Neuroscience

A conserved code for anatomy: Neurons throughout the brain embed robust signatures of their anatomical location into spike trains

Reviewed Preprint
v1 • November 19, 2024
Not revised

Gemechu B Tolossa, Aidan M Schneider, Eva L Dyer, Keith B Hengen

Department of Biology, Washington University in Saint Louis, Saint Louis, United States • Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, United States

eLife Assessment

This article reports a **useful** set of findings on how electrophysiological response properties of neurons correlate with their position in the brain. The evidence currently remains **incomplete**, with reviewers making specific suggestions for how clustering needs to be redone. The manuscript would also benefit from a more focused presentation of results and the removal of incorrect claims about recording biases.

https://doi.org/10.7554/eLife.101506.1.sa3

Abstract

Neurons in the brain are known to encode diverse information through their spiking activity, primarily reflecting external stimuli and internal states. However, whether individual neurons also embed information about their own anatomical location within their spike patterns remains largely unexplored. Here, we show that machine learning models can predict a neuron's anatomical location across multiple brain regions and structures based solely on its spiking activity. Analyzing high-density recordings from thousands of neurons in awake, behaving mice, we demonstrate that anatomical location can be reliably decoded from neuronal activity across various stimulus conditions, including drifting gratings, naturalistic movies, and spontaneous activity. Crucially, anatomical signatures generalize across animals and even across different research laboratories, suggesting a fundamental principle of neural organization. Examination of trained classifiers reveals that anatomical information is enriched in specific interspike intervals as well as responses to stimuli. Within the visual isocortex, anatomical embedding is robust at the level of layers and primary versus secondary but does not robustly separate individual secondary structures. In contrast, structures within the hippocampus and thalamus are robustly separable based on their spike patterns. Our findings reveal a generalizable dimension of the neural code, where anatomical information is multiplexed with the encoding of external stimuli and internal states. This discovery provides new insights into the relationship between brain structure and function, with broad implications for neurodevelopment, multimodal integration, and the

interpretation of large-scale neuronal recordings. Immediately, it has potential as a strategy for in-vivo electrode localization.

Introduction

Foundational to any effort towards understanding the brain is a singular question: what information is carried in a neuron's spiking? It is widely understood that the action potential is the unit of information exchange in the central nervous system. Adrian first recorded a single neuron's action potential in 1928, establishing a rate code in sensory neurons [1]. In 1943, McCulloch and Pitts demonstrated that neuronal circuits could compute Boolean algebra [2], and 16 years later showed that the visual environment is conveyed to the brain by way of neuronal spike patterns [3]. Since then, the concept of a neural code has emerged— neuronal spiking is determined by inputs, including stimuli, and noise [4, 5, 6, 7]. In parallel works spanning 1899 to 1951, Cajal, Brodmann, and Penfield demonstrated diverse neuronal types that organize into anatomical loci, each with distinct functional contributions to sensation, perception, and behavior [8, 9, 10]. In other words, information carried by neighboring neurons is likely to be similar in content, be it visual or interoceptive. However, much of our understanding of the neural code is derived from experimental designs that manipulate stimuli or measure behavior. This approach leaves innumerable other forms of biologically relevant features as potentially latent variables, such as reliable identifying information about the neuron itself. A complete understanding of these latent variables is essential for a complete understanding of a neuron's role in the brain.

The null hypothesis is that the impact of anatomy on a neuron's activity is either nonexistent or unremarkable. This is supported by three observations at the level of models. First, from a computational perspective, neurons' outputs primarily reflect their inputs along with noise [7, 11, 12]. Second, artificial recurrent neural networks, which also perform input integration, can emulate patterns of neural circuit activity and functions central to biology, including motor control [13] and visual processing [14]. Yet, such networks do not require the formalized structure that is a hallmark of brain organization. Finally, in deep neural networks, anatomy is only relevant insofar as progressively deeper layers are functionally distinct. Taken together, complex information processing is achievable with neither a strict concept of anatomy nor a computational encoding of anatomy.

However, there are three principal reasons to justify asking if neurons reliably embed their anatomical location in their spiking. First, recent work suggests that, in addition to stimulus information, neurons transmit their genetic identity [15]. Second, in contrast to artificial systems, there are conceptual reasons why a brain might benefit from a reliable neuron-level code for anatomy. For example, anatomical information is presumably crucial during neurodevelopment [16]. Likewise, such information could be valuable when parsing inputs during multimodal integration [17]. Finally, within a species, brain anatomy is highly stereotyped. Thus, it stands to reason that, should a neuron's anatomy be embedded in its spike train, this embedding might generalize across individuals.

Historically, the classification of brain regions has utilized anatomical landmarks, functional outputs, and, more recently, genetic markers, creating a comprehensive map of neural diversity. There is evidence that neural activity may vary by region, although such observations are restricted to population-level statistics [18]. Specifically, patterns of neural activity evolve gradually across the anatomical landscape, influenced by gradients of synaptic weight [19, 20], the autocorrelation timescale [21], and connectivity patterns [22, 23, 24]. While indicative of the brain's complex architecture, these observations comprise subtle, statistical differences between populations rather than stark, unique signatures that could be used to classify an individual neuron with any confidence. Thus it is unclear whether different computational

rules have evolved in functionally distinct circuits. Alternatively, subtle statistical changes could reflect methodological limitations to capturing the full spectrum of activity patterns that define distinct brain regions. Should robust, circuit-specific rules exist, their description would enable a more precise understanding of how regional variations contribute to the broader neural code, and sharpen our understanding of the brain's computational organization.

Ultimately, the question can be distilled as, what is the minimum spatial scale at which neural activity patterns are reliably structured by the region of origin? Such patterns, should they exist, are most meaningful when identified at the level of single neurons. However, given the variance observed in any number of neuronal features—from tuning properties, to genetic cell type, to connectivity — it is unlikely that anatomy reliably determines neural activity at an obvious level, if at all. To address this, we employed a supervised machine learning approach to analyze publicly available datasets of high density, multi-region, single unit recordings in awake and behaving mice. To specifically evaluate whether anatomical information is embedded in the neuronal code, we examined the timing of spikes generated by well-isolated single units. We examined the possibility of anatomically-defined computational rules at four levels: 1) large-scale brain regions (hippocampus, midbrain, thalamus, and visual isocortex), 2) hippocampal structures (CA1, CA3, dentate gyrus, prosubiculum, and subiculum), 3) thalamic structures (ethmoid nucleus, dorsal lateral geniculate, lateral posterior nucleus, ventral medial geniculate, posterior complex, suprageniculate nucleus, ventral posteromedial nucleus), and 4) visual cortical structures (primary, anterolateral, anteromedial, lateral, posteromedial, rostrolateral). We find that traditional measures of neuronal activity (e.g. firing rate) alone are unable to reliably distinguish anatomical location. In contrast, using a multi-layer perceptron (MLP), we reveal that information about brain regions and structure is recoverable from more complete representations of single unit spiking (e.g. interspike interval distribution). Further, we demonstrate that learning the computational anatomical rules in one animal is sufficient to decode another, both within and across distinct research groups. These observations suggest a conserved code for anatomical origin in the output of individual neurons, and that, in modern datasets that span multiple regions, electrode localization can be assisted by spike timing.

Results

Dataset and inclusion criteria

To evaluate whether individual neurons embed reliable information about their structural localization in their spike trains, we required consistently reproduced recordings of large numbers of single units distributed throughout numerous cortical and subcortical structures. Further, we reasoned that candidate recordings should involve only awake and behaving animals. Finally, while highly restricted and repetitive stimuli are frequently leveraged to average out noise [25], this approach increases the likelihood of overfitting in our case. Thus, we only considered datasets that contained diverse stimuli as well as spontaneous activity. Two open datasets from the Allen Institute meet these criteria, specifically 1) Brain Observatory and 2) Functional Connectivity [18]. These datasets comprise tens of thousands of neurons recorded with high density silicon probes (Neuropixels) in a total of $N = 58$ mice (BO $N = 32$, FC $N = 26$). We used the selections of drifting gratings and naturalistic movie [26] found in the Brain Observatory dataset, which were distinct from those in the Functional Connectivity dataset. Additionally, we used recordings of spontaneous activity during blank screen presentations from both the Brain Observatory and Functional Connectivity datasets. Raw data were spike sorted at the Allen Institute; because information transmission in the brain arises primarily from the timing of spiking and not waveforms (etc)., our studies involve only the timestamps of individual spikes from well-isolated units (**Fig. 1A**). These units were filtered according to objective quality metrics such as ISI violations, presence ratio and amplitude cutoff (see Methods). At the largest spatial division, neurons were assigned to four distinct brain regions; hippocampus, midbrain, thalamus, and

visual cortices. Within regions, neurons were assigned to fine grained structures, for example CA1 hippocampus, lateral posterior thalamus, and anteromedial visual cortex (**Fig. 1B**). Note that midbrain neurons were not further classified by structure due to the relatively low number of neurons recorded there (to be considered at the structure level, we required a minimum of n = 150 neurons). We tested the possibility of a computational embedding of anatomical location at two levels of generalizability. In the transductive approach, all neurons from all animals were merged before splitting into a training set and a testing set. This arrangement preserves the capacity of a model to learn some within-animal features. In contrast, for the inductive approach, model training is performed on all neurons from a set of animals, and model testing is performed on neurons from entirely withheld animals. In this case, any successful learning must, by definition, be generalizable across animals (**Fig. 1A**).

Exploratory trends in spiking by region and structure across the population

Many studies have highlighted statistical differences in the activity of populations of neurons as a function of brain region and structure [21, 27, 28, 29]. At face value, this raises the possibility that anatomy may have a powerful role in shaping spiking patterns. However, the existence of population-level differences does not necessarily imply that the anatomical location of an individual neuron can be reliably inferred from its spike train alone; especially in large datasets, even subtle regional differences in spiking may be significant. Put simply, statistically separable populations can exhibit extensive overlap [30]. Prior to considering patterns embedded in individual spike trains, we asked whether and to what extent anatomical location explains the diversity of spiking patterns when considered across the entire population of included neurons (n = 18,691 neurons from N = 32 mice). Specifically, we applied unsupervised analyses in which the dominant trends in data are learned first, and afterwards, the correspondence of trends to anatomical location is examined.

We examined three distinct representations of spiking activity: 1) a previously described collection of 14 well-established spiking metrics (e.g. firing rate, coefficient of variation, etc.) [15], 2) the distribution of interspike intervals (ISIs) (0-3 s, 100 uniform bins), and 3) the trial averaged PSTHs for each drifting grating condition (the collection of gratings included 8 orientations and 5 frequencies, PSTH was calculated in 100 msec bins across a 3 s trial). To ensure generalizability across unsupervised analyses, we employed three dimensionality reduction methods: 1) Principal Components Analysis (PCA) [31], 2) t-distributed Stochastic Neighbor Embedding (t-SNE) [32], and 3) Uniform Manifold Approximation and Projection (UMAP) [33]. The resultant scatter plots suggest that, depending on the combination of analytical method and included features, there are subtle but statistically significant influences of brain region and structure in these dimensionally-reduced spaces (**Fig. 1C**, **S1**). However, across all 36 combinations of anatomical tasks, examined features, and analytical approaches, there was not a single example of a prominent grouping at the level of brain structure or brain region. These data may suggest that, examined as large populations, various features of neuronal spike timing may differ between regions, but that such differences are neither distinct nor robust. In this context, it is interesting to note that, despite rigidly stereotyped anatomy and structures defined by cell-type specific enrichment [34, 35], hippocampal populations stood out for their anomalous lack of statistically significant partitioning between structures (**Fig. S1**).

Separability of brain structures using established spiking metrics

Conservatively, it is possible that our unsupervised analysis (**Fig. 1**) reflects a true limit of the separability of structures based on spiking activity. However, by definition, unsupervised approaches reflect the most prominent trends (e.g. the largest sources of variance, in the case of PCA) in neuronal activity, which may or may not be related to anatomical location. As a result, it is possible that an analysis whose effectiveness is defined by the separability of region and structure

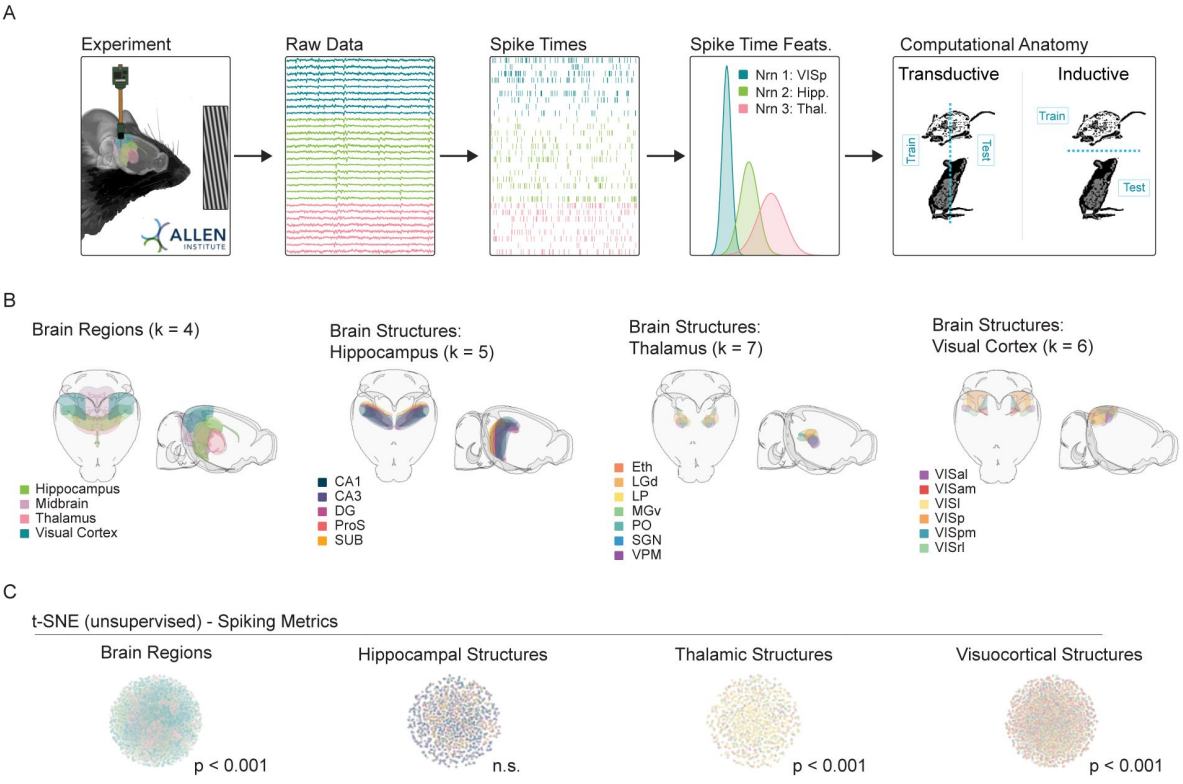


Figure 1

Dimensionality reduction suggests a limited relationship between neuroanatomy and spike timing.

A. Experimental pipeline. Left to right: Recording, raw data, extracted spike times, spiking time features (e.g., rates, CV), and model training protocols. The Allen Institute Visual Coding dataset comprises high density silicon extracellular recordings that span multiple brain regions and structures. During recording, mice were headfixed and presented with multiple visual stimuli, including drifting gratings. For supervised experiments, classifiers were either transductive—all neurons from all animals were mixed, and divided into train and test sets, or inductive— train and test sets were divided at the level of the animal. **B.** Brain regions and structures included in our analyses. (Left to right) Brain Regions: Hippocampus, Midbrain, Thalamus and Visual Cortex. Hippocampal Structures: CA1, CA3, Dentate Gyrus (DG), Prosubiculum (ProS), Subiculum (SUB). Thalamic Structures: Ethmoid Nucleus (Eth), Dorsal Lateral Geniculate (LGd), Lateral Posterior Nucleus (LP), Ventral Medial Geniculate (MGv), Posterior Complex (PO), Suprageniculate Nucleus (SGN), Ventral Posteromedial Nucleus (VPM). Visuocortical Structures: Anterolateral (VISal), Anteromedial (VISam), Lateral (VISl), Primary (VISp), Posteromedial (VISpm), Rostrolateral (VISrl). **C.** Unsupervised t-SNE plot of units recorded in each set of regions/structures. For each unit, 14 spiking metrics (see methods) describe the spike train, which is then placed in t-SNE, 2D scatterplot. Color scheme follows 1B. P-values are derived from a permutation test (shuffled control) of structure/region classifiability on the dimensionally-reduced space-see Unsupervised Analysis section of Methods

may reveal more reliable anatomical rules. To test this possibility, we turned to a supervised approach. We asked whether, given the spiking activity of a single unit, could its region and/or structure be successfully predicted by a machine-learning model?

Broadly, in our initial supervised experiments we pooled units from all animals. Pooled neurons were divided into non-overlapping sets for training (60%), validation (20%), and testing (20%). Crucially, equal proportions of each animal's neurons were maintained in each set. This architecture allowed us to seek patterns across all animals that generalize to unseen neurons from the same set of animals (**Fig. 1A** Computational Anatomy). Here, borrowed from machine-learning, we use the term *transductive* to describe this approach which can transfer patterns learned from neighboring neurons to unseen ones [36, 37].

Model performance is quantified with balanced accuracy, which is the average prediction accuracy for each class (i.e., region or structure) [38]. Error is ± SEM, and chance is 1/number of classes. To clearly indicate how models make errors between classes, we use confusion matrices—square heat maps where rows indicate true class labels and columns indicate predicted class labels. Note that diagonal entries represent the portion of units whose region/structure was correctly predicted. Entries off the diagonal display instances of the model's "confusion". An effective model will comprise a high balanced accuracy and show strong diagonal structure in the corresponding confusion matrix.

We first passed standard measures of neuronal activity to a simple model. Specifically, we trained logistic regressions to predict a unit's region/structure from 14 statistical measures that comprised three categories: 1) standard statistics, such as mean or maximum firing rate, 2) measures of local variance, such as CV2, a temporally-constrained alternative to the coefficient of variation (CV) [39], and 3) spectral power, such as the delta band (0.1-4 Hz). Models were trained and tested in each of four prediction tasks: 1) brain regions, 2) hippocampal structures, 3) thalamic structures, and 4) visuocortical structures (**Fig. 2A-D**). In each of these tasks, we evaluated the effectiveness of the 14 features individually as well as in combination to predict the anatomical localization of individual neurons (**Fig.2**).

In all four tasks, the majority of individual spiking metrics supported weak but above-chance classification (**Fig. 2A-D**). In contrast to brain regions and hippocampal structures, the ability of the 14 metrics to reliably indicate different visuocortical structures was notably low. Ordered by their classification performance (**Fig. 2A-D**, x-axes), the arrangement of spiking metrics was highly task dependent. We examined the correlation of metric ordering— the order of classification effectiveness for the spiking metrics, from greatest to least— between the region task and the three structure tasks. We found a range of correlation from 0.03 to 0.73 (Spearman Rank Correlations-regions/VC: $p = 0.0336$, regions/TH: $p = 0.0814$, regions/HC: $p = 0.7253$), suggesting that the characteristics of the spike train that best discriminate regions are in some cases distinct from the characteristics used to discriminate structure. In combination, the 14 metrics substantially increased the overall balanced accuracy for each of the four tasks, although the visuocortical task remained challenging (**Fig. 2E-H**; Brain Regions Balanced Accuracy = 52.91 ±1.24; Hippocampal Structures = 44.10 ±1.99; Thalamic Structures = 37.14 ±2.57; Visuocortical Structures = 24.09 ±1.46). Perhaps unsurprisingly, classification of regions was more effective than the structures within them. Taken together, these results demonstrate an intermediate divisibility of regions and structures by logistic regression based on spiking metrics. Given that spiking metrics were predetermined and that logistic regression can only learn linearly discriminable features, it is possible that more robust anatomical information could emerge from the spike train with more complex learned representations.

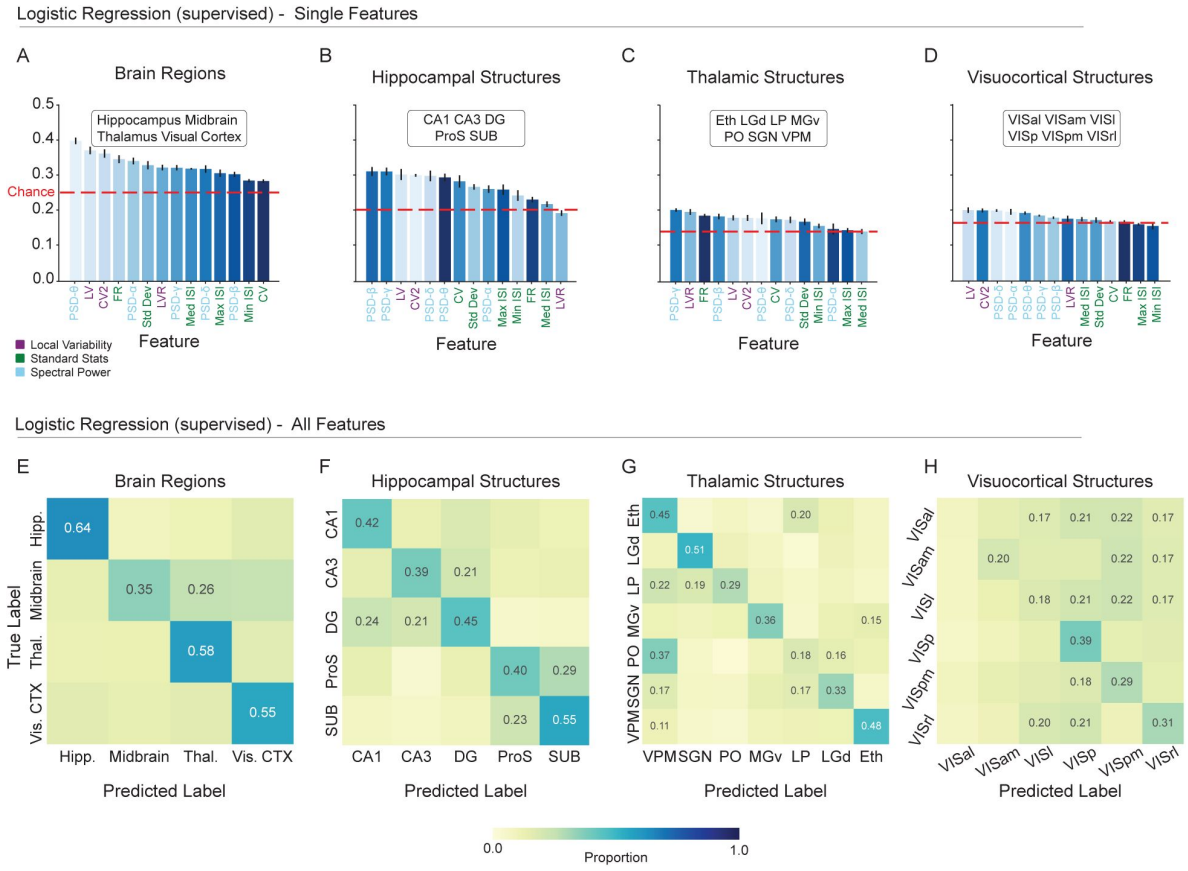


Figure 2

A linear classifier can learn to predict single neuron location based on standard spiking metrics.

A-D., Balanced accuracy of logistic regression models trained to predict the anatomical location of single units based on each of 14 individual spiking metrics: Coefficient of Variation 2 (CV2), Local Variation (LV), Revised Local Variation (LVR), Mean Firing Rate (FR), Standard Deviation (Std Dev) of interspike intervals (ISIs), Coefficient of Variation (CV), Minimum ISI (Min ISI), Median ISI (Med ISI), Maximum ISI (Max ISI), power spectral density (PSD)-*δ* (Delta Band 0.1-4 Hz), PSD-*θ* (Theta Band 4-8 Hz), PSD-*α* (Alpha Band 8-12 Hz), PSD-*β* (Beta Band 12-40 Hz), PSD-*γ* (Gamma Band 40-100 Hz). Balanced accuracy expected by chance varies by task and is indicated by the dashed red line. Features on the x-axis are ordered by performance. Feature (bar) colors are assigned by the ordering in A (brain region task) and maintained for the structure tasks (B-D). This shows the extent to which individual features maintain their relative importance across tasks. **E-H.**, Confusion matrices from logistic regression models trained to predict unit location from the combination of all 14 spiking metrics. Each confusion matrix shows the average of 5 train/test splits of the data. The proportion is printed only in cells where proportion was greater than chance level (1/number of classes). Balanced accuracy for each task: Brain Regions = 52.91 ±1.24; Hippocampal Structures = 44.10 ±1.99; Thalamic Structures = 37.14 ±2.57; Visuocortical Structures = 24.09 ±1.46 (error is the SEM across 5 splits).

Flexible, non-linear embedding of anatomical information in neuronal spiking

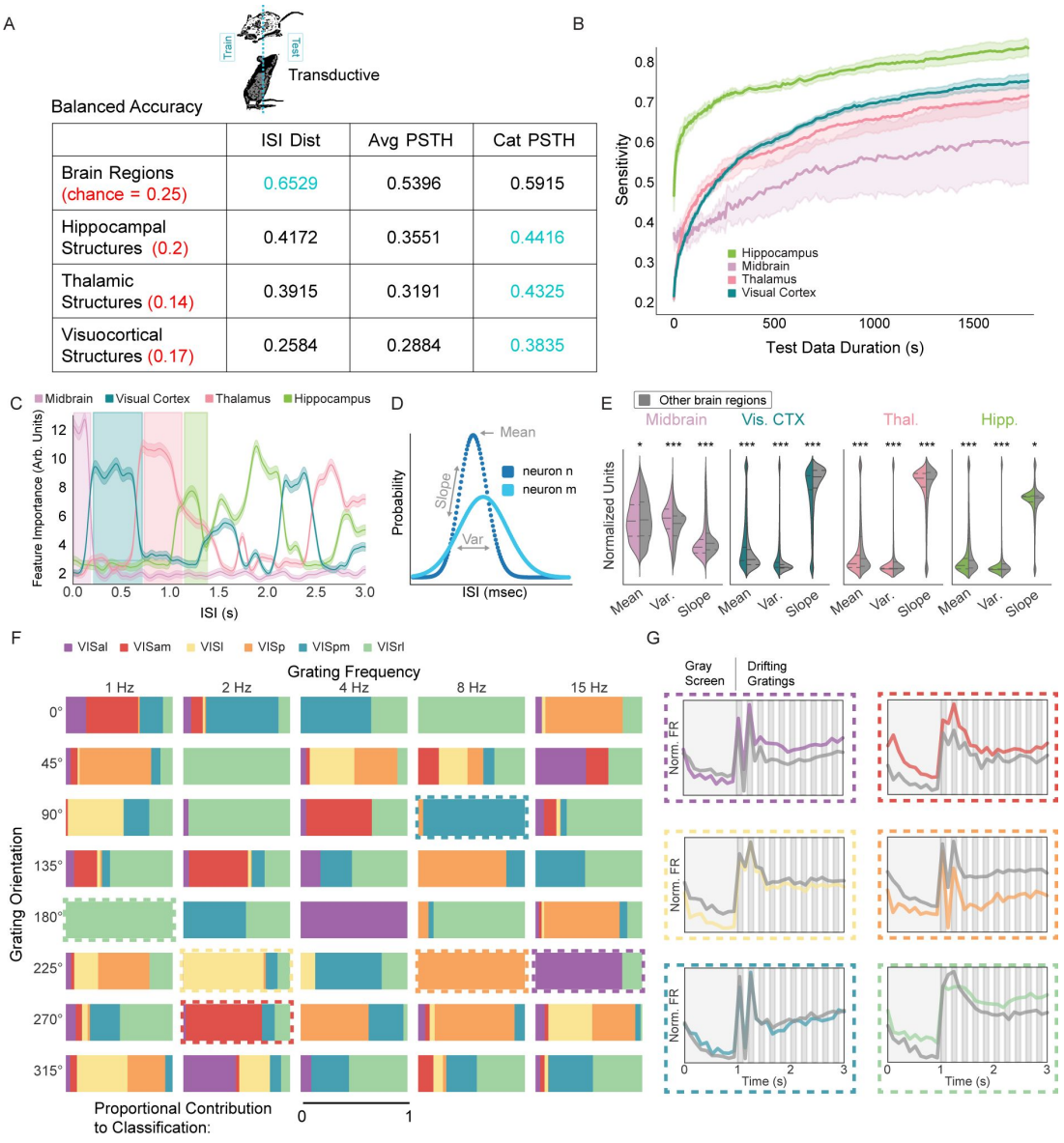
The 14 spiking metrics were selected based on prior literature [15]. However, spike timing can vary along an enormous number of axes, many of which are not indicated by preselected features. To test the possibility that non-parametric, data-driven representations of a neuron's spiking might contain additional, valuable information regarding anatomy, we considered the full collection of each neuron's interspike intervals (ISIs), i.e., the time between two consecutive spikes. Prior work suggests the increased complexity of the ISI distribution is best captured by a nonlinear classification model [15]. Thus, we employed a multilayer perceptron (MLP). MLPs are relatively simple class of feedforward artificial neural network. Briefly, an MLP consists of an input layer, one or more hidden layers, and an output layer. Each layer is composed of interconnected nodes (artificial neurons) that process and transmit information via weighted connections and nonlinear activation functions. This enables MLPs to learn and model complex relationships between input features and output targets.

Broadly, MLPs were more successful in extracting anatomical information from spike trains than logistic regressions trained on spiking metrics (**Fig. 3A**). Specifically, MLPs were trained transductively (which enables transfer of patterns learned from neurons to unseen neighbors) on each of three arrangements of spike times: 1) the entire distribution of ISIs generated during the presentation of drifting gratings (0-3 sec in 100 uniform bins), 2) each neuron's average PSTH across all stimuli—essentially describing a cell's mean response to stimuli, and 3) the concatenated mean PSTHs for each of the 40 combinations of drifting grating orientation and frequency. Across all four classification tasks (brain regions and three sets of structures), MLPs trained on the ISI distribution and concatenated PSTHs outperformed logistic regression models. Conversely, MLPs trained on the average PSTH (averaged across all drifting gratings without respect to orientation or frequency) exhibited reduced balanced accuracy (**Fig. 3**). Together, these data suggest that the embedding of anatomical information in a neuron's activity is recoverable when some form of precise information about spike time is available, and that anatomical information is degraded by averaging.

Due to their reliance on hidden layers, MLPs are prone to the black-box effect; deep models often offer limited insight into the underlying mechanisms of their performance. To elucidate the features of ISI distributions and PSTHs that MLPs use to infer a neuron's anatomical location, we strategically manipulated three features of our data: 1) we varied input duration, 2) we permuted specific bands of the ISI distribution, and 3) we permuted neuronal responses to specific drifting grating orientation/frequency combinations.

We passed progressively reduced lengths of input data, ranging from 1,800 seconds down to 0.1 seconds, into the fully trained model and evaluated its sensitivity (true positive rate). This approach allowed us to assess the timescale of the anatomically informative patterns learned by the MLP. If short intervals of data reliably indicated anatomy, it would suggest that models learn to detect acute and reliable signatures. Conversely, if model sensitivity increased monotonically with time, it would indicate that anatomical information is temporally distributed and cumulative. Our manipulation of the test set duration robustly supported the latter hypothesis; sensitivity for all four brain regions increased as input data duration increased. In all conditions, the slope of the input duration versus sensitivity line was still positive at 1,800 seconds (**Fig. 3B**).

ISIs vary extensively, from the biophysical minimum of the absolute refractory period to many seconds of silence. To test whether anatomical information might be enriched in specific temporal ISI windows, we assembled each neuron's entire ISI distribution (0 - 3 s in 10 ms bins) and selectively permuted subsets of intervals before passing the distribution to the trained MLP for classification. We shuffled ISI counts across neurons within specific temporal windows (e.g., ISIs



	ISI Dist	Avg PSTH	Cat PSTH
Brain Regions (chance = 0.25)	0.6529	0.5396	0.5915
Hippocampal Structures (0.2)	0.4172	0.3551	0.4416
Thalamic Structures (0.14)	0.3915	0.3191	0.4325
Visuocortical Structures (0.17)	0.2584	0.2884	0.3835

Figure 3

Anatomical information in spike trains is captured by nonlinear models that learn patterns in ISIs and stimulus-specific responses.

A. Average balanced accuracy of transductive multi-layer perceptrons (MLPs) in classifying unit location in each task (rows) based on three representations of the spike train (columns). Chance is red, and peak balanced accuracy is blue. ISI dist—full distribution of ISIs; Avg PSTH—mean peristimulus time histogram across all trials; Cat PSTH—concatenation of PSTHs from all 40 stimuli (see methods). **B.** MLP sensitivity as a function of test data duration. Model was trained normally and tested on varying amounts of data from each of the four brain regions. **C.** Feature importance from models that classified anatomy based on ISI dist. Features are ISI ranges between 0 and 3 s in 10 msec bins. 5 splits and 100 iterations. Error is ±1 SEM across 100 shuffles. High value ranges for each region are highlighted. **D.** Illustration of ISI mean, slope, and variance. **E.** Regional distributions of mean, slope, and variance within the highlighted range (in C) compared to averages from all other regions within that range (gray). Multiple comparisons corrected t-tests: $p >= 0.05$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$). **F.** Visuocortical structure information is enriched in subsets of stimuli. Each rectangle represents one of the 40 stimulus parameter combinations. Colors correspond to visuocortical structures, and the area of color shows the relative importance of that stimulus to model classification of the given structure. Cat PSTH model. Dashed boxes—exemplar stimulus conditions. **G.** PSTHs corresponding to structure/stimulus pair indicated in F. Specific structure PSTH shown in color. Average PSTH of all other structures shown in gray.

between 50 and 60 ms). This approach tested whether the MLP's performance depended on viewing the entire ISI distribution or was enriched in a subset of patterns. Consistent with temporal tuning, certain subsets of ISIs were more important than others for MLP performance, with the critical time windows varying by brain region (**Fig. 3C**). For instance, very fast ISIs (<~ 150 ms) were particularly valuable for identifying midbrain neurons, while ISIs of 200 - 600 ms made an outsized contribution to visual cortical classification.

To gain insight into how regional information could be enriched within these ISI ranges, we examined ISI distribution properties within each identified range, contrasting the region preferentially embedded in that range with all others. Across every region and range, there were subtle but statistically significant differences in the mean, variance, and slope of distributions (**Fig. 3D,E**). This suggests the presence of coarse features in regionally-relevant spiking patterns but does not rule out the contribution of a combination of multiple ISI ranges for reliable embedding.

Brain region, the coarsest anatomical classification task, was most effectively extracted from broad ISI distributions (65.29% Balanced Accuracy vs. 59.15% concatenated PSTH vs. 52.91% in spike metrics-based logistic regression, chance = 25%; **Fig. 3A**). In contrast, the discriminability of smaller and more numerous structures was greatest when MLPs were trained on detailed spike timing information (i.e., concatenated PSTHs; **Fig. 3A**, **S2**). Somewhat intuitively, the most robust embedding of visual cortical structures was obtained in conjunction with visual stimulus information (38.35% Balanced Accuracy vs. 25.84% ISI distribution vs. 24.09% in spike metrics-based logistic regression, chance = 17%; **Fig. 3A**, **S2**).

We next asked whether structure information might be conveyed by differing responses to specific stimuli. Analogous to our methods for ISIs, we shuffled the values of particular PSTHs (with specified orientation and frequency) across neurons prior to passing PSTHs to the trained MLP for structure classification. We found that information about each of the visuocortical structures was embedded in the response to multiple stimulus variants. In other words, the MLP learning of visuocortical structure was not characterised by responses to a single set of stimulus parameters (**Fig. 3F**). However, despite the distributed nature of structure information, some stimuli carried more predictive value than others (**Fig. 3G**). Interestingly, despite PSTHs being defined by the presentation of visual stimuli, similar improvements were observed in hippocampal and thalamic structure tasks (Hippocampal structures: 44.16% Balanced Accuracy vs. 41.72% ISI distribution vs. 44.10% in spike metrics-based logistic regression, chance = 20%; Thalamic structures: 43.25% Balanced Accuracy vs. 39.15% ISI distribution vs. 37.14% in spike metrics-based logistic regression, chance = 14%; **Fig. 3A**, **Fig.S2**). This finding is noteworthy because, with the exception of the LGN, none of the constituent structures are primarily associated with visual processing.

Anatomical codes generalize across animals and conditions

Our results demonstrate that anatomical information is embedded in the spike trains of single neurons. However, thus far, our results employ transductive models (which can learn patterns for unseen neurons from their immediate neighbors in the same animal). This raises the possibility that MLPs learned to identify animal-specific spiking patterns that do not generalize across individuals. As an example, if hippocampal neurons from animal *n* happen to be characterized by a specific rhythm, it is reasonable to assume that their neighbors, who happen to be withheld for the test set, exhibit similar activity. Transductive models are agnostic to whether such a pattern is anatomically meaningful and simply a one-off correlation that happens to be localized. To disambiguate these possibilities, we adopted an inductive approach in which the train- and test-sets are divided at the level of the entire animal. In other words, patterns are learned in one set of animals, and tested on another group of animals that is entirely new to the model (**Fig. 1A**).

We trained and tested inductive models on the four classification tasks— brain regions, hippocampal structures, thalamic structures, and visuocortical structures– across two input conditions, general ISI distributions and concatenated PSTHs. The null hypothesis is that

anatomical information generalizes within the animal but not to new animals. In support of the alternate—that anatomical embedding is a universal feature in the spike train–inductive models performed significantly above chance in seven out of eight conditions, and were statistically indistinguishable from the performance of transductive models for all four ISI-based tasks. In the four concatenated PSTH tasks, inductive models exhibited significantly lower balanced accuracy than transductive models, although they were still above chance in all but the hippocampal structures task (**Fig. 4A**). These results suggest that stimulus dependent representations of anatomical location are predominately specific to the animal, while ISI distribution-based anatomical information is general across animals.

Single neuron classification errors can be corrected by population context

While the ISI-based inductive models demonstrate that generalizable anatomical information is carried in single unit spike trains, classification is imperfect. Note that the model classifies individual neurons via a winner-take-all approach. Even for correctly labeled neurons, the probability for the chosen class only needs to narrowly exceed the probability of other options. We next asked to what extent such uncertainty could be mitigated if the consensus was taken across a group of neurons, analogous to how the brain processes noisy information [40, 41, 42]. One possibility is that uncertainty is driven by shared noise amongst neighboring cells. Errors of this form could be amplified by a consensus vote. Alternatively, if incorrect classifications are stochastically distributed across a recording array, errors should be trivially correctable by considering the surrounding ensemble—a stray CA1 neuron should not be detected in the middle of CA3, for example. To test this, we added a second step following single unit classification. This step comprised the use of a Gaussian kernel to smooth anatomical probabilities across neurons recorded on neighboring electrodes.

In support of the stochastic error hypothesis, smoothing dramatically increased the balanced accuracy of MLPs in all four tasks. The brain region identification task improved from 65.10 ± 1.77% to 89.47 ± 2.98% compared to non-smoothed inductive models. The smoothed regional prediction probabilities across an individual probe from an example animal is shown in **Fig. 4B**. Similarly, the hippocampal task improved from 35.89 ± 1.42% to 51.01 ± 4.50%, the thalamic task improved from 32.79 ± 2.37% to 53.21 ± 7.59%, and the visuocortical task improved from 25.52 ± 0.73% to 38.48 ± 3.31% (**Fig. S5**, **S6**). This suggests that erroneously labeled neurons do not share anatomically-relevant spike patterns with nearby cells, and thus are amenable to consensus-based correction.

To this point, MLPs were trained and tested only on neurons related to an individual task. For example, MLPs trained on visuocortical structures were never exposed to thalamic neurons. To evaluate inductive models across the full set of regional and structural comparisons while capitalizing on smoothing of errors, we trained a hierarchical model that combined smoothed brain region classification with subsequent smoothed structure identification. Across all 19 labels (18 structures and midbrain), the hierarchical inductive model achieved a balanced accuracy of 46.91 ± 1.90% (**Fig. 4C**). An observable effect of a hierarchical approach is that, because smoothed brain region classification is so effective, errors generally occur only between similar structures (VISl vs. VISal) rather than highly divergent ones (VISl vs. CA1). Here, only 5 out of the total 250 possible cross-regional errors occurred above chance (**Fig. 4C**). These data demonstrate that, with smoothing, a hierarchical model can extract surprisingly effective anatomical structure information from single neuron spike timing that generalizes to unseen animals. This further suggests that there is a latent neural code for anatomical location embedded within the spike train, a feature that could be practically applied to determining the brain region of a recording electrode without the need for post-hoc histology.

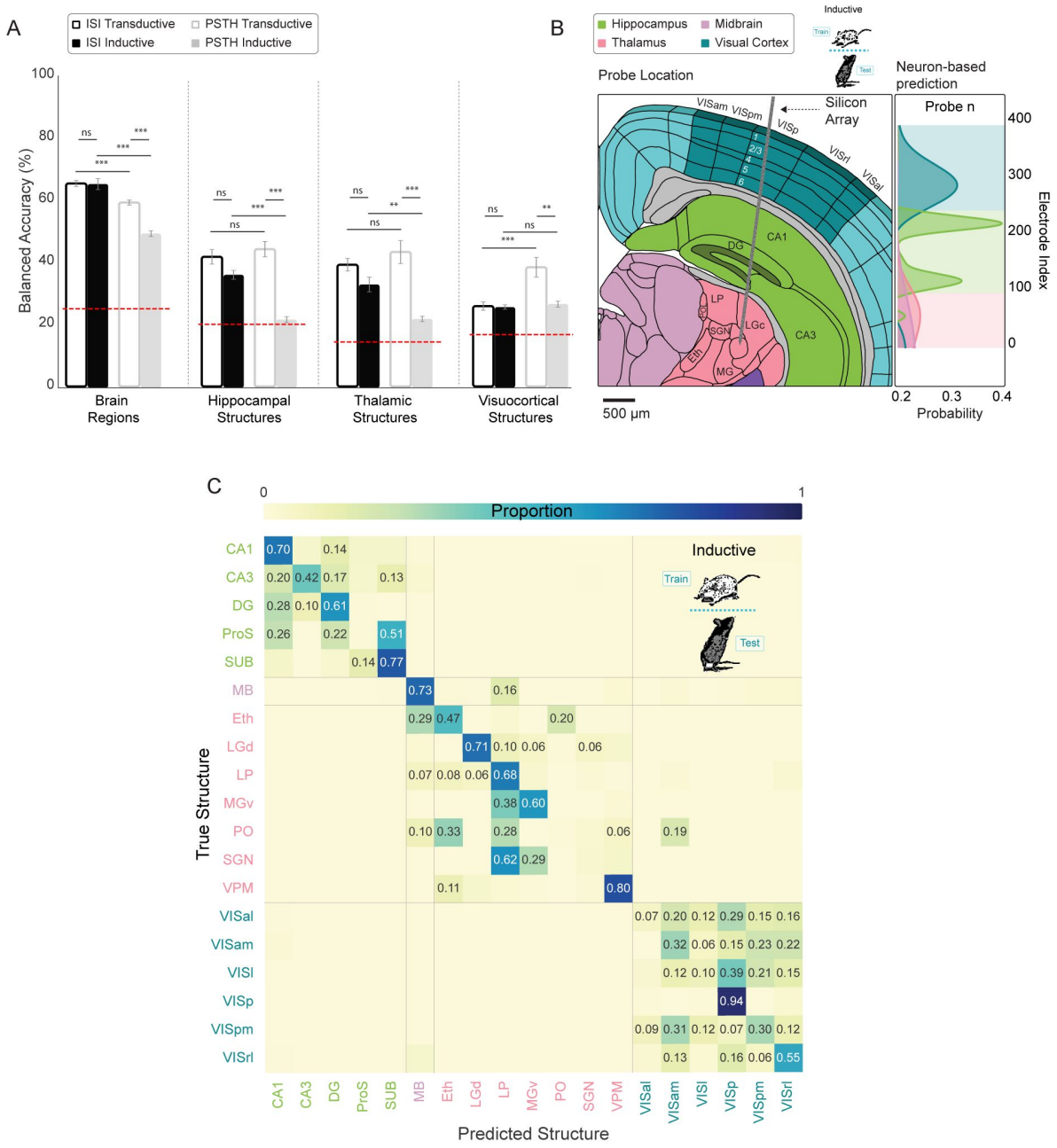


Figure 4

Spike time features that predict anatomy generalize across animals.

A. Balanced accuracy (± SE) of multi layer perceptrons (MLPs) trained on ISI distributions and concatenated PSTHs in both inductive (withhold entire animals for testing) and transductive splits. Chance (red dashed line) varies by task. Linear mixed effects regression (ns/not significant: $p >= 0.05$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$). **B.** Left: Illustration of an example implanted silicon array spanning isocortex, hippocampus, and thalamus. Right: Brain region probability for the example implant shown on the left calculated by smoothing across neuron-based classifications. Colored background shows the consensus prediction as a function of neuron location (electrode number). **C.** Confusion matrix resulting from hierarchical (region then structure) inductive classification with smoothing. Matrix cells with proportion less than chance (1/number of classes) contain no text. Average balanced accuracy after smoothing (by task): Brain Regions = 89.47 ± 2.98%; Hippocampal Structures = 51.01 ± 4.50%; Thalamic Structures = 53.21 ± 7.59%; Visuocortical Structures = 38.48 ± 3.31% (error is the SEM across 5 splits). Overall balanced accuracy: 46.91± 1.90%.

Spike train embedding of visual superstructure and cortical layer

Even with smoothing, visuocortical structures remained the most challenging to classify. There are two possible explanations for this. First, it may be that there are truly no consistent differences in the organization of single unit spiking across visuocortical structures. Alternatively, there are true differences but our MLP-based approach fails to learn them. Although there are clear cytoarchitectural boundaries between primary visual cortex (VISp) and secondary visual areas, the differentiation of secondary visual areas is functionally determined [43, 44]. We asked whether broadly defined superstructures, VISp and VISs—a combination of all secondary areas (VISal, VISam, VISl, VISpm, and VISrl)–increases the effectiveness of spike time-based classification (**Fig. 5A**). We trained an MLP on VISp versus VISs, and, in this binarized context, observed a balanced accuracy of 79.98 ±3.03 % (**Fig. 5B**). This effect which was driven by VISs, which achieved 96 % sensitivity. To understand if converting six visual structures into two superstructures truly improved discriminability, it is necessary to make the equivalent comparison in the more complex model. To achieve this, we simply took the results of the visuocortical task across 6 structures (**Fig. S6**, bottom left) and collapsed the secondary regions into a single class. Intriguingly, this yielded a balanced accuracy of 91.02 ±0.95%, driven by VISp at 94.4 % sensitivity. A likely explanation is that, despite using resampling strategies to correct for class imbalance, VISp initially has more true units than any other structure in the 6-class task, incentivizing accurate classification of it. In contrast, for the binary classification, VISp has fewer units than all secondary structures combined (VISs), which instead incentivises classification of VISs. It seems VISp is a more effective classification target for this problem. These results support the notion the meaningful computational division in murine visuocortical regions is at the level of VISp versus secondary areas.

Another major axis along which cortical neurons exhibit anatomical structure is layer. This aligns with recent evidence that indicates that cortical layers can be distinguished computationally, particularly in terms of cell types and tuning preferences [15, 45, 46]. Thus, we hypothesized that cortical layer might be more reliably embedded in single unit spike times than visuocortical structure. Consistent with this, models trained on visuocortical neurons from all structures were able to robustly recover layer information (**Fig. 5C,D**). It is noteworthy that transductive and inductive models achieved similar balanced accuracies on both ISI distribution (inductive unsmoothed balanced accuracy: 46.43 ±0.97%, inductive smoothed: 62.59 ±1.13%, transductive unsmoothed: 46.52 ±0.63%, transductive smoothed: 60.37 ±2.64%) and concatenated PSTHs (inductive unsmoothed: 41.13 ±1.28%, inductive smoothed: 52.16 ±2.38%, transductive unsmoothed: 41.66 ±0.51%, transductive smoothed: 52.35 ±1.69%). However, cortical layer information was more robust in the broad ISI distribution, which outperformed concatenated PSTHs, even in transductive models. Also noteworthy is the fact that layer IV exhibited the greatest confusion (specifically, with adjacent layers), achieving a sensitivity of only 28%. These results suggest that general embeddings are more readily available for cortical layers than cortical structures.

Anatomical embeddings generalize across experimental conditions

Drifting gratings, while widely embraced across decades of research in the visual system, are highly stereotyped and low dimensional compared to natural visual environments. As a result, the anatomical embeddings described thus far could require the repeated presentation of drifting gratings. In other words, it is reasonable to suggest that anatomical information embedded in single unit activity would be immediately obscured by a complex visual environment. To evaluate this, we trained new, inductive MLPs on single unit activity in the context of drifting gratings as well as two additional conditions: 1) the presentation of naturalistic movies (ten repeated presentations of a 120 s long excerpt from the film *Touch of Evil*), and 2) spontaneous activity (an unchanging gray screen) (**Fig. 6**). We then tested these MLPs on activity from all three

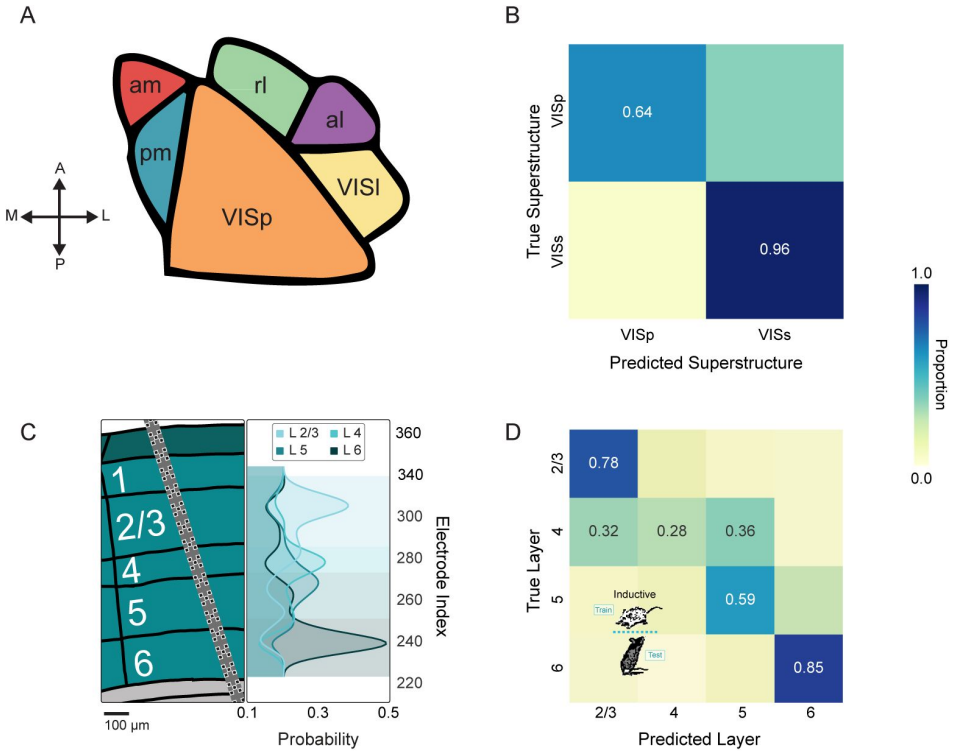


Figure 5

Primary vs secondary distinction and cortical layer are more evident in spike timing than individual structures.

A. Illustration of visuocortical structures that can be grouped into primary versus secondary superstructures: VIS-am, -pm, -l, -al, -rl are grouped into secondary visual cortex (VISs) while VISp is primary visual cortex. **B.** Confusion matrix resulting from inductive classification of superstructure (with smoothing). Cells with proportion less than chance (1/number of classes) contain no text. Balanced accuracy is 79.98 ±3.03% **C.** Left: Illustration of example array implanted across cortical layers. Right: Layer probability for the example implant shown on the left calculated by smoothing across neuron-based classifications. Colored background shows the consensus prediction as a function of neuron location (electrode number). **D.** Confusion matrix resulting from smoothed inductive classification of layer. Balanced accuracy is 62.59 ±1.12%

conditions. This allowed us to ascertain a) whether anatomical information is available outside of the context of drifting gratings, and b) whether the anatomical information learned in one condition can generalize to another.

We tested every pairing of train/test condition on each of 6 tasks: brain regions, hippocampal structures, thalamic structures, visuocortical structures, visuocortical layers, and visuocortical superstructures. Chance levels varied between 14.3 and 50 % across tasks. To facilitate inter-task comparisons, we quantified accuracy by employing Matthews Correlation Coefficient (MCC). MCC is a balanced measure that takes into account true and false positives and negatives, providing a reliable statistical measure especially for imbalanced datasets [47, 48]. MCC values range from −1 to +1, with +1 representing perfect prediction, 0 indicating performance no better than random chance, and −1 signifying total disagreement between prediction and observation. Importantly, MCC is normalized by the number of categories, allowing for comparisons across tasks with different numbers of classes [49].

MLPs across almost every task and every pairing of train/test condition performed far above chance (MCC chance = 0) (**Fig. 6**, left column). MLPs tasked with brain region showed remarkable generalizability across stimuli (mean MCC = 0.90). This was followed by hippocampal structures (mean MCC = 0.46), visuocortical layers (mean MCC = 0.41), thalamic structures (mean MCC = 0.30), visuocortical superstructures (mean MCC = 0.26), and visuocortical structures (mean MCC = 0.18). Interestingly, the only instances of chance-level performance arose in the visuocortical superstructure task—MLPs chance (MCC = 0) when trained on drifting gratings and tested on either spontaneous activity or naturalistic movies. While this appears consistent with stimulus-specific embeddings, visuocortical MLPs trained on spontaneous activity were above chance when tested on drifting gratings. The same was true for those trained on naturalistic movies. Taken together, these data suggest that the embedding of anatomical information in single neuron activity is not abolished by complex visual stimuli or spontaneous activity, and that the embeddings learned in one context are not absent in other contexts. The visuocortical superstructure task results imply that, in some contexts, complex stimuli may produce more generalizable results than simple stimuli.

In each of the 54 combinations of task and train/test condition, there were diverse underlying patterns of MLP learning (**Fig. 6**, right column). Intuitively, instances with high MCC, such as those in the brain regions task, produced strong diagonal structure (where the predicted class aligns with the true class). Tasks resulting in lower MCC, such as the visuocortical structures, yielded slight diagonal structure in some cases, but were driven by a small number of accurate points. Across all six tasks, within-stimulus models (e.g., trained on spontaneous, tested on spontaneous) tended to produce significantly greater MCCs than across-stimulus models, although all were above chance (**Fig. S7**).

Anatomical embeddings generalize across research laboratories

The results of inductive models trained and tested in mismatched conditions suggest that, amidst stimulus information, neuronal spike trains carry universal signatures of their anatomical location. If this were true, it should be possible to apply models trained on animals from one research group—in this case, the Allen Brain Observatory team— and decode neuronal anatomy from data generated at an independent location. In other words, the Allen-based model should predict a neuron's anatomical location based on its spike train even if the recording was conducted in a different location and experimental context.

To test this, we required independently generated data that maintained two key features; 1) the spatiotemporal micro-scale resolution of the Allen Institute's recordings, and 2) the anatomical breadth of the Allen Institute's recordings. These criteria were satisfied by an open dataset from Steinmetz et al. [50], which comprise high density silicon recordings that span many of the same regions and structures examined in the Allen data. However, the Steinmetz et al. [50]

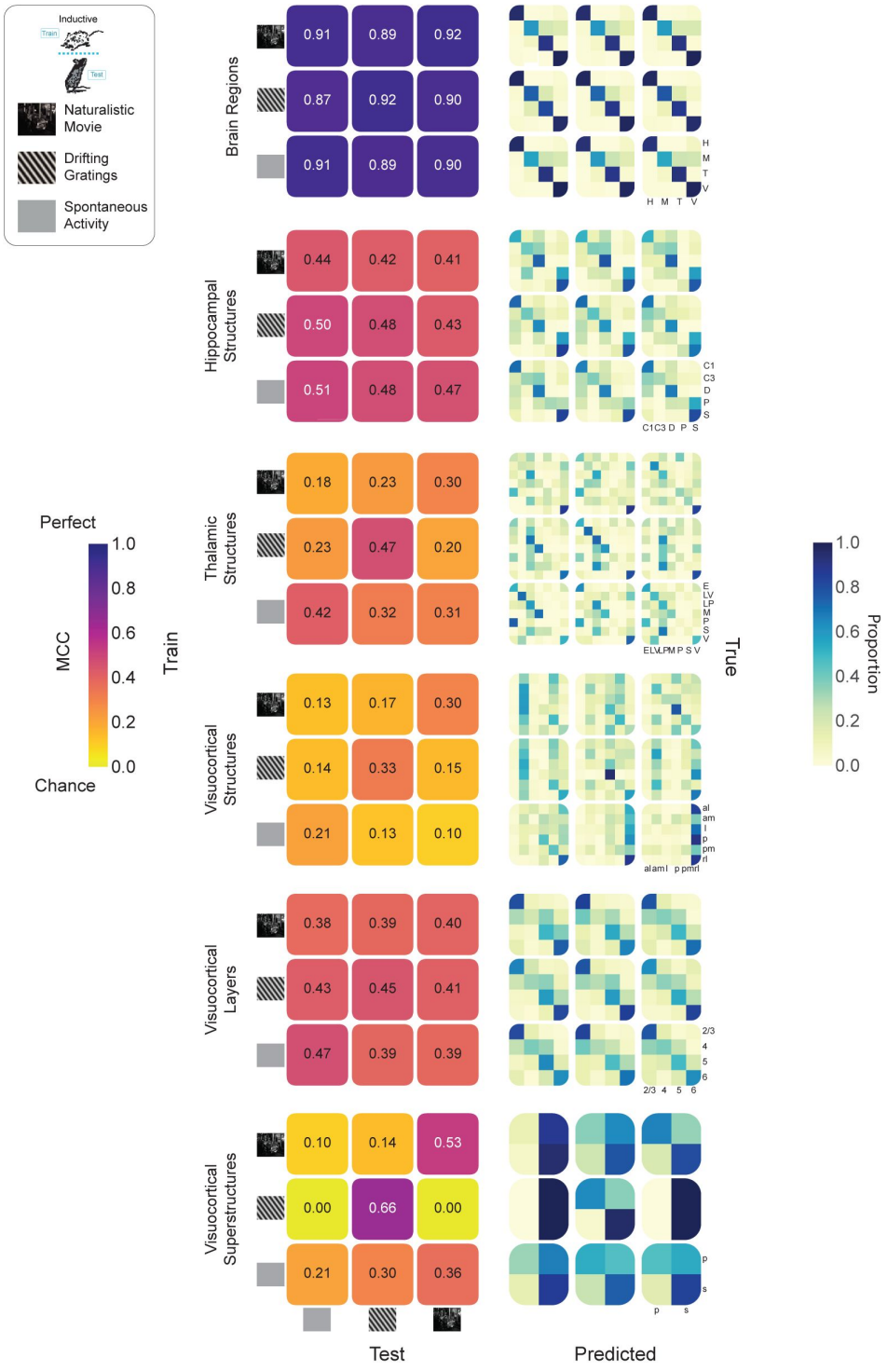


Figure 6

Anatomical information embedding in spike trains generalizes across diverse stimuli.

(Top Left) Schematic showing inductive train/test split along with visual stimuli: naturalistic movie, drifting gratings, spontaneous activity (i.e. gray screen). (Left Column) Grids showing Matthew's Correlation Coefficient (MCC) values for pairs of training stimuli (grid rows) and testing stimuli (grid columns). Grid diagonals (top right to bottom left) represent train/test within the same stimuli. (Right Column) Confusion matrices corresponding to the each of the MCC grids in the left column.

experiments are markedly different, in that they comprise mice trained to carry out a decision making task. Summarily, the task involved observing a Gabor patch of fixed orientation (45°) and spatial frequency (0.1 cycles per degree) with varying contrast on either the left or right side, in response to which the mouse must turn a wheel toward the side with higher contrast[50]. Recall that the Allen data were recorded in the context of passive viewing (**Fig. 7A**).

We classified anatomy in Steinmetz et al. [50] data using two variants of our Allen-based models; one trained on ISI distributions recorded during the presentation of drifting gratings, and one trained on ISI distributions recorded in three conditions (drifting gratings, naturalistic movies, and spontaneous activity). Both models successfully predicted brain region above chance in every Steinmetz et al. animal (N = 10), with the exception of one animal in the drifting gratings model (mean balanced accuracy calculated across neurons, drifting gratings model—80.46%; combined stimuli—81.28%; **Fig. 7B**). At the level of structures, two general principles emerged: 1) models trained on combined stimuli were generally more effective than drifting gratings models— this is particularly evident when examining diagonal structure in the confusion matrices (**Fig. 7B,C**, **S8**), and 2) the visuocortical structures task did not transfer effectively between laboratories (drifting gratings model—28.41%, mixed model—28.34%). Hippocampal and thalamic structures as well as visuocortical layers were identifiable well above chance, especially in mixed models (drifting gratings/mixed models: hippocampal structures—40.07 / 69.89%; thalamic structures—21.52 / 58.22%; visuocortical layers—46.36 / 49.01%). Visuocortical superstructures were marginally identified, but only in the combined stimuli model (drifting gratings model—58.03%, mixed model—59.08%), as only one animal was above chance in the drifting gratings condition (**Fig. 7B**). Note that, due to differences in the recordings and number of units from the structures, each Steinmetz et al. task involved 4 classes, such that chance is 25 % with the exception of visuocortical structures (k = 5, chance = 20%) and visuocortical superstructures (k = 2, chance = 50%). Finally, error is not reported as there are no splits to test across in the Steinmetz et al. data; balanced accuracy is reported for the entire applicable dataset.

Taken together, these data suggest that information about a neuron's anatomical location can be extracted from the time series of its spiking. This principle appears to apply to neurons from diverse structures across the telencephalon, diencephalon, and mesencephalon. Further, some of the rules by which this information is organized are shared across animals, and are not an artifact of a task or local protocol.

Discussion

Understanding the information carried within a neuron's spiking has been the subject of investigation for a century [1]. While it is well established that neuronal activity encodes stimuli, behavior, and cognitive processes [4, 5], the possibility that spike trains might also carry information about a neuron's own anatomical identity has remained largely unexplored. Our study provides compelling evidence that such anatomical information is indeed embedded within neuronal spike patterns. Using machine learning approaches, we demonstrate that this embedding is robust across multiple spatial scales, from broad brain regions to specific structures and cortical layers. Crucially, these anatomical signatures generalize across animals, experimental conditions, and even research laboratories, suggesting a fundamental principle of neural organization. Our findings reveal a previously unrecognized dimension of the neural code, one that is multiplexed with the encoding of external stimuli and internal states [51, 52, 53]. These data advance an understanding of the relationship between structure and function in neural circuits, raising the possibility that structure is not ancillary to neuronal computation, but intrinsically embedded within it. Beyond fundamental scientific insight, our findings may be of benefit in various practical applications, such as the continued development of brain-machine interfaces and neuroprosthetics, as well as for the interpretation of large-scale neuronal recordings. Note, however, that a purely utilitarian goal, e.g., electrode localization in

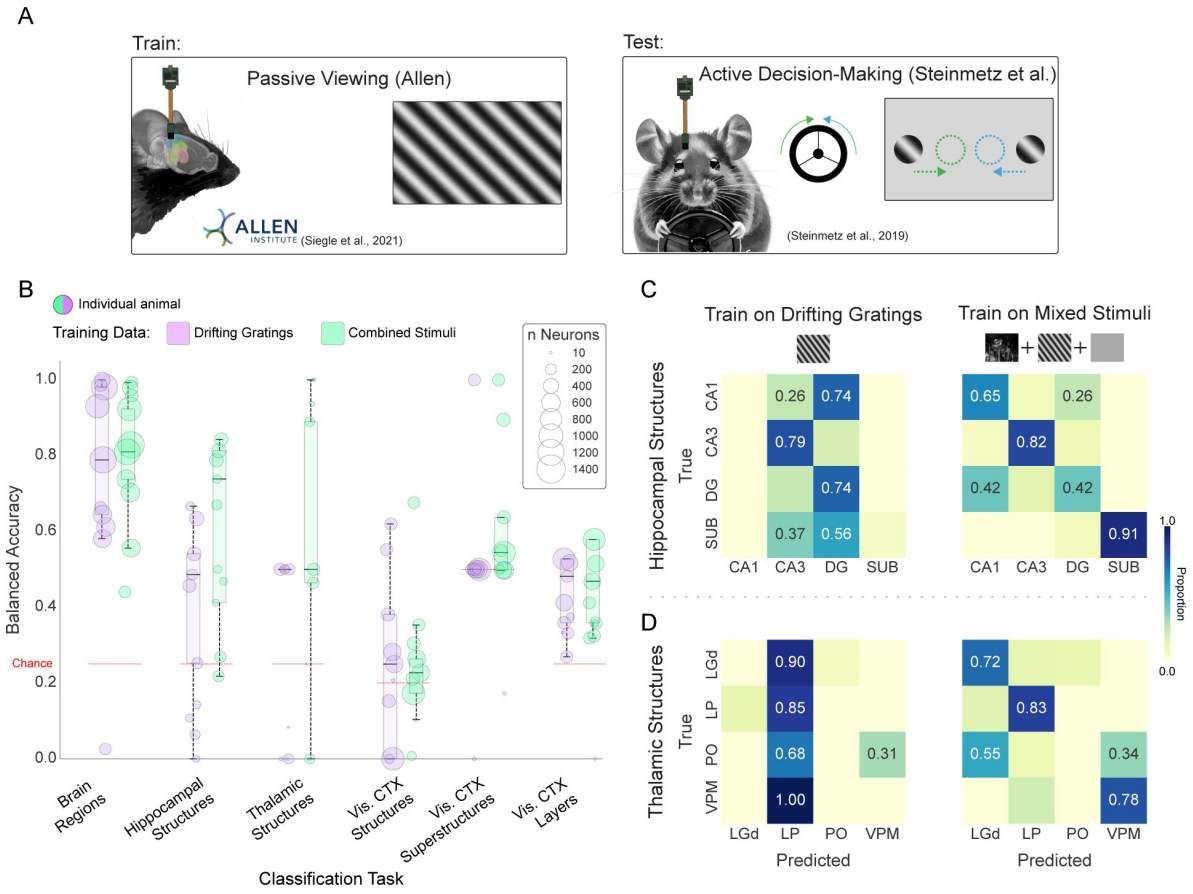


Figure 7

Anatomical information embedded in spike trains generalizes across laboratories and protocols.

A. Illustration of behavioral tasks employed by two laboratories: (left) Passive Viewing (Allen Institute) vs. (right) Active Decision-Making (Steinmetz et al.). In Active Decision-Making, the mouse spins a wheel in response to the location of the drifting grating presented on the screen. **B.** Bubble chart showing balanced accuracy of individual test set animals from the Steinmetz et al. The model was trained on Allen Institute data. Bubble size indicates the number of units recorded in the test animal. Black horizontal lines represent the median of the balanced accuracy distributions across the animals. Models are trained with either drifting gratings alone (purple) or a combination of drifting gratings, natural movies, and spontaneous/gray screen (green). Balanced accuracy across all Steinmetz et al. neurons: BR DG=80.46%, BR Mix: 81.28 %, HS DG: 40.07 %, HS Mix: 69.89 %, TS DG: 21.52 %, TS Mix: 58.22 %, VCS DG: 28.41 %, VCS Mix: 28.34 %, VCSS DG: 58 %, VCSS Mix: 59 %, VC Layers DG: 46.36 %, VC Layers Mix: 49.01 % where BR stands for brain regions, HS hippocampal structures, TS thalamic structures, VCS visual cortex structures, VCSS visual cortex superstructures, DG drifting gratings and Mix denotes the combined stimuli. **C.** Confusion matrices for prediction of hippocampal structures in Steinmetz et al. test set when trained on drifting gratings stimulus (left) vs. mixed stimuli (right) from the Allen Institute. **D.** Confusion matrices for prediction of thalamic structures in Steinmetz et al. test set when trained on drifting gratings stimulus (left) vs. mixed stimuli (right) from the Allen Institute.

electrophysiological recordings, would be well served by considering additional features, such as extracellular waveforms [54, 55], in addition to spike timing. While this approach holds promise for practical application, the inclusion of waveform information subverts the question of whether a neuron's output—the timing of its spiking–contains an embedding of its location.

Clearly, there are powerful differences throughout the brain as a function of anatomy. There is extensive literature to describe this diversity at many levels, from gene expression gradients in hippocampal pyramidal neurons [35] to distinct connectivity patterns across cortical layers [24]. A complementary albeit smaller literature provides some indication that, at a coarse-grained level, similar gradients can be identified in neuronal activity. For example, examined at the level of the entire population, there are subtle but reliable differences in isocortical neuronal variability in functionally distinct cortical areas [27]. Similarly, considered as a population-level statistic, there is a cortical gradient of neuronal intrinsic timescale (a measure of how long a neuron's activity remains correlated with itself over time) [21]. However, while the ability to detect statistical differences across a population demonstrates that anatomy bears some influence on brain activity, it does not suggest that anatomy can be extracted from a single neuron's activity. This is exemplified by placing two observations side-by-side. First, single neuron firing rates are log-normally distributed. This means that within any given brain region, there is a wide range of firing rates, with a long tail of high-firing neurons. Second, the population-level distribution of firing rates varies slightly but significantly as a function of region [56, 57]. The combination of these facts implies that while there may be detectable differences between regions at the population level, the extensive overlap in firing rate distributions makes it challenging, if not impossible, to determine a neuron's anatomical origin based solely on its firing rate.

Here, we approach this problem from a fundamentally different perspective than the population-based paradigm. Capitalizing on 1) the availability of open, high density recordings across a plurality of brain regions, and 2) machine learning methods capable of learning complex, nonlinear relationships, we approach the problem at the level of single neuron classification. This approach has strengths and weaknesses. The principal strength is the revelation that anatomical information is embedded in the activity of individual neurons throughout the brain. The principal weakness is limited interpretability, a criticism widely raised when contemplating machine learning (even when addressing this issue, the ground-truth patterns can themselves be complex and non-intuitive, e.g.,**Fig. 3**). However, in exchange for simplicity, our approach is founded on establishing whether there exist spike train-based anatomical signatures that generalize across animals, experimental conditions, and research laboratories. That such signatures are readily identifiable indicates neurons are not mere conduits of stimulus-related information. While neurons can be described as ratevarying Poisson processes [7], it is increasingly clear that a more complete description should incorporate additional streams of information, such as transcriptomic cell type and anatomy across varying time scales [15, 58].

Our findings reveal a striking difference in the generalizability of anatomical information encoded in ISI distributions versus PSTHs. Specifically, inductive models trained on ISI distributions maintained performance levels comparable to their transductive counterparts across all tasks, while PSTH-based models showed a significant drop in performance when tested on new animals. This is perhaps surprising, given that stimulus response properties (such as receptive field size and response latency) are generally understood to exhibit some consistency across animals recorded under equivalent conditions [59, 60, 18]. Viewed through the lens of stimulus and response, PSTHs in the Allen Institute datasets are robust to slight differences in retinotopic locations across animals [18]. Despite this, our PSTH-based models performed poorly when trying to predict anatomical location in new animals under the same stimulus conditions. Thus, it is likely that stimulus-response information is semi-orthogonal to the embedding of anatomical location.

The difference in generalizability when comparing PSTH- and ISI-based models may reflect a fundamental distinction in the nature of information captured by these two representations of neural activity. ISI distributions encapsulate intrinsic properties of neuronal firing patterns that appear to be conserved across animals, potentially reflecting stable, anatomy-specific computational features. These might include cell-types and their resultant ion channel compositions [61, 62], local circuit motifs [63, 64], or homeostatic mechanisms that shape firing statistics independent of stimuli [65, 66, 67]. Crucially, these features are not inherently tied to a stimulus. In contrast, PSTHs involve studying the repeated presentation of the same stimulus. Despite this repetition, there is neuron and trial level variability [68, 69]. Neuronal response variability to repeated stimuli often correlates among nearby neurons[70, 71] and is influenced by local circuit connectivity and the animal's cognitive factors such as attention [72, 73]. Thus, transductive models may learn from variability in the training dataset that is highly informative in the test dataset. However, such variability would not carry across animals. The robustness of ISI-based anatomical embeddings across animals and even across laboratories underscores the fundamental nature of these anatomical fingerprints in neural activity, transcending individual differences and specific experimental paradigms.

Our analysis of visual isocortical structures offers intriguing insights into the computational organization of the murine visual system. While PSTH-based models performed well in classifying visual areas, suggesting stimulus-specific differences in neuronal responses across these regions, the overall classification accuracy remained relatively low compared to other brain areas. This finding aligns with previous studies that have demonstrated differences in orientation and spatial frequency tuning across mouse visual areas [74, 75, 76, 77]. However, our results suggest that these differences, while statistically significant at the population level, may not translate into robust, neuron-specific computational signatures. The murky distinction between primary and secondary visual areas in mice is reflected in our data, with the most reliable discrimination occurring between primary visual cortex (VISp) and grouped secondary areas, rather than among individual secondary regions. Even this distinction, however, was not dramatic. Interestingly, we found that isocortical layers within visual areas were more readily distinguishable than the areas themselves, suggesting that laminar organization may play a more fundamental role in shaping neuronal computations than areal boundaries in mouse visual cortex. This hierarchical organization - from broad regions to superstructures (e.g., VISp vs. secondary areas) to substructures (layers) - provides a nuanced view of functional specialization in the mouse visual system. While numerous recent studies have indicated functional specialization of mouse higher visual areas beyond VISp [43, 44, 78], our data suggest that these specializations may not manifest as significant differences in single neuron spiking features across areas. This observation raises the possibility that secondary visual areas in mice may not be computationally distinct, at least in terms of their constituent neurons' fundamental spiking patterns. The relative inability to classify the structure of visual cortical neurons may, in fact, reflect a genuine neurobiological feature: the computational properties of neurons in different visual areas may be largely indistinguishable based on spike timing alone.

It is reasonable to consider the possible mechanism and function of an anatomical information stream. It is not necessary that neurons explicitly impart anatomical location on the structure of spike timing. If all neurons are otherwise identical until cell type, connectivity, and stimulus are considered, then these would neatly describe the mechanism by which anatomy is imprinted in spiking. In this case, our ability to extract anatomical information from the spike train is reflective of cell type and connectivity—each important features of a neuron's contribution to brain function. Independent of its mechanistic original, information about anatomical location, cell type, and/or connectivity in a spike train raises questions about possible utility. In other words, just because it is there, does it matter? The null hypothesis, in this case, is that anatomical information is epiphenomenal, an inevitable byproduct. However, the robustness of anatomical information across stimuli, animals, and even laboratories suggests a consistent selective pressure rather than a local consequence. Consider: recorded neurons represent a severe subsampling of the local

population (< 1%) [79]. Further, recordings are random in their local sampling, each neuron's connectivity is unique [80], and vertebrate brains lack "identified neurons" characteristic of simple organisms [81]. Taken together, it seems unlikely that a single neuron's happenstance imprinting of its unique connectivity should generalize across stimuli and animals. Could neurons use this information? Computational modeling demonstrates that individual neurons could learn to discriminate between complex temporal patterns of inputs [17]. This suggests that neurons could distinguish and selectively respond to inputs carrying different types of information, such as selectively attending to anatomically distinct inputs. The timescale of anatomical location embedding is also consistent with molecular mechanisms by which neurons integrate slow information, e.g., CaMKII [82].

While the technical barriers are daunting, the obvious experiment is to manipulate a neuron's anatomical embedding while leaving stimulus information intact. Should this disrupt any aspect of sensation, perception, cognition, or behavior, the answer would be clear. If not, there is still great practical utility in an experimenter's capacity to ascertain anatomy based on neuronal activity.

Our findings demonstrate that individual neurons embed information about their anatomical location in their spike patterns, with these anatomical embeddings conserved across animals and experimental conditions. This previously unrecognized aspect of neural coding raises questions about the relationship between brain structure and function at the single-neuron level. Future research will be crucial in determining whether and how different streams of information, including these anatomical embeddings, contribute to neural computation and the variegated functions of the brain.

Methods

Datasets

Neurodata Without Borders (NWB) files for the Allen Institute's Visual Coding Neuropixels dataset were retrieved with AllenSDK [83]. Units passing the default filtering criteria of ISI violations < 0.5, amplitude cutoff < 0.1, and presence ratio > 0.9 were selected for further analysis. Units with a firing rate below 0.1 Hz over the session were excluded.

Mice were head fixed and presented with visual stimuli for a roughly three-hour session [18]. The mice were shown one of the two slightly different visual stimulus sets:"Brain Observatory 1.1" (N = 32 animals) or "Functional Connectivity" (N = 28 animals) [84, 85]. From these sessions, we retrieved spike times coincident with the presentation of drifting gratings and natural movie three from "Brain Observatory 1.1" and gray screen (spontaneous) from both sets. For each animal, 30 minutes of drifting gratings were presented. Each trial consisted of one second of gray screen followed by two seconds of drifting gratings. Gratings were varied in spatial orientation (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°) and temporal frequency (1, 2, 4, 8, and 15 Hz), but consistent in spatial frequency (0.04 cycles/degree) and contrast (80%) with 15 equivalent presentations of each particular stimulus. Separately, a sustained mean-luminance blank (gray) screen was presented during intervals between blocks of stimuli. In total, gray screen was presented in this manner for approximately 20 minutes to each animal. Natural movie three consisted of a 120-second clip from the opening scene of the movie Touch of Evil repeated 10 times. After passing through the criteria above, there were 18,961 units for drifting gratings and natural movie three (visual cortex = 9,402, hippocampus = 4,301, thalamus = 4,068, and midbrain= 920) while the spontaneous stimulus contains 37,458 units (visual cortex = 18,514, hippocampus = 10,337, thalamus = 6,625, and midbrain = 1,982). In order to have at least 30 units in the test for each split, we removed classes (regions or structures) from the dataset if they contained less than 150 units. We also removed units with ambiguous structure assignment such as 'VIS' or 'TH'. The final number of units in each

structure (for drifting gratings and natural movie three) is as summarized in **Table 1**. The spontaneous (blank screen) stimulus has more units from the "Functional Connectivity" dataset in addition to the units from the "Brain Observatory."

The dataset provides labels for each unit's brain structure as ecephys structure acronym, determined through a combination of stereotactic targeting, histology, common coordinate framework (CCF) mapping, and intrinsic signal imaging (particularly for visual cortical areas) [18, 85]. For our larger brain region prediction tasks, we grouped the provided brain structures into a higher hierarchy based on the Allen Brain Atlas ontology's structure tree [86]. However, the publicly available units' metadata does not include layer information. To address this for cortical units, we mapped the CCF coordinates of each unit onto a finer-grained parcellation (25 μm) level within the ontology, enabling the extraction of layer labels. There were 1,229 units, 1,601 units, 3,046 units, and 1,057 units in layer 2, 4, 5, and 6 respectively for drifting gratings and natural movie three.

To generalize across experimental conditions, we used a dataset from Steinmetz et al. [50], which generally inspired the experiments used in the International Brain Laboratory (IBL) datasets. The mice were shown drifting gratings (oriented at 45° and spatial frequency of 0.1 cycles per degree) of varying contrast. The tasks were divided into an average of four second trials and involved presentation of auditory cue, wheel turning and reward presentation with inter-trial intervals of gray screen [50]. The experimental sessions also involved passive stimulus presentation after the behavioral sessions [50]. We did not restrict our analysis to any interval, but used all the available spikes in the experiments. The data is collected from 10 mice and 39 sessions with each session potentially resulting in different units from the same mice. The units (or clusters) have structure labels or anatomical locations which was determined by combining electrophysiological features, histology and CCF mapping [50]. To extract layer information, we used the same method as with the Allen Institute's dataset: mapping the CCF coordinates to higher resolution parcellation using the Allen CCF reference space. We observed that some brain structure labels provided in the dataset did not match the coordinate mappings, possibly due to manual adjustments. Since accurate layer extraction required the provided structures to overlap with the CCF-mapped structures, we decided to use only those units where the labels matched. In addition, we removed the structures that were not in the training set, i.e., structures not found in our final Allen dataset. This resulted in 8,219 units with the number of units units in each structure as summarized in **Table 2**.

Unsupervised Analysis

Unsupervised dimensionality reductions across spiking features utilized python implementations of principal component analysis (PCA) [87], t-distributed Stochastic Neighbor Embedding (TSNE) [87], and Uniform Manifold Approximation and Projection (UMAP) [33]. Features to be dimensionally reduced were either a collection of established spiking metrics [15] (see Logistic Regression section for more details), ISI distributions (100 uniform bins between 0 and 3 s; see Multi-Layer Perceptron section), and averaged PSTHs (100 uniform bins over the 3 second span of a trial; see Multi-Layer Perceptron section) for each stimulus condition grouped/concatenated together.

To appropriately tune the hyperparameters of t-SNE and UMAP, we performed a grid search over a broad range of values and selected the hyperparameter combination which maximized the separability of region/structures for a task. Specifically, we selected parameters which yielded optimal training balanced accuracy (no units held out) of a logistic regression classifier on the primary two dimensions of the reduced space.

Table 1

Number of included single units as a function of brain structure: Allen Institute data

Region	Structure	Number of Units
Hippocampus	CA1	2,552
	CA3	367
	DG	713
	ProS	176
	SUB	470
Midbrain	MB	920
Thalamus	Eth	225
	LGd	903
	LP	1,462
	MGv	175
	PO	289
	SGN	178
	VPM	261
Visual Cortex	VISal	1,595
	VISam	1,537
	VISl	971
	VISp	2,076
	VISpm	904
	VISrl	1,467

Table 2

Number of included single units as a function of brain structure: Steinmetz et al. data

Region	Structure	Number of Units
Hippocampus	CA1	1,078
	CA3	511
	DG	548
	SUB	541
Midbrain	MB	108
Thalamus	LGd	134
	LP	395
	PO	619
	VPM	68
Visual Cortex	VISam	1,358
	VISl	535
	VISp	1,535
	VISpm	540
	VISrl	249

To evaluate the significance of region/structure separation in these spaces we compared these optimal balanced accuracies with the balanced accuracies obtained after 1,000 random shuffling of the region/structure labels prior to training the logistic regression. We sought to evaluate whether the relationship (relative distances) between regions/structures was preserved between classes across feature sets and dimensionality reduction methods. First, we created histograms of the distribution of scores along the primary axis of the dimensionality reduction for each region/structure. For each pair of regions/structures we calculated the Wasserstein distance between their histograms. We plotted the results as triangular distance matrices to show that these relationships are largely unconserved across feature sets and dimensionality reduction methods.

Dataset Splitting

To properly optimize our supervised models while ensuring generalizability to unseen data we performed a standard split into train, validation, and test sets (with an approximate 60/20/20 ratio with respect to the number of neurons). To ensure the full dataset was evaluated, we extracted five stratified samples such that each unit was included in the validation set in one of these splits, and separately included in the test set in another split. This is intuitively similar to a 5-fold cross validation.

In a transductive split, for each animal, 60% of the neurons were allocated to train, 20% of the neurons were in the validation set, and 20% were allocated to the test set. We sought to ensure classes (i.e. regions/structures) were represented appropriately in each set. If 20% of the dataset was class A, we attempted to stratify our sets such that ~20% of the train set was class A, ~20% of the validation set was class A, and ~20% of the test set was class A.

In an inductive split, all neurons from 60% of the animals were allocated to train, all neurons from 20% of the animals were allocated to validation, and all neurons from 20% of the animals were allocated to the test set.

For the generalization across experimental conditions, 80% of the Allen units were allocated to a training set, while 20% were used as validation set. All of the Steinmetz et al. units in the final dataset were used as test set.

Logistic Regression

Summary statistics based on spike times during drifting gratings presentation were obtained and used as features for a logistic regression implemented in scikit-learn [87]. This workflow largely follows our prior work [15]. The logistic regression used L2-regularization and 1E-4 tolerance stopping criterion. When possible, interspike intervals were used to compute these statistics instead of binned spike counts. The following standard statistics were computed: mean firing rate, standard deviation of ISI, median ISI, maximum ISI, minimum ISI, coefficient of variation. Oscillatory activity of the spike train was captured using the power spectral density (PSD) from a periodogram using scipy.signal [88]. The mean value was calculated within each of the following bands (< 4 Hz, 4-8 Hz, 8-12 Hz, 12-40 Hz, 40-100 Hz) and the ratio of each band's power to the power across all bands < 100 Hz was used as a feature. Local variability in the spike train was captured through the CV2, LV, and LVR statistics (as implemented in Elephant) [89]. These statistics were used first individually, and second aggregated to predict each class (structure/region) for a task.

Multi-layer Perceptron

For each task, an MLP was implemented in scikit-learn [87] with a designated set of features to represent the spike train. Features were either: 1) interspike intervals (ISIs) distribution 2) averaged peri-stimulus histogram (PSTH) or 3) concatenated PSTHs.

To calculate ISI distribution for a neuron under specific stimulus, we converted all the spike times of that neuron during the stimulus's presentation into ISIs. Then, we created 100 uniform bins between 1 ms and 3 s and counted the neuron's ISIs in each bin. This resulted in 300 features of ISI "distribution" for each neuron. Finally, we performed min-max normalization, where for each neuron, the maximum value across all bins was set to 1, and the minimum value was set to 0.

The average and concatenated PSTHs were calculated for the drifting grating stimulus. For the average PSTH, we aligned all spikes around the stimulus presentation (1 s before the stimulus presentation and 2 s of stimulus presentation) for each neuron. Then, we created 100 uniform bins over 3 s and counted the number of spikes in each bin for each trial. Then, we averaged the values in each bin across all trials. This resulted in 300 features for each neuron. Finally, we performed min-max normalization, where the maximum value across all bins was set to 1, and the minimum value was set to 0. For the concatenated PSTH, a PSTH was calculated separately for each combination of the drifting grating's parameters, i.e., temporal frequency and orientation. Similar to the average PSTH, we aligned the spikes around the stimulus presentation and performed a binned spike count for 30 uniform bins between 0 and 3 s. We then constructed the PSTH by taking mean across the number of trials. As each combination of the drifting grating's parameter was repeated 15 times during the experiment, the PSTHs are mean of 15 trials. Similar to the ISI distribution and average PSTH, we performed a min-max normalization for each neuron. This operation took place on each PSTH prior to concatenation. Given that there were 40 parameter combinations and 30 bins for each combination, this resulted in 1,200 features per neuron.

For each task, a Bayesian hyperparameter optimization (HyperOpt package) [90] was used to tune hyperparameters (**Table 3**) such that they maximized balanced accuracy on the validation set. To mitigate the effect of class imbalance in our dataset, we treated resampling as a hyperparameter and evaluated undersampling, oversampling, and data augmentation. We found that the synthetic minority oversampling technique (SMOTE) [91], a data augmentation approach, yielded optimal performance on validation sets. Therefore classes were resampled with SMOTE in all cases unless otherwise specified. A different model was trained and tuned for each of five splits to avoid data leakages. For our inductive predictions across datasets (Allen to Steinmetz et al.) tasks, to ensure that the training data encompassed the relevant anatomical representations present in the Steinmetz et al. dataset, we trained models on a subset of Allen dataset that included only brain structures that are also present in the Steinmetz et al. dataset.

Consistent with prior work [15], an MLP with a single hidden layer was generally found to perform better than an MLP with multiple hidden layers. The number of nodes for this hidden layer was optimized. Alpha is the L2 regularization term. Beta 1 is the exponential decay rate for estimates of first moment vector in the Adam optimizer. All unmentioned hyperparameters took on default values from sklearn.

To identify the features contributing to the models' learning, we employed permutation feature importance, which measures the extent to which shuffling a specific feature across samples increases the model's prediction error [92]. To quantify the feature importance of a specific inter-spike interval (ISI) distribution bin for predicting a particular brain region or structure, we first estimated the model's error for the original features, given that the neurons originated from that specific brain region or structure. Then, for those neurons, we randomly permuted (shuffled) a specific bin of the ISI distribution across all samples (neurons) and estimated the model's error for the permuted features. The difference between the original and the permuted features' error represents the feature importance for that specific bin and region/structure. We repeated this process 100 times for all bins in the ISI distribution, each region/structure, and across the five splits. The feature importance is the mean across the split and repetitions. We followed a similar approach for the concatenated PSTH.

Hyperparameter	ISI Distribution	Avg. PSTH	Cat. PSTH
Number of Nodes	50:600:50	30:300:50	50:600:50
Learning Rate	1E-7,1E-6,1E-5,1E-4,1E-3,1E-2,1E-1	1E-5,1E-4,1E-3,1E-2,1E-1	1E-5,1E-4,1E-3,1E-2,1E-1
Batch Size	25:500:50	50:600:50	50:600:50
Alpha	1E-4,1E-3,1E-2,1E-1,1E0,5,10	1E-4,1E-3,1E-2,1E-1,1E0	1E-4,1E-3,1E-2,1E-1,1E0
Beta_1	0.5:0.9:0.1	0.5:0.9:0.1	0.5:0.9:0.1

Table 3

MLP Hyperparameters

Smoothing

For some tasks, predictions were spatially smoothed to improve predictions by leveraging the spatial arrangements of recording channels to weight the influence of nearby units on each prediction. First, neurons (units) were grouped by session and probe to ensure the smoothing was applied within the same probe and session. Second, the predicted probabilities for each class were averaged for all neurons sorted to a particular electrode. Electrodes with no neurons detected were ignored for the purposes of this smoothing. Third, for each electrode, the spatial proximity of nearby units was leveraged by calculating a Gaussian weight. For an electrode, the weight for each nearby unit was determined using a normal distribution centered on the current unit's electrode index (i.e. a linear approximation of the electrode geometry) with a standard deviation equal to the smoothing window. The probability density function (PDF) of the normal distribution was used to compute these weights. The class probabilities for each nearby unit were multiplied by their respective weights. The weighted probabilities were then summed and normalized by the sum of the weights to produce the smoothed prediction for each class for that electrode. The process was repeated for all electrodes. To determine an optimal standard deviation of the Gaussian kernel, we used Hyperopt to maximize balanced accuracy on the validation set.

Models performance metrics and statistical tests

In order to quantify the models' performances, we mainly used balanced accuracy, which is the macro average of recall (true positive rate) per class [38]. Chance-level is 1/number of classes. In some cases, we used the Matthews Correlation Coefficient (MCC). This measure considers true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). It provides a balanced measure, useful even with imbalanced datasets, with a value ranging from −1 (total disagreement) to 1 (perfect prediction) [47, 48]. 0 represents chance-level (which is adjusted to the number of classes). As most of our tasks were multiclass classification, we used a multi-class version of MCC [87, 93].

Given a confusion matrix C for N classes, the following intermediate variables are defined:

$$t_n = \sum_{i=1}^{N} C_{in} \quad \text{(total occurrences of class } n \text{ in the actual data)},$$

$$p_n = \sum_{i=1}^{N} C_{ni} \quad \text{(total occurrences of class } n \text{ in the predictions)},$$

$$c = \sum_{n=1}^{N} C_{nn} \quad \text{(total number of correctly predicted samples)},$$

$$s = \sum_{i=1}^{N} \sum_{j=1}^{N} C_{ij} \quad \text{(total number of samples)}.$$

The multi-class MCC is defined as:

$$\mathrm{MCC} = \frac{c \times s - \sum_{n=1}^{N} p_n \times t_n}{\sqrt{\left(s^2 - \sum_{n=1}^{N} p_n^2\right) \times \left(s^2 - \sum_{n=1}^{N} t_n^2\right)}}.$$

In all cases, the statistical test and level of significance are indicated in the relevant sections of the main text and figure legends. In most cases, linear mixed effects models are employed with subsequent ANOVA for main effect and post-hoc EMMeans with Tukey test for pairwise

comparisons. The implementation was in R (lmer) [94]. In some cases, multiple T-tests with Bonferroni correction (as appropriate) was employed. The implementation was in Python (scipy.stats) [88].

Acknowledgements

We would like to thank the Allen Institute for generating and sharing the Visual Coding datasets. We would like to thank Dr. Josh Siegle for technical insights into the Allen datasets and his helpful perspective on our work. We would also like to thank Dr. Nick Steinmetz for sharing data and technical advice.

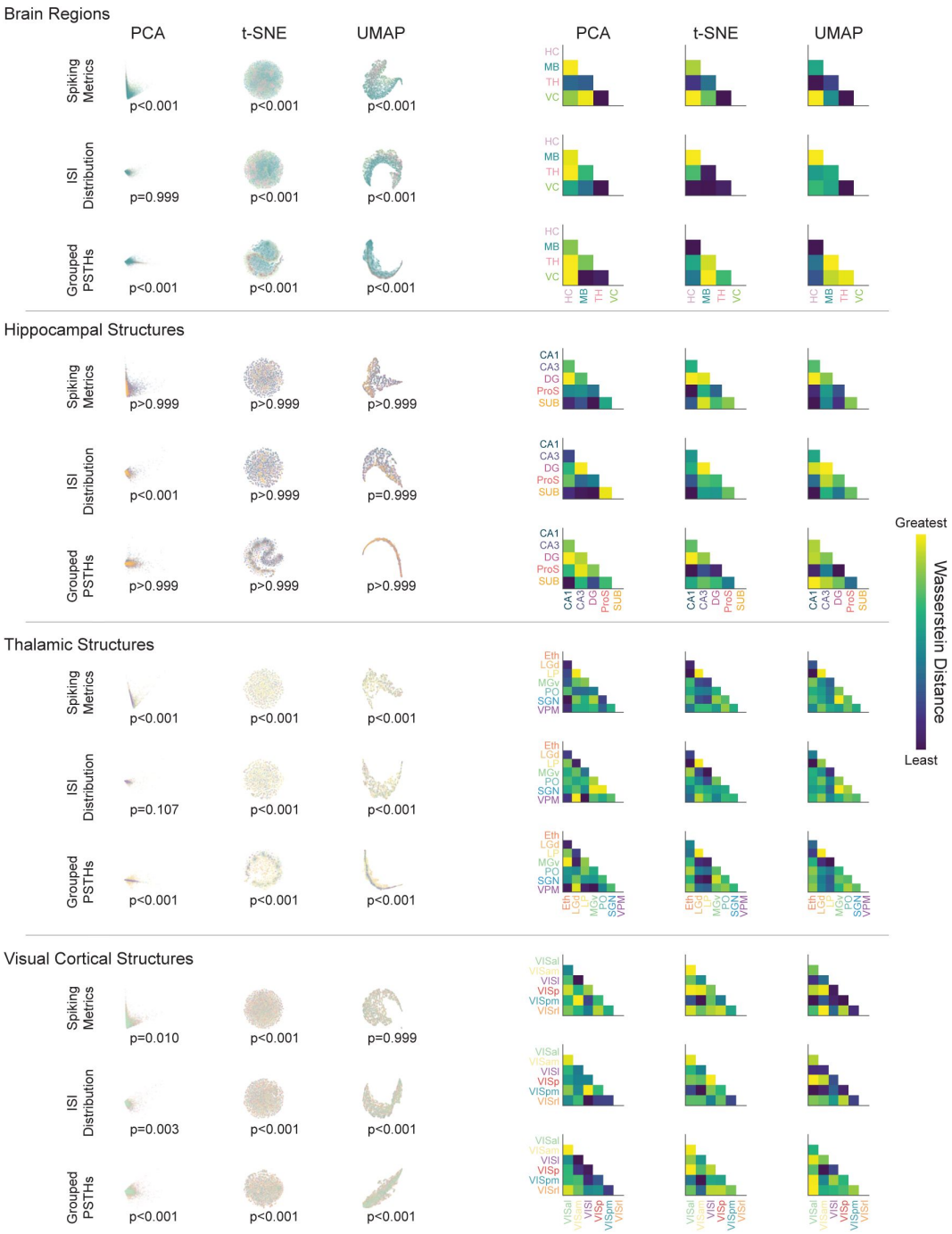


Figure S1

Unsupervised analysis of spiking activity.

Left columns (scatters) show three methods of unsupervised dimensionality reduction (2D): principal component analysis (PCA), t-distributed Stochastic Neighbor Embedding (t-SNE), and Uniform Manifold Approximation and Projection (UMAP). Right columns (Wasserstein distance matrices) show quantification of the distance between anatomical groupings in the corresponding scatter plots on the left. Distance is calculated in the primary dimension for each plot. Rows denote three separate representations of unit activity: 14 pre determined spike metrics (e.g., firing rate), the full ISI distribution, or concatenated PSTHs. Note that within a task (e.g., Brain Regions), there is not a consistent pattern in the distance matrices, suggesting that any structure in the scatterplots is circumstantial. Colors in scatter plots correspond to anatomical labels on the matrices. Horizontal lines separate tasks. P-values in bottom right corner of plots refer to a permutation test of a logistic regression classifier's training accuracy for accurate prediction of region/structure labels (relative to shuffled) based on plot coordinates for all individual units.

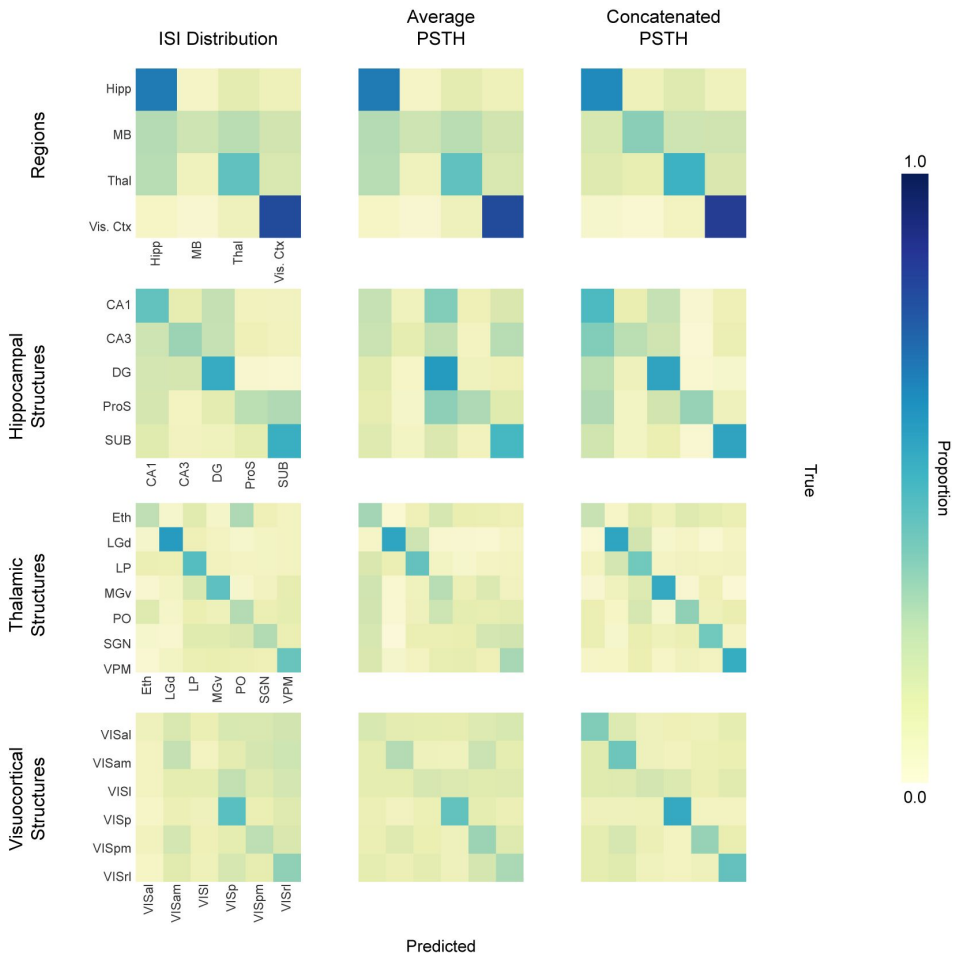


Figure S2

Unsmoothed transductive classification: confusion matrices.

Confusion matrices for MLP models using a transductive train/test split are shown. Matrices for four classification tasks (rows) and three different spiking activity representations (columns) are shown. Note that in the transductive condition, all neurons across animals are pooled and divided into train/test splits, such that there is a within-animal aspect to classifier learning. In individual matrices, the proportion of true instances of a particular class (columns) are distributed across a row reflecting the distribution of MLP predictions. Correct classifications fall along the diagonal from top left to bottom right. The confusion matrices' labels, provided in the first column, are consistent in subsequent columns. Average (across the 5 splits) balanced accuracy (expressed as percent) for each task across the features from left to right: Regions (65.29, 53.96, 59.15), Hippocampus (41.72, 35.51, 44.16), Thalamus (39.15, 31.91, 43.25), Visual Ctx (25.84, 28.84, 38.35)

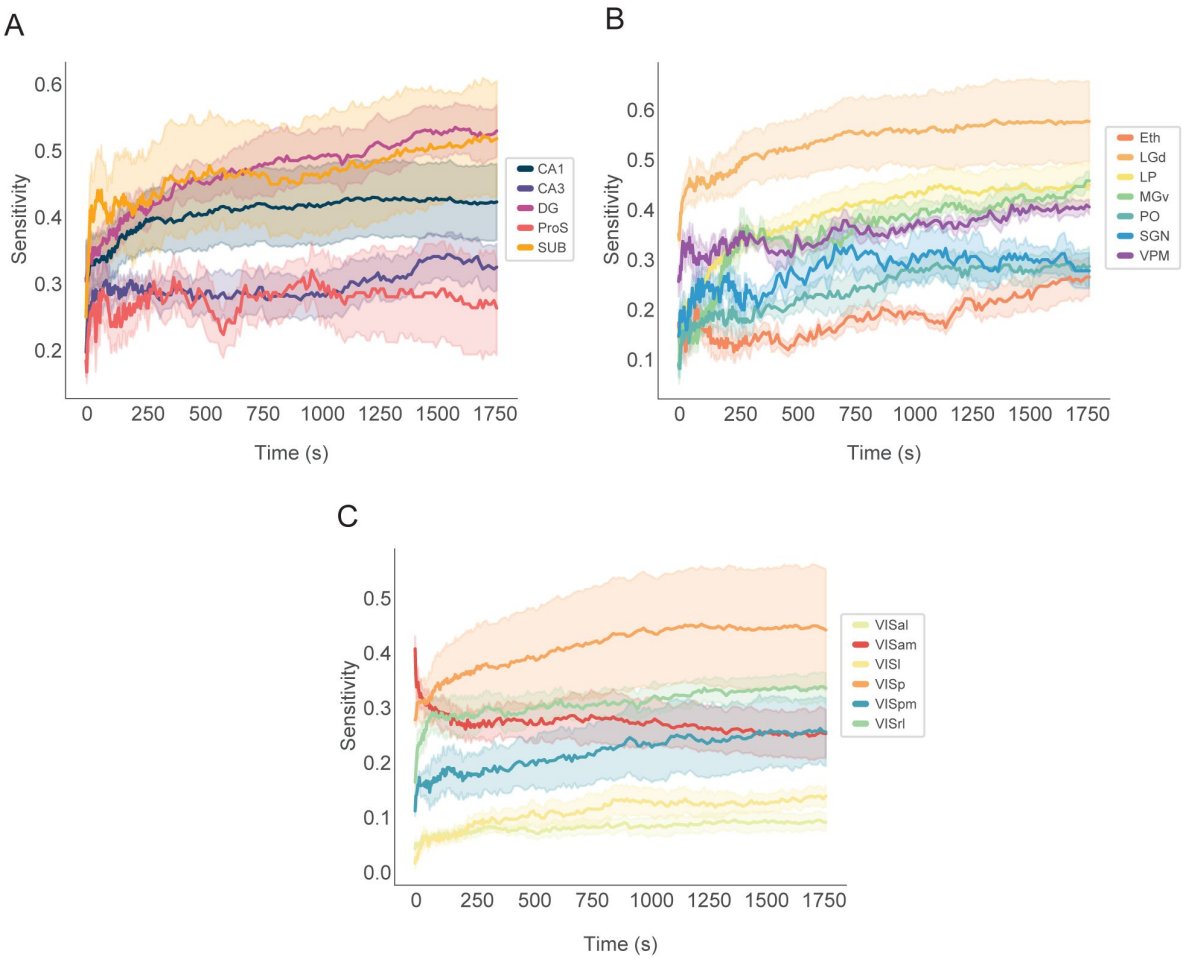


Figure S3

MLP-based model sensitivity as a function of input duration.

Models were trained normally as tested with varying amounts (durations) of input data. In all but one example, model performance increases as a function of input size (duration). The exception is in the visuocortical task, the secondary structure, VISam, declines with time. This reflects the model's failure to learn a robust pattern. Three tasks are shown: **A.** Hippocampal Structures, **B.** Thalamic structures, and **C.** Visuocortical structures.

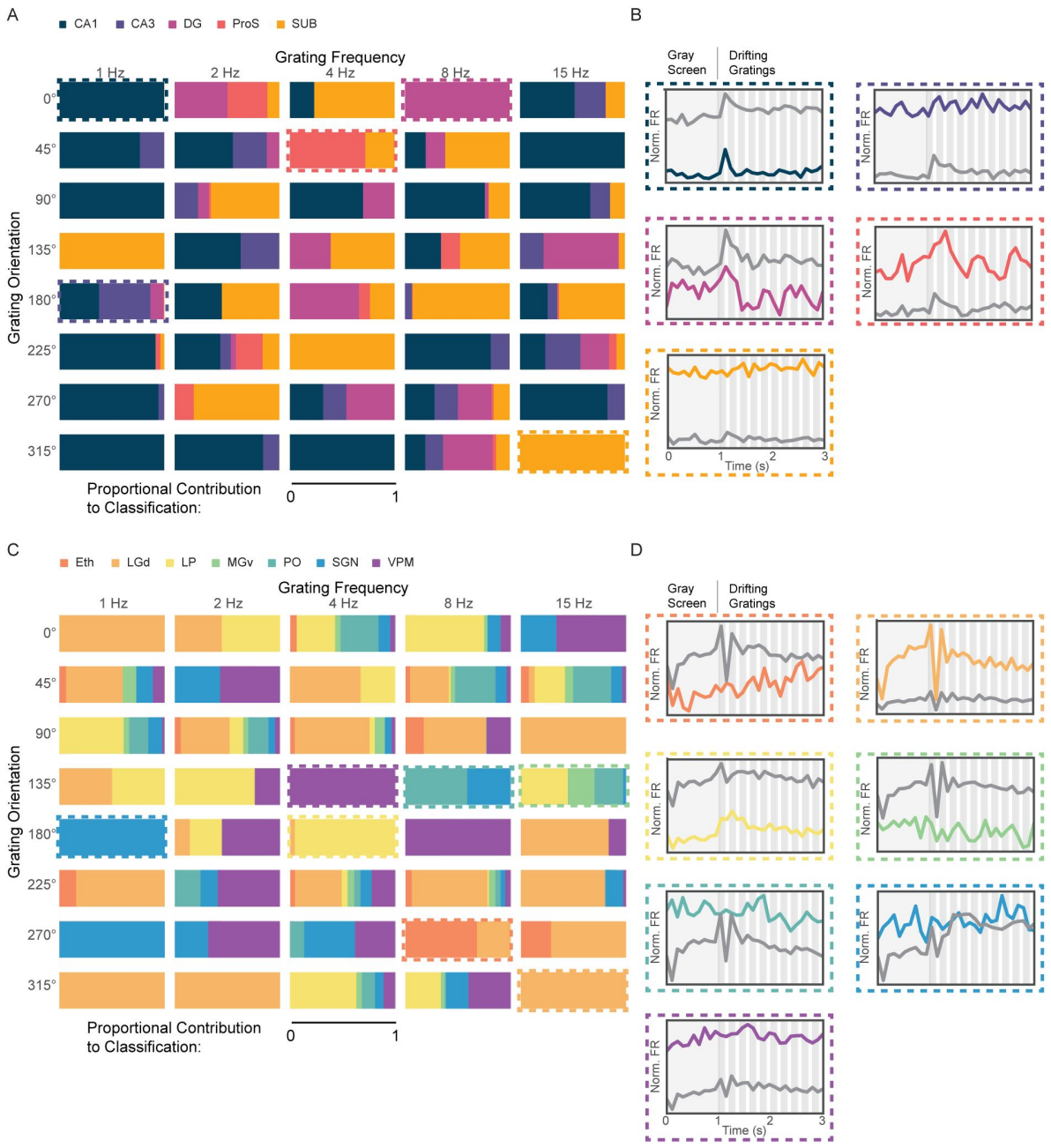


Figure S4

Interpretation of hippocampal and thalamic MLPs.

A Hippocampal structure information is enriched in subsets of stimuli. Each rectangle represents one of the 40 stimulus parameter combinations. Colors correspond to hippocampal structures, and the area of color shows the relative importance of that stimulus to model classification of the given structure. Cat PSTH model. Dashed boxes—exemplar stimulus conditions. **B.** PSTHs corresponding to structure/stimulus pair indicated in A. Specific structure PSTH shown in color. Average PSTH of all other structures shown in gray. **C,D.** Same as A and B except for thalamic neurons and structures.

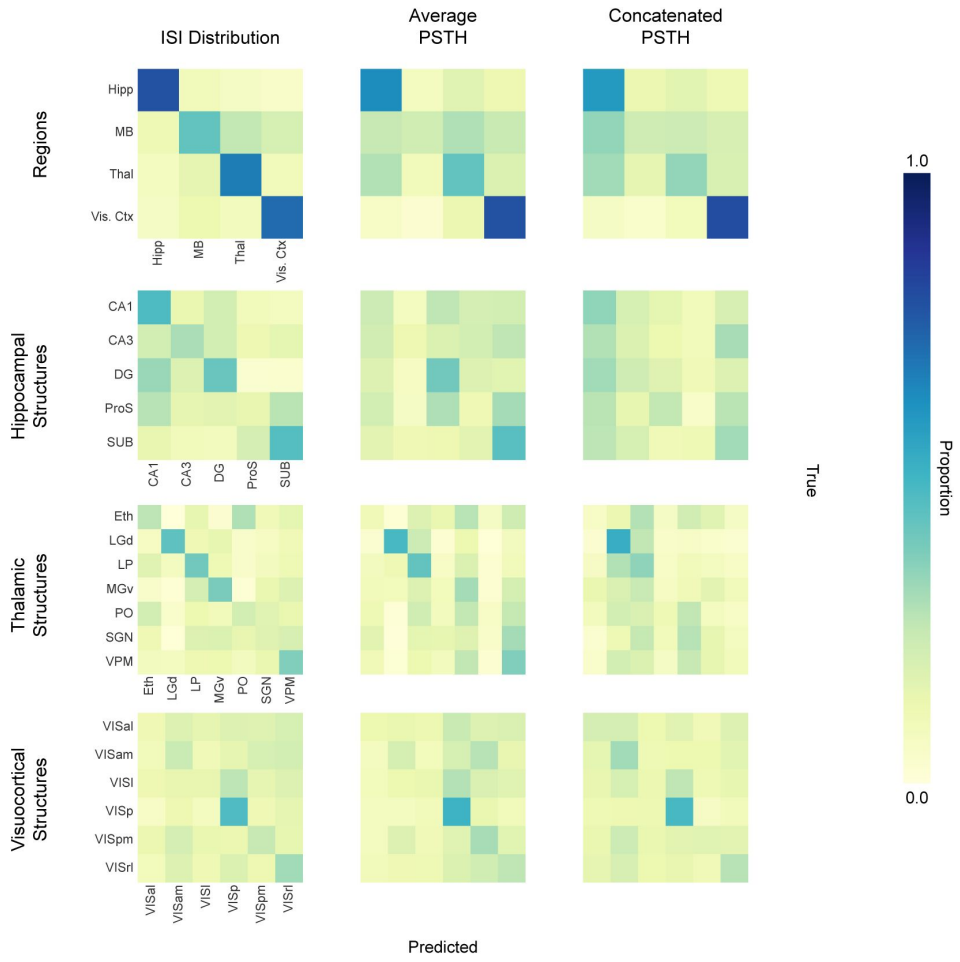


Figure S5

Unsmoothed inductive classification: confusion matrices.

Confusion matrices for MLP models using an inductive train/test split are shown. Note that in the inductive condition, models are trained on all neurons from a subset of animals, and tested on all neurons from a withheld group of animals. This eliminates any possibility of learning a local solution within animal. Matrices for four classification tasks (rows) and three different spiking activity representations (columns) are shown. In individual matrices, the proportion of true instances of a particular class (columns) are distributed across a row reflecting the distribution of MLP predictions. Correct classifications fall along the diagonal from top left to bottom right. The confusion matrices' labels, provided in the first column, are consistent in subsequent columns. Average (across the 5 splits) balanced accuracy values (in %) for each task across the features from left to right: Regions (65.10, 51.72, 49.16), Hippocampus (35.89, 26.39, 21.50), Thalamus (32.79, 26.28, 21.72), Visual Ctx (25.52, 25.83, 26.51)

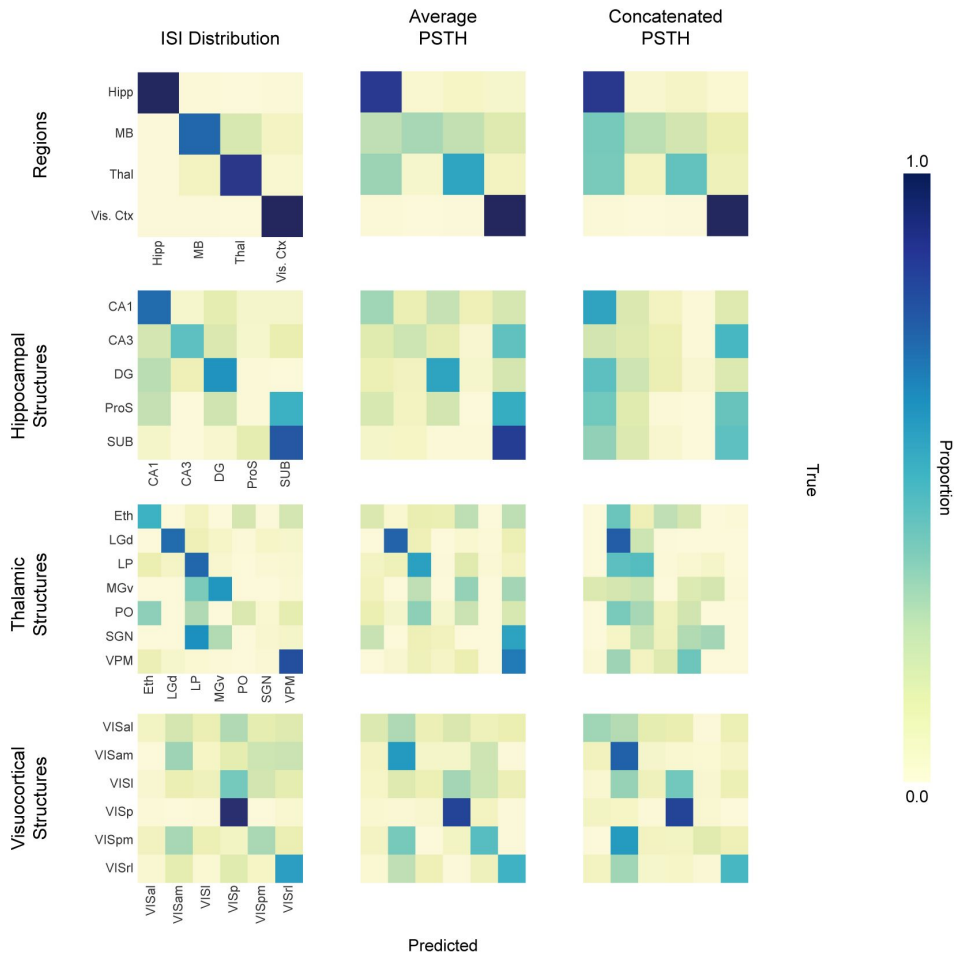


Figure S6

Smoothed inductive classification: confusion matrices.

Confusion matrices for MLP models using an inductive (across animals) train/test split are shown. Note that in the inductive condition, models are trained on all neurons from a subset of animals, and tested on all neurons from a withheld group of animals. This eliminates any possibility of learning a local solution within animal. Matrices for four classification tasks (rows) and three different spiking activity representations (columns) are shown. In individual matrices, the proportion of true instances of a particular class (columns) are distributed across a row reflecting the distribution of MLP predictions. Correct classifications fall along the diagonal from top left to bottom right. The confusion matrices' labels, provided in the first column, are consistent in subsequent columns. Average (across the 5 splits) balanced accuracy values (in %) for each task across the features from left to right: Regions (89.47, 67.95, 63.92), Hippocampus (51.01, 39.31, 25.84), Thalamus (53.21, 35.38, 24.81), Visual Ctx (38.48, 44.66, 43.97)

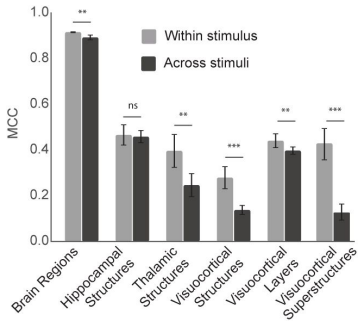


Figure S7

The effect of training and testing within and across stimulus conditions.

For each classification task (e.g., Brain Regions), the bar chart shows the mean MCC value with standard error for models trained and tested on the same stimulus condition (e.g. train: drifting gratings & test: drifting gratings; train: natural movie & test: natural movie) in gray. In black, MCC of models trained and tested on different stimuli (e.g. train: drifting gratings & test: spontaneous; train: natural movies & test: drifting gratings). Linear mixed effects. (ns/not significant: $p >= 0.05$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$)

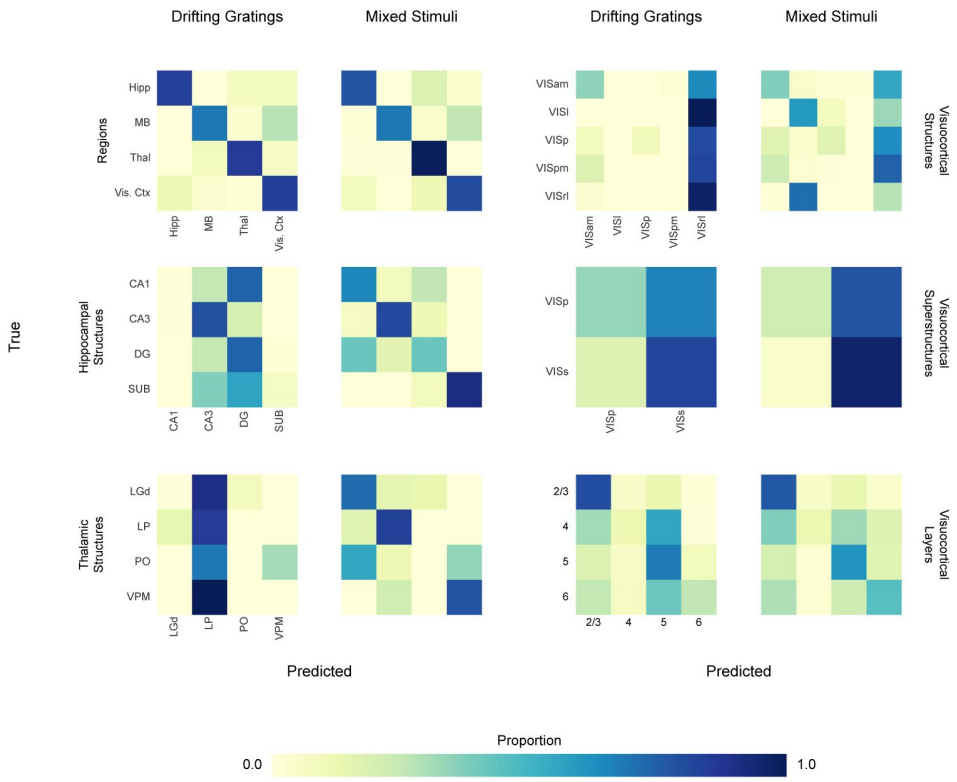


Figure S8

Train/test across laboratories: Allen-to-Steinmetz et al. confusion matrices.

6 tasks are displayed. The y-axis label denotes the task (e.g, Regions), and the column label denotes the training condition (drifting gratings or mixed stimuli). All models involve training on the Allen Institute data and testing on Steinmetz et al. data. Mixed stimuli models are trained on neuronal activity recorded during the presentation of drifting gratings, natural movies, and spontaneous activity. Balanced accuracy corresponding to each confusion matrix: BR DG=80.46%, BR Mix: 81.28 %, HS DG: 40.07 %, HS Mix: 69.89 %, TS DG: 21.52 %, TS Mix: 58.22 %, VCS DG: 28.41 %, VCS Mix: 28.34 %, VCSS DG: 58 %, VCSS Mix: 59 %, VC Layers DG: 46.36 %, VC Layers Mix: 49.01 % where BR stands for brain regions, HS hippocampal structures, TS thalamic structures, VCS visual cortex structures, VCSS visual cortex superstructures, DG drifting gratings and Mix denotes the combined stimuli.

References

[1] Adrian E. D., Bronk D. W. (1928) **The discharge of impulses in motor nerve fibres: Part I. Impulses in single fibres of the phrenic nerve** *The Journal of Physiology* **66**:81–101 https://doi.org/10.1113/jphysiol.1928.sp002509

[2] McCulloch Warren S., Pitts Walter (1943) **A logical calculus of the ideas immanent in nervous activity** *The Bulletin of Mathematical Biophysics* **5**:115–133 https://doi.org/10.1007/bf02478259

[3] Lettvin J, et al. (1959) **What the frog's eye tells the frog's brain** *Proc. IRE* **47**:1940–1951

[4] Aldo Faisal A., Selen Luc P. J., Wolpert Daniel M. (2008) **Noise in the nervous system** *Nature Reviews Neuroscience* **9**:292–303 https://doi.org/10.1038/nrn2258

[5] Gerstner Wulfram, et al. (2014) **Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition** Cambridge University Press https://doi.org/10.1017/cbo9781107447615

[6] Mainen Zachary F., Sejnowski Terrence J. (1995) **Reliability of Spike Timing in Neocortical Neurons** *Science* **268**:1503–1506 https://doi.org/10.1126/science.7770778

[7] Softky WR, Koch C (1993) **The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs** *The Journal of Neuroscience* **13**:334–350 https://doi.org/10.1523/jneurosci.13-01-00334.1993

[8] Ramón y Cajal Santiago (1904) **Textura del Sistema Nervioso del Hombre y de los Vertebrados** Madrid: Moya

[9] Brodmann Korbinian (1909) **Vergleichende Lokalisationslehre der Großhirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues** Leipzig: Barth

[10] Penfield Wilder, Jasper Herbert (1951) **Epilepsy and the Functional Anatomy of the Human Brain** Boston: Little, Brown and Company

[11] Shadlen Michael N., Newsome William T. (1994) **Noise, neural codes and cortical organization** *Current Opinion in Neurobiology* **4**:569–579 https://doi.org/10.1016/0959-4388(94)90059-0

[12] London Michael, et al. (2010) **Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex** *Nature* **466**:123–127 https://doi.org/10.1038/nature09086

[13] Pandarinath Chethan, et al. (2018) **Inferring single-trial neural population dynamics using sequential auto-encoders** *Nature Methods* **15**:805–815 https://doi.org/10.1038/s41592-018-0109-9

[14] Sawant Yash, et al. (2022) **A Midbrain Inspired Recurrent Neural Network Model for Robust Change Detection** *The Journal of Neuroscience* **42**:8262–8283 https://doi.org/10.1523/jneurosci.0164-22.2022

[15] Schneider Aidan, et al. (2023) **Transcriptomic cell type structures in vivo neuronal activity across multiple timescales** *Cell Reports* **42** https://doi.org/10.1016/j.celrep.2023.112318

[16] Patel N, Poo MM (1982) **Orientation of neurite growth by extracellular electric fields** *The Journal of Neuroscience* **2**:483–496 https://doi.org/10.1523/jneurosci.02-04-00483.1982

[17] Gütig Robert, Sompolinsky Haim (2006) **The tempotron: a neuron that learns spike timing–based decisions** *Nature Neuroscience* **9**:420–428 https://doi.org/10.1038/nn1643

[18] Siegle Joshua H., et al. (2021) **Survey of spiking in the mouse visual system reveals functional hierarchy** *Nature* **592**:86–92 https://doi.org/10.1038/s41586-020-03171-x

[19] Elston G. N., Kaas Jon H. (2007) **Specialization of the neocortical pyramidal cell during primate evolution** *Evolution of Nervous Systems* Oxford: Elsevier :191–242

[20] Chaudhuri Rajan, et al. (2015) **A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex** *Neuron* **88**:419–431 https://doi.org/10.1016/j.neuron.2015.09.008

[21] Murray John D, et al. (2014) **A hierarchy of intrinsic timescales across primate cortex** *Nature Neuroscience* **17**:1661–1663 https://doi.org/10.1038/nn.3862

[22] Maunsell J. H., Van Essen D. C. (1983) **The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey** *Journal of Neuroscience* **3**:2563–2586

[23] Gămănůt Răzvan, et al. (2018) **The mouse cortical connectome, characterized by an ultra-dense cortical graph, maintains specificity by distinct connectivity profiles** *Neuron* **97**:698–715 https://doi.org/10.1016/j.neuron.2018.01.010

[24] Harris Julie A., et al. (2019) **Hierarchical organization of cortical and thalamic connectivity** *Nature* **575**:195–202 https://doi.org/10.1038/s41586-019-1716-z

[25] Millman Daniel J, et al. (2020) **VIP interneurons in mouse primary visual cortex selectively enhance responses to weak but specific stimuli** *eLife* **9** https://doi.org/10.7554/elife.55130

[26] (1958) **Touch of Evil**

[27] Shinomoto Shigeru, et al. (2009) **Relating Neuronal Firing Patterns to Functional Differentiation of Cerebral Cortex** *PLoS Computational Biology* **5** https://doi.org/10.1371/journal.pcbi.1000433

[28] Mochizuki Yasuhiro, et al. (2016) **Similarity in Neuronal Firing Regimes across Mammalian Species** *The Journal of Neuroscience* **36**:5736–5747 https://doi.org/10.1523/jneurosci.0230-16.2016

[29] Wang Xiao-Jing (2022) **Theory of the Multiregional Neocortex: Large-Scale Neural Dynamics and Distributed Cognition** *Annual Review of Neuroscience* **45**:533–560 https://doi.org/10.1146/annurev-neuro-110920-035434

[30] Hastie Trevor, Tibshirani Robert, Friedman Jerome (2009) **The Elements of Statistical Learning** Springer New York https://doi.org/10.1007/978-0-387-84858-7

[31] Pearson Karl (1901) **On Lines and Planes of Closest Fit to Systems of Points in Space** *Philosophical Magazine* **2**:559–572 https://doi.org/10.1080/14786440109462720

[32] van der Maaten L.J.P., Hinton G.E. (2008) **Visualizing Data using t-SNE** *Journal of Machine Learning Research* **9**:2579–2605

[33] McInnes Leland, Healy John, Melville James (2018) **UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction** *arXiv*

[34] Jinno Shozo, et al. (2007) **Neuronal Diversity in GABAergic Long-Range Projections from the Hippocampus** *The Journal of Neuroscience* **27**:8790–8804 https://doi.org/10.1523/jneurosci.1847-07.2007

[35] Cembrowski Mark S., et al. (2016) **Spatial Gene-Expression Gradients Underlie Prominent Heterogeneity of CA1 Pyramidal Neurons** *Neuron* **89**:351–368 https://doi.org/10.1016/j.neuron.2015.12.013

[36] Pan Sinno Jialin, Yang Qiang (2010) **A Survey on Transfer Learning** *IEEE Transactions on Knowledge and Data Engineering* **22**:1345–1359 https://doi.org/10.1109/TKDE.2009.191

[37] Vapnik Vladimir N. (1998) **Statistical Learning Theory** Wiley-Interscience

[38] Brodersen Kay H., et al. (2010) **The balanced accuracy and its posterior distribution** *Proceedings of the 20th International Conference on Pattern Recognition (ICPR)* :3121–3124 https://doi.org/10.1109/ICPR.2010.764

[39] Holt G.R., et al. (1996) **Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons** *Journal of Neurophysiology* **75**:1806–1814 https://doi.org/10.1152/jn.1996.75.5.1806

[40] Riehle Alexa, et al. (1997) **Spike Synchronization and Rate Modulation Differentially Involved in Motor Cortical Function** *Science* **278**:1950–1953 https://doi.org/10.1126/science.278.5345.1950

[41] Maass Wolfgang (2000) **On the Computational Power of Winner-Take-All** *Neural Computation* **12**:2519–2535 https://doi.org/10.1162/089976600300014827

[42] Paz Luciano, et al. (2016) **Confidence through consensus: a neural mechanism for uncertainty monitoring** *Scientific Reports* **6** https://doi.org/10.1038/srep21830

[43] Wang Quanxin, Burkhalter Andreas (2007) **Area map of mouse visual cortex** *Journal of Comparative Neurology* **502**:339–357 https://doi.org/10.1002/cne.21286

[44] Glickfeld Lindsey L., Olsen Shawn R. (2017) **Higher-Order Areas of the Mouse Visual Cortex** *Annual Review of Vision Science* **3**:251–273 https://doi.org/10.1146/annurev-vision-102016-061331

[45] Lee Eric Kenji, et al. (2024) **PhysMAP - interpretablein vivoneuronal cell type identification using multi-modal analysis of electrophysiological data** *bioRxiv* https://doi.org/10.1101/2024.02.28.582461

[46] Wang Helen, et al. (2022) **Diversity in spatial frequency, temporal frequency, and speed tuning across mouse visual cortical areas and layers** *Journal of Comparative Neurology* **530**:3226–3247 https://doi.org/10.1002/cne.25404

[47] Chicco Davide, Jurman Giuseppe (2020) **The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation** *BMC genomics* **21**

[48] Boughorbel Sabri, Jarray Fethi, El-Anbari Mohammed (2017) **Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric** *PloS one* **12**

[49] Jurman Giuseppe, Riccadonna Samantha, Furlanello Cesare (2012) **A comparison of MCC and CEN error measures in multi-class prediction** *PloS one* **7**

[50] Steinmetz Nicholas A., et al. (2019) **Distributed coding of choice, action and engagement across the mouse brain** *Nature* **576**:266–273 https://doi.org/10.1038/s41586-019-1787-x

[51] Olshausen B, Field D (2004) **Sparse coding of sensory inputs** *Current Opinion in Neurobiology* **14**:481–487 https://doi.org/10.1016/j.conb.2004.07.007

[52] Harris Kenneth D., Thiele Alexander (2011) **Cortical state and attention** *Nature Reviews Neuroscience* **12**:509–523 https://doi.org/10.1038/nrn3084

[53] Parks David F., et al. (2023) **A non-oscillatory, millisecond-scale embedding of brain state provides insight into behavior** *bioRxiv* https://doi.org/10.1101/2023.06.09.544399

[54] Buccino Alessio P., et al. (2018) **Combining biophysical modeling and deep learning for multielectrode array neuron localization and classification** *Journal of Neurophysiology* **120**:1212–1232 https://doi.org/10.1152/jn.00210.2018

[55] Jia Xiaoxuan, et al. (2019) **High-density extracellular probes reveal dendritic backpropagation and facilitate neuron classification** *Journal of Neurophysiology* **121**:1831–1847 https://doi.org/10.1152/jn.00680.2018

[56] Roxin Alex, et al. (2011) **On the Distribution of Firing Rates in Networks of Cortical Neurons** *The Journal of Neuroscience* **31**:16217–16226 https://doi.org/10.1523/jneurosci.1677-11.2011

[57] Mizuseki Kenji, Buzsáki György (2013) **Preconfigured, Skewed Distribution of Firing Rates in the Hippocampus and Entorhinal Cortex** *Cell Reports* **4**:1010–1021 https://doi.org/10.1016/j.celrep.2013.07.039

[58] Mi Lu, et al., et al., Oh A. (2023) **Learning Time-Invariant Representations for Individual Neurons from Population Dynamics** *Advances in Neural Information Processing Systems* Curran Associates, Inc :46007–46026

[59] Niell Cristopher M., Stryker Michael P. (2008) **Highly Selective Receptive Fields in Mouse Visual Cortex** *The Journal of Neuroscience* **28**:7520–7536 https://doi.org/10.1523/jneurosci.0623-08.2008

[60] Schmolesky Matthew T., et al. (1998) **Signal Timing Across the Macaque Visual System** *Journal of Neurophysiology* **79**:3272–3278 https://doi.org/10.1152/jn.1998.79.6.3272

[61] Zeisel Amit, et al. (2018) **Molecular Architecture of the Mouse Nervous System** *Cell* **174**:999–1014 https://doi.org/10.1016/j.cell.2018.06.021

[62] Saunders Arpiar, et al. (2018) **Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain** *Cell* **174**:1015–1030 https://doi.org/10.1016/j.cell.2018.07.028

[63] Braganza Oliver, Beck Heinz (2018) **The Circuit Motif as a Conceptual Tool for Multilevel Neuroscience** *Trends in Neurosciences* **41**:128–136 https://doi.org/10.1016/j.tins.2018.01.002

[64] Luo Liqun (2021) **Architectures of neuronal circuits** *Science* **373** https://doi.org/10.1126/science.abg7285

[65] Hengen Keith B., et al. (2013) **Firing Rate Homeostasis in Visual Cortex of Freely Behaving Rodents** *Neuron* **80**:335–342 https://doi.org/10.1016/j.neuron.2013.08.038

[66] Hengen Keith B., et al. (2016) **Neuronal Firing Rate Homeostasis Is Inhibited by Sleep and Promoted by Wake** *Cell* **165**:180–191 https://doi.org/10.1016/j.cell.2016.01.046

[67] Ma Zhengyu, et al. (2019) **Cortical Circuit Dynamics Are Homeostatically Tuned to Criticality In Vivo** *Neuron* **104**:655–664 https://doi.org/10.1016/j.neuron.2019.08.031

[68] Renart Alfonso, Machens Christian K (2014) **Variability in neural activity and behavior** *Current Opinion in Neurobiology* **25**:211–220 https://doi.org/10.1016/j.conb.2014.02.013

[69] Cohen Marlene R, Kohn Adam (2011) **Measuring and interpreting neuronal correlations** *Nature Neuroscience* **14**:811–819 https://doi.org/10.1038/nn.2842

[70] Smith Matthew A., Kohn Adam (2008) **Spatial and Temporal Scales of Neuronal Correlation in Primary Visual Cortex** *The Journal of Neuroscience* **28**:12591–12603 https://doi.org/10.1523/jneurosci.2929-08.2008

[71] Smith Matthew A., Sommer Marc A. (2013) **Spatial and Temporal Scales of Neuronal Correlation in Visual Area V4** *The Journal of Neuroscience* **33**:5422–5432 https://doi.org/10.1523/jneurosci.4782-12.2013

[72] Rosenbaum Robert, et al. (2016) **The spatial structure of correlated neuronal variability** *Nature Neuroscience* **20**:107–114 https://doi.org/10.1038/nn.4433

[73] Ruff Douglas A, Cohen Marlene R (2014) **Attention can either increase or decrease spike count correlations in visual cortex** *Nature Neuroscience* **17**:1591–1597 https://doi.org/10.1038/nn.3835

[74] Marshel James H., et al. (2011) **Functional Specialization of Seven Mouse Visual Cortical Areas** *Neuron* **72**:1040–1054 https://doi.org/10.1016/j.neuron.2011.12.004

[75] Andermann Mark L., et al. (2011) **Functional Specialization of Mouse Higher Visual Cortical Areas** *Neuron* **72**:1025–1039 https://doi.org/10.1016/j.neuron.2011.11.013

[76] Roth M. M., Helmchen F., Kampa B. M. (2012) **Distinct Functional Properties of Primary and Posteromedial Visual Area of Mouse Neocortex** *Journal of Neuroscience* **32**:9716–9726 https://doi.org/10.1523/jneurosci.0110-12.2012

[77] Ayzenshtat Inbal, Jackson Jesse, Yuste Rafael (2016) **Orientation Tuning Depends on Spatial Frequency in Mouse Visual Cortex** *eneuro* **3** https://doi.org/10.1523/eneuro.0217-16.2016

[78] Kumar Mari Ganesh, et al. (2021) **Functional parcellation of mouse visual cortex using statistical techniques reveals response-dependent clustering of cortical processing areas** *PLOS Computational Biology* **17** https://doi.org/10.1371/journal.pcbi.1008548

[79] Steinmetz Nicholas A, et al. (2018) **Challenges and opportunities for large-scale electrophysiology with Neuropixels probes** *Current Opinion in Neurobiology* **50**:92–100 https://doi.org/10.1016/j.conb.2018.01.009

[80] Sebastian Seung H (2011) **Neuroscience: Towards functional connectomics** *Nature* **471**:170–172

[81] Marder Eve, Bucher Dirk (2007) **Understanding circuit dynamics using the stomatogastric nervous system of lobsters and crabs** *Annu. Rev. Physiol* **69**:291–316

[82] Lisman John, Schulman Howard, Cline Hollis (2002) **The molecular basis of CaMKII function in synaptic and behavioural memory** *Nature Reviews Neuroscience* **3**:175–190 https://doi.org/10.1038/nrn753

[83] Allen Institute for Brain Science (2019) **AllenSDK**

[84] Allen Institute MindScope Program (2019) **Allen Brain Observatory – Neuropixels Visual Coding [dataset]**

[85] Allen Institute MindScope Program (2019) **Allen Brain Observatory – Neuropixels Visual Coding** Allen Institute for Brain Science

[86] Wang Quanxin, et al. (2020) **The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas** *Cell* **181**:936–953 https://doi.org/10.1016/j.cell.2020.04.007

[87] Pedregosa F., et al. (2011) **Scikit-learn: Machine Learning in Python** *Journal of Machine Learning Research* **12**:2825–2830

[88] Virtanen Pauli, et al. (2020) **SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python** *Nature Methods* **17**:261–272 https://doi.org/10.1038/s41592-019-0686-2

[89] Denker M., Yegenoglu A., Grün S. (2018) **Collaborative HPC-enabled workflows on the HBP Collaboratory using the Elephant framework** *Neuroinformatics* **2018** https://doi.org/10.12751/incf.ni2018.0019

[90] Bergstra James, Yamins Daniel, Cox David D. (2013) **Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures** *Proceedings of the 30th International Conference on Machine Learning (ICML 2013)* :I-115–I-123

[91] Chawla Nitesh V., et al. (2002) **Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD 2002)** ACM :106–113

[92] Fisher Aaron, Rudin Cynthia, Dominici Francesca (2018) **All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously** *arXiv* https://doi.org/10.48550/ARXIV.1801.01489

[93] Gorodkin J. (2004) **Comparing two K-category assignments by a K-category correlation coefficient** *Computational Biology and Chemistry* **28**:367–374 https://doi.org/10.1016/j.compbiolchem.2004.09.006

[94] Bates Douglas, et al. (2015) **Fitting Linear Mixed-Effects Models Using lme4** *Journal of Statistical Software* **67**:1–48 https://doi.org/10.18637/jss.v067.i01

Author information

Gemechu B Tolossa*

Department of Biology, Washington University in Saint Louis, Saint Louis, United States
ORCID iD: 0000-0001-6405-2908

*These authors contributed equally to this work

Aidan M Schneider*

Department of Biology, Washington University in Saint Louis, Saint Louis, United States
ORCID iD: 0000-0001-8774-2303

*These authors contributed equally to this work

Eva L Dyer

Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, United States

Keith B Hengen

Department of Biology, Washington University in Saint Louis, Saint Louis, United States
ORCID iD: 0000-0001-5017-4090

For correspondence: khengen@wustl.edu

Editors

Reviewing Editor
Tatyana Sharpee
Salk Institute for Biological Studies, La Jolla, United States of America

Senior Editor
Sacha Nelson
Brandeis University, Waltham, United States of America

Reviewer #1 (Public review):

Summary:

The paper by Tolossa et al. presents classification studies that aim to predict the anatomical location of a neuron from the statistics of its in-vivo firing pattern. They study two types of statistics (ISI distribution, PSTH) and try to predict the location at different resolutions (region, subregion, cortical layer).

Strengths:

This paper provides a systematic quantification of the single-neuron firing vs location relationship.

The quality of the classification setup seems high.

The paper uncovers that, at the single neuron level, the firing pattern of a neuron carries some information on the neuron's anatomical location, although the predictive accuracy is not high enough to rely on this relationship in most cases.

Weaknesses:

As the authors mention in the Discussion, it is not clear whether the observed differences in firing are epiphenomenal. If the anatomical location information is useful to the neuron, to what extent can this be inferred from the vicinity of the synaptic site, based on the neurotransmitter and neuromodulator identities? Why would the neuron need to dynamically update its prediction of the anatomical location of its pre-synaptic partner based on activity when that location is static, and if that information is genetically encoded in synaptic proteins, etc (e.g., the type of the synaptic site)? Note that the neuron does not need to classify all possible locations to guess the location of its pre-synaptic partner because it may only receive input from a subset of locations. If an argument on activity-based estimation being more advantageous to the neuron than synaptic site-based estimation cannot be made, I believe limiting the scope of the paper (e.g., in the Introduction) to an epiphenomenal observation and its quantification will improve the scientific quality.Life Assessment

This article reports a useful set of findings on how electrophysiological response properties of neurons correlate with their position in the brain. The evidence currently remains incomplete, with reviewers making specific suggestions for how clustering needs to be redone. The manuscript would also benefit from a more focused presentation of results and the removal of incorrect claims about recording biases.

https://doi.org/10.7554/eLife.101506.1.sa2

Reviewer #2 (Public review):

Summary:

In this manuscript, Tolossa et al. analyze Inter-spike intervals from various freely available datasets from the Allen Institute and from a dataset from Steinmetz et al. They show that they can modestly decode between gross brain regions (Visual vs. Hippocampus vs. Thalamus), and modestly separate sub-areas within brain regions (DG vs. CA1 or various visual brain areas).

Strengths:

The paper is reasonably well written, and the definitions are quite well done. For example, the authors clearly explained transductive vs. inductive inference in their decoders. E.g., transductive learning allows the decoder to learn features from each animal, whereas inductive inference focuses on withheld animals and prioritizes the learning of generalizable features.

Weaknesses:

However, even with some of these positive aspects, I still found the manuscript to be a laundry list of results, where some results are overly explained and not particularly compelling or interesting, whereas interesting results are not strongly described or emphasized. The overall problem is that the study is not cohesive, and the authors need to either come up with a tool or demonstrate a scientific finding. The current version attempts to split the middle and thus is not as impactful as it could be.

https://doi.org/10.7554/eLife.101506.1.sa1

Author response:

> ***Reviewer #1 (Public review):***
>
> *Summary:*
>
> *The paper by Tolossa et al. presents classification studies that aim to predict the anatomical location of a neuron from the statistics of its in-vivo firing pattern. They study two types of statistics (ISI distribution, PSTH) and try to predict the location at different resolutions (region, subregion, cortical layer).*
>
> *Strengths:*
>
> *This paper provides a systematic quantification of the single-neuron firing vs location relationship.*
>
> *The quality of the classification setup seems high.*
>
> *The paper uncovers that, at the single neuron level, the firing pattern of a neuron carries some information on the neuron's anatomical location, although the predictive accuracy is not high enough to rely on this relationship in most cases.*

Thank you for your thoughtful feedback. The level of predictive accuracy offered by our current approach, while far above chance, is insufficient for electrode localization in most cases. Although, we speculate that our results represent a lower limit on possible performance—future improvements are almost certain as larger datasets are generated, more diverse features of neural activity are employed, and more advanced ML tools are implemented. We note that the current performance indicates a far more reliable embedding of anatomy in spiking than precedented by the modest statistical significance previously described in the literature. It would have been impossible to achieve this without the tremendous resources provided by the Allen Institute. In our revision, we will clarify that major performance improvements are both possible and probable.

> *Weaknesses:*
>
> *As the authors mention in the Discussion, it is not clear whether the observed differences in firing are epiphenomenal. If the anatomical location information is useful to the neuron, to what extent can this be inferred from the vicinity of the synaptic site, based on the neurotransmitter and neuromodulator identities? Why would the neuron need to dynamically update its prediction of the anatomical location of its pre-synaptic partner based on activity when that location is static, and if that information is genetically encoded in synaptic proteins, etc (e.g., the type of the synaptic site)? Note that the neuron does not need to classify all possible locations to guess the location of its pre-synaptic partner because it may only receive input from a subset of locations. If an argument on activity-based estimation being more advantageous to the neuron than synaptic site-based estimation cannot be made, I believe limiting the scope of the paper (e.g., in the Introduction) to an epiphenomenal observation and its quantification will improve the scientific quality.*

Summarily, in response to the two reviewers, we will minimize our discussion of this question in the revision. However, given that our results are either epiphenomenal or functional, we feel that it is important to indicate these possibilities, even if this indication is succinct and conservative.

In pursuit of a more concise revision, we will not expand our discussion to accommodate this interesting conversation with the reviewer, but we are excited to briefly offer our perspective

here.

Regarding the epiphenomenal nature of our observations: this is a complex question that would be challenging but not impossible to validate experimentally. It has been previously established that neurons, especially those that integrate inputs from a variety of regions and are involved in diverse functions, could benefit from mechanisms for dynamically parsing inputs (Gutig, Sompolinsky 2006). Neurotransmitter and neuromodulator identities may indeed convey some information about presynaptic neuron location (e.g., NE may originate from the locus coeruleus). However, hypothetically, the binding of a neurotransmitter only bears on the postsynaptic neuron via ionic current, or second messenger activity. Postsynaptic neurons do not consume or otherwise endocytose the neurotransmitter, thus the ability of a neuron to "know" the presynaptic identity is a function of induced postsynaptic activity. Certainly, there are multiple streams of information that can provide insight into anatomical location all taking the ultimate form of neural activity and membrane dynamics. This would be broadly consistent with (for example) reward prediction error which is evident in dopamine release, firing rates, spiking patterns, and oscillatory rhythms.

We could imagine a possible role for the embedding of location in spiking patterns. It is important to note that many neurons in neighboring areas share common neurotransmitters (e.g., glutamate, GABA). Neurons receiving input from multiple regions with similar neurotransmitter profiles could benefit from additional information in the spiking patterns for distinguishing input sources, especially for multimodal integration. For instance, an inferior parietal lobule neuron or microcircuit could be downstream from both auditory cortex (listening) and Broca's area (speaking). Imagine an individual is in a crowded coffee shop waiting for their drink order to be called while speaking to their friend. In this scenario, it may be important to recognize region-specific activity and thus selectively attend to it. Thus, it is unlikely that neurons actively update a "location prediction," but rather that location-related information is passively embedded in spike patterning and this might be dynamically leveraged in computation. We emphasize that this is a simplified conceptual example and not a hypothesis that we test in the paper. This conversation, however, is a wonderful example of the thought experiments that we hope will grow from this type of work.

> ***Reviewer #2 (Public review):***
>
> *Summary:*
>
> *In this manuscript, Tolossa et al. analyze Inter-spike intervals from various freely available datasets from the Allen Institute and from a dataset from Steinmetz et al. They show that they can modestly decode between gross brain regions (Visual vs. Hippocampus vs. Thalamus), and modestly separate sub-areas within brain regions (DG vs. CA1 or various visual brain areas).*
>
> *Strengths:*
>
> *The paper is reasonably well written, and the definitions are quite well done. For example, the authors clearly explained transductive vs. inductive inference in their decoders. E.g., transductive learning allows the decoder to learn features from each animal, whereas inductive inference focuses on withheld animals and prioritizes the learning of generalizable features.*

Thank you!

> *Weaknesses:*

> *However, even with some of these positive aspects, I still found the manuscript to be a laundry list of results, where some results are overly explained and not particularly compelling or interesting, whereas interesting results are not strongly described or emphasized. The overall problem is that the study is not cohesive, and the authors need to either come up with a tool or demonstrate a scientific finding. The current version attempts to split the middle and thus is not as impactful as it could be*

In our revision, we will endeavor to present our results in line with your suggestions. Thank you for the careful and thorough feedback that will improve the readability of our manuscript. We strove to be complete in establishing the logic leading to our ultimate finding —that a robust code for anatomical location can be extracted from single neuron spike trains, but not from more traditional descriptions of neural activity. Our detection of this code, albeit not perfect in performance, is, in most cases, both far above chance levels and is robust to animal identity and laboratory of origin. Our presentation of these results is cohesive in as much as we sequentially establish a series of results that build towards a concluding set of experiments. We start by establishing a baseline via standard measurements and then explore more challenging problems through more complex models that build toward our final test. Based on your feedback, we will contract and expand elements of this sequence.

While our findings raise the possibility of developing a computational tool for electrode localization, pending additional features and/or datasets, our current focus is on establishing the neurobiological principle of anatomical embedding in spike trains. The purpose of briefly mentioning a possible application is that we hope to encourage those engaged in machine-learning on multi-modal neural data that this problem is tractable, yet still open. Based on your feedback, we will clarify that the focus of our current work is not an introduction of a new tool.

https://doi.org/10.7554/eLife.101506.1.sa0