics*, 2017, 109(5): 438-445. DOI: [10.1016/j.ygeno.2017.06.009](https://doi.org/10.1016/j.ygeno.2017.06.009).
- [88] YU M, HILLEBRAND A, TEWARIE P, et al. Hierarchical clustering in minimum spanning trees[J/OL]. *Chaos*, 2015, 25(2): 023107. DOI: [10.1063/1.4908014](https://doi.org/10.1063/1.4908014).
- [89] KIMES P K, LIU Y, NEIL HAYES D, et al. Statistical significance for hierarchical clustering[J/OL]. *Biometrics*, 2017, 73(3): 811-821. DOI: [10.1111/biom.12647](https://doi.org/10.1111/biom.12647).
- [90] USOSKIN D, FURLAN A, ISLAM S, et al. Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing[J/OL]. *Nature Neuroscience*, 2015, 18(1): 145-153. DOI: [10.1038/nn.3881](https://doi.org/10.1038/nn.3881).
- [91] LI J, KLUGHAMMER J, FARLIK M, et al. Single-cell transcriptomes reveal characteristic features of human pancreatic islet cell types[J/OL]. *EMBO reports*, 2016, 17(2): 178-187. DOI: [10.15252/embr.201540946](https://doi.org/10.15252/embr.201540946).
- [92] žurauskienė J, YAU C. pcaReduce: Hierarchical clustering of single cell transcriptional profiles[J/OL]. *BMC Bioinformatics*, 2016, 17(1): 140. DOI: [10.1186/s12859-016-0984-y](https://doi.org/10.1186/s12859-016-0984-y).
- [93] INTARAPANICH A, SHAW P J, ASSAWAMAKIN A, et al. Iterative pruning PCA improves resolution of highly structured populations[J/OL]. *BMC Bioinformatics*, 2009, 10(1): 382. DOI: [10.1186/1471-2105-10-382](https://doi.org/10.1186/1471-2105-10-382).
- [94] KONISHI T. Principal component analysis for designed experiments[J/OL]. *BMC Bioinformatics*, 2015, 16(18): S7. DOI: [10.1186/1471-2105-16-S18-S7](https://doi.org/10.1186/1471-2105-16-S18-S7).
- [95] MENDONÇA P R, Vargas-Caballero M, ERDÉLYI F, et al. Stochastic and deterministic dynamics of intrinsically irregular firing in cortical inhibitory interneurons[J/OL]. *eLife*, 2016, 5: e16475. DOI: [10.7554/eLife.16475](https://doi.org/10.7554/eLife.16475).

Acknowledgements

Three years ago, I was fortunate to come here after the rigorous preparation of the postgraduate entrance examination. I am very grateful to have spent three meaningful years here. The late nights of solitude over the past three years have taught me to introspect, communicate with myself, calmly contemplate and delve into problems. The three years of tempering have surprised me with my potential. Visible growth fills me with great joy. All of these will become the most precious things in my life.

I am extremely grateful for the careful guidance of Teacher Yu over these three years, providing sufficient space and conditions for my learning and growth. Encouraging me to constantly experiment and explore my potential, demanding rigorously during the research process, pushing me to progress continuously. Many thanks to Senior Sister Shaona for her assistance, encouragement, discussing problems with me during our project on intermediate neuron classification, helping me search for literature, patiently explaining electrophysiology knowledge, teaching me the essence of cooperation, and experiencing the joy of jointly discussing and solving a problem, even if it's just a small one. Also, I deeply appreciate Teacher Fu's help in research projects, carefully revising my papers, providing suggestions, as well as caring and looking after me in daily life. I also want to thank Teacher Ganzhe for his care in research projects and daily life, as well as the concern and help from senior brothers and sisters in the laboratory, such as Brother Wang, Sister Wang Min, Brother Cao, Sister Youning, Shen Dan, Xu Yeqian, and others.

Furthermore, I want to express gratitude to Zheping for studying together, discussing problems together, staying up late writing programs and papers together. Thanks to Xiaoyu, Shumin, and Zhang Qi for their understanding and tolerance when I return to the dormitory late at night. I am fortunate to have met you all and spent these three years together, playing, growing, and creating beautiful memories of graduate life.

Finally, I want to thank my family and all those who have always cared about me. Thank you for your understanding and support, your unconditional love, tolerance, and encouragement. Whenever I encounter difficulties and confusion, you always patiently listen to me, analyze problems, and give me advice.

复旦大学 学位论文独创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。论文中除特别标注的内容外，不包含任何其他个人或机构已经发表或撰写过的研究成果。对本研究做出重要贡献的个人和集体，均已在论文中作了明确的声明并表示了谢意。本声明的法律结果由本人承担。

作者签名: 周莹 日期: 2019.12.9

复旦大学 学位论文使用授权声明

本人完全了解复旦大学有关收藏和利用博士、硕士学位论文的规定，即：学校有权收藏、使用并向国家有关部门或机构送交论文的印刷本和电子版本；允许论文被查阅和借阅；学校可以公布论文的全部或部分内容，可以采用影印、缩印或其它复制手段保存论文。涉密学位论文在解密后遵守此规定。

作者签名: 周莹 导师签名: 张永华 日期: 2019.12.9