NEOCORTICAL CELL TYPE CLASSIFICATION FROM ELECTROPHYSIOLOGY RECORDINGS USING DEEP NEURAL NETWORKS

Raymond Wang[1*], **Sang Min Han**[1*], **Marta Gajowa**[2], **Chunlei Liu**[1,3]
[1]Department of Electrical Engineering and Computer Sciences
[2]Department of Molecular and Cell Biology
[3]Helen Wills Neuroscience Institute
University of California, Berkeley
Berkeley, CA 94720, USA
{raymondwang, smhan, martagajowa, chunlei.liu}@berkeley.edu

ABSTRACT

Understanding the neural code requires identifying different functional units involved in the neural circuits. One way to identify these functional units is to solve a neuron type classification problem. For decades, current clamp electrophysiology recordings have provided the means to classify the neurons based on subtle differences in action potential shapes and spiking patterns. However, significant variations in neuronal type definitions, classification pipelines, and variability in the neuronal activities make unambiguous determination of neuron type challenging. Previous solutions to this electrophysiology-based cell type classification problem consisted of dimensionality reduction juxtaposed with clustering using hand-crafted action potential features. Recent discoveries have allowed genetic-based cell-type classifications, which have fewer ambiguities, but they are less practical in vivo and have even lower throughput. Leveraging the unprecedented ground truth data published in the Allen Institute Cell Types Database, which contains anatomical, genetic, and electrophysiology characterizations of neurons in the mouse neocortex, we construct a robust and efficient convolutional neural network (CNN) that successfully classifies neurons according to their genetic label or broad type (excitatory or inhibitory) solely using current-clamp electrophysiology recordings. The CNN is configured as a multiple-input single-output network consisting of three subnetworks that take in the raw time series electrophysiology recording as well as the real and imaginary components of its Fourier coefficients. Our single pipeline method is fast and streamlined while simultaneously outperforming previous methods and achieving more classification classes using only single current-clamp trace as the input. This end-to-end convolutional neural network-based classification method removes the need for hand-crafted features, specific knowledge, or human intervention for quick identification of the cell type with high accuracy, enabling interpretation of the experimental data in a bias-free manner and a much broader scientific context.

1 INTRODUCTION

The neuronal type classification problem has been present in neuroscience since Ramón y Cajal's presentation of the neuron doctrine, which highlighted the ample diversity of neurons. Neuroscientists hypothesized that morphology differences play a functional role in the neural circuit. This intuition was extended to the investigation of differences in neuronal activity with the appearance of the current clamp technique, which allowed observations of various action potential shapes and patterns in neurons.

[*]Equal contribution.

Features such as action potential (AP) threshold and frequency, full width at half maximum (FWHM) of an action potential (AP width), or afterhyperpolarization values and their common ratios in subsequent action potentials in a train, which are easily distinguishable by a person, aimed to describe the differences among the observed variabilities in neuronal activities. Although no two neurons have the same activity, these hand-crafted features were used to define electrophysiological types of neurons (Beierlein et al., 2003; Nowak et al., 2008). Due to intrinsic neuron-to-neuron variabilities and lack of established features that definitively separate neuron types, no single pipeline is sufficient enough for unambiguous classification.

Several electrophysiology-based classifications including methods that consider more features and ones that focus on a single neuronal subpopulation only coexist today (Petilla Interneuron Nomenclature Group; PING and others, 2008; Markram et al., 2004). Previous solutions to this cell type classification consisted of dimensionality reduction juxtaposed with clustering using calculated action potential features and cell morphology. These approaches also suffered in classification accuracy as they relied on AP shape, spiking pattern, or cell shape parameters that span a continuous feature space, which do not have clear separation borders.

In addition to the aforementioned classification methods based on morphology and electrophysiology, one based on genetic makeup appeared recently in systems and circuit neuroscience (Tasic et al., 2018). A vast majority of neurons can be clustered according to the genetic encoding characteristic of the group proteins. This gene-based consideration of neuron types allows a less ambiguous classification pipeline. The most commonly used genetic types are neuron-derived neurotrophic factor (Ndnf), parvalbumin (Pvalb), somatostatin (Sst), and vasoactive intestinal peptide (Vip) interneurons, and excitatory (Exc) neurons that are predominantly pyramidal cells.

Furthermore, the appearance of genetically modified transgenic lines of animals, where cells of a selected genetic type can be marked by fluorophore, enabled studies relating electrophysiological activity to the genetic type of neurons (Taniguchi et al., 2011). Attempts to characterize the electrophysiological features of specific genetic types of neurons showed that the majority of Pvalb interneurons correspond to Fast Spiking (FS) electrophysiological cell type, but both Vip and Sst interneurons as well as excitatory neurons can appear as Regular Spiking (RS) or Low Threshold Spiking (LTS) types (Contreras, 2004). Therefore, we can conclude that there does not exist a clear mapping from one classification scheme to another, partially due to low throughput of existing data and differences in definitions and experimental pipelines.

To address the above neuron type classification and improve on the existing classification pipeline architectures, a robust and efficient convolutional neural network (CNN) is developed that successfully classifies neurons according to their genetic label or broad type (excitatory or inhibitory) solely using current clamp electrophysiology recordings. The method presented in this paper relies on the ground truth data published in the Allen Institute Cell Types Database, which contains anatomical and morphological descriptions, genetic types, and electrophysiology features of thousands of neurons in the mouse neocortex (Gouwens et al., 2019). We use this open access database, which is one of the flagship initiatives of the Allen Institute for Brain Science, to obtain the genetic type label of a given neuron based on a short snippet of its action potential activity, bypassing supervised definitions of signal features as well as the electrophysiology classification scheme. The following sections detail the related state-of-the-art methods, the specifications of our novel deep neural network-based neuron type classification method, as well as the results achieved using our CNN architecture.

2 Related Work

Gouwens et al. (2019) built and made publicly available the Allen Cell Types Database. From this database, 17 electrophysiological, 38 morphological, and 46 morpho-electric neuron types were identified using a custom classification pipeline. The authors employed biocytin-based neuronal reconstruction to extract morphological features and used raw current clamp electrophysiology recordings of cells from the mouse visual cortex in vitro as inputs for electrophysiological features. After computing handcrafted single action potential features including action potential height, threshold, upstroke speed, and downstroke speed, as well as features corresponding to action potential trains, such as interspike intervals and spiking frequency, principal component analysis and t-distributed stochastic neighbor embedding techniques were applied to project the high-dimensional electrophysiological feature space into two dimensions. With clustering, the authors were able to identify

17 electrophysiological neuron types, 4 of which were classified as excitatory subtypes and 13 inhibitory. The 13 inhibitory subtypes were further mapped to the four inhibitory interneuron type based on genetic tags: Sst, Vip, Pvalb, and Ndnf.

Ghaderi et al. (2018) successfully developed a semi-supervised technique to classify neuron types using limited in vivo electrophysiology recordings data. The authors considered only 3 types of neurons: excitatory pyramidal (Pyr) cells, parvalbumin positive (Pvalb) interneurons, and somatostatin positive (Sst) interneurons from layer 2/3 of the mouse primary visual cortex. After extracting single spikes, they extracted discriminative action potential features by computing the Discrete Cosine Transform of the recorded electrophysiology traces. Principal component analysis and fuzzy c-mean clustering were then performed, and neurons were classified using minimum distance classifier. The authors achieved accuracies of 91.59 ± 1.69, 97.47 ± 0.67, and 89.06 ± 1.99 for Pvalb, Pyr, and Sst, respectively, which yielded an overall accuracy of $92.67 \pm 0.54\%$. This classification algorithm pipeline was further applied to the in vitro data from the Allen Institute Cell Types Database containing Pvalb, Sst, 5HT3a, and Vip genetic types. Testing on a dataset comprised of a pool of 50 neurons where multiple electrophysiology traces have been recorded for each neuron, the authors achieved accuracies of $93.57 \pm 0.59\%$, $89.15 \pm 0.63\%$, $81.69 \pm 0.56\%$, $79.23 \pm 1.38\%$, and $77.02 \pm 0.91\%$ for Pvalb, Sst, Vip, 5HT3a, and Pyr, respectively.

3 METHODS

In this paper, we present a neuron type classification technique based on a simple convolutional neural network (CNN) architecture. Using the in vitro current clamp electrophysiology recording traces of neurons in the mouse neocortex, the CNN is configured as a multiple-input single-output network consisting of three subnetworks. The first subnetwork takes in a portion of the raw time series recording that is 50 ms in duration and contains at least one action potential. The real and imaginary components of the trace's Fourier coefficients are fed into the second and third subnetworks, respectively.

3.1 DATASET

We use data collected from 1947 cells in the Allen Institute Cell Types Database (Gouwens et al., 2019). Of these, we omitted 81 cells containing only morphological features and lacking electrophysiology recordings, and only the remaining 1866 cells containing electrophysiology recordings were used to build our training, validation, and test sets. Every one of these 1866 cells was obtained from transgenic animal lines, and thus is associated with a genetic label (Exc, Ndnf, Vip, Sst, and Pvalb). Each neuron contains approximately 50 electrophysiology trace recordings that are responses to multiple current clamp stimuli including short square, long square, ramp, and noise. Of these traces, we only considered responses to the short square stimulus, which are 3 ms in duration that is just long enough to induce a single action potential. Each of these recordings are collected either at 200 kHz or 50 kHz sampling rate. Because each neuron is recorded at a different level of stimulus, we only take the traces which contain an action potential.

To obtain the most useful information about the neuron type being assessed, the type of classification task that our architecture solves can be dictated by the needs of a neurophysiologist. One task is to distinguish neuronal activity coming from an excitatory (Exc) or an inhibitory (Inh) neuron. This task is less informative, but provides helpful explanation for analyzing electrophysiological recordings. The more interesting task is to discriminate among the 5 genetic types (Pvalb, Sst, Vip, and Ndnf inhibitory types and Exc), which constitute the broader excitatory and inhibitory categories. We train our network to perform the two aforementioned tasks. Accurate discrimination of the 5 genetic types would be highly valuable to the neuroscience community.

3.2 PREPROCESSING

One of the main objectives of our approach is to remove as much supervised overhead in the data processing stage of the neuron classification pipeline as possible. The only preprocessing steps we perform on the raw time series data is removing excess portions of recordings that do not provide useful information. We are only interested in the portion of the recording containing an action potential, so only 50 ms of recording, 25 ms before and after the onset of the short square stimulus,

was considered. The 25 ms of pre- and post-stimulus time duration was chosen to ensure that potential discriminative features that may be present before and after the onset of stimulus would be captured by the representation learning performed by our convolutional neural network. The 25 ms of post-stimulus time guarantees that a single action potential has returned to its resting membrane potential after depolarization and hyperpolarization.

We also take the fast Fourier transform of these 50 ms time series traces. The real and imaginary components of the resulting Fourier coefficients are used as inputs to the subnetworks of our convolutional neural network architecture.

For model selection and performance, we divide the collection of aforementioned data from the Allen Institute Cell Types Database based on the unique cell identification numbers. The ratio between training and validation sets was fixed at 8:2. Once the best generalized performing model was identified, we independently split the dataset again based on the unique cell identification numbers. Eighty percent of the data was reserved for training and the remaining 20% was set aside as the test set data. The test set was further split as follows: 80% test and 20% validation. The validation set was used to tune the network hyperparameters. This dataset separation by cell prevents overfitting, and provides a significant advantage and improvement to existing methods. Due to natural cell-to-cell variations, basic action potential features like resting membrane potential can vary considerably among cells of the same genetic type. We therefore use and report the maximum validation accuracy we obtain over 100 epochs.

3.3 Network Architecture

We use a multiple-input single-output convolutional neural network (CNN) for training. To remove the need for handcrafted features, our deep neural network uses a one-dimensional convolutional neural network as a feature encoder and employs dense layers to output class predictions. The standard one-dimensional CNN encoder is implemented using PyTorch. The encoder contains 6 convolutional layers, and each layer is passed through a Rectified Linear Unit (ReLU) activation function. Batch normalization introduced by Ioffe & Szegedy (2015) is applied to each activated layer. The exact specifications of each layer is shown in c) of Figure 3.3.

For training, we employ the Adam optimizer with the learning rate set to 10^{-3} and ℓ_2 regularization parameter set to 10^{-5} to minimize the cross entropy loss with sum reduction (Kingma & Ba, 2014). The initial weights were randomly generated.

An 8:2 training-validation data set split was used to select the optimal network model configuration. We first tried a single-input single-output configuration, where only the top stream in c) of Figure 3.3 was used. This architecture resulted in a validation accuracy of 88.52% for classification among the 5 genetic neuron types: Exc, Pvalb, Sst, Ndnf, and Vip; and a validation accuracy of 96.35% for classification between excitatory and inhibitory cells.

We then tested using a dual-input single-output configuration. Features were trained independently on each subnetwork and were concatenated at the final step for classification. Using only the real component of the Fourier coefficients as input to the additional subnetwork resulted in a validation accuracies of 91.43% for classification among the 5 genetic neuron types, and 99.38% for classification between excitatory and inhibitory broad types. Similarly, using only the complex component of the Fourier coefficients resulted in validation accuracies of 89.13% for classification among the 5 genetic types, and 96.58% for classification between excitatory and inhibitory broad types.

Finally, we used a triple-input single-output configuration, where we use the raw time series trace in addition to both the real and imaginary components of the Fourier coefficients. Features were also trained independently on each subnetwork and were concatenated at the final step for classification. This configuration was our best performing architecture, which resulted in a validation accuracies of 92.05% for classification among the 5 genetic types and 98.10% for classification between excitatory and inhibitory broad types. This triple-input network architecture's ℓ_2 regularization hyperparameter was tuned to 10^{-5}.

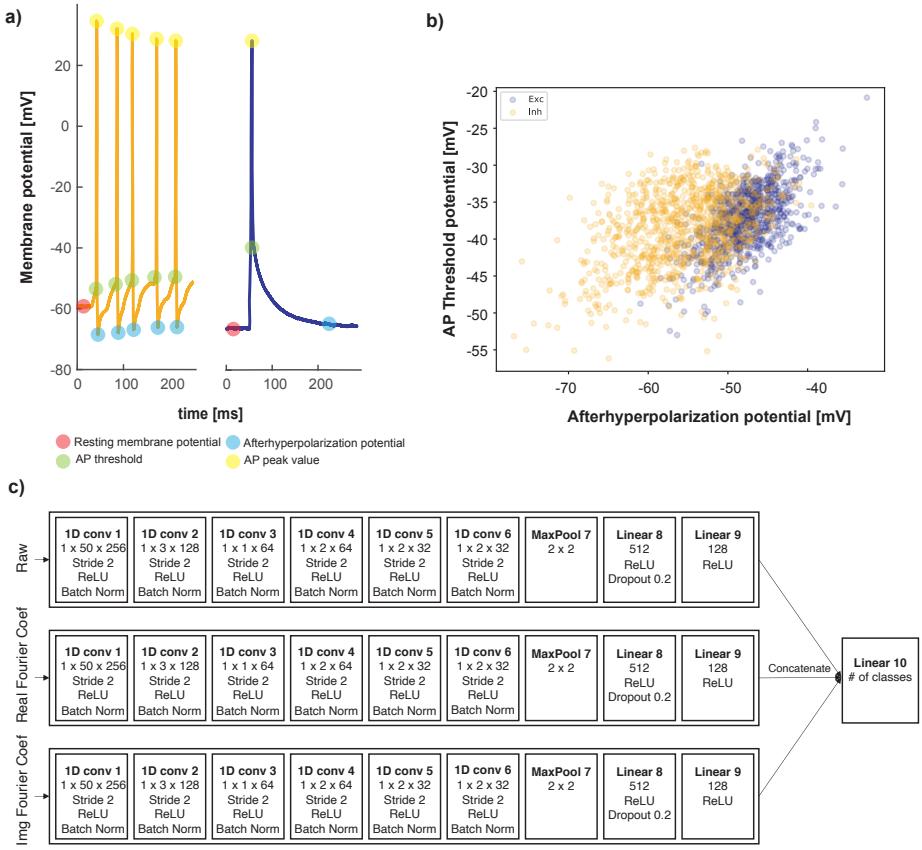

Figure 1: **a)** Diagram of neuronal activity featuring typical FS, Inh neuron (left) and RS, Exc neuron (right). Some exemplary AP features are marked on the traces. **b)** Example plot showing continuous values of AP features characteristic of Exc and Inh groups of neurons. **c)** Schematic diagram of the CNN architecture design developed.

3.4 Transfer Learning

One of the biggest challenges in neuroscience is obtaining sufficient amount of data for each neuron type for training. Supervised learning requires a large number of sample points to train efficient classifiers. We attempt to solve this problem using transfer learning. From the results of our triple-input single-output network, we can conclude that the 2 class classification task of identifying excitatory versus inhibitory neurons achieves high accuracy. This observation inspired us to first train the model on the task it does well, and then fine tune. We therefore trained the aforementioned CNN network for the two broad type classification task first for 100 epochs. We then changed the task to the 5 genetic type classification using the network trained on the two broad type classification task. The network uses the same CNN weights from the 2 type classification task, but has an output layer changed to the 5 type classification task. Then, the updated network was fine-tuned on the 5 type classification task for 100 epochs.

The same Adam optimizer with the learning rate of 10^{-3} and ℓ_2 regularization parameter of 10^{-5} was used for training. We tested the network's ability to fine tune on the data by carrying out 3:7, 5:5, and 8:2 training-test set splits. This choice was made to investigate the model's ability to fine tune on limited data points. For the 5 class classification task, at the epoch with the highest validation accuracy, the computed test set accuracies were 86.87%, 87.46%, and 89.22%, for the above three training-test set split ratios, respectively. Details are shown in Table 2.

In addition, we fine tuned the network on the 5 type classification task, used a 5 type classification validation set, selected the epoch with the highest validation accuracy, and evaluated a 2 class classification task's test set accuracy. This particular network has never seen this 2 class classification test set data during training, but resulted in a test set accuracy of 98.30%.

4 RESULTS

Our best performing triple-input single-output network architecture resulted in a test set accuracy of 89.76% for classification among the 5 genetic neuron types: excitatory (Exc), parvalbumin (Pvalb), somatostatin (Sst), neuron-derived neurotrophic factor (Ndnf), and vasoactive intestinal peptide (Vip) cells. This finalized network architecture resulted in a test set accuracy of 98.28% for classification between excitatory and inhibitory neurons. For both tasks, the individual classes' precisions, recalls, and f1-scores are reported in Table 1.

Table 1: Triple-input CNN precision

Type	precision (%)	recall (%)	f1-score (%)	support
		2 Class Validation Set		
Exc	98.78	98.78	98.78	490
Inh	98.52	98.52	98.52	406
weighted avg	98.66	98.66	98.66	896
		2 Class Test Set		
Exc	97.28	99.51	98.38	611
Inh	99.42	96.80	98.10	532
weighted avg	98.28	98.25	98.25	1143
		5 Class Validation Set		
Exc	96.27	99.79	98.00	466
Ndnf	84.62	95.65	89.80	46
Pvalb	99.17	98.76	98.96	241
Sst	88.24	87.21	87.72	86
Vip	87.50	46.67	60.87	45
weighted avg	95.23	95.36	94.94	884
		5 Class Test Set		
Exc	94.93	98.76	96.80	644
Ndnf	57.63	89.47	70.10	38
Pvalb	93.08	93.08	93.08	260
Sst	77.54	82.31	79.85	130
Vip	70.59	16.67	26.97	72
weighted avg	89.76	90.12	88.75	1144

Table 2 reports the test set accuracies of transfer learning.

Table 2: Transfer learning precision

Type	3:7	5:5	8:2
Exc	97.65	98.08	94.60
Ndnf	96.59	89.43	90.00
Pvalb	80.99	79.93	91.25
Sst	71.76	72.26	84.13
Vip	44.65	51.70	45.65
weighted avg	86.87	87.46	89.22

5 CONCLUSION AND FUTURE WORK

We presented in this paper a solution to the neuron classification problem that avoids using imperfect, non-standardized, and cumbersome electrophysiological classification schemes. In turn, the constructed neural network maps the pool of neurons based on their activity into less ambiguous genetic classification that is normally not widely accessible or practical in experimental pipelines. After training, our streamlined end-to-end convolutional neural network-based classification method does not require any domain specific knowledge or human intervention for quick identification of the neuron's cell type with high accuracy. The method presented in this paper provides an efficient and standardized tool for the neuroscience community to use, thus enabling data analysis in a broader scientific context. We further showed that the network architecture learns representations that successfully distinguishes the neuron types, even when these features are not immediately recognizable upon inspection as shown by the histograms of the data in the Appendix.

Although we achieved the state-of-the-art results, there exists potential source of noise in our data. The genetic labels we use in both the training, validation, and test sets are currently based on the animal type used to obtain the electrophysiology traces. The transgenic animal lines however are not guaranteed to be accurate. It is known that some portion of recordings coming from a given type will in fact belong to another genetic type (Hu et al., 2013). It is worth quantifying how much error is propagated from this potential noise source across all the genetic types for future work, and obtaining more accurate genetic labels from genetic sequencing data will be considered.

Furthermore, we plan to augment our dataset with in vivo electrophysiology recordings data, which is characterized by a more noisy background. Such extension will be particularly beneficial for rare but highly valuable in vivo–often blind to the cell type–current clamp recordings. Moreover, completely removing the need for tedious handcrafted features computation, which requires scientific domain expertise of an experimenter, will contribute to better reproducibility as well as faster, less biased scientific outcomes. On the other hand, studying the network outputs can contribute to improved understanding of what neuronal activity features are best for defining a given neuron type. It is a common assumption that encrypted in the genes is the ion channel repertoire that defines the electrophysiologically recorded neuronal activity (Nandi et al., 2020). Finding correlations or causal relationship between the genetic and electrophysiology classifications is one of the ultimate longings of a neurophysiologist, which follows the original intuition of Ramón Cajal that various groups of neurons differ from one another.

ACKNOWLEDGMENTS

The authors thank Lyle Graham for valuable feedback and advice. Sang Min Han was partially supported by the National Science Foundation Graduate Research Fellowship.

REFERENCES

Michael Beierlein, Jay R Gibson, and Barry W Connors. Two dynamically distinct inhibitory networks in layer 4 of the neocortex. *Journal of neurophysiology*, 90(5):2987–3000, 2003.

Diego Contreras. Electrophysiological classes of neocortical neurons. *Neural Networks*, 17(5-6): 633–646, 2004.

Parviz Ghaderi, Hamid Reza Marateb, and Mir-Shahram Safari. Electrophysiological profiling of neocortical neural subtypes: a semi-supervised method applied to in vivo whole-cell patch-clamp data. *Frontiers in neuroscience*, 12:823, 2018.

Nathan W Gouwens, Staci A Sorensen, Jim Berg, Changkyu Lee, Tim Jarsky, Jonathan Ting, Susan M Sunkin, David Feng, Costas A Anastassiou, Eliza Barkan, et al. Classification of electrophysiological and morphological neuron types in the mouse visual cortex. *Nature neuroscience*, 22(7):1182–1195, 2019.

Hang Hu, John Z Cavendish, and Ariel Agmon. Not all that glitters is gold: off-target recombination in the somatostatin–ires-cre mouse line labels a subset of fast-spiking interneurons. *Frontiers in neural circuits*, 7:195, 2013.

Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pp. 448–456. PMLR, 2015.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Henry Markram, Maria Toledo-Rodriguez, Yun Wang, Anirudh Gupta, Gilad Silberberg, and Caizhi Wu. Interneurons of the neocortical inhibitory system. *Nature reviews neuroscience*, 5(10):793–807, 2004.

Anirban Nandi, Tom Chartrand, Werner Van Geit, Anatoly Buchin, Zizhen Yao, Soo Yeun Lee, Yina Wei, Brian Kalmbach, Brian Lee, Ed Lein, et al. Single-neuron models linking electrophysiology, morphology and transcriptomics across cortical cell types. *bioRxiv*, 2020.

Lionel G Nowak, Maria V Sanchez-Vives, and David A McCormick. Lack of orientation and direction selectivity in a subgroup of fast-spiking inhibitory interneurons: cellular and synaptic mechanisms and comparison with other electrophysiological cell types. *Cerebral Cortex*, 18(5): 1058–1078, 2008.

Hiroki Taniguchi, Miao He, Priscilla Wu, Sangyong Kim, Raehum Paik, Ken Sugino, Duda Kvitsani, Yu Fu, Jiangteng Lu, Ying Lin, et al. A resource of cre driver lines for genetic targeting of gabaergic neurons in cerebral cortex. *Neuron*, 71(6):995–1013, 2011.

Bosiljka Tasic, Zizhen Yao, Lucas T Graybuck, Kimberly A Smith, Thuc Nghi Nguyen, Darren Bertagnolli, Jeff Goldy, Emma Garren, Michael N Economo, Sarada Viswanathan, et al. Shared and distinct transcriptomic cell types across neocortical areas. *Nature*, 563(7729):72–78, 2018.

Petilla Interneuron Nomenclature Group; PING and others. Petilla terminology: nomenclature of features of gabaergic interneurons of the cerebral cortex. *Nature reviews. Neuroscience*, 9(7):557, 2008.

A Histograms of Electrophysiology Recording Data

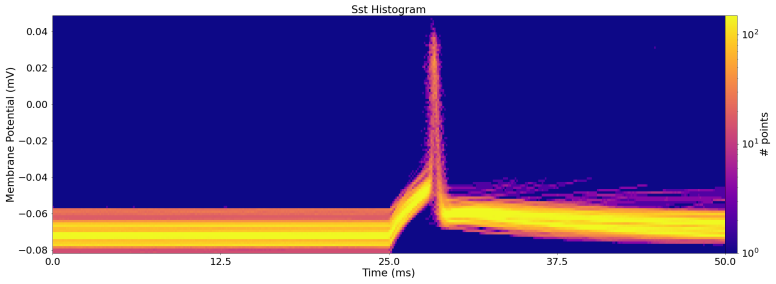

Figure 2: Time series histograms of the current clamp electrophysiology recordings for Sst

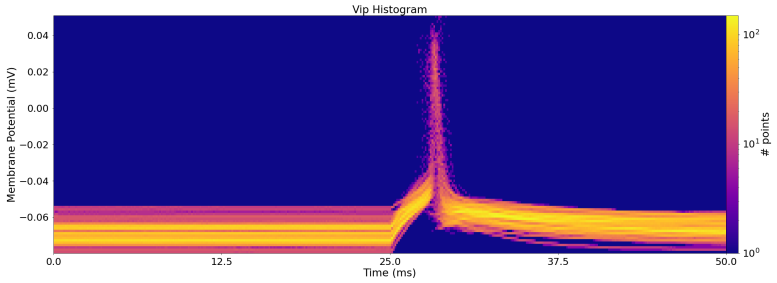

Figure 3: Time series histograms of the current clamp electrophysiology recordings for Vip

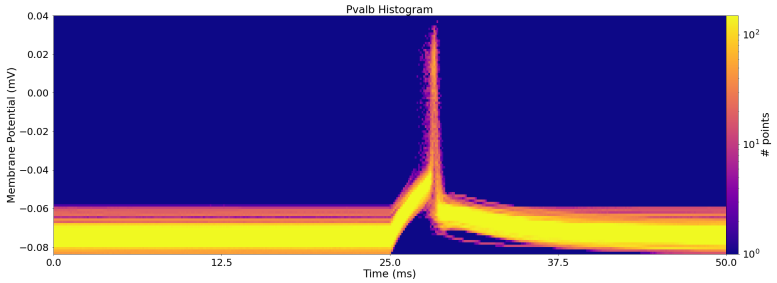

Figure 4: Time series histograms of the current clamp electrophysiology recordings for Pvalb

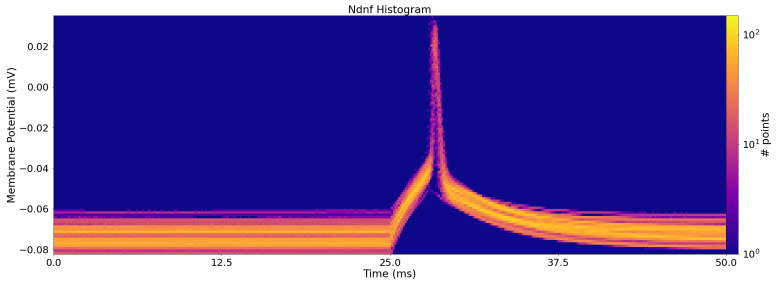

Figure 5: Time series histograms of the current clamp electrophysiology recordings for Ndnf

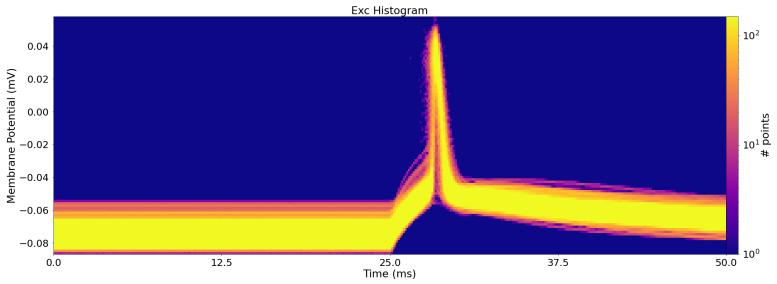

Figure 6: Time series histograms of the current clamp electrophysiology recordings for Exc

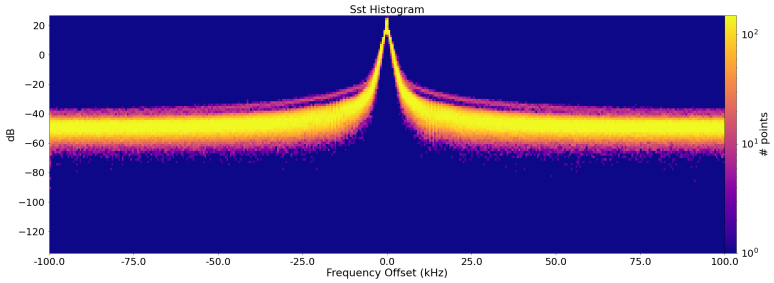

Figure 7: Fourier spectra histograms of the time series data for Sst

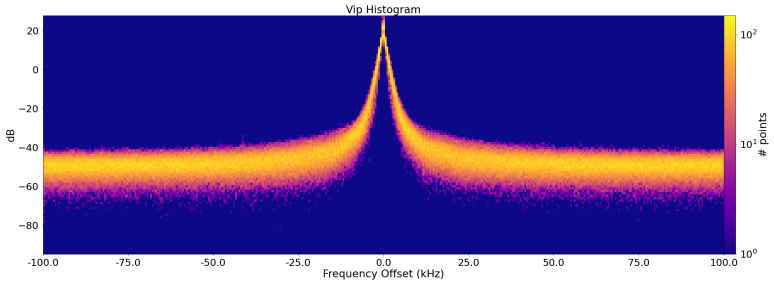

Figure 8: Fourier spectra histograms of the time series data for Vip

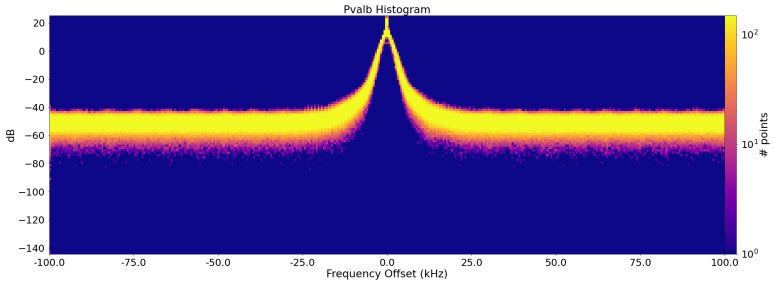

Figure 9: Fourier spectra histograms of the time series data for Pvalb

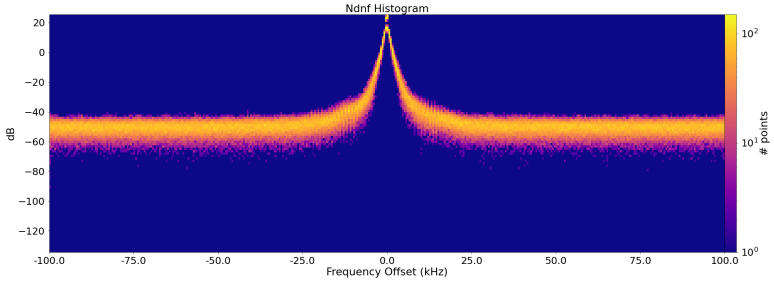

Figure 10: Fourier spectra histograms of the time series data for Ndnf

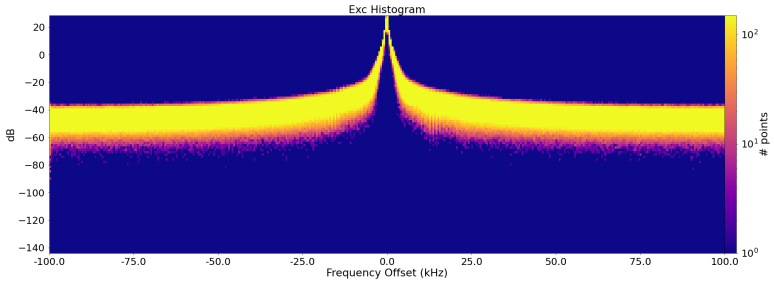

Figure 11: Fourier spectra histograms of the time series data for Exc