Non-linear Dimensionality Reduction on Extracellular Waveforms Reveals Physiological, Functional, and Laminar Diversity in Premotor Cortex

Eric Kenji Lee[a], Hymavathy Balasubramanian[e], Alexandra Tsolias[b], Stephanie Anakwe[f], Maria Medalla[b], Krishna V. Shenoy[g,i,h,j,k,l], Chandramouli Chandrasekaran[a,b,c,d,*]

[a]*Department of Psychological and Brain Sciences, Boston University, Boston, MA, USA*

[b]*Department of Anatomy & Neurobiology, Boston University School of Medicine, Boston, MA, USA*

[c]*Center for Systems Neuroscience, Boston University, Boston, MA, USA*

[d]*Department of Biomedical Engineering, Boston University, Boston, MA, USA*

[e]*Bernstein Center for Computational Neuroscience, Technische Universität, Berlin, Germany*

[f]*Undergraduate Program in Neuroscience, Boston University, Boston, MA, USA*

[g]*Department of Electrical Engineering, Stanford University, CA, USA*

[h]*Department of Neurobiology, Stanford University, CA, USA*

[i]*Department of Bioengineering, Stanford University, CA, USA*

[j]*Wu Tsai Neurosciences Institute, Stanford University, CA, USA*

[k]*Bio-X Institute, Stanford University, CA, USA*

[l]*Howard Hughes Medical Institute, Stanford University, CA, USA*

Abstract

Cortical circuits involved in decision-making are thought to contain a large number of cell types—each with different physiological, functional, and laminar distribution properties—that coordinate to produce behavior. Current *in vivo* methods rely on clustering of specified features, such as trough to peak duration of extracellular spikes, to identify putative cell types, but these can only capture a small amount of variation. Here, we develop a new method (*WaveMAP*) that combines non-linear dimensionality reduction with graph clustering to identify putative cell types. We apply *WaveMAP* to extracellular waveforms recorded from dorsal premotor cortex of macaque monkeys performing a decision-making task. Using *WaveMAP*, we robustly establish eight waveform clusters and show that these clusters recapitulate previously identified narrow- and broad-spiking types while also revealing undocumented diversity within these sub types. The eight clusters exhibited distinct laminar distributions, characteristic firing rate patterns, and decision-related dynamics.

*Corresponding Author
Email address: cchandr1@bu.edu (Chandramouli Chandrasekaran)

By revealing additional cell type diversity, *WaveMAP* facilitates a more nuanced understanding of how the dynamics of cell types unfolds across cortical layers during decision-making.

Keywords:
dimensionality reduction, cell types, extracellular waveform, unsupervised clustering

---

Significance (120 words)

How different cell types sculpt activity patterns in brain areas associated with decision-making is a fundamentally unresolved problem in neuroscience. In monkeys, and other species where transgenic access is not yet possible, identifying physiological types *in vivo* relies on only a few discrete user-specified features of extracellular waveforms to identify cell types. Here, we show that non-linear dimensionality reduction with graph clustering applied to the entire extracellular waveform can delineate many different putative cell types and does so in an interpretable manner. We show that this method reveals previously undocumented physiological, functional, and laminar diversity in the dorsal premotor cortex of monkeys, a key brain area implicated in decision-making.

Introduction

The processes involved in decision-making, such as deliberation on sensory evidence and the preparation and execution of motor actions, are thought to emerge from the coordinated dynamics within and between cortical layers[1,2], cell types[3–6] and brain areas[7,8]. A large body of research has described differences in decision-related dynamics across brain areas[9–12] and a smaller set of studies has provided insight into layer-dependent dynamics during decision-making[1,2,13,14]. However, we currently do not understand how decision-related dynamics emerge across putative cell types. Here, we address this open question by first developing a new method, *WaveMAP*, that combines non-linear dimensionality reduction and graph-based clustering. We apply this technique to extracellular waveforms to identify putative cell classes and examine their physiological, functional, and laminar distribution properties.

In mice, and to some extent in rats, transgenic tools allow the *in vivo* detection of particular cell types[3,5] while *in vivo* studies in primates are largely restricted to using features of the extracellular action potential (EAP) such as shape and cell firing rate (FR). Early *in vivo* monkey work[15] introduced the importance of EAP features—such as spike duration and action potential (AP) width—in identifying cell types. These experiments introduced the concept of broad- and narrow-spiking neurons. Later experiments in the guinea pig[16] and cat[17] then established the idea that these broad- and narrow-spiking waveform shapes mostly correspond to excitatory pyramidal and inhibitory cells respectively and this has been assumed to also be true in primates. This method of identifying cell types in mammalian cortex *in vivo* is widely used in neuroscience but it is insufficient to capture the known structural and transcriptomic diversity of cell types in the monkey and the mouse[18,19]. Furthermore, recent observations in the monkey defy this simple classification of broad- and narrow-spiking cells corresponding to excitatory and inhibitory cells respectively. Two such examples in the primate that have resisted this idea are narrow-spiking pyramidal tract neurons in deep layers of M1 (Betz cells)[20,21] and narrow-spiking excitatory cells in layer III of V1, V2 and MT[22–25].

To capture a more representative diversity of cell types *in vivo*, more recent studies have incorporated additional features of EAPs (beyond AP width) such as trough to peak duration[26], repolarization time[27,28], and triphasic waveform shape[29,30]. Although these user-specified methods are amenable to human intuition, they often inadequately distinguish between previously identified cell types[20,31,32]. It is also unclear how to choose these user-specified features in a principled manner (i.e., one that maximizes explanatory power) as they are often highly correlated with one another. This results in different studies choosing between different sets of specified features each yielding different inferred cell classes[27,33–35]—thus it is difficult to compare putative cell types across literature. Some studies even conclude that there is no single set of specified features that is a reliable differentiator of type[32,36]. These issues led us to investigate techniques that don't require feature specification but are designed to find patterns in complex datasets through non-linear dimensionality reduction. Such methods have seen usage in diverse neuroscientific contexts such as single-cell transcriptomics[37,38], in analyzing models of biological neural networks[39,40], the identification of behavior[41–43], and in electrophysiology[44–47].

Here, in a novel technique that we term *WaveMAP*, we combine a non-linear dimensionality reduction method (Universal Manifold Approximation and Projection [UMAP][48]) with graph community detection (Louvain clustering[49]) to understand the physiological, functional, and laminar distribution diversity of cell types during decision-making. We applied *WaveMAP* to extracellular waveforms collected from neurons in macaque dorsal premotor cortex (PMd) in a decision-making task using laminar multi-channel probes (16 electrode "U probes"). We found that *WaveMAP* significantly outperformed current approaches without need for user-specification of waveform features like trough to peak duration. This data-driven approach exposed more diversity in extracellular waveform shape than any single constructed spike feature in isolation or in combination. Using interpretable machine learning, we also show that *WaveMAP* picks up on nuanced and meaningful biological variability in waveform shape.

WaveMAP revealed three broad-spiking and five narrow-spiking waveform types that differed in shape, physiological, functional, and laminar distribution properties. Although most narrow-spiking cells had the high maximum firing rates typically associated with inhibitory neurons, some had firing rates similar to broad-spiking neurons which are typically considered to be excitatory. The time at which choice selectivity ("discrimination time") emerged for many narrow-spiking cell classes was earlier than broad-spiking neuron classes—except for the narrow-spiking cells that had broad-spiking like maximum firing rates. Finally, many clusters had distinct laminar distributions that appear layer-dependent in a manner matching certain anatomical cell types. This clustering explains variability in discrimination time over and above previously reported laminar differences[1]. Together, this constellation of results reveals previously undocumented relationships between waveform shape, physiological, functional, and laminar distribution properties that are missed by traditional approaches. Our results provide powerful new insights into how candidate cell classes can be better identified and how these types coordinate with specific timing, across layers, to shape decision-related dynamics.

Results

Task and Behavior

Two male rhesus macaques (T and O) were trained to perform a red-green reaction time decision-making task (Fig. 1A). The task was to discriminate the dominant color of a central static red-green checkerboard cue and to report their decision with an arm movement towards one of two targets (red or green) on the left or right (Fig. 1A).

The timeline of the task is as follows: a trial began when the monkey touched the center target and fixated on a cross above it. After a short period, two targets red and green appeared on the either side of the center target (see Fig. 1B, top). The target configuration was randomized: sometimes the left target was red and the right target was green or vice versa. Another short viewing period later, a red-green checkerboard appeared in the center of the screen with a variable mixture of red and green squares. We parameterized the variability of the checkerboard by its color coherence (C). Larger coherence indicates that there is more of one color than the other (an easy trial) versus a low coherence which indicates that the two colors are more equal in number (a difficult trial). In signed color coherence (SC), large positive numbers indicate the presence of more red squares while large negative numbers indicate more green squares. Numbers close to zero (positive or negative) indicate an almost even number of either. The signed color coherence provides an estimate of whether there are more red or green squares in the checkerboard whereas coherence describes the difficulty of the stimulus (Fig. 1B, bottom).

Our monkeys demonstrated the range of behaviors typically observed in decision-making tasks: monkeys made more errors and were slower for more ambiguous checkerboards compared to less ambiguous completely red or green checkerboards (Fig. 1C, D). We used coherence and reaction times (RT) to analyze the structure of decision-related neural activity.

Recordings and Single Neuron Identification

While monkeys performed this task, we recorded single neurons from the caudal aspect of dorsal premotor cortex (PMd; Fig. 1E, top) using single tungsten or linear multi-contact electrodes (Plexon U-Probes, 625 neurons, 490 U-probe waveforms; Fig. 1E, right) and a BlackRock Microsystems Acquisition System (Cerebus). In this study, we analyzed the average extracellular action potential (EAP) waveforms of these neurons. All waveforms were analyzed after being filtered by a 4th order high-pass Butterworth filter (250 Hz). A 1.6 ms snippet of the waveform was recorded for each spike and used in these analyses, which is a duration longer than many studies of waveform shape. We restricted our analysis to well-isolated single neurons. Extracellular waveforms were identified as single neurons by only accepting waveforms with minimal ISI violations ($1.5\% < 1.5$ ms) and also by careful online inspection of the extracellular waveforms and subsequent offline spike sorting (Plexon Offline Sorter). Every effort was made online to ensure thresholds were conservative so that the best quality spikes were included in the analysis.

We used previously reported approaches to align, average, and normalize spikes[32,50]. Spikes were aligned in time via their depolarization trough and normalized between -1 and 1 in normalized amplitude units. Other common pre-processing methods, such as trough alignment and normalization to trough depth, showed no differences in downstream analyses (Fig. S5B). "Positive spiking" units with large positive amplitude pre-hyperpolarization spikes were dropped from the analysis

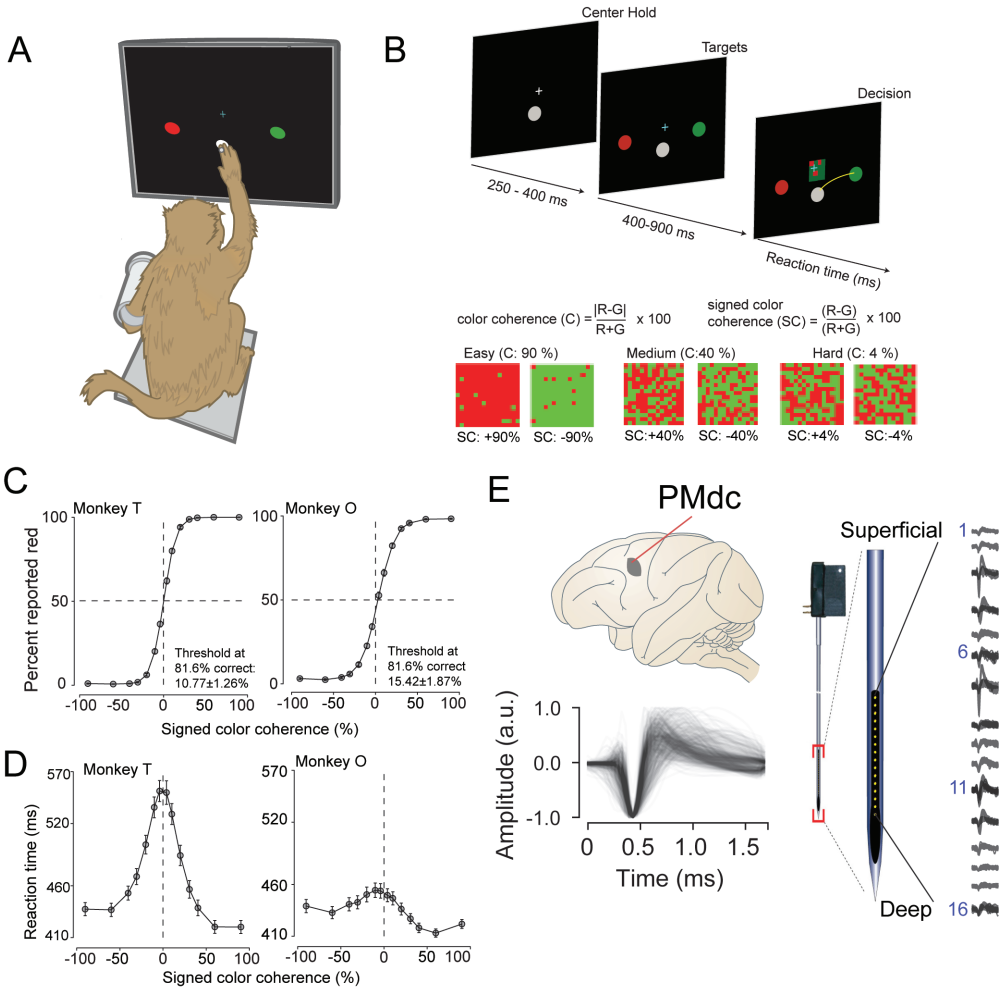

Figure 1: Recording locations, waveform shapes, techniques, task, and discrimination behavior (A) An illustration of the behavioral setup in the discrimination task. The monkey was seated with one arm free and one arm gently restrained in a plastic tube via a cloth sling. An infrared-reflecting (IR) bead was taped to the forefinger of the free hand and was used in tracking arm movements. This gave us an instantaneous readout of the hand's position and allowed us to mimic a touch screen. **(B)** A timeline of the decision-making task (top). At bottom is defined the parametrization of difficulty in the task in terms of color coherence and signed color coherence. **(C)** Average discrimination performance and **(D)** RT over sessions of the two monkeys as a function of the signed color coherence of the checkerboard cue. RT plotted here includes both correct and incorrect trials for each session and then averaged across sessions. Gray markers show measured data points along with 2 $\times$ S.E.M. estimated over sessions. For many data points in (C), the error bars lie within the marker. X-axes in both (C), (D) depict the SC in %. Y-axes depict the percent responded red in (C) and RT in (D). Also shown in the inset of (C) are discrimination thresholds (mean $\pm$ S.D. over sessions) estimated from a Weibull fit to the overall percent correct as a function of coherence. The discrimination threshold is the color coherence (CC) at which the monkey made 81.6% correct choices. 75 sessions for monkey T (128,989 trials) and 66 sessions for monkey O (108,344 trials) went into the averages. **(E)** The recording location in caudal PMd (top); normalized and aligned isolated single-unit waveforms (n = 625, 1.6 ms each, bottom); and schematic of the 16-ch Plexon U-probe (right) used during the behavioral experiment.

due to their association with dendrites and axons[29,35,51]. Recordings were pooled across monkeys to increase statistical power for *WaveMAP*.

Non-linear Dimensionality Reduction with Graph Clustering Reveals Robust Low-Dimensional Structure in Extracellular Waveform Shape

In *WaveMAP*, we use a three step strategy for the analysis of extracellular waveforms: We first parsed intrinsic waveform structure using a non-linear dimensionality reduction method called universal manifold approximation and projection (UMAP)[48]. Second, we used the high-dimensional graph (a fuzzy simplical complex) found by UMAP (the approximation step, the "A" in UMAP) and passed it through a graph-based community detection method, Louvain clustering,[49] to delineate high-dimensional clusters. Third, we then used UMAP to project the high-dimensional data into two dimensions (projection step, the "P" in UMAP). We colored the data points in this projected space according to their Louvain cluster membership found in step two.

The time series for each normalized and trough-aligned waveform (see Fig. 1E, bottom) was passed directly into the UMAP algorithm without any prior dimensionality reduction (except for normalization) and without any feature specification. To identify the presence of clusters in the high-dimensional graph (produced by UMAP), we applied Louvain clustering which finds communities of highly inter-connected data points[49]. These communities appear as color-coded clusters in the UMAP space after projection (Fig. 2A). Louvain clustering was used because it operates directly in high dimensions (Fig. S1). This algorithm is also fairly deterministic (after fixing a random seed) and not susceptible to the variability (in local optima) induced by clustering directly on data in the projected space (see Supplementary Information and Fig. S2).

Louvain clustering identified eight clusters in total. The UMAP algorithm then non-linearly projected the waveforms into a two-dimensional space in a way that preserved local and global variation. Fig. 2A shows results of this analysis. UMAP provides a clear organization without the need for prior specification of important features.

For expository reasons and to link to prior literature[16,52], we loosely subdivide based on the trough to peak duration these eight clusters into "narrow-spiking" and "broad-spiking" supraclusters. The broad-spiking clusters had a median trough to peak duration of 0.76 $\pm$ 0.24 ms (median $\pm$ S.D.) and the narrow-spiking clusters had median trough to peak duration of 0.36 $\pm$ 0.1 ms (median $\pm$ S.D.). The narrow-spiking neurons are shown in warm colors (including green) at right in Fig. 2A and the broad-spiking neurons are shown in cool colors at left in the same figure. The narrow-spiking supracluster was composed of five clusters with "narrow-spiking" waveforms (clusters ①, ②, ③, ④, ⑤) and comprised ~12%, ~12%, ~18%, ~7%, and ~19% (n = 72, 78, 113, 43, and 116) of the total waveforms, for ~68% of total waveforms. The broad-spiking supracluster was composed of three "broad-spiking" waveform clusters (⑥, ⑦, and ⑧) comprising ~13%, ~5%, and ~15% (n = 80, 29, and 94) respectively and collectively ~32% of total waveforms.

To validate that *WaveMAP* finds a "real" representation of the data, we examined if a very different method could pick on the same representation. We trained a gradient-boosted random forest classifier on the waveform data using five-fold cross-validation with the cluster indices as labels. Hyperparameters were tuned with a grid search and final parameters are shown in Table S1. After training, the held-out classification accuracy averaged over clusters was 91%. Fig. 2C shows the associated confusion matrix which contains accuracies for each class along the main diagonal and misclassification rates on the off-diagonals. Such successful classification at high

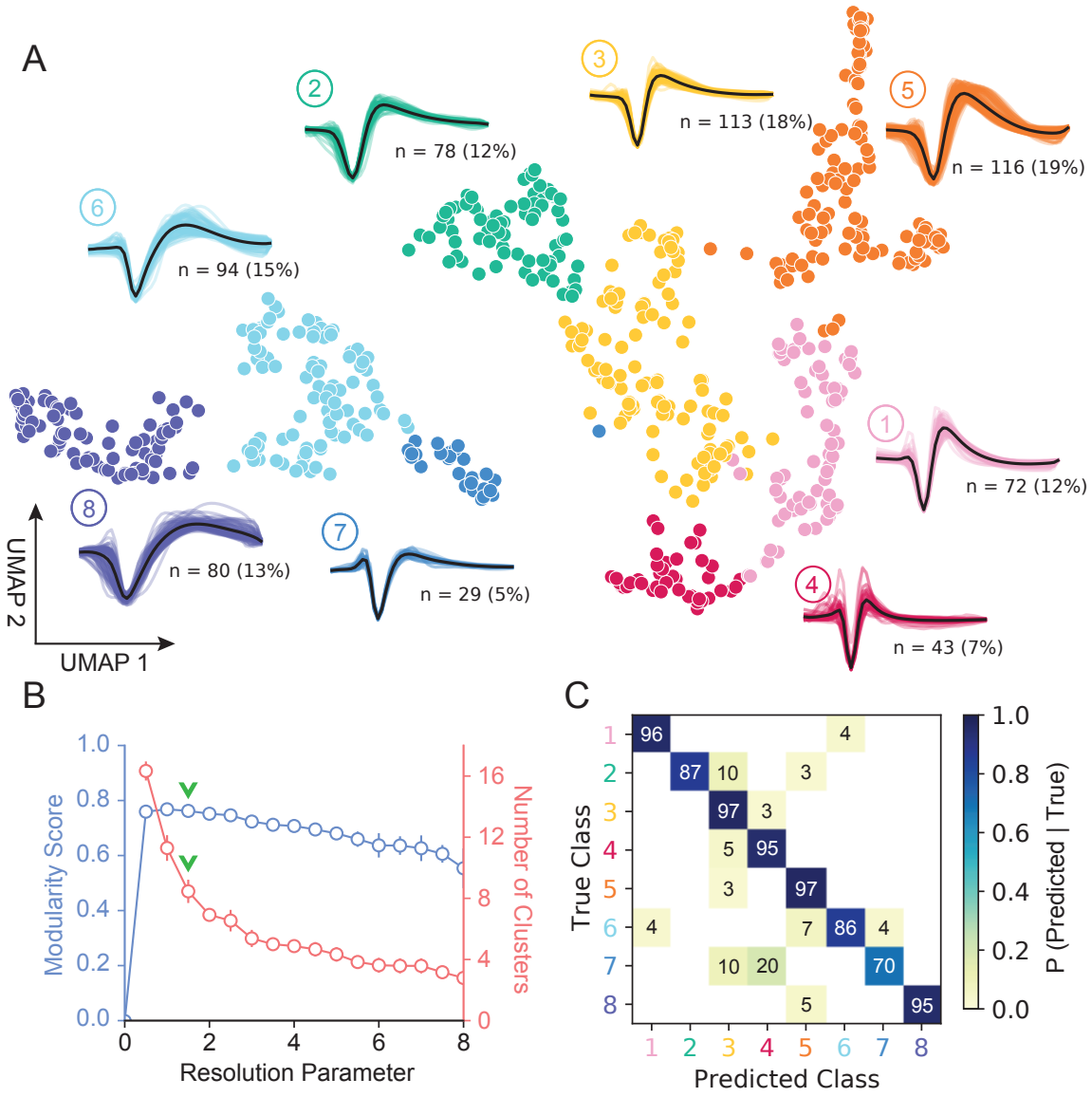

Figure 2: UMAP and Louvain clustering reveal a robust diversity of averaged single-unit waveform shapes (A) Scatter plot of normalized EAP waveforms in UMAP space colored by Louvain cluster membership. Adjacent to each numbered cluster (① through ⑧) is shown all member waveforms and the average waveform shape (in black). Each waveform is 1.6 ms. Percentages do not add to 100% due to rounding. **(B)** Louvain clustering resolution parameter versus modularity score (in blue, at left) and the number of clusters (communities) found (in red, at right). This was averaged over 25 random seeds for UMAP using the full dataset at each resolution parameter from 0 to 8 in 0.5 unit increments. Each data point is the mean ± S.D. with many S.D. bars smaller than the marker size. Green chevrons indicate the resolution parameter of 1.5 chosen and its position along both curves. **(C)** The confusion matrix of a random forest classifier with five-fold cross-validation. The main diagonal shows the held-out classification accuracy for each cluster and the off-diagonals show the misclassification rates for each cluster to each other cluster. The average accuracy for each cluster was 91% across all clusters.

levels of accuracy was only possible because there were learnable clusterings of similar waveform shapes in the high-dimensional space revealed by UMAP.

Traditional Clustering Methods with Specified features Fail to Capture Full Variability

Our unsupervised approach (Fig. 2) generates a stable clustering of waveforms. However, is our method better than the traditional approach of using specified features[27,32,50,53,54]? To compare how *WaveMAP* performs relative to traditional clustering methods built on specified features, we applied a Gaussian mixture model (GMM) to the three-dimensional space produced by three commonly used waveform features using the same dataset. In accordance with previous work, the features we chose (Fig. 3A) were action potential (AP) width of the spike (width in milliseconds of the full-width half minimum of the depolarization trough, a common feature in intra-cellular studies of waveforms[20]), the peak ratio which is the ratio of pre-hyperpolarization peak (A1) to the post-hyperpolarization peak (A2)[29], and the trough to peak duration (time in ms from the depolarization trough to post-hyperpolarization peak) which is the most common feature used in analysis of extracellular recordings[32,50].

The GMM result applied to these three measures is shown in Fig. 3B. This method identified four waveform clusters that roughly separated into broad-spiking (BS, ∼33%, n = 208) which were putatively excitatory, narrow-spiking (NS, ∼43%, n = 269) which are putatively inhibitory, broad-spiking triphasic (BST, ∼9%, n = 55), and narrow-spiking triphasic (NST, ∼15%, n = 93) (Fig. 3B). Broad- and narrow-spiking types are well-documented on the basis of AP width and trough to peak duration. Triphasic waveforms, thought to be neurons with myelinated axons or neurites[29,30,35,56,57], contain an initial positive spike before the trough and can be identified by taking the ratio of the pre-hyperpolarization peak with the post-hyperpolarization peak around 1.0 ms (Fig. 3A). These GMM clusters are similar to those obtained from other clusterings of EAP's in macaque cortex[27,58]. We selected four clusters by calculating the Bayesian information citerion (BIC) statistic as a function of the number of clusters (Fig. 3D). For this dataset and number of features, we saw no meaningful decrease in the BIC statistic beyond 4 clusters (green chevron in Fig. 3C).

To compare the learnability of this representation with the representation provided by UMAP, we trained the same random forest classifier on the waveform data (after separate hyperparameter tuning, Fig. S1) but this time using the four GMM classes as target labels. After training, the accuracy across all four classes averaged ∼78% with no classification accuracy over 95% and misclassifications between every class (Fig. 3D). The classifier trained on specified features under-performed the classifier trained on the whole waveform found by *WaveMAP*. In *WaveMAP*, the individual classification accuracy of most classes exceeded 95% with few misclassifications between groups even though there were double the number of clusters. This result suggests that the clusters based on specified features are less differentiable than *WaveMAP* clusters even when a much lower cluster number is considered.

This deficit can be understood as an inability of the GMM to fully capture the structure of the variability of clusters. If we examine the gray data point shadows (Fig. 3B), no features contain clear clusters and neither do they contain Gaussian distributions which is an assumption of the GMM model. Examining the marginal distributions in Fig. 3B, none of the features induce a clear separability between the clusters alone or in conjunction. Furthermore, the reproducible clusters

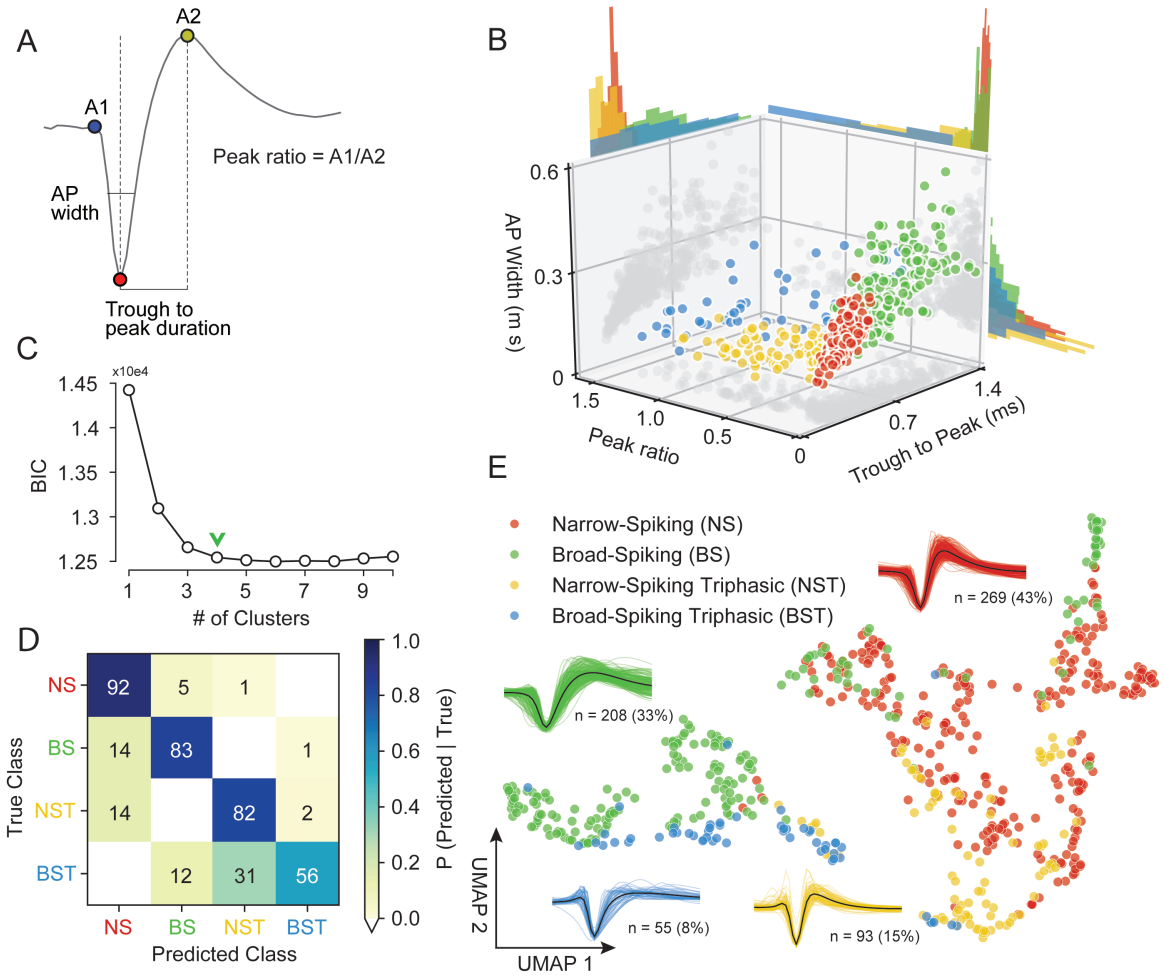

Figure 3: Gaussian mixture model clustering on specified features fails to capture the breadth of waveform diversity (A) The three EAP waveform landmarks used to generate the specified features passed to the GMM on a sample waveform. ● is the pre-hyperpolarization peak (A1); ● is the depolarization trough; and ● is the post-hyperpolarization peak (A2). **(B)** A three-dimensional scatter plot with marginal distributions of waveforms and GMM classes on the three specified features in (A). Narrow-spiking (NS) are in red; broad-spiking (BS) in green; narrow-spiking triphasic (NST) in yellow; and broad-spiking triphasic (BST) types are in blue. Trough to peak was calculated as the time between ● and ●; peak ratio was determined as the ratio between the heights of ● and ● (A1/A2); and AP width was determined as the width of the depolarization trough ● using the MLIB toolbox[55]. **(C)** The optimal cluster number in the three-dimensional feature space in (B) was determined to be 4 clusters using the Bayesian information criterion (BIC)[27]. The number of clusters was chosen to be at the "elbow" of the BIC curve (green chevron). **(D)** A confusion matrix for a gradient-boosted random classifier with five-fold cross-validation. The main diagonal contains the classification accuracy percentages across the four GMM clusters and the off-diagonal contains the misclassification rates. The average accuracy per each class was 78%. **(E)** The same scatter plot of normalized EAP waveforms in UMAP space as in Fig. 2E but now colored by GMM category.

found by *WaveMAP* are linearly inseparable in the feature space of the three GMM features. Labeling the data according to *WaveMAP* cluster identity yields an inseparable collection of points (Fig. S3A).

Note, this is not an artifact of using a lower cluster number in the GMM as opposed to the eight

found by *WaveMAP*. Even if the GMM is instantiated with eight clusters (Fig. S3B), a classifier is still unable to learn this clustering with high accuracy (Fig. S3C) even if the waveforms found by the GMM with eight clusters seem somewhat sensible (Fig. S3D). Thus, it is a deficit of the constructed feature-based approach to capture the full diversity of waveforms and not a peculiarity of the model parameters chosen.

We also investigated the representation of specified features in the projected UMAP space. We color coded the waveforms in UMAP, as in Fig. S4, according to each point's feature values across AP width (Fig. S4B), trough to peak duration (Fig. S4C), and peak ratio (Fig. S4D). We find that *WaveMAP* implicitly captures each of these specified feature shown as a monotonic gradation of values. Our method also exposes the correlation between certain specified features: the gradient (direction of steepest change) between trough to peak duration and AP width points roughly in the same direction thus both features are highly correlated. This correlation between features is another reason why traditional approaches fail to capture the full diversity of waveform shapes.

To obtain a clearer picture of how *WaveMAP* captures clustering missed by specified features, we color the points in UMAP space by their GMM cluster identity in Fig. 3E. Here, *WaveMAP* is able to recapitulate the same structure observed by specified features as a gradation from triphasic to biphasic along the UMAP-2 direction. Our technique also captures the transition from broad- to narrow-spiking along the UMAP-1 direction. *WaveMAP* is also able to capture clusters that occupy an intermediate identity between GMM classes. For instance, *WaveMAP* cluster ② (Fig. 2A) is nearly equal parts broad- and narrow-spiking in the GMM clustering (Fig. 3E). If a GMM were used, ② would be split between two classes despite it having a distinct waveform shape characterized by a small pre-hyperpolarization peak, a moderate post-hyperpolarization peak, and relatively constant repolarization slope.

WaveMAP interpretably recapitulates and expands upon known waveform features

We have established that *WaveMAP* has the ability to discover extracellular waveform clusters but a common contention with such methods is that these approaches are uninterpretable. Here using an interpretable machine learning approach, we show that *WaveMAP* is sensible[59,60]. To identify the features our algorithm is paying attention to, we first computed the inverse mapping of the UMAP transform to probe the projected space in a systematic way. Second, we leveraged the random forest classifier in Fig. 2C and used SHapley Additive exPlanations (SHAP values) to reveal what waveform features are implicitly used to differentiate clusters[61].

To quantify the differences between Louvain clusters, we applied a grid of "test points" to the UMAP projected space (Fig. 4A, top) and inverted the transform at each location; each of these test points is a coordinate on a grid (black **x**'s) and shows the waveform associated with every point in the projected space (Fig. 4A, bottom). On the bottom is shown the waveform that corresponds to each point in UMAP space color-coded to the nearest cluster or to gray if there were no nearby clusters. As UMAP-1 increases, there is a smooth transition in the sign of the inflection of the repolarization slope (the second derivative) from negative to positive (slow to fast repolarization rate). That is, the post-hyperpolarization peak becomes more sharp as we increase in the UMAP-1 direction. As UMAP-2 increases, we see a widening of the post-hyperpolarization slope distinct from the change in its inflection (UMAP-1). These two UMAP dimensions recapitulate the known importance of hyperpolarization properties in clustering waveforms. Both hyperpolarization

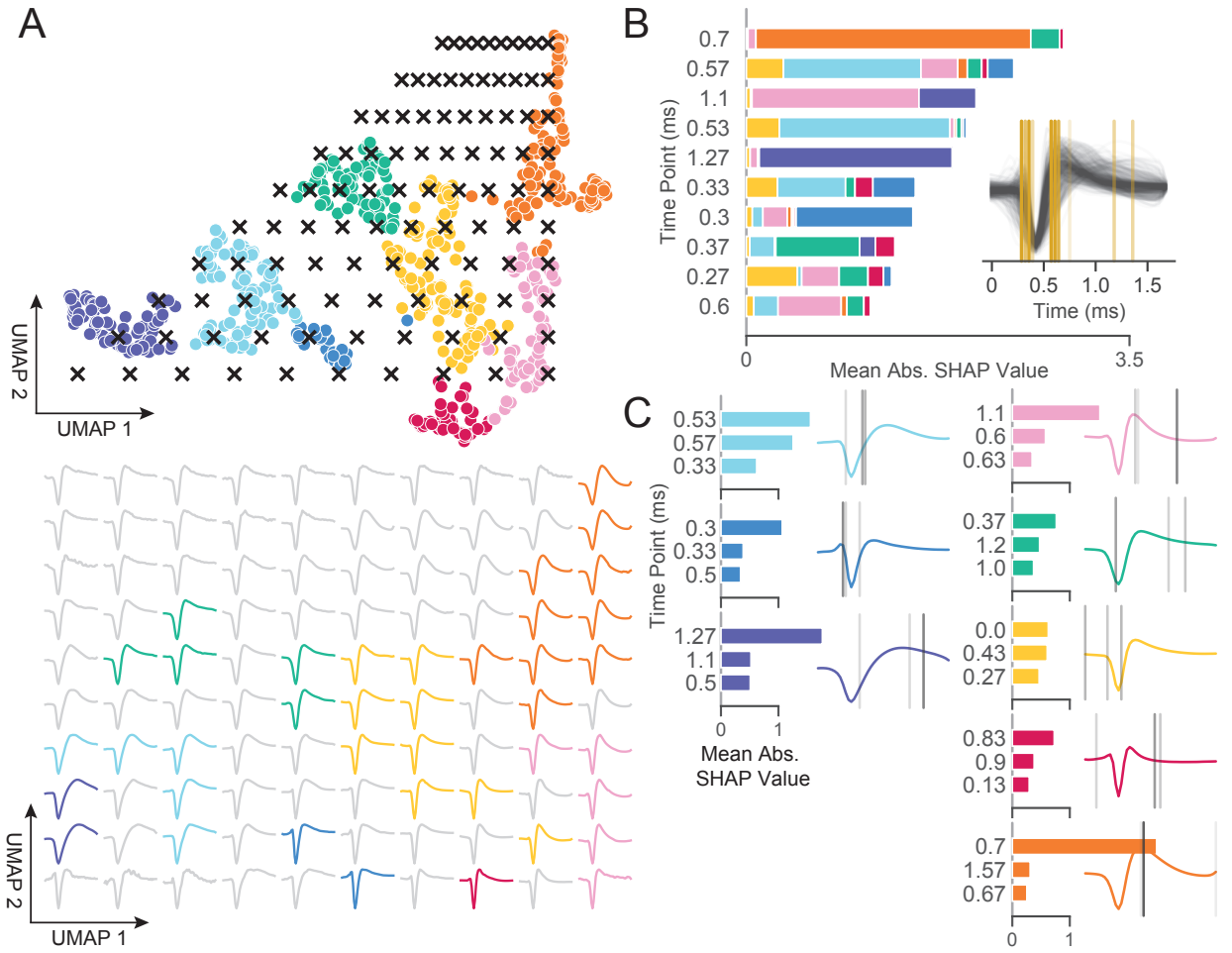

Figure 4: *WaveMAP* provides interpretable representations that both validate and extend known and unknown features importances (A) *WaveMAP* applied to the EAP's as in Fig. 2A overlaid with a grid of test points (black **x**'s, top). At bottom, the inverse UMAP transform is used to show the predicted waveform at each test point. These are plotted and assigned the color of the nearest cluster or in gray if no cluster is nearby. Note that there exists instability in the waveform shape (see waveforms at corners) as test points leave the learned manifold. **(B)** The mean absolute SHAP values for ten time points along all waveforms subdivided according to the SHAP values contributed by each *WaveMAP* cluster. These SHAP values were informed by a random forest classifier trained on the waveforms with the *WaveMAP* clusters as labels. In the inset, all waveforms are shown and in goldenrod are shown the time points for which the SHAP values are shown on the left. Each vertical line is such that the most opaque line contains the greatest summed mean absolute SHAP value across *WaveMAP* clusters; the least opaque, the smallest summed mean absolute SHAP value. **(C)** Each averaged Louvain waveform cluster is shown with the three time points containing the greatest SHAP values for each cluster individually. As before, the SHAP value at each time point is proportional to the opacity of the gray vertical line but now also shown as a bar graph; this bar graph shows the relative SHAP values for the three locations specified by the gray lines.

rate (proportional to trough to peak width) and hyperpolarization slope inflection (proportional to repolarization time) are separate but highly informative properties[26,27]. Furthermore, since repolarization rate and post-hyperpolarization width associate with different UMAP dimensions, this implies that these two processes are somewhat independent factors shaping the waveform and that potassium channel dynamics—which principally govern repolarization rates—might play an important role in differentiating cell types via waveform shape. Thus, *WaveMAP* not only finds an interpretable and smoothly varying low-dimensional space (a "manifold") it also offers possibly yielding biological insights, in this case how cell types might differ according to channel protein expression and dynamics.

In Fig. 4B, we made use of SHAP values to identify which aspects of waveform shape the random forest classifier utilizes in assigning what waveform to which cluster[61,62]. SHAP values build off of the game theoretic quantity of Shapley values[63,64] which poses that each feature (point in time along the waveform) is of variable importance in influencing the classifier to decide whether the data point belongs to a specific class or not. Operationally, SHAP values are calculated by examining the change in classifier performance as each feature is obscured (the waveform's amplitude at each time point in this case), one-by-one[61]. Fig. 4B shows the top-10 time points (in terms of total SHAP value, i.e. importance in explaining a cluster) and their location. It is important to note that not every time point is equally informative for distinguishing every cluster individually and thus each bar is subdivided into the mean absolute SHAP value contribution of the eight constituent waveform classes. For instance, the 0.7 ms location is highly informative for cluster ⑤ and the 0.3 ms point is highly informative for cluster ⑦.

In the inset is shown all waveforms along with each of the top ten time points (in goldenrod) with higher SHAP value shown with more opacity. The time points with highest SHAP value tend to cluster around two different locations giving us an intuition for which locations are most informative for telling apart the Louvain clusters. For instance, the 0.5 to 0.65 ms region contains high variability amongst waveforms and is important in separating out broad- from narrow-spiking clusters. This region roughly contains the post-hyperpolarization peak which is a feature of known importance and incorporated into nearly every study of EAP waveform shape (see Table 2 in ref.[20]). Similarly, SHAP values implicate the region around 0.3 ms to 0.4 ms as time points that are also of importance and these correspond to the pre-hyperpolarization peak which is notably able to partition out triphasic waveforms[29]. Importance is also placed on the location at 0.6 ms corresponding to the inflection point which is similarly noted as being informative[27,28]. These methods also implicate other regions of interest that have not been previously noted in the literature to the best of our knowledge: two other locations are highlighted farther along the waveform at 1.1 and 1.27 ms and are important for differentiating ⑧ and ① from the other waveforms. This result suggests that using only up to 1.0 ms or less of the waveform may obscure diversity.

In Fig. 4C, we show the three locations that are most informative for delineating a specific cluster; these appear as gray lines with their opacity proportional to their importance. These individually-informative features often do align with those identified as globally-informative but do so with cluster-specific weights. Put another way, not every time point is equally informative for identifying waveforms individually and these "most informative" parts of each waveform don't always perfectly align with globally-informative features. In summary, *WaveMAP* independently

and sensibly arrived at a more nuanced incorporation of the very same features identified in previous work—and several novel ones—using a completely unsupervised framework which obviated the need to specify waveform features. In the second half of the paper, we investigate whether these clusters have distinct physiological (in terms of firing rate), functional, and laminar distribution properties which would give evidence that they connect to real cell types.

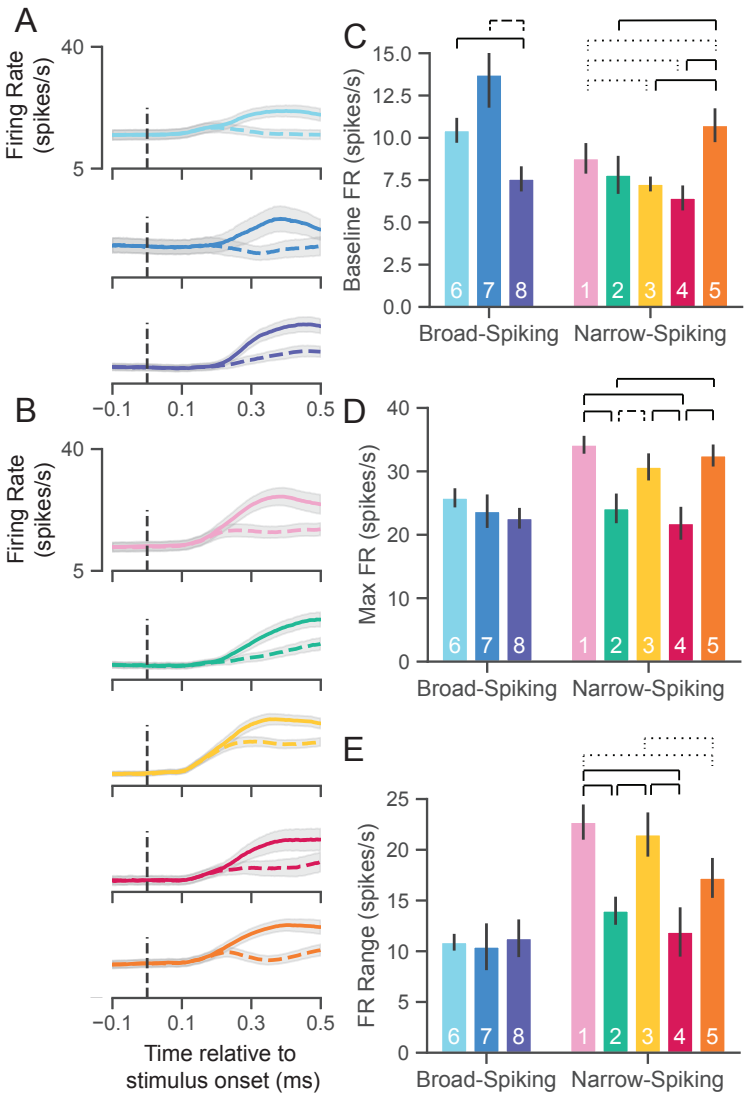

Figure 5: UMAP clusters exhibit distinct physiological properties (A) Stimulus-aligned trial-averaged firing rate activity in PMd for broad-spiking *WaveMAP* clusters. The traces shown are separated into trials for PREF direction reaches (solid lines) and NONPREF direction reaches (dashed lines) and across the corresponding *WaveMAP* clusters. Shaded regions correspond to bootstrapped standard error of the mean. **(B)** The same plots as in (A) but for narrow-spiking *WaveMAP* clusters. **(C)** Baseline median firing rates with S.E.M. for the neurons in the eight different classes. Baselines were calculated as the average firing rate from the first 200 ms of recording before the stimulus appeared. **(D)** Maximum median firing rates with S.E.M. for the neurons in the eight different clusters. This was calculated as the median maximum firing rate for each neuron across the entire trial. **(E)** FR range with S.E.M. calculated as the median difference, per neuron, between its baseline and max FR. ····· $p < 0.05$; - - - - $p < 0.01$; —— $p < 0.005$; Mann-Whitney *U* test, FDR adjusted.

WaveMAP clusters have distinct physiological properties

A defining aspect of cell types is that they vary in their physiology and especially firing rate properties[15,16,52,65–67]. However, these neuronal characterizations via waveform *ex vivo* are not always conserved when the same waveform types are observed *in vivo* during behavior[68,69]. To connect our waveform clusters to physiological cell types *in vivo*, beyond waveform shape, we identified each cluster's firing rate properties. We performed several analyses using the firing rate (FR) in spikes per second (spikes/s) for each cluster during the decision-making task described in Fig. 1.

The trial-averaged FRs are aligned to stimulus onset (stim-aligned) and separated into preferred (PREF, solid trace) or non-preferred (NONPREF, dashed trace) reach direction trials. This is shown for both broad- (Fig. 5A) and narrow-spiking (Fig. 5B) clusters. A neuron's preferred direction (right or left) was determined as the reach direction in which it had a higher FR on average in the 100 ms time period before movement onset.

To further quantify the FR differences between clusters we calculated three properties of the FR response to stimulus: baseline firing rate, max firing rate, and FR range.

Baseline FR: Cell types are thought to demonstrate different baseline firing rates. We estimated baseline firing rate (baseline FR, Fig. 5C) as the median FR across the 200 ms time period before the appearance of the red-green checkerboard and during the hold period after targets appeared for the broad (Fig. 5A), and narrow-spiking clusters (Fig. 5B). The broad-spiking clusters showed significant differences in baseline FR when pooled and compared against the pooled narrow-spiking clusters ($p = 0.0028$, Mann-Whitney *U* test). Similar patterns were observed in another study of narrow vs. broad spiking neurons in PMd during an instructed delay task[53]. We also found that not all broad-spiking neurons had low baseline firing rates and not all narrow-spiking neurons had high firing rates. The broad-spiking clusters ⑥ and ⑦ were not significantly different but both differed significantly from ⑧ in that their baseline FR was much higher (10.3 ± 0.7 and 13.2 ± 1.9 spikes/s vs. 7.6 ± 0.75 spikes/s [median $\pm$ bootstrap S.E.]; $p = 0.0052$, $p = 0.0029$ respectively, Mann-Whitney *U* test, FDR adjusted). The narrow-spiking clusters (Fig. 5B, right) ②, ③, and ④ had relatively low median baseline FRs (7.5 ± 1.1, 7.4 ± 0.4, 6.5 ± 0.7 spikes/s, median $\pm$ bootstrap S.E.) and were not significantly different from one another but all were significantly different from ① and ⑤ ($p = 0.04$, $p = 2.8\text{e-}4$, $p = 2.8\text{e-}7$, $p = 4.9\text{e-}5$ respectively, Mann-Whitney *U* test, FDR adjusted; see Fig. 5C).

Maximum FR: A second important property of cell types is their maximal firing rate[15,16,65]. We estimated the maximum FR for a cluster as the median of the maximum FR of neurons in the cluster in a 1200 ms period aligned to movement onset (800 ms before and 400 ms after movement onset; Fig. 5D). In addition to significant differences in baseline FR, pooled broad- vs. pooled narrow-spiking neurons showed a significant difference in max FR ($p = 1.60\text{e-}5$, Mann-Whitney *U* test). Broad-spiking clusters were fairly homogeneous with low median max FR (24.3 ± 1.0, median $\pm$ bootstrap S.E.) and no significant differences between distributions. In contrast, there was significant heterogeneity in the FR's of narrow-spiking neurons: three clusters (①, ③, and ⑤) had uniformly higher max FR (33.1 ± 1.1, median $\pm$ bootstrap S.E.) while two others (② and ④) were uniformly lower in max FR (23.0 ± 1.4, median $\pm$ bootstrap S.E.) and were comparable to the broad-spiking clusters. Nearly each of the higher max FR narrow-spiking clusters were

significantly different than each of the lower max FR clusters (all pairwise relationships $p < 0.001$ except ② to ③ which was $p = 0.007$, Mann-Whitney *U* test, FDR adjusted).

FR Range: Many neurons, especially inhibitory types, display a sharp increase in firing rates during behavior relative to their baseline. In addition, their FR's span a wide range during behavior[1,32,53,70,71]. To examine this change over the course of a trial, we took the median difference across trials between the max FR and baseline FR per neuron to calculate the FR range. We again found the group difference between pooled broad- and pooled narrow-spiking clusters to be significant ($p = 0.0002$, Mann-Whitney *U* test). Each broad-spiking cluster (⑥, ⑦, and ⑧) had a median increase of around 10.8 spikes/s (10.8 ± 0.8, 10.7 ± 2.3, and 10.9 ± 1.9 spikes/s respectively, median ± bootstrap S.E.) and each was nearly identical in FR range differing by less than 0.2 spikes/s. In contrast, the narrow-spiking clusters differed greatly in their FR range showing, in the same manner as max FR, with significant differences between the clusters ①, ③, and ⑤ having a large range (20.3 ± 1.1 spikes/s, median ± bootstrap S.E.) and the clusters ② and ④ having a relatively smaller range (13.4 ± 1.3 spikes/s, median ± bootstrap S.E.). This demonstrates that some narrow-spiking clusters, in addition to having high baseline firing rates, highly modulated their firing rates over the course of a behavioral trial.

Such physiological heterogeneity in narrow-spiking cells has been noted before in their FR properties[26,28,72] and in some cases attributed to different subclasses of a single inhibitory class[73,74]. Other work also strongly suggests that narrow-spiking cells contain excitatory neurons with distinct FR properties which may contribute to this diversity[20,24].

Furthermore, if *WaveMAP* has truly arrived at a closer delineation of underlying cell types compared to previous methods, it should produce a "better" clustering of physiological properties beyond just a better clustering of waveform shape. To address this issue, we calculate the same firing rate traces and physiological properties as in Fig. 5 but with the GMM clusters (Fig. S6). While the FR traces maintain the same trends (BS does not increase its FR prior to the split into PREF and NONPREF while NS does; compare to *WaveMAP* broad-spiking vs. narrow-spiking clusters respectively), much of the significant differences between clusters is lost across all physiological measures even though less groups are compared (Fig. S6B, C, and D). We also quantitatively reify these differences by calculating the effect sizes (Cohen's f^2) across the *WaveMAP* and GMM clusterings with a one-way ANOVA. The effect size was larger for *WaveMAP* vs. GMM clustering respectively for every physiological property: baseline firing rate (0.070 vs. 0.013), maximum firing rate (0.035 vs. 0.011), and firing rate range (0.055 vs. 0.034).

UMAP clusters have distinct decision-related dynamics

Our analysis in the previous section showed that there is considerable heterogeneity in their physiological properties. Are these putative cell types also functionally different? Prior literature suggests and argues that neuronal cell types have distinct functional roles during cortical computation with precise timing. For instance, studies of macaque premotor[75], interior temporal (IT)[76], and frontal eye field (FEF)[9] areas show differences in decision-related functional properties: between broad- and narrow-spiking neurons, narrow-spiking neurons exhibit choice-selectivity earlier than broad-spiking neurons. In the mouse, specific aspects of behavior are directly linked with inhibitory cell types[3,4]. This would suggest that inhibitory cell types in the monkey are also implicated in behavior although this has yet to be directly assessed. Here we examine the functional properties

of each cluster based on two inferred statistics: choice-related dynamics and discrimination time.

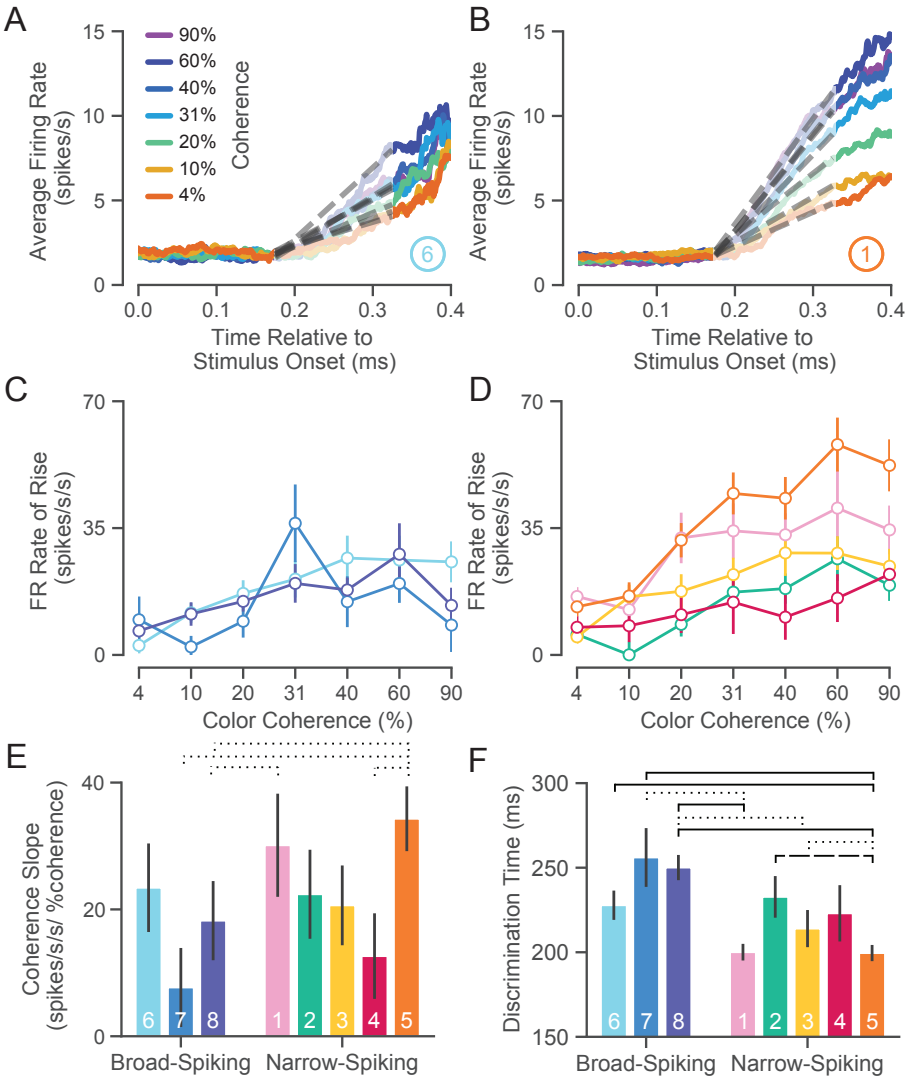

Figure 6: UMAP clusters exhibit distinct functional properties (A) Average FR over time for ⑥ across different coherences. **(B)** Average FR over time for ① across different coherences. **(C)** FR rate of rise vs. coherence for broad- and **(D)** narrow-spiking clusters. **(E)** Bootstrapped median coherence slope is shown with the bootstrapped standard error of the median for each cluster on a per-neuron basis. Coherence slope is a linear regression of the cluster-specific lines in the previous plots Fig. 6C and D. **(F)** Bootstrapped median discrimination time for each cluster with bootstrapped standard error of the median. Discrimination time was calculated as the earliest time at which the choice-selective signal could be differentiated from the FR or the hold period[1]. ····· $p < 0.05$; - - - - $p < 0.01$; —— $p < 0.005$; Mann-Whitney *U* test, FDR adjusted.

Choice-related Dynamics: The first property we assessed for these *WaveMAP* clusters was the dynamics of the choice-selective signal. The neural prediction made by computational models of decision-making (for neurons that covary with an evolving decision) is the build-up of average neural activity in favor of a choice is faster for easier compared to harder color coherences[1,9,11]. Build-up activity is measured by analyzing the rate of change of choice-selective activity vs. time. We therefore examined the differences in averaged stimulus-aligned choice-selectivity signals

(defined as `|left - right|`) for different checkerboard color coherences for each cluster.

In Fig. 6A and B, we show average choice-selectivity signals across seven coherence levels for an example broad- (⑥) and narrow-spiking cluster (①). For ⑥ (Fig. 6A), easier stimuli (higher coherence) only led to modest increases in the rate at which the choice selectivity signal increases. In contrast, ① (Fig. 6B) shows faster rates for the choice-selective signal as a function of coherence. We summarized these effects by measuring the rate of change for the choice-selective signal between 175 and 325 ms for stimulus-aligned trials in each coherence condition (dashed lines in Fig. 6A,B). This rate of rise for the choice-selective signal (spikes/s/s) vs. coherence is shown for broad- (Fig. 6C) and narrow-spiking (Fig. 6D) clusters. The broad-spiking clusters demonstrate fairly similar coherence-dependent changes with each cluster being somewhat indistinguishable and only demonstrating a modest increase with respect to coherence. In contrast, the narrow-spiking clusters show a diversity of responses with ① and ⑤ demonstrating a stronger dependence of choice-related dynamics on coherence compared to the other three narrow-spiking clusters which were more similar in response to broad-spiking neurons.

We further summarized these plots by measuring the dependence of the rate of rise of the choice-selective signal as a function of coherence measured as the slope of a linear regression performed on the rate of rise vs. coherence for each cluster (Fig. 6E). The coherence slope for broad-spiking clusters was moderate and similar to ②, ③, and ④ while the coherence slope for ① and ⑤ was steeper. Consistent with Fig. 6C,D, the choice selective signal for ① and ⑤ showed the strongest dependence on stimulus coherence.

Discrimination time: The second property that we calculated was the discrimination time for clusters which is defined as the first time in which the choice-selective signal (again defined as `|left - right|`) departed from the FR of the hold period. We calculated the discrimination time on a neuron-by-neuron basis by computing the first time point in which the difference in FR for the two choices was significantly different from baseline using a bootstrap test (at least 25 successive time points significantly different from baseline FR corrected for multiple comparisons[1]). Discrimination time for broad-spiking clusters (255 $\pm$ 94 ms, median $\pm$ bootstrap S.E.) was significantly later than narrow-spiking clusters (224 $\pm$ 89 ms, $p < 0.005$, median $\pm$ bootstrap S.E., Mann-Whitney *U* test). Clusters ① and ⑤, with the highest max FRs (34.0 $\pm$ 1.4 and 33.0 $\pm$ 1.8 spikes/s, median $\pm$ S.E.) and most strongly modulated by coherence, had the fastest discrimination times as well (200.0 $\pm$ 4.9 and 198.5 $\pm$ 4.9 ms, median $\pm$ S.E.).

Together the analysis of choice-related dynamics and discrimination time showed that there is considerable heterogeneity in the properties of narrow-spiking neuron types. Not all narrow-spiking neurons are faster than broad-spiking neurons and choice-selectivity signals have similar dynamics for many broad-spiking and narrow-spiking neurons. ① and ⑤ have the fastest discrimination times and strongest choice dynamics. In contrast, the broad-spiking neurons have uniformly slower discrimination times and weaker choice-related dynamics.

WaveMAP clusters contain distinct laminar distributions

In addition to having certain physiological properties and functional roles, numerous studies have shown that cell types across phylogeny, verified by single-cell transcriptomics, are defined by distinct patterns of laminar distribution in cortex[18,77]. Here we examined the laminar distributions of *WaveMAP* clusters and compared them to laminar distributions of GMM clusters. The number

of waveforms from each cluster was counted at each of sixteen U-probe channels separately. These channels were equidistantly spaced every 0.15 mm between 0.0 to 2.4 mm. This spanned the entirety of PMd which is 2.5 mm in depth from the pial surface to white matter[78]. However, making absolute statements about layers is difficult with these measurements because of errors in aligning superficial electrodes with layer I across different days. This could lead to shifts in estimates of absolute depth; up to 0.15 mm (the distance between the first and second electrode) of variability is induced in the alignment process (see Methods). However, relative comparisons are likely better preserved. Thus, we use relative comparisons to describe laminar differences between distributions and in comparison to anatomical counts in fixed tissue in later sections.

Above each column of Fig. 7A, B are the laminar distributions for all waveforms in the associated supracluster (in gray); below these are the laminar distributions for each supracluster's constituent clusters. On the right (Fig. 7C), we show the distribution of all waveforms collected at top in gray with each GMM cluster's distribution shown individually below.

The overall narrow- and broad-spiking populations did not differ significantly according to their distribution ($p = 0.24$, Kolmogorov-Smirnov test). The broad-spiking supracluster of neurons (⑥, ⑦, and ⑧) are generally thought to contain cortical excitatory pyramidal neurons enriched in middle to deep layers[16,79]. Consistent with this view, we found these broad-spiking clusters (Fig. 7A) were generally centered around middle to deep layers with broad distributions and were not significantly distinguishable in laminarity (all comparisons $p > 0.05$, two-sample Kolmogorov-Smirnov test, FDR adjusted).

In contrast, narrow-spiking clusters (Fig. 7B) were distinctly varied in their distribution such that almost every cluster had a unique laminar distribution. Cluster ① contained a broad distribution. It was significantly different in laminar distribution from clusters ② and ④ ($p=0.002$ and $p = 0.013$ respectively, two-sample Kolmogorov-Smirnov, FDR adjusted).

Cluster ② showed a strongly localized concentration of neurons at a depth of 1.1 ± 0.33 mm (mean ± S.D.). It was significantly different from almost all other narrow-spiking clusters ($p = 0.002$, $p = 1e\text{-}5$, $p = 0.010$ for ①, ④, and ⑤ respectively; two-sample Kolmogorov-Smirnov test, FDR adjusted). Similarly, cluster ③ also showed a strongly localized laminar distribution but was situated more superficially than ② with a heavier tail (1.0 ± 0.6 mm, mean ± S.D.). This cluster was more superficial than ②.

Cluster ④ was restricted uniquely deep in its cortical distribution (1.70 ± 0.44, mean ± S.D.). These neurons had a strongly triphasic waveform shape characterized by a large pre-hyperpolarization peak. These waveforms have been implicated as arising from myelinated excitatory pyramidal cells[29] especially dense in this caudal region of PMd[80].

The last cluster, ⑤, like ① was characterized by a broad distribution across cortical depths unique among narrow-spiking neurons and was centered around a depth of 1.3 ± 0.65 mm (mean ± S.D.) and present in all layers[78].

Such laminar differences were not observed when we used GMM clustering. Laminar distributions for BS, BST, NS, and NST did not significantly differ from each other (Fig. 7C; BS vs. BST had $p = 0.067$, all other relationships $p > 0.2$; two-sample Kolmogorov-Smirnov test, FDR adjusted). Each GMM cluster also exhibited broad distributions across cortex which is at odds with our

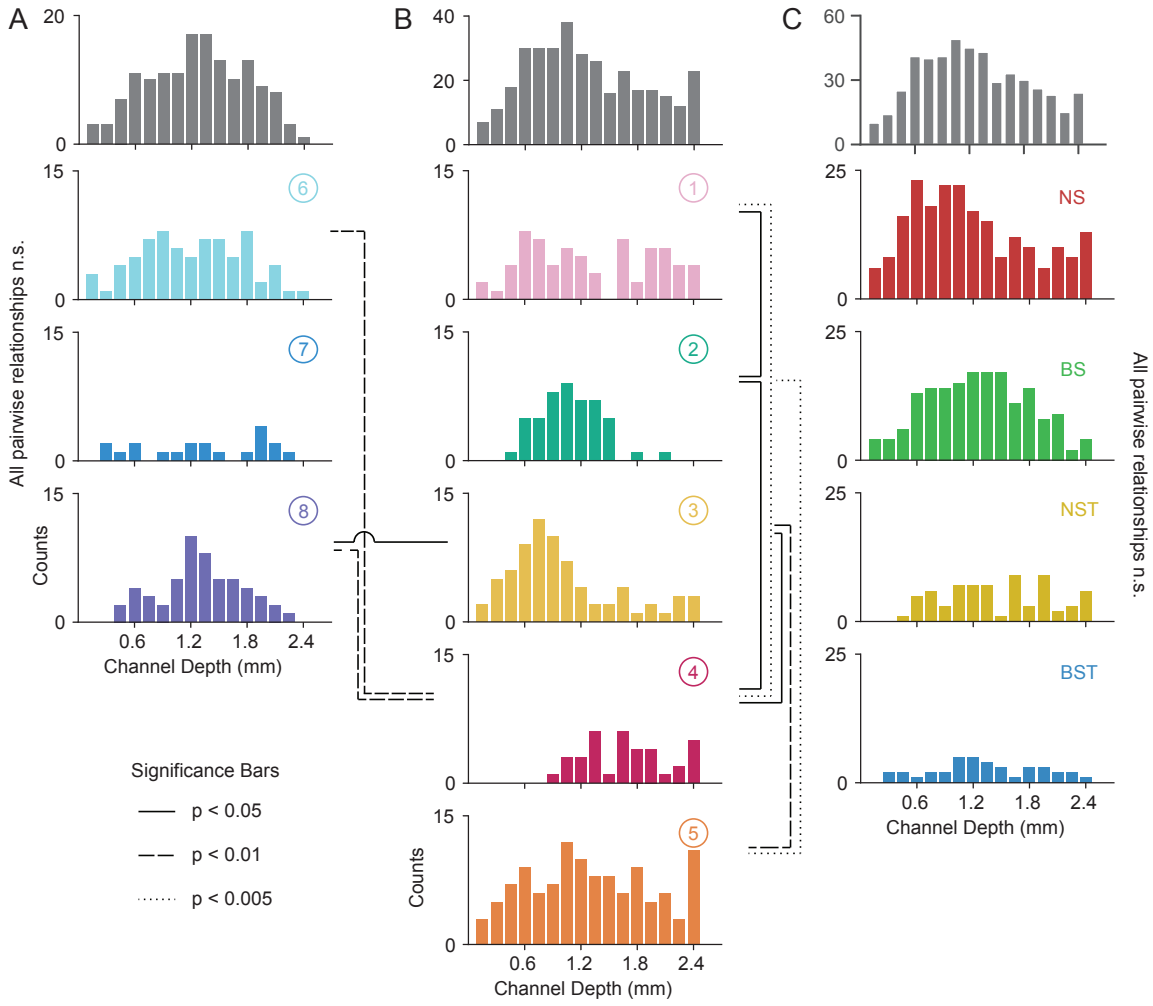

Figure 7: Laminar distribution of *WaveMAP* waveform clusters (A, B) The overall histogram for the broad- and narrow-spiking waveform clusters are shown at top across cortical depths on the left and right respectively (in gray); below are shown histograms for their constituent *WaveMAP* clusters. These waveforms are shown sorted by the cortical depth at which they were recorded from the (0.0 mm (pial surface) to 2.4 mm in 0.15 mm increments). Broad-spiking clusters were generally centered around middle layers and were less distinct in their differences in laminar distribution. Narrow-spiking clusters are shown on the right and were varied in their distribution with almost every cluster significantly varying in laminar distribution from every other. **(C)** Depth histograms for all waveforms collected (top, in gray) and every GMM cluster (below). ····· $p < 0.05$; - - - - $p < 0.01$; —— $p < 0.005$; two-sample Kolmogorov-Smirnov Test, FDR adjusted.

understanding of cell types using histology.

Some narrow-spiking WaveMAP cluster laminar distributions align with inhibitory subtypes

We have shown that *WaveMAP* clusters have more distinct laminarity than GMM clusters. If *WaveMAP* clusters are consistent with cell type, we should expect their distributions to be relatively consistent with distributions from certain anatomical types visualized via immunohistochemistry (IHC). An especially well-studied set of non-overlapping anatomical inhibitory neuron types in the monkey are parvalbumin-, calretinin-, and calbindin-positive GABAergic interneurons (PV^+, CR^+, and CB^+ respectively)[81]. Using IHC, we examined tissue from macaque rostral PMd stained for

each of these three interneuron types. We then conducted stereological counting of each type averaged across six exemplars to quantify cell type distribution across cortical layers (see Fig. 8A and B)[82] and compared it to the distributions in Fig. 7.

Both CB^+ and CR^+ cells (Fig. 8C and D respectively) exhibited a similarly restricted superficial distribution most closely resembling ③. In addition, CR^+ and CB^+ cells are known to have very similar physiological properties and spike shape[83]. An alternative possibility is that one of CR^+ or CB^+ might correspond to ② and the other to ③ but this is less likely given their nearly identical histological distributions (Fig. 8C and D) and similar physiology[83].

In contrast, *WaveMAP* cluster ①, had laminar properties consistent with PV^+ neurons (Fig. 8B): both were concentrated superficially but proliferated into middle layers (Fig. 8E). In addition, there were striking physiological and functional similarities between ① and PV^+ cells. In particular, both ① and PV^+ cells have low baseline firing rate, early responses to stimuli and robust modulation of firing rate. This feature of PV^+ cells was elicited through optogenetic studies in the mouse that show PV^+ neurons in M1 have low baseline firing rate, early responsiveness to stimuli, high peak firing rate, and short discrimination time[4]. Cluster ⑤ also had similar properties to ① and could also correspond to PV^+ cells.

Together, these results from immunohistochemistry suggest that the narrow-spiking clusters identified from *WaveMAP* potentially map on to different inhibitory types.

Heterogeneity in decision-related activity emerges from both cell type and layer

Our final analysis examines whether these *WaveMAP* clusters can explain some of the heterogeneity observed in decision-making responses in PMd above previous methods[1]. Heterogeneity in decision-related activity can emerge from two sources: laminar location that presumably reflects inputs and outputs or variability that can emerge from different cell types within each layer. To examine whether *WaveMAP* clusters and neuron depth captured the same or different aspects of the microcircuit, we regressed discrimination time on both separately and together. To compare these, we examined the change in variance explained (adjusted R^2). We then compared this against the GMM clusters with depth information to show that *WaveMAP* better explains the heterogeneity of decision-related responses.

We previously showed that some of the variability in decision-related responses is explained by the layer from which the neurons are recorded[1]. However, as our results here show, there is considerable variability within cortical layers. We therefore examined if adding cell type information in addition to cortical depth improved our explained variability in decision-related activity. Consistent with previous work, we found that cortical depth explains some variability in discrimination time (1.7%). We next examined if the *WaveMAP* clusters identified explain variability in discrimination time: a categorical regression between *WaveMAP* clusters and discrimination time, explained a much larger 6.6% of variance. Including both cortical depth and cluster identity in the regression explained 7.3% of variance in discrimination time.

Repeating this analysis for GMM clusters, we found that GMM clusters regressed against discrimination time only explained 3.3% of variance and the inclusion of both GMM cluster and cortical depth only explained 4.6% of variance. Thus, we find that *WaveMAP* explains a much larger variance explained relative to depth alone or with traditional methods. This demonstrates

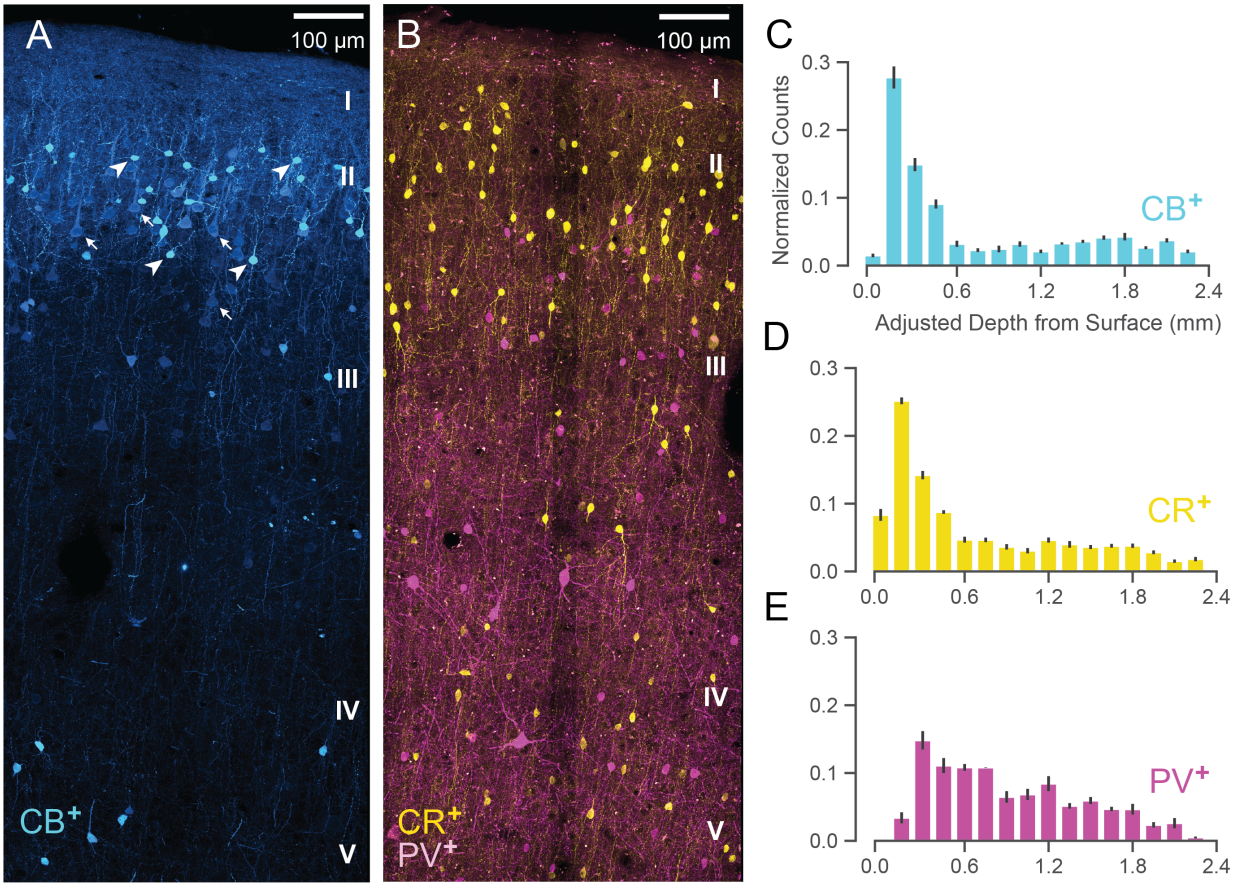

Figure 8: Anatomical labeling of three inhibitory interneuron types in PMd: **(A)** Maximum intensity projection of immunohistological staining of rostral PMd CB^+ interneurons in blue. Note the many weakly-positive excitatory pyramidal neurons (arrows) in contrast to the strongly-positive interneurons (arrowheads). Only the interneurons were considered in stereological counting. In addition, only around first 1.5 mm of tissue is shown (top of layer V) but the full tissue area was counted down to 2.4 mm. Layer IV exists as a thin layer in this area. Layer divisions were estimated based on depth and referencing Arikuni *et al.*[78]. **(B)** Maximum intensity projection of immunohistological staining of PMd CR^+ and PV^+ interneurons in yellow and fuschia respectively. The same depth of tissue and layer delineations were used as in (A). **(C, D, E)** Stereological manual counts[82] (mean ± S.D.) of CB^+, CR^+, PV^+ cells in PMd respectively. Counts were collected from six specimens and normalized at each depth.

that *WaveMAP* clusters come closer to cell types than previous efforts and are not artifacts of layer-dependent decision-related inputs. That is, both the cortical layer in which a cell type is found as well *WaveMAP* cluster membership contributes to the variability in decision-related responses. Furthermore, they outperform GMM clusters as regressors of a functional property associated with cell types. These results further highlight the power of *WaveMAP* to separate out putative cell types and help better understand decision-making circuits.

Discussion

Our goal in this study was to further understand the relationship between waveform shape and the physiological, functional, and laminar distribution diversity of cell populations in dorsal premotor cortex during perceptual decision-making. Our approach was to develop a new method *WaveMAP* that combines a recently developed non-linear dimensionality reduction technique (UMAP) with graph clustering (Louvain community detection) to uncover hidden diversity in extracellular waveforms. We found this approach not only replicated previous studies by distinguishing between narrow- and broad-spiking neurons but did so in a way that 1) revealed additional diversity , and 2) obviated the need to examine particular waveform features. In this way, our results demonstrate how traditional feature-based methods obscure biological detail, which is more faithfully revealed by our *WaveMAP* method. Furthermore, through interpretable machine learning, we show our approach not only leverages many of the features already established as important in the literature but expands upon them in a more nuanced manner—all with minimal supervision or stipulation of priors. Given the successes of standard UMAP across a variety of data domains, we envision that *WaveMAP* will provide improvements in these areas with its additions of graph-based clustering and interpretable machine learning. These diverse applications include computational ethology[41,42,84], analyzing multi-scale population structure[85], and even metascientific analyses of the literature[86].

Advantages of WaveMAP over traditional methods

At the core of *WaveMAP* is UMAP which has important advantages over other non-linear dimensionality reduction methods that have been applied in this context, such as t-distributed stochastic neighborhood embedding (t-SNE)[44]. While providing visualization, these methods are difficult to cluster upon because they return a different mapping on every initialization. UMAP, in contrast, is a manifold learning method that projects efficiently into arbitrary dimensionalities while also returning an invertible transform. That is, we can pass new data into the projected space without having to recompute the mapping. This property provides two advantages over other dimensionality reduction approaches: First, it allows exploration of any region of the manifold. This yields an intuitive understanding of how UMAP non-linearly transforms the data which might be related to underlying biological phenomena. Thus, UMAP allows *WaveMAP* to go beyond a "discriminative model" typical of other clustering techniques and function as a "generative model" with which to make predictions. Second, it enables cross-validation of a classifier trained on cluster labels. This is impossible with methods that don't return an invertible transform. To cross-validate unsupervised methods, unprocessed test data must be passed into the transform computed *only* on training data[87]. This is only possible if a proper transform is admitted by the method of dimensionality reduction as in UMAP.

A final advantage of UMAP is that it inherently allows for not just unsupervised but supervised and semi-supervised learning whereas other methods do not[88]. This key difference enables "transductive

inference" which is making predictions on unlabeled test points based upon information gleaned from labeled training points. This opens up a diverse number of novel applications in neuroscience through informing the manifold learning process with biological ground truths (in what is called "metric learning")[89,90]. Experimentalists could theoretically pass biological ground truths (e.g., cell type information derived using opto- or pharmaco-genetic methods *in vivo* or transcriptomics post-experiment) to *WaveMAP* as training labels. This would in effect "teach" *WaveMAP* to produce a manifold that more closely hews to true underlying diversity. For instance, if experimentalists "opto-tag" neurons of a particular cell type[44,91–94], this information can be passed along with the extracellular waveform to *WaveMAP* which would, in a semi-supervised manner, learn manifolds better aligned to biological truth. A learned manifold could be used in future experiments to identify cell types in real-time without opto-tagging. This could be done by projecting the averaged waveforms found within an experiment into the learned *WaveMAP* manifold. This method would be especially useful in a scenario in which the number of electrodes exceeds the number of channels available to record from simultaneously and not all cell types are of equal interest to record (e.g., Neuropixels probes which have 960 electrodes but simultaneously record from only 384 channels[95,96]). We believe this is a rich area that can be explored in future work.

We use a fully-unsupervised method for separating and clustering waveform classes associated with distinct laminar distributions and functional properties in a decision-making task. One concern with fully unsupervised methods is that the features used for separation are unclear. However, by applying an interpretable machine learning method[61,63], we showed that our unsupervised methods utilized many of the same waveform features derived by hand in previous work but did so in a single unifying framework. Our interpretable machine learning approach shows how each waveform feature delineates certain waveform clusters at the expense of others and—more importantly—shows how they can be optimally combined to reveal the full diversity of waveform shapes.

Our novel approach of using non-linear dimensionality reduction with graph clustering on the population of extracellular action potentials compared to specified waveform features has parallels with the evolution of new approaches for the analysis of neuronal firing rates in relevant brain areas[97–101]. Classically, the approach to analyzing firing rates involved in cognition was to develop simple metrics that separated neurons recorded in relevant brain areas. For instance, tuning is used to separate neurons in the motor[102] and visual cortex[103]. Similarly, visuomotor indices that categorize neurons along a visual to motor continuum are used to understand firing rates during various tasks in the frontal eye fields[104] and premotor cortex[1]. However, these specified features quash other aspects of a firing rate profile in favor of focusing on only a few other aspects. New approaches to analyze firing rates use dimensionality reduction techniques such as principal component analysis[97,98,105], tensor component analysis[106], demixed principal component analysis[107], targeted dimensionality reduction[99], and autoencoder neural networks[108]. These methods have provided insight into heterogeneous neural activity patterns in many brain areas without the need for specified features like tuning or a visuomotor index. Our study adopts a similar approach for analysis of extracellular waveforms by examining whole populations and strongly suggests that non-linear dimensionality reduction methods applied to the entire extracellular waveform are better than using hand-derived waveform features such as trough to peak duration, repolarization time, spike width and other metrics. This progression also follows similar trends in machine learning towards data-driven approaches.

Waveform cluster shapes are unlikely to arise from electrode placement

It is a possibility is that the diversity of waveforms we observe is just an artifact of electrode placement relative to the site of discharge. This supposes that waveform shape changes with with respect to the distance between the neuron and electrode. This is unlikely because both *in vitro* studies[56] and computational simulations[109] show distance from the soma induces changes in amplitude but not shape. We controlled for this variation in amplitude by normalizing spikes during preprocessing.

It is another possibility that different neuronal structures present different waveform shapes. Put another way, the diversity we see could be due to recording from different morphological structures (dendrites, soma, or axons) rather than different cell types. While it is true that there are some cell structures associated with different waveform shapes (such as triphasic waveforms near neurites, especially axons[29,30,35,56]), highly-controlled *in vitro* studies show that a large majority of EAP's are from somata (86%)[56]. In concordance with this, we only observed one cluster (④, 6% of all EAP's) with a triphasic shape. These waveforms were only found in deep layers where myelination is prevalent. Thus we believe that almost all of our waveforms come from somata, with the possible exclusion of ④.

It is also unlikely that the waveform diversity seen is due to the location of the electrode relative to the neuron given the observed differences in firing rate (Fig. 5), functional role (Fig. 6), and laminarity (Fig. 7) associated with each *WaveMAP* cluster.

Given that electrode location has little effect on waveform shape, we might then ask what about a neuron's waveform shape, in terms of cellular physiology, is captured by *WaveMAP*? We propose that the space found by UMAP-1 and UMAP-2 sensibly covaries according to documented properties of K^+ ion channel dynamics. As UMAP-1 increases, we observe a smooth transition of the inflection of the repolarization slope from negative to positive (slow to fast repolarization rate). Said differently, the post-hyperpolarization peak becomes sharper as we increase in the UMAP-1 direction. These observations are consistent with the same gradual change in intracellular AP repolarization slope facilitated by the kinetics of the fast voltage-gated Kv3 potassium-channel in an activity-dependent manner[110]. These channels are necessary for sustained high-frequency firing[111]. In the UMAP-2 direction, there is a smooth decrease in the width of the post-hyperpolarization peak and this direction roughly traverses from broad- to narrow-spiking to triphasic waveforms. This gradual change too has been noted as being associated with the kinetics of the Kv3 potassium-channel: blocking this channel in a dose-dependent manner with tetraethylammonium induces a gradual widening of post-hyperpolarization peak width[112,113]. Both of these changes in intracellular waveform shape likely have a strong effect on the shape of extracellular waveforms[114].

Better parcellation of waveform variability leads to biological insight

Our results show a greater proportion of narrow- (putatively inhibitory) vs. broad-spiking (putatively excitatory) neurons (69% vs. 31% respectively); this appears inconsistent with anatomical studies. These studies demonstrate, through direct labeling of cell type, that in the macaque cortex, 65-80% of neurons are excitatory while 20-35% are inhibitory. We are not the only study to report this puzzling result: Kaufman and colleagues[53] note a proportion of "intermediary" cells that could be classified as either narrow- or broad-spiking. Onorato and colleagues[24] also report greater numbers of narrow-spiking compared to broad-spiking neurons in monkey V1. Thus, care must be

taken when attempting links between spike waveform metrics and cell type. A resolution to this discrepancy is to rethink equating narrow-spiking to inhibitory cells and broad-spiking to excitatory cells. In the monkey, not all inhibitory neurons have narrow spikes[58,115] and not all excitatory neurons have broad spikes[20,21,24]. Anatomical studies show that a substantial number of excitatory neurons in the monkey motor and visual cortices express the Kv3.1b potassium channel which is known to confer neurons with the ability to produce action potentials of narrow spike width and high firing rate[22,25,116,117]. Furthermore, *in vivo* studies show narrow-spiking neurons can be excitatory in motor[20,21] and visual cortices[22,24,25,116] of the macaque.

We therefore believe prior studies have underexplored the diversity of classes accessed by their physiological recordings. Histograms of peak width (and other specified features) across literature are often not cleanly bimodal[27,31,53], and the relative proportions of narrow vs. broad is often dependent on the cutoff chosen for narrow- vs. broad-spiking neurons which span a wide range[20]. Analyses like ours which look at entire waveforms—rather than a few specified features—extract this diversity from extracellular recordings wheres specified features mix waveform classes.

We also find that many narrow-spiking subtypes in PMd signal choice earlier than broad-spiking neurons in our decision-making task (Fig. 6F). These observations are consistent with another study of PMd in monkeys in reach target selection and movement production[75]. In this study, narrow-spiking neurons signaled the selected target 25 ms earlier than broad-spiking neurons. Our results are also consistent with other studies of narrow- vs. broad-spiking neurons in the frontal eye fields (FEF)[9] and inferior temporal area (IT)[76] during decision-making. In these studies, narrow-spiking neurons had higher firing rates before movement onset compared to broad-spiking neurons—a result consistent with our observations for some "narrow-spiking" PMd neurons. Our analyses recapitulate these results and provide additional insights into how different narrow-spiking cell types correlate with decisions. We reproduce the result that narrow-spiking cells, as a whole, have a lower discrimination time than broad-spiking cells but in addition we show that certain narrow-spiking cells respond as slowly as broad-spiking cells (② and ④; Fig. 6F). This lends further evidence to our theory that ② and ④ are likely narrow-spiking excitatory cells.

We observed further heterogeneity in narrow-spiking cells with different clusters occupying different discrete areas of cortex. One of our narrow-spiking clusters (cluster ②) was localized to more superficial layers (Fig. 7B) and had functional properties—low baseline firing rate and longer discrimination times—which are thought to be more closely aligned to properties of excitatory neurons. Furthermore, it is known that there are large populations of narrow-spiking excitatory cells in other areas (PMd is as of yet unstudied in this manner)[21]. Similarly, another narrow-spiking cluster (④) exhibited physiological and functional properties to ② (all comparisons not significant in Fig. 5C, D, and E or Fig. 6E and F) but with a distinct laminar distribution Fig. 7B and highly triphasic waveform shape Fig. S7B. In contrast to ② which concentrated in layer III, ④ was restricted to deep layers which are known to contain large pyramidal cells, often corticospinal. This is also where large-amplitude triphasic waveforms appear most often[29,30,35]. We suspect that cluster ② is similar to the one reported in a recent study of the primary visual cortex[24] while ④ likely contains axons or initial segments of deep narrow-spiking pyramidal cells[20,21,117].

Neuropixels and optogenetics can provide better insight into candidate cell classes in the monkey

Our recordings here were performed with 16 channel U-probes and provided reasonable estimates of

laminar organization for these different putative cell classes. Use of high-density electrophysiological methods providing higher electrode counts perpendicular to the cortical surface would provide further insight into the laminar organization of different cell types[96]. High-density recording would allow us to perform *WaveMAP* in an additional dimension (across multiple electrodes) to increase confidence in identified cell classes and localization of signal to somata[44]. Sensitive electrodes providing spatial access to neural activity[96] can also improve our understanding of how these cell classes are organized both parallel and perpendicular to cortical surface[118,119]. This would allow for the identification of "me-types" through electromorphology[37,45].

Another powerful tool that has been leveraged in the study of cell types during behavior is optogenetics[3,5,6]. Although in its infancy relative to its use in the mouse, optogenetics in monkeys offers direct interrogation of cell types. Future studies that combine opto-tagging in monkeys with simultaneous application of *WaveMAP* and other waveform classification methods will allow us to more precisely link our putative cell classes to function[120] and do so *in vivo*. NHP optogenetics is slowly advancing and efforts in many research groups around the world are producing new methods for *in vivo* optogenetics[121]. We expect future experiments using the promising new mDlx[122] and h56d[123] promoter sequences to selectively opto-tag inhibitory neurons or PV^{+} neurons directly[124] will greatly benefit validation of these derived cell classes.

Author Contributions

CC trained monkeys and recorded in PMd using multi-contact electrodes under the mentorship of KVS. EKL developed the UMAP and Louvain Clustering method (*WaveMAP*) with interpretable machine learning under the mentorship of CC. HB, EKL, and CC expanded the previous GMM approach for clustering based on waveform features. EKL and CC analyzed *WaveMAP* clusters for their properties. AT and MM performed immunohistochemistry experiments and imaging. SA, EKL, and AT performed stereological cell counting. EKL and CC wrote initial drafts of the paper. All authors refined further drafts contributing analyses, insights, and writing.

Acknowledgments

We thank M. Noichl and L. McInnes for their suggestions on graph community detection and UMAP visualization. We also thank Dr. Kathleen Rockland, Dr. Jennifer Luebke and Dr. Jonathan Kao for detailed comments on the previous versions of the manuscript.

CC was supported by a NIH/NINDS R00 award R00NS092972, the Moorman-Simon Interdisciplinary Career Development Professorship from Boston University, the Whitehall foundation, and the Young Investigator Award from the Brain and Behavior Research Foundation. MM was supported by a NIH-NIMH K99/R00MH101234 grant and a NIH-NIMH R01MH116008 grant. KVS was supported by the following awards: NIH Director's Pioneer Award 8DP1HD075623, NIDCD R01-DC014034, NIDCD U01-DC017844, NINDS UH2-NS095548, NINDS UO1-NS098968, DARPA-BTO 'REPAIR' Award N66001-10-C-2010, DARPA-BTO 'NeuroFAST' award W911NF-14-2-0013, Simons Foundation Collaboration on the Global Brain awards 325380 and 543045, Office of Naval Research award N000141812158, Larry and Pamela Garlick, Wu Tsai Neurosciences Institute at Stanford, the Hong Seh and Vivian W. M. Lim endowed professorship and the Howard Hughes Medical Institute. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Declaration of interests

K.V.S. consults for Neuralink Corp. and CTRL-Labs Inc. (part of Facebook Reality Labs) and is on the scientific advisory boards of MIND-X Inc., Inscopix Inc. and Heal Inc. All other authors have no competing interests. These companies provided no funding and had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Methods (3,000 words)

Subjects and Surgery

Our experiments were conducted using two adult male macaque monkeys (*Macaca mulatta*; monkey T, 7 years, 14 kg; O, 11 years, 15.5 kg) that were trained to reach to visual targets for a juice reward. Our Monkeys were housed in a social vivarium with a normal day/night cycle. All experimental protocols were approved by the Stanford University Institutional Animal Care and Use Committee (IACUC). After initial training to come out of the cage and sit comfortably in a chair, monkeys underwent sterile surgery for implantation of head restraint holders (Crist Instruments, cylindrical head holder) and standard recording cylinders (Crist Instruments, Hagerstown, MD).

We placed our cylinders over caudal PMd (+16, 15 stereotaxic coordinates) and surface normal to the cortex. We covered the skull within the cylinder with a thin layer of dental acrylic/PALACOS bone cement.

Apparatus

Monkeys sat in a customized chair (Crist Instruments, Snyder Chair) with the head restrained via the surgical implant. The arm not used for reaching was loosely restrained using a tube and a cloth sling. Experiments were controlled and data were collected under a custom computer control system (xPC target and Psychtoolbox-3[125]). Visual stimuli were displayed on an Acer HN2741 computer screen placed approximately 30 cm from the monkey and a photodetector (Thorlabs PD360A) was used to record the onset of the visual stimulus at a 1 ms resolution. Every session, we taped a small infrared reflective bead (11.5 mm, NDI Digital passive spheres) 1 cm from the tip of the middle digit of the right hand (left hand, monkey O). The position of this bead was tracked optically in the infrared (60 Hz, 0.35 mm root mean square accuracy; Polaris system; Northern Digital).

Eye position was tracked with an overhead infrared camera (estimated accuracy of 1°, Iscan, Burlington, MA). To get a stable eye image for the overhead infrared camera, an infrared dichroic mirror was positioned at a 45° angle (facing upward) immediately in front of the nose. This mirror reflected the image of the eye in the infrared range while letting visible light pass through. A visor placed around the chair prevented the monkey from touching the infrared mirror, the juice tube, or bringing the bead to his mouth.

Behavioral training

Our animals were trained using the following operant conditioning protocol. First, the animal was rewarded for arm movements toward the screen and learnt to take pieces of fruit on the screen. Once the animal acquired the association between reaching and reward, the animal was conditioned to reach and touch a target for a juice reward. The position, as well as the color of this target, was then randomized as the monkey learned to touch targets of various colors at different locations on the screen. We then used a design in which the monkey first held the central hold for a brief period, and then a checkerboard cue, which was nearly 100% red or 100% green, appeared for 400–600 ms and finally the two targets appeared. The monkey received a reward for making a reach to the color of the target that matched the checkerboard cue. Two-target "Decision" blocks were interleaved with single target blocks to reinforce the association between checkerboard color and the correct target. After two weeks of training with this interleaved paradigm, the animal reliably reached to the target matching the color of the central checkerboard cue. We switched the paradigm around by adopting a design in which the targets appeared before the checkerboard cue onset. We initially trained on holding periods (where the monkeys view targets) from 300 to 1800 ms. We trained the animal to maintain the hold on the center until the checkerboard cue appeared by providing small amounts of juice at rough time intervals. When the animal reliably avoided breaking central hold during the holding period, we stopped providing small amounts of juice. After the animal learned to stay still during the target viewing period, we introduced more difficult checkerboard cues to the animal while reducing the maximal holding period to 900 ms. We then trained the animal to discriminate the checkerboard as accurately and as fast as possible while discouraging impulsivity by adopting timeouts.

Electrophysiological recordings

Known response properties of PMd and M1, the neural responses to muscle palpation, and stereotactic coordinates were used as our guides for electrophysiological recordings. Our chambers were placed surface normal to cortex to align with the skull of the monkey and recordings were performed perpendicular to the surface of the brain. Recordings were made anterior to the central sulcus, lateral to the spur of the arcuate sulcus, and lateral to the precentral dimple. For both monkeys, we were able to identify the upper and lower arm representation by repeated palpation at a large number of sites to identify muscle groups associated with the sites. Recordings were performed in the PMd and M1 contralateral to the arm used by the monkey. Monkey T used his right arm (O used his left arm) to perform tasks.

A subset of the electrophysiological recordings were performed using traditional single electrode recording techniques. Briefly, we made small burr holes through the PALACOS/acrylic using handheld drills. We then used a Narishige drive with a blunt guide tube placed in firm contact with the dura. Recordings were obtained using FHC electrodes to penetrate the overlying dura. Every effort was made to isolate single units during the recordings with FHC electrodes by online monitoring and seeking out well-isolated signals.

We performed linear multi-contact electrode (U-probe) recordings in the same manner as single electrode recordings with some minor modifications. We used a slightly sharpened guide tube to provide more purchase on dura. We also periodically scraped away, under ketamine-dexmetotomidine anesthesia, any overlying tissue on the dura. Both these modifications greatly facilitated penetration of the U-probe. We typically penetrated the brain at very slow rates (~2–5 μm/s). Once we felt we had a reasonable sample population of neurons, potentially spanning different cortical layers, we stopped and waited for 45–60 min for the neuronal responses to stabilize. The experiments then progressed as usual. We used 180-μm thick 16-electrode U-probes with an inter-electrode spacing of 150 μm; electrode contacts were 100 KΩ in impedance.

We attempted to minimize the variability in U-probe placement on a session-by-session basis. Our approach was to place the U-probe so that the most superficial electrodes (electrodes 1, 2 on the 16 channel probe) were in layer I and able to record multi-unit spiking activity. Any further movement of the electrode upwards resulted in the spiking activity disappearing and a change in the overall activity pattern of the electrode (suppression of overall LFP amplitudes). Similarly, driving the electrodes deeper resulted in multiphasic extracellular waveforms and also a change in auditory markers which were characterized by decreases in overall signal intensity and frequency content. Both markers suggested that the electrode entered white matter. Recording yields and electrode placement were in general much better in monkey T (average of ~16 units per session) than monkey O (average of ~9 units per session). We utilized these physiological markers as a guide to place electrodes and thus minimize variability in electrode placement on a session-by-session basis. Importantly, the variability in placement would act against our findings of depth-related differences shown in Fig. 7.

Preprocessing of single-unit recordings

We obtained 996 extracellularly recorded single units (778 units recorded with the U-probe) from PMd across two monkeys (450 from Monkey O and 546 from Monkey T). Of these, we identified 801 units whose ISI violations (refractory period $\leq$ 1.5 ms) $\leq$ 1.5%[1]. Our waveforms were filtered

with a 4th-order 250 Hz high-pass Butterworth filter. The waveforms for each of the units were extracted for a duration of 1.6 ms with a pre-trough period of 0.4 ms, sampled at 30 kHz.

Alignment and normalization of waveforms

In order to calculate the mean waveform for each single unit, we upsampled individual waveforms calculated over different trials by a factor of 10 and aligned them based on the method proposed in[32]. For each waveform, we calculated its upswing slope (slope between trough to peak) and the downswing slope (slope to the trough) and re-aligned to the midpoint of the slope that exceeded the other by a factor of 1.5. Following this alignment, we chose the best set of waveforms for calculating the mean as those that satisfied the criteria (1) less the 2 standard deviations (S.D.) from the mean at each point and (2) average deviation from the mean across time is less than 0.4[32]. The final set of waveforms for each unit was averaged and downsampled to 48 time points. Upon visual inspection, we then identified 761 units (625 single units with 490 U-probe recorded units) whose average waveforms qualified the criteria of exhibiting a minimum of two phases with a first trough trajectory and the remaining waveforms, unless stated otherwise here, were removed from the analysis. We excluded positive-spiking waveforms because of their association with axons[35]. Finally, we normalized the waveforms by dividing the extreme value of the amplitude such that the maximum deviation is ± 1 unit[50].

It is important to note that the preprocessing we use, individual mean subtraction and ± 1 unit normalization, operates independently of the data. Using another commonly used preprocessing normalization, normalization to trough depth, we obtained extremely similar results. We found ± 1 unit trough to peak normalization had virtually the same number of clusters as normalization to trough (8.29 ± 0.84 vs. 8.16 ± 0.65 clusters, mean $\pm$ S.D.; Fig. S5A and C). Furthermore, both normalizations picked out the same structure (Fig. S5B and D; the normalization to trough did have a 9th cluster splitting off of ⑤ but this was something also seen with ± 1 unit trough to peak normalization in certain data subsets as well.

WaveMAP: UMAP and Louvain clustering

The normalized extracellular waveforms were passed into the Python package `umap 0.4.0rc3`[48] with the parameters shown in S1. The `n_neighbors` value was increased to 20 to induce more emphasis on global structure. UMAP utilizes a stochastic k-nearest neighbor search to establish the graph and stochastic gradient descent to arrive at the embedding thus it produces similar but different embeddings in the projected space. For reproducibility reasons, the `random_state` was fixed in the algorithm and in `numpy`. The choice of random seed only impacted the projection and not the clustering (Fig. S2A). From here, the graph provided by `umap.graph_` was passed into the Louvain community detection algorithm to generate the clustering seen in Fig. 2A. For details of the UMAP algorithm, see Supplementary Information.

Graph networks are often hierarchical and it has been recommended that the Louvain resolution parameter be chosen to elicit the phenomenon of interest[126,127]. To select the resolution parameter t, we chose a value that best balanced a maximization of modularity score (a measure of the ratio between connections within a cluster vs. incoming from outside of it; see Supplementary Information) while still returning an interpretable number of clusters given our data size. We selected a resolution parameter at the "elbow" of the number of clusters and at the top of the modularity score plateau of maximum values (yellow marker on Fig. 2B). Choosing a lower

resolution parameter resulted in communities of size lower than 20 units which precluded statistical comparison against other communities Fig. S2A. These scores were calculated over 25 random UMAP instantiations of the full dataset. For algorithmic details of Louvain clustering, see Supplementary Information.

To validate that our parameter selection was stable and produced the same number of clusters reliably, we used a bootstrap and applied the *WaveMAP* procedure to random subsets of the full data set Fig. S2B. We obtained 100 random samples from 10% to 100% of the full data set in 10% increments while simultaneously choosing a different random seed for the UMAP algorithm each time. Thus the variances due to sampling and random instantiation are compounded and shown together in Fig. S2B.

Ensemble clustering for graphs (ECG)[128,129] was used to validate the clusters found in Fig. 2A (see Fig. S2C). We added the algorithm (https://github.com/ftheberge/Ensemble-Clustering-for-Graphs) into the `python-igraph` package[130] and passed UMAP graphs into it directly. We set the number of partitions k to be 10 to produce the plot in Fig. S2C. This algorithm uses k different randomized instantiations of the clusters in the graph followed by one round of Louvain clustering (Fig. S1B. Each of these k (called level-1 partitions since one round of Louvain was performed) graphs are then combined as a single graph such that when edges co-occur between nodes in one of the k graphs, it is more heavily weighted. This ensembling of several graphs via the weight function $W_{\mathcal{P}}$ (see Supplemental Methods) yields the final ECG graph.

Gradient-boosted random forest classifier

We then trained a gradient boosted random forest classifier in `xgboost 1.0.2`[131] with five-fold cross-validation on the *WaveMAP* cluster labels in Fig. 2A. A 30% - 70% test-train split was used. Optimal hyperparameters were obtained after a grid search using `scikit-learn`'s `GridSearchCV` function with parameters in Table S1. The default objective function `binary:logistic` was also used. The percent accuracy for each cluster against all others is plotted as a confusion matrix in Fig. 2C.

The same procedure was used when training on the GMM labels found in Fig. 3D and for the eight cluster GMM labels in Fig. S3B. Each of these classifiers also separately underwent hyperparameter tuning using `scikit-learn`'s `GridSearchCV` function as well with final hyperparameters shown in S1.

It is important to note that cross-validation was done after the cluster labels were generated which results in data leakage[87,132] hurting out-of-dataset performance and thus classifier performance only demonstrates UMAP's ability to sensibly separate waveforms within-dataset relative to traditional GMM methods (Fig. 3D). Fig. 2C is not meant to show how such a classifier would perform out-of-dataset.

Specified waveform shape features

To compute specified features for each normalized waveforms (Fig. 3A), we first up-sampled the waveforms from 48 to 480 time points using a cubic spline interpolation method. We then used this up-sampled waveform to compute three separate features: trough to peak duration, AP width, and peak ratio. Trough to peak is the time from the bottom of the depolarization trough (global minimum) to the post-hyperpolarization peak (subsequent local maximum). AP width was

calculated as the width of the depolarization trough at the full-width half-minimum point. Both these measures were found using the `mwave` function from the `MLIB 1.7.0.0` toolbox. Peak ratio was the ratio of heights (above baseline) between the pre-hyperpolarization (maximum before trough) and the post-hyperpolarization peak (maximum after trough).

Gaussian mixture model clustering

Using the specified feature values (trough to peak, AP width, and peak ratio), the normalized waveforms were clustered in the three-dimensional feature space using a Gaussian mixture model (GMM) with hard-assignment (each data point belongs to one cluster) through MATLAB's `fitgmdist` function across 50 replicates (Fig. 3B). Each replicate is a different random instantiation of the GMM algorithm and the model with the largest log likelihood is chosen.

The Bayesian information criterion (BIC) was used to determine the optimal cluster number and is defined as

$$\mathrm{BIC} = -2\ln P(X|\theta) + K\ln(n) \tag{1}$$

where the first term $-2\ln P(X|\theta)$ is the negative log likelihood function i.e. the conditional probability of observing the sample X given a vector of parameters θ. In the partular case of GMM, the function $P(X)$ is a sum of multivariate Gaussians. The second term $K\ln(n)$ is a penalty on model complexity (a regularization) thus capturing the idea that "simple is better" and ultimately constraining the number of Gaussians used to fit the data. It is a function of the number of free model parameters K and the size of the dataset n.

Assuming we have N_f features and N_c clusters we can calculate K using the following framework. For each Gaussian, the total number of parameters is N_f means and $N_f(N_f+1)/2$ covariance parameters. Another free parameter that is learned is the weight for each Gaussian that sums up to 1, leaving us with $N_c - 1$ unique weights. Thus the K which is the effective number of parameters is,

$$K = N_c\left(N_f + \frac{N_f(N_f+1)}{2}\right) + N_c - 1 \tag{2}$$

The "best" model in a BIC-sense will have the set of parameters θ maximizing the likelihood function (thus minimizing the negative log likelihood) for a given model or model family—a number of multivariate Gaussians in a three-dimensional feature space in this case. To arrive at the parameters best approximating the Gaussian distribution giving rise to the data (Maximum Likelihood Estimation or MLE), the Expectation-Maximization (EM) algorithm was used. The optimal cluster number was selected as the lowest number of clusters between 1 and 10 at which the change in BIC was minimal (at the "elbow" in Fig. 3C).

Interpretable machine learning: UMAP inverse transform and SHAP

To facilitate interpretability, we used the invertibility of the UMAP transform (which itself is based on Delauney triangulation) to generate test waveforms tiling the manifold of the projected space Fig. 4A. 100 evenly-spaced test coordinates were generated spanning the manifold and passed backwards through the UMAP transform using `umap`'s built-in `inverse_transform` function. The waveform generated at each test point is shown color-coded to the nearest cluster color or in gray if the distance exceeds 0.5 units in UMAP space.

Using the package `shap` (https://github.com/slundberg/shap), SHAP values were calculated for the classifier trained on *WaveMAP* identified clusters. The trained `XGBoost` model was passed directly into the model-agnostic `shap.TreeExplainer`[62] which then calculated SHAP values for all waveform time points (features). TreeExplainer assigned SHAP values for every time point class-by-class and these were used to generate the class-specific SHAP plots (Fig. 4C). The SHAP values for each time point, across classes, was summed to generate the overall SHAP values for each time point (Fig. 4B).

Choice-selective signal

We use an approach developed by Meister and Huk to estimate the choice-selective signal[133]. We chose such an approach because decision-related activity of PMd neurons is not simply increases and decreases in firing rate and often shows considerable temporal modulation. We estimated for each neuron a choice-selective signal on a time point-by-time point basis as absolute value of the firing rate difference between left and right choice trials (`|left - right|`) or equivalently `|PREF-NONPREF|`. We use this choice-selective signal to understand choice-related dynamics and estimate discrimination time.

Discrimination time

We identified the discrimination time, that is the time at which the neuron demonstrated significant choice selectivity, on a neuron-by-neuron basis. We compared the choice-selective signal at each point to the 95th percentile of the bootstrap estimates of baseline choice-selective signal (i.e., before checkerboard stimulus onset). We enforced the condition that the choice-selective signal should be significantly different from the baseline for at least 25 ms after this first identified time to be included as an estimate of a time of significant choice selectivity for that neuron. Using longer windows provided very similar results.

Choice-related dynamics

To understand the dynamics of the choice-selectivity signal as a function of the unsigned checkerboard coherence, we performed the following analysis. As described above, we first estimated the choice-selectivity signal in spikes/s for each neuron and each checkerboard coherence as shown for example in Fig. 6A,B. We then estimated the slope of this choice-selectivity signal in the 175-325 millisecond period after checkerboard onset. Repeating this analysis for each unsigned checkerboard coherence provided us with an estimate of the rate of change of the choice selectivity signal (η) for all the coherences is spikes/s/s. Averaging over neurons for each cluster provided us with the graphs in Fig. 6C,D. We then estimated the dependence of η on unsigned coherence by regressing η and coherence to estimate how strong choice-selectivity signals in a particular cluster are modulated by the checkerboard input and are summarized in Fig. 6E.

Experimental subjects (anatomical data)

Archived tissues were harvested from six young rhesus macaques of both sexes (9 $\pm$ 1.13 years, *Macaca mulatta*). These subjects were close in age to the macaques used in the main study and were part of part of a larger program of studies on aging and cognition led by Dr. Douglas Rosene. These monkeys were obtained from the Yerkes National Primate Center and housed individually in the Laboratory Animal Science Center at the Boston University School of Medicine; these facilities are fully accredited by the Association for Assessment and Accreditation of Laboratory Animal

Care (AAALAC). Research was conducted in strict accordance with the guidlines of the National Institutes of Health's Guide for the Care and Use of Laboratory Animals and Public Health Service Policy on the Humane Care and Use of Laboratory Animals.

Perfusion and fixation

All brain tissue for histological studies was fixed and harvested using our well-established two-stage perfustion protocol as described[134]. After sedation with ketamine hydrochloride (10 mg/ml) and deep anesthetization with sodium pentobarbital (to effect, 15 mg/kg i.v.), monkeys were perfused through the ascending aorta with ice-cold Krebs–Henseleit buffer containing (in mM): 6.4 Na_2HPO_4, 1.4 NA_2PO_4, 137.0 NaCl, 2.7 KCl, and 1.0 $MgCl_2$ at pH 7.4 (Sigma-Aldrich) followed by fixation with 4% paraformaldehyde in 0.1M phosphate buffer (PB, pH 7.4, 37 °C). The fixed brain sample was blocked, *in situ*, in the coronal plane, removed from the skull and cryoprotected in a series of glycerol solutions, and flash frozen in 70 °C isopentane[135]. The brain was cut on a freezing microtome in the coronal plate at 30 µm and were systematically collected into 10 matched series and stored in cryoprotectant (15% glycerol, in 0.1M PB, pH 7.4) at -80 °C[136].

Immunohistochemistry

To assess the laminar distribution of interneurons we batch processed 30 µm coronal sections through the rostral dorsal premotor cortex area (PMdr) from six specimens. Sections were immunolabelled for inhibitory neuronal subtypes based on their expression of calcium binding proteins, calbindin (CB), calretinin (CR), and parvalbumin (PV), which label non-overlapping populations in primates[81]. Free floating sections were first rinsed (3 x 10 min, 4 °C) in 0.01M phosphate-buffered saline (PBS) and incubated in 50 mM glycine for 1 hr at 4 °C. Sections were then rinsed in 0.01M PBS (3 x 10 min, 4 °C), and antigen retrieval was performed with 10 mM sodium citrate (pH 8.5) in a 60-70 °C water bath for 20 min. Sections were then rinsed in 0.01M PBS (3 x 10 min, 4 °C) and incubated in pre-block (0.01M PBS, 5% bovine serum albumin [BSA], 5% normal donkey serum [NDS], 0.2% Triton X-100) to reduce any non-specific binding of secondary antibodies. Primary antibodies (Fig. S2) were diluted in 0.1 M PB, 0.2% acetylated BSA (BSA-c), 1% NDS, 0.1% Triton X-100. To increase the penetration of the antibody, two microwave incubation sessions (2 × 10 min at 150 watts) using the Pelco Biowave Pro (Ted Pella), followed by a 48 hour incubation at 4 °C with gentle agitation. After rinsing (3 x 10 min) in 0.01M PBS at 4 °C, sections were co-incubated with secondary antibodies diluted in incubation buffer (see S2), microwaved 2 × 10 min at 150 W, and placed at 4 °C for 24 hours with gentle agitation. Sections were then rinsed (3 x 10min) in 0.1M PB, mounted onto slides and coverslipped with prolong anti-fade gold mounting medium (ThermoFisher) and cured at room temperature in the dark.

Confocal microscopy and visualization of immunofluorescent labeling

Immunofluorescent labeling was imaged using a laser-scanning confocal microscope (Leica SPE) using 488 and 561 nm diode lasers. For each coronal section, two sets of tile scan images of a cortical column, ~200 µm wide and spanning, from pia to the white matter boundary, were obtained in the PMdr. This corresponded to the area 6FR in cytoarchitectrual maps[80,137,138] and area F7 in several functional maps[139,140]. The two columns were spaced 200 µm apart. All images were acquired using a plain apochromat 40x/1.3 NA oil-immersion objective at a resolution of 0.268 x 0.268 x 0.5 µm voxel size. The resulting image stacks were deconvolved and converted to 8-bit images using AutoQuant (Media Cybernetics) to improve the signal to noise ratio[134].

Stereological cell counting

Due to its demonstrated ability in producing minimally-biased results, 3D stereologic cell counting[82] was utilized to count parvalbumin- (PV^+), calretinin- (CR^+) and calbindin-positive (CB^+) cells. Using the CellCounter plugin in `Fiji`[141] on each image stack after maximum intensity projection, the inhibitory cells were counted slice by slice, recognized by their round shape (as opposed to pyramids), lack of apical dendrite, and relatively high uniform intensity. Cells at the bottom slice of each image stack and touching the left image border were excluded to avoid double-counting.

Statistics

All statistical tests (Kolmogorov-Smirnov, Kruskal-Wallis, and Mann-Whitney *U*) were conducted using the package `scipy.stats`[142]. Multiple comparisons were corrected for using false detection-rate adjusted p-values (Benjamini-Hochberg); this was done using `scipy.stats.multitest` and `scikit-posthocs`[143]. Ordinary least squares regressions were conducted in the package `statsmodels`[144]. Bootstrapped standard errors of the median were calculated by taking 5,000 random samples with replacement (a bootstrap) of a dataset and then the standard deviation of each bootstrap was taken. Effect sizes were given as adjusted R^2 values or Cohen's f^2 (of a one-way ANOVA) using `statsmodels.formula.api.ols` and `statsmodels.stats.oneway` respectively.

References

1. Chandrasekaran C, Peixoto D, Newsome WT, Shenoy KV (2017) Laminar differences in decision-related neural activity in dorsal premotor cortex. *Nature Communications* 8(1):614.
2. Finn ES, Huber L, Jangraw DC, Molfese PJ, Bandettini PA (2019) Layer-dependent activity in human prefrontal cortex during working memory. *Nature Neuroscience* 22(10):1687–1695.
3. Pinto L, Dan Y (2015) Cell-type-specific activity in prefrontal cortex during goal-directed behavior. *Neuron* 87(2):437–450.
4. Estebanez L, Hoffmann D, Voigt BC, Poulet JFA (2017) Parvalbumin-Expressing GABAergic Neurons in Primary Motor Cortex Signal Reaching. *Cell Reports* 20(2):308–318.
5. Lui JH, et al. (2020) Differential encoding in prefrontal cortex projection neuron classes across cognitive tasks. *Cell*.
6. Kvitsiani D, et al. (2013) Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature* 498(7454):363–366.
7. Gold JI, Shadlen MN (2007) The neural basis of decision making. *Annual review of neuroscience* 30.
8. Cisek P (2012) Making decisions through a distributed consensus. *Current opinion in neurobiology* 22(6):927–936.
9. Ding L, Gold JI (2012) Neural correlates of perceptual decision making before, during, and after decision commitment in monkey frontal eye field. *Cerebral Cortex* 22(5):1052–1067.
10. Thura D, Cisek P (2014) Deliberation and commitment in the premotor and primary motor cortex during dynamic decision making. *Neuron* 81(6):1401–1416.
11. Roitman JD, Shadlen MN (2002) Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of neuroscience* 22(21):9475–9489.
12. Hanks TD, et al. (2015) Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature* 520(7546):220–223.

13. Chandrasekaran C, Bray IE, Shenoy KV (2019) Frequency shifts and depth dependence of premotor beta band activity during perceptual decision-making. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 39:1420–1435.
14. Bastos AM, Loonis R, Kornblith S, Lundqvist M, Miller EK (2018) Laminar recordings in frontal cortex suggest distinct layers for maintenance and control of working memory. *Proceedings of the National Academy of Sciences* 115(5):1117–1122.
15. Mountcastle VB, Talbot WH, Sakata H, Hyvärinen J (1969) Cortical neuronal mechanisms in flutter-vibration studied in unanesthetized monkeys. neuronal periodicity and frequency discrimination. *Journal of neurophysiology* 32(3):452–484.
16. McCormick DA, Connors BW, Lighthall JW, Prince DA (1985) Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *Journal of Neurophysiology* 54(4):782–806.
17. Azouz R, Gray CM, Nowak LG, McCormick DA (1997) Physiological properties of inhibitory interneurons in cat striate cortex. *Cerebral Cortex* 7(6):534–545.
18. Hodge RD, et al. (2019) Conserved cell types with divergent features in human versus mouse cortex. *Nature* pp. 1–8.
19. Krienen FM, et al. (2020) Innovations present in the primate interneuron repertoire. *Nature* pp. 1–8.
20. Vigneswaran G, Kraskov A, Lemon RN (2011) Large Identified Pyramidal Cells in Macaque Motor and Premotor Cortex Exhibit "Thin Spikes": Implications for Cell Type Classification. *The Journal of Neuroscience* 31(40):14235–14242.
21. Soares D, et al. (2017) Expression of Kv3.1b potassium channel is widespread in macaque motor cortex pyramidal cells: A histological comparison between rat and macaque. *Journal of Comparative Neurology* 525(9):2164–2174.
22. Constantinople CM, Disney AA, Maffie J, Rudy B, Hawken MJ (2009) Quantitative analysis of neurons with Kv3 potassium channel subunits, Kv3.1b and Kv3.2, in macaque primary visual cortex. *Journal of Comparative Neurology* 516(4):291–311.
23. Amatrudo JM, et al. (2012) Influence of Highly Distinctive Structural Properties on the Excitability of Pyramidal Neurons in Monkey Visual and Prefrontal Cortices. *The Journal of Neuroscience* 32(40):13644–13660.
24. Onorato I, et al. (2019) A Distinct Class of Bursting Neurons with Strong Gamma Synchronization and Stimulus Selectivity in Monkey V1. *Neuron*.
25. Kelly JG, García-Marín V, Rudy B, Hawken MJ (2018) Densities and Laminar Distributions of Kv3.1b-, PV-, GABA-, and SMI-32-Immunoreactive Neurons in Macaque Area V1. *Cerebral Cortex* 29(5):1921–1937.
26. Ardid S, et al. (2015) Mapping of functionally characterized cell classes onto canonical circuit operations in primate prefrontal cortex. *The Journal of Neuroscience* 35(7):2975–2991.
27. Trainito C, Nicolai Cv, Miller EK, Siegel M (2019) Extracellular Spike Waveform Dissociates Four Functionally Distinct Cell Classes in Primate Cortex. *Current Biology*.
28. Boroujeni KB, Tiesinga P, Womelsdorf T (2020) Interneuron specific gamma synchronization encodes uncertain cues and prediction errors in lateral prefrontal and anterior cingulate cortex. *bioRxiv*.
29. Barry JM (2015) Axonal activity in vivo: technical considerations and implications for the exploration of neural circuits in freely moving animals. *Frontiers in Neuroscience* 9:153.
30. Robbins AA, Fox SE, Holmes GL, Scott RC, Barry JM (2013) Short duration waveforms

recorded extracellularly from freely moving rats are representative of axonal activity. *Frontiers in Neural Circuits* 7:181.

31. Krimer LS, et al. (2005) Cluster Analysis–Based Physiological Classification and Morphological Properties of Inhibitory Neurons in Layers 2–3 of Monkey Dorsolateral Prefrontal Cortex. *Journal of Neurophysiology* 94(5):3009–3022.
32. Kaufman MT, Churchland MM, Shenoy KV (2013) The roles of monkey M1 neuron classes in movement preparation and execution. *Journal of neurophysiology* 110(4):817–25.
33. Viskontas IV, Ekstrom AD, Wilson CL, Fried I (2007) Characterizing interneuron and pyramidal cells in the human medial temporal lobe in vivo using extracellular recordings. *Hippocampus* 17(1):49–57.
34. Katai S, et al. (2010) Classification of extracellularly recorded neurons by their discharge patterns and their correlates with intracellularly identified neuronal types in the frontal cortex of behaving monkeys. *European Journal of Neuroscience* 31(7):1322–1338.
35. Sun SH, et al. (2021) Analysis of extracellular spike waveforms and associated receptive fields of neurons in cat primary visual cortex. *The Journal of Physiology*.
36. Weir K, Blanquie O, Kilb W, Luhmann HJ, Sinning A (2015) Comparison of spike parameters from optically identified GABAergic and glutamatergic neurons in sparse cortical cultures. *Frontiers in Cellular Neuroscience* 8:460.
37. Tasic B, et al. (2018) Shared and distinct transcriptomic cell types across neocortical areas. *Nature* 563(7729):72–78.
38. Becht E, et al. (2018) Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology* 37(1):38–44.
39. Maheswaranathan N, Williams AH, Golub MD, Ganguli S, Sussillo D (2019) Universality and individuality in neural dynamics across large populations of recurrent networks.
40. Kleinman M, Chandrasekaran C, Kao JC (2019) Recurrent neural network models of multi-area computation underlying decision-making. *bioRxiv*.
41. Bala PC, et al. (2020) OpenMonkeyStudio: Automated Markerless Pose Estimation in Freely Moving Macaques. *bioRxiv* p. 2020.01.31.928861.
42. Hsu AI, Yttri EA (2020) B-soid: An open source unsupervised algorithm for discovery of spontaneous behaviors. *bioRxiv*.
43. Dolensek N, Gehrlach DA, Klein AS, Gogolla N (2020) Facial expressions of emotion states and their neuronal correlates in mice. *Science (New York, N.Y.)* 368(6486):89–94.
44. Jia X, et al. (2019) High-density extracellular probes reveal dendritic backpropagation and facilitate neuron classification. *Journal of Neurophysiology* 121(5):1831–1847.
45. Gouwens NW, et al. (2020) Integrated Morphoelectric and Transcriptomic Classification of Cortical GABAergic Cells. *Cell* 183(4):935–953.e19.
46. Klempíř O, et al. (2020) Application of spike sorting algorithm to neuronal signals originated from boron doped diamond micro-electrode arrays. *Physiological Research* 69(3).
47. Markanday A, et al. (2020) Using deep neural networks to detect complex spikes of cerebellar Purkinje cells. *Journal of Neurophysiology* 123(6):2217–2234.
48. McInnes L, Healy J, Melville J (2018) Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv*.
49. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008(10):P10008.
50. Snyder AC, Morais MJ, Smith MA (2016) Dynamics of excitatory and inhibitory networks are

differentially altered by selective attention. *Journal of Neurophysiology* 116(4):1807–1820.
51. Gold C, Girardin CC, Martin KAC, Koch C (2009) High-Amplitude Positive Spikes Recorded Extracellularly in Cat Visual Cortex. *Journal of Neurophysiology* 102(6):3340–3351.
52. Connors BW, Gutnick MJ, Prince DA (1982) Electrophysiological properties of neocortical neurons in vitro. *Journal of Neurophysiology* 48(6):1302–1320.
53. Kaufman MT, et al. (2010) Roles of monkey premotor neuron classes in movement preparation and execution. *Journal of neurophysiology* 104(2):799–810.
54. Mitchell JF, Sundberg KA, Reynolds JH (2007) Differential attention-dependent response modulation across cell classes in macaque visual area v4. *Neuron* 55(1):131–141.
55. Stuttgen M (2020) Mlib - toolbox for analyzing spike data.
56. Deligkaris K, Bullmann T, Frey U (2016) Extracellularly Recorded Somatic and Neuritic Signal Shapes and Classification Algorithms for High-Density Microelectrode Array Electrophysiology. *Frontiers in Neuroscience* 10:421.
57. Bakkum DJ, et al. (2013) Tracking axonal action potential propagation on a high-density microelectrode array across hundreds of sites. *Nature Communications* 4(1):2181.
58. Gur M, Beylin A, Snodderly DM (1999) Physiological Properties of Macaque V1 Neurons are Correlated With Extracellular Spike Amplitude, Duration, and Polarity. *Journal of Neurophysiology* 82(3):1451–1464.
59. Molnar C (2020) *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable.* (LULU COM).
60. Azodi CB, Tang J, Shiu SH (2020) Opening the Black Box: Interpretable Machine Learning for Geneticists. *Trends in Genetics* 36(6):442–455.
61. Lundberg S, Lee SI (2017) A Unified Approach to Interpreting Model Predictions. *arXiv*.
62. Lundberg SM, et al. (2020) From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence* 2(1):56–67.
63. Shapley LS (1988) *A value for n-person games*, ed. Roth AE. (Cambridge University Press), p. 31–40.
64. Štrumbelj E, Kononenko I (2014) Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems* 41(3):647–665.
65. Connors BW, Gutnick MJ (1990) Intrinsic firing patterns of diverse neocortical neurons. *Trends in Neurosciences* 13(3):99–104.
66. Nowak LG, Azouz R, Sanchez-Vives MV, Gray CM, McCormick DA (2003) Electrophysiological Classes of Cat Primary Visual Cortical Neurons In Vivo as Revealed by Quantitative Analyses. *Journal of Neurophysiology* 89(3):1541–1566.
67. Contreras D (2004) Electrophysiological classes of neocortical neurons. *Neural Networks* 17(5-6):633–646.
68. Steriade M (2004) Neocortical cell classes are flexible entities. *Nature Reviews Neuroscience* 5(2):121–134.
69. Steriade M, Timofeev I, Dürmüller N, Grenier F (1998) Dynamic Properties of Corticothalamic Neurons and Local Cortical Interneurons Generating Fast Rhythmic (30–40 Hz) Spike Bursts. *Journal of Neurophysiology* 79(1):483–490.
70. Johnston K, DeSouza JFX, Everling S (2009) Monkey Prefrontal Cortical Pyramidal and Putative Interneurons Exhibit Differential Patterns of Activity Between Prosaccade and Antisaccade Tasks. *Journal of Neuroscience* 29(17):5516–5524.
71. Hussar CR, Pasternak T (2009) Flexibility of Sensory Representations in Prefrontal Cortex

Depends on Cell Type. *Neuron* 64(5):730–743.
72. Quirk MC, Sosulski DL, Feierstein CE, Uchida N, Mainen ZF (2009) A defined network of fast-spiking interneurons in orbitofrontal cortex: responses to behavioral contingencies and ketamine administration. *Frontiers in Systems Neuroscience* 3:13.
73. Povysheva NV, Zaitsev AV, Gonzalez-Burgos G, Lewis DA (2013) Electrophysiological Heterogeneity of Fast-Spiking Interneurons: Chandelier versus Basket Cells. *PLoS ONE* 8(8):e70553.
74. Zaitsev AV, et al. (2009) Interneuron Diversity in Layers 2–3 of Monkey Prefrontal Cortex. *Cerebral Cortex* 19(7):1597–1615.
75. Song JH, McPeek RM (2010) Roles of Narrow- and Broad-Spiking Dorsal Premotor Area Neurons in Reach Target Selection and Movement Production. *Journal of Neurophysiology* 103(4):2124–2138.
76. Mruczek REB, Sheinberg DL (2012) Stimulus selectivity and response latency in putative inhibitory and excitatory neurons of the primate inferior temporal cortex. *Journal of Neurophysiology* 108(10):2725–2736.
77. Tosches MA, et al. (2018) Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. *Science* 360(6391):881–888.
78. Arikuni T, Watanabe K, Kubota K (1988) Connections of area 8 with area 6 in the brain of the macaque monkey. *Journal of Comparative Neurology* 277(1):21–40.
79. Nandy AS, Nassi JJ, Reynolds JH (2017) Laminar Organization of Attentional Modulation in Macaque Visual Area V4. *Neuron* 93(1):235–246.
80. Barbas H, Pandya DN (1987) Architecture and frontal cortical connections of the premotor cortex (area 6) in the rhesus monkey. *The Journal of Comparative Neurology* 256(2):211–228.
81. DeFelipe J (1997) Types of neurons, synaptic connections and chemical characteristics of cells immunoreactive for calbindin-D28K, parvalbumin and calretinin in the neocortex. *Journal of Chemical Neuroanatomy* 14(1):1–19.
82. Schmitz C, et al. (2014) Current automated 3D cell detection methods are not a suitable replacement for manual stereologic cell counting. *Frontiers in Neuroanatomy* 8(27):1–14.
83. Zaitsev A, et al. (2005) Localization of Calcium-binding Proteins in Physiologically and Morphologically Characterized Interneurons of Monkey Dorsolateral Prefrontal Cortex. *Cerebral Cortex* 15(8):1178–1186.
84. Ali M, Jones MW, Xie X, Williams M (2019) TimeCluster: dimension reduction applied to temporal data for visual analytics. *The Visual Computer* 35(6-8):1013–1026.
85. Diaz-Papkovich A, Anderson-Trocmé L, Ben-Eghan C, Gravel S (2019) UMAP reveals cryptic population structure and phenotype heterogeneity in large genomic cohorts. *PLOS Genetics* 15(11):e1008432.
86. Noichl M (2019) Modeling the structure of recent philosophy. *Synthese* pp. 1–12.
87. Moscovich A, Rosset S (2019) On the cross-validation bias due to unsupervised pre-processing. *arXiv*.
88. Sainburg T, McInnes L, Gentner TQ (2020) Parametric UMAP: learning embeddings with deep neural networks for representation and semi-supervised learning. *arXiv*.
89. Bellet A, Habrard A, Sebban M (2013) A Survey on Metric Learning for Feature Vectors and Structured Data. *arXiv*.
90. Yang L, Jin R (2006) Distance metric learning: A comprehensive survey.
91. Roux L, Stark E, Sjulson L, Buzsáki G (2014) In vivo optogenetic identification and manipu-

lation of GABAergic interneuron subtypes. *Current opinion in neurobiology* 26:88–95.
92. Deubner J, Coulon P, Diester I (2019) Optogenetic approaches to study the mammalian brain. *Current Opinion in Structural Biology* 57:157–163.
93. Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N (2012) Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* 482(7383):85–88.
94. Hangya B, Ranade SP, Lorenc M, Kepecs A (2015) Central Cholinergic Neurons Are Rapidly Recruited by Reinforcement Feedback. *Cell* 162(5):1155–1168.
95. Trautmann EM, et al. (2019) Accurate Estimation of Neural Population Dynamics without Spike Sorting. *Neuron*.
96. Jun JJ, et al. (2017) Fully integrated silicon probes for high-density recording of neural activity. *Nature* 551(7679):232–236.
97. Shenoy KV, Sahani M, Churchland MM (2013) Cortical control of arm movements: a dynamical systems perspective. *Annual review of neuroscience* 36.
98. Churchland MM, et al. (2012) Neural population dynamics during reaching. *Nature* 487(7405):51–56.
99. Mante V, Sussillo D, Shenoy KV, Newsome WT (2013) Context-dependent computation by recurrent dynamics in prefrontal cortex. *nature* 503(7474):78–84.
100. Remington ED, Narain D, Hosseini EA, Jazayeri M (2018) Flexible sensorimotor computations through rapid reconfiguration of cortical dynamics. *Neuron* 98(5):1005 – 1019.e5.
101. Wang J, Narain D, Hosseini EA, Jazayeri M (2017) Flexible timing by temporal scaling of cortical responses. *Nature Neuroscience* 21(1):102–110.
102. Georgopoulos A, Schwartz A, Kettner R (1986) Neuronal population coding of movement direction. *Science* 233(4771):1416–1419.
103. Hubel DH, Wiesel TN (1959) Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology* 148(3):574.
104. Bruce CJ, Goldberg ME (1985) Primate frontal eye fields. i. single neurons discharging before saccades. *Journal of neurophysiology* 53(3):603–635.
105. Cunningham JP, Byron MY (2014) Dimensionality reduction for large-scale neural recordings. *Nature neuroscience* 17(11):1500–1509.
106. Williams AH, et al. (2018) Unsupervised Discovery of Demixed, Low-Dimensional Neural Dynamics across Multiple Timescales through Tensor Component Analysis. *Neuron* 98(6):1099–1115.e8.
107. Kobak D, et al. (2016) Demixed principal component analysis of neural population data. *Elife* 5:e10989.
108. Pandarinath C, et al. (2018) Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature methods* 15(10):805–815.
109. Lindén H, et al. (2014) LFPy: a tool for biophysical simulation of extracellular potentials generated by detailed model neurons. *Frontiers in Neuroinformatics* 7.
110. Kaczmarek LK, Zhang Y (2017) Kv3 Channels: Enablers of Rapid Firing, Neurotransmitter Release, and Neuronal Endurance. *Physiological Reviews* 97(4):1431–1468.
111. Ding S, Matta SG, Zhou FM (2011) Kv3-Like Potassium Channels Are Required for Sustained High-Frequency Firing in Basal Ganglia Output Neurons. *Journal of Neurophysiology* 105(2):554–570.
112. Erisir A, Lau D, Rudy B, Leonard CS (1999) Function of Specific K+ Channels in Sustained High-Frequency Firing of Fast-Spiking Neocortical Interneurons. *Journal of Neurophysiology*

82(5):2476–2489.
113. Bean BP (2007) The action potential in mammalian central neurons. *Nature Reviews Neuroscience* 8(6):451–465.
114. Henze DA, et al. (2000) Intracellular Features Predicted by Extracellular Recordings in the Hippocampus In Vivo. *Journal of Neurophysiology* 84(1):390–400.
115. Torres-Gomez S, et al. (2020) Changes in the Proportion of Inhibitory Interneuron Types from Sensory to Executive Areas of the Primate Neocortex: Implications for the Origins of Working Memory Representations. *Cerebral cortex (New York, N.Y. : 1991)*.
116. Kelly JG, Hawken MJ (2020) GABAergic and non-GABAergic subpopulations of Kv3.1b-expressing neurons in macaque V2 and MT: laminar distributions and proportion of total neuronal population. *Brain Structure and Function* 225(3):1135–1152.
117. Ichinohe N, et al. (2004) A voltage-gated potassium channel, Kv3.1b, is expressed by a subpopulation of large pyramidal neurons in layer 5 of the macaque monkey cortex. *Neuroscience* 129(1):179–185.
118. Saleh MS, et al. (2019) Cmu array: A 3d nano-printed, customizable ultra-high-density microelectrode array platform. *bioRxiv*.
119. Mosher CP, et al. (2020) Cellular classes in the human brain revealed in-vivo by heartbeat-related modulation of the extracellular action potential waveform. *Cell Reports* 30(10):3536 – 3551.e6.
120. Courtin J, et al. (2014) Prefrontal parvalbumin interneurons shape neuronal activity to drive fear expression. *Nature* 505(7481):92–96.
121. Tremblay S, et al. (2020) An Open Resource for Non-human Primate Optogenetics. *Neuron*.
122. De A, El-Shamayleh Y, Horwitz GD (2020) Fast and reversible neural inactivation in macaque cortex by optogenetic stimulation of gabaergic neurons. *Elife* 9:e52658.
123. Mehta P, et al. (2019) Functional access to neuron subclasses in rodent and primate forebrain. *Cell Reports* 26(10):2818 – 2832.e8.
124. Vormstein-Schneider D, et al. (2020) Viral manipulation of functionally distinct interneurons in mice, non-human primates and humans. *Nature Neuroscience* pp. 1–8.
125. Kleiner M, et al. (2007) What's new in psychtoolbox-3. *Perception* 36(14):1–16.
126. Porter MA, Onnela JP, Mucha PJ (2009) Communities in Networks. *arXiv*.
127. Lambiotte R, Delvenne JC, Barahona M (2008) Laplacian dynamics and multiscale modular structure in networks. *arXiv preprint arXiv:0812.1770*.
128. Poulin V, Théberge F (2018) Ensemble Clustering for Graphs. *arXiv*.
129. Poulin V, Théberge F (2019) Ensemble clustering for graphs: comparisons and applications. *Applied Network Science* 4(1):51.
130. Csardi G, Nepusz T (2006) The igraph software package for complex network research. *InterJournal* Complex Systems:1695.
131. Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. arxiv 2016. *arXiv preprint arXiv:1603.02754*.
132. Kaufman S, Rosset S, Perlich C (2011) Leakage in data mining: formulation, detection, and avoidance in *In Proceedings of the 2011 conference on Knowledge Discovery in Data Mining*. pp. 556–563.
133. Meister MLR, Hennig JA, Huk AC (2013) Signal multiplexing and single-neuron computations in lateral intraparietal area during decision-making. *Journal of Neuroscience* 33(6):2254–2267.
134. Medalla M, Luebke JI (2015) Diversity of Glutamatergic Synaptic Strength in Lateral

Prefrontal versus Primary Visual Cortices in the Rhesus Monkey. *The Journal of Neuroscience* 35(1):112–127.

135. Rosene DL, Roy NJ, Davis BJ (1986) A cryoprotection method that facilitates cutting frozen sections of whole monkey brains for histological and histochemical processing without freezing artifact. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* 34(10):1301–1315.
136. Estrada LI, et al. (2016) Evaluation of Long-Term Cryostorage of Brain Tissue Sections for Quantitative Histochemistry. *Journal of Histochemistry & Cytochemistry* 65(3):153–171.
137. Morecraft R, Cipolloni P, Stilwell [U+2010] Morecraft K, Gedney M, Pandya D (2004) Cytoarchitecture and cortical connections of the posterior cingulate and adjacent somatosensory fields in the rhesus monkey. *Journal of Comparative Neurology* 469(1):37–69.
138. Morecraft RJ, et al. (2019) Terminal organization of the corticospinal projection from the lateral premotor cortex to the cervical enlargement (C5–T1) in rhesus monkey. *Journal of Comparative Neurology* 527(16):2761–2789.
139. Matelli M, Luppino G (1996) Thalamic input to mesial and superior area 6 in the macaque monkey. *Journal of Comparative Neurology* 372(1):59–87.
140. Rizzolatti G, Luppino G, Matelli M (1998) The organization of the cortical motor system: new concepts. *Electroencephalography and Clinical Neurophysiology* 106(4):283–296.
141. Shindelin J, et al. (2012) Fiji: An Open-source Platform for Biological-image Analysis. *Nature Methods* 9(7):676—-682.
142. Virtanen P, et al. (2020) SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* 17:261–272.
143. Terpilowski M (2019) scikit-posthocs: Pairwise multiple comparison tests in python. *The Journal of Open Source Software* 4(36):1169.
144. Seabold S, Perktold J (2010) statsmodels: Econometric and statistical modeling with python in *9th Python in Science Conference*.
145. Verleysen M, Francois D, Simon G, Wertz V (2003) Artificial Neural Nets Problem Solving Methods, 7th International Work-Conference on Artificial and Natural Neural Networks, IWANN2003 Maó, Menorca, Spain, June 3–6, 2003 Proceedings, Part II. pp. 105–112.
146. Beyer KS, Goldstein J, Ramakrishnan R, Shaft U (1999) When is "nearest neighbor" meaningful? in *Proceedings of the 7th International Conference on Database Theory*, ICDT '99. (Springer-Verlag, Berlin, Heidelberg), p. 217–235.
147. McInnes L (2019) Umap uniform manifold approximation and projection for dimension reduction.
148. Tenenbaum JB, Silva Vd, Langford JC (2000) A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science* 290(5500):2319–2323.
149. Maaten Lvd, Hinton G (2008) Visualizing data using t-sne. *Journal of machine learning research* 9(Nov):2579–2605.
150. Lee JA, Verleysen M, eds. (2007) *Nonlinear Dimensionality Reduction*. (Springer New York).
151. Spivak DI (2009) Metric realization of fuzzy simplicial sets.
152. Dong W, Moses C, Li K (2011) Efficient k-nearest neighbor graph construction for generic similarity measures in *Proceedings of the 20th international conference on World wide web*. pp. 577–586.
153. Belkin M, Niyogi P (2002) Laplacian eigenmaps and spectral techniques for embedding and clustering in *Advances in Neural Information Processing Systems 14*, eds. Dietterich TG,

Becker S, Ghahramani Z. (MIT Press), pp. 585–591.

154. Bengio Y, Vincent P, Paiement J (2003) Spectral clustering and kernel pca are learning eigenfunctions.
155. Kobak D, Linderman GC (2019) UMAP does not preserve global structure any better than t-SNE when using the same initialization. *bioRxiv* p. 2019.12.19.877522.
156. Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. *Phys. Rev. E* 69(2):026113.
157. Lambiotte R (2007) Finding communities at different resolutions in large networks.

Supplementary Information

Clustering in high-dimensionality and the curse of dimensionality

Clustering in high-dimensions is a difficult problem. In particular, "concentration of measure" results in Euclidean distances, used by k-Means clustering, becoming meaningless as a measure of distance. Specifically, as dimensionality increases, the difference between distances of randomly chosen points all converge to the same constant distance[145,146]. UMAP counters this by assuming "connectedness" between each point and its nearest neighbors. This results in the local unit distance metric around each data point stretching or contracting[147] (see Supplementary Information). In addition, traditionally used clustering methods like a Gaussian mixture model (GMM) or k-Means do not consistently capture these clusters because they assume normality in cluster shape. This is violated by the perturbations introduced by the non-linearity of the UMAP projection even if distributions are Gaussian in the original high-dimensional space (which they are also likely not either).

UMAP dimensionality reduction

UMAP is among the class of non-linear dimensionality reductions known as manifold learning algorithms which also includes other well-known methods in neuroscience such as Isomap[148] and t-SNE[149] (see[150] for a review of methods). Key to this algorithm is the presumption that although data may not be uniformly spaced in the ambient space, it is uniform on some low-dimensional manifold. It is also assumed that the underlying manifold is locally connected (i.e. doesn't have any breaks or isolated points). This leads to the conclusion that the underlying notion of distance (Riemannian metric) changes in each region of the manifold: the notion of a unit distance "stretches" in areas of sparser density and "shortens" in areas of higher density. This is formalized beginning with defining how a local Riemannian metric g should be constructed by **Lemma 1** of[48]:

> **Lemma 1**: Let $\mathcal{M}$ be a Riemannian manifold equipped with a metric g in ambient space $\mathbb{R}^n$. Let $p \in \mathcal{M}$ be a point in this space. If g is locally constant about the point p in an open neighborhood U such that g is a constant diagonal matrix in diagonal coordinates, then in a ball $B \subseteq U$ centered at p with a volume $\frac{\pi^{n/2}}{\Gamma(n/2+1)}$ with respect to g, the geodesic distance from p to any point $q \in B$ is $\frac{1}{r} d_{\mathbb{R}^n}(p, q)$, where r is the radius of the ball in the ambient space and $d_{\mathbb{R}^n}$ is the existing metric on the ambient space.

Using this definition of g, each ball B of fixed volume should contain the same number of data points in X regardless of where on the manifold B is located. This also implies that a ball centered on data point x_i should contain the k-nearest neighbors of x_i in a fixed volume no matter which $x_i \in X$ is chosen. Thus the geodesic distance around each data point is normalized by its distance to its k^{th} nearest neighbor.

To compensate for the impact of certain nearest neighbors in the ball lying much further than those closer by, the normalizing distances are transformed by the exponential function,

$$\sum_{j=1}^{k} \exp\left(\frac{-|x_i - x_{i_j}|}{r_i}\right) = \log_2(k)$$

To unite these disparate metric spaces (each data point has a possibly unique local notion of distance), category theory is used to convert this representation to that of a fuzzy simplicial set via an adjunction that will not be defined here (see **Section 3** of[48] and the Nerve Theorem). In this way, the topological structure of the data can be equivalently represented as a metric space or a set of fuzzy simplices. One large benefit of this construction is that while normalized distances in high-dimensional spaces suffer from the Curse of Dimensionality in the form of concentration of measure, normalized nearest neighbor distances do not[147]. The end result of this process is an approximation of the topology of the manifold by fuzzy simplicial sets in the form of a Čech complex.

With this fuzzy topological representation, the low-dimensional representation can be found through an optimization procedure that minimizes the cross-entropy of fuzzy simplicial sets containing the same objects. Given two fuzzy sets with the same members A and separate membership strength functions μ and ν (of Spivak's characteristic form[151]), the cross-entropy $C((A,\mu),(A,\nu))$ is defined as,

$$C((A,\mu),(A,\nu)) \triangleq \sum_{a\in A}\left(\mu(a)\log\frac{\mu(a)}{\nu(a)} + \big(1-\mu(a)\big)\log\frac{1-\mu(a)}{1-\nu(a)}\right).$$

The first term $\mu(a)\log\frac{\mu(a)}{\nu(a)}$ captures the attractive force minimised if short edges in high-dimension correspond to short edges in low-dimension and $(1-\mu(a))\log\frac{1-\mu(a)}{1-\nu(a)}$ is the repulsive forces that are minimised if long edges in high-dimension correspond to long edges in low-dimension.

From a computational perspective, this whole UMAP process proceeds in two steps: construction of a k-nearest neighbor graph and layout of the graph into a low-dimensional manifold. Note that after the first step, the k-nearest neighbor graph is passed to Louvain community detection and thus the clustering is not dependent on the embedding of the graph, just on its construction and associated UMAP parameters such as `n_neighbors` and `metric` but not `min_dist`. The embedding *is* however used for visualization and interpretability.

Graph construction: Given a set of data points $X = \{x_1,\ldots,x_N\}$ and a metric d, the construction of an undirected weighted k-nearest neighbor graph (captured by an adjacency matrix capturing the connection weights between nodes) is conducted using a nearest neighbor descent algorithm[152]. For each data point $x_i \in X$ and fixed nearest neighbor hyperparameter k, we have the set $\{x_{i_1}\ldots x_{i_k}\}$ the set of k-nearest neighbors of x_i under the local Riemannian metric d. We define ρ_i and σ_i such that,

$$\rho_i = \min\{d(x_i,x_{i_j})|1\le j\le k, d(x_i,x_{i_j})>0\}$$

and setting σ_i such that,

$$\log_2(k) = \sum_{j=1}^{k}\exp\left(\frac{-\max\big(0,d(x_i,x_{i_j})-\rho_i\big)}{\sigma_i}\right).$$

The weighted graph $\overline{G} = (V,E,w)$ is defined in terms of the vertices V, edges

$E = \{(x_i, x_{i_j}) | 1 \leq j \leq k, 1 \leq i \leq N\}$, and weight function w as,

$$w((x_i, x_{i_j})) = \sum_{j=1}^{k} \exp\left(\frac{-\max\left(0, d(x_i, x_{i_j}) - \rho_i\right)}{\sigma_i}\right)$$

. If A is the weighted adjacency matrix of $\overline{G}$, we can get the undirected weighted graph B by the relationship,

$$B = A + A^{\intercal} - A \circ A^{\intercal}$$

where $\circ$ is the Hadamard product.

Graph layout: The UMAP algorithm finds a low-dimensional projection (manifold) of the high-dimensional data by a force directed layout of the constructed graph. Before this is done though, the graph is spectrally embedded to aid in consistency and convergence of the algorithm through initialization[153–155]. The symmetric normalized Laplacian L of the graph is calculated for the 1-skeleton of the weighted graph which is analogous to the Laplace-Beltrami operator (divergence of the gradient, $\Delta f = \nabla^2 f$) on a manifold. If D is the degree matrix (a diagonal matrix containing the degree of each vertex) of the adjacency matrix A, we compute the Laplacian matrix as,

$$L = D^{\frac{1}{2}}(D - A)D^{\frac{1}{2}}$$

with associated eigenvectors y and eigenvalues λ,

$$Ly = \lambda Dy.$$

After the spectral embedding of the graph with Laplacian eigenmaps, the force directed graph layout iteratively applies attractive and repulsive forces on the edges and vertices. The attractive force between two vertices i and j at coordinates $\mathbf{y}_i$ and $\mathbf{y}_j$ with tunable hyperparameters a and b and is determined by,

$$\frac{-2ab\left\|\mathbf{y}_i - \mathbf{y}_j\right\|_2^{2(b-1)}}{1 + \left\|\mathbf{y}_i - \mathbf{y}_j\right\|_2^2} w\Big((x_i, x_j)\Big)(\mathbf{y}_i - \mathbf{y}_j)$$

and the repulsive forces with hyper-parameter $\epsilon = 0.001$ to prevent division by zero,

$$\frac{b}{\left(\epsilon + \left\|\mathbf{y}_i - \mathbf{y}_j\right\|_2^2\right)\left(1 + \left\|\mathbf{y}_i - \mathbf{y}_j\right\|_2^2\right)}\Big(1 - w\Big((x_i, x_j)\Big)\Big)(\mathbf{y}_i - \mathbf{y}_j).$$

This optimization procedure is then completed using stochastic gradient descent to arrive at the final embedding.

Louvain method for community detection

The Louvain method for community detection (here called clustering)[49] operates on a weighted network and locates highly-interconnected nodes called a community. This "connectedness" is measured by their modularity Q with added resolution parameter t[156,157] (taking real values between -1 and 1 inclusive) defined as,

$$Q_t = (1-t) + \frac{1}{2m}\sum_{i,j}\left[tA_{i,j} - \frac{k_i k_j}{2m}\right]\delta(c_i, c_j)$$

where t is a parameter controlling the "characteristic scale" of the communities[127]. The larger the resolution parameter, the fewer the number of communities and the larger their size. Smaller values of resolution parameter results in more communities smaller in size. Note that when $t = 1$, the simplified definition of modularity is given as appears in[49]. $A_{i,j}$ is an adjacency matrix with the weights of the edges between the nodes indexed by i and j. $k_i = \sum_j A_{i,j}$ is the sum of weights of the edges connected to the node i. c_i is the community to which the node i belongs to and the function $\delta(c_i, c_j)$ is an indicator function that is 1 if $c_i = c_j$ and 0 otherwise. The value $m = \frac{1}{2}\sum_{i,j} A_{i,j}$ which is the sum of all the weights of all the edges in the network. This equation also serves as an objective function for the iterative procedure in Blondel *et al* 2008[49] which proceeds in two steps,

Modularity Optimization: Each node is assigned to its own community (in the initial step) and then each node is moved into a community with a random neighbor and the change in modularity is examined. The equation for this change in modularity is,

$$\Delta Q_t = \left[\frac{\sum_{\text{in}} + k_{i,\text{in}}}{2m} - \left(\frac{\sum_{\text{tot}} + {k_i}^2}{2m}\right)\right] - \left[\frac{\sum_{\text{in}}}{2m} - \left(\frac{\sum_{\text{tot}}}{2m}\right) - \left(\frac{k_i}{2m}\right)^2\right].$$

where $\sum_{\text{in}}$ is the sum of all the weights inside the community that the node i is moving into. $\sum_{\text{tot}}$ is the sum of all the weights of the edges to nodes in the community i is moving into. k_i is the sum of all the weighted edges incident on i. $k_{i,\text{in}}$ is the sum of the weights from the edges of i to nodes in the cluster.

Once the change in modularity is examined for a node before and after joining a neighboring cluster, the neighbor joins (or stays with) the community with the largest positive increase in modularity; if no increase can be found, the node remains a part of its current community. Once there can be found no increase in modularity for any points, the algorithm proceeds to the second step.

Community Aggregation: Every node in each community in the previous step is then collapsed into a single node and their edges combined to form a new graph. The process is then repeated from the previous step.

The graph produced by UMAP was passed into this Louvain clustering algorithm[49] using `cylouvain 0.2.3` with parameters in Table S1. This clustering method requires no prior specification of the number of clusters that should be present but its number does depend on the resolution parameter. To choose this parameter, a sensitivity analysis was conducted across various values of

the resolution parameter with the number of communities and total modularity compared Fig. 2B. Each waveform was then plotted in UMAP space and color-coded to its associated cluster label found by Louvain Clustering Fig. 2A.

Ensemble clustering for graphs (ECG)

ECG is a consensus clustering method for graphs and was used to validate the Louvain clustering algorithm. ECG consists of two steps: generation and integration.

Generation: This step instantiates the ECG algorithm by using Louvain clustering to produce a randomized set of k level-1 partitions $\mathcal{P} \in \{\mathcal{P}_\infty, \ldots, \mathcal{P}_\|\}$. The randomization comes from the randomization of vertices in the initial step of Louvain clustering.

Integration: Once the k randomized level-1 Louvain partitions are obtained, Louvain is run on a weighted version of the initial graph $G = (V, E)$. These weights $W_\mathcal{P}(u, v)$ are obtained via co-association of edges $e = (u, v) \in E$. These weights are defined as,

$$W_\mathcal{P}(u,v) \triangleq \begin{cases} w_* + (1 - w_*) \cdot \left(\frac{\sum_{i=1}^{k} v_{\mathcal{P}_i}(u,v)}{k} \right) & \text{, if } (u,v) \text{ is in the 2-core of } G \\ w_* & \text{, otherwise} \end{cases}$$

where we have the minimum ECG weight $0 < w_* < 1$ and the co-occurence of edges u and v as $v_{P_i} = \sum_{j=1}^{l_i} \mathbb{1}_{C_i^j}(u) \cdot \mathbb{1}_{C_i^j}(v)$ where $\mathbb{1}_{C_i^j}(u)$ is an indicator function of if the edge u occurs in the cluster of P_i or not.

With this function, ECG combines these level-1 Louvain partitions as a single weighted graph which serves as the result.

SHapley Additive exPlanations (SHAP)

SHAP values build off of Shapley values[63] and provides interpretability to machine learning models by computing the contributions of each feature towards the overall model. These explanations of machine learning models are models in and of themselves and are referred to as "additive feature attribution methods". These explanation models use simplified inputs x' which are mapped to original inputs through a function $x = h_x(x')$ and try to ensure that $g(z') \approx f(h_x(z'))$ whenever $z' \approx x'$ where f is the machine learning model and g is the explanation model. This yields the additive form which is a linear combination of binary variables:

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z_i'$$

where $z' \in \{0, 1\}^M$ is the binary value specifying the inclusion or exclusion of a number of simplified feature inputs M; $\phi_i \in \mathbb{R}$ is the effect of each feature.

Work in[61] devises such a model satisfying three important properties within this framework:

Local accuracy/efficiency: the explanation's features with their effects ϕ_i must sum for each feature i to the output $f(x)$.

$$f(x) = \phi_0(f) + \sum_{i=1}^{M} \phi_i(f, x)$$

where $\phi_0(f) = \mathbb{E}\Big[f(z)\Big] = f_x(\varnothing)$

Consistency/Monotonicity: If a model changes so that the effect of a feature increases of stays the same regardless of other inputs, that input's attribution should not decrease. For any two models f and f' if,

$$f'_x(S) - f'_x(S\backslash i) \geq f_x(S) - f_x(S\backslash i)$$

where $S \in \mathcal{F}$ are subsets of all features and $S\backslash i$ is the setting of feature i to zero (or some background reference value intended to be negligible) then $\phi_i(f', x) \geq \phi_i(f, x)$.

Missingness: This is the idea that features with no effect on f_x should have no assigned impact ϕ_i. This is expressed as,

$$f_x(S \cup i) = f_x(S)$$

for all subsets of features $S \in \mathcal{F}$, then $\phi_i(f, x) = 0$.

The authors prove that the only possible additive feature attribution method that satisfies these three criteria is SHAP whose values are computed as the following,

$$\phi_i(f, x) = \sum_{R \in \mathbb{R}} \frac{1}{M!} \Big[f_x(P_i^R \cup i) - f_x(P_i^R) \Big]$$

where R is the set of all feature orderings and P_i^R is the set of all features that come before the i^{th} one in ordering R and M is the number of input features.

Extending SHAP values to tree classifiers, the authors create `shap.TreeExplainer`[62] to calculate SHAP values by using path-dependent feature perturbation to yield the plots in Fig. 4B and C.

WaveMAP is stable over hyperparameter choice and random seed

Hyperparameter Choice: Louvain clustering[49] requires the specification of a resolution hyperparameter, t, which controls the "characteristic scale" by which network "communities" are identified; the larger this parameter, the fewer the number of clusters (communities) detected and vice versa[157]. We selected a resolution parameter that balanced a maximal modularity score with a number of clusters selected at the "elbow" of the curve of number of clusters vs. resolution parameter. Modularity (the "connectedness" of a community, see Methods) is a summed pairwise measure between nodes in a cluster. It is defined as the difference between the weights of the edges within a cluster and the edges incoming from any other node outside of the cluster. Maximizing this value over the whole graph finds communities (sub-graphs) with high amounts of inter-connectivity. We chose t to be 1.5 which resulted in an average modularity score of 0.761 ± 0.004 (mean $\pm$ S.D.) and an average of 8.29 ± 0.84 (mean $\pm$ S.D.) clusters across 25

random UMAP seeds (Fig. 2B). In this manner, we found a set of waveform clusters the balanced the diversity found by UMAP and statistical interpretability.

Robustness of solutions: To show this hierarchy of clustering resolutions along the curve in Fig. 2C and to demonstrate *WaveMAP*'s robustness to random seed initialization, we plotted three different plots for several different resolution parameters in Fig. S2A. Each random seed produced the same clustering with only slight perturbations of scale and rotation. To validate these clusters were reliable and not an artifact of our particular data sample, we counted the number of clusters from 100 randomly permuted subsets of the full dataset at varying proportions (from 10% to 100% in 10% increments) and also set each with a different UMAP random seed (Fig. S2B). As the data portion increased, we found that the number of clusters increased then tapered off to around eight clusters at ∼60% of the full dataset. At 100% of the dataset, we found 8.5 ± 0.61 (mean ± S.D.) clusters with residual variability induced by the UMAP random seed. This analysis is reassuring and demonstrates that we have an adequate number of waveforms to describe the diversity of cell types in monkey PMd.

Stability: Louvain clustering is sometimes unstable. That is, results from successive runs on the same data can show considerable variation on some datasets[129]. To test whether these eight clusters consistently contained the same constituent data points run-to-run, we used ensemble clustering for graphs (ECG)[128,129]. ECG generates k randomized level-1 (one round of Louvain clustering, Fig. S1B) partitions and combines together their graph structure via the co-occurrence of edges between nodes across partitionings. Hence the "ensemble" in the name (also called "consensus clustering"). Performing ECG with k=10, 100 times on UMAP graphs with different random seeds produced an average of 8.87 ± 0.74 (mean ± S.D.) clusters which was similar to that found by Louvain clustering with resolution parameter set to 1.5. In addition, the runs of ECG that yielded eight clusters had an almost exact structure to that produced by *WaveMAP* (compare Fig. S2C to Fig. 2A). The runs of ECG with more than eight clusters contained small clusters ($n \leq 20$) splitting off from ⑤ which were too small to allow us to make rigorous conclusions statistically. We therefore chose the more conservative eight cluster solution that balanced maximizing cluster number while ensuring adequate cluster sizes.

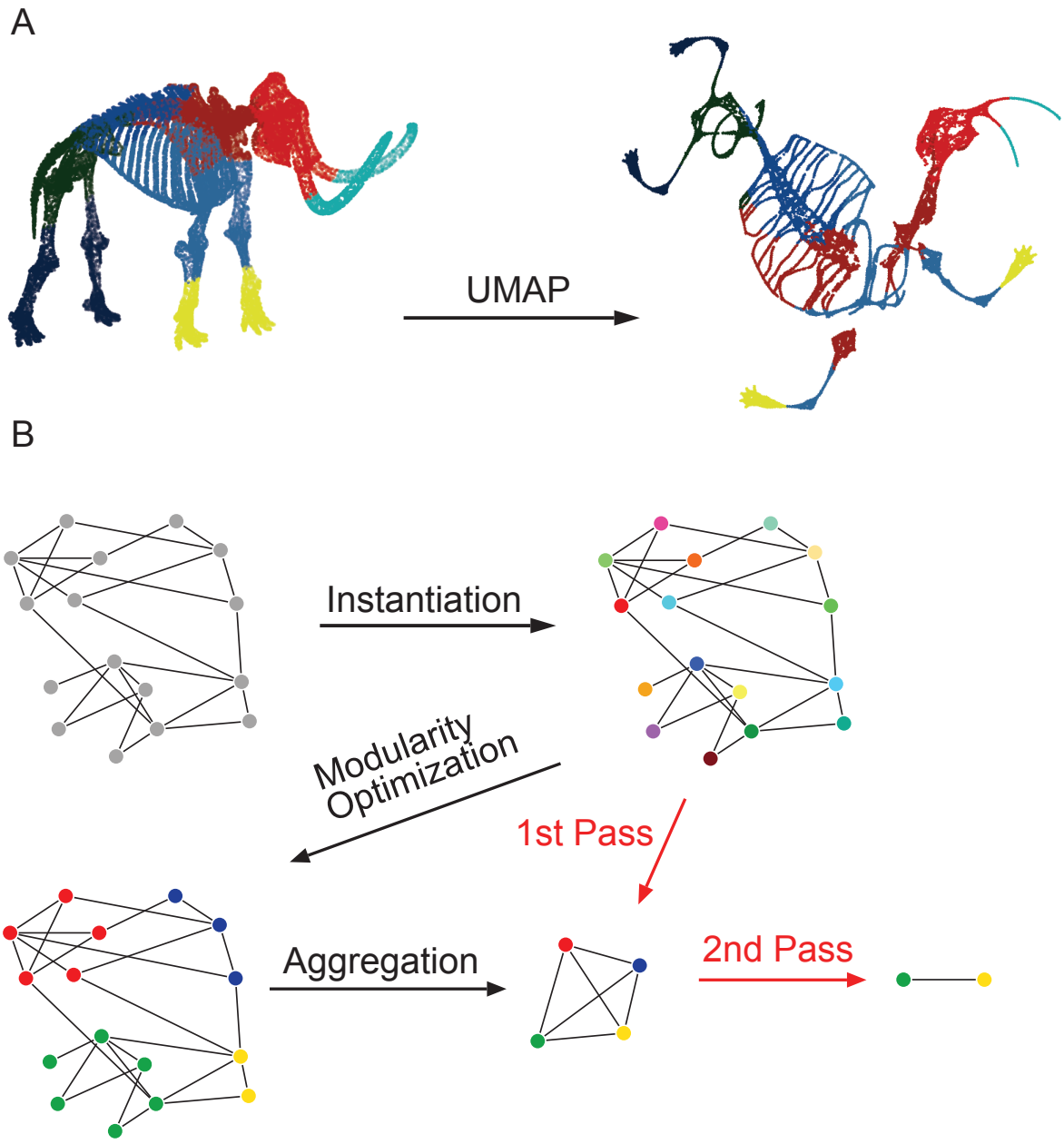

Figure S1: Diagrams of UMAP and Louvain Community Detection: **(A)** A demonstration of UMAP projection on a 3D point cloud skeleton of a wooly mammoth. Local and global structures are incorporated as evident in the preservation of the individual bone shapes and the sensible spatial relationships between the body parts. Idea from Maximilian Noichl (https://github.com/MNoichl/UMAP-examples-mammoth-) and mammoth skeleton from the Smithsonian Institute's Smithsonian 3D (https://3d.si.edu/). **(B)** The Louvain community detection algorithm is applied to weighted undirected graphs and proceeds in three steps which are said to be one "pass" of the algorithm: (1) each node is assigned to its own cluster; (2) each node is randomly moved into a neighboring cluster and if modularity increases, it becomes a member of that cluster; (3) once modularity no longer increases, each cluster is collapsed into one node. This process repeats for multiple passes until a stopping condition is reached.

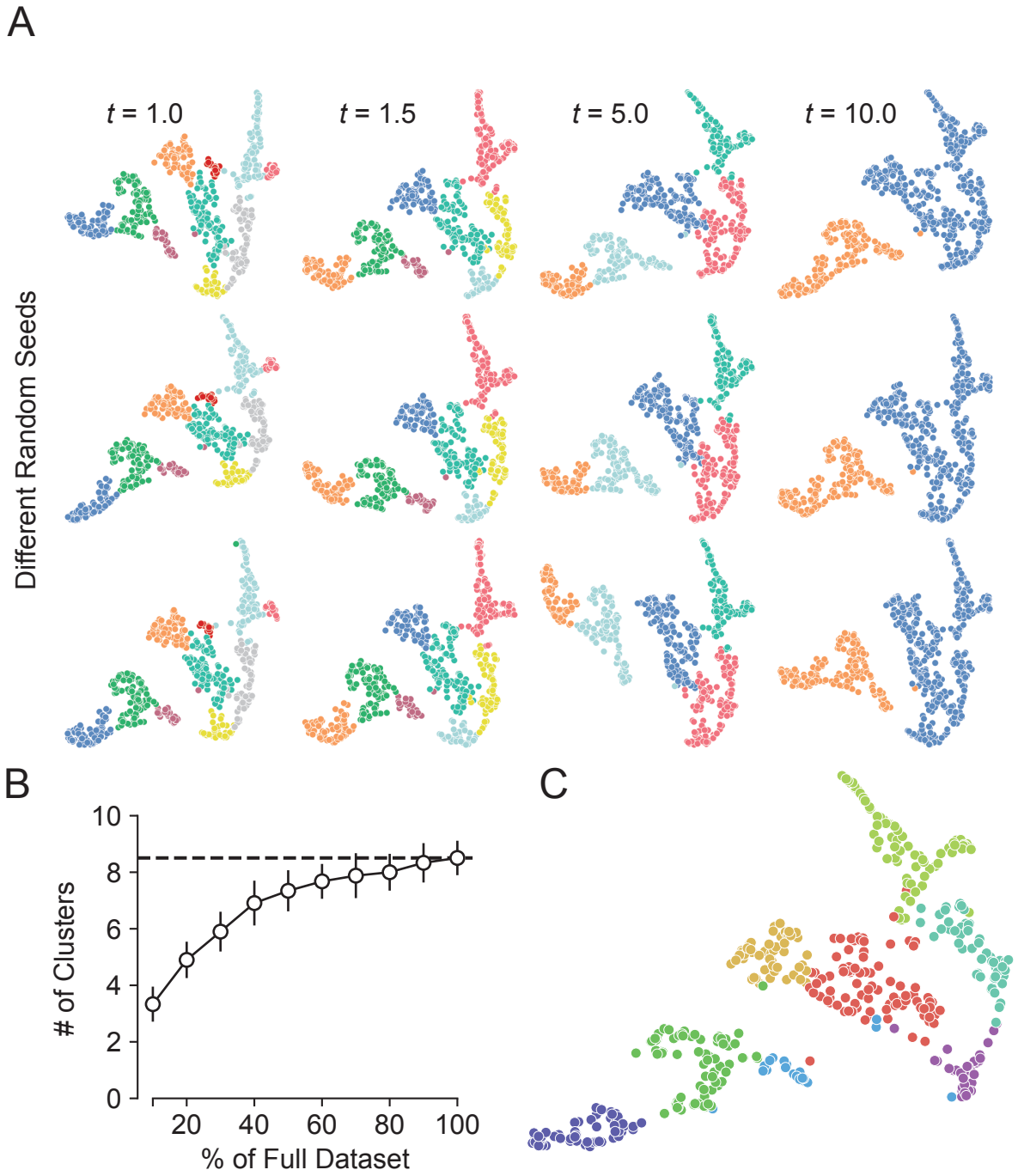

Figure S2: Stability analysis of *WaveMAP*: **(A)** *WaveMAP* instantiated with several different UMAP random seeds and also Louvain resolution parameters. **(B)** The mean ± S.D. number of clusters produced by *WaveMAP* across various proportions of the full dataset. **(C)** An eight cluster result of ECG applied to the UMAP graph.

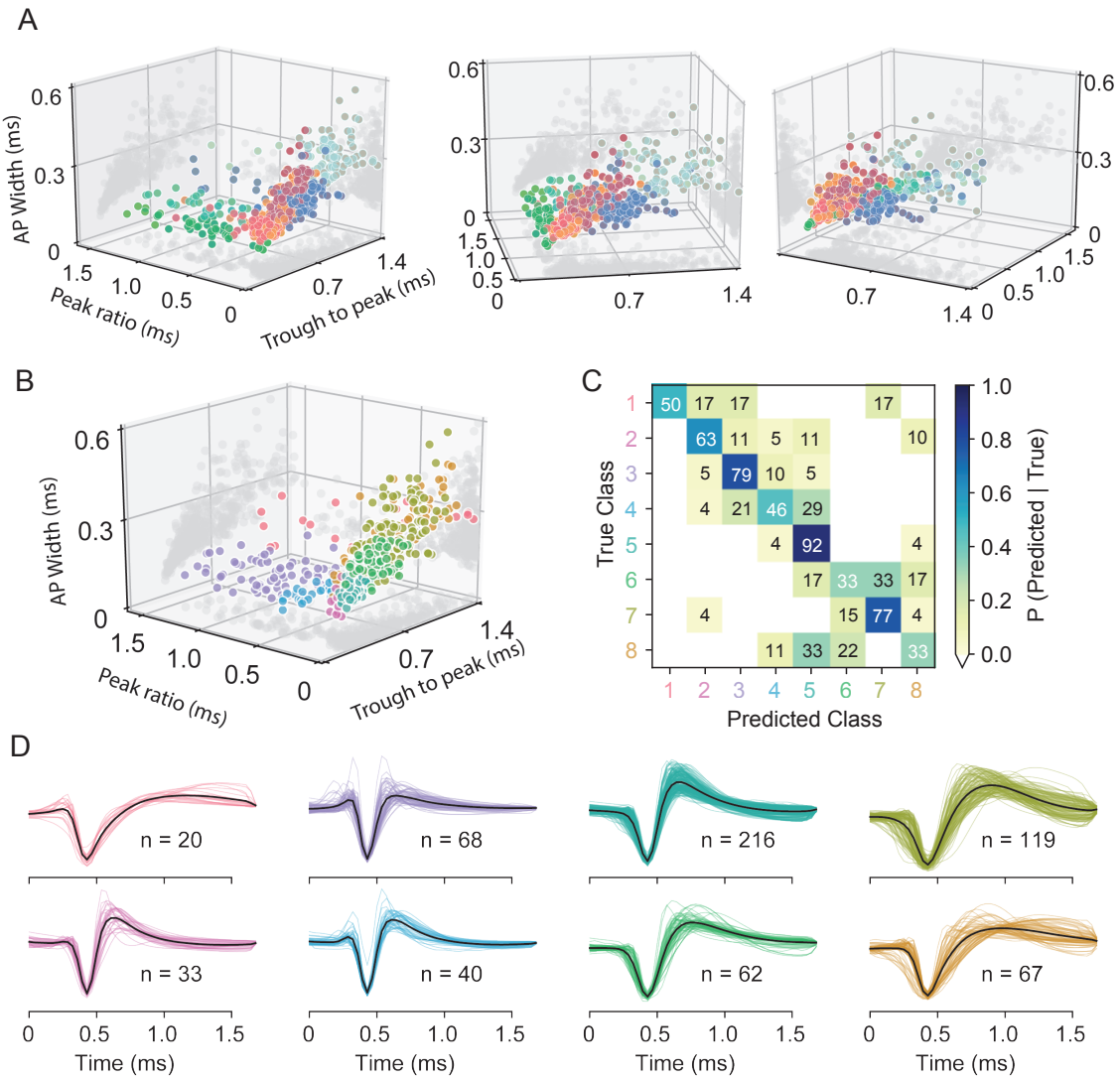

Figure S3: Comparison of GMM and UMAP in the constructed feature space: **(A)** Three views of the eight *WaveMAP* clusters shown in the constructed feature space. The clusters maintain some structure but are largely mixed and linearly inseparable. **(B)** A GMM instantiated with eight clusters in the constructed feature space. **(C)** Confusion matrix for a random forest classifier with the same hyperparameters as the one trained on four GMM classes S1. Numbers listed are in percent accuracy on the main diagonal and misclassification rate percentage on the off-diagonals against held-out data. **(D)** Each cluster of waveforms in the eight class GMM.

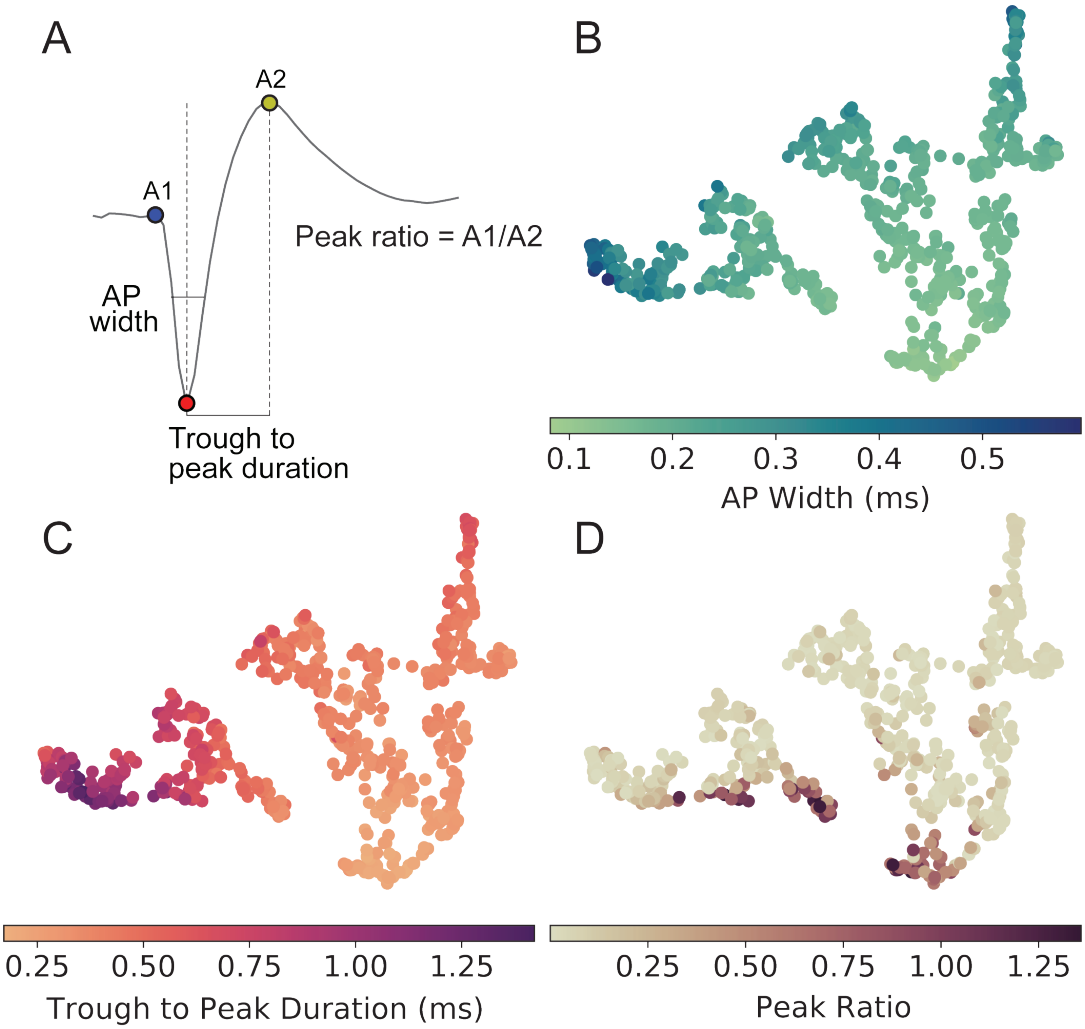

Figure S4: *WaveMAP* implicitly captures waveform features without the need for specification: **(A)** Three waveform shape features used in traditional clustering approaches. The three EAP waveform landmarks used to generate the specified features passed to the GMM on a sample waveform. ● is the pre-hyperpolarization peak (A1); ● is the depolarization trough; and ● is the post-hyperpolarization peak (A2). AP width is the distance in time between the falling and rising phase of the depolarization trough at its full-width half minimum. The trough to peak duration is the distance between the minimum of the depolarization trough and the peak of the post-hyperpolarization peak. The peak ratio is the height (above zero) of the pre-hyperpolarization peak over the height (again, above zero) of the post-hyperpolarization peak. The same diagram as in Fig. 3A but repeated here. **(B, C, D)** The waveform data points in the projected UMAP space and color coded according to their AP width, trough to peak duration, and peak ratio respectively.

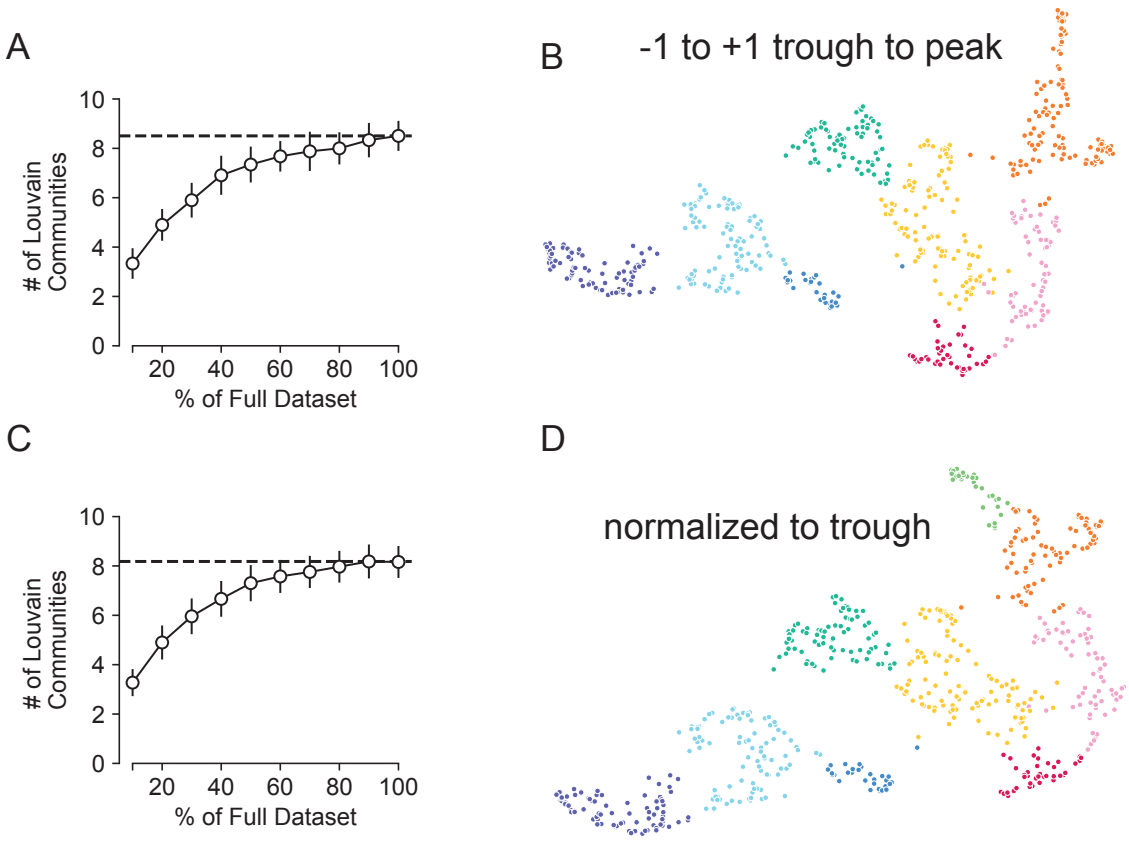

Figure S5: Different normalizations have little effect on *WaveMAP* structure: **(A)** As in Fig. S2B, the number of Louvain communities found across various random subsets and random seeds. The mean number of clusters shown on the full dataset with a dashed line. **(B)** the *WaveMAP* clusters on waveforms with ± 1 unit trough to peak normalization (used in the paper). **(C)** The same random subsetting and random seed strategy in (C) applied to waveform data normalized to trough depth. **(D)** *WaveMAP* clusters applied to waveform data normalized to trough depth.

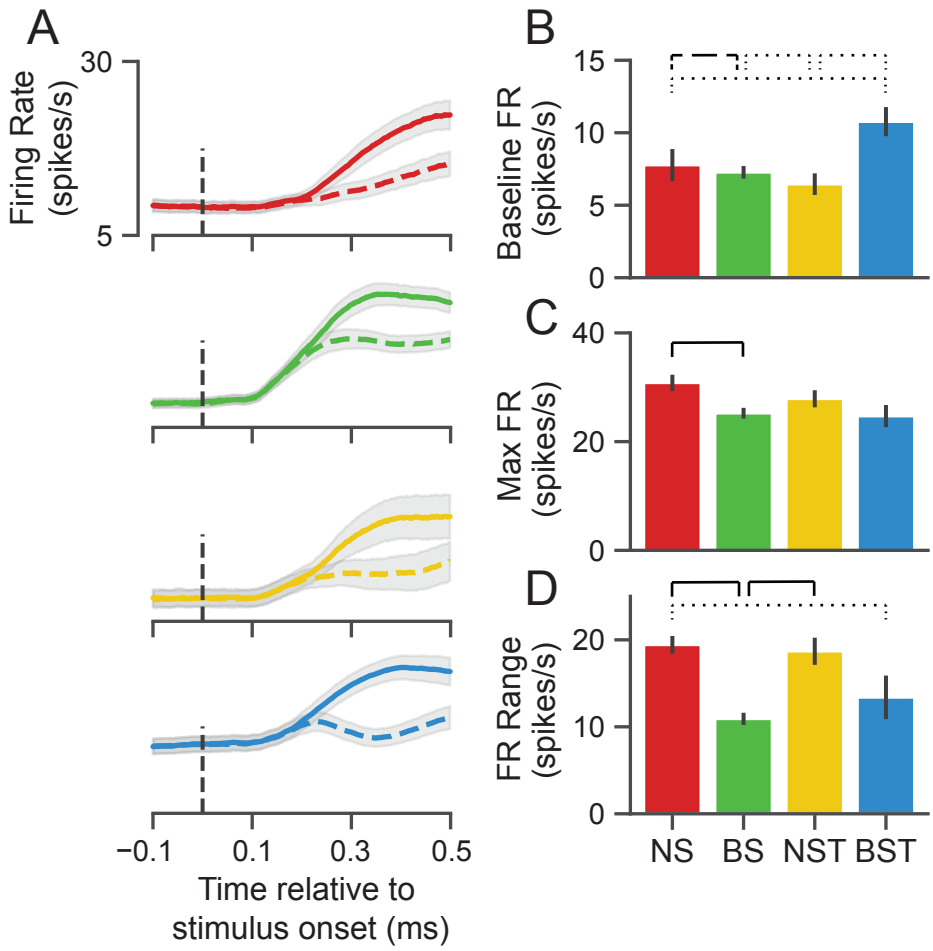

Figure S6: GMM clusters are less physiologically distinguishable than *WaveMAP* clusters: **(A)** Stimulus-aligned trial-averaged firing rate activity in PMd for GMM clusters. As in Fig. 5, the traces are separated into `PREF` and `NONPREF` trials with solid and dashed lines respectively. Shaded regions correspond to bootstrapped standard error of the mean. **(B)** Baseline median firing rates for the four GMM clusters. Baselines were calculated as the average firing rate during the first 200 ms of the trial. ····· $p < 0.05$; - - - - $p < 0.01$; —— $p < 0.005$; Mann-Whitney *U* test, FDR adjusted.

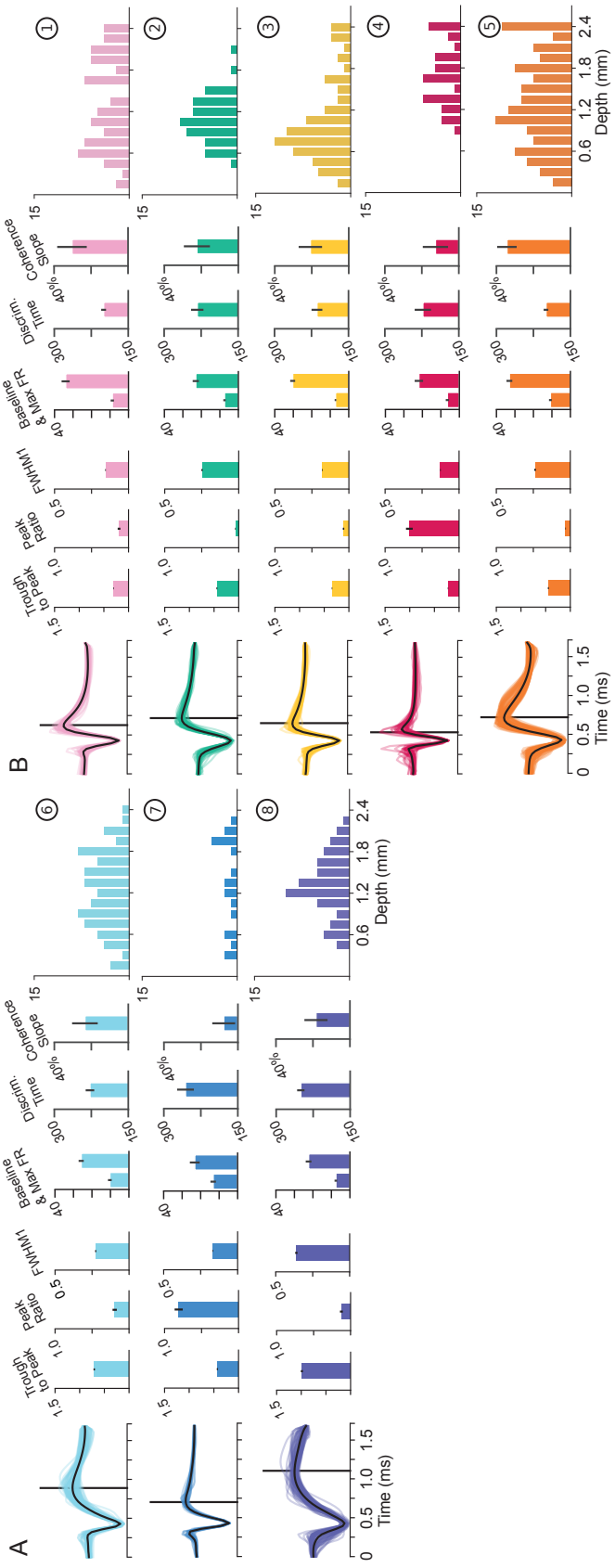

Figure S7: Detailed summary of each UMAP cluster and features: **(A, B)** A detailed summary of broad- (A) and narrow-spiking (B) cluster waveform shapes, physiological measures, and laminar distribution. Each waveform shape is shown at left with the average waveform shown as a black trace. The average post-hyperpolarization peak position is shown with a black line. The three waveform features used in the GMM classification (Fig. 3A) are shown in the middle as the mean ± S.E. The baseline and max FR for each cluster are subsequently shown in spikes/s (median ± bootstrap S.E.). Functional properties, discrimination time and coherence slope, are shown in milliseconds and spikes/s/s/% coherence (both shown in median ± bootstrap S.E.). Laminar distributions are also shown with each column in the histogram being the number of each waveform found at each channel location. Channels are spaced every 0.15 mm apart from 0.0 to 2.4 mm.

Function	Function Name	Parameters	Value
UMAP Algorithm (Python)	umap.UMAP	n_neighbors	30
		min_dist	0.1
		random_state	42
		metric	'euclidean'
Louvain Clustering (Python)	cylouvain.best_partition	resolution	1.5
UMAP Random Forest (Python)	xgboost.XGBClassifier	max_depth	4
		min_child_weight	2.5
		n_estimators	100
		learning_rate	0.3
		objective	'binary:logistic'
		rand_state	42
GMM Random Forest (Python)	xgboost.XGBClassifier	max_depth	10
		min_child_weight	2.5
		n_estimators	110
		learning_rate	0.05
		objective	'binary:logistic'
		seed	42
8-Class GMM Random Forest (Python)	xgboost.XGBClassifier	max_depth	4
		min_child_weight	2.5
		n_estimators	100
		learning_rate	0.3
		objective	'binary:logistic'
		seed	42
Gaussian Mixture Model (MATLAB)	fitgmdist	k	4
		start	'randsample'
		replicates	50
		statset('MaxIter')	200

Table S1: Non-default model hyperparameters used

Primary Antibody	Host	Dilution	Vendor	Catalog	Secondary Antibody Conjugate	Host-Antigen	Dilution	Vendor	Catalog
Calbindin D-28k (CB)	Rabbit	1:2000	Swant	CB38	Alexa 546	Donkey anti-Rabbit	1:200	ThermoFisher	A10040
Calretinin D-28k (CR)	Rabbit	1:2000	Swant	7697	Alexa 546	Donkey anti-Rabbit	1:200	ThermoFisher	A10040
Parvalbumin (PV)	Guinea Pig	1:2000	Swant	GP72	Alexa 488	Donkey anti-Guinea Pig	1:200	Jackson	706-545-148

Table S2: Primary and secondary antibodies used in immunohistochemical staining